

# MODULATING PROKARYOTIC LIFESTYLE BY DNA-BINDING PROTEINS

EDITED BY : Tatiana Venkova, Antonio Juarez and Manuel Espinosa  
PUBLISHED IN : Frontiers in Molecular Biosciences







# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2017 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-105-0

DOI 10.3389/978-2-88945-105-0

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)



# MODULATING PROKARYOTIC LIFESTYLE BY DNA-BINDING PROTEINS

Topic Editors:

**Tatiana Venkova**, University of Texas Medical Branch, USA

**Antonio Juarez**, Universidad de Barcelona, Spain

**Manuel Espinosa**, Centro de Investigaciones Biologicas, Spain



Grand Prismatic Spring - Yellowstone  
Courtesy of Stephen Kendall [www.purplekitephoto.com](http://www.purplekitephoto.com)

## The Overview of the Topic was the following:

“One of the most active areas of research in molecular microbiology has been the study of how bacteria modulate their genetic activity and its consequences. The prokaryotic world has received much interest not only because the resulting phenomena are important to cells, but also because many of the effects often can be readily measured. Contributing to the interest of the present topic is the fact that modulation of gene activity involves the sensing of intra- and inter-cellular conditions, DNA binding and DNA dynamics, and interaction with the replication/transcription machinery of the cell. All of these processes are fundamental to the operation of a genetic entity and condition their lifestyle. Further, the discoveries achieved in the bacterial world



have been of ample use in eukaryotes. In addition to the fundamental interest of understanding modulation of prokaryotic lifestyle by DNA-binding proteins, there is an added interest from the healthcare point of view. As it is well known the antibiotic-resistance strains of pathogenic bacteria are a major world problem, so that there is an urgent need of innovative technologies to tackle it. Most of the acquired resistances are spread by processes of horizontal gene transfer mediated by mobile elements in which DNA replication and gene expression are of basic interest. There is an imperative of finding new alternatives to the 'classical' way of treatment of bacterial infections and these new alternatives include the discovery of new drugs and of new bacterial targets. Nevertheless, these new alternatives will find a dead-end if we are unable to obtain a better understanding of the basic processes modulating bacterial gene expression. Our goal to achieve with this Topic of Frontiers is to accelerate our understanding of protein-DNA interactions. First, the topic will bring together several very active researchers in the study of gene replication, gene regulation, the strategies applied by the different proteins that participate in these processes, and their consequences. We will also acquire an in-depth knowledge of some of the mechanisms of gene regulation, gene transfer and gene replication. Further, the readers of the papers will realize the importance of the topic and will learn the most recent thinking, results, and approaches in the area".

We are fully confident that we have exceeded our expectations. Now we are proud to present the final output of the Topic, which is the eBook. It includes 24 articles contributed by 118 authors. As of today, Monday, 16th, January 2017, the total number of readings has reached 19,284, 14,921 article views, and 2,944 article downloads.

**Citation:** Venkova, T., Juarez, A., Espinosa, M., eds. (2017). *Modulating Prokaryotic Lifestyle by DNA-Binding Proteins*. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-105-0



# Table of Contents

**08 Editorial: Modulating Prokaryotic Lifestyle by DNA-Binding Proteins: Learning from (Apparently) Simple Systems**

Tatiana Venkova, Antonio Juárez and Manuel Espinosa

## **Chapter 1: Bacterial Replication and its control**

This Chapter compiles five articles dealing with DNA replication control in terms of timing (Riber et al., 2016), assembly of the machinery at the origin of replication (Wegrzyn et al., 2016), the mechanisms involved in the opening of the DNA strands, a fundamental fact in all forms of life (Jha et al., 2016), ingenious solutions to initiate DNA replication (Salas et al., 2016), and a new form of regulation of protein involved in replication by inorganic compounds (Ruiz-Masó et al., 2016).

**11 Opening the Strands of Replication Origins—Still an Open Question**

Jyoti K. Jha, Revathy Ramachandran and Dhruba K. Chattoraj

**24 Multiple DNA Binding Proteins Contribute to Timing of Chromosome Replication in *E. coli***

Leise Riber, Jakob Frimodt-Møller, Godefroid Charbon and Anders Løbner-Olesen

**33 Metal-Induced Stabilization and Activation of Plasmid Replication Initiator RepB**

José A. Ruiz-Masó, Lorena Bordanaba-Ruiseco, Marta Sanz, Margarita Menéndez and Gloria del Solar

**48 DNA-Binding Proteins Essential for Protein-Primed Bacteriophage  $\Phi$ 29 DNA Replication**

Margarita Salas, Isabel Holguera, Modesto Redrejo-Rodríguez and Miguel de Vega

**69 Replisome Assembly at Bacterial Chromosomes and Iteron Plasmids**

Katarzyna E. Wegrzyn, Marta Gross, Urszula Uciechowska and Igor Konieczny

## **Chapter 2: Partition of genetic information into daughter cells**

This Chapter includes two very relevant reviews written by a well-known scientist in the field (Funnell, 2016), as well as another by a young investigator, both exploring the assembly of the nucleoprotein complex known as segrosome during DNA segregation during cell division (Oliva, 2016).

**86 ParB Partition Proteins: Complex Formation and Spreading at Bacterial and Plasmid Centromeres**

Barbara E. Funnell



### Chapter 3: Gene-expression control at the transcriptional level

Three important articles have been included within this Chapter, dealing with interaction of the bacterial RNA Polymerase with its DNA targets and their possible role in drug design (Lee and Borukhov, 2016), sensing of the environmental signals (Hartman et al., 2016), and a very important paper on genome-wide regulation via protein–DNA interactions at a distance (Qian et al., 2016).

**101 *Modulating Salmonella Typhimurium’s Response to a Changing Environment through Bacterial Enhancer-Binding Proteins and the RpoN Regulon***

Christine E. Hartman, David J. Samuels and Anna C. Karls

**110 *Bacterial RNA Polymerase-DNA Interaction—The Driving Force of Gene Expression and the Target for Drug Action***

Jookyung Lee and Sergei Borukhov

**125 *Genome-Wide Transcriptional Regulation and Chromosome Structural Arrangement by GalR in E. coli***

Zhong Qian, Andrei Trostel, Dale E. A. Lewis, Sang Jun Lee, Ximiao He, Anne M. Stringer, Joseph T. Wade, Thomas D. Schneider, Tim Durfee and Sankar Adhya

### Chapter 4: Global regulatory networks

Bacteria control the expression of their genes by regulatory networks that participate in their lifestyle. This subject has been tackled here by six important papers dealing with the global regulation (Brandi et al., 2016; Cech et al., 2016; Erill et al., 2016; Solano-Collado et al., 2016), post-transcriptional control of gene expression (Amin et al., 2016), and importantly the control of the expression of bacterial operons encoding toxin antitoxin genes, a very risky business in the bacterial survival when encountering stress situations (Chan et al., 2016).

**136 *Post-translational Serine/Threonine Phosphorylation and Lysine Acetylation: A Novel Regulatory Aspect of the Global Nitrogen Response Regulator GlnR in S. coelicolor M145***

Rafat Amin, Mirita Franz-Wachtel, Yvonne Tiffert, Martin Heberer, Mohamed Meky, Yousra Ahmed, Arne Matthews, Sergii Krysenko, Marco Jakobi, Markus Hinder, Jane Moore, Nicole Okoniewski, Boris Maček, Wolfgang Wohlleben and Agnieszka Bera

**150 *The Escherichia Coli Hfq Protein: An Unattended DNA-Transactions Regulator***

Grzegorz M. Cech, Agnieszka Szalewska-Pałasz, Krzysztof Kubiak, Antoine Malabirade, Wilfried Grange, Veronique Arluison and Grzegorz Węgrzyn

**156 *Keeping the Wolves at Bay: Antitoxins of Prokaryotic Type II Toxin-Antitoxin Systems***

Wai Ting Chan, Manuel Espinosa and Chew Chieng Yeo

**176 *The Verrucomicrobia LexA-Binding Motif: Insights into the Evolutionary Dynamics of the SOS Response***

Ivan Erill, Susana Campoy, Sefa Kılıç and Jordi Barbé



**189 *MgaSpn and H-NS: Two Unrelated Global Regulators with Similar DNA-Binding Properties***

Virtu Solano-Collado, Mário Hüttener, Manuel Espinosa, Antonio Juárez and Alicia Bravo

**201 *An Interplay among FIS, H-NS, and Guanosine Tetraphosphate Modulates Transcription of the Escherichia coli cspA Gene under Physiological Growth Conditions***

Anna Brandi, Mara Giangrossi, Anna M. Giuliodori and Maurizio Falconi

## **Chapter 5: Horizontal gene transfer**

The Chapter includes four major contributions dealing with bacterial conjugation studied under different points of view, but all of them relying on how horizontal gene transfer, that leads to the spread of antimicrobial resistance can be studied in-depth with the focus on basic protein-DNA transactions during transfer (Grohmann et al., 2016; Gruber et al., 2016), cross-talks between conjugative relaxases and their target DNAs (Fernández-González et al., 2016) as well as by the use of comparative genomics to develop models to framework the role of plasmids in the bacterial world (Fernandez-Lopez et al., 2016).

**213 *DNA-Binding Proteins Regulating pIP501 Transfer and Replication***

Elisabeth Grohmann, Nikolaus Goessweiner-Mohr and Sabine Brantl

**229 *Conjugative DNA Transfer Is Enhanced by Plasmid R1 Partitioning Proteins***

Christian J. Gruber, Silvia Lang, Vinod K. H. Rajendra, Monika Nuk, Sandra Raffl, Joel F. Schildbach and Ellen L. Zechner

**246 *A Functional oriT in the Ptw Plasmid of Burkholderia cenocepacia Can Be Recognized by the R388 Relaxase TrwC***

Esther Fernández-González, Sawsane Bakioui, Margarida C. Gomes, David O'Callaghan, Annette C. Vergunst, Félix J. Sangari and Matxalen Llosa

**257 *Comparative Genomics of the Conjugation Region of F-like Plasmids: Five Shades of F***

Raul Fernandez-Lopez, Maria de Toro, Gabriel Moncalian, M. Pilar Garcillan-Barcia and Fernando de la Cruz

## **Chapter 6: Bacteria-host relationships**

Extensive research on two related human bacterial pathogens such as *Salmonella* (Lobato-Márquez et al., 2016), and *Shigella* (Di Martino et al., 2016) are covered by these two leading laboratories in the field, providing new insights into the subject.

**271 *Stabilization of the Virulence Plasmid pSLT of Salmonella Typhimurium by Three Maintenance Systems and Its Evaluation by Using a New Stability Test***

Damián Lobato-Márquez, Laura Molina-García, Inma Moreno-Córdoba, Francisco García-del Portillo and Ramón Díaz-Orejas

**282 *The Multifaceted Activity of the VirF Regulatory Protein in the Shigella Lifestyle***

Maria Letizia Di Martino, Maurizio Falconi, Gioacchino Micheli, Bianca Colonna and Gianni Prosseda



## Chapter 7: Hot out of the oven

This Chapter includes two very relevant review contributions by leaders in their respective fields: the membrane world of Gram-positive bacteria (Albanesi and de Mendoza, 2016), and a theoretical and extremely interesting view on the CRISPR role of plasmid and bacteriophage copy number in the process (Severinov et al., 2016).

**293    *FapR: From Control of Membrane Lipid Homeostasis to a Biotechnological Tool***

Daniela Albanesi and Diego de Mendoza

**302    *The Influence of Copy-Number of Targeted Extrachromosomal Genetic Elements on the Outcome of CRISPR-Cas Defense***

Konstantin Severinov, Iaroslav Ispolatov and Ekaterina Semenova



# Editorial: Modulating Prokaryotic Lifestyle by DNA-Binding Proteins: Learning from (Apparently) Simple Systems

Tatiana Venkova<sup>1</sup>, Antonio Juárez<sup>2,3</sup> and Manuel Espinosa<sup>4\*</sup>

<sup>1</sup> University of Texas Medical Branch (UTMB), Galveston, TX, USA, <sup>2</sup> Departamento de Genética, Microbiología y Estadística, Universidad de Barcelona, Barcelona, Spain, <sup>3</sup> Institut de Bioenginyeria de Catalunya (IBEC), Barcelona, Spain, <sup>4</sup> Centro de Investigaciones Biológicas (CSIC), Madrid, Spain

**Keywords:** DNA-protein interactions, gene regulation in prokaryotes, replication control, regulation of bacterial gene expression, global regulatory networks

## Editorial on the Research Topic

### Modulating Prokaryotic Lifestyle by DNA-Binding Proteins: Learning from (Apparently) Simple Systems

## OPEN ACCESS

### Edited by:

Alexey V. Onufriev,  
Virginia Tech, USA

### Reviewed by:

Saeed Izadi,  
Genentech, USA

### \*Correspondence:

Manuel Espinosa  
mespinosa@cib.csic.es

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 02 November 2016

**Accepted:** 19 December 2016

**Published:** 06 January 2017

### Citation:

Venkova T, Juárez A and Espinosa M  
(2017) Editorial: Modulating  
Prokaryotic Lifestyle by DNA-Binding  
Proteins: Learning from (Apparently)  
Simple Systems.  
Front. Mol. Biosci. 3:86.  
doi: 10.3389/fmolb.2016.00086

Within the research in Molecular Biology, one important field along the years has been the analyses on how prokaryotes regulate the expression of their genes and what the consequences of these activities are. Prokaryotes have attracted the interests of researchers not only because the processes taking place in their world are important to cells, but also because many of the effects often can be readily measured, both at the single cell level and in large populations. Contributing to the interest of the present topic is the fact that modulation of gene activity involves the sensing of intra- and inter-cellular conditions, DNA binding and DNA dynamics, and interaction with the replication/transcription machinery of the cell. All of these processes are fundamental to the operation of a biological entity and they condition its lifestyle. Further, the discoveries achieved in the bacterial world have been of ample use in eukaryotes. In addition to the fundamental interest of understanding modulation of prokaryotic lifestyle by DNA-binding proteins, there is an added interest from the healthcare point of view. As it is well-known the antibiotic-resistance strains of pathogenic bacteria are a major world problem, so that there is an urgent need of innovative approaches to tackle it. Human and animal infectious diseases impose staggering costs worldwide in terms of loss of human life and livestock, diminished productivity, and the heavy economic burden of disease. The global dimension of international trade, personal travel, and population migration expands at an ever-accelerating rate. This increasing mobility results in broader and quicker dissemination of bacterial pathogens and in rapid spread of antibiotic resistance. The majority of the newly acquired resistances are horizontally spread among bacteria of the same or different species by processes of lateral (horizontal) gene transfer, so that discovery of new antibiotics is not the definitive solution to fighting infectious diseases. There is an absolute need of finding novel alternatives to the “classical” approach to treat infections by bacterial pathogens, and these new ways must include the exploration and introduction of novel antibacterials, the development of alternative strategies, and the finding of novel bacterial targets. However, all these approaches will result in a stalemate if we, researchers, are not able to achieve a better understanding of the mechanistic processes underlying bacterial gene expression. It is, then, imperative to continue



gaining insight into the basic mechanisms by which bacterial cells regulate the expression of their genes. That is why our Research Topic hosted by Frontiers in Molecular Biosciences was timely, and the output of it offers novel and up-to-date points of view to the “simple” bacterial world.

## THE RESEARCH TOPIC

From the beginning of announcing the present Topic, we stated that “our goal to achieve with this Topic of Frontiers was to accelerate and to broaden our understanding of gene regulation in prokaryotes through knowledge on protein–DNA interactions.” Thus, it was our aim that the topic would provide us with the opportunity of bringing together a number of very active researchers that were dealing with DNA replication, genetic regulation, as well as the approaches taken by the various proteins that play a role in these processes and their consequences. The papers compiled in this Research Topic will allow us to gain an up-to-date knowledge of some of the mechanisms of genetic regulation, DNA transfer, and DNA replication as well. Further, the readers will understand the influence of the topic on the scientific world, and will learn the most contemporary reasoning, results, and experimental approaches in the area. The authors of the articles which compose the Research Topic represent some of the top laboratories in the world actively working on the Topic of DNA-binding proteins from prokaryotic origin. We are aware that the list of contributors is far from complete. But logical limitations must be introduced. The present Research Topic brings together important articles dealing with the study of gene regulation and the mechanisms underlying, including replication, transfer, segregation, and transcriptional and post-transcriptional control of gene expression as well as global regulatory networks in prokaryotes. Finally, we envisage that a large number of people will notice the topic subject and will thereby come to realize that it is of great importance, broadening the expected impact. Hence, it seems important to reunite, in a single and intense Research Topic the articles written by researchers that give the latest information from the various fields of structural analysis with those who provide genetic, biochemical, and biophysical information.

We have gathered a total of 24 articles written by a total of 114 authors, thus we believe the joint contributions has reached a high level of contributors that have achieved a high scientific standard that will be difficult to overcome in future attempts in this particular field. We are pleased to present articles covering not only basic research, but also applied science. The readers will be able to look at the bacterial replication control in terms of time (Riber et al.), mechanistic concepts (Węgrzyn et al.), novel ways to achieve DNA replication (Salas et al.), basic principles of DNA strands’ opening (Jha et al.), and a novel regulation of a replication initiator protein by inorganic compounds (Ruiz-Masó et al.). Segregation and partition of DNA copies to daughter cells (Oliva; Funnell et al.) will help the readers to understand the basis of this fundamental process: how cells receive copies of DNA lies at the basis of bacterial

life. Control of bacterial gene expression has been tackled at different levels. Firstly, gene-expression control at the level of transcription covers various aspects including RNA Polymerase interactions (Lee and Borukhov), environmental signal sensing (Hartman et al.), and genome-wide indirect regulation via protein–DNA interactions at a distance causing large-scale restructuring of the chromosome (Qian et al.). Secondly, global regulatory networks (Solano-Collado et al.; Erill et al.; Cech et al.; Brandi et al.) as well as post-transcriptional control (Amin et al.) that lead to intra-chromosomal communications and structural arrangement by DNA-interacting proteins have been a subject of very active research. And thirdly, the role of bacterial operons encoding toxin-antitoxin genes, and how they regulate their expression has been reviewed from the structural and functional points of view by Chan et al. Deeper understanding of general biological processes like replication and transcription in bacteria favors the quest for novel anti-bacterial agents, a life-saving area of research that is discussed by the Lee and Borukhov. The bacterial conjugation studied as a major pathway for horizontal gene transfer that leads to the spread of antimicrobial resistance, has been approached by Grohmann et al., Gruber et al., Fernández-González et al., and discussed in detail by Fernandez-Lopez et al. On the other hand, extensive research on human pathogens as *Salmonella* (Lobato-Márquez et al.), *Shigella* (Letizia Di Martino et al.), and Gram-positive bacteria (Albanesi and de Mendoza) provides not only better understanding of their molecular genetics, but also could become novel diagnostics, anti-virulence, and biotechnological tools. The relatively recently discovered bacterial immune system CRISPR, applied in eukaryotic genomic engineering and soon in human medicine, is also tackled in a theory article presenting a mathematical model on the role of plasmid/phage copy number in the process (Severinov et al.).

## WHERE DO WE GO FROM HERE?

We have arrived in the post genomic era in a somewhat embarrassing position. We now know the genomic content of a large number of organisms and the gene expression patterns in many of them. We indeed are in the situation of trying to handle big data buried within the many little crevices and nooks of the Internet resources. However, for the minority of gene products whose cellular function is presently known, only occasionally we know how the protein product of a gene actually works or how it interacts with other components in the cell. This situation is equivalent to being able to understand Morse code, but not yet understanding the language of the person sending us the telegraphic message. We should aim to a comprehensive view of regulatory systems which for the most part are not yet fully understood. In addition, we should learn more from the mechanisms of action of some of proteins that play central roles in global regulation of cellular function. Prokaryotic systems need to be emphasized because in general these systems are considerably better understood than their eukaryotic counterparts are. This is because work with prokaryotes allows highly sophisticated genetic analyses,

complex and sensitive physiological measurements, and detailed biochemical and biophysical studies of the components of regulatory systems. It is with prokaryotic systems that the goal, now distantly visible on the horizon, of achieving a basic understanding of all the biological processes within a cell and their interactions, will first be achieved. Novel search engines must be developed, so that we can use the many data that are nowadays at our disposal, but we must find ways and procedures on how to dig them out, how to order them and clarify their meaning and how to be able to develop models in which we can express our views of the prokaryotic world. Sharing this information among scientists that try to understand the complexity of the prokaryotic world should constitute the next generation of knowledge and toward the today science, in our opinion, should devote the next coming years.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

While this Editorial was written, authors were funded by Spanish MINECO, grants CSD2008-00013, BIO2013-49148-C2-2-R, BIO2013-49148-C2-1-R, BIO2015-69085-REDC, BIO2016-76412-C2-2-R-AEI/FEDER, UE, and BIO2016-76412-C2-1-R-AEI/FEDER, UE.

## ACKNOWLEDGMENTS

Thanks are due to the contributors to this Research Topic as well as the Editorial support of the Journal.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2017 Venkova, Juárez and Espinosa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*





# Opening the Strands of Replication Origins—Still an Open Question

Jyoti K. Jha, Revathy Ramachandran and Dhruva K. Chattoraj\*

Laboratory of Biochemistry and Molecular Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas Medical Branch,  
USA

### Reviewed by:

Grzegorz Węgrzyn,  
University of Gdańsk, Poland  
Kurt Henry Piepenbrink,  
The University of Maryland School of  
Medicine, USA  
Rafael Giraldo,  
Spanish National Research Council,  
Spain  
Julia Grimwade,  
Florida Institute of Technology, USA  
Anders Løbner-Olesen,  
University of Copenhagen, Denmark

### \*Correspondence:

Dhruva K. Chattoraj  
chattoraj@nih.gov

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 30 July 2016

**Accepted:** 16 September 2016

**Published:** 30 September 2016

### Citation:

Jha JK, Ramachandran R and  
Chattoraj DK (2016) Opening the  
Strands of Replication Origins—Still an  
Open Question.  
Front. Mol. Biosci. 3:62.  
doi: 10.3389/fmolb.2016.00062

The local separation of duplex DNA strands (strand opening) is necessary for initiating basic transactions on DNA such as transcription, replication, and homologous recombination. Strand opening is commonly a stage at which these processes are regulated. Many different mechanisms are used to open the DNA duplex, the details of which are of great current interest. In this review, we focus on a few well-studied cases of DNA replication origin opening in bacteria. In particular, we discuss the opening of origins that support the theta ( $\theta$ ) mode of replication, which is used by all chromosomal origins and many extra-chromosomal elements such as plasmids and phages. Although the details of opening can vary among different origins, a common theme is binding of the initiator to multiple sites at the origin, causing stress that opens an adjacent and intrinsically unstable A+T rich region. The initiator stabilizes the opening by capturing one of the open strands. How the initiator binding energy is harnessed for strand opening remains to be understood.

**Keywords:** replication origins, DNA melting, bacterial origins, lambda origin, plasmid origins

## INTRODUCTION

A remarkable feature of double stranded DNA (dsDNA) is its ability to undergo denaturation, whereby its strands can be completely separated into single strands, and renaturation, whereby the two complementary strands can be annealed back to form dsDNA. *In vitro*, DNA can undergo denaturation or renaturation simply in response to a change in salt concentration, temperature, pH or the presence of mild reagents such as formamide (Inman, 1966; Westmoreland et al., 1969). The reversibility of strand separation is the basis of hybridization techniques such as Southern blotting and PCR.

Strand opening usually refers to situations where the stability of duplex DNA is altered locally and for a limited period by DNA binding proteins. Complementary strands of DNA are most stable in the double helical B-form as modeled by Watson and Crick. Opening of the strands is thus energetically unfavorable. Active processes are involved in making the opening site-specific, and of significant length and duration so that the downstream events become feasible. In the case of replication initiation, the immediate downstream event is the loading of the replicative helicase. The helicase enlarges the opening and mediates loading of the primase and the replisome machinery that are required for duplicating the DNA (Bell and Kaguni, 2013).

Among different origins, the structure and the process of strand opening vary significantly, but there are several commonalities (Bramhill and Kornberg, 1988b; **Figure 1**). Common elements include: (1) The presence of multiple initiator protein binding sites (9-mers) within the origin. The binding of the initiator allows site-specific opening, which enables helicase loading. (2) The presence of A+T-rich DNA sequences (13-mers) within the origin where the opening

initiates. A stretch of ~20 bp A+T-rich region (called a DNA unwinding element, or DUE) is common within replication origins, most likely due to the fact that A+T-rich regions are easier to melt than G+C rich sequences (Inman, 1966; Kowalski and Eddy, 1989). (3) Remodeling (bending/folding/stretching) of origin DNA upon initiator binding, which is often facilitated by additional binding of nucleoid associated proteins (NAPs, e.g., HU; Stenzel et al., 1987; Hwang and Kornberg, 1992a; Dorman, 2009). (4) A requirement for the DNA to be negatively supercoiled, which is an under-wound and unstable state, that can make the DNA prone to opening but not open enough for helicase loading (Bramhill and Kornberg, 1988a). (5) Opening at DUE resulting from its intrinsic instability, and stress from DNA remodeling and negative supercoiling (Bowater et al., 1991). (6) Stabilization of the open state by the single stranded DNA (ssDNA) binding activity of the initiator, which captures one specific single strand of the open DNA so that the other is available for helicase loading. It is worth emphasizing that *in vivo* the aggregate of the A+T rich DUE, NAPs and negative supercoiling are not enough, and that the initiator binding to the origin provides an essential contribution to the energetics of opening. Additional regulatory factors are usually involved to modulate the frequency and timing of opening. Below we elaborate on the core features of opening for a few specific origins.

## OPENING OF AN AAA+ PROTEIN CONTROLLED ORIGIN, *ORIC*, OF *ESCHERICHIA COLI*

The opening of the *E. coli* origin, *oriC*, has been studied in the most depth. The opening was demonstrated *in vitro* at a time when DNA replication could be separated into discrete stages, with each step dependent on the previous one: initiator binding to the origin, strand opening at DUE, loading of the helicase, and finally, loading of the primase and the rest of the replisome (Figure 1; Bramhill and Kornberg, 1988b; O'Donnell, 2006). The ability to delineate the replication initiation process into discrete stages revealed that origin opening is not only a critical first step, but also a key replicon-specific event, as the players in subsequent steps seem common to all replicons.

Decades of genetic, biochemical and structural studies have generated a wealth of information on the structure-function relationship of the *E. coli* initiator, DnaA. DnaA is a highly conserved initiator protein in bacteria with structural similarity to initiators in the other domains of life (Giraldo, 2003). DnaA belongs to the AAA+ superfamily of ATPases (Neuwald et al., 1999) and has four domains (Ozaki and Katayama, 2009; Figure 2A): An N-terminal domain for homo-oligomerization and interactions with other replication related proteins, a non-conserved linker domain between the N-terminal domain and the large AAA+ domain for binding and hydrolyzing ATP, and a C-terminal DnaA binding domain (DBD) containing a helix-turn-helix (HTH) motif and a proximal basic loop for specific binding to dsDNA (Erzberger et al., 2002). The AAA+ domain mediates ATP dependent DnaA oligomerization that is independent of the N-terminal domain, which allows the AAA+

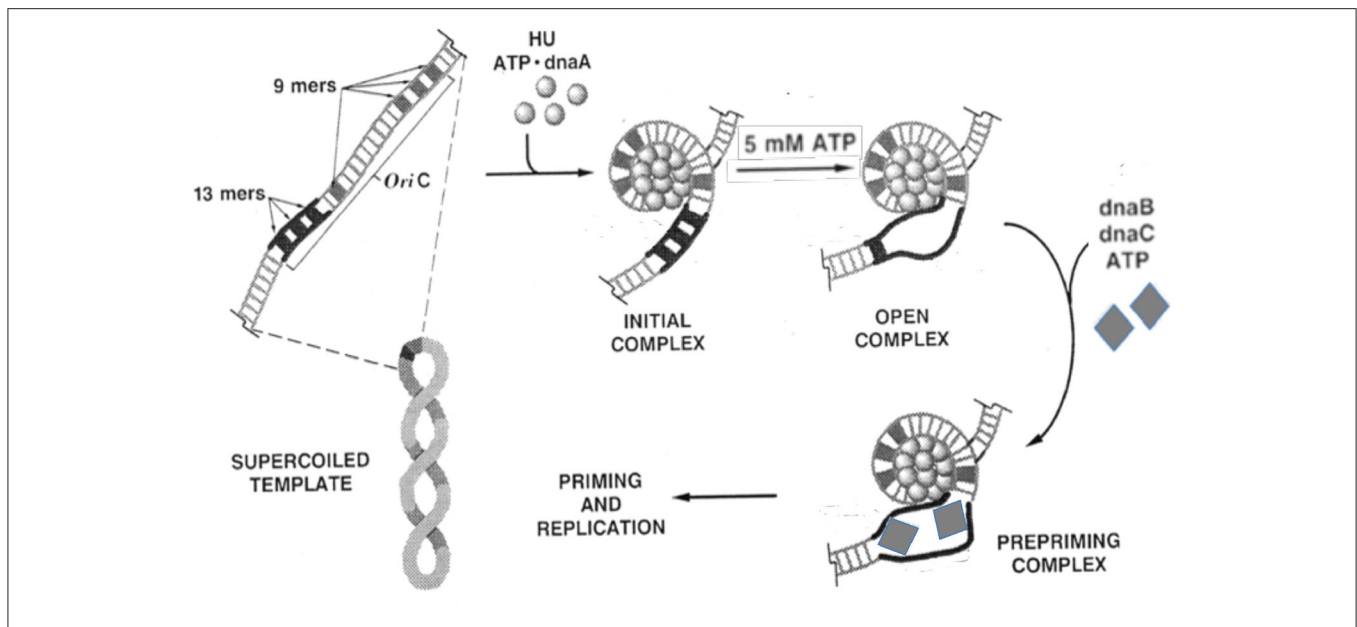
domain to bind to ssDNA (Duderstadt et al., 2011). DnaA thus uses two different domains to bind to ds- and ss-DNA. DnaA has several binding sites in *oriC*. The organization of the sites and their interactions with DnaA are complex (Leonard and Grimwade, 2010). Models have been proposed to explain how these interactions may give rise to strand opening, as we discuss below.

## Formation of Nucleoprotein Complexes at *oriC* of *E. coli*

DnaA binds through its C-terminal HTH motif to *dsoriC* at eleven 9-mer sites (Figure 2B). The three high affinity sites (R1, R2, and R4, Kd < 20 nM) remain bound throughout the cell cycle, and have equal affinity for DnaA-ATP and DnaA-ADP (Nievera et al., 2006). Binding to the remaining sites requires cooperative interactions with DnaA bound to the high affinity sites, with most requiring higher concentrations of DnaA-ATP. Binding to these weaker sites is cell-cycle specific and peaks immediately before the time of initiation, when the DnaA-ATP concentration reaches a maximum (Kurokawa et al., 1999; Nievera et al., 2006).

Two NAPs, Fis and IHF, regulate the timing of DNA-ATP binding. When bound to *oriC*, Fis inhibits saturation of DnaA-ATP binding to the weaker sites. Upon release of Fis, IHF binding facilitates saturation of binding (Ryan et al., 2004). These studies indicate that saturation of binding is a highly regulated process in the cell cycle, and is achieved by controlling the DnaA-ATP concentration. The increase in the DnaA-ATP concentration promotes oligomerization of DnaA-ATP from R4 to C3, which is believed to cause dissociation of Fis from its site that overlaps C3 (Rozgaja et al., 2011). Fis dissociation is believed to remove the steric barrier to IHF association, although the exact mechanism remains to be determined (Kaur et al., 2014; Leonard and Grimwade, 2015). The involvement of NAPs suggests the presence of long range interactions in the formation of nucleoprotein complexes at *oriC*. The importance of the relative distances between DnaA binding sites and their helical phasing is also suggestive of higher order structure formation (Woelker and Messer, 1993). Neither Fis nor IHF are essential *in vivo* or for replication *in vitro*; they are, however, required for regulating replication initiation in the cell cycle (Ryan et al., 2004). IHF can efficiently substitute for HU *in vitro*, indicating redundancy in NAPs requirement *in vivo* (Hwang and Kornberg, 1992a).

In addition to the eleven 9-mer sites, *oriC* contains three repeating 13-mer sequences that comprise the DUE, to which DnaA-ATP binds (Figures 1, 2; Speck and Messer, 2001). The 13-mers have adenine methylation sites, which, when methylated, are expected to favor strand separation (Gotoh and Tagashira, 1981). DnaA binding to DUE most likely requires the DUE to be single stranded, although initial binding may occur on dsDUE (Figure 3A; Duderstadt et al., 2011). The binding is mediated through the AAA+ domain of DnaA oligomers (Ozaki et al., 2008; Duderstadt et al., 2011). These details were obtained from X-ray crystallographic structures of N-terminal deleted DnaA from thermophilic bacteria (Erzberger et al., 2002; Ozaki et al., 2008). The DNA-protein and protein-protein contacts seen in



**FIGURE 1 | Biochemical steps leading to initiation of replication from the *E. coli* origin, *oriC*.** The figure is adapted from (Bramhill and Kornberg, 1988a) with permission from the publisher. The initiator DnaA initially binds to the 9-mers (R sites), and then saturates the origin in the presence of ATP and HU. Under favorable reaction conditions (high ATP, negative supercoiling, 38°C, appropriate concentration of HU), the region with three 13-mers (DUE) opens. The opening allows loading of the DnaB-DnaC complex (the helicase and the helicase loader), and subsequently the rest of the replisome to initiate and complete duplication of the minichromosome. Note that (i) the entire process can be achieved with purified components (Kaguni and Kornberg, 1984; Bramhill and Kornberg, 1988a); (ii) the top strand of the open region contacts the DnaA-ATP bundle, which stabilizes the open complex; (iii) DnaA-ATP binding to *oriC* can be achieved at <1  $\mu$ M concentrations of ATP, whereas the opening requires mM concentrations of ATP, most likely for reasons other than DNA binding (Saxena et al., 2015); (iv) The DNA spiral ("writhing") shown surrounding the DnaA-ATP bundle is left-handed. Current evidence suggests it to be right-handed, which is more logical in terms of opening (Erzberger et al., 2006); and (v) HU stimulates opening, and helps to localize it to 13-mers, but is not essential (Bramhill and Kornberg, 1988a).

crystals are also functionally significant in *E. coli* (Ozaki et al., 2008; Duderstadt et al., 2011).

Several important key findings have emerged from structural studies. Whereas DnaA-ADP is monomeric, DnaA-ATP is oligomeric (Erzberger et al., 2002, 2006; Ozaki et al., 2008; Ozaki and Katayama, 2012). The oligomerization is dependent on ATP, which bridges neighboring DnaA protomers at the interface between neighboring subunits by making contact with the Walker A and B motifs of one subunit and a conserved arginine ("arginine finger") of the neighboring subunit through its  $\gamma$ -phosphate. The involvement of the  $\gamma$ -phosphate explains why DnaA-ADP fails to oligomerize. Mutating the arginine finger abolishes ATP-dependent binding of DnaA to *oriC* and initiation activity. Thus, oligomerization appears to be the mechanism to allow sequential binding to weak dsDNA sites and to the DUE (Cheng et al., 2015).

## Models for Opening at *oriC*

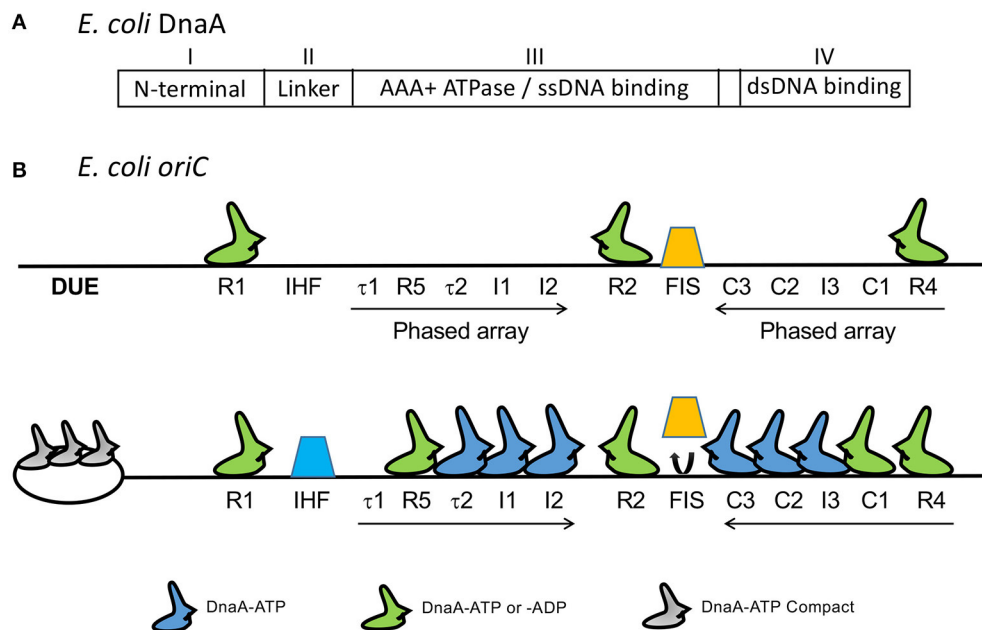
DnaA-ATP in the crystal and in solution is a polymeric, right-handed spiral filament, which affords a ready explanation for how it could facilitate opening: Wrapping of DNA around a right-handed spiral is the same as introducing positive supercoils that could spontaneously induce compensatory negative supercoils in the adjoining DNA (Erzberger et al., 2006; Zorman et al., 2012). Although the negative supercoils can diffuse out of the origin, their proximity to wrapped DNA and propensity for melting

render them more likely to be absorbed by unwinding of the DUE (Bowater et al., 1991; Polaczek et al., 1998). In this scenario, the wrapping of dsDNA around DnaA not only generates the unwinding force but also helps to confine the unwinding within the DUE. A stronger barrier to supercoil diffusion out of the A+T-rich DUE is suggested by the finding that the open state of DUE is quite stable in the absence of helicase loading *in vitro* and *in vivo* (Odegrip et al., 2000). Capturing one of the single strands by the AAA+ domain of DnaA oligomers as found in co-crystals could be a straightforward way to retain the DUE in the open state (Duderstadt et al., 2011).

DnaA may also directly open dsDUE (Figure 3A). This model is based on the structural similarities of ssDNA in complex with DnaA-ATP or RecA-ATP, and the biochemical evidence that DnaA can unwind short stretches of dsDNA (Duderstadt et al., 2011). RecA can transfer a single strand to homologous dsDNA (Shibata et al., 2001). Although cocrystals of DnaA-ATP with dsDNA are yet to be obtained, such structures were obtained with the archaeal initiator, Cdc6/Orc1, which is an AAA+ protein with significant homology to DnaA (Giraldo, 2003). The archaeal initiator was found to distort dsDNA, and DnaA also bends DNA upon binding (Schaper and Messer, 1995). Thus, similar to RecA, DnaA oligomers may initially contact dsDUE and distort the region enough to initiate ssDUE binding.

Structural studies indicate two distinct states of DnaA-ATP for ds- and ss-DNA binding. For contact with dsDNA, the





**FIGURE 2 | (A)** The four domains of DnaA. **(B)** Map of *oriC* showing several DnaA binding sites that were identified subsequent to the 9-mers shown in **Figure 1**. The figure is adapted from (Rozgaja et al., 2011) with permission from the publisher. Three of the 9-mers are high-affinity R sites (R1, R2, and R4) that bind DnaA-ADP and DnaA-ATP with nearly equal affinity and they remain bound throughout the cell cycle. The remaining sites R5,  $\tau$ 1,  $\tau$ 2, I1-I3, and C1-C3 are present in two phased arrays and preferentially or exclusively bind to DnaA-ATP, except for R5 which binds both DnaA-ADP and DnaA-ATP, as other R sites do (McGarry et al., 2004; Kawakami et al., 2005; Rozgaja et al., 2011). [Note that DnaA binding site R3 is not shown. The identification of R3 could have been a misinterpretation of binding to C2 and C3, which overlap with R3 (Rozgaja et al., 2011)]. Upon accumulation of DnaA-ATP, the R1 and R4 sites nucleate sequential binding of DnaA to the two arrays. The DnaA oligomer extension could displace Fis (orange trapezoid), possibly removing some steric hindrance to IHF binding (blue trapezoid; Kaur et al., 2014).

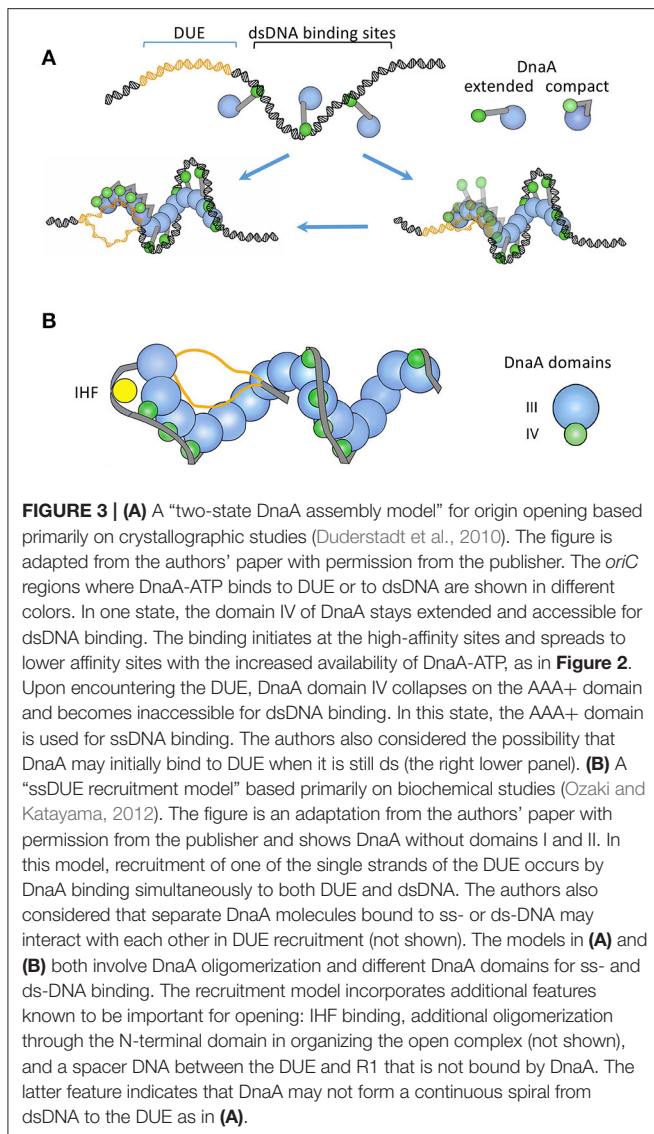
C-terminal HTH domains of DnaA oligomers stick out of the spiral and are free to contact dsDNA as it wraps around the spiral from the outside (**Figure 3A**). For contact with ssDNA, the HTH domain collapses on the AAA+ domain of the partner protomer and can no longer contact dsDNA. The interaction between the collapsed HTH domain and the AAA+ domain is required for oligomerization-mediated ssDNA binding, origin opening, and initiation *in vivo* (Duderstadt et al., 2010). In other words, the HTH domain also contributes to DnaA oligomerization. What triggers the HTH domain to change its conformation from an extended to a collapsed state in DnaA oligomers is not understood. Another study suggested that ds- and ss-DNA binding can occur simultaneously on the same DnaA-ATP oligomer (Ozaki and Katayama, 2012; **Figure 3B**). When DUE sequences were provided as single-stranded oligos together with a DUE-deleted *dsoriC* fragment, the oligos could contact specific pore residues of the DnaA-ATP spiral. Mutating the contacting residues (V211A and R245A) prevented DUE binding and opening. Although it remains to be resolved whether the ds- and ss-DNA binding occur with separate or the same DnaA molecules, it is clear that DnaA oligomerization is important for origin opening and ssDNA binding. The importance of DnaA oligomerization has also been demonstrated in *Bacillus subtilis* (Scholefield et al., 2012).

Weak DnaA binding sites are clustered into two phased arrays that are oriented opposite to each other (Rozgaja et al., 2011; **Figure 2B**). The wrapping model appears inconsistent with this

finding, because the handedness of wrapping is expected to be opposite for the two arrays and the torsional stress generated by wrapping of one array would be neutralized by wrapping of the other. However, the contribution of the two arrays to stress may not be equivalent. The DUE proximal array may be more important and can suffice for the opening. In fact, the deletion of the DUE distal array from *oriC* does not affect viability, and can achieve DUE opening, ssDNA binding and some DnaB loading *in vitro* (Stepankiw et al., 2009; Ozaki and Katayama, 2012). The DUE distal array becomes crucial during rapid growth and is required for enhancing helicase loading both *in vitro* and *in vivo* (Weigel et al., 2001; Stepankiw et al., 2009). Some elements of *oriC* might not be essential but they are there for improving its efficiency.

## Regulation of *oriC* Opening

So far, we have discussed the importance of DnaA-ATP in regulating the opening, and the involvement of NAPs in this process. There are also other regulatory proteins that influence the opening by controlling DnaA interactions with *oriC*. Many of these regulators interact with the N-terminal domain of DnaA and modulate its oligomerization activity. Proteins HU and DiaA promote oligomerization and unwinding by DnaA (Hwang and Kornberg, 1992a; Chodavarapu et al., 2008a; Keyamura et al., 2009). There are also N-terminal domain binding proteins L2 and Dps that impede oligomerization and origin opening (Chodavarapu et al., 2008b, 2011). These



N-terminal domain activities help in timing replication during the cell cycle and in maintaining replication synchrony during rapid growth, but are not essential for origin opening. Indeed, several studies have concluded that the essential role of the N-terminal domain is in the loading of the helicase (Sutton et al., 1998; Sharma et al., 2001; Speck and Messer, 2001; Simmons et al., 2003). However, DnaA cannot be loaded to low-affinity sites without an intact N-terminal domain, which would imply an essential role of the domain in opening (Miller et al., 2009). These apparent contradictions highlight the importance of clarifying the role of the N-terminal domain-mediated oligomerization.

There are also regulators that can indirectly control the opening of the DUE. Some regulators, such as SeqA and IciA, bind directly to the DUE and prevent opening by interfering with DnaA binding (Hwang and Kornberg, 1992b; Lu et al., 1994). SeqA also prevents DnaA binding to some of the low affinity sites that have overlapping SeqA binding sites (Nievera et al., 2006).

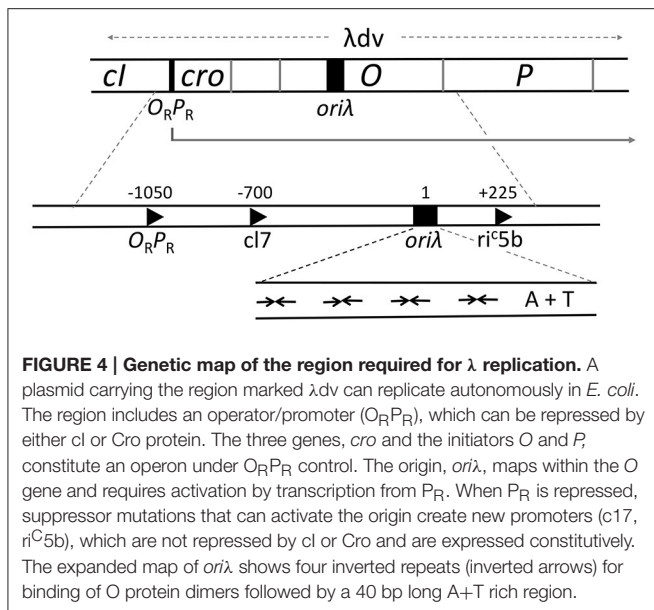
Several other regulators control the DnaA-ATP level. These regulators have been reviewed comprehensively elsewhere, and will not be discussed here (Katayama et al., 2010; Skarstad and Katayama, 2013). Finally, for unknown reasons, transcription is required for replication initiation (Skarstad et al., 1990). The act of transcription elongation induces negative supercoiling of the upstream DNA. An appropriately oriented promoter may thus help origin opening by increasing negative supercoiling. This is further discussed below.

## OPENING OF THE BACTERIOPHAGE LAMBDA ORIGIN BY TRANSCRIPTIONAL ACTIVATION

In phage lambda ( $\lambda$ ), DNA replication has been extensively studied and is fairly well-understood. Before the days of cloning, the small size of the phage genome (about 50 kb, one 1/100th the size of the *E. coli* chromosome) made physical manipulation possible, allowing isolation and characterization of intact replication intermediates. This led to the first unambiguous demonstration that replication starts from a unique origin, and that two replication forks proceed from the origin in opposite directions (bidirectional replication), as was conceived in the replicon model (Jacob et al., 1964; Inman and Schnös, 1971).

Genetic characterization of lambda replication has provided quite a few alternate strategies for replication initiation. For example, instead of DnaA as the initiator, two phage-encoded initiators, the O and P proteins, are used (Ogawa and Tomizawa, 1968). The initiation depends on transcription within or nearby the origin region (Dove et al., 1969). A more remarkable finding was the discovery of three chaperone proteins in *E. coli* (DnaJ, DnaK, and GrpE) and their participation in replication initiation (Georgopoulos and Herskowitz, 1971; Saito and Uchida, 1977; Friedman et al., 1984). *In vitro* studies with purified components reproduced the salient features of the system as determined *in vivo*: bidirectional replication, and requirements for transcription and chaperone proteins (Learn et al., 1993). We elaborate on these features in the context of our general scheme of origin opening.

The minimal region that retains the replication characteristics of the entire genome is contained within 2.4 kb ( $\lambda$ dv, **Figure 4**). It comprises a promoter  $P_R$  and four genes *cro*, *cII*, *O*, and *P* that are transcribed from  $P_R$  (Matsubara, 1981). The origin (*ori* $\lambda$ ) maps within the *O* gene and transcription from  $P_R$  activates the origin, in addition to its role in providing the mRNA for O and P synthesis. *ori* $\lambda$  contains nearly perfect inverted repeats of a 19-bp sequence that bind O protein dimers (Grosschedl and Hobom, 1979; Moore et al., 1979; Tsurimoto and Matsubara, 1981). The initiator binding repeats of the origin were given a special name, iterons (Moore et al., 1979). O binding to *ori* $\lambda$  in a negatively supercoiled DNA causes a significant structural change that includes the opening of the neighboring 40 bp A+T rich region (Dodson et al., 1986; Schnos et al., 1988). The iteron DNA is bent in solution and bends further upon O binding, and it has been proposed that the “free energy of bending is trapped



in the *oriλ*-*O* complex” (Zahn and Blattner, 1987). The opening is dependent on negative supercoiling and binding of *O* protein copies to each of the multiple iterons.

*In vitro* studies suggest that additional proteins are involved in stabilizing the opening to allow loading of the host-encoded helicase, DnaB. As in *oriC*, the initiation of  $\lambda$  replication is separated into several stages. The formation of the *oriλ*-*O* complex is the initial stage, followed by the formation of *oriλ*-*O*-*P*-DnaB, *oriλ*-*O*-*P*-DnaB-DnaJ and *oriλ*-*O*-*P*-DnaB-DnaJ-DnaK complexes (Alfano and McMacken, 1989a,b; Dodson et al., 1989). All these complexes are functional as they support replication when supplemented with the missing replisome components. In the *oriλ*-*O*-*P*-DnaB complex, *P* plays matchmaker by binding simultaneously to *O* and DnaB (Mallory et al., 1990; Osipiuk et al., 1993). DnaB is kept inactive at this stage through interaction with *P*, until chaperone proteins disassemble the complex to activate the helicase (Mensa-Wilmot et al., 1989b; Zylicz et al., 1989). The disassembly requires the ATPase activity of DnaK. The chaperones thus participate in initiation after the origin has opened.

*O*, *P*, and DnaB all harbor cryptic ssDNA binding activity (Learn et al., 1997). Interactions between *O* and *P*, and between *P* and DnaB, which suppress the intrinsic ssDNA binding activity of DnaB, are all required to form a stable ssDNA-*O*-*P*-DnaB complex. Both *O* and *P* of the complex contact the open DUE and stabilize the initial open structure. *O* also stabilizes the *P*-DnaB interaction, perhaps ensuring that DnaB is loaded only at the *O*-bound origin.

Although the chaperones provide a crucial activation function to the helicase, they do not control the efficiency of initiation or, most likely, strand-opening; these are controlled by transcription from the  $P_R$  (Thomas and Bertani, 1964; Dove et al., 1969). Hence, the repressors that control  $P_R$  activity are the regulators of replication. Within the minimal ~2.4 kb replicon ( $\lambda$ dv), *Cro* serves as the repressor for  $P_R$ , and in the intact

phage, repression is enforced by the *cI* protein (Matsubara, 1981; Womble and Rownd, 1986). The two repressors bind to the same operator ( $O_R$ ) sequences. In the prophage state, when  $P_R$  is repressed by *cI*, replication does not initiate even when *O* and *P* are supplied *in trans* (Thomas and Bertani, 1964). Mutations that activate replication under the above conditions (Dove et al., 1969) were found to create new promoters that are not controlled by *cI* (e.g., *cI7*, Figure 4). This observation led to the proposal that the  $\lambda$  origin requires activation by transcription. The transcription requirement has also been confirmed *in vitro* (Mensa-Wilmot et al., 1989a). In an RNA Polymerase dependent purified system, addition of *cI* abrogates replication initiation, but not in the presence of the *cI7* promoter. Later studies showed that the promoter could be downstream and directed away from the origin (e.g., *ri<sup>C</sup>5b*, Figure 4), implying that the origin region itself need not be transcribed (Furth et al., 1982).

The finding that new promoters located on either side of *oriλ* can activate the origin can be explained by the “twin supercoiled domain” model, where a transcribing RNA polymerase generates positive supercoils ahead of it and negative supercoils behind it (Liu and Wang, 1987). A common feature of the new promoters, regardless of whether they are ahead of or behind the origin, is that they are all oriented for rightward transcription, similar to  $P_R$ . In other words, they are all disposed to increase negative superhelicity of the origin region; this is straightforward in the case of *ri<sup>C</sup>5b* which is downstream of *oriλ*, but when the promoter is upstream, as in the case of  $P_R$  or *cI7*, transcription needs to proceed past the origin.

The requirement for transcriptional activation may be indirectly tied to increasing negative superhelicity. Notably, RNA polymerase is not required for *in vitro*  $\lambda$  replication with purified proteins (Mensa-Wilmot et al., 1989b). In the presence of HU, however, the purified system becomes dependent on RNA polymerase and transcription (Mensa-Wilmot et al., 1989a), which can sweep off HU from DNA. HU is known to constrain (reduce) negative supercoils, which could inhibit origin opening (Drlica and Rouviere-Yaniv, 1987; Mensa-Wilmot et al., 1989a). The superhelical density of plasmids isolated from cells appears adequate for replication *in vitro* but when it is reduced by HU binding, the role of transcription becomes obligatory. [Similarly, for *oriC*, transcription can activate replication under some conditions but is not required when purified proteins are used (Funnell et al., 1986)]. Transcription not only counters HU, but also makes replication initiation bidirectional (Learn et al., 1993). Without transcription, replication in the purified system almost always initiates unidirectionally, although *in vivo* it is primarily bidirectional (Schnos and Inman, 1970; Mensa-Wilmot et al., 1989b). How transcription significantly improves the frequency of bidirectional replication remains to be determined (Learn et al., 1993). Transcription and negative supercoiling may also contribute in additional ways (Szambowska et al., 2011). The RNA polymerase  $\beta$  subunit makes a direct contact with the *O* protein, and this interaction is stimulated by negative supercoiling. Thus, lowering the energy required for DNA strand separation may not be the only role of negative supercoiling.



## OPENING OF ORIGINS IN PLASMIDS WITH REPEATED INITIATOR BINDING SITES (ITERONS)

The basic feature of the lambda origin, namely, an array of repeats of replicon-specific initiator binding sites (iterons), can be found in the origins of a large family of bacterial plasmids (**Figure 5**; Chatteraj and Schneider, 1997). Unlike  $\lambda$  iterons, which bind O dimers, plasmid iterons bind monomeric initiators, a feature that is important for regulating replication, as discussed later. Plasmid iterons are generally present in phase with the helical repeat of B-DNA, and disturbing the phasing can inactivate the origin (Brendler et al., 1997; Doran et al., 1998). The presence of phased iterons indicates that the plasmid origins assume a higher order structure, as appears to be the case for *oriC* and *ori $\lambda$* .

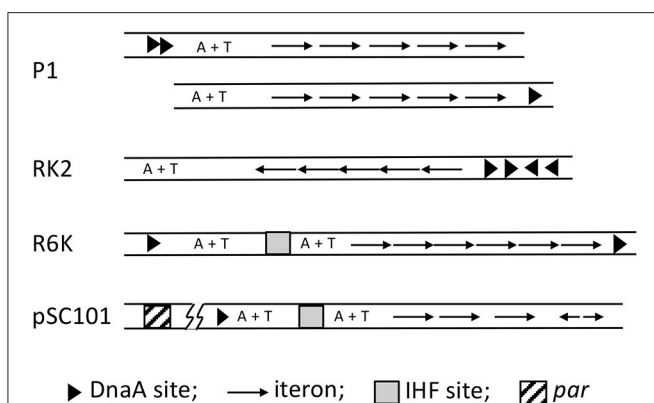
Apart from iterons, the plasmid origins have binding sites for DnaA and a NAP, both of which are required for the origin function. However, the AAA+ (ATPase) domain of DnaA is not required for plasmid replication, suggesting DnaA plays a less crucial role in plasmid replication than in chromosomal replication (Lu et al., 1998; Sharma et al., 2001). Plasmid replication is controlled instead by dimerization of plasmid specific initiators (Paulsson and Chatteraj, 2006). Chaperones are involved in plasmid replication, but unlike their role in  $\lambda$  replication, they control the dimerization efficiency of the initiator and are not involved in activation of the replicative helicase (Wickner et al., 1991). In spite of these differences, the origin opening mechanism is believed to follow the  $\lambda$  paradigm, namely, distortion of the origin by initiator binding to the iteron array with cooperation from DnaA and NAP binding, resulting in opening the A+T rich region. However, unlike  $\lambda$  replication,

transcription is not known to be a requirement for *in vivo* plasmid replication.

Origin opening has been studied in several of the iteron-based plasmids, including P1 (Mukhopadhyay et al., 1993), F (Kawasaki et al., 1996), RK2 (Konieczny et al., 1997), pSC101 (Sharma et al., 2001), and R6K (Lu et al., 1998; Krüger et al., 2001). The roles of the plasmid initiator (usually called Rep), DnaA and NAP vary depending upon the plasmid. In plasmid P1, DnaA alone can initiate opening, but it is greatly facilitated by the addition of RepA (Mukhopadhyay et al., 1993). RepA alone is ineffective. In plasmid F also DnaA alone can open the origin but neither the initiator RepE nor the NAP (HU) alone can do so (Kawasaki et al., 1996). Together, RepE and HU are efficient in opening. Addition of DnaA further increases the efficiency of opening and extends the open region. In plasmid RK2, Rep (TrfA) can open if either HU or DnaA is present (Konieczny et al., 1997). Opening by TrfA together with HU is significantly improved when DnaA is also present. In pSC101, cooperation of DnaA, RepA, and IHF is required to open the origin efficiently (Sharma et al., 2001). All three, Rep (pi), DnaA and a NAP (IHF) are also required to open *ori $\gamma$*  of R6K (Krüger et al., 2001; Lu et al., 1998). However, with a hyperactive variant of pi, DnaA and IHF are not required, indicating that their roles are mostly facilitatory (Krüger and Filutowicz, 2003). The general picture that emerges is that although some opening might be seen without the full complement of the three proteins, the efficiency and/or the extent of the opening are usually different in such cases.

The above studies indicate a direct correlation between the efficiency of origin opening and replication initiation. In P1 and F, situations that increase or decrease initiation due to changes in Rep or iteron concentration also correspondingly enhance or reduce opening (Kawasaki et al., 1996; Park et al., 1998; Park and Chatteraj, 2001; Zzaman and Bastia, 2005). In RK2, whose DUE comprises 13-mer A+T rich repeats like the DUE of *oriC*, changing their sequence, arrangement, or number reduces the stability of the open DUE as well as the origin firing efficiency (Rajewska et al., 2008; Wegrzyn et al., 2014). In pSC101, a RepA mutant specifically defective in interactions with DnaA and replication initiation is also defective in origin opening (Sharma et al., 2001). In R6K, both the monomer and dimer forms of pi bind and bend iterons almost equally, but only the monomer-bound origins can open, which is the form that is proficient in initiation (Krüger et al., 2001; Krüger and Filutowicz, 2003). As mentioned earlier, the facilitators of opening of *ori $\gamma$* , IHF and DnaA, are also required for initiation. The pi mutants that can open without the facilitators are also hyperactive (copy-up) for initiation. These results suggest that initiators control initiation efficiency at the DNA-opening step (Krüger et al., 2001).

In plasmids RK2 and F, the initiators bind ssDUE, as we have described for DnaA binding to *oriC*-DUE, and O and P binding to *ori $\lambda$* -DUE (Wegrzyn et al., 2014). In iteron-bearing plasmids, similar to *oriC* or *ori $\lambda$* , the A+T region is not always flanked by protein binding sites that might prevent migration of the opening away from the origin. Even in cases where the DUE is flanked by DnaA and RepA binding sites, the DnaA binding sites can be moved to the other end of the origin, so that the plasmid origin now mimics *oriC* or *ori $\lambda$*  (P1 origin, **Figure 5**; Abeles et al.,



**FIGURE 5 | Maps of iteron-bearing plasmids.** The plasmid iterons are mostly direct repeats, as opposed to the inverted repeats in *ori $\lambda$* . Two P1 maps are shown, the top one being the wild type and the one below with the DnaA binding sites (arrow heads) deleted and one of them moved next to the iterons. Both of the origins are functional, indicating that the A+T rich region does not have to be bounded by protein binding sites. In the RK2 plasmid, the A+T rich region naturally lacks protein binding sites in one of its flanks. By contrast, the R6K *ori $\gamma$*  is bounded by DnaA sites, which most likely interact directly. pSC101 uniquely includes a *par* locus (about 200 bp away from the origin as indicated by the line breaks), which binds gyrase and specifically changes the negative superhelicity of the origin, and thereby enhances replication of the plasmid.

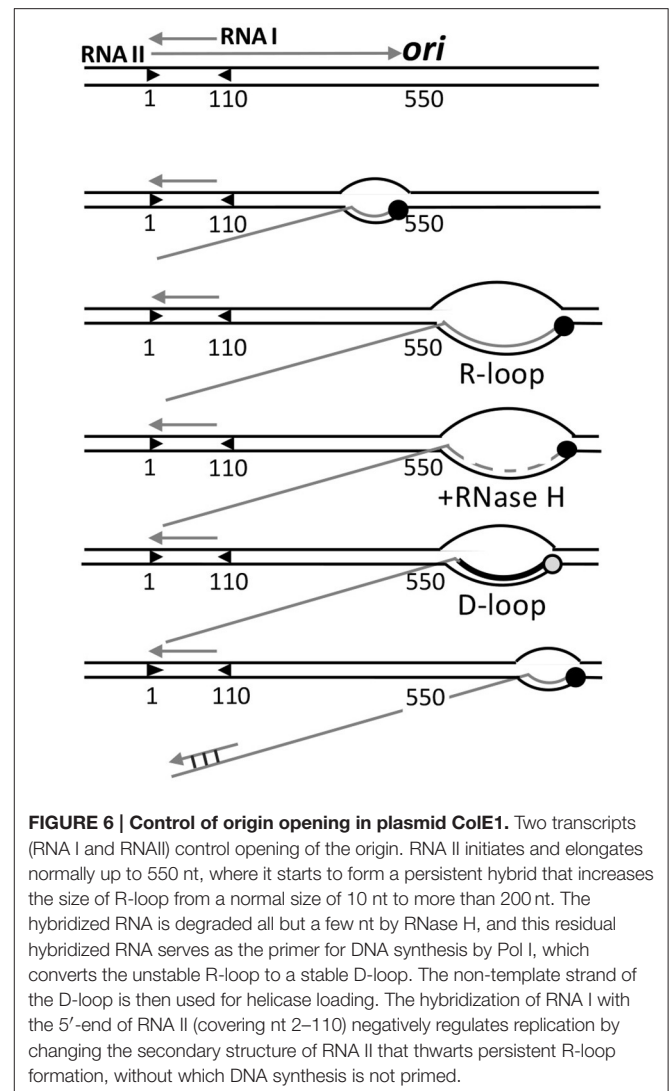
1990; Park and Chattoraj, 2001). These results argue in favor of active anchoring mechanisms. Indeed in RK2 and F, iteron-bound initiators can bind simultaneously to an oligo from DUE, as they can at *oriC* (Ozaki and Katayama, 2012; Wegrzyn et al., 2014). While DnaA uses two different domains for binding to ds- and ss-DNA, it is not known whether that is also the case for plasmid initiators.

In many iteron-based plasmids, chaperones improve initiator–iteron binding that leads to opening. This is in contrast to their roles in  $\lambda$ , where the chaperones come into play after origin opening. In plasmids, the chaperones increase the availability of initiator monomers in a form that binds to iterons (Wickner et al., 1991; Ishiai et al., 1994; Toukdarian et al., 1996; Zzaman et al., 2004b). The increase in monomer results from dimer dissociation *in vitro* and this results from refolding of misfolded subunits that apparently reduces dimerization affinity (Giraldo et al., 2003; Nakamura et al., 2007). The chaperones could be either DnaK and its cohorts DnaJ and GrpE for P1 RepA (Wickner et al., 1991), F RepE (Ishiai et al., 1994) and R6K pi (Zzaman et al., 2004b), or only ClpA for P1 RepA (Wickner et al., 1994), or ClpB+DnaK+DnaJ+GrpE as for RK2 TrfA (Konieczny and Liberek, 2002). Although replication of these plasmids require the monomers, prevention of over-replication requires the dimers (Paulsson and Chattoraj, 2006). The chaperones thus play an important role in maintaining the proper balance between the activator (monomer) and inhibitor (dimer) forms of the initiators.

Origins of iteron-based plasmids are generally not dependent on transcription. As in *E. coli* and  $\lambda$  replication, RNA polymerase is not required for R6K and F replication when purified components are used (Abhyankar et al., 2003; Zzaman et al., 2004a). However, as in other systems, RNA polymerase has been shown to play a beneficial role in pSC101 replication *in vivo*. The plasmid has a locus, *par*, for gyrase binding that increases negative supercoiling specifically at the plasmid origin (Miller et al., 1990; Conley and Cohen, 1995). The supercoiling and replication defects of  $\Delta par$  strains are suppressed by transcription from a suitably positioned and oriented promoter (Beaucage et al., 1991). These results are consistent with the “twin supercoiled domain” model but also support the view that superhelicity can be changed locally without changing the overall superhelical density of the plasmid (Rahmouni and Wells, 1989). The localized changes are believed to improve initiator interactions with the origin and thereby its activity (Ingmer and Cohen, 1993).

## OPENING OF *ColE1* PLASMID ORIGIN BY FORMATION OF A PERSISTENT RNA-DNA HYBRID

Replication initiation of plasmid *ColE1* differs from that of the replicons described above. *ColE1* does not use a plasmid-encoded initiator. Rather, initiation depends on the host RNA polymerase, which synthesizes a non-coding RNA (RNA II) from a promoter 550 bp upstream of the origin. This serves to open the origin and provides the primer for DNA synthesis (Figure 6). This



**FIGURE 6 | Control of origin opening in plasmid *ColE1*.** Two transcripts (RNA I and RNAII) control opening of the origin. RNA II initiates and elongates normally up to 550 nt, where it starts to form a persistent hybrid that increases the size of R-loop from a normal size of 10 nt to more than 200 nt. The hybridized RNA is degraded all but a few nt by RNase H, and this residual hybridized RNA serves as the primer for DNA synthesis by Pol I, which converts the unstable R-loop to a stable D-loop. The non-template strand of the D-loop is then used for helicase loading. The hybridization of RNA I with the 5'-end of RNA II (covering nt 2–110) negatively regulates replication by changing the secondary structure of RNA II that thwarts persistent R-loop formation, without which DNA synthesis is not primed.

diverges from the norm for bacterial replicons, where the primer is synthesized by the primase, DnaG, which is brought to the open origin by DnaB-DnaG protein-protein interactions (Bell and Kaguni, 2013).

In transcription elongation, normally about 10 nt at the 3'-end of transcripts stay hybridized to the template strand, forming a three-stranded bubble called the R-loop. As new nucleotides are added at the 3'-end, the hybridized nucleotides at the 5'-end of R-loops leave the template strand, thereby maintaining the size of the translocating R-loops. This canonical scenario is maintained in the case of RNA II for the first 550 nt, after which the RNA does not exit from the R-loop as new nucleotides are added. The persistence of hybridized RNA causes the R-loop to grow in size to even more than 200 bp. [Persistence of RNA-DNA hybrid was also found in transcriptional activation of *oriC* (Baker and Kornberg, 1988)]. RNase H then almost completely degrades the RNA from the R-loop except for the 4 to 5 hybridized nucleotides at the 5'-end side. The residual hybridized RNA suffices to serve as the primer to start DNA

synthesis by DNA polymerase I (Pol I). *In vivo*, RNase H most likely prevents the R-loops to expand much in length, which needs to be at least 40 nt to allow helicase loading (Masukata et al., 1987). The three stranded D-loop synthesized by Pol I apparently provides sufficient opening for the helicase. The open (D-loop) region does not have to be A+T rich in this case because the opening is caused by DNA synthesis, not from the intrinsic instability of the region. The D-loop is a stable structure wherein the newly synthesized strand prevents the non-template strand from hybridizing back to the template strand.

Initiation is controlled by the prevention of persistent RNA-DNA hybridization. A shorter RNA (RNA I) about 110 nt long and fully complementary to the 5'-end of RNA II is responsible for preventing R-loop expansion. RNA I and RNAII are both constitutively synthesized. As their concentrations increase with increasing plasmid copy number, hybridization becomes increasingly significant. Hybridization changes the secondary structure of RNA II that thwarts persistent R-loops formation, and hence, priming.

In sum, although the initiation of ColE1 replication is mechanistically distinct, it espouses the two features highlighted here: the need for stabilizing the open state and the use of the origin opening stage to control initiation. It should be noted that in ColE1, transcription plays a direct and essential role in initiation by providing the primer, whereas in other cases transcription helps indirectly by increasing mainly negative superhelicity and is not obligatory. Finally, the study of ColE1 replication provided the first example of control by a non-coding antisense RNA, which is now recognized to be widespread in biology (Tomizawa et al., 1981; Eguchi et al., 1991).

## ORIGIN OPENING IN EUKARYOTES

The bacterial program of first opening the origin and then loading two hexameric helicases sequentially for bidirectional replication is not conserved in archaea and eukaryotes. In the latter, both the helicases are loaded together as a double hexamer to an unopened ds-origin (Bell and Kaguni, 2013). The loading otherwise follows the basic bacterial paradigm: the helicase (a hetero-hexamer, MCM2-7) in association with the helicase loader Cdt1, is recruited to the origin bound by the initiator (ORC) as a complex with another factor called Cdc6. The double hexamer is loaded in the post-mitotic, early G1 phases of the cell cycle in an inactive state and as a ring that encircles the ds-origin in its central core. The helicase activation and strand separation occur later in S-phase where the double hexamer is converted to single hexameric rings, each encircling one single strand for bidirectional movement. These major transitions require S-phase kinases and several additional factors, the details of which are under current investigation (Yardimci and Walter, 2014; Bochman and Schwacha, 2015; Petojevic et al., 2015). Loading and activation of the helicase at different stages of the cell cycle help to restrict initiation to only once per cell cycle (Nguyen et al., 2001; Arias and Walter, 2007). Since no new helicases can be loaded in S-phase, new origin firing

cannot happen either. Thus, although the mechanisms of helicase loading have been largely conserved, the mechanisms of helicase activation and origin opening have diverged in different domains of life.

## CONCLUSIONS AND FUTURE CONSIDERATIONS

Here we have provided a few examples of how bacterial origins open, permitting loading of the replicative helicase. As some opening is possible without initiators, it is likely that the origin is inherently unstable (Gille and Messer, 1991; Mukhopadhyay and Chattoraj, 1993; Polaczek et al., 1998). Initiator binding pushes the propensity of opening over the threshold. Of all the requirements for opening, the free energy of negative supercoiling of the origin region appears to be the most basic requirement (Miller et al., 1990). Many of the facilitators of opening (e.g., transcription, NAPs, and pSC101*par*) work through changing superhelicity of the origin region. The activated state of supercoiled DNA facilitates changes of the DNA structure, events that are less likely to occur with linear DNA.

Initiator multimerization appears to be a general contributor to origin opening. This can allow wrapping of DNA with initiator and consequently changing the superhelicity of neighboring DNA. DNA binding proteins usually bend the DNA, and the initiators are no exception (Mukherjee et al., 1985; Mukhopadhyay and Chattoraj, 1993; Schaper and Messer, 1995). Stress from DNA bending can induce base-pair opening (Kahn et al., 1994). [The NAP binding can also untwist DNA (Teter et al., 2000)]. The opening by bending is initially local but the unwinding may migrate. Reducing the number of initiator binding sites in DNA generally makes the origin inefficient or inactive, depending upon the degree of binding site reduction.

Multimerization is also involved in ssDNA binding, which either stabilizes the open region or promotes actual unwinding (or both; **Figure 3A**). What triggers the conformational switch in initiators that allows them to bind ssDNA remains a challenging question (Duderstadt et al., 2011). More structural studies of the complexes are in order to get further insights into the opening process. This is now a realizable goal given the recent progress in cryo-EM (Merk et al., 2016). Even in the relatively clear case of ColE1, the trigger that causes the RNA polymerase to start forming persistent RNA-DNA hybrids when it encounters the origin sequence remains speculative.

Although we have referred to the DUE simply as A+T rich, the exact sequence of the region also matters (Hwang and Kornberg, 1992b; Ozaki et al., 2008). In RK2, the A+T rich region has 13-mers that bear partial homology to the *E. coli* 13-mers, but they are not interchangeable (Kowalczyk et al., 2005). In  $\lambda$ , the sequence of the A+T rich region is highly strand asymmetric: almost all purines are in one strand (Schnos et al., 1988). Paradoxically, such extreme asymmetric distribution of purine and pyrimidines stiffens the DNA more than DNA with more random sequences (Wells et al., 1970). In most cases, a specific strand is captured in the open region (Mukhopadhyay et al., 1993; Rajewska et al., 2008; Wegrzyn et al., 2014). Strand



capture preference is also observed in experiments where the single strands are supplied *in trans* (Ozaki et al., 2008; Rajewska et al., 2008; Wegrzyn et al., 2014). A recent study has revealed a repeating trinucleotide motif that is conserved in bacterial DUEs and is required for origin opening (Richardson et al., 2016).

A recent study has provided a new insight into how supercoiling-induced DNA opening of A+T rich sequences is favored by G+C rich flanks (Vlijm et al., 2015). Such flanks are found in A+T rich stretches of some replication origins but their role has remained unclear (Brendler et al., 1991; Richardson et al., 2016). G+C rich stretches were suggested to resist transmission of torsional stress along the DNA (Skarstad et al., 1990). The recent study suggests further that the G+C stretches because of their stiffness resist (plectoneme formation) supercoiling of under-wound DNA and thereby help to concentrate the unwinding to A+T rich regions.

ATP not only plays an essential role in origin opening by AAA+ initiators, it also plays a rather mysterious role in opening origins where the initiators have no known ATP binding domain or ATPase activity. For example,  $\pi$  binding and bending are not sufficient to open R6K *ori $\gamma$*  without the presence of ATP (Krüger and Filutowicz, 2003). ATP, although not required, enlarges opening in RK2 even when a hyperactive mutant TrfA is used in conjunction with the facilitators DnaA and HU (Konieczny et al., 1997). Also at *oriC*, ATP requirements for DNA binding and origin opening by DnaA can vary by orders of magnitude (nM vs. mM; Figure 1; Bramhill and Kornberg, 1988a). A high

ATP concentration can cause a conformation change in DnaA that appears likely to be required for opening (Saxena et al., 2015).

In closing, we prefer the view that the opening proceeds in steps rather than by a highly cooperative transition (the two models in Figure 3A). Initiator binding initiates the opening and it is further enhanced by multimerization of the initiator, facilitators like DnaA and NAPs (in plasmids) and by factors such as the helicase loaders DnaC and  $\lambda$ P that have ssDNA binding activity. The opening by initiators alone may not be sufficient for helicase loading. The involvement of multiple factors provides multiple opportunities for regulation.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

Our funding is from the Intramural Research Program of the Center for Cancer Research, NCI, NIH.

## ACKNOWLEDGMENTS

The authors are grateful to Deepak Bastia, Julia Grimwade, David Levens, Roger McMacken, and Michael Yarmolinsky for thoughtful comments, and to Jemima Barrowman for editing.

## REFERENCES

- Abeles, A. L., Reaves, L. D., and Austin, S. J. (1990). A single DnaA box is sufficient for initiation from the P1 plasmid origin. *J. Bacteriol.* 172, 4386–4391.
- Abhyankar, M. M., Zzaman, S., and Bastia, D. (2003). Reconstitution of R6K DNA replication *in vitro* using 22 purified proteins. *J. Biol. Chem.* 278, 45476–45484. doi: 10.1074/jbc.M308516200
- Alfano, C., and McMacken, R. (1989a). Heat shock protein-mediated disassembly of nucleoprotein structures is required for the initiation of bacteriophage lambda DNA replication. *J. Biol. Chem.* 264, 10709–10718.
- Alfano, C., and McMacken, R. (1989b). Ordered assembly of nucleoprotein structures at the bacteriophage lambda replication origin during the initiation of DNA replication. *J. Biol. Chem.* 264, 10699–10708.
- Arias, E. E., and Walter, J. C. (2007). Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. *Genes Dev.* 21, 497–518. doi: 10.1101/gad.1508907
- Baker, T. A., and Kornberg, A. (1988). Transcriptional activation of initiation of replication from the *E. coli* chromosomal origin: an RNA-DNA hybrid near *oriC*. *Cell* 55, 113–123. doi: 10.1016/0092-8674(88)90014-1
- Beaucage, S. L., Miller, C. A., and Cohen, S. N. (1991). Gyrase-dependent stabilization of pSC101 plasmid inheritance by transcriptionally active promoters. *EMBO J.* 10, 2583–2588.
- Bell, S. P., and Kaguni, J. M. (2013). Helicase loading at chromosomal origins of replication. *Cold Spring Harb. Perspect. Biol.* 5:a010124. doi: 10.1101/cshperspect.a010124
- Bochman, M. L., and Schwacha, A. (2015). DNA replication: strand separation unravelled. *Nature* 524, 166–167. doi: 10.1038/nature14643
- Bowater, R., Aboul-ela, F., and Lilley, D. M. (1991). Large-scale stable opening of supercoiled DNA in response to temperature and supercoiling in (A + T)-rich regions that promote low-salt cruciform extrusion. *Biochemistry* 30, 11495–11506. doi: 10.1021/bi00113a003
- Bramhill, D., and Kornberg, A. (1988a). Duplex opening by DnaA protein at novel sequences in initiation of replication at the origin of the *E. coli* chromosome. *Cell* 52, 743–755.
- Bramhill, D., and Kornberg, A. (1988b). A model for initiation at origins of DNA replication. *Cell* 54, 915–918.
- Brendler, T., Abeles, A., and Austin, S. (1991). Critical sequences in the core of the P1 plasmid replication origin. *J. Bacteriol.* 173, 3935–3942.
- Brendler, T. G., Abeles, A. L., Reaves, L. D., and Austin, S. J. (1997). The iteron bases and spacers of the P1 replication origin contain information that specifies the formation of a complex structure involved in initiation. *Mol. Microbiol.* 23, 559–567. doi: 10.1046/j.1365-2958.1997.d01-1869.x
- Chattoraj, D. K., and Schneider, T. D. (1997). Replication control of plasmid P1 and its host chromosome: the common ground. *Prog. Nucleic Acid Res. Mol. Biol.* 57, 145–186. doi: 10.1016/S0079-6603(08)60280-9
- Cheng, H. M., Groger, P., Hartmann, A., and Schlierf, M. (2015). Bacterial initiators form dynamic filaments on single-stranded DNA monomer by monomer. *Nucleic Acids Res.* 43, 396–405. doi: 10.1093/nar/gku1284
- Chodavarapu, S., Felczak, M. M., and Kaguni, J. M. (2011). Two forms of ribosomal protein L2 of *Escherichia coli* that inhibit DnaA in DNA replication. *Nucleic Acids Res.* 39, 4180–4191. doi: 10.1093/nar/gkq1203
- Chodavarapu, S., Felczak, M. M., Yaniv, J. R., and Kaguni, J. M. (2008a). *Escherichia coli* DnaA interacts with HU in initiation at the *E. coli* replication origin. *Mol. Microbiol.* 67, 781–792. doi: 10.1111/j.1365-2958.2007.06094.x
- Chodavarapu, S., Gomez, R., Vicente, M., and Kaguni, J. M. (2008b). *Escherichia coli* Dps interacts with DnaA protein to impede initiation: a model of adaptive mutation. *Mol. Microbiol.* 67, 1331–1346. doi: 10.1111/j.1365-2958.2008.06127.x
- Conley, D. L., and Cohen, S. N. (1995). Effects of the pSC101 partition (*par*) locus on *in vivo* DNA supercoiling near the plasmid replication origin. *Nucleic Acids Res.* 23, 701–707. doi: 10.1093/nar/23.4.701
- Dodson, M., Echols, H., Wickner, S., Alfano, C., Mensa-Wilmot, K., Gomes, B., et al. (1986). Specialized nucleoprotein structures at the origin of

- replication of bacteriophage lambda: localized unwinding of duplex DNA by a six-protein reaction. *Proc. Natl. Acad. Sci. U.S.A.* 83, 7638–7642. doi: 10.1073/pnas.83.20.7638
- Dodson, M., McMacken, R., and Echols, H. (1989). Specialized nucleoprotein structures at the origin of replication of bacteriophage lambda. Protein association and disassociation reactions responsible for localized initiation of replication. *J. Biol. Chem.* 264, 10719–10725.
- Doran, K. S., Konieczny, I., and Helinski, D. R. (1998). Replication origin of the broad host range plasmid RK2. Positioning of various motifs is critical for initiation of replication. *J. Biol. Chem.* 273, 8447–8453. doi: 10.1074/jbc.273.14.8447
- Dorman, C. J. (2009). Nucleoid-associated proteins and bacterial physiology. *Adv. Appl. Microbiol.* 67, 47–64. doi: 10.1016/S0065-2164(08)01002-2
- Dove, W. F., Hargrove, E., Ohashi, M., Haugli, F., and Guha, A. (1969). Replicator activation in lambda. *Japan J Genet* 44, 11–22.
- Drica, K., and Rouviere-Yaniv, J. (1987). Histone-like proteins of bacteria. *Microbiol. Rev.* 51, 301–319.
- Duderstadt, K. E., Chuang, K., and Berger, J. M. (2011). DNA stretching by bacterial initiators promotes replication origin opening. *Nature* 478, 209–213. doi: 10.1038/nature10455
- Duderstadt, K. E., Mott, M. L., Crisone, N. J., Chuang, K., Yang, H., and Berger, J. M. (2010). Origin remodeling and opening in bacteria rely on distinct assembly states of the DnaA initiator. *J. Biol. Chem.* 285, 28229–28239. doi: 10.1074/jbc.M110.147975
- EGUCHI, Y., ITOH, T., and TOMIZAWA, J. (1991). Antisense RNA. *Annu. Rev. Biochem.* 60, 631–652. doi: 10.1146/annurev.bi.60.070191.003215
- Erzberger, J. P., Mott, M. L., and Berger, J. M. (2006). Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat. Struct. Mol. Biol.* 13, 676–683. doi: 10.1038/nsmb1115
- Erzberger, J. P., Pirruccello, M. M., and Berger, J. M. (2002). The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. *EMBO J.* 21, 4763–4773. doi: 10.1093/emboj/cdf496
- Friedman, D. I., Olson, E. R., Georgopoulos, C., Tilly, K., Herskowitz, I., and Banuett, F. (1984). Interactions of bacteriophage and host macromolecules in the growth of bacteriophage lambda. *Microbiol. Rev.* 48, 299–325.
- Funnell, B. E., Baker, T. A., and Kornberg, A. (1986). Complete enzymatic replication of plasmids containing the origin of the *Escherichia coli* chromosome. *J. Biol. Chem.* 261, 5616–5624.
- Furth, M. E., Dove, W. F., and Meyer, B. J. (1982). Specificity determinants for bacteriophage lambda DNA replication. III. Activation of replication in lambda *ri<sup>c</sup>* mutants by transcription outside of *ori*. *J. Mol. Biol.* 154, 65–83. doi: 10.1016/0022-2836(82)90417-X
- Georgopoulos, C., and Herskowitz, I. (1971). “*Escherichia coli* mutants blocked in lambda DNA synthesis,” in *The Bacteriophage Lambda*, ed A. D. Hershey (New York, NY: Cold Spring Harbor laboratory), 553–564.
- Gille, H., and Messer, W. (1991). Localized DNA melting and structural perturbations in the origin of replication, *oriC*, of *Escherichia coli* *in vitro* and *in vivo*. *EMBO J.* 10, 1579–1584.
- Giraldo, R. (2003). Common domains in the initiators of DNA replication in Bacteria, Archaea and Eukarya: combined structural, functional and phylogenetic perspectives. *FEMS Microbiol. Rev.* 26, 533–554. doi: 10.1111/j.1574-6976.2003.tb00629.x
- Giraldo, R., Fernández-Tornero, C., Evans, P. R., Díaz-Orejas, R., and Romero, A. (2003). A conformational switch between transcriptional repression and replication initiation in the RepA dimerization domain. *Nat. Struct. Biol.* 10, 565–571. doi: 10.1038/nsb937
- Gotoh, O., and Tagashira, Y. (1981). Locations of frequently opening regions on natural DNAs and their relation to functional loci. *Biopolymers* 20, 1043–1058. doi: 10.1002/bip.1981.360200514
- Grosschedl, R., and Hobom, G. (1979). DNA sequences and structural homologies of the replication origins of lambdoid bacteriophages. *Nature* 277, 621–627. doi: 10.1038/277621a0
- Hwang, D. S., and Kornberg, A. (1992a). Opening of the replication origin of *Escherichia coli* by DnaA protein with protein HU or IHF. *J. Biol. Chem.* 267, 23083–23086.
- Hwang, D. S., and Kornberg, A. (1992b). Opposed actions of regulatory proteins, DnaA and IcaA, in opening the replication origin of *Escherichia coli*. *J. Biol. Chem.* 267, 23087–23091.
- Ingmer, H., and Cohen, S. N. (1993). The pSC101 *par* locus alters protein-DNA interactions *in vivo* at the plasmid replication origin. *J. Bacteriol.* 175, 6046–6048.
- Inman, R. B. (1966). A denaturation map of the lambda phage DNA molecule determined by electron microscopy. *J. Mol. Biol.* 18, 464–476. doi: 10.1016/S0022-2836(66)80037-2
- Inman, R. B., and Schnös, M. (1971). Structure of branch points in replicating DNA: presence of single-stranded connections in lambda DNA branch points. *J. Mol. Biol.* 56, 319–325. doi: 10.1016/0022-2836(71)90467-0
- Ishiai, M., Wada, C., Kawasaki, Y., and Yura, T. (1994). Replication initiator protein RepE of mini-F plasmid: functional differentiation between monomers (initiator) and dimers (autogenous repressor). *Proc. Natl. Acad. Sci. U.S.A.* 91, 3839–3843. doi: 10.1073/pnas.91.9.3839
- Jacob, F., Brenner, S., and Cuzin, F. (1964). On the regulation of DNA replication in bacteria. *Cold Spring Harb. Symp. Quant. Biol.* 28, 329–348. doi: 10.1101/SQB.1963.028.01.048
- Kaguni, J. M., and Kornberg, A. (1984). Replication initiated at the origin (*oriC*) of the *E. coli* chromosome reconstituted with purified enzymes. *Cell* 38, 183–190. doi: 10.1016/0092-8674(84)90539-7
- Kahn, J. D., Yun, E., and Crothers, D. M. (1994). Detection of localized DNA flexibility. *Nature* 368, 163–166. doi: 10.1038/368163a0
- Katayama, T., Ozaki, S., Keyamura, K., and Fujimitsu, K. (2010). Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and *oriC*. *Nat. Rev. Microbiol.* 8, 163–170. doi: 10.1038/nrmicro2314
- Kaur, G., Vora, M. P., Czerwonka, C. A., Rozgaja, T. A., Grimwade, J. E., and Leonard, A. C. (2014). Building the bacterial orisome: high affinity DnaA recognition plays a role in setting the conformation of *oriC* DNA. *Mol. Microbiol.* 91, 1148–1163. doi: 10.1111/mmi.12525
- Kawakami, H., Keyamura, K., and Katayama, T. (2005). Formation of an ATP-DnaA-specific initiation complex requires DnaA Arginine 285, a conserved motif in the AAA+ protein family. *J. Biol. Chem.* 280, 27420–27430. doi: 10.1074/jbc.M502764200
- Kawasaki, Y., Matsunaga, F., Kano, Y., Yura, T., and Wada, C. (1996). The localized melting of mini-F origin by the combined action of the mini-F initiator protein (RepE) and HU and DnaA of *Escherichia coli*. *Mol. Gen. Genet.* 253, 42–49. doi: 10.1007/s004380050294
- Keyamura, K., Abe, Y., Higashi, M., Ueda, T., and Katayama, T. (2009). DiaA dynamics are coupled with changes in initial origin complexes leading to helicase loading. *J. Biol. Chem.* 284, 25038–25050. doi: 10.1074/jbc.M109.002717
- Konieczny, I., Doran, K. S., Helinski, D. R., and Blasina, A. (1997). Role of TrfA and DnaA proteins in origin opening during initiation of DNA replication of the broad host range plasmid RK2. *J. Biol. Chem.* 272, 20173–20178. doi: 10.1074/jbc.272.32.20173
- Konieczny, I., and Liberek, K. (2002). Cooperative action of *Escherichia coli* ClpB protein and DnaK chaperone in the activation of a replication initiation protein. *J. Biol. Chem.* 277, 18483–18488. doi: 10.1074/jbc.M107580200
- Kowalczyk, L., Rajewska, M., and Konieczny, I. (2005). Positioning and the specific sequence of each 13-mer motif are critical for activity of the plasmid RK2 replication origin. *Mol. Microbiol.* 57, 1439–1449. doi: 10.1111/j.1365-2958.2005.04770.x
- Kowalski, D., and Eddy, M. J. (1989). The DNA unwinding element: a novel, *cis*-acting component that facilitates opening of the *Escherichia coli* replication origin. *EMBO J.* 8, 4335–4344.
- Krüger, R., and Filutowicz, M. (2003). pi protein- and ATP-dependent transitions from ‘closed’ to ‘open’ complexes at the gamma *ori* of plasmid R6K. *Nucleic Acids Res.* 31, 5993–6003. doi: 10.1093/nar/gkg809
- Krüger, R., Konieczny, I., and Filutowicz, M. (2001). Monomer/dimer ratios of replication protein modulate the DNA strand-opening in a replication origin. *J. Mol. Biol.* 306, 945–955. doi: 10.1006/jmbi.2000.4426
- Kurokawa, K., Nishida, S., Emoto, A., Sekimizu, K., and Katayama, T. (1999). Replication cycle-coordinated change of the adenine nucleotide-bound forms of DnaA protein in *Escherichia coli*. *EMBO J.* 18, 6642–6652. doi: 10.1093/emboj/18.23.6642
- Learn, B. A., Um, S. J., Huang, L., and McMacken, R. (1997). Cryptic single-stranded-DNA binding activities of the phage lambda P and *Escherichia coli* DnaC replication initiation proteins facilitate the transfer of *E. coli*

- DnaB helicase onto DNA. *Proc. Natl. Acad. Sci. U.S.A.* 94, 1154–1159. doi: 10.1073/pnas.94.4.1154
- Learn, B., Karzai, A. W., and McMacken, R. (1993). Transcription stimulates the establishment of bidirectional lambda DNA replication *in vitro*. *Cold Spring Harb. Symp. Quant. Biol.* 58, 389–402. doi: 10.1101/SQB.1993.058.01.046
- Leonard, A. C., and Grimwade, J. E. (2010). Initiation of DNA replication. *EcoSal Plus* 4. doi: 10.1128/ecosalplus.4.4.1
- Leonard, A. C., and Grimwade, J. E. (2015). The orisome: structure and function. *Front. Microbiol.* 6:545. doi: 10.3389/fmicb.2015.00545
- Liu, L. F., and Wang, J. C. (1987). Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. U.S.A.* 84, 7024–7027. doi: 10.1073/pnas.84.20.7024
- Lu, M., Campbell, J. L., Boye, E., and Kleckner, N. (1994). SeqA: a negative modulator of replication initiation in *E. coli*. *Cell* 77, 413–426. doi: 10.1016/0092-8674(94)90156-2
- Lu, Y. B., Datta, H. J., and Bastia, D. (1998). Mechanistic studies of initiator-initiator interaction and replication initiation. *EMBO J.* 17, 5192–5200. doi: 10.1093/emboj/17.17.5192
- Mallory, J. B., Alfano, C., and McMacken, R. (1990). Host virus interactions in the initiation of bacteriophage lambda DNA replication. Recruitment of *Escherichia coli* DnaB helicase by lambda P replication protein. *J. Biol. Chem.* 265, 13297–13307.
- Masukata, H., Dasgupta, S., and Tomizawa, J. (1987). Transcriptional activation of ColE1 DNA synthesis by displacement of the nontranscribed strand. *Cell* 51, 1123–1130. doi: 10.1016/0092-8674(87)90598-8
- Matsubara, K. (1981). Replication control system in lambda dv. *Plasmid* 5, 32–52. doi: 10.1016/0147-619X(81)90076-7
- McGarry, K. C., Ryan, V. T., Grimwade, J. E., and Leonard, A. C. (2004). Two discriminatory binding sites in the *Escherichia coli* replication origin are required for DNA strand opening by initiator DnaA-ATP. *Proc. Natl. Acad. Sci. U.S.A.* 101, 2811–2816. doi: 10.1073/pnas.0400340101
- Mensa-Wilmot, K., Carroll, K., and McMacken, R. (1989a). Transcriptional activation of bacteriophage lambda DNA replication *in vitro*: regulatory role of histone-like protein HU of *Escherichia coli*. *EMBO J.* 8, 2393–2402.
- Mensa-Wilmot, K., Seaby, R., Alfano, C., Wold, M. C., Gomes, B., and McMacken, R. (1989b). Reconstitution of a nine-protein system that initiates bacteriophage lambda DNA replication. *J. Biol. Chem.* 264, 2853–2861.
- Merk, A., Bartesaghi, A., Banerjee, S., Falconieri, V., Rao, P., Davis, M. I., et al. (2016). Breaking Cryo-EM resolution barriers to facilitate drug discovery. *Cell* 165, 1698–1707. doi: 10.1016/j.cell.2016.05.040
- Miller, C. A., Beaucage, S. L., and Cohen, S. N. (1990). Role of DNA superhelicity in partitioning of the pSC101 plasmid. *Cell* 62, 127–133. doi: 10.1016/0092-8674(90)90246-B
- Miller, D. T., Grimwade, J. E., Betteridge, T., Rozgaja, T., Torgue, J. J., and Leonard, A. C. (2009). Bacterial origin recognition complexes direct assembly of higher-order DnaA oligomeric structures. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18479–18484. doi: 10.1073/pnas.0909472106
- Moore, D. D., Denniston-Thompson, K., Kruger, K. E., Furth, M. E., Williams, B. G., Daniels, D. L., et al. (1979). Dissection and comparative anatomy of the origins of replication of lambdoid phages. *Cold Spring Harb. Symp. Quant. Biol.* 43(Pt 1), 155–163. doi: 10.1101/SQB.1979.043.01.022
- Mukherjee, S., Patel, I., and Bastia, D. (1985). Conformational changes in a replication origin induced by an initiator protein. *Cell* 43, 189–197. doi: 10.1016/0092-8674(85)90023-6
- Mukhopadhyay, G., Carr, K. M., Kaguni, J. M., and Chatteraj, D. K. (1993). Open-complex formation by the host initiator, DnaA, at the origin of P1 plasmid replication. *EMBO J.* 12, 4547–4554.
- Mukhopadhyay, G., and Chatteraj, D. K. (1993). Conformation of the origin of P1 plasmid replication. Initiator protein induced wrapping and intrinsic unstacking. *J. Mol. Biol.* 231, 19–28. doi: 10.1006/jmbi.1993.1253
- Nakamura, A., Wada, C., and Miki, K. (2007). Structural basis for regulation of bifunctional roles in replication initiator protein. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18484–18489. doi: 10.1073/pnas.0705623104
- Neuwald, A. F., Aravind, L., Spouge, J. L., and Koonin, E. V. (1999). AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* 9, 27–43.
- Nguyen, V. Q., Co, C., and Li, J. J. (2001). Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature* 411, 1068–1073. doi: 10.1038/35082600
- Nievera, C., Torgue, J. J., Grimwade, J. E., and Leonard, A. C. (2006). SeqA blocking of DnaA-oriC interactions ensures staged assembly of the *E. coli* pre-RC. *Mol. Cell* 24, 581–592. doi: 10.1016/j.molcel.2006.09.016
- Odegrip, R., Schoen, S., Haggard-Ljungquist, E., Park, K., and Chatteraj, D. K. (2000). The interaction of bacteriophage P2 B protein with *Escherichia coli* DnaB helicase. *J. Virol.* 74, 4057–4063. doi: 10.1128/JVI.74.9.4057-4063.2000
- O'Donnell, M. (2006). Replisome architecture and dynamics in *Escherichia coli*. *J. Biol. Chem.* 281, 10653–10656. doi: 10.1074/jbc.R500028200
- Ogawa, T., and Tomizawa, J. (1968). Replication of bacteriophage DNA. I. Replication of DNA of lambda phage defective in early functions. *J. Mol. Biol.* 38, 217–225. doi: 10.1016/0022-2836(68)90407-5
- Osipiuk, J., Georgopoulos, C., and Zyllic, M. (1993). Initiation of lambda DNA replication. The *Escherichia coli* small heat shock proteins, DnaJ and GrpE, increase DnaK's affinity for the lambda P protein. *J. Biol. Chem.* 268, 4821–4827.
- Ozaki, S., and Katayama, T. (2009). DnaA structure, function, and dynamics in the initiation at the chromosomal origin. *Plasmid* 62, 71–82. doi: 10.1016/j.plasmid.2009.06.003
- Ozaki, S., and Katayama, T. (2012). Highly organized DnaA-oriC complexes recruit the single-stranded DNA for replication initiation. *Nucleic Acids Res.* 40, 1648–1665. doi: 10.1093/nar/gkr832
- Ozaki, S., Kawakami, H., Nakamura, K., Fujikawa, N., Kagawa, W., Park, S. Y., et al. (2008). A common mechanism for the ATP-DnaA-dependent formation of open complexes at the replication origin. *J. Biol. Chem.* 283, 8351–8362. doi: 10.1074/jbc.M708684200
- Park, K., and Chatteraj, D. K. (2001). DnaA boxes in the P1 plasmid origin: the effect of their position on the directionality of replication and plasmid copy number. *J. Mol. Biol.* 310, 69–81. doi: 10.1006/jmbi.2001.4741
- Park, K., Mukhopadhyay, S., and Chatteraj, D. K. (1998). Requirements for and regulation of origin opening of plasmid P1. *J. Biol. Chem.* 273, 24906–24911. doi: 10.1074/jbc.273.38.24906
- Paulson, J., and Chatteraj, D. K. (2006). Origin inactivation in bacterial DNA replication control. *Mol. Microbiol.* 61, 9–15. doi: 10.1111/j.1365-2958.2006.05229.x
- Petojevic, T., Pesavento, J. J., Costa, A., Liang, J., Wang, Z., Berger, J. M., et al. (2015). Cdc45 (cell division cycle protein 45) guards the gate of the Eukaryote Replisome helicase stabilizing leading strand engagement. *Proc. Natl. Acad. Sci. U.S.A.* 112, E249–E258. doi: 10.1073/pnas.1422003112
- Polaczek, P., Kwan, K., and Campbell, J. L. (1998). Unwinding of the *Escherichia coli* origin of replication (*oriC*) can occur in the absence of initiation proteins but is stabilized by DnaA and histone-like proteins IHF or HU. *Plasmid* 39, 77–83. doi: 10.1006/plas.1997.1328
- Rahmouni, A. R., and Wells, R. D. (1989). Stabilization of Z DNA *in vivo* by localized supercoiling. *Science* 246, 358–363. doi: 10.1126/science.2678475
- Rajewska, M., Kowalczyk, L., Konopa, G., and Konieczny, I. (2008). Specific mutations within the AT-rich region of a plasmid replication origin affect either origin opening or helicase loading. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11134–11139. doi: 10.1073/pnas.0805662105
- Richardson, T. T., Harran, O., and Murray, H. (2016). The bacterial DnaA-trio replication origin element specifies single-stranded DNA initiator binding. *Nature* 534, 412–416. doi: 10.1038/nature17962
- Rozgaja, T. A., Grimwade, J. E., Iqbal, M., Czerwonka, C., Vora, M., and Leonard, A. C. (2011). Two oppositely oriented arrays of low-affinity recognition sites in *oriC* guide progressive binding of DnaA during *Escherichia coli* pre-RC assembly. *Mol. Microbiol.* 82, 475–488. doi: 10.1111/j.1365-2958.2011.07827.x
- Ryan, V. T., Grimwade, J. E., Camara, J. E., Crooke, E., and Leonard, A. C. (2004). *Escherichia coli* prereplication complex assembly is regulated by dynamic interplay among Fis, IHF and DnaA. *Mol. Microbiol.* 51, 1347–1359. doi: 10.1046/j.1365-2958.2003.03906.x
- Saito, H., and Uchida, H. (1977). Initiation of the DNA replication of bacteriophage lambda in *Escherichia coli* K12. *J. Mol. Biol.* 113, 1–25. doi: 10.1016/0022-2836(77)90038-9
- Saxena, R., Vasudevan, S., Patil, D., Ashoura, N., Grimwade, J. E., and Crooke, E. (2015). Nucleotide-induced conformational changes in *Escherichia coli* DnaA protein are required for bacterial ORC to Pre-RC conversion



- at the chromosomal origin. *Int. J. Mol. Sci.* 16, 27897–27911. doi: 10.3390/ijms161126064
- Schaper, S., and Messer, W. (1995). Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. *J. Biol. Chem.* 270, 17622–17626. doi: 10.1074/jbc.270.29.17622
- Schnos, M., and Inman, R. B. (1970). Position of branch points in replicating lambda DNA. *J. Mol. Biol.* 51, 61–73. doi: 10.1016/0022-2836(70)90270-6
- Schnos, M., Zahn, K., Inman, R. B., and Blattner, F. R. (1988). Initiation protein induced helix destabilization at the lambda origin: a prepriming step in DNA replication. *Cell* 52, 385–395. doi: 10.1016/S0092-8674(88)80031-X
- Scholefield, G., Errington, J., and Murray, H. (2012). Soj/ParA stalls DNA replication by inhibiting helix formation of the initiator protein DnaA. *EMBO J.* 31, 1542–1555. doi: 10.1038/emboj.2012.6
- Sharma, R., Kachroo, A., and Bastia, D. (2001). Mechanistic aspects of DnaA-RepA interaction as revealed by yeast forward and reverse two-hybrid analysis. *EMBO J.* 20, 4577–4587. doi: 10.1093/emboj/20.16.4577
- Shibata, T., Nishinaka, T., Mikawa, T., Aihara, H., Kurumizaka, H., Yokoyama, S., et al. (2001). Homologous genetic recombination as an intrinsic dynamic property of a DNA structure induced by RecA/Rad51-family proteins: a possible advantage of DNA over RNA as genomic material. *Proc. Natl. Acad. Sci. U.S.A.* 98, 8425–8432. doi: 10.1073/pnas.111005198
- Simmons, L. A., Felczak, M., and Kaguni, J. M. (2003). DnaA Protein of *Escherichia coli*: oligomerization at the *E. coli* chromosomal origin is required for initiation and involves specific N-terminal amino acids. *Mol. Microbiol.* 49, 849–858. doi: 10.1046/j.1365-2958.2003.03603.x
- Skarstad, K., Baker, T. A., and Kornberg, A. (1990). Strand separation required for initiation of replication at the chromosomal origin of *E. coli* is facilitated by a distant RNA–DNA hybrid. *EMBO J.* 9, 2341–2348.
- Skarstad, K., and Katayama, T. (2013). Regulating DNA replication in bacteria. *Cold Spring Harb. Perspect. Biol.* 5:a012922. doi: 10.1101/cshperspect.a012922
- Speck, C., and Messer, W. (2001). Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. *EMBO J.* 20, 1469–1476. doi: 10.1093/emboj/20.6.1469
- Stenzel, T. T., Patel, P., and Bastia, D. (1987). The integration host factor of *Escherichia coli* binds to bent DNA at the origin of replication of the plasmid pSC101. *Cell* 49, 709–717. doi: 10.1016/0092-8674(87)90547-2
- Stepankiw, N., Kaidow, A., Boye, E., and Bates, D. (2009). The right half of the *Escherichia coli* replication origin is not essential for viability, but facilitates multi-forked replication. *Mol. Microbiol.* 74, 467–479. doi: 10.1111/j.1365-2958.2009.06877.x
- Sutton, M. D., Carr, K. M., Vicente, M., and Kaguni, J. M. (1998). *Escherichia coli* DnaA protein. The N-terminal domain and loading of DnaB helicase at the *E. coli* chromosomal origin. *J. Biol. Chem.* 273, 34255–34262. doi: 10.1074/jbc.273.51.34255
- Szabowska, A., Pierchod, M., Wegrzyn, G., and Glinkowska, M. (2011). Coupling of transcription and replication machineries in lambda DNA replication initiation: evidence for direct interaction of *Escherichia coli* RNA polymerase and the lambdaO protein. *Nucleic Acids Res.* 39, 168–177. doi: 10.1093/nar/gkq752
- Teter, B., Goodman, S. D., and Galas, D. J. (2000). DNA bending and twisting properties of integration host factor determined by DNA cyclization. *Plasmid* 43, 73–84. doi: 10.1006/plas.1999.1443
- Thomas, R., and Bertani, L. E. (1964). On the control of the replication of temperate bacteriophages superinfecting immune hosts. *Virology* 24, 241–253.
- Tomizawa, J., Itoh, T., Selzer, G., and Som, T. (1981). Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. *Proc. Natl. Acad. Sci. U.S.A.* 78, 1421–1425. doi: 10.1073/pnas.78.3.1421
- Toukdarian, A., Helinski, D. R., and Perri, S. (1996). The plasmid RK2 initiation protein binds to the origin of replication as a monomer. *J. Biol. Chem.* 271, 7072–7078. doi: 10.1074/jbc.271.12.7072
- Tsurimoto, T., and Matsubara, K. (1981). Purified bacteriophage lambda O protein binds to four repeating sequences at the lambda replication origin. *Nucleic Acids Res.* 9, 1789–1799. doi: 10.1093/nar/9.8.1789
- Vlijm, R., v. d. Torre, J., and Dekker, C. (2015). Counterintuitive DNA sequence dependence in supercoiling-induced DNA melting. *PLoS ONE* 10:e0141576. doi: 10.1371/journal.pone.0141576
- Wegrzyn, K., Fuentes-Perez, M. E., Bury, K., Rajewska, M., Moreno-Herrero, F., and Konieczny, I. (2014). Sequence-specific interactions of Rep proteins with ssDNA in the AT-rich region of the plasmid replication origin. *Nucleic Acids Res.* 42, 7807–7818. doi: 10.1093/nar/gku453
- Weigel, C., Messer, W., Preiss, S., Welzcek, M., Morigen, and Boye, E. (2001). The sequence requirements for a functional *Escherichia coli* replication origin are different for the chromosome and a minichromosome. *Mol. Microbiol.* 40, 498–507. doi: 10.1046/j.1365-2958.2001.02409.x
- Wells, R. D., Larson, J. E., Grant, R. C., Shortle, B. E., and Cantor, C. R. (1970). Physicochemical studies on polydeoxyribonucleotides containing defined repeating nucleotide sequences. *J. Mol. Biol.* 54, 465–497. doi: 10.1016/0022-2836(70)90121-X
- Westmoreland, B. C., Szybalski, W., and Ris, H. (1969). Mapping of deletions and substitutions in heteroduplex DNA molecules of bacteriophage lambda by electron microscopy. *Science* 163, 1343–1348. doi: 10.1126/science.163.3873.1343
- Wickner, S., Gottesman, S., Skowrya, D., Hoskins, J., McKenney, K., and Maurizi, M. R. (1994). A molecular chaperone, ClpA, functions like DnaK and DnaJ. *Proc. Natl. Acad. Sci. U.S.A.* 91, 12218–12222. doi: 10.1073/pnas.91.25.12218
- Wickner, S., Hoskins, J., and McKenney, K. (1991). Monomerization of RepA dimers by heat shock proteins activates binding to DNA replication origin. *Proc. Natl. Acad. Sci. U.S.A.* 88, 7903–7907. doi: 10.1073/pnas.88.18.7903
- Woelker, B., and Messer, W. (1993). The structure of the initiation complex at the replication origin, *oriC*, of *Escherichia coli*. *Nucleic Acids Res.* 21, 5025–5033. doi: 10.1093/nar/21.22.5025
- Womble, D. D., and Rownd, R. H. (1986). Regulation of lambda dv plasmid DNA replication. A quantitative model for control of plasmid lambda dv replication in the bacterial cell division cycle. *J. Mol. Biol.* 191, 367–382. doi: 10.1016/0022-2836(86)90133-6
- Yardimci, H., and Walter, J. C. (2014). Prereplication-complex formation: a molecular double take? *Nat. Struct. Mol. Biol.* 21, 20–25. doi: 10.1038/nsmb.2738
- Zahn, K., and Blattner, F. R. (1987). Direct evidence for DNA bending at the lambda replication origin. *Science* 236, 416–422. doi: 10.1126/science.2951850
- Zorman, S., Seitz, H., Sclavi, B., and Strick, T. R. (2012). Topological characterization of the DnaA-oriC complex using single-molecule nanomanipulation. *Nucleic Acids Res.* 40, 7375–7383. doi: 10.1093/nar/gks371
- Zylicz, M., Ang, D., Liberek, K., and Georgopoulos, C. (1989). Initiation of lambda DNA replication with purified host- and bacteriophage-encoded proteins: the role of the DnaK, DnaJ and GrpE heat shock proteins. *EMBO J.* 8, 1601–1608.
- Zzaman, S., Abhyankar, M. M., and Bastia, D. (2004a). Reconstitution of F factor DNA replication *in vitro* with purified proteins. *J. Biol. Chem.* 279, 17404–17410. doi: 10.1074/jbc.M400021200
- Zzaman, S., and Bastia, D. (2005). Oligomeric initiator protein-mediated DNA looping negatively regulates plasmid replication *in vitro* by preventing origin melting. *Mol. Cell* 20, 833–843. doi: 10.1016/j.molcel.2005.10.037
- Zzaman, S., Reddy, J. M., and Bastia, D. (2004b). The DnaK-DnaJ-GrpE chaperone system activates inert wild type pi initiator protein of R6K into a form active in replication initiation. *J. Biol. Chem.* 279, 50886–50894. doi: 10.1074/jbc.M407531200

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Jha, Ramachandran and Chatteraj. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Multiple DNA Binding Proteins Contribute to Timing of Chromosome Replication in *E. coli*

Leise Riber<sup>†</sup>, Jakob Frimodt-Møller<sup>†</sup>, Godefroid Charbon and Anders Løbner-Olesen<sup>\*</sup>

Section for Functional Genomics and Center for Bacterial Stress Response and Persistence, Department of Biology, University of Copenhagen, Copenhagen, Denmark

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas Medical  
Branch-Galveston, USA

### Reviewed by:

Dhruba Chattoraj,  
National Institutes of Health, USA  
Martin Marinus,  
University of Massachusetts, USA

### \*Correspondence:

Anders Løbner-Olesen  
lobner@bio.ku.dk

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 24 May 2016

**Accepted:** 14 June 2016

**Published:** 28 June 2016

### Citation:

Riber L, Frimodt-Møller J, Charbon G  
and Løbner-Olesen A (2016) Multiple  
DNA Binding Proteins Contribute to  
Timing of Chromosome Replication in  
*E. coli*. *Front. Mol. Biosci.* 3:29.  
doi: 10.3389/fmolb.2016.00029

Chromosome replication in *Escherichia coli* is initiated from a single origin, *oriC*. Initiation involves a number of DNA binding proteins, but only DnaA is essential and specific for the initiation process. DnaA is an AAA+ protein that binds both ATP and ADP with similar high affinities. DnaA associated with either ATP or ADP binds to a set of strong DnaA binding sites in *oriC*, whereas only DnaA<sup>ATP</sup> is capable of binding additional and weaker sites to promote initiation. Additional DNA binding proteins act to ensure that initiation occurs timely by affecting either the cellular mass at which DNA replication is initiated, or the time window in which all origins present in a single cell are initiated, i.e. initiation synchrony, or both. Overall, these DNA binding proteins modulate the initiation frequency from *oriC* by: (i) binding directly to *oriC* to affect DnaA binding, (ii) altering the DNA topology in or around *oriC*, (iii) altering the nucleotide bound status of DnaA by interacting with non-coding chromosomal sequences, distant from *oriC*, that are important for DnaA activity. Thus, although DnaA is the key protein for initiation of replication, other DNA-binding proteins act not only on *oriC* for modulation of its activity but also at additional regulatory sites to control the nucleotide bound status of DnaA. Here we review the contribution of key DNA binding proteins to the tight regulation of chromosome replication in *E. coli* cells.

**Keywords:** *E. coli*, chromosome replication, DNA binding proteins, cell mass, initiation synchrony

## TIMING OF INITIATION OF CHROMOSOME REPLICATION IN *E. COLI*

Chromosome replication in *Escherichia coli* is initiated from a single replication origin, *oriC*. The *oriC*-encoded structural and functional instructions for initiation are well-described (Leonard and Mechali, 2013; Skarstad and Katayama, 2013). In brief, the minimal *oriC* contains two functional regions: the Duplex Unwinding Element (DUE), which comprises three AT-rich repeat sequences of each 13 bp, and the flanking DnaA Assembly Region (DAR) (Figure 1; Mott and Berger, 2007; Ozaki and Katayama, 2012). DnaA is the initiator protein responsible for DUE opening and for the recruitment of replisome components and is the only protein that is both essential and specific for the initiation process (Kaguni, 2011; Leonard and Grimwade, 2011). DnaA belongs to the AAA+ proteins (ATPases Associated with diverse Activities) and can bind both ATP and ADP with similar high affinities (Sekimizu et al., 1987). The DAR region contains high affinity DnaA Boxes (R1, R4, and R2) that bind both DnaA<sup>ATP</sup> and DnaA<sup>ADP</sup>, along with multiple low affinity sites (R3, R5/M, I1, I2, I3, C1, C2, C3,  $\tau$ 1, and  $\tau$ 2) that bind DnaA<sup>ATP</sup>.

(McGarry et al., 2004; Kawakami et al., 2005; Rozgaja et al., 2011). The DAR region also contains recognition sequences for two additional DNA binding proteins; IHF and Fis (Figure 1; Polaczek, 1990; Gille et al., 1991).

Throughout most of the cell cycle *oriC* is bound by DnaA located at R1, R2, and R4. This origin recognition complex (ORC) serves dual purposes in setting the stage for proper orisome assembly and preventing premature DNA unwinding. The ratio of DnaA<sup>ATP</sup> to DnaA<sup>ADP</sup> varies through the cell cycle and the peak at about 70–80% DnaA<sup>ATP</sup> coincides with replication initiation (Kurokawa et al., 1999). In the current model for orisome formation, two converging DnaA<sup>ATP</sup> filaments are formed (Rozgaja et al., 2011). One filament originates from R4 and grows leftward. This R4-filament displaces Fis from its binding site next to R2, which allows IHF to bind its recognition sequence next to R1. IHF bends the DNA 180° thereby bringing R1 in proximity of R5 and allows for the formation of the rightward filament responsible for duplex opening at the DUE, DnaC assisted helicase loading and assembly of the replisome (Leonard and Grimwade, 2011, 2015; Ozaki et al., 2012). Following initiation, DnaA<sup>ATP</sup> is converted to DnaA<sup>ADP</sup> primarily by a process called regulatory inactivation of DnaA (RIDA), which is dependent on the Hda protein bound to ADP and the DNA-loaded  $\beta$ -clamp of the polymerase III holoenzyme (Kato and Katayama, 2001), and by the less efficient *datA*-dependent DnaA<sup>ATP</sup> hydrolysis (DDAH). DDAH takes place at *datA* and is dependent on IHF (Figure 1; Kasho and Katayama, 2013).

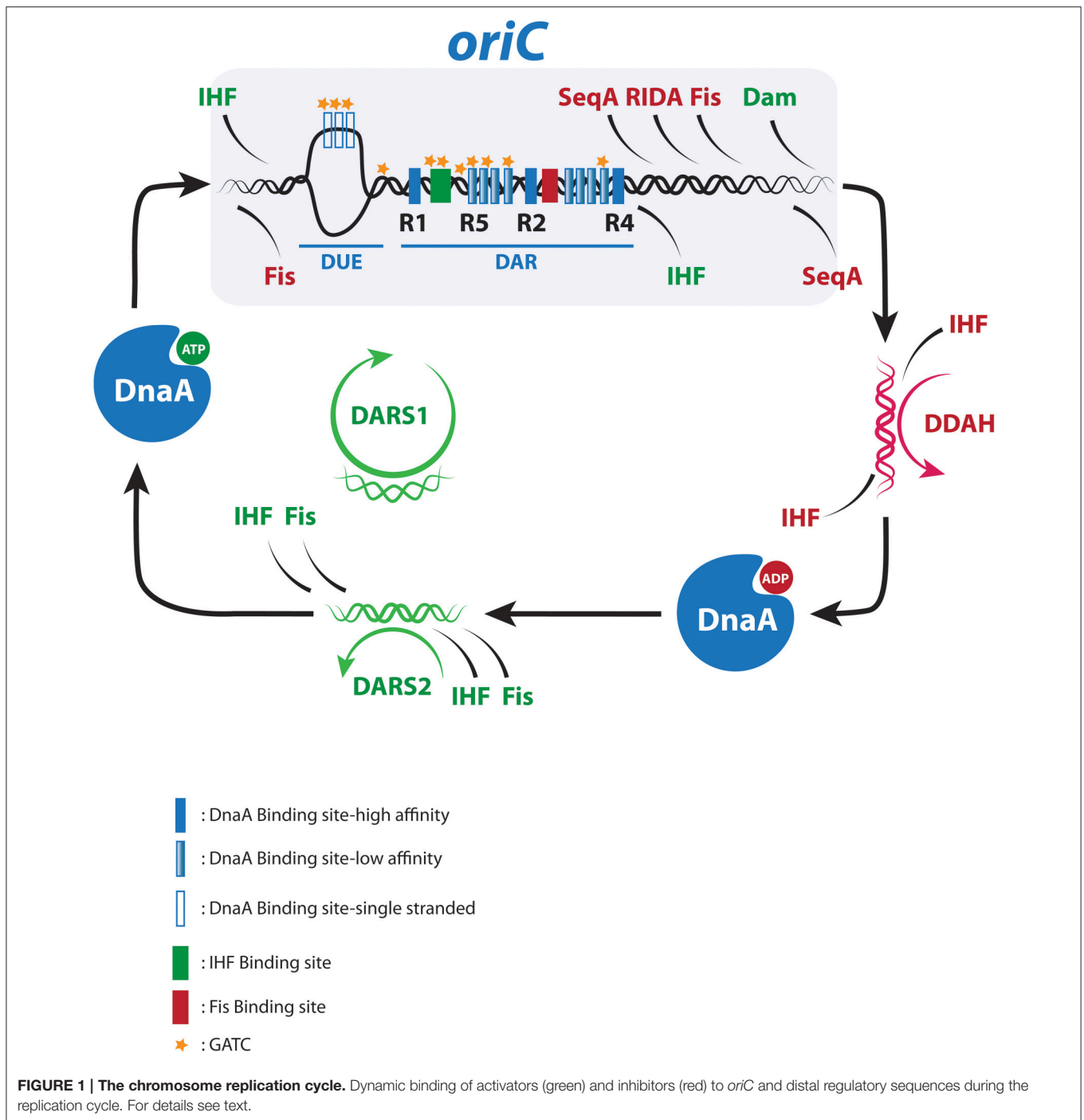
## Coordination of Initiations with Cell Mass Increase

A long standing observation is that initiation of chromosome replication occurs when a certain cellular mass per origin, the initiation mass, is reached (Donachie, 1968; Hill et al., 2012). This coupling of replication initiation to cell growth depends on the DnaA protein. Earlier studies indicate that accumulation of DnaA protein sets the time of initiation in the cell cycle especially around or below wild-type level (Løbner-Olesen et al., 1989). On the other hand, a coordinated increase in DnaA<sup>ATP</sup> and DnaA<sup>ADP</sup> does not significantly increase initiation (Kurokawa et al., 1999; Flatten et al., 2015), suggesting that accumulation of DnaA<sup>ATP</sup> is insufficient to trigger initiation. However, in the absence of RIDA, where DnaA is mainly ATP bound, a modest increase in DnaA<sup>ATP</sup> level leads to excessive initiations from *oriC* (Riber et al., 2006; Fujimitsu et al., 2008), as does expression of a DnaA mutant protein insensitive to RIDA (Simmons et al., 2004). Together, this indeed suggests that accumulation of DnaA<sup>ATP</sup> triggers initiation, whereas this effect can be offset by a similar increase in DnaA<sup>ADP</sup> (Donachie and Blakely, 2003). The participation of DnaA<sup>ADP</sup> in orisome formation remains unclear (Leonard and Grimwade, 2015), but the above observations suggest that it affects initiation negatively. Overall, accumulation of DnaA protein during steady-state growth, along with the cell cycle specific peak in DnaA<sup>ATP</sup>/DnaA<sup>ADP</sup> ratio, determines the onset of initiation with little variation between individual cells.

## Coordination of initiations within a Single Cell

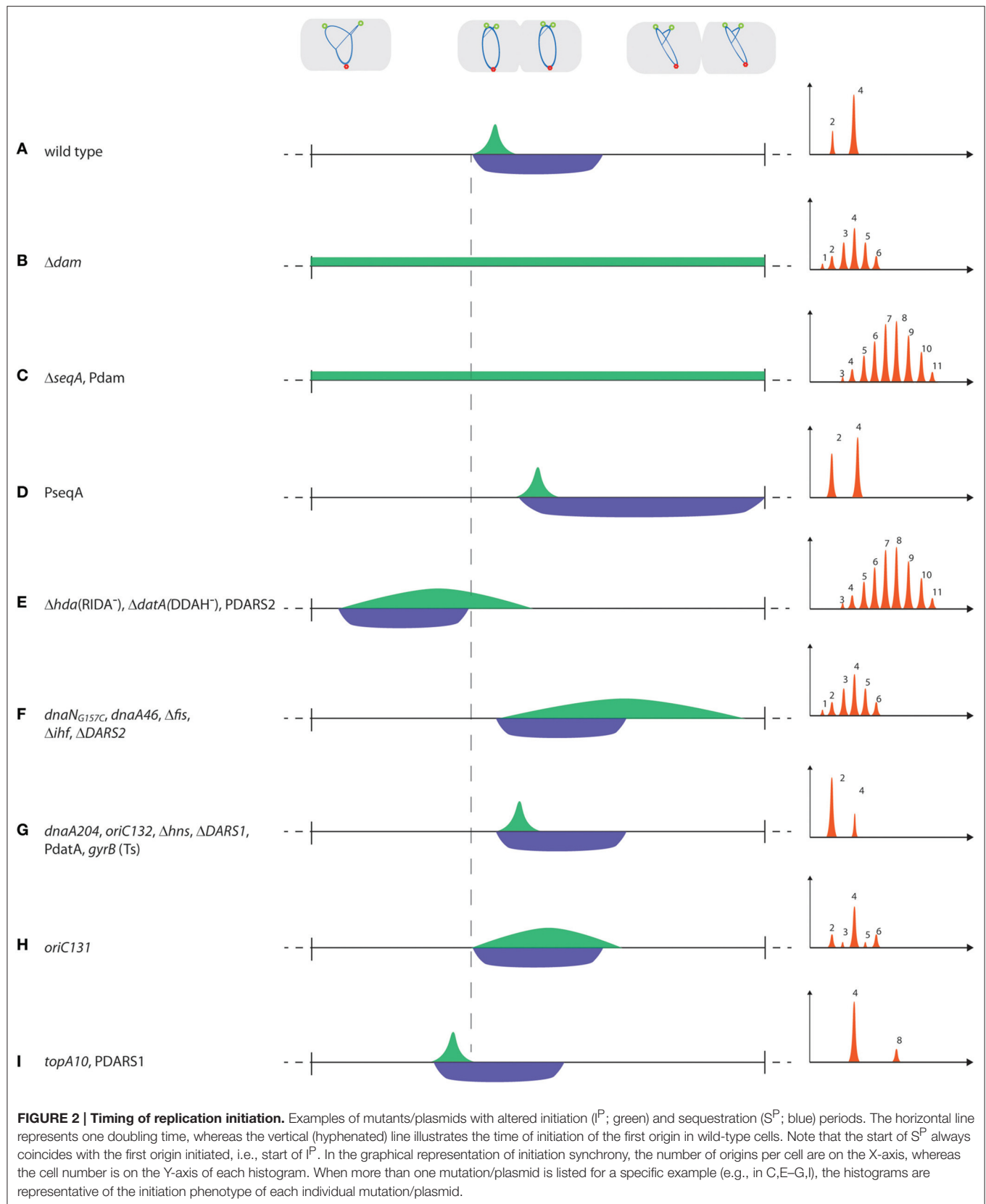
In individual cells, initiation at all origins occurs within approximately 1/10 of the doubling time (Initiation period,  $I^P$ ; Figure 2A). Rapidly growing cells with overlapping replication cycles therefore predominantly contain  $2^n$  ( $n = 1, 2, 3$ ) copies of *oriC*, referred to as initiation synchrony (Skarstad et al., 1986). Initiation synchrony depends on the immediate inactivation of newly replicated origins by sequestration. *oriC* contain 11 copies of the sequence GATC that are methylated by Dam methyltransferase and bound, i.e., sequestered, by SeqA when hemimethylated. Sequestration prevents DnaA binding to its weak sites in *oriC* (Nievera et al., 2006) for approximately 1/3 generation (Sequestration period,  $S^P$ ; Figure 2A) and serves to keep track of which origins have been initiated (Boye and Løbner-Olesen, 1990; Campbell and Kleckner, 1990; Lu et al., 1994). The ability to initiate all origins in synchrony could result from maintaining a high DnaA<sup>ATP</sup> level throughout  $I^P$ . Alternatively the first origin initiated may release its DnaA<sup>ATP</sup> to assist in triggering successive initiations at remaining origins in a cascade-like manner to ensure that free DnaA<sup>ATP</sup> increases through  $I^P$  and enforces synchrony (Løbner-Olesen et al., 1994). These models predict different outcomes for sequestration deficient cells. A high DnaA<sup>ATP</sup> level throughout  $I^P$  would result in re-initiation(s) within  $I^P$ , asynchrony and overinitiation. The cascade model predicts a delay between successive initiations due to newly initiated origins competing with old origins for a limited amount of DnaA<sup>ATP</sup>. The initiation frequency would be directly proportional with accumulation of DnaA<sup>ATP</sup> resulting in asynchrony but an unchanged overall initiation frequency, which is in accordance with experimental observations for Dam deficient cells (Boye and Løbner-Olesen, 1990; Løbner-Olesen et al., 1994).

Synchrony is only observed when  $I^P < S^P$  (Figure 2A). In cells with aberrant timing of initiation, the  $I^P$  and  $S^P$  periods change, i.e., either start earlier in the cell cycle at a decreased initiation mass, i.e., overinitiation, or are delayed with an increased initiation mass, i.e., underinitiation. Alternatively, the duration of  $I^P$  and  $S^P$  may change relative to each other, and when  $I^P > S^P$ , newly initiated origins, released from sequestration, compete with origins not yet initiated. Consequently, some origins are re-initiated while others are not initiated at all, leading to loss of synchrony (Olsson et al., 2003; Skarstad and Løbner-Olesen, 2003). This is exemplified by *dam* mutants without a sequestration period that initiate throughout the cell cycle (Figure 2B; Boye and Løbner-Olesen, 1990; Lu et al., 1994). *seqA* mutants are also asynchronous but have a higher origin concentration, possibly because DnaA is increased, relative to *dam* mutants (Figure 2C; Campbell and Kleckner, 1990; von Freiesleben et al., 1994). Increased levels of Dam will, due to faster re-methylation rates, reduce  $S^P$  and when this becomes shorter than  $I^P$ , asynchrony follows (Figure 2C; Boye and Løbner-Olesen, 1990; von Freiesleben et al., 2000a). Excess SeqA protein delays initiation, prolongs the sequestration period but does not affect synchrony (Figure 2D; Bach et al., 2003; Charbon



et al., 2011). During sequestration the activity of DnaA is lowered by RIDA and DDAH. RIDA is presumably accelerated by generation of new replication forks at initiation and hence more DNA loaded  $\beta$ -clamps (Moolman et al., 2014). Similarly, DDAH is increased shortly after initiation when the *data* locus is duplicated and together they ensure a post-initiation decrease in the  $\text{DnaA}^{\text{ATP}}/\text{DnaA}^{\text{ADP}}$  ratio (Figure 1). RIDA ( $\Delta hda$ ) and to a lesser degree DDAH ( $\Delta data$ ) deficient cells fail to lower the ratio of  $\text{DnaA}^{\text{ATP}}/\text{DnaA}^{\text{ADP}}$  to prevent re-initiation following

sequestration. This results in asynchrony and early initiation at a reduced cell mass (Figure 2E; Kitagawa et al., 1998; Fujimitsu et al., 2008; Kasho and Katayama, 2013). On the other hand, the *dnaN*<sub>G157C</sub> mutant, which is more active in RIDA (*dnaN* encodes the  $\beta$ -clamp), or extra copies of *data*, results in delayed initiation and, for *dnaN*<sub>G157C</sub> cells, also produces asynchrony (Figures 2F,G; Morigen et al., 2001; Gon et al., 2006; Charbon et al., 2011; Johnsen et al., 2011). During sequestration, the overall level of free DnaA is reduced by titration (Hansen et al., 1991;





Kitagawa et al., 1996, 1998; Ogawa et al., 2002) and by arrest of *de novo* DnaA synthesis (Campbell and Kleckner, 1990).

## MODULATION OF TIMING OF REPLICATION INITIATION BY DNA BINDING PROTEINS

Several DNA binding proteins affect either the cell mass at initiation, the initiation synchrony, or both. These proteins either bind specifically to *oriC* to affect DnaA binding, non-specifically to DNA to alter *oriC* topology, or they bind sequences important for the nucleotide bound status of DnaA.

### Proteins That Specifically Interact with *oriC* Prior to Initiation

The most important protein to interact with *oriC* prior to initiation is DnaA. Mutations in DnaA that affect nucleotide binding, such as *dnaA46*, are presumably somewhat deficient in formation of DnaA multimers on *oriC*, which results in delayed initiation and a prolonged initiation period (Skarstad et al., 1988; Boye et al., 1996). As sequestration remains unchanged ( $I^P > S^P$ ), *dnaA46* cells are asynchronous (**Figure 2F**; Skarstad and Løbner-Olesen, 2003). Mutations in DnaA that affect DNA binding, but not nucleotide binding (e.g., *dnaA204*), lead to late but synchronous initiation (**Figure 2G**; Skarstad et al., 1988; Torheim et al., 2000). The ability to form DnaA<sup>ATP</sup> filaments on *oriC* therefore seems of greater importance for initiation synchrony than a tight anchoring to DnaA binding sites.

Conflicting data exist on the role of Fis for timing of initiation. Binding Fis to *oriC* *in vitro* is reported to either inhibit initiation of replication by inducing conformational changes at *oriC* that prevent orisome formation (Wold et al., 1996; Ryan et al., 2002, 2004), or have no effect on initiation (Margulies and Kaguni, 1998). Cells with a mutated primary Fis binding in *oriC* (*oriC131*) have an origin concentration similar to wild-type (**Figure 2H**; Weigel et al., 2001; Riber et al., 2009; Flatten and Skarstad, 2013). Fis-deficient cells, on the other hand, have a lowered origin concentration (Flatten and Skarstad, 2013; Kasho et al., 2014), suggesting that initiation is delayed (**Figure 2F**). However, because Fis affects multiple cellular processes due to its involvement in DNA organization one should be careful in assessing its role in initiation solely based on the behavior of Fis-deficient cells. Both Fis deficiency or loss of its primary *oriC* binding site result in initiation asynchrony (**Figures 2F,H**; Riber et al., 2009; Flatten and Skarstad, 2013), indicating that these cells are deficient for proper orisome assembly and/or for preventing premature DNA unwinding. The role of IHF in replication timing is less controversial. An *oriC* mutant with a disrupted IHF binding site (*oriC132*) is somewhat deficient in orisome formation and has delayed but synchronous initiation (**Figure 2G**; Weigel et al., 2001; Skarstad and Løbner-Olesen, 2003; Riber et al., 2009). *ihf* mutant cells also initiate replication at an increased mass per origin consistent with a stimulatory role of IHF on initiation. Cells deficient in IHF are on the other hand asynchronous (**Figure 2F**; von Freiesleben et al., 2000b).

This is in agreement with an additional role of IHF for DnaA<sup>ATP</sup> generation at *DARS2* (see below).

A number of proteins negatively regulate initiation of replication *in vitro*. These include ArcA that binds to 13 mer AT rich repeats, to DnaA box R1 and to the IHF binding site in *oriC*, and IciA that binds to 13-mer AT-rich repeats in *oriC* (Hwang and Kornberg, 1990; Lee et al., 2001). The impact of ArcA and IciA on replication initiation *in vivo* is modest (Nystrom et al., 1996) or not known, respectively. The stationary-phase induced CspD protein binds ssDNA to inhibit replication initiation and elongation *in vitro*, whereas no *in vivo* data are available (Yamanaka et al., 2001). Upon association with Cnu and/or Hha, H-NS (see below) binds to a specific sequence in *oriC* that overlaps DnaA box R5 (Kim et al., 2005; Yun et al., 2012). Cells deficient in Cnu and/or Hha are, however, similar to wild-type (Kim et al., 2005). Finally, the protein Rob binds to a single site in *oriC* *in vitro*, but does not affect initiation *in vivo* (Skarstad et al., 1993).

### DNA Binding Proteins That Affect Topology of *oriC*

In *E. coli* the genomic DNA is mostly negatively supercoiled (Wang et al., 2013). Unconstrained supercoiling of *oriC* contributes to the ease of duplex opening and is determined by transcription (not covered here; for review see Magnan and Bates, 2015) along with the actions of topoisomerase I and DNA gyrase enzymes (Wu et al., 1988). Mutations in topoisomerase I, which removes negative supercoils, result in initiation at a slightly reduced mass while synchrony is maintained (**Figure 2I**; von Freiesleben and Rasmussen, 1992; Olsson et al., 2003). Conversely, temperature sensitive *gyrB* mutant cells, with moderately reduced negative superhelicity of the chromosome, enhance the temperature sensitivity of a *dnaA46* mutant (Filutowicz, 1980) and show delayed synchronous initiations (**Figure 2G**; von Freiesleben and Rasmussen, 1991; Usongo et al., 2013). This suggests that initiation is facilitated by an increase in negative superhelicity of the chromosome. However, *topA-gyr* mutations influence chromosome segregation, R-loop formation and possibly induce stable DNA replication independent of *oriC* (Usongo et al., 2013, 2016) making it difficult to assess the effect of large changes in overall supercoiling on replication initiation. *In vivo*, nucleoid-associated proteins (NAPs; Dillon and Dorman, 2010), such as IHF, Fis, H-NS, HU, and MukFEB constrain negative supercoils to condense the chromosome and could therefore affect initiation of chromosome replication (Badrinarayanan et al., 2015; Lal et al., 2016). H-NS deficient cells have an increased negative superhelicity of the genome (Mojica and Higgins, 1997; Hardy and Cozzarelli, 2005). Yet, genetic evidence suggests that loss of H-NS hampers initiation (Katayama et al., 1996), and H-NS deficient cells initiate replication in synchrony at an increased cell mass (**Figure 2G**; Kaidow et al., 1995; Atlung and Hansen, 2002). The HU protein can substitute for IHF in DnaA-mediated unwinding of *oriC* *in vitro* (Hwang and Kornberg, 1992) although their mechanisms of action differ (Ryan et al., 2002). *In vivo*, genetic evidence suggests that loss of HU stimulates initiation despite decreased negative

supercoiling (Louarn et al., 1984). Loss of MukB, involved in condensation of the bacterial chromosome (Hiraga et al., 1989; Cui et al., 2008), results in reduced negative supercoiling (Weitao et al., 2000), but initiations remain synchronous (Weitao et al., 1999). It is not known whether MukB affects the initiation mass. Finally, the starvation-induced NAP, Dps, binds non-specifically to *oriC*, and interacts with the N-terminus of DnaA, inhibiting DNA unwinding *in vitro*. Loss of Dps does not result in loss of synchrony, but increases the cellular origin content somewhat (Chodavarapu et al., 2008). In summary, it seems that NAPs modulate replication initiation but that the effect is not solely mediated through an effect on DNA supercoiling.

## GETTING READY FOR THE NEXT ROUND OF REPLICATION

At later cell cycle stages DnaA<sup>ATP</sup> is regenerated for the next initiation to take place (Figure 1). *E. coli* can rejuvenate DnaA<sup>ADP</sup> to DnaA<sup>ATP</sup> in a process assisted by acidic phospholipids (Saxena et al., 2013) or at two non-coding chromosomal sites called *DARS1* and *DARS2* (Fujimitsu et al., 2009). *DARS1* and *DARS2* are located in each replicore halfway between *oriC* and *terC*, and are duplicated after the end of sequestration. Multiple DnaA<sup>ADP</sup> molecules form complexes with *DARS* to facilitate release of ADP resulting in apo-DnaA, which will primarily rebind ATP as this is more abundant than ADP within the cell (Petersen and Møller, 2000).

*DARS1* is not known to be regulated by any proteins, whereas rejuvenation at the more efficient *DARS2* locus is dependent on binding of both IHF and Fis (Kasho et al., 2014). While Fis binds *DARS2* throughout the cell cycle, IHF provides cell cycle specificity to *DARS2* activity by only binding and activating *DARS2* immediately prior to initiation to ensure an increase in DnaA<sup>ATP</sup> level (Fujimitsu et al., 2009; Kasho et al., 2014). Extra copies of *DARS1* or *DARS2* will increase the overall DnaA<sup>ATP</sup> level, which results in early initiation (Figures 2E,I) and for *DARS2* also extends I<sup>P</sup>, thereby resembling RIDA deficiency (Figure 2E; Fujimitsu et al., 2009; Charbon et al., 2011). Deletion of *DARS1*, *DARS2*, or both reduces the ability to reactivate DnaA for new initiations in the following cell cycle and results in delayed initiation (Figures 2F,G; Fujimitsu et al., 2009; Kasho et al., 2014; Frimodt-Møller et al., 2015). Loss of *DARS2* also increases the relative duration of the initiation period, leading to initiation asynchrony (Figure 2F; Fujimitsu et al., 2009; Frimodt-Møller et al., 2015). This suggests that both *DARS1* and *DARS2* are important for coupling initiation to cell mass increase, whereas only the cell-cycle regulated *DARS2* is crucial for maintaining initiation synchrony.

## REFERENCES

Atlung, T., and Hansen, F. G. (2002). Effect of different concentrations of H-NS protein on chromosome replication and the cell cycle in *Escherichia coli*. *J. Bacteriol.* 184, 1843–1850. doi: 10.1128/JB.184.7.1843-1850.2002

## CONCLUDING REMARKS

Overall, timing of chromosome replication in *E. coli* takes place at least at two levels. First, initiation of replication is tightly coupled to cell mass increase through accumulation of DnaA<sup>ATP</sup>. Second, synchrony of initiations within the single cell is not necessarily connected to initiation mass but results from each origin being simultaneously initiated only once per generation, with asynchrony originating from failure to obey this once-and-only-once rule. DnaA remains the only replication protein solely required for initiation at *oriC*, but additional proteins act on *oriC* and elsewhere to assist in coupling of replication to cell growth and synchrony. In particular IHF and Fis display complex functions, targeting several regulatory sites. IHF has a dual role on replication initiation, acting both positively (i.e., binding to *DARS2* and *oriC*) and negatively (i.e., binding to *datA*). Also, IHF binds *oriC* at the pre-initiation stage and interacts with *datA* and *DARS2* following initiation. Binding of IHF to these regions is suggested to be temporally regulated so that IHF binds to *oriC*, to *datA* and to *DARS2* in a successive manner during cell cycle progression (Kasho and Katayama, 2013; Kasho et al., 2014). *In vivo*, *ihf* mutants display an initiation-compromised phenotype, indicating that the overall role of IHF on initiation of replication appears positive.

For a long time, the contribution of Fis in initiation regulation has been questioned. Recent studies do, however, suggest an overall positive role of Fis in replication initiation (Flatten and Skarstad, 2013; Kasho et al., 2014), which likely results from ensuring ordered orisome formation by preventing premature IHF binding and DNA unwinding (Leonard and Grimwade, 2015) and from stimulating DnaA<sup>ATP</sup> rejuvenation at *DARS2*. As the cellular Fis level depends on both growth-rate and phase, it could adjust chromosome replication to the bacterial growth rate through its activity on *DARS2* (Kasho et al., 2014).

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

This research was part of the Center for Bacterial Stress Response and Persistence (BASP) funded by a grant from the Danish National Research Foundation (DNRF120) and by a grant from the Novo Nordisk Foundation.

Bach, T., Krekling, M. A., and Skarstad, K. (2003). Excess SeqA prolongs sequestration of *oriC* and delays nucleoid segregation and cell division. *EMBO J.* 22, 315–323. doi: 10.1093/emboj/cdg020

Badrinarayanan, A., Le, T. B., and Laub, M. T. (2015). Bacterial chromosome organization and segregation. *Annu. Rev. Cell Dev. Biol.* 31, 171–199. doi: 10.1146/annurev-cellbio-100814-125211

- Boye, E., and Løbner-Olesen, A. (1990). The role of Dam methyltransferase in the control of DNA replication in *E. coli*. *Cell* 62, 981–989. doi: 10.1016/0092-8674(90)90272-G
- Boye, E., Stokke, T., Kleckner, N., and Skarstad, K. (1996). Coordinating DNA replication initiation with cell growth: differential roles for DnaA and SeqA proteins. *Proc. Natl. Acad. Sci. U.S.A.* 93, 12206–12211. doi: 10.1073/pnas.93.22.12206
- Campbell, J. L., and Kleckner, N. (1990). *E. coli* *oriC* and the *dnaA* gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. *Cell* 62, 967–979. doi: 10.1016/0092-8674(90)90271-F
- Charbon, G., Riber, L., Cohen, M., Skovgaard, O., Fujimitsu, K., Katayama, T., et al. (2011). Suppressors of DnaA(ATP) imposed overinitiation in *Escherichia coli*. *Mol. Microbiol.* 79, 914–928. doi: 10.1111/j.1365-2958.2010.07493.x
- Chodavarapu, S., Gomez, R., Vicente, M., and Kaguni, J. M. (2008). *Escherichia coli* Dps interacts with DnaA protein to impede initiation: a model of adaptive mutation. *Mol. Microbiol.* 67, 1331–1346. doi: 10.1111/j.1365-2958.2008.06127.x
- Cui, Y., Petruschenko, Z. M., and Rybenkov, V. V. (2008). MukB acts as a macromolecular clamp in DNA condensation. *Nat. Struct. Mol. Biol.* 15, 411–418. doi: 10.1038/nsmb.1410
- Dillon, S. C., and Dorman, C. J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* 8, 185–195. doi: 10.1038/nrmicro2261
- Donachie, W. D. (1968). Relationship between cell size and time of initiation of DNA replication. *Nature* 219, 1077–1079. doi: 10.1038/2191077a0
- Donachie, W. D., and Blakely, G. W. (2003). Coupling the initiation of chromosome replication to cell size in *Escherichia coli*. *Curr. Opin. Microbiol.* 6, 146–150. doi: 10.1016/S1369-5274(03)00026-2
- Filutowicz, M. (1980). Requirement of DNA gyrase for the initiation of chromosome replication in *Escherichia coli* K-12. *Mol. Gen. Genet.* 177, 301–309. doi: 10.1007/BF00267443
- Flatten, I., Fossum-Raunehaug, S., Taipale, R., Martinsen, S., and Skarstad, K. (2015). The DnaA protein is not the limiting factor for initiation of replication in *Escherichia coli*. *PLoS Genet.* 11:e1005276. doi: 10.1371/journal.pgen.1005276
- Flatten, I., and Skarstad, K. (2013). The Fis protein has a stimulating role in initiation of replication in *Escherichia coli* in vivo. *PLoS ONE* 8:e83562. doi: 10.1371/journal.pone.0083562
- Frimodt-Møller, J., Charbon, G., Krogfelt, K. A., and Løbner-Olesen, A. (2015). Control regions for chromosome replication are conserved with respect to sequence and location among *Escherichia coli* strains. *Front. Microbiol.* 6:1011. doi: 10.3389/fmicb.2015.01011
- Fujimitsu, K., Senriuchi, T., and Katayama, T. (2009). Specific genomic sequences of *E. coli* promote replicational initiation by directly reactivating ADP-DnaA. *Genes Dev.* 23, 1221–1233. doi: 10.1101/gad.1775809
- Fujimitsu, K., Suetsugu, M., Yamaguchi, Y., Mazda, K., Fu, N., Kawakami, H., et al. (2008). Modes of overinitiation, dnaA gene expression, and inhibition of cell division in a novel cold-sensitive hda mutant of *Escherichia coli*. *J. Bacteriol.* 190, 5368–5381. doi: 10.1128/JB.00044-08
- Gille, H., Egan, J. B., Roth, A., and Messer, W. (1991). The FIS protein binds and bends the origin chromosomal DNA replication, *oriC*, of *Escherichia coli*. *Nucleic Acids Res.* 19, 4167–4172. doi: 10.1093/nar/19.15.4167
- Gon, S., Camara, J. E., Klungsoyr, H. K., Crooke, E., Skarstad, K., and Beckwith, J. (2006). A novel regulatory mechanism couples deoxyribonucleotide synthesis and DNA replication in *Escherichia coli*. *EMBO J.* 25, 1137–1147. doi: 10.1038/sj.emboj.7600990
- Hansen, F. G., Christensen, B. B., and Atlung, T. (1991). The Initiator titration model: computer simulation of chromosome and minichromosome control. *Res. Microbiol.* 142, 161–167. doi: 10.1016/0923-2508(91)90025-6
- Hardy, C. D., and Cozzarelli, N. R. (2005). A genetic selection for supercoiling mutants of *Escherichia coli* reveals proteins implicated in chromosome structure. *Mol. Microbiol.* 57, 1636–1652. doi: 10.1111/j.1365-2958.2005.04799.x
- Hill, N. S., Kadoya, R., Chatteraj, D. K., and Levin, P. A. (2012). Cell size and the initiation of DNA replication in bacteria. *PLoS Genet* 8:e1002549. doi: 10.1371/journal.pgen.1002549
- Hiraga, S., Niki, H., Ogura, T., Ichinose, C., Mori, H., Ezaki, B., et al. (1989). Chromosome partitioning in *Escherichia coli*: novel mutants producing Anucleate Cells. *J. Bacteriol.* 171, 1496–1505.
- Hwang, D. S., and Kornberg, A. (1990). A novel protein binds a key origin sequence to block replication of an *E. coli* minichromosome. *Cell* 63, 325–331. doi: 10.1016/0092-8674(90)90165-B
- Hwang, D. S., and Kornberg, A. (1992). Opening of the replication origin of *Escherichia coli* by DnaA protein with protein HU or IHF. *J. Biol. Chem.* 267, 23083–23086.
- Johnsen, L., Flatten, I., Morigen, Dalhus, B., Bjørås, M., Waldminghaus, T., and Skarstad, K. (2011). The G157C mutation in the *Escherichia coli* sliding clamp specifically affects initiation of replication. *Mol. Microbiol.* 79, 433–446. doi: 10.1111/j.1365-2958.2010.07453.x
- Kaguni, J. M. (2011). Replication initiation at the *Escherichia coli* chromosomal origin. *Curr. Opin. Chem. Biol.* 15, 606–613. doi: 10.1016/j.cbpa.2011.07.016
- Kaidow, A., Wachi, M., Nakamura, J., Magae, J., and Nagai, K. (1995). Anucleate cell production by *Escherichia coli* delta hns mutant lacking a histone-like protein, H-NS. *J. Bacteriol.* 177, 3589–3592.
- Kasho, K., Fujimitsu, K., Matoba, T., Oshima, T., and Katayama, T. (2014). Timely binding of IHF and Fis to DARS2 regulates ATP-DnaA production and replication initiation. *Nucleic Acids Res.* 42, 13134–13149. doi: 10.1093/nar/gku1051
- Kasho, K., and Katayama, T. (2013). DnaA binding locus datA promotes DnaA-ATP hydrolysis to enable cell cycle-coordinated replication initiation. *Proc. Natl. Acad. Sci. U.S.A.* 110, 936–941. doi: 10.1073/pnas.1212070110
- Katayama, T., Takata, M., and Sekimizu, K. (1996). The nucleoid protein H-NS facilitates chromosome DNA replication in *Escherichia coli* dnaA mutants. *J. Bacteriol.* 178, 5790–5792.
- Kato, J., and Katayama, T. (2001). Hda, a novel DnaA-related protein, regulates the replication cycle in *Escherichia coli*. *EMBO J.* 20, 4253–4262. doi: 10.1093/emboj/20.15.4253
- Kawakami, H., Keyamura, K., and Katayama, T. (2005). Formation of an ATP-DnaA-specific initiation complex requires DnaA Arginine 285, a conserved motif in the AAA+ protein family. *J. Biol. Chem.* 280, 27420–27430. doi: 10.1074/jbc.M502764200
- Kim, M. S., Bae, S. H., Yun, S. H., Lee, H. J., Ji, S. C., Lee, J. H., et al. (2005). Cnu, a novel oriC-binding protein of *Escherichia coli*. *J. Bacteriol.* 187, 6998–7008. doi: 10.1128/JB.187.20.6998-7008.2005
- Kitagawa, R., Mitsuki, H., Okazaki, T., and Ogawa, T. (1996). A novel DnaA protein-binding site at 94.7 min on the *Escherichia coli* chromosome. *Mol. Microbiol.* 19, 1137–1147. doi: 10.1046/j.1365-2958.1996.453983.x
- Kitagawa, R., Ozaki, T., Moriya, S., and Ogawa, T. (1998). Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for *Escherichia coli* DnaA protein. *Genes Develop.* 12, 3032–3043. doi: 10.1101/gad.12.19.3032
- Kurokawa, K., Nishida, S., Emoto, A., Sekimizu, K., and Katayama, T. (1999). Replication cycle-coordinated change of the adenine nucleotide-bound forms of DnaA protein in *Escherichia coli*. *EMBO J.* 18, 6642–6652. doi: 10.1093/emboj/18.23.6642
- Lal, A., Dhar, A., Trostel, A., Kouzine, F., Seshasayee, A. S., and Adhya, S. (2016). Genome scale patterns of supercoiling in a bacterial chromosome. *Nat. Commun.* 7:11055. doi: 10.1038/ncomms11055
- Lee, Y. S., Han, J. S., Jeon, Y., and Hwang, D. S. (2001). The arc two-component signal transduction system inhibits *in vitro* *Escherichia coli* chromosomal initiation. *J. Biol. Chem.* 276, 9917–9923. doi: 10.1074/jbc.M008629200
- Leonard, A. C., and Grimwade, J. E. (2011). Regulation of DnaA assembly and activity: taking directions from the genome. *Annu. Rev. Microbiol.* 65, 19–35. doi: 10.1146/annurev-micro-090110-102934
- Leonard, A. C., and Grimwade, J. E. (2015). The orisome: structure and function. *Front. Microbiol.* 6:545. doi: 10.3389/fmicb.2015.00545
- Leonard, A. C., and Mechali, M. (2013). DNA replication origins. *Cold Spring Harb. Perspect. Biol.* 5:a010116. doi: 10.1101/cshperspect.a010116



- Løbner-Olesen, A., Hansen, F. G., Rasmussen, K. V., Martin, B., and Kuempel, P. L. (1994). The initiation cascade for chromosome replication in wild-type and Dam methyltransferase deficient *Escherichia coli* cells. *EMBO J.* 13, 1856–1862.
- Løbner-Olesen, A., Skarstad, K., Hansen, F. G., von Meyenburg, K., and Boye, E. (1989). The DnaA protein determines the initiation mass of *Escherichia coli* K-12. *Cell* 57, 881–889. doi: 10.1016/0092-8674(89)90802-7
- Louarn, J., Bouche, J. P., and Louarn, J. M. (1984). Genetic inactivation of topoisomerase I suppresses a defect in initiation of chromosome replication in *E. coli*. *Mol. Gen. Genet.* 195, 170–174. doi: 10.1007/BF00332741
- Lu, M., Campbell, J. L., Boye, E., and Kleckner, N. (1994). SeqA: a negative modulator of replication initiation in *E. coli*. *Cell* 77, 413–426. doi: 10.1016/0092-8674(94)90156-2
- Magnan, D., and Bates, D. (2015). Regulation of DNA replication initiation by Chromosome structure. *J. Bacteriol.* 197, 3370–3377. doi: 10.1128/JB.00446-15
- Margulies, C., and Kaguni, J. M. (1998). The FIS protein fails to block the binding of DnaA protein to oriC, the *Escherichia coli* chromosomal origin. *Nucleic Acids Res.* 26, 5170–5175. doi: 10.1093/nar/26.22.5170
- McGarry, K. C., Ryan, V. T., Grimwade, J. E., and Leonard, A. C. (2004). Two discriminatory binding sites in the *Escherichia coli* replication origin are required for DNA strand opening by initiator DnaA-ATP. *Proc. Natl. Acad. Sci. U.S.A.* 101, 2811–2816. doi: 10.1073/pnas.0400340101
- Mojica, F. J., and Higgins, C. F. (1997). *In vivo* supercoiling of plasmid and chromosomal DNA in an *Escherichia coli* hns mutant. *J. Bacteriol.* 179, 3528–3533.
- Moolman, M. C., Krishnan, S. T., Kerssemakers, J. W., van den Berg, A., Tulinski, P., Depken, M., et al. (2014). Slow unloading leads to DNA-bound beta2-sliding clamp accumulation in live *Escherichia coli* cells. *Nat. Commun.* 5, 5820. doi: 10.1038/ncomms6820
- Morigen, Boye, E., Skarstad, K., and Løbner-Olesen, A. (2001). Regulation of chromosomal replication by DnaA protein availability in *Escherichia coli*: effects of the *datA* region. *Biochim. Biophys. Acta.* 1521, 73–80. doi: 10.1016/S0167-4781(01)00292-5
- Mott, M. L., and Berger, J. M. (2007). DNA replication initiation: mechanisms and regulation in bacteria. *Nat. Rev. Microbiol.* 5, 343–354. doi: 10.1038/nrmicro1640
- Nievera, C., Torgue, J. J., Grimwade, J. E., and Leonard, A. C. (2006). SeqA blocking of DnaA-oriC interactions ensures staged assembly of the *E. coli* pre-RC. *Mol. Cell* 24, 581–592. doi: 10.1016/j.molcel.2006.09.016
- Nystrom, T., Larsson, C., and Gustafsson, L. (1996). Bacterial defense against aging: role of the *Escherichia coli* ArcA regulator in gene expression, readjusted energy flux and survival during stasis. *EMBO J.* 15, 3219–3228.
- Ogawa, T., Yamada, Y., Kuroda, T., Kishi, T., and Moriya, S. (2002). The *datA* locus predominantly contributes to the initiator titration mechanism in the control of replication initiation in *Escherichia coli*. *Mol. Microbiol.* 44, 1367–1375. doi: 10.1046/j.1365-2958.2002.02969.x
- Olsson, J. A., Nordstrom, K., Hjort, K., and Dasgupta, S. (2003). Eclipse-synchrony relationship in *Escherichia coli* strains with mutations affecting sequestration, Initiation of replication and superhelicity of the bacterial chromosome. *J. Mol. Biol.* 334, 919–931. doi: 10.1016/j.jmb.2003.10.029
- Ozaki, S., and Katayama, T. (2012). Highly organized DnaA-oriC complexes recruit the single-stranded DNA for replication initiation. *Nucleic Acids Res.* 40, 1648–1665. doi: 10.1093/nar/gkr832
- Ozaki, S., Noguchi, Y., Hayashi, Y., Miyazaki, E., and Katayama, T. (2012). Differentiation of the DnaA-oriC subcomplex for DNA unwinding in a replication initiation complex. *J. Biol. Chem.* 287, 37458–37471. doi: 10.1074/jbc.M112.372052
- Petersen, C., and Møller, L. B. (2000). Invariance of the nucleoside triphosphate pools of *Escherichia coli* with growth rate. *J. Biol. Chem.* 275, 3931–3935. doi: 10.1074/jbc.275.6.3931
- Polaczek, P. (1990). Bending of the origin of replication of *E. coli* by binding of IHF at a specific site. *New Biol.* 2, 265–271.
- Riber, L., Fujimitsu, K., Katayama, T., and Løbner-Olesen, A. (2009). Loss of Hda activity stimulates replication initiation from I-box, but not R4 mutant origins in *Escherichia coli*. *Mol. Microbiol.* 71, 107–122. doi: 10.1111/j.1365-2958.2008.06516.x
- Riber, L., Olsson, J. A., Jensen, R. B., Skovgaard, O., Dasgupta, S., Marinus, M. G., et al. (2006). Hda-mediated inactivation of the DnaA protein and *dnaA* gene autoregulation act in concert to ensure homeostatic maintenance of the *Escherichia coli* chromosome. *Genes Dev.* 20, 2121–2134. doi: 10.1101/gad.379506
- Rozgaja, T. A., Grimwade, J. E., Iqbal, M., Czerwonka, C., Vora, M., and Leonard, A. C. (2011). Two oppositely oriented arrays of low-affinity recognition sites in oriC guide progressive binding of DnaA during *Escherichia coli* pre-RC assembly. *Mol. Microbiol.* 82, 475–488. doi: 10.1111/j.1365-2958.2011.07827.x
- Ryan, V. T., Grimwade, J. E., Camara, J. E., Crooke, E., and Leonard, A. C. (2004). *Escherichia coli* prereplication complex assembly is regulated by dynamic interplay among Fis, IHF and DnaA. *Mol. Microbiol.* 51, 1347–1359. doi: 10.1046/j.1365-2958.2003.03906.x
- Ryan, V. T., Grimwade, J. E., Nievera, C. J., and Leonard, A. C. (2002). IHF and HU stimulate assembly of pre-replication complexes at *Escherichia coli* oriC by two different mechanisms. *Mol. Microbiol.* 46, 113–124. doi: 10.1046/j.1365-2958.2002.03129.x
- Saxena, R., Fingland, N., Patil, D., Sharma, A. K., and Crooke, E. (2013). Crosstalk between DnaA protein, the initiator of *Escherichia coli* chromosomal replication, and acidic phospholipids present in bacterial membranes. *Int. J. Mol. Sci.* 14, 8517–8537. doi: 10.3390/ijms14048517
- Sekimizu, K., Bramhill, D., and Kornberg, A. (1987). ATP activates *dnaA* protein in initiating replication of plasmids bearing the origin of the *E. coli* chromosome. *Cell* 50, 259–265. doi: 10.1016/0092-8674(87)90221-2
- Simmons, L. A., Breier, A. M., Cozzarelli, N. R., and Kaguni, J. M. (2004). Hyperinitiation of DNA replication in *Escherichia coli* leads to replication fork collapse and inviability. *Mol. Microbiol.* 51, 349–358. doi: 10.1046/j.1365-2958.2003.03842.x
- Skarstad, K., Boye, E., and Steen, H. B. (1986). Timing of initiation of chromosome replication in individual *E. coli* cells. *EMBO J.* 5, 1711–1717.
- Skarstad, K., and Katayama, T. (2013). Regulating DNA replication in bacteria. *Cold Spring Harb. Perspect. Biol.* 5:a012922. doi: 10.1101/cshperspect.a012922
- Skarstad, K., and Løbner-Olesen, A. (2003). Stable co-existence of separate replicons in *Escherichia coli* is dependent on once-per-cell-cycle initiation. *EMBO J.* 22, 140–150. doi: 10.1093/emboj/cdg003
- Skarstad, K., Thöny, B., Hwang, D. S., and Kornberg, A. (1993). A novel binding protein of the origin of the *Escherichia coli* chromosome. *J. Biol. Chem.* 268, 5365–5370.
- Skarstad, K., von Meyenburg, K., Hansen, F. G., and Boye, E. (1988). Coordination of chromosome replication initiation in *Escherichia coli*: effects of different *dnaA* alleles. *J. Bacteriol.* 170, 852–858.
- Torheim, N. K., Boye, E., Løbner-Olesen, A., Stokke, T., and Skarstad, K. (2000). The *Escherichia coli* SeqA protein destabilizes mutant DnaA204 protein. *Mol. Microbiol.* 37, 629–638. doi: 10.1046/j.1365-2958.2000.02031.x
- Usongo, V., Martel, M., Balleydier, A., and Drolet, M. (2016). Mutations reducing replication from R-loops suppress the defects of growth, chromosome segregation and DNA supercoiling in cells lacking topoisomerase I and RNase HI activity. *DNA Repair* 40, 1–17. doi: 10.1016/j.dnarep.2016.02.001
- Usongo, V., Tanguay, C., Nolent, F., Bessong, J. E., and Drolet, M. (2013). Interplay between type IA topoisomerases and gyrase in chromosome segregation in *Escherichia coli*. *J. Bacteriol.* 195, 1758–1768. doi: 10.1128/JB.02001-12
- von Freiesleben, U., Krekling, M. A., Hansen, F. G., and Løbner-Olesen, A. (2000a). The eclipse period of *Escherichia coli*. *EMBO J.* 19, 6240–6248. doi: 10.1093/emboj/19.22.6240
- von Freiesleben, U., and Rasmussen, K. V. (1991). DNA replication in *Escherichia coli* *gyrB*(Ts) mutants analysed by flow cytometry. *Res. Microbiol.* 142, 223–227. doi: 10.1016/0923-2508(91)90034-8
- von Freiesleben, U., and Rasmussen, K. V. (1992). The level of supercoiling affects the regulation of DNA replication in *Escherichia coli*. *Res. Microbiol.* 143, 655–663. doi: 10.1016/0923-2508(92)90060-2
- von Freiesleben, U., Rasmussen, K. V., Atlung, T., and Hansen, F. G. (2000b). Rifampicin resistant initiation of chromosome replication from oriC in *ihf* mutants. *Mol. Microbiol.* 37, 1–8. doi: 10.1046/j.1365-2958.2000.02060.x



- von Freiesleben, U., Rasmussen, K. V., and Schaechter, M. (1994). SeqA limits DnaA activity in replication from *oriC* in *Escherichia coli*. *Mol. Microbiol.* 14, 763–772. doi: 10.1111/j.1365-2958.1994.tb01313.x
- Wang, X., Montero, L. P., and Rudner, D. Z. (2013). Organization and segregation of bacterial chromosomes. *Nat. Rev. Genet.* 14, 191–203. doi: 10.1038/nrg3375
- Weigel, C., Messer, W., Preiss, S., Welzcek, M., Boye, M., and Boye, E. (2001). The sequence requirements for a functional *Escherichia coli* replication origin are different for the chromosome and a minichromosome. *Mol. Microbiol.* 40, 498–507. doi: 10.1046/j.1365-2958.2001.02409.x
- Weitao, T., Nordstrom, K., and Dasgupta, S. (1999). Mutual suppression of mukB and seqA phenotypes might arise from their opposing influences on the *Escherichia coli* nucleoid structure. *Mol. Microbiol.* 34, 157–168. doi: 10.1046/j.1365-2958.1999.01589.x
- Weitao, T., Nordström, K., and Dasgupta, S. (2000). *Escherichia coli* cell cycle control genes affect chromosome superhelicity. *EMBO Rep.* 1, 494–499. doi: 10.1093/embo-reports/kvd106
- Wold, S., Crooke, E., and Skarstad, K. (1996). The *Escherichia coli* Fis protein prevents initiation of DNA replication from *oriC* in vitro. *Nucleic Acids Res.* 24, 3527–3532. doi: 10.1093/nar/24.18.3527
- Wu, H. Y., Shyy, S. H., Wang, J. C., and Liu, L. F. (1988). Transcription generates positively and negatively supercoiled domains in the template. *Cell* 53, 433–440. doi: 10.1016/0092-8674(88)90163-8
- Yamanaka, K., Zheng, W., Crooke, E., Wang, Y. H., and Inouye, M. (2001). CspD, a novel DNA replication inhibitor induced during the stationary phase in *Escherichia coli*. *Mol. Microbiol.* 39, 1572–1584. doi: 10.1046/j.1365-2958.2001.02345.x
- Yun, S. H., Ji, S. C., Jeon, H. J., Wang, X., Kim, S. W., Bak, G., et al. (2012). The CnuK9E H-NS complex antagonizes DNA binding of DnaA and leads to temperature-dependent filamentous growth in *E. coli*. *PLoS ONE* 7:e45236. doi: 10.1371/journal.pone.0045236

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MM declared a past co-authorship with the author ALO to the handling Editor, who ensured that the process met the standards of a fair and objective review.

Copyright © 2016 Riber, Frimodt-Møller, Charbon and Løbner-Olesen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Metal-Induced Stabilization and Activation of Plasmid Replication Initiator RepB

José A. Ruiz-Masó<sup>1</sup>, Lorena Bordanaba-Ruiseco<sup>1</sup>, Marta Sanz<sup>1</sup>, Margarita Menéndez<sup>2,3\*</sup> and Gloria del Solar<sup>1\*</sup>

<sup>1</sup> Molecular Biology of Gram-Positive Bacteria, Molecular Microbiology and Infection Biology, Centro de Investigaciones Biológicas (Consejo Superior de Investigaciones Científicas), Madrid, Spain, <sup>2</sup> Biological Physical Chemistry, Protein Structure and Thermodynamics, Instituto de Química-Física Rocasolano (Consejo Superior de Investigaciones Científicas), Madrid, Spain, <sup>3</sup> CIBER of Respiratory Diseases, Madrid, Spain

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas Medical Branch,  
USA

### Reviewed by:

Gabriel Moncalian,  
University of Cantabria, Spain  
Jan Nesvera,  
Institute of Microbiology (ASCR),  
Czech Republic

### \*Correspondence:

Margarita Menéndez  
mmenendez@iqfr.csic.es  
Gloria del Solar  
gdelsolar@cib.csic.es

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 03 August 2016

**Accepted:** 02 September 2016

**Published:** 21 September 2016

### Citation:

Ruiz-Masó JA, Bordanaba-Ruiseco L,  
Sanz M, Menéndez M and del Solar G  
(2016) Metal-Induced Stabilization and  
Activation of Plasmid Replication  
Initiator RepB.  
Front. Mol. Biosci. 3:56.  
doi: 10.3389/fmolb.2016.00056

Initiation of plasmid rolling circle replication (RCR) is catalyzed by a plasmid-encoded Rep protein that performs a Tyr- and metal-dependent site-specific cleavage of one DNA strand within the double-strand origin (*dso*) of replication. The crystal structure of RepB, the initiator protein of the streptococcal plasmid pMV158, constitutes the first example of a Rep protein structure from RCR plasmids. It forms a toroidal homohexameric ring where each RepB protomer consists of two domains: the C-terminal domain involved in oligomerization and the N-terminal domain containing the DNA-binding and endonuclease activities. Binding of Mn<sup>2+</sup> to the active site is essential for the catalytic activity of RepB. In this work, we have studied the effects of metal binding on the structure and thermostability of full-length hexameric RepB and each of its separate domains by using different biophysical approaches. The analysis of the temperature-induced changes in RepB shows that the first thermal transition, which occurs at a range of temperatures physiologically relevant for the pMV158 pneumococcal host, represents an irreversible conformational change that affects the secondary and tertiary structure of the protein, which becomes prone to self-associate. This transition, which is also shown to result in loss of DNA binding capacity and catalytic activity of RepB, is confined to its N-terminal domain. Mn<sup>2+</sup> protects the protein from undergoing this detrimental conformational change and the observed protection correlates well with the high-affinity binding of the cation to the active site, as substituting one of the metal-ligands at this site impairs both the protein affinity for Mn<sup>2+</sup> and the Mn<sup>2+</sup>-driven thermostabilization effect. The level of catalytic activity of the protein, especially in the case of full-length RepB, cannot be explained based only on the high-affinity binding of Mn<sup>2+</sup> at the active site and suggests the existence of additional, lower-affinity metal binding site(s), missing in the separate catalytic domain, that must also be saturated for maximal activity. The molecular bases of the thermostabilizing effect of Mn<sup>2+</sup> on the N-terminal domain of the protein as well as the potential location of additional metal binding sites in the entire RepB are discussed.

**Keywords:** HUH endonucleases, plasmid-encoded Rep proteins, metal-dependent catalytic activity, RepB thermostability, Mn<sup>2+</sup> affinity

## INTRODUCTION

The rolling circle replication (RCR) mechanism is used by transposons, small plasmids, phages, and viruses that replicate autonomously in a wide range of organisms, from prokaryotes to humans (Campos-Olivas et al., 2002). Plasmids that use this mechanism for their replication are termed RCR plasmids, and they are found in bacteria, archaea, and mitochondria (Novick, 1998; Khan, 2000; Ruiz-Masó et al., 2015). Initiation of plasmid RCR requires site-specific cleavage of one plasmid DNA strand within the double-strand origin (*dso*) of replication. This reaction is catalyzed by the metal-dependent endonucleolytic activity of the plasmid-encoded Rep protein, which yields a free 3'-OH end that serves as primer for initiation of the leading-strand synthesis by a host DNA polymerase. The initiator Rep also mediates the endonuclease and strand-transfer reactions that take place at the termination of the leading-strand replication process (Novick, 1998).

RCR plasmids have been classified into several replicon families based on sequence similarities at the Rep and *dso* level (del Solar et al., 1998; Khan, 2005; Ruiz-Masó et al., 2015). The replicon of pMV158, a small (5541 bp) multicopy promiscuous plasmid originally isolated from *Streptococcus agalactiae* and involved in antibiotic resistance spread, has been studied in depth and is considered as the prototype of a family of RCR plasmids isolated from several eubacteria (del Solar et al., 1993). RepB, the replication initiator protein of pMV158, carries out metal ion-dependent DNA cleavage and rejoining reactions as part of its replication function. Upon specific binding to the *dso*, RepB cleaves one strand of the DNA at a specific dinucleotide of the nick sequence (TACTACG/AC; / indicating the nick site) located on the apical loop of a hairpin formed by an inverted repeat (IR-I) (Moscoso et al., 1995; Ruiz-Masó et al., 2007). The nature of the cleavage reaction demands that the DNA substrate is in an unpaired configuration, which is achieved by IR-I hairpin extrusion on supercoiled DNA. *In vitro*, RepB contacts with its primary binding site (the *bind* locus) and with a region of the *nic* locus that includes the right arm of IR-I. Binding of RepB to the *bind* locus seems to facilitate binding of the protein to the *nic* locus, which promotes extrusion of the IR-I hairpin containing the substrate DNA to be cleaved (Ruiz-Masó et al., 2007). The nucleophilic attack on the scissile phosphodiester bond of the DNA is most likely exerted by the catalytic Tyr99 of RepB (Moscoso et al., 1997). Like other RCR Rep initiators from plasmids and bacteriophages, RepB lacks ATPase and helicase activities (de la Campa et al., 1990; Moscoso et al., 1995). Thus, apart from the DNA polymerase, other host proteins such as a superfamily 1 (SF1) DNA helicase and a single-stranded DNA (ssDNA)-binding protein are expected to be recruited to participate in the early stages of initiation and elongation.

RepB is a 210 amino acid polypeptide that is purified as a hexamer (RepB<sub>6</sub>, Ruiz-Masó et al., 2004). X-ray crystallography revealed the structure of full-length RepB<sub>6</sub>, which forms a toroidal homohexameric ring (Ruiz-Masó et al., 2004; Boer et al., 2009). Each RepB protomer comprises an N-terminal endonuclease domain, referred to as the origin binding domain (OBD), and a C-terminal oligomerization domain (OD) that

forms a cylinder with a six-fold symmetry in the hexamer (Supplementary Figure 1). The conformational ensemble of RepB<sub>6</sub> is characterized by a rigid cylindrical scaffold, formed by the ODs, to which the OBDs are attached as highly flexible appendages. The intrinsic flexibility allows RepB to adopt multiple conformational states and might be involved in the specific recognition of the *dso* (Boer et al., 2016). The N-terminal 131-residue OBD domain retains the DNA-binding and nuclease functions of the protein (Boer et al., 2009). This domain belongs to the superfamily of HUH endonucleases (in which U is a hydrophobic residue), which includes proteins of the Rep class, involved in replication of bacteriophages, plasmids, and plant and animal viruses, and of the Mob class, also known as relaxases, involved in the conjugal transfer of plasmid DNA (Ilyina and Koonin, 1992). The overall structure of the endonuclease domain of the HUH endonuclease superfamily is very similar despite the low level of sequence identity, and is characterized by a five-stranded antiparallel  $\beta$ -sheet flanked by a variable number of  $\alpha$ -helices (Dyda and Hickman, 2003; Chandler et al., 2013). Moreover, the entire superfamily appears to follow a common endonucleolytic mechanism based on a catalytic Tyr and a divalent metal coordinated by a His cluster (Dyda and Hickman, 2003). The conserved HUH sequence motif, present in Rep and Mob proteins (Ilyina and Koonin, 1992), was confirmed as part of the metal binding site from structural data (Campos-Olivas et al., 2002; Hickman et al., 2002; Boer et al., 2009). Another conserved motif, designated UXXYUXK in Rep proteins, includes the catalytic Tyr (Ilyina and Koonin, 1992).

RepB OBD central  $\beta$ -sheet is flanked by helices  $\alpha$ 1 and  $\alpha$ 2 at one face, and by helix  $\alpha$ 3, which provides the catalytic residue Tyr99, and the short helix  $\alpha$ 4 at the opposite side. In addition, a Mn<sup>2+</sup> cation is found close to Tyr99 in the active site (Supplementary Figure 1B). This metal ion is coordinated by five ligands, namely the RepB residues His39, Asp42, His55, and His57 (the latter two residues forming the HUH motif) and a single solvent molecule, in an octahedral-minus-one or square-based pyramidal geometry (Boer et al., 2009). All four RepB residues ligating the Mn<sup>2+</sup> cation are placed in sequence motifs that are conserved in the Rep proteins of the pMV158 RCR plasmid family (del Solar et al., 1993), as is also the case with catalytic Tyr99 and with Tyr115, which hydrogen bonds to the Asp42 carboxyl group (Supplementary Figure 1). *In vitro*, only Mn<sup>2+</sup> and Co<sup>2+</sup>, among various divalent cations tested, are able to promote RepB-mediated nicking-closing of supercoiled plasmid DNA (Boer et al., 2009). Thus, the presence of Mn<sup>2+</sup> in the active site is consistent with these requirements. Although in DNA cleavage reactions where the hydroxyl group of a tyrosine or a serine acts as a nucleophile there is no apparent need of a metal cation for activation, the simultaneous presence of a tyrosine and a metal cation in the active site seems to be a common feature in the HUH endonucleases studied so far. In fact, it is generally accepted to attribute a structural role to the cation bound in the active site (Hickman et al., 2002; Larkin et al., 2005; Boer et al., 2006). In RepB, the Mn<sup>2+</sup> ion probably interacts with the oxygen atoms of the scissile DNA phosphate, polarizing the bond and favoring the nucleophilic attack by the catalytic Tyr99 (Boer et al., 2009). The presence of additional divalent

cation binding sites at the interface of OBD and OD domains has been reported for the C2 crystal structure of RepB<sub>6</sub> (Boer et al., 2016).

Current structural information about full length Rep proteins from RCR plasmids is restricted to RepB, although the structure of a chimeric initiator Rep protein of staphylococcal plasmids belonging to the pT181 family has been recently solved (Carr et al., 2016). In addition, little information on biochemical and biophysical parameters has been reported for these proteins. In this work we have analyzed the effect of Mn<sup>2+</sup> on both the thermostability and the catalytic activity of RepB. We demonstrate that the manganese cation strongly protects the protein from undergoing a thermal transition that otherwise takes place between 32 and 45°C. We also show that the conformational change associated with this transition is confined to the OBD and renders the protein catalytically inactive and unable to recognize the plasmid replication origin. Mn<sup>2+</sup>-driven thermostabilization of RepB most likely results from binding of the divalent cation to the active site of the protein, and is compatible with the metal affinity values obtained by isothermal titration calorimetry (ITC) for different protein variants. On the other hand, the analysis of the Mn<sup>2+</sup> concentration dependence of the catalytic activity of the protein indicates that maximal activity of full-length RepB<sub>6</sub> would require saturation of both the high affinity site in the active center and additional lower affinity site(s).

## RESULTS

### Characterization of the RepB Thermal Transitions and Their Effects on the Protein Activity

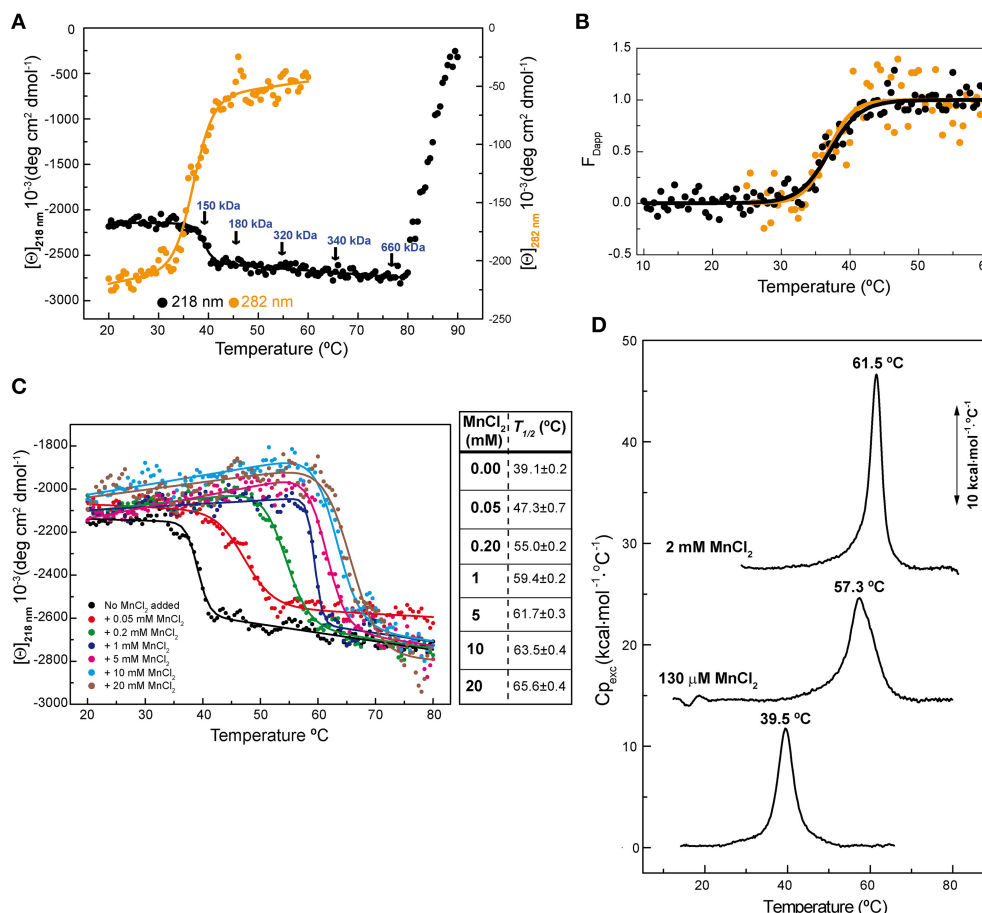
Previous circular dichroism (CD) studies on hexameric RepB<sub>6</sub> revealed the presence of a temperature-induced irreversible transition between 32 and 45°C leading to a small, but significant, increase of the protein  $\alpha$ -helical content, whereas a second transition occurring above 80°C resulted in RepB precipitation (Ruiz-Masó et al., 2004). We now show that the first transition also induced a decline in the ellipticity signal at 282 nm (**Figure 1A**), indicative that RepB<sub>6</sub> tertiary/quaternary structure was also modified, and that the transition advance estimated from the CD thermal profiles at 282 and 218 nm fully overlapped (**Figure 1B**). The irreversibility of such conformational change allowed us to analyze, by analytical ultracentrifugation, the oligomerization state of RepB<sub>6</sub> heated to different temperatures in the range from 25 to 75°C. As indicated in **Figure 1A**, the average molecular weight remained close to that of the hexamer (145.5 kDa) up to the end of the first transition ( $M_{app}/M_0 = 1.2$  at 45°C;  $M_0$  being the hexamer molecular weight). However, a clear increase in the oligomerization state was observed as the temperature was further increased, followed by the protein precipitation above 80°C.

To investigate the effect of the first thermal transition on RepB<sub>6</sub> activity as RCR initiator, we tested the nicking/closing

and DNA binding capacities of RepB<sub>6</sub> after being heated or not to 45°C. The results showed that the protein heated to 45°C was unable to relax the supercoiled (sc) cognate plasmid DNA (**Figure 2A**) and had also lost its ability to bind to the target dsDNA (**Figure 2B**). In contrast with this, the RepB<sub>6</sub> nicking/closing activity on scDNA was maximal at 60°C in the presence of 10–20 mM MnCl<sub>2</sub>, whereas it decreased to about 50% when the reaction was carried out at the same Mn<sup>2+</sup> concentrations but at 37°C, the optimal growth temperature of the pMV158 pneumococcal host (Moscoso et al., 1995; **Figure 2C**). It is noteworthy that the enhancement of RepB<sub>6</sub> activity at 60°C is restricted to sc plasmid DNA and was not observed on ssDNA substrates unable to form the IR-I cruciform (**Figure 3**). Therefore, the higher activity at 60°C is most likely due to the high temperature facilitating the extrusion of the cruciform that renders the nick sequence a single-stranded substrate. Be that as it may, preservation of the activity at 60°C required a factor specifically present in the reaction mixture that protected the protein from the thermal inactivation. Mn<sup>2+</sup> cations were next shown to account for this role, as the presence of 20 mM MnCl<sub>2</sub> during RepB<sub>6</sub> heating to 45°C prevented its inactivation and kept intact its endonuclease (**Figure 2A**) and DNA binding activities (not shown). Regardless of the presence of MnCl<sub>2</sub>, the sample heated up to 70°C was completely inactive (**Figure 2A**). To explore the influence of Mn<sup>2+</sup> on the structure and thermal stability of RepB<sub>6</sub>, we first compared the CD spectra (far- and near-UV regions) of the protein in the presence and in the absence of MnCl<sub>2</sub>. Their coincidence indicated that no significant changes occurred in either the secondary structure or the tertiary/quaternary structure upon binding of Mn<sup>2+</sup> (not shown). Next we carried out thermal denaturation experiments in the presence of Mn<sup>2+</sup> at concentrations ranging from 0.05 to 20 mM and the first thermal transition was assessed by monitoring RepB<sub>6</sub> ellipticity at 218 nm. The results showed a strong stabilization of RepB<sub>6</sub> by Mn<sup>2+</sup>, shifting the apparent half-transition temperature ( $T_{1/2}$ ) by around 25°C (from ~39 to ~66°C) at the maximum concentration of MnCl<sub>2</sub> tested (**Figure 1C**). In contrast, MgCl<sub>2</sub> or CaCl<sub>2</sub> addition had no effect on RepB<sub>6</sub> stability (not shown). The steepest variation of  $T_{1/2}$  occurred below 1 mM and the progression of  $T_{1/2}$  at higher ligand concentrations followed the trend expected for ligand binding domains (Brandts et al., 1989).

The influence of Mn<sup>2+</sup> in RepB<sub>6</sub> structural stability was also examined by differential scanning calorimetry (DSC; **Figure 1D**). In the absence of cation, the thermogram shows a peak with a transition temperature ( $T_m$ ) of 39.5°C, very close to the  $T_{1/2}$  obtained for the first CD transition, and a transition enthalpy change of 71 kcal/mol of protomer, which supported a protein denaturation event. Above 80°C the baseline dropped drastically due to RepB<sub>6</sub> precipitation, in agreement with CD results. The visible peak was drastically shifted to higher temperatures upon Mn<sup>2+</sup> addition (57.3 and 61.5°C for 130  $\mu$ M and 2 mM of Mn<sup>2+</sup>, respectively; **Figure 1D**). The cation addition also increased the transition enthalpies to 89 kcal/mol (130  $\mu$ M Mn<sup>2+</sup>) and 129 kcal/mol (2 mM Mn<sup>2+</sup>).





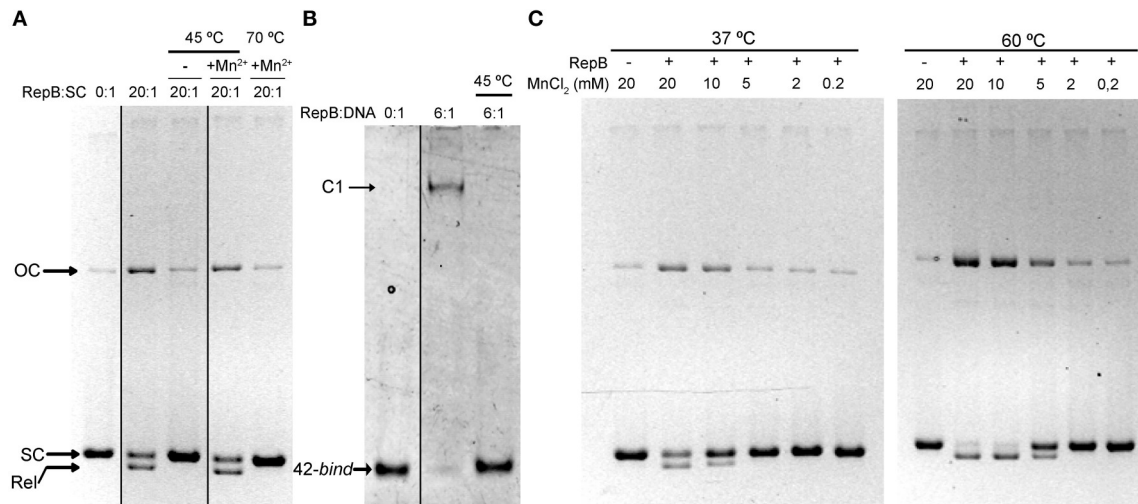
**FIGURE 1 | RepB<sub>6</sub> low-temperature transition involves global changes in the protein structure. (A)** Temperature-induced conformational changes of RepB<sub>6</sub> as monitored by the variation of the ellipticity at 218 and 282 nm ([ $\Theta$ ] represents protein molar ellipticity). Samples heated at the indicated temperatures (arrows) were analyzed by analytical ultracentrifugation (sedimentation equilibrium), and the estimated average molecular masses are displayed on the protein CD thermal profile. **(B)** Apparent fraction of modified protein ( $F_{Dapp}$ ) calculated from the transition curves registered at 218 nm (●) and 282 nm (●) between 10 and 60°C. The solid line shows the fit of Equation (1) to  $F_{Dapp}$  values. **(C)** Temperature transition curves of RepB<sub>6</sub> (12  $\mu$ M) in the presence of increasing concentrations of MnCl<sub>2</sub> (indicated inside the graph) measured by CD at 218 nm. The table shows the apparent half-transition temperatures of RepB<sub>6</sub> derived from fit of Equation (1) to the figure experimental curves (solid lines). **(D)** DSC profile of the first thermal transition of RepB<sub>6</sub> (30  $\mu$ M) monitored in the absence and in the presence of 130  $\mu$ M or 2 mM Mn<sup>2+</sup>. The position of the maximum of the heat capacity function ( $T_m$ ) is indicated.

## The First Thermal Transition Is Confined to the Catalytic Domain

To investigate whether the first thermal transition affects a particular protein domain, we analyzed the thermal stability of the separate RepB domains, purified from *Escherichia coli* as described (Boer et al., 2009). The N-terminal OBD, which has been shown by analytical ultracentrifugation to be in a monomeric state, contains the endonuclease and DNA binding activities, and retains these abilities when separated from the C-terminal OD, which maintains its hexameric structure (Boer et al., 2009, 2016). Of note, the near-UV CD spectra of the separate domains correlate fairly well with that of RepB<sub>6</sub> (i.e., the RepB<sub>6</sub> spectrum approximately matches the curve obtained by addition of the spectra of the separate domains weighted by the fractional contribution of their amino acid number to the complete protein), evidencing the

conservation of tertiary/quaternary structure in both domains (Figure 4).

Thermal stability of OBD and OD was studied by CD spectroscopy following the procedure used for RepB<sub>6</sub>. The CD thermal profile of OD at 218 nm showed that the oligomerization domain suffered a single thermal transition when the temperature was raised above 78°C, which correlated with the observable precipitation of the sample and the reduction of the spectrum intensity (Supplementary Figure 2). Although the three-dimensional structure of RepB<sub>6</sub> did not reveal a metal binding site specific of the OD (Boer et al., 2009), we decided to assess the effect of Mn<sup>2+</sup> in the stability of the domain. The presence of 1 mM MnCl<sub>2</sub> during the heating of the sample delayed the thermal transition about 5°C. This effect was not specific of Mn<sup>2+</sup> and MgCl<sub>2</sub> produced the same stabilization (not shown).



**FIGURE 2 | The presence of Mn<sup>2+</sup> activates RepB and protects it against thermal inactivation. (A)** Activity assays on supercoiled DNA. Samples of RepB unheated, heated to 45°C in the absence of Mn<sup>2+</sup>, or heated to 45 or 70°C in the presence of 20 mM Mn<sup>2+</sup> were mixed, at the indicated molecular ratios, with 6.7 nM of pMV158 supercoiled DNA and then incubated for 30 min at 37°C in the presence of 20 mM MnCl<sub>2</sub>. The supercoiled (SC), closed relaxed circles (Rel), and open circular (OC) plasmid forms were separated by electrophoresis on agarose gels containing ethidium bromide. Images from different parts of the same gel have been grouped and indicated by dividing lines. **(B)** RepB<sub>6</sub> recognition of the *bind* DNA measured by EMSA. Unheated or 45°C-heated samples of RepB<sub>6</sub> were mixed, at the indicated molecular ratios, with 0.4 μM of a 42-bp DNA fragment containing the three 11 bp-direct repeats that constitute the *bind* region of the pMV158 *dso*. Positions of the free (42-*bind*) and complexed (C1) DNA are indicated. Images from different parts of the same gel have been grouped and indicated by dividing lines. **(C)** Supercoiled pMV158 DNA (6.7 nM) was incubated, either at 37°C or at 60°C, with the indicated concentrations of Mn<sup>2+</sup>, in the absence (–) or in the presence (+) of purified RepB<sub>6</sub> (RepB:pMV158 DNA molecular ratio of 20:1).

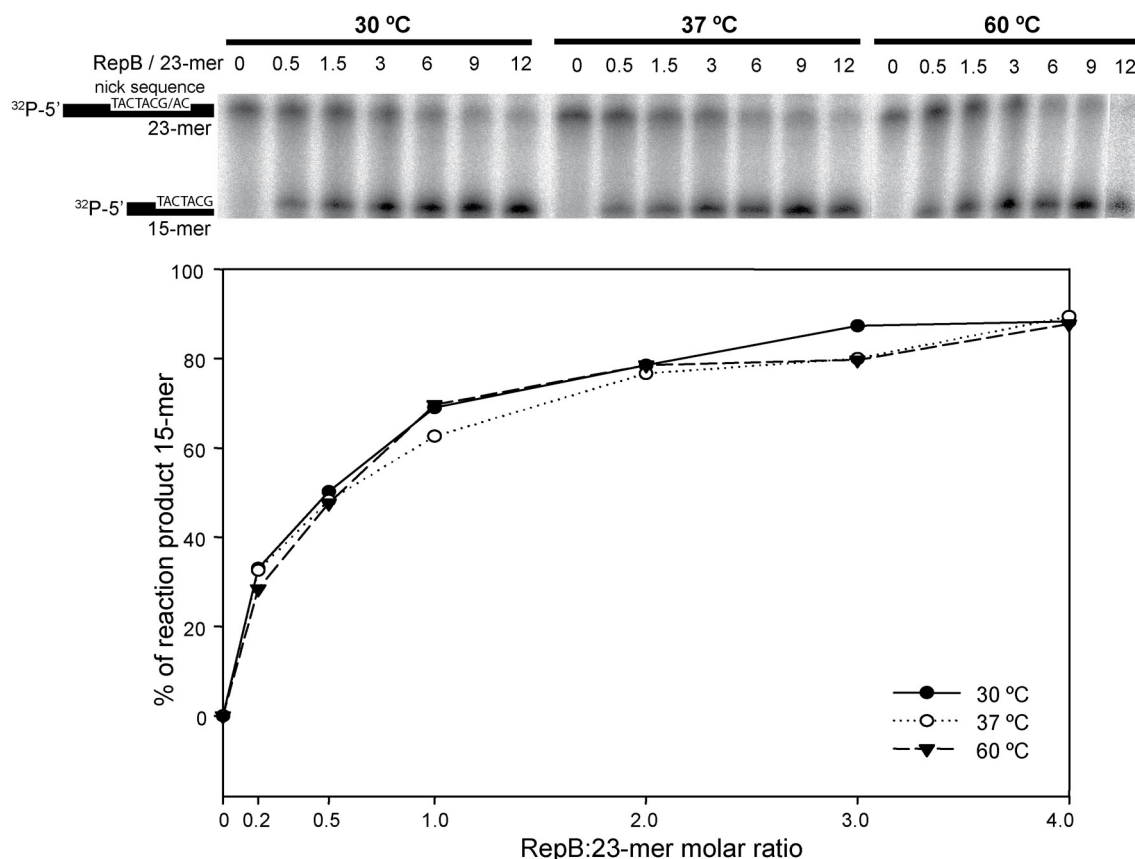
Prior to its thermal characterization, purified OBD, which carried a His-tag, was subjected to an extra-chelating treatment aimed to eliminate trace amounts of divalent cations from the purification steps. The OBD thermal profile at 218 nm shows a single irreversible structural change that takes place with a  $T_{1/2}$  of ~51.5°C. The structural change increased by 66% the ellipticity value at 218 nm, though the intensity of the whole far-UV spectrum decreased when the temperature was raised above 60°C due to OBD precipitation (Figure 5A and Supplementary Figure 3A). Of note, the first structural change of RepB<sub>6</sub> has the same magnitude in protomer molar ellipticity units than the transition of OBD, whereas the cooperativity of the process appears to be somewhat different (Figure 5A). The presence of MnCl<sub>2</sub> during the heating of the OBD sample stabilized the domain structure, increasing by around 13°C the on-set of the thermal transition at 5 mM MnCl<sub>2</sub> (OBD precipitation after denaturation hampered the estimation of  $T_{1/2}$  values above 2 mM Mn<sup>2+</sup>; Figure 5).

Contribution of the active site cation to the stability and the catalytic activity of OBD was evaluated by replacing the acidic residue Asp42, involved in Mn<sup>2+</sup> binding to the active center, by alanine. As for OBD, the protein mutant was treated with EDTA, prior to its equilibration in CD buffer, to eliminate any trace of divalent cations from the purification steps. The far-UV CD spectra of OBD and OBD<sup>D42A</sup> acquired at 20°C were very similar, if not identical (Supplementary Figure 3), and the presence of MnCl<sub>2</sub> did not modify the spectra (not shown). Generation of the mutant OBD<sup>D42A</sup> resulted in a protein variant whose thermal stability was comparable to that of the wild type

domain in the absence of Mn<sup>2+</sup>. In fact, both the magnitude of the ellipticity change at 218 nm and the  $T_{1/2}$  of OBD<sup>D42A</sup> and OBD (50.5 and 51.5°C, respectively) were similar (Figure 5A). Despite removal of a Mn<sup>2+</sup> ligand in the active site, OBD<sup>D42A</sup> still has the capacity to bind Mn<sup>2+</sup>, as shown by the ability of Mn<sup>2+</sup> to up shift the thermal denaturation of the mutant domain (Figures 5B–D). However, at low MnCl<sub>2</sub> concentrations the transition shift was lower in the mutant, probably due to the loss of one of the metal ligand and the consequent Mn<sup>2+</sup> affinity decrease (Figure 5B). Together, these results evidenced that first thermal transition displayed by RepB<sub>6</sub> corresponds to the OBD catalytic domain, and that OBD is the receptor of Mn<sup>2+</sup> cations accounting for RepB<sub>6</sub> stabilization. Besides, the magnitude of the enthalpy change associated to this transition strongly indicates that it implies OBD denaturation.

### Determination of RepB<sub>6</sub>-Mn<sup>2+</sup> Binding Affinity by ITC

The affinity of RepB<sub>6</sub>, OBD and OBD<sup>D42A</sup> for Mn<sup>2+</sup> was examined by ITC. Titrations were performed at 25°C. The analysis of the binding isotherms (Figure 6) showed that each protomer of RepB<sub>6</sub> binds one Mn<sup>2+</sup> cation with high affinity ( $K_b = (2.5 \pm 0.7) \times 10^7 \text{ M}^{-1}$ ;  $N = 0.87 \pm 0.01$  sites/protomer) and a binding enthalpy of  $-3.00 \pm 0.01 \text{ kcal mol}^{-1}$ . The affinity of Mn<sup>2+</sup> for the isolated OBD domain was rather similar ( $K_b = (2.4 \pm 0.8) \times 10^7 \text{ M}^{-1}$ ), but the number of titrable sites was drastically reduced ( $N = 0.47 \pm 0.01$  sites/monomer). In contrast, the  $N$ -value of 0.82 obtained for



**FIGURE 3 | *In vitro* cleavage activity of RepB<sub>6</sub> on ssDNA oligos at different temperatures.** A 23-mer oligo substrate (100 nM), radioactively labeled in 5', was incubated with RepB<sub>6</sub> at three different temperatures (30, 37, and 60°C) in the presence of 20 mM of MnCl<sub>2</sub> and at the indicated protein:oligo substrate molar ratios. The products were separated on 20% PAA, 8M urea denaturing gels (upper part). Nicking activity of RepB<sub>6</sub> was quantified as the percentage of 15-mer product formed (lower part).

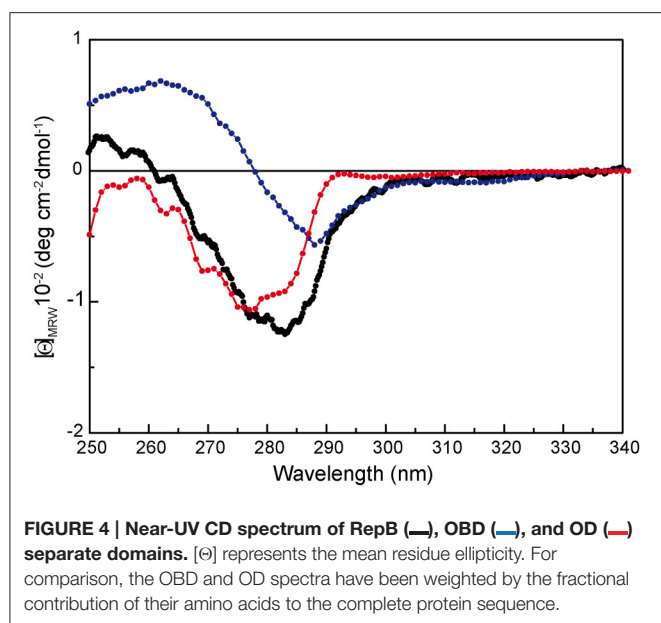
the OBD<sup>D42A</sup>-Mn<sup>2+</sup> complex compared well with that of the complete protein, and the binding affinity was reduced to about one-thirtieth ( $K_b = (8.63 \pm 0.08) \times 10^5 \text{ M}^{-1}$ ). As shown by the thermodynamic parameters displayed in **Figure 6**, Mn<sup>2+</sup> binding to RepB active site is entropically driven, which suggests that primarily occurs through electrostatic interactions and possible removal of bound solvent molecules from the binding interface. However, substitution of Asp42 by alanine made the enthalpy of binding  $-3.37 \text{ kcal mol}^{-1}$  more favorable but almost canceled the entropic contribution, evidencing that Mn<sup>2+</sup> binding to OBD<sup>D42A</sup> implies hydrogen bond formation and/or an entropically unfavorable reorganization of the domain structure.

The high affinity of OBD and OBD<sup>D42A</sup> for Mn<sup>2+</sup> is consistent with the strong metal-dependent stabilization observed in the CD thermal profiles and the DSC thermograms (**Figures 1, 5**). On the other hand, the reduced binding capacity of OBD could denote a high proportion of non-functional domain or, alternatively, previous occupation of Mn<sup>2+</sup> binding sites. This latter possibility could also explain the higher stability of OBD in CD buffer without Mn<sup>2+</sup> compared to RepB<sub>6</sub>.

Titration of RepB<sub>6</sub> with 1 mM MgCl<sub>2</sub> produced neither heat uptake nor release, and supplementation of ITC buffer with 2 mM MgCl<sub>2</sub> did not change Mn<sup>2+</sup> affinity for RepB<sub>6</sub> (not shown). These results, together with the failure of MgCl<sub>2</sub> to stabilize the OBD domain, strongly indicates that Mg<sup>2+</sup> cannot substitute Mn<sup>2+</sup> at the active site.

### The Effect of Metal Binding on OBD and RepB<sub>6</sub> Catalytic Activity

RepB<sub>6</sub> and OBD are able to catalyze the joining of the 5'-phosphate end of the cleavage reaction product with a new 3'-OH end (Moscato et al., 1995). The effect of different divalent metals on the activity of OBD and RepB<sub>6</sub> on single-stranded oligonucleotides (oligos), as well as the influence of the D42A mutation, was assessed by performing cleavage and strand-transfer assays. For these experiments, OBD and OBD<sup>D42A</sup> proteins were subjected to the extra chelating treatment indicated above after removal of their His-tags. In order to reveal the total fraction of reaction products, the reaction mixtures contained 10 pmol of a Cy5 3'-labeled 27-mer substrate carrying



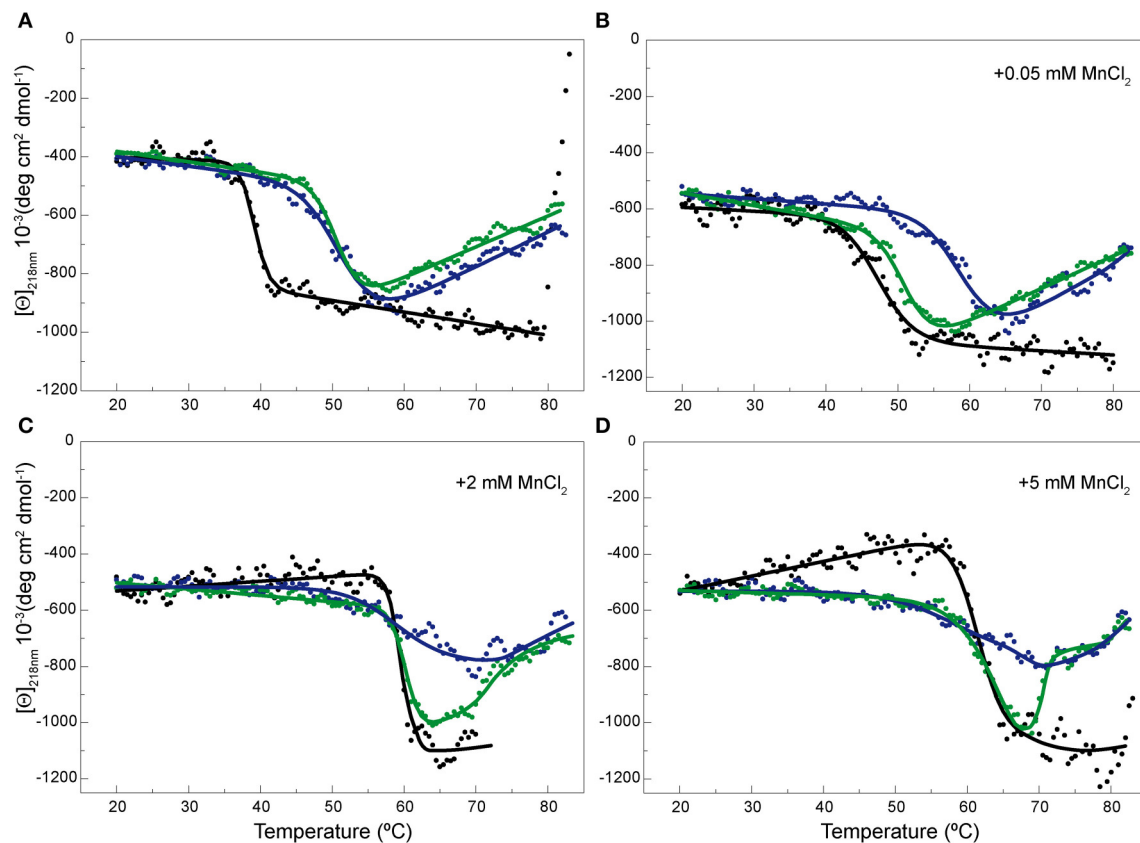
the specific nick sequence, and a 10-fold molar excess of an unlabeled 30-mer that provided the 3'-OH substrate for strand transfer, thus avoiding re-joining of the 27-mer oligo. The mixture of oligos was treated with OBD or OBD<sup>D42A</sup> as indicated in the Experimental Procedures and, subsequently, the reaction products were analyzed by electrophoresis in PAA-urea sequencing gels. Cleavage and strand-transfer activities resulted in the generation of two new fluorescent bands corresponding to 12- and 42-mer products, respectively. In addition, incubation of the samples with SDS and proteinase K, used to stop the reaction, allowed the detection of a covalent complex between OBD and the 12-mer oligo, which appeared as a third fluorescent band corresponding to a small peptide linked to the 5' end of the 12-mer oligo (Figure 7A). The fraction of labeled DNA in each of the three reaction products was calculated and used to determine the protein total activity. Under these conditions of substrate excess, the strand-transfer activity of OBD and OBD<sup>D42A</sup> was prevalent regardless of the Mn<sup>2+</sup> concentration and of the protein variant used, and the main reaction product was the 42-mer (Figure 7A). The effect of adding increasing concentrations of MnCl<sub>2</sub> on the level of substrate conversion by OBD or OBD<sup>D42A</sup> is displayed in Figure 7B. It should be noted that the catalytic activity of OBD was fully dependent on the metal ion, as deduced from the absence of reaction products in the presence of 10 mM EDTA (not shown). However, in the absence of EDTA and at 0.1 μM of MnCl<sub>2</sub>, the lowest metal cation concentration added, the reaction products amounted to ~36 and 41% of the 27-mer total added, respectively; that is, ~66–76% of the maximal activity, which was reached at around 40 μM MnCl<sub>2</sub>. These values reflect the high binding affinity of OBD for Mn<sup>2+</sup>, for which an apparent dissociation constant of 0.5 ± 0.3 μM was estimated assuming that the activity increase above the background reflected the saturation of the cation available sites. The catalytic activity of OBD<sup>D42A</sup> also augmented

upon increasing MnCl<sub>2</sub> concentrations. At 0.1 μM Mn<sup>2+</sup> the percentage of reaction products (~25% of the initial substrate) was, again, very close to the value with no MnCl<sub>2</sub> added, and represented a 51% of substrate conversion under conditions of maximal activity (Figure 7B). The apparent dissociation constant for the Mn<sup>2+</sup> cations accounting for this activity increase (2.6 ± 0.6 μM) was around five-folds higher than for wild-type OBD. The Mn<sup>2+</sup> apparent dissociation constant inferred for OBD<sup>D42A</sup> from the activity assays (~2.6 μM) matched quite well the value of K<sub>d</sub> (~2 μM) obtained by extrapolation of the ITC constant to 37°C, whereas that of OBD (~0.5 μM) was around nine-fold higher than the ITC-derived value (K<sub>d, 37°C</sub> ≅ 57 nM). This apparent discrepancy was probably due to the errors of the activity measurements and to the small net increment of OBD activity at saturation by Mn<sup>2+</sup> with relation to the background without Mn<sup>2+</sup>, whose high value likely reflects the capture of Mn<sup>2+</sup> traces present in the reaction mixture by the active site.

The catalytic activity of RepB<sub>6</sub> on single-stranded oligos at MnCl<sub>2</sub> concentrations ranging from 0.1 μM to 1 mM was analyzed using also a RepB protomer:27-mer substrate DNA molecular ratio of 1:10. As for OBD and OBD<sup>D42A</sup>, the catalytic activity of RepB<sub>6</sub> increased with MnCl<sub>2</sub> concentration (Figure 8A) and the strand-transfer activity was prevalent under conditions of substrate excess (not shown). By contrary, no product formation was observed in the absence of added Mn<sup>2+</sup> and RepB<sub>6</sub> half-maximal activity was reached at ~60 μM of MnCl<sub>2</sub>, a value that is three orders of magnitude higher than that extrapolated from ITC data (K<sub>d, 37°C</sub> ≅ 56 nM). To examine the specificity of such high cation concentration requirement for nicking and strand-transfer activities, we measured the activity of RepB<sub>6</sub> in the same MnCl<sub>2</sub> concentration range but supplementing the reaction mixture with 0.2 mM MgCl<sub>2</sub>. In the presence of only MgCl<sub>2</sub>, the activity of RepB<sub>6</sub> became measurable and product formation represented ~5% of the 27-mer added. Moreover, the presence of MgCl<sub>2</sub> enhanced significantly the activity of RepB<sub>6</sub> at non-saturating concentrations of MnCl<sub>2</sub> without varying the maximal activity of the protein (Figure 8A). The increase of RepB<sub>6</sub> catalytic activity upon Mg<sup>2+</sup> addition is unlikely to be due to trace amounts of Mn<sup>2+</sup> in the MgCl<sub>2</sub> solution, as they should represent less than 4 nM. Besides, the following experimental data suggest that MgCl<sub>2</sub> does not bind to the active site, although they do not discard that it can partially replace Mn<sup>2+</sup> in activating nicking and strand-transfer. First, 0.2 mM MgCl<sub>2</sub> does not stabilize the OBD domain against thermal denaturation in the complete RepB<sub>6</sub> protein. Second, we have failed to find any evidence of Mg<sup>2+</sup> high-affinity binding to RepB<sub>6</sub> through Mg<sup>2+</sup> direct titration or Mg<sup>2+</sup>/Mn<sup>2+</sup> competition assays by ITC (not shown). Moreover, Mn<sup>2+</sup> was the cation found in the active center of the C3 crystals of RepB<sub>6</sub> even though the crystallization buffer contained 200 mM MgCl<sub>2</sub> and theoretically lacked Mn<sup>2+</sup> (Boer et al., 2009).

We have also analyzed the pattern of the reaction products generated by RepB<sub>6</sub> under conditions of protein excess (10:1 protein:27-mer molar ratio) and observed that it varied depending on the concentration of Mn<sup>2+</sup> added (Figure 8B). At 7.5 μM MnCl<sub>2</sub> the main reaction product resulted from the strand-transfer activity of RepB<sub>6</sub>; the observed protein activation





**FIGURE 5 | Temperature-induced changes in the secondary structure of OBD and OBD<sup>D42A</sup>.** Temperature transition curves of RepB<sub>6</sub> (●), OBD (●), and OBD<sup>D42A</sup> (●) measured by CD at 218 nm in the absence of MnCl<sub>2</sub> (A) or in the presence of 0.05 (B), 2 (C), and 5 (D) mM of MnCl<sub>2</sub> ([θ] represents the protein molar ellipticity). The CD thermal profile of RepB<sub>6</sub> has been shifted along the ordinate axis to facilitate the comparison with those OBD and OBD<sup>D42A</sup>. In the case of OBD and OBD<sup>D42A</sup>, the continuous lines represent an average of the experimental data. Measurements were carried out at 12 μM RepB<sub>6</sub> and 19 μM OBD or OBD<sup>D42A</sup>.

relied on the divalent cation as it was not achieved when 7.5 μM NaCl was added instead. Interestingly, at 1 mM MnCl<sub>2</sub> the proportion of strand transfer product was perceptibly decreased and the reaction was shifted to the formation of nicking product and covalent adduct (Figure 8B). The same effect was achieved by supplementing with 1 mM MgCl<sub>2</sub>, although the advance of the reaction was significantly lower (not shown). By contrast, an excess of OBD protein relative to the 27-mer substrate (molar ratio of 20:1) rendered, both at low and high MnCl<sub>2</sub> concentration, a pattern of reaction products where the two types of products coexisted (Figure 8B).

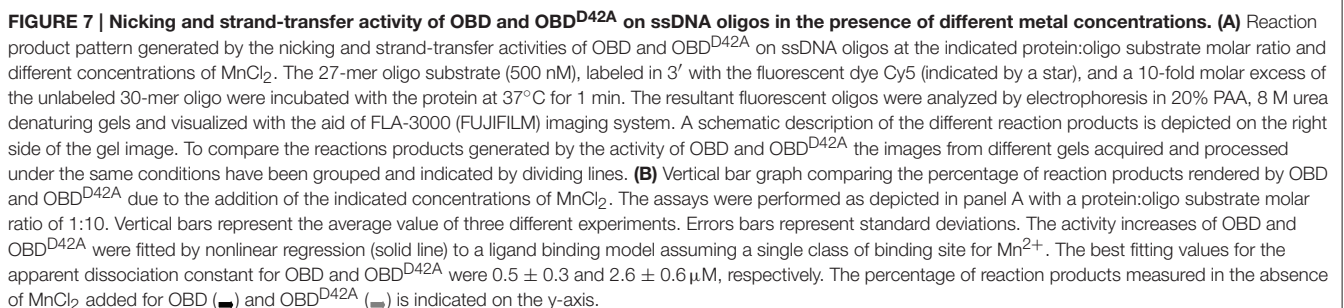
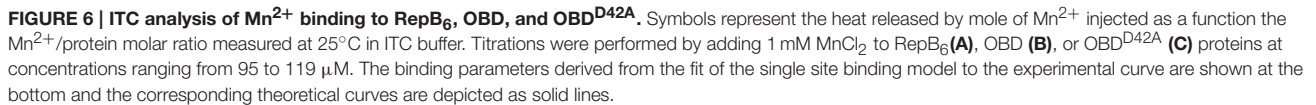
## DISCUSSION

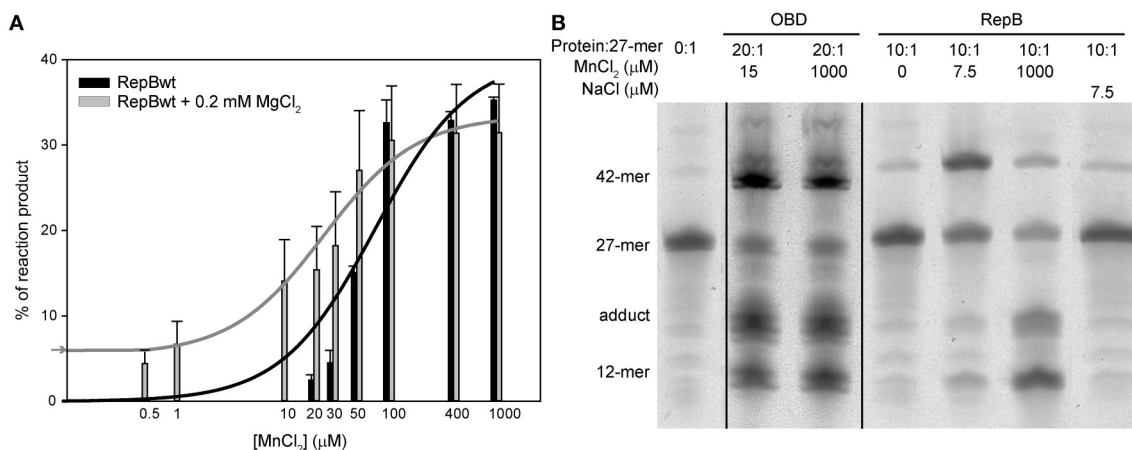
### Influence of Mn<sup>2+</sup> in the Structural Stability of OBD and RepB<sub>6</sub>

Thermal denaturation of RepB<sub>6</sub> takes place in two irreversible steps. The first one leads to an inactive form of the protein, and the second one results in protein precipitation (Figure 1). Further characterization of RepB<sub>6</sub> and of its separate OBD and OD domains showed that the first conformational change

exclusively affects the endonuclease domain, impairing its dsDNA binding and ssDNA catalytic activities (Figures 2, 5). The process reflects OBD denaturation, based on DSC data and near-UV CD spectroscopic changes, although in overall RepB the domain secondary structure seems to be largely preserved (Figure 1). The low stability of the OBD domain, whose thermal denaturation takes place with a T<sub>m</sub> of 39.5°C, contrasts with the high thermostability of the oligomerization domain, which maintains its native structure at temperatures as high as 80°C (Figure 1 and Supplementary Figure 2).

Mn<sup>2+</sup> binding results in a strong thermal stabilization of the endonuclease domain, both in its separate form (OBD protein) and in full-length RepB<sub>6</sub> (Figures 1, 5), which likely correlates with saturation of one high affinity site of RepB per protomer, as measured by ITC (*K*<sub>d</sub> ~40 nM; Figure 6). Other divalent cations, like Mg<sup>2+</sup> and Ca<sup>2+</sup>, failed to stabilize RepB<sub>6</sub> against thermal denaturation and Mg<sup>2+</sup> binding was not observable by ITC, which pointed to their incapacity to bind RepB<sub>6</sub> with high affinity. The protective effect of Mn<sup>2+</sup> binding on the RepB structure likely results from the stabilization of the four protein ligands at the active site (His39, Asp42, His55, and His57; Boer et al., 2009) and of their coordination spheres. Three out of





**FIGURE 8 | Nicking and strand-transfer activity of RepB<sub>6</sub> on ssDNA oligos in the presence of different metal cation concentrations. (A)** The vertical bar graph shows the percentage of reaction products rendered by RepB<sub>6</sub>, in the presence or in the absence of 0.2 mM of MgCl<sub>2</sub>, when the indicated concentrations of MnCl<sub>2</sub> were added. The 27-mer oligo substrate (500 nM), labeled in 3' with the fluorescent dye Cy5, and a 10-fold molar excess of the unlabeled 30-mer oligo were incubated with the protein at 37°C for 1 min. The assays were performed at a protein:oligo substrate molar ratio of 1:10. Vertical bars represent the average value of three different measurements and error bars are standard deviations. The activity curves of RepB<sub>6</sub> supplemented or not with MgCl<sub>2</sub> were fitted by nonlinear regression (solid lines) to a ligand binding model assuming a single class of binding site for Mn<sup>2+</sup>. The best fit values for the apparent dissociation constant for RepB<sub>6</sub> was 24.1 ± 6.9 μM in the presence of 0.2 mM of MgCl<sub>2</sub>, and 80.5 ± 26.2 μM in the absence of MgCl<sub>2</sub>. The percentage of reaction products measured in the presence of 0.2 mM of MgCl<sub>2</sub> with no MnCl<sub>2</sub> added (▬) is indicated on the y-axis. **(B)** Reaction product pattern generated by the nicking and strand-transfer activities of RepB<sub>6</sub> and OBD on ssDNA oligos. The assays were performed as those depicted in (A), at the protein:oligo substrate molar ratio and MnCl<sub>2</sub> concentration indicated on the top of each lane. As a control, the reaction was also carried out with NaCl instead of manganese salt. The resultant fluorescent oligos were analyzed, visualized and quantified as in Figure 7. To compare the reactions products generated by the activity of OBD and RepB<sub>6</sub> the images from different gels acquired and processed under the same conditions have been grouped and indicated by dividing lines.

these four ligand residues are linked through several hydrogen bonds. Namely His39 and His55 main chains are interconnected through two H-bonds, whereas the carboxyl group of Asp42 is hydrogen-bonded to the side chains of His55 and Tyr115. Besides, His39 side chain makes a hydrogen bond with the carbonyl oxygen of Leu100, and the carbonyl group of Asp42 is hydrogen-bonded to the main-chain amide-N of Ser44, whose hydroxyl oxygen is connected, in turn, to the main-chain amide-N of Lys50. Additionally, His57 and Ser36 residues form three hydrogen bonds through their main chains and side chains (Supplementary Figure 1). Hence, by stabilizing this network of polar contacts, the Mn<sup>2+</sup> cation contributes as well to held in place the flexible 21-residue loop that connects strands β2 and β3, and the region comprised between helix α3 and 3<sub>10</sub>-helix η2, both of them flanking the active site (Boer et al., 2009). Moreover, the global conformation of this region might be altered in the metal-free form of OBD<sup>D42A</sup>, thereby explaining the affinity decrease derived from the loss of a metal ligand, as well as the differences found in the enthalpy and entropy of Mn<sup>2+</sup> binding to OBD<sup>D42A</sup> with respect to OBD (Figure 6). The D42A variant of OBD not only retains the Mn<sup>2+</sup> binding capacity but also the catalytic activity, which amounted to ~90% of wild-type OBD under Mn<sup>2+</sup> saturating concentrations (Figure 7). Therefore, the Asp42 moiety, although not essential for metal binding, contributes significantly to the high affinity of the cation and helps to maintain the architecture of the catalytic groove. Of note, the Tyr115 moiety, hydrogen-bonded to Asp42, is conserved among the Rep proteins of the pMV158 family. The architectural role of

this interaction would be also consistent with the lack of nicking activity showed by the Rep protein variant Y116W of pJB01, a *Enterococcus faecium* plasmid belonging to the pMV158 replicon family, which was formerly attributed to the involvement of Tyr116 (equivalent to Tyr 115 of pMV158 RepB) in the catalytic reaction (Kim et al., 2006).

Mn<sup>2+</sup> also binds tightly to other HUH endonucleases like Rep of AAV5, TraI of F, minMobA of R1162, and MobM of pMV158, though with about one-twentieth the affinity for RepB (Hickman et al., 2002; Larkin et al., 2007; Xia and Robertus, 2009; Lorenzo-Díaz et al., 2011). As for OBD<sup>D42A</sup>, Mn<sup>2+</sup> binding to MobM was enthalpically driven, which indicated that cation binding triggered a conformational rearrangement of MobM structure (Lorenzo-Díaz et al., 2011). Metal-induced stabilization has been proved also in some HUH endonucleases of the Mob class. Mn<sup>2+</sup> gave the greatest stabilization of minMobA and MobM (Xia and Robertus, 2009; Lorenzo-Díaz et al., 2011), although their protection was significantly lower than that induced in RepB at equal cation concentration.

## Conservation of Mn<sup>2+</sup> Binding Traits within the HUH Endonuclease Superfamily

Configuration of the active site of HUH endonucleases results from the spatial arrangement of a divalent cation and amino acids from several conserved motifs. The presence of an acidic residue involved either directly or indirectly in metal coordination seems to be a common feature in Mg<sup>2+</sup> or Mn<sup>2+</sup> binding

proteins. Three neutral His side chains coordinating the metal is the configuration most widely conserved among relaxases characterized so far, with the exception of MbeA from plasmid ColE1, with a HEN signature substituting the canonical His triad (Varsaki et al., 2003). Moreover, the interaction through a hydrogen bond between a conserved Asp residue (Asp81) and a His of the 3-His cluster in the active site of relaxase TraI of F seems to do more than orient the His to coordinate the metal. It probably modulates the charge of the His on the metal, allowing a greater polarization of the scissile phosphate bond (Larkin et al., 2007). Substitution of any of the residues of the 3-His cluster by Ala results in no detectable metal binding in TraI of F or minMobA of R1162 (Larkin et al., 2007; Xia and Robertus, 2009). By contrary, the D81A variant of TraI of F binds Mn<sup>2+</sup> with lower affinity than the wild type enzyme and displays a conditional phenotype, exhibiting minimal activity with MgCl<sub>2</sub> but wild-type activity with MnCl<sub>2</sub> (Larkin et al., 2007), which reminds our results for the OBD<sup>D42A</sup> mutant. In this line, substitution of any of the three His residues of the RepB<sub>6</sub> metal binding pocket yielded unstable protein variants that precipitated irreversibly upon being overproduced (not shown). In the case of the viral Rep initiators, the His residue which does not belongs to the HUH motif is replaced by an acidic residue. Thus, the metal bound at the active site of Rep of AAV5 is coordinated by two His (89 and 91) and the acidic side chain of Glu82, whose independent substitution results in no detectable binding of Mn<sup>2+</sup> (Hickman et al., 2002). Similarly, substitution of Glu83 of AAV2-Rep68 (equivalent to Glu82 in AAV5-Rep) by alanine severely impaired the nicking activity of AAV2-Rep68, but residual activity was observed in the presence of Mn<sup>2+</sup> (Yoon-Robarts and Linden, 2003).

## Role of Metal Cations on RepB Activity

Divalent metals could play a role in the proper positioning of the substrate within the catalytic cavity by neutralizing the charges of the ssDNA substrate. They could help also to orient the catalytic residue/s or enhance the polarization of the scissile phosphate bond. Despite Mn<sup>2+</sup> high binding affinity, no nicking or strand-transfer activity were detected in RepB<sub>6</sub> at cation concentrations below 20 μM, which were yet expected to saturate the metal site located at the active center, considering the  $K_{d, 37^{\circ}\text{C}}$  value extrapolated from ITC titration data. Moreover, the concentration of MnCl<sub>2</sub> required for RepB<sub>6</sub> half-maximal activity exceeded by three orders of magnitude the  $K_{d, 37^{\circ}\text{C}}$  value (Figure 8). One possibility to explain this apparent discrepancy was that, at these quite low Mn<sup>2+</sup> concentrations, rejoining of the cleaved 23-mer substrate by full-length RepB<sub>6</sub> predominated over the strand transfer activity, even when the strand-transfer oligo substrate was in a 10-fold molar excess relative to that harboring the nicking sequence. However, a further increase of the strand-transfer substrate up to a 100-fold molar excess did not increase RepB<sub>6</sub> catalytic activity (not shown), making this hypothesis unlikely. As such inconsistency did not exist in the OBD<sup>D42A</sup> mutant and was largely attenuated in the separate endonuclease domain OBD, the distinct protein configuration inherent to each structure might underlie their different behavior. In this context, binding of metal cations to secondary binding

sites located at the interface of the RepB<sub>6</sub> domains and/or protomers, or even between RepB<sub>6</sub> and the substrate DNA, so that full nicking activity would be reached only when high and low affinity sites become saturated, could explain the apparent inconsistency between the Mn<sup>2+</sup> binding affinity of RepB<sub>6</sub> calculated from ITC and the enzymatic assays. In contrast with this, the enhancement of OBD (or OBD<sup>D42A</sup>) activity promoted by Mn<sup>2+</sup> (Figure 7) most probably derives from the cation binding to the active site.

The structure of the RepB hexamer reveals a high degree of conformational plasticity, allowing differences of up to 55° in the orientation of the OBDs relative to the ODs (Boer et al., 2009, 2016). As a result, RepB<sub>6</sub> can exist at least in two distinct structural conformations (C2 and C3 structures). The movement of the OBDs and their position relative to the ODs is mainly determined by the flexibility of the hinge region connecting both domains and by the distinct interactions created between the OBD and the hinge region of a protomer and the OD helix α5 of a neighboring protomer (Boer et al., 2016). Interestingly, a divalent cation can bind to this region through the backbone of the hinge region of a protomer and side chains of residues from the own OD and that of an adjacent protomer in an OBD conformation-dependent way. Indeed, null, half or full site occupancy by Mg<sup>2+</sup> or Ba<sup>2+</sup> has been observed, respectively, for the inward, intermediate and outward positions of the OBD domains in RepB<sub>6</sub> C2 structure (Boer et al., 2016). The role of this metal binding site in the orientation of the OBD domains or RepB<sub>6</sub> activity is presently unknown. However, C2 and C3 structures of RepB<sub>6</sub> were obtained in crystallization buffers with different divalent metal conditions. So, it is tempting to speculate on the possibility that the metal bound to this second site, which does not exist in separate OBD, could influence the activity of RepB<sub>6</sub>, accounting for both the high cation concentrations required for full activity on ssDNA oligos and the ability of Mg<sup>2+</sup> to enhance the activity of RepB<sub>6</sub> at non-saturating concentrations of Mn<sup>2+</sup> (Figure 8). Characterization of Mob class proteins like MobM or minMobA also suggested the uptake of additional cations for maximal nicking activity (Xia and Robertus, 2009; Lorenzo-Díaz et al., 2011). The OBD movements and the structural elements controlling its relative orientation within RepB<sub>6</sub> have been suggested to play an important role in the adaptive capacity of RepB to bind diverse DNA structures within the replication origin (Boer et al., 2009). Moreover, the presence of the hinge region in other initiators suggests that it may be a common, crucial structural element for the binding and manipulation of DNA (Boer et al., 2016).

Notably, the pattern of reaction products generated by the activity of RepB<sub>6</sub> on ssDNA oligos when using an excess of protein respect to the substrate depends on MnCl<sub>2</sub> concentration. The reaction shifted from favoring formation of the strand transfer product at low (7.5 μM) MnCl<sub>2</sub> concentration toward formation of the nicking product and the covalent adduct at 1 mM of MnCl<sub>2</sub> (Figure 8). We hypothesize that a high concentration of divalent metals may reduce the retention of the strand-transfer oligo substrate near the active site of RepB<sub>6</sub>, thereby avoiding the strand transfer reaction. The fact that OBD activity, at the same protein:substrate molar ratio used for RepB<sub>6</sub>,



resulted in similar proportions of strand transfer and nicking products, independently of the MnCl<sub>2</sub> concentration, further support the notion that the path followed by the substrate oligo is different in RepB<sub>6</sub> and OBD, either due to the presence of the OD domain or to the structural/mechanistic implications of its incorporation into the RepB<sub>6</sub> hexamer. Interestingly, very high concentrations of MnCl<sub>2</sub> or MgCl<sub>2</sub> ( $\geq 10$  mM) decreased the activity of OBD and RepB<sub>6</sub> (not shown) probably because they prevented the interaction of the protein with the substrate ssDNA.

## Biological Relevance of Mn<sup>2+</sup> in pMV158 Replication

The physiologically relevant metal for Rep and Mob proteins of the HUH endonuclease superfamily is uncertain, as illustrates the variety of metal cations (Mg<sup>2+</sup>, Mn<sup>2+</sup>, Zn<sup>2+</sup>, or Ni<sup>2+</sup>, among others) found in the active site of the HUH endonucleases whose structure has been solved (Hickman et al., 2002; Datta et al., 2003; Boer et al., 2006, 2009; Monzingo et al., 2007; Vega-Rocha et al., 2007; Nash et al., 2010; Francia et al., 2013).

Inside the pMV158 family of replication initiators, information on metal ion recognition has been provided for RepB of pMV158 (this work) and RepB of pJB01 (Kim et al., 2006), which is also active with Mn<sup>2+</sup>. In addition, MobM, the other nucleotidyl-transferase encoded by pMV158, also requires Mn<sup>2+</sup> for its optimal activity (Lorenzo-Díaz et al., 2011). This paucity of information about the preference for cation usage makes difficult to discern whether the selection of the cation reflects either a preference of the Rep proteins of the pMV158 replicon family or a greater availability of Mn<sup>2+</sup> in the particular cellular environment. In this sense, inductively coupled plasma mass spectrometry (ICP-MS) analysis revealed milimolar concentrations of cell-associated Mn<sup>2+</sup> in *Streptococcus pneumoniae* (Jacobsen et al., 2011). Mn<sup>2+</sup> cations are known to be required *in vivo* for several cellular processes of this bacterium, like capsule formation, metabolism and detoxification, and its cellular homeostasis is maintained even when the extracellular Mn<sup>2+</sup> is depleted (Jacobsen et al., 2011). Therefore, the availability of such a high concentration of intracellular Mn<sup>2+</sup> is consistent with the relevance of this cation for certain DNA transactions, such as replication, conjugation or recombination, performed in this bacterium.

## CONCLUDING REMARKS

Here we report the characterization of the activity and thermal stability of the endonuclease domain of RepB, the initiator protein representative of the pMV158 replicon family of RCR plasmids. RepB is shown to consist of a thermolabile (N-terminal catalytic OBD) and a thermostable (C-terminal hexamerization OD) domain. Binding of Mn<sup>2+</sup> to the active center of the protein protects the OBD from undergoing a conformational change that implies loss of its tertiary structure and renders the protein both catalytically inactive and unable to recognize the plasmid origin. The Asp42 residue, which is one of the Mn<sup>2+</sup> ligands in the active center of RepB, was found to be involved in high affinity binding

of the divalent cation. Saturation of both the high affinity Mn<sup>2+</sup> binding site at the active center and the lower affinity additional site(s) seems to be required for maximal activity of full-length hexameric RepB.

## EXPERIMENTAL PROCEDURES

### Construction of OBD<sup>D42A</sup>

GeneTailor™ System (Invitrogen) was used to perform site-directed mutagenesis. The mutants were generated by replacement of Asp42 by Ala in the active site of OBD. DNA of plasmid pQE1-OBD (Boer et al., 2009), employed to overproduce OBD, was used as template in the mutagenesis reactions. Overlapping primers were designed following the manufacturer's specifications. The expected mutation was confirmed by DNA sequencing and the resultant mutant was purified as indicated below.

### Protein Purification

RepB<sub>6</sub>, OBD, and OD were purified as described previously (Ruiz-Masó et al., 2004; Boer et al., 2009). OBD<sup>D42A</sup> was purified following the protocol used for the respective wild type form. To study OBD and OBD<sup>D42A</sup> Mn<sup>2+</sup>-binding affinities by ITC, as well as the effect of Mn<sup>2+</sup> addition on their catalytic activities, the N-terminal His-tags of OBD and OBD<sup>D42A</sup> were completely removed by using the exoproteolytic enzymes of the TAGZyme system (Unizyme). Protein concentrations were measured spectrophotometrically using the theoretical molar absorption coefficients. Concentrations of RepB<sub>6</sub> given throughout the text refer to total protomers.

### Activity of RepB on Supercoiled DNA

Mixtures of RepB protein (4 pmol) and pMV158 DNA (0.2 pmol) were incubated in a total volume of 30  $\mu$ l of buffer B (20 mM Tris-HCl, pH 8.0, 5 mM DTT) supplemented with 100 mM of KCl and different concentrations of MnCl<sub>2</sub> (ranging from 0.2 to 20 mM) for 30 min at 37°C or 60°C. After incubation, samples were treated with Proteinase K (125  $\mu$ g/ml) for 10 min at 23°C and mixed with sample loading buffer. Reaction products were analyzed by electrophoresis in 1% agarose gels with 0.5  $\mu$ g/ml ethidium bromide in TBE buffer. DNA bands were visualized with a GelDoc system (Bio-Rad) and the QuantityOne software (Bio-Rad) was used for the quantitative analysis of the fluorescence intensities given by the different plasmid forms.

### Nicking and Strand-Transfer Activities on Single-Stranded Oligonucleotides

For cleavage and strand-transfer assays, 10 pmol of the 27-mer oligo substrate 5'-TGCTTCCGTAACG/ACCCCCATTAA-3' (where "/" indicates the RepB nick-site) fluorescently labeled with Cy5 were mixed with 100 pmol of an unlabeled 30-mer oligo 5'-TACTGCGGAATTCTGCTTCCATCTACTACG-3' that provided the 3'-OH substrate for strand transfer, thus avoiding the re-joining of the 27-mer oligo. The mixture was incubated for 1 min at 37°C with RepB<sub>6</sub> (1 pmol of protomers),

OBD, or OBD<sup>D42A</sup> (1 pmol) in 20 µl of buffer B supplemented with a final concentration of 300 mM NaCl and containing different concentrations of divalent metal salts. Protein samples were previously diluted in 20 mM of Tris-HCl buffer (pH 8.0) supplemented with 430 mM of NaCl and 0.2 mg/ml of BSA. After incubation for 1 min at 37°C, the reaction mixtures were treated with proteinase K (60 µg/ml) and 0.05% of SDS for 10 min at 37°C. Prior to electrophoresis, the samples were mixed with 10× DNA loading buffer without dye and denatured by heating at 95°C for 3 min. The products were separated on 20% PAA (19:1 acrylamide:bis-acrylamide), 8 M urea denaturing gels. After electrophoresis, the gels were analyzed by using a FLA-3000 (FUJIFILM) imaging system and the QuantityOne software (Bio-Rad) to quantify the reaction products.

For cleavage assays at different temperatures, 3 pmol of the 23-mer oligo substrate 5'-TGCTTCGTACTACG/ACCCCCCA-3' (where "/" indicates the RepB nick-site) labeled with <sup>32</sup>P at the 5' end using T4 polynucleotide kinase (Sambrook et al., 1989) was incubated for 10 min at 30, 37, and 60°C with different amounts of RepB<sub>6</sub>, ranging from 0.6 to 12 pmol of protomers, in 30 µl of buffer B supplemented with a final concentration of 300 mM NaCl and containing 10 mM of MnCl<sub>2</sub>. After incubation, the reaction mixtures were treated with 60 mM of EDTA and immediately frozen in a mixture of ethanol and dry ice. The reaction products were recovered by ethanol precipitation in the presence of 0.3 M of sodium acetate pH 7. The pellet was washed with 70% ethanol and dissolved in 2× loading buffer (95% formamide, 100 mM EDTA, 0.5% bromophenol blue, 2.5% xylene cyanol). The samples were denatured by heating at 95°C for 3 min and separated as described above.

## EMSA Assays

Reactions to analyze the dsDNA binding capacity of RepB<sub>6</sub> after being heated or not to 45°C were performed in buffer B supplemented with 300 mM of KCl containing 2.4 µM of RepB<sub>6</sub> and 0.4 µM of 42-bp oligonucleotide (42-bind) carrying the *bind* locus (coordinates 529–570 of the pMV158 DNA sequence). After 30 min at 25°C, free and bound DNAs were separated by electrophoresis on native 5% PAA gels. The gels were stained with ethidium bromide and the DNA bands were visualized by fluorescence.

## CD Assays

CD measurements were performed in a J-810 spectropolarimeter (Jasco Corp.) fitted with a peltier temperature controller, using 1-mm or 10-mm path-length cells for far- and near-UV data acquisition, respectively. To analyze the temperature-associated changes in secondary structure, purified proteins were dialyzed at 4°C against buffer CD (20 mM HEPES, pH 8.0, 4.5% ammonium sulfate, 5% ethylene glycol) containing Chelex-100 (0.14% w/v), and then supplemented with various concentrations of Mn<sup>2+</sup> by addition of small volumes of concentrated MnCl<sub>2</sub> stocks prepared in the same buffer. His-tagged OBD and OBD<sup>D42A</sup> were subjected to an extra-chelating treatment to eliminate trace amounts of divalent metals. Briefly, after purification, the protein samples were incubated with 10-fold molar excess of EDTA

for 1 h at 4°C and then dialyzed against buffer CD containing Chelex-100 (0.14% w/v).

CD spectra (average of 4 scans) were acquired using a scan rate of 20 nm min<sup>-1</sup>, a response time of 4 s and a bandwidth of 1 nm. Thermal denaturation experiments were carried out by increasing the temperature from 20 to 95°C at a heating rate of 40°C/h and allowing the cell to equilibrate for 60 s before recording the ellipticity at the selected wavelength. Spectra were recorded in parallel from 20 to 95°C with temperature increments of 10°C, allowing the temperature to equilibrate for 1 min before spectrum acquisition. Buffer contribution was subtracted from the experimental data, and the corrected ellipticity was converted to mean residue ellipticity unless otherwise stated. Data acquisition and processing were carried out using Jasco Spectra-Manager software. Phenomenological description of thermal denaturation profiles was carried out by means of Equation (1) using the Origin software (Microcal Inc.):

$$\Theta = \Theta_D(T) - [\Theta_D(T) - \Theta_N(T)] / \{1 + \exp[A(T - T_{1/2})/RTT_{1/2}]\} \quad (1)$$

where  $\Theta_D(T)$  and  $\Theta_N(T)$  are the ellipticities values of the denatured and native states of the protein at the absolute temperature  $T$ ,  $T_{1/2}$  is the half-transition temperature,  $R$  is the gas constant, and  $A$  accounts for the transition cooperativity.  $\Theta_D(T)$  and  $\Theta_N(T)$  values in Equation (1) were approximated as linear functions of  $T$  (Ruiz et al., 2014).

## Calorimetric Studies

Mn<sup>2+</sup> binding to RepB, OBD and OBD<sup>D42A</sup> was studied at 25°C by ITC using a VP-ITC microcalorimeter (GE Healthcare, Madrid, Spain). Before measurements, the proteins were exhaustively dialysed at 4°C against buffer ITC (20 mM HEPES, pH 7.6, 400 mM KCl) containing Chelex-100 (0.14% w/v) and MnCl<sub>2</sub> solutions were prepared in the final dialysate after removing Chelex-100. Titrations were performed by stepwise injection of 1 mM MnCl<sub>2</sub> solution into the reaction cell loaded with the protein at concentrations of 95–119 µM. Typically, 13 × 7 µl injections followed by several 15 µl injections were performed for RepB<sub>6</sub> and OBD, and 27 × 10 µl for OBD<sup>D42A</sup>, while stirring at 307 rpm. The heat of MnCl<sub>2</sub> dilution was determined in separate runs and subtracted from the total heat produced following each injection. The experiments were carried out at 25°C. Data acquisition and analysis were carried out using the ITC-Viewer and Origin-ITC softwares (GE Healthcare). Mn<sup>2+</sup> dissociation constants at 37°C were extrapolated from ITC data by means of the van't Hoff equation assuming that binding occurred without heat capacity change.

DSC measurements were performed at a heating rate of 60°C/h in a VP-DSC microcalorimeter (Microcal Inc.), at a constant pressure of 2 atm. RepB was equilibrated in CD buffer supplemented with the required Mn<sup>2+</sup> concentration. Microcal DSC-Viewer and Origin-DSC software was used for data acquisition and analysis. Excess heat capacity functions were obtained after subtraction of the buffer-buffer base line and transformed into molar heat capacities dividing by the number of moles of RepB in the DSC cell.

## AUTHOR CONTRIBUTIONS

JR and LB purified RepB and OBD. MS cloned and purified the mutant OBD<sup>D42A</sup>. MM and JR performed the CD and calorimetric assays and analyzed the data. JR performed the protein activity assays with supercoiled plasmid DNA. JR and LB performed the protein activity assays with single stranded oligos. JR, MM, and GdS designed the study and wrote the article. All authors discussed the results, edited, and approved the manuscript.

## FUNDING

This work was supported by grants from the Spanish Ministry of Economy and Competitiveness (BFU2015-70052-R to

MM; BFU2010-19597 to GdS, AGL2012-40084-C03 and AGL2015-71923-REDT to GdS). Additional funding to MM was provided by the CIBER de Enfermedades Respiratorias (CIBERES), an initiative of the Instituto de Salud Carlos III (ISCIII).

## ACKNOWLEDGMENTS

Thanks are due to Prof. M. Espinosa for fruitful discussions.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmolb.2016.00056>

## REFERENCES

- Boer, D. R., Ruiz-Masó, J. A., Gomez-Blanco, J. R., Blanco, A. G., Vives-Llàcer, M., Chacón, P., et al. (2009). Plasmid replication initiator RepB forms a hexamer reminiscent of ring helicases and has mobile nuclease domains. *EMBO J.* 28, 1666–1678. doi: 10.1038/emboj.2009.125
- Boer, D. R., Ruiz-Masó, J. A., Rueda, M., Petoukhov, M. V., Machón, C., Svergun, D. I., et al. (2016). Conformational plasticity of RepB, the replication initiator protein of promiscuous streptococcal plasmid pMV158. *Sci. Rep.* 6:20915. doi: 10.1038/srep20915
- Boer, R., Russi, S., Guasch, A., Lucas, M., Blanco, A. G., Pérez-Luque, R., et al. (2006). Unveiling the molecular mechanism of a conjugative relaxase: the structure of TrwC complexed with a 27-mer DNA comprising the recognition hairpin and the cleavage site. *J. Mol. Biol.* 358, 857. doi: 10.1016/j.jmb.2006.02.018
- Brandts, J. F., Hu, C. Q., Lin, L. N., and Mos, M. T. (1989). A simple model for proteins with interacting domains. Applications to scanning calorimetry data. *Biochemistry* 28, 8588–8596. doi: 10.1021/bi00447a048
- Campos-Olivas, R., Louis, J. M., Clerot, D., Gronenborn, B., and Gronenborn, A. M. (2002). The structure of a replication initiator unites diverse aspects of nucleic acid metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10310–10315. doi: 10.1073/pnas.152342699
- Carr, S. B., Phillips, S. E., and Thomas, C. D. (2016). Structures of replication initiation proteins from staphylococcal antibiotic resistance plasmids reveal protein asymmetry and flexibility are necessary for replication. *Nucleic Acids Res.* 44, 2417–2428. doi: 10.1093/nar/gkv1539
- Chandler, M., de la Cruz, F., Dyda, F., Hickman, A. B., Moncalian, G., and Ton-Hoang, B. (2013). Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat. Rev. Microbiol.* 11, 525–538. doi: 10.1038/nrmicro3067
- Datta, S., Larkin, C., and Schilbach, J. F. (2003). Structural insights into single-stranded DNA binding and cleavage by F factor TraI. *Structure* 11, 1369–1379. doi: 10.1016/j.str.2003.10.001
- de la Campa, A. G., del Solar, G. H., and Espinosa, M. (1990). Initiation of replication of plasmid pLS1. The initiator protein RepB acts on two distant regions. *J. Mol. Biol.* 213, 247–262. doi: 10.1016/S0022-2836(05)80188-3
- del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M., and Díaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.* 62, 434–464.
- del Solar, G., Moscoso, M., and Espinosa, M. (1993). Rolling circle-replicating plasmids from gram-positive and -negative bacteria: a wall falls. *Mol. Microbiol.* 8, 789–796. doi: 10.1111/j.1365-2958.1993.tb01625.x
- Dyda, F., and Hickman, A. B. (2003). A mob of reps. *Structure* 11, 1310. doi: 10.1016/j.str.2003.10.010
- Francia, M. V., Clewell, D. B., de la Cruz, F., and Moncalian, G. (2013). Catalytic domain of plasmid pAD1 relaxase TraX defines a group of relaxases related to restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13606–13611. doi: 10.1073/pnas.1310037110
- Hickman, A. B., Ronning, D. R., Kotin, R. M., and Dyda, F. (2002). Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep. *Mol. Cell* 10, 327. doi: 10.1016/s1097-2765(02)00592-0
- Ilyina, T. V., and Koonin, E. V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes, and archaeobacteria. *Nucleic Acids Res.* 20, 3279–3285. doi: 10.1093/nar/20.13.3279
- Jacobsen, F. E., Kazmierczak, K. M., Lisher, J. P., Winkler, M. E., and Giedroc, D. P. (2011). Interplay between manganese and zinc homeostasis in the human pathogen *Streptococcus pneumoniae*. *Metallomics* 3, 38–41. doi: 10.1039/C0MT00050G
- Khan, S. A. (2000). Plasmid rolling-circle replication: recent developments. *Mol. Microbiol.* 37, 477–484. doi: 10.1046/j.1365-2958.2000.02001.x
- Khan, S. A. (2005). Plasmid rolling-circle replication: highlights of two decades of research. *Plasmid* 53, 126–136. doi: 10.1016/j.plasmid.2004.12.008
- Kim, S. W., Jeong, E. J., Kang, H. S., Tak, J. I., Bang, W. Y., Heo, J. B., et al. (2006). Role of RepB in the replication of plasmid pJB01 isolated from *Enterococcus faecium* JC1. *Plasmid* 55, 99–113. doi: 10.1016/j.plasmid.2005.08.002
- Larkin, C., Datta, S., Harley, M. J., Anderson, B. J., Ebie, A., Hargreaves, V., et al. (2005). Inter- and intramolecular determinants of the specificity of single-stranded DNA binding and cleavage by the F factor relaxase. *Structure* 13, 1533. doi: 10.1016/j.str.2005.06.013
- Larkin, C., Haft, R. J. F., Harley, M. J., Traxler, B., and Schilbach, J. F. (2007). Roles of active site residues and the HUH motif of the F plasmid TraI relaxase. *J. Biol. Chem.* 282, 33707–33713. doi: 10.1074/jbc.M703210200
- Lorenzo-Díaz, F., Dostál, L., Coll, M., Schilbach, J. F., Menéndez, M., and Espinosa, M. (2011). The MobM relaxase domain of plasmid pMV158: thermal stability and activity upon Mn<sup>2+</sup> and specific DNA binding. *Nucleic Acids Res.* 39, 4315–4329. doi: 10.1093/nar/gkr049
- Monzingo, A. F., Ozburn, A., Xia, S., Meyer, R. J., and Robertus, J. D. (2007). The structure of the minimal relaxase domain of MobA at 2.1 Å resolution. *J. Mol. Biol.* 366, 165–178. doi: 10.1016/j.jmb.2006.11.031
- Moscoso, M., del Solar, G., and Espinosa, M. (1995). Specific nicking-closing activity of the initiator of replication protein RepB of plasmid pMV158 on supercoiled or single-stranded DNA. *J. Biol. Chem.* 270, 3772–3779. doi: 10.1074/jbc.270.8.3772
- Moscoso, M., Eritja, R., and Espinosa, M. (1997). Initiation of replication of plasmid pMV158: mechanisms of DNA strand-transfer reactions mediated by the initiator RepB protein. *J. Mol. Biol.* 268, 840. doi: 10.1006/jmbi.1997.1012
- Nash, R. P., Habibi, S., Cheng, Y., Lujan, S. A., and Redinbo, M. R. (2010). The mechanism and control of DNA transfer by the conjugative relaxase of resistance plasmid pCU1. *Nucleic Acids Res.* 38, 5929–5943. doi: 10.1093/nar/gkq303

- Novick, R. P. (1998). Contrasting lifestyles of rolling-circle phages and plasmids. *Trends Biochem. Sci.* 23, 434–438. doi: 10.1016/S0968-0004(98)01302-4
- Ruiz, F. M., Scholz, B. A., Buzamet, E., Kopitz, J., André, S., Menéndez, M., et al. (2014). Natural single amino acid polymorphism (F19Y) in human galectin-8: detection of structural alterations and increased growth-regulatory activity on tumor cells. *FEBS J.* 281, 1446–1464. doi: 10.1111/febs.12716
- Ruiz-Masó, J. A., López-Zumel, C., Menéndez, M., Espinosa, M., and del Solar, G. (2004). Structural features of the initiator of replication protein RepB encoded by the promiscuous plasmid pMV158. *Biochim. Biophys. Acta* 1696, 113–119. doi: 10.1016/j.bbapap.2003.09.010
- Ruiz-Masó, J. A., Lurz, R., Espinosa, M., and del Solar, G. (2007). Interactions between the RepB initiator protein of plasmid pMV158 and two distant DNA regions within the origin of replication. *Nucleic Acids Res.* 35, 1230–1244. doi: 10.1093/nar/gkl1099
- Ruiz-Masó, J. A., Machón, C., Bordanaba-Ruiseco, L., Espinosa, M., Coll, M., and del Solar, G. (2015). Plasmid rolling-circle replication. *Microbiol. Spectr.* 3:PLAS-0035-2014. doi: 10.1128/microbiolspec.plas-0035-2014
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning. A Laboratory Manual*. 2nd Edn. New York, NY: Cold Spring Harbor Laboratory Press.
- Varsaki, A., Lucas, M., Afendra, A. S., Drainas, C., and de la Cruz, F. (2003). Genetic and biochemical characterization of MbeA, the relaxase involved in plasmid ColE1 conjugative mobilization. *Mol. Microbiol.* 48, 481–493. doi: 10.1046/j.1365-2958.2003.03441.x
- Vega-Rocha, S., Byeon, I. J., Gronenborn, B., Gronenborn, A. M., and Campos-Olivas, R. (2007). Solution structure, divalent metal and DNA binding of the endonuclease domain from the replication initiation protein from porcine circovirus 2. *J. Mol. Biol.* 367, 473. doi: 10.1016/j.jmb.2007.01.002
- Xia, S., and Robertus, J. D. (2009). Effect of divalent ions on the minimal relaxase domain of MobA. *Arch. Biochem. Biophys.* 488, 42–47. doi: 10.1016/j.abb.2009.06.004
- Yoon-Roberts, M., and Linden, R. M. (2003). Identification of active site residues of the adeno-associated virus type 2 Rep endonuclease. *J. Biol. Chem.* 278, 4912–4918. doi: 10.1074/jbc.M209750200

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Ruiz-Masó, Bordanaba-Ruiseco, Sanz, Menéndez and del Solar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# DNA-Binding Proteins Essential for Protein-Primed Bacteriophage $\Phi$ 29 DNA Replication

Margarita Salas\*, Isabel Holguera, Modesto Redrejo-Rodríguez and Miguel de Vega

Centro de Biología Molecular Severo Ochoa (Consejo Superior de Investigaciones Científicas-UAM), Universidad Autónoma de Madrid, Madrid, Spain

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Spanish National Research Council -  
Centro de Investigaciones Biológicas,  
Spain

### Reviewed by:

Juan Carlos Alonso,  
National Center for Biotechnology  
(Consejo Superior de Investigaciones  
Científicas), Spain  
Enrique Viguera Mínguez,  
University of Málaga, Spain

### \*Correspondence:

Margarita Salas  
msalas@cbm.csic.es

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 15 June 2016

**Accepted:** 20 July 2016

**Published:** 05 August 2016

### Citation:

Salas M, Holguera I,  
Redrejo-Rodríguez M and de Vega M  
(2016) DNA-Binding Proteins Essential  
for Protein-Primed Bacteriophage  
 $\Phi$ 29 DNA Replication.  
Front. Mol. Biosci. 3:37.  
doi: 10.3389/fmolb.2016.00037

*Bacillus subtilis* phage  $\Phi$ 29 has a linear, double-stranded DNA 19 kb long with an inverted terminal repeat of 6 nucleotides and a protein covalently linked to the 5' ends of the DNA. This protein, called terminal protein (TP), is the primer for the initiation of replication, a reaction catalyzed by the viral DNA polymerase at the two DNA ends. The DNA polymerase further elongates the nascent DNA chain in a processive manner, coupling strand displacement with elongation. The viral protein p5 is a single-stranded DNA binding protein (SSB) that binds to the single strands generated by strand displacement during the elongation process. Viral protein p6 is a double-stranded DNA binding protein (DBP) that preferentially binds to the origins of replication at the  $\Phi$ 29 DNA ends and is required for the initiation of replication. Both SSB and DBP are essential for  $\Phi$ 29 DNA amplification. This review focuses on the role of these phage DNA-binding proteins in  $\Phi$ 29 DNA replication both *in vitro* and *in vivo*, as well as on the implication of several *B. subtilis* DNA-binding proteins in different processes of the viral cycle. We will revise the enzymatic activities of the  $\Phi$ 29 DNA polymerase: TP-deoxynucleotidylation, processive DNA polymerization coupled to strand displacement, 3'-5' exonucleolysis and pyrophosphorolysis. The resolution of the  $\Phi$ 29 DNA polymerase structure has shed light on the translocation mechanism and the determinants responsible for processivity and strand displacement. These two properties have made  $\Phi$ 29 DNA polymerase one of the main enzymes used in the current DNA amplification technologies. The determination of the structure of  $\Phi$ 29 TP revealed the existence of three domains: the priming domain, where the primer residue Ser232, as well as Phe230, involved in the determination of the initiating nucleotide, are located, the intermediate domain, involved in DNA polymerase binding, and the N-terminal domain, responsible for DNA binding and localization of the TP at the bacterial nucleoid, where viral DNA replication takes place. The biochemical properties of the  $\Phi$ 29 DBP and SSB and their function in the initiation and elongation of  $\Phi$ 29 DNA replication, respectively, will be described.

**Keywords:** bacteriophage  $\Phi$ 29, DNA replication, DNA polymerase, terminal protein, DNA binding proteins

## INTRODUCTION

Bacteriophages are the most abundant biological entities on earth (Brüssow and Hendrix, 2002). Approximately 96% of the reported bacteriophages belong to the order *Caudovirales*, which is composed of three families: *Myoviridae*, *Siphoviridae*, and *Podoviridae* (Ackermann, 2003). *Bacillus subtilis* phage  $\Phi$ 29 belongs to the *Podoviridae* family and to the  $\Phi$ 29-like genus, together with

phages  $\Phi$ 15, PZA, BS32, B103, Nf, M2Y, and GA-1 (Ackermann, 1998). These are the smallest phages that infect *Bacillus*, and they are among the smallest known phages that possess a dsDNA genome (Anderson and Reilly, 1993). Based on its relatedness, these phages have been classified in three groups: group I includes phages  $\Phi$ 29, PZA,  $\Phi$ 15, and BS32; group II contains phages B103, Nf and M2Y; and group III has GA-1 as its only member (Yoshikawa et al., 1985, 1986; Pecenkova and Paces, 1999).

Bacteriophage  $\Phi$ 29 genome consists of a linear dsDNA  $\sim$ 19 Kb-long with a terminal protein (TP) covalently linked to each 5' end (Salas, 1991).  $\Phi$ 29 has served as a model system for studying the protein-priming mechanism of DNA replication, being the TP-primed replication system best characterized *in vitro*. The use of a TP as primer for viral DNA replication has also been described for other bacteriophages (e.g., *Escherichia coli* and *Streptococcus pneumoniae* phages PRD1 and Cp-1, respectively), eukaryotic viruses (adenovirus), and some *Streptomyces* spp. (Chang and Cohen, 1994; Bao and Cohen, 2001). In addition, the presence of TPs has been described or suggested in viruses infecting *Archaea* (Bath et al., 2006; Peng et al., 2007), some linear plasmids of bacteria, fungi, and higher plants (Salas, 1991; Meinhardt et al., 1997; Chaconas and Chen, 2005), transposable elements (Kapitonov and Jurka, 2006) and mitochondrial DNA (Fricova et al., 2010).

Besides the essential role of priming DNA replication, TPs can perform additional functions. It has been shown that adenovirus TP is important for the anchoring of the viral genome to the nuclear matrix, which enhances transcription of the viral DNA (Schaack et al., 1990). TPs have also been shown to be required for DNA packaging (Bjornsti et al., 1982, 1983), transfection (Hirokawa, 1972; Ronda et al., 1983; Porter and Dyall-Smith, 2008), and nucleoid and nuclear targeting (Tsai et al., 2008; Muñoz-Espín et al., 2010; Redrejo-Rodríguez et al., 2012). Furthermore, biochemical studies have suggested that  $\Phi$ 29 TP is endowed with peptidoglycan-hydrolytic activity (Moak and Molineux, 2004).

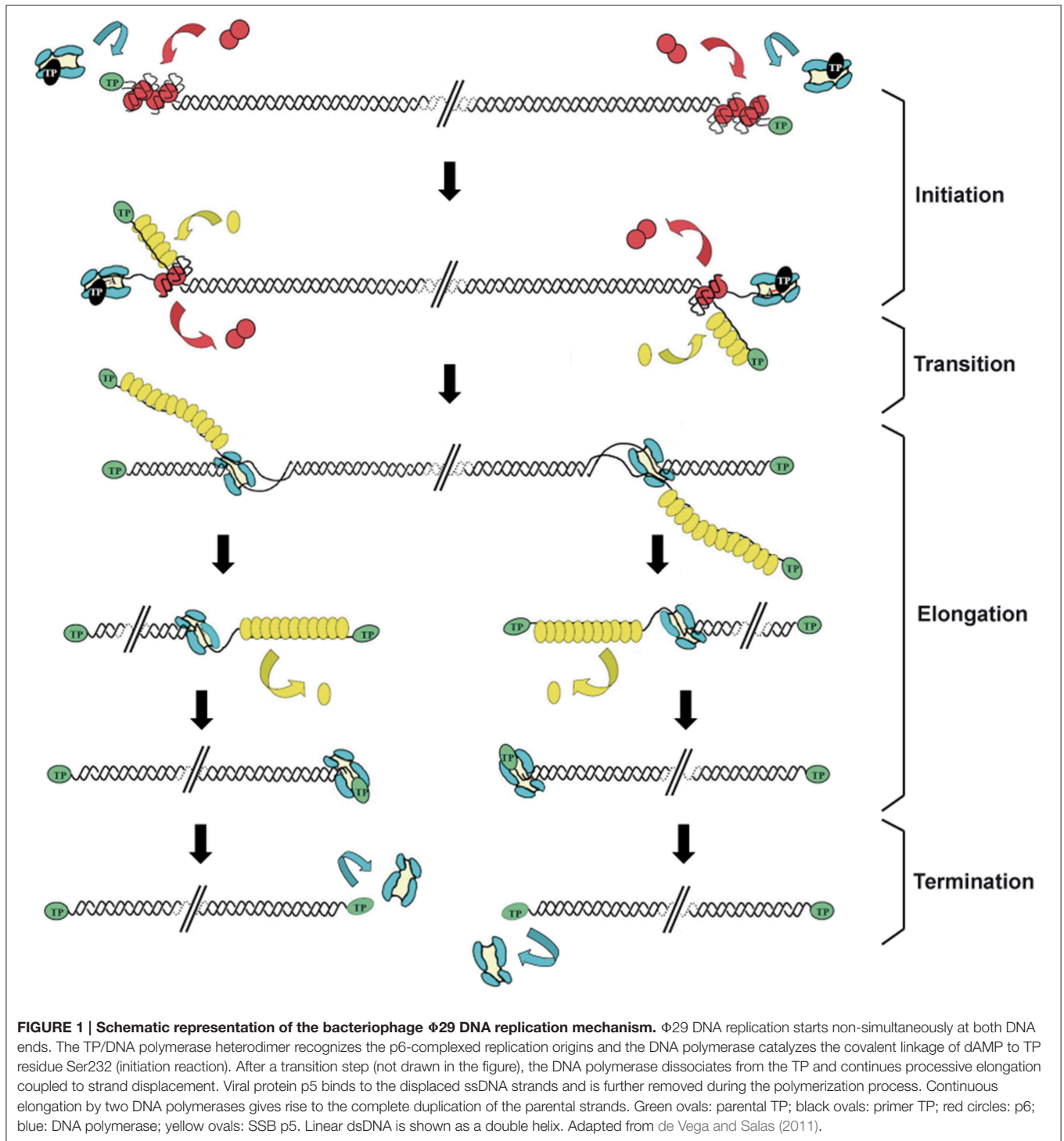
## $\Phi$ 29 TERMINAL PROTEIN

Replication of the  $\Phi$ 29 genome takes place by a process of symmetrical replication in which both origins are used for initiation in a non-simultaneous manner (Blanco et al., 1989; **Figure 1**). The protein that acts as primer for the initiation of  $\Phi$ 29 DNA replication, the so-called TP, is a 266 amino acids protein encoded by the early viral gene 3. The first step of  $\Phi$ 29 DNA replication is the formation of a heterodimer between a free molecule of TP (primer TP) and the DNA polymerase ( $\Phi$ 29 DNAP) (Blanco et al., 1987). Then, this complex recognizes the replication origins, located at both ends of the viral genome, by specific interactions with both the TP that is linked to the genome ends by a previous round of replication (parental TP) and DNA sequences (García et al., 1984; Gutiérrez et al., 1986a,b; González-Huici et al., 2000a,b). The parental TP is the major signal for replication origin recognition by the heterodimer (Gutiérrez et al., 1986b; González-Huici et al., 2000b) and both, DNA polymerase and

primer TP, are involved in such recognition through specific interactions with the parental TP (Freire et al., 1996; Illana et al., 1998; González-Huici et al., 2000a; Serna-Rico et al., 2000; Pérez-Arnáiz et al., 2007). The  $\Phi$ 29 double-stranded DNA binding protein p6 (DBP) (see below) binds all along  $\Phi$ 29 DNA forming a nucleoprotein complex that causes the unwinding of the DNA helix at the ends, facilitating the initiation step (Serrano et al., 1994). After origin recognition, the viral DNA polymerase catalyzes the formation of a phosphoester between the first dAMP and the hydroxyl group of the primer TP residue Ser232 (Blanco and Salas, 1984; Hermoso et al., 1985). The initiation reaction is directed by the second T at the 3' end of the template (3' TTTCAT 5'), after which the TP-dAMP complex translocates one position backwards to recover the information corresponding to the first T of the template strand. Then, the second T will serve again as template for the incorporation of the following nucleotide (Méndez et al., 1992). This backward translocation of the TP-dAMP complex is known as sliding-back mechanism and requires a terminal repetition of at least 2 nucleotides in the template strand to guarantee the fidelity of the initiation reaction (Méndez et al., 1992) (see below). The TP/DNA polymerase heterodimer is not dissociated immediately after initiation. There is a transition stage in which the DNA polymerase synthesizes a 5 nt-long elongation product while complexed with TP, undergoes some structural changes during the incorporation of nucleotides 6 to 9, and finally dissociates from the TP after the incorporation of the 10th nucleotide (Méndez et al., 1997). Then, the viral DNA polymerase continues DNA elongation in a processive manner, which occurs coupled to the displacement of the non-template strand (Blanco et al., 1989). DNA elongation by one  $\Phi$ 29 DNAP coming from each origin generates type I replicative intermediates, consisting of full-length  $\Phi$ 29 dsDNA molecules with two branches of ssDNA. These stretches of ssDNA are bound by the viral single-stranded DNA-binding protein p5 (SSB) (see below) (Gutiérrez et al., 1991), which will be further removed during the polymerization process. When the two replication forks meet, the type I replicative intermediate gives rise to two physically separated type II replicative intermediates. These molecules consist of full-length  $\Phi$ 29 DNA in which a portion of the DNA starting from one end is dsDNA and the portion spanning to the other end is ssDNA (Harding and Ito, 1980; Inciarte et al., 1980). Termination of viral DNA replication occurs when the DNA polymerase reaches the template end, and after replication of the last nucleotide, dissociates from the viral genome.

The 3.0 Å resolution crystallographic structure of the heterodimer formed between  $\Phi$ 29 DNA polymerase and TP revealed that the latter is composed of three structural domains (Kamtekar et al., 2006; see **Figure 2**):

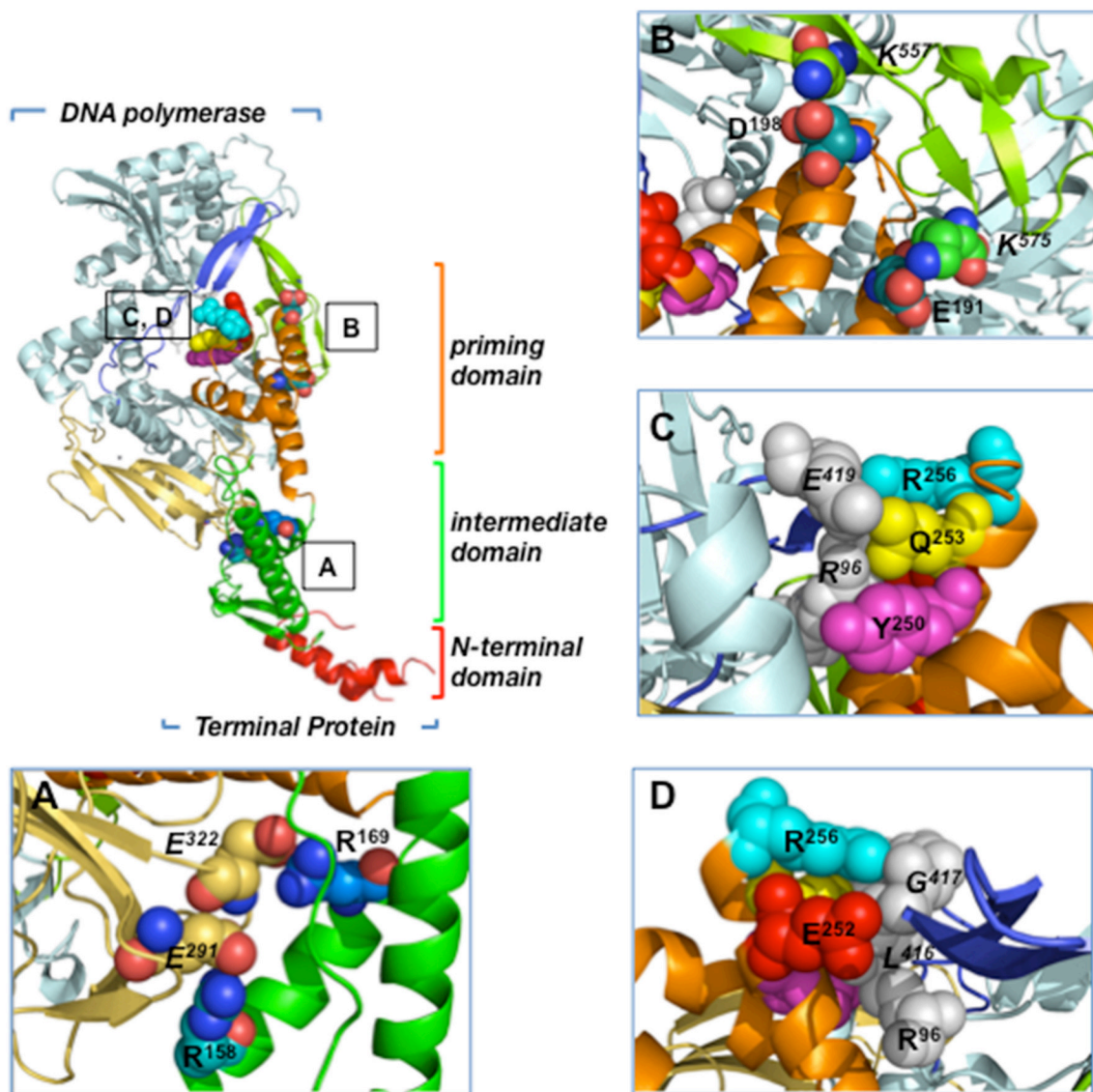
- The TP N-terminal domain comprises residues 1 to 73 and its tertiary structure is unknown because it was disordered in the crystal lattice. Circular dichroism experiments have shown that this domain has a high content in  $\alpha$ helix (60%), and secondary structure predictions determined two  $\alpha$ helices connected by a disordered loop (Holguera et al., 2014). This



domain is responsible for non-sequence specific DNA binding (Zaballos and Salas, 1989) and for the localization of the protein at the bacterial nucleoid (Muñoz-Espín et al., 2010). In addition, a role in origin unwinding has been proposed for the TP N-terminal domain, since this domain is not required for the initiation reaction at a partially open origin (Pérez-Arnáiz et al., 2007; Gella et al., 2014).

- The TP intermediate domain (residues 74–172) is composed of two long  $\alpha$ helices and a short  $\beta$ -turn- $\beta$  structure. This domain makes extensive contacts with the DNA polymerase (mainly with the TPR1 subdomain), being the main responsible for the specificity of the interaction with the DNA polymerase and for the stability of the heterodimer (Pérez-Arnáiz et al., 2007; del Prado et al., 2012).





**FIGURE 2 | Residues involved in  $\Phi$ 29 TP/DNA polymerase interaction.** Ribbon representation of the  $\Phi$ 29 TP/DNA polymerase heterodimer. DNA polymerase is colored in blue except for TPR1 (pale yellow), TPR2 (dark blue) and thumb (light green) subdomains. TP N-terminal, intermediate and priming domains are colored in red, green and orange, respectively. **(A)** Close-up view of the TP intermediate domain residues R158 and R169, proposed to make salt bridges with DNA polymerase residues E291 and E322 of the TPR1 subdomain, respectively. **(B)** Close-up view of the TP priming domain residues E191 and D198, proposed to interact with the DNA polymerase thumb subdomain residues K575 and K557, respectively. **(C)** The proposed stacking interactions between TP priming domain residues R256, Q253, and Y250, and DNA polymerase residues R96 (exonuclease domain) and E419 (TPR2 subdomain) are indicated. **(D)** Close-up view of the proposed interaction between TP priming domain residues R256 and E252 with DNA polymerase residues R96 (exonuclease domain) and L416 and G417 (TPR2 subdomain). Coordinates of the  $\Phi$ 29 TP/DNA polymerase heterodimer are from PDB ID 2EX3 (Kamtekar et al., 2006). The figure was made using the Pymol software (<http://www.pymol.org>).

- The TP C-terminal domain or priming domain (residues 173–266) is connected to the intermediate domain through a hinge region and it is comprised of a four-helix bundle. The TP priming residue Ser232 lies in the so-called priming loop, a disordered loop comprising residues 227–233. The TP priming domain makes extensive interactions with the TPR2 and thumb subdomains of the DNA polymerase (Kamtekar et al., 2006; del Prado et al., 2012) and is responsible for dictating the nucleotide used as template during initiation of viral DNA

replication (Longás et al., 2008). The TP priming domain has been proposed to mimic duplex product DNA in its electrostatic profile and binding site in the DNA polymerase, as both occupy the same binding cleft in the DNA polymerase (de Vega et al., 1998a; Kamtekar et al., 2006).

There are not proteins in structural databases with sufficient structural homology with  $\Phi$ 29 TP. Genes encoding TPs from other  $\Phi$ 29-like phages such as B103, PZA, Nf, and GA-1 have been sequenced (Paces et al., 1985; Leavitt and Ito, 1987; Illana



et al., 1996; Pecenková et al., 1997; Meijer et al., 2001). The amino acid sequence comparison of these TPs has revealed a high degree of conservation between PZA and  $\Phi$ 29 TPs (97.7% identity). In fact,  $\Phi$ 29 TP can functionally substitute for PZA TP *in vivo* (Bravo et al., 1994a). The conservation is lower in the case of Nf (62.4% identity) and B103 (62% identity) TPs (Leavitt and Ito, 1987; Pecenková et al., 1997). GA-1 TP is the most distantly related one, being the percentage of identity with  $\Phi$ 29 TP of 40% (Illana et al., 1996).

## TP Residues Involved in Priming Activity

Site-directed mutagenesis has been carried out at the TP priming residue Ser232. The change of Ser232 to Thr gives rise to a protein completely inactive in the initiation reaction (Garmendia et al., 1988). Similarly, the change of Ser232 into Cys almost completely abolishes the initiation capacity of the TP mutant, being its initiation capacity about 1% of that of the wild-type TP (Garmendia et al., 1990). These TP mutants interacted in a wild-type manner with both  $\Phi$ 29 DNAP and DNA (Garmendia et al., 1988, 1990). Furthermore, mutation of TP priming-loop residues Leu220 and Ser226 into Pro highly impaired the initiation activity but did not affect either DNA polymerase or DNA-binding, suggesting the implication of these residues in the initiation reaction (Garmendia et al., 1990).

## TP Residues Involved in DNA-Binding

$\Phi$ 29 TP binds to both single-stranded and double-stranded DNA *in vitro* (Prieto, 1986; Zaballos and Salas, 1989). As mentioned above, the TP domain responsible for non-specific DNA-binding is the N-terminal domain (Zaballos and Salas, 1989). As in many non-sequence specific DNA-binding proteins, TP N-terminal domain basic residues are implicated in its DNA-binding capacity (Holguera et al., 2014).

Viral DNA replication in prokaryotes takes place at specific subcellular locations. In this sense, the use of host organizing structures seems to be essential to provide an appropriate scaffold for viral DNA replication.  $\Phi$ 29 TP localizes at the bacterial nucleoid along the infective cycle, being the N-terminal domain responsible for this localization (Muñoz-Espín et al., 2010). Additionally, parental TP (and therefore TP-DNA) localizes at the bacterial nucleoid, independently of primer TP (Muñoz-Espín et al., 2010). Importantly, the TP N-terminal domain is essential for an efficient viral DNA replication *in vivo* (Muñoz-Espín et al., 2010). To determine the TP residues involved in nucleoid targeting, each basic residue of the TP N-terminal domain was replaced independently by alanine, and the subcellular localization of the resulting proteins fused to YFP was analyzed. Lys27 was the only TP residue that, changed individually, impaired the TP nucleoid localization (Holguera et al., 2014). By using X-Chip techniques, it was shown that wild-type  $\Phi$ 29 TP, but not mutant K27A, binds *B. subtilis* genome *in vivo*, establishing a correlation between nucleoid localization and DNA-binding (Holguera et al., 2014). During the infective cycle both TP and viral DNA polymerase localize at the bacterial nucleoid, being the nucleoid localization of the DNA polymerase dependent on the expression of TP (Muñoz-Espín et al., 2010). The subcellular localization of the viral

DNA replication machinery at the bacterial nucleoid has been proposed to serve as a compartmentalization mechanism to make the replication process more efficient, as well as a means of taking advantage of the bacterial chromosome segregation dynamics (Muñoz-Espín et al., 2010). The impact of bacterial chromosome TP binding on host processes such as DNA replication and transcription remains to be investigated.

Interestingly,  $\Phi$ 29 TP localizes at the bacterial nucleoid when expressed in the distantly related bacterium *E. coli*, being the TP N-terminal domain the one responsible for this localization (Muñoz-Espín et al., 2010; Redrejo-Rodríguez et al., 2013). Furthermore, the TP from phage PRD1, which infects *E. coli* among other bacteria, localizes at the *E. coli* nucleoid independently of other viral components. TPs from other phages such as Cp-1, Nf, and GA-1 also localize at the *E. coli* nucleoid, although localization in their host systems remains to be determined (Redrejo-Rodríguez et al., 2013). Altogether, these results suggest that nucleoid localization is a functional property conserved in phage TPs. Importantly, a Nuclear Localization Signal (NLS) has been described in  $\Phi$ 29 TP, as well as in a variety of other TPs such as those from Nf, PRD1, Bam35, and Cp-1 phages (Redrejo-Rodríguez et al., 2012).

## TP Residues Involved in DNA Polymerase-Binding

The extensive interactions of  $\Phi$ 29 TP intermediate and priming domains with the DNA polymerase account for the high stability of the heterodimer (Lázaro et al., 1995; Kamtekar et al., 2006). The crystallographic structure of the heterodimer shows that the TP intermediate domain is structurally complementary to the DNA polymerase TPR1 subdomain; this interface has many charged residues that include two salt bridges between TP residues Arg158 and Arg169, and DNA polymerase residues Glu291 and Glu322, respectively (Kamtekar et al., 2006; **Figure 2A**). In the case of the highly electronegative TP priming domain, the structure shows interactions between TP residues Glu191 and Asp198, and DNA polymerase thumb subdomain residues Lys575 and Lys557, respectively (**Figure 2B**). In addition, TP residues Gln253 and Tyr250 would interact with DNA polymerase exonuclease domain residue Arg96 through a hydrogen bond and a stacking interaction, respectively (Kamtekar et al., 2006; **Figure 2C**). In this sense, mutation of DNA polymerase residue Arg96 to alanine was shown to impair the interaction with TP (Rodríguez et al., 2004). Similarly, TP residues Glu252, Gln253, and Arg256 from the C-terminal helix of the priming domain would pack against DNA polymerase TPR2 subdomain residues Leu416, Gly417, and Glu419, respectively (Kamtekar et al., 2006; **Figures 2C,D**). In fact, by biochemical analysis of TP mutants, TP residues Arg158, Arg169, Glu191, Asp198, Tyr250, Glu252, Gln253, and Arg256 were shown to be involved in the interaction between TP and DNA polymerase (del Prado et al., 2012). Additionally, biochemical studies using both TP and DNA polymerase mutant proteins strongly suggest that TP priming loop residue Glu233 interacts directly with the DNA polymerase palm subdomain residue Lys529

during the first step of TP-DNA replication (del Prado et al., 2013).

## TP Interaction with Other Viral Proteins

Apart from the DNA polymerase,  $\Phi$ 29 TP interacts with other viral proteins. By means of *in vitro* chemical crosslinking, it has been shown that  $\Phi$ 29 TP interacts with the viral early protein p1, which is a membrane-associated protein. Based on these results, a model of membrane anchorage of the viral replication machinery mediated by p1 has been proposed (Bravo et al., 2000). In addition, TP interacts with the membrane protein p16.7 *in vitro* (Serna-Rico et al., 2003), presenting another anchoring point to the bacterial membrane. This interaction has also been proposed to facilitate the binding of p16.7 to the displaced strands of the viral genome, favoring their recruitment to the bacterial membrane (Serna-Rico et al., 2003). Mutations introduced at several residues of the TP N-terminal and intermediate domains impaired DNA replication when TP acted simultaneously as primer and parental TP, suggesting that a proper interaction between primer and parental TP is important for origin recognition (Illana et al., 1999; Serna-Rico et al., 2000; del Prado et al., 2012; Holguera et al., 2015).

## $\Phi$ 29 DNA POLYMERASE

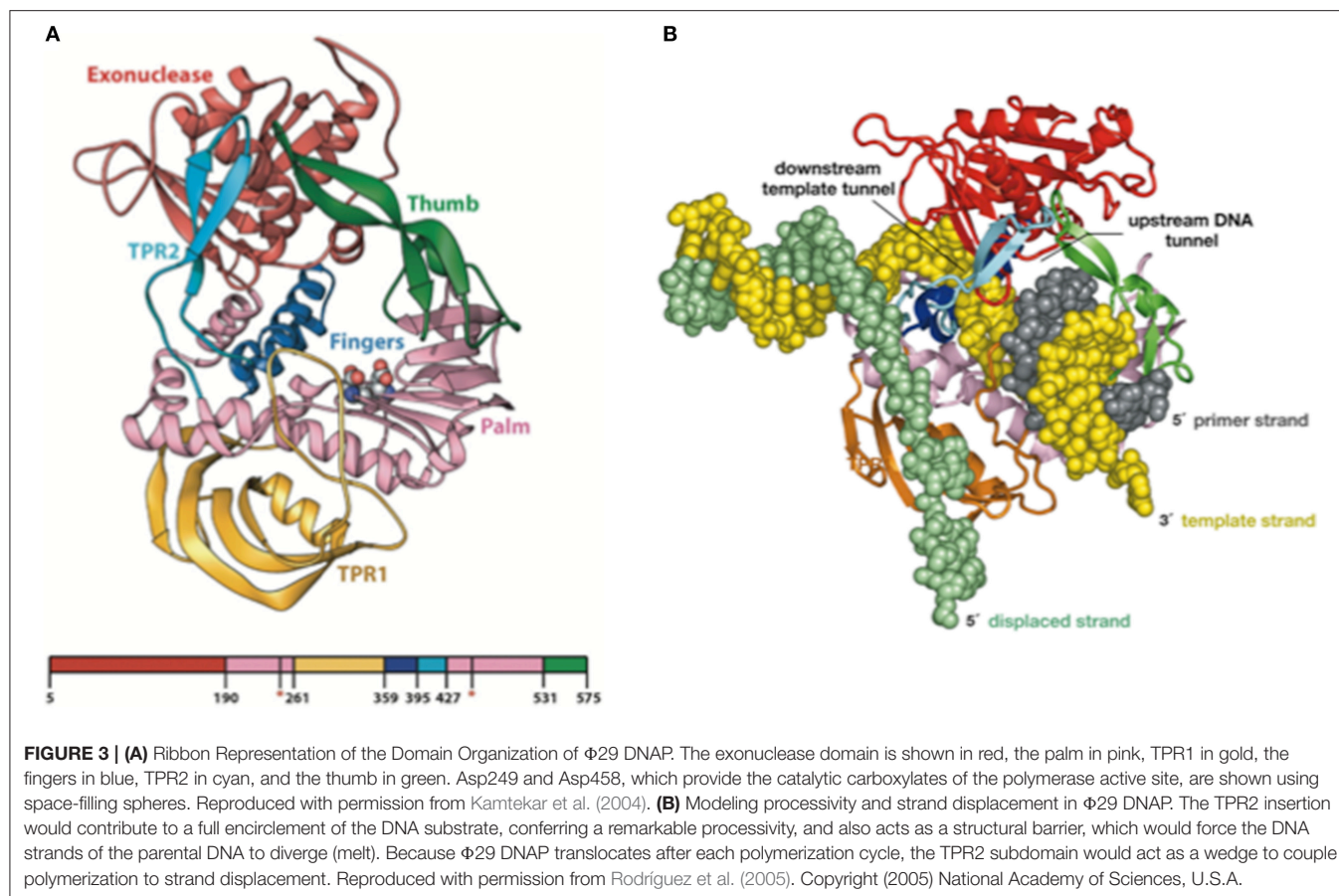
### Processive Polymerization Coupled to Strand Displacement: Two Specific Attributes of $\Phi$ 29 DNA Polymerase

$\Phi$ 29 DNAP is a small (66 kDa) single subunit enzyme, the product of the viral gene 2, characterized as the viral DNA replicase (Blanco and Salas, 1984, 1985b; Salas, 1991), and belonging to the family B (eukaryotic-type) of DNA-dependent DNA polymerases (Blanco and Salas, 1986; Bernad et al., 1987). As any other conventional DNA polymerase,  $\Phi$ 29 DNAP catalyzes the sequential template-directed addition of dNMP units onto the 3'-OH group of a growing DNA chain in a faithful manner as it shows discrimination values of  $10^4$ – $10^6$ , and a poor mismatch elongation efficiency (Esteban et al., 1993). Extensive site directed mutagenesis studies in  $\Phi$ 29 DNAP described the function of specific amino acids at motifs YxGG, Dx<sub>2</sub>SLYP, LEXE, Kx<sub>3</sub>NSxYG, Tx<sub>2</sub>GR, YxDTDS, and KxY, placed at the C-terminal domain (residues 190–572; polymerization domain) and highly conserved among the eukaryotic DNA polymerases from family B (Blanco and Salas, 1995, 1996; Pérez-Arnaiz et al., 2006, 2009, 2010; Salas and de Vega, 2006; del Prado et al., 2013; Santos et al., 2014). These investigations allowed the identification of the catalytic residues responsible for coordinating the metal ions and the ones acting as ligands of the substrates (DNA, TP, and dNTP).

In contrast to the complexity of other *in vitro* replication systems, efficient synthesis of full-length  $\Phi$ 29 TP-DNA can be accomplished *in vitro* with only the presence of TP and  $\Phi$ 29 DNAP (Blanco and Salas, 1985b). The efficiency of this minimal replication system relies on three unique catalytic features of  $\Phi$ 29 DNAP: (1) ability to initiate DNA replication by using a TP as primer (Salas, 1991), thus bypassing the need for a primase (see below). (2) an extremely high processivity (>70 kb,

measured by rolling circle replication, the highest described for a DNA polymerase), allowing replication of the entire genome from a single binding (and priming) event, without the assistance of processivity factors (Blanco et al., 1989); (3) unlike most replicases,  $\Phi$ 29 DNAP efficiently couples DNA polymerization to strand displacement, without the need of helicase-like proteins (Blanco et al., 1989). These three aforementioned exceptional properties are essential to allow the symmetric DNA replication mode of bacteriophage  $\Phi$ 29 mentioned above, by which the two DNA strands are synthesized continuously from both ends of the linear molecule (Blanco et al., 1989). In the case of  $\Phi$ 29 TP-DNA amplification, the single-stranded DNA binding protein p5 and the double-stranded DNA binding protein p6 are essential.

Resolution of the  $\Phi$ 29 DNAP structure, in collaboration with Thomas Seitz's lab (Yale University), gave the insights into these three unique properties of the enzyme, use of TP as primer, processivity, and strand displacement capacity (Kamtekar et al., 2004, 2006). Thus, the  $\Phi$ 29 DNAP structure is formed by an N-terminal exonuclease domain, harboring the 3'-5' exonuclease active site, and a C-terminal polymerization domain (see **Figure 3A**) that has the universally conserved palm (containing the catalytic residues as well as DNA ligands), fingers (mainly involved in binding the incoming dNTP), and thumb (containing DNA ligands which confer stability to the primer-terminus) subdomains (Kamtekar et al., 2004). Although *a priori* this bimodular structure would be a common theme among proofreading DNA polymerases, the main structural novelty was the presence in the polymerization domain of  $\Phi$ 29 DNAP of two subdomains called TPR1 and TPR2, specifically present in the protein-primed subgroup of DNA polymerases (Blasco et al., 1990; Dufour et al., 2000). TPR1 is placed at the edge of the palm, while TPR2 contains a  $\beta$ -hairpin and forms with the apex of the thumb subdomain an arch-like structure. Palm, thumb, TPR1, and TPR2 subdomains form doughnut-shaped structure at the polymerization active site that encircles the growing DNA product (Berman et al., 2007), acting as an internal clamp that confers the DNA-binding stability responsible for the inherent processivity of the enzyme (Rodríguez et al., 2005), similar to the sliding clamps used by other replicative polymerases (see **Figure 3B**). On the other hand, TPR2, palm and fingers subdomains, together with the exonuclease domain, encircle the downstream template strand (Berman et al., 2007), forming a narrow tunnel whose dimensions ( $\sim 10$  Å) do not allow dsDNA binding. This fact forces the unwinding of the downstream dsDNA to allow threading of the template strand through this tunnel to reach the polymerization site, using the same topological mechanism as hexameric helicases to open dsDNA regions (see **Figure 3B**), providing the structural basis for the strand displacement capacity of  $\Phi$ 29 DNAP (Kamtekar et al., 2004; Rodríguez et al., 2005). The use of optical tweezers has allowed to conclude that the DNA polymerase destabilizes the two nearest base pairs of the fork by maintaining a sharp bending of the template and the complementary strands at a closed fork junction (Morin et al., 2012). Therefore, the polymerase, instead of behaving as a "passive" unwinding motor that would imply that translocation of the protein traps transient unwinding fluctuations of the fork, behaves as an



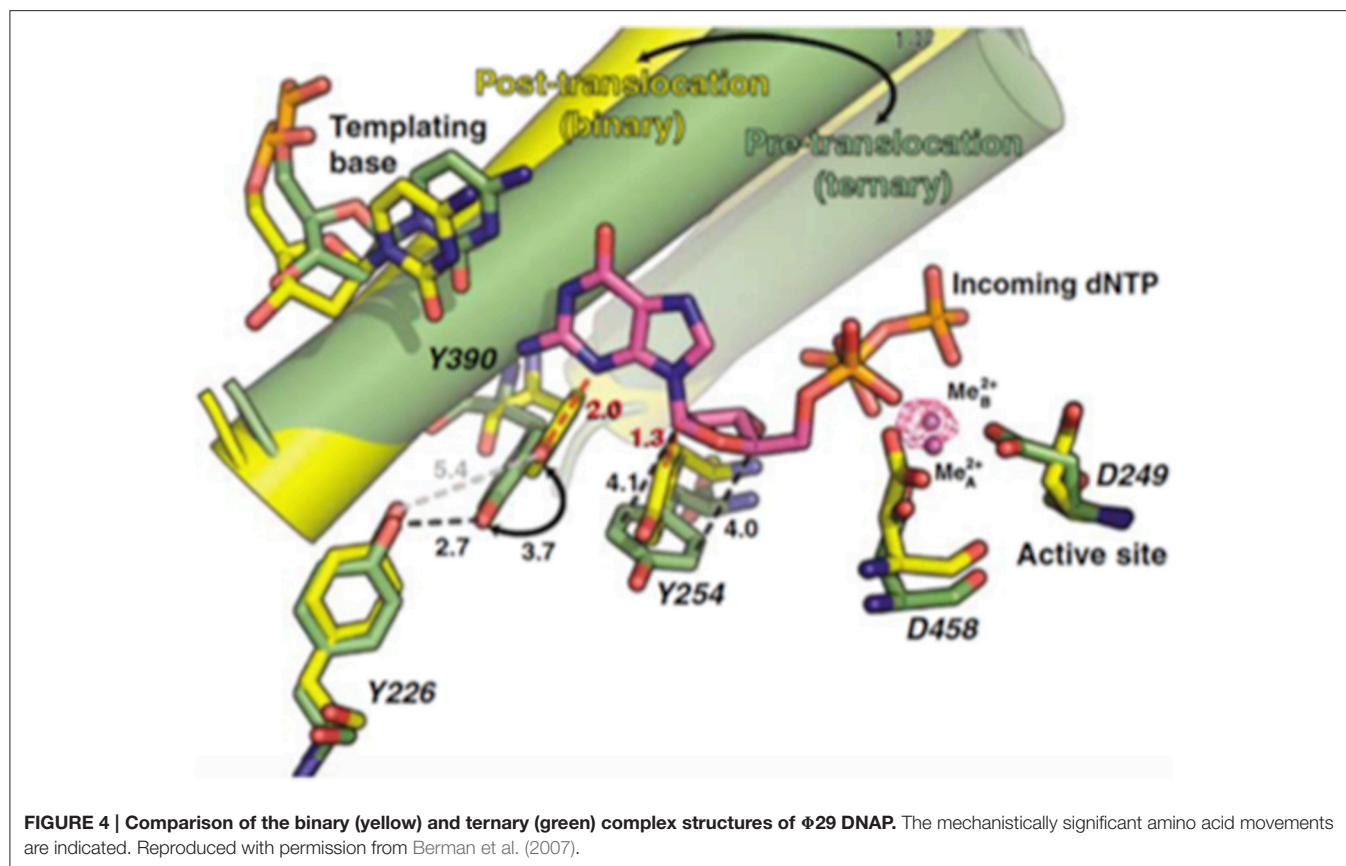
“active” motor, actively destabilizing the duplex DNA at the junction.

## On the Translocation Mechanism of $\Phi$ 29 DNA Polymerase

As any other replicative DNA polymerase, after inserting a dNMP,  $\Phi$ 29 DNAP has to translocate the growing DNA one position backwards to allow the next insertion step to occur, a process called translocation. The structures of the binary and ternary complexes of  $\Phi$ 29 DNAP provided a structural basis for comprehending the mechanism of translocation (Berman et al., 2007). The dNTP insertion site is initially occupied by the aromatic ring of two conserved residues, Tyr390 (from the fingers subdomain) and Tyr254 (from the palm subdomain; see **Figure 4**). Once the incoming nucleotide gains access and binds at the polymerization active site it triggers a  $14^\circ$  rotation of the fingers subdomain toward the polymerization active site, going from an open to a closed state and allowing electropositively charged residues from the fingers subdomain to bind the  $\alpha$ - and  $\beta$ -phosphates of the dNTP. Closing of the fingers provokes Tyr390 and Tyr254 to abandon the nucleotide insertion site to form part of the nascent base pair binding pocket, allowing the base moiety of the incoming nucleotide to form a Watson-Crick base pair with the templating nucleotide, whereas the deoxyribose ring stacks on the phenolic group of Tyr254. Once the phosphoester

bond formation between the  $\alpha$ -phosphate of the incoming dNTP and the OH-group of the priming nucleotide has taken place (pre-translocation state), the pyrophosphate produced leaves the DNA polymerase, breaking the electrostatic crosslink that kept the fingers subdomain in the closed state. Concomitantly to the fingers opening, residues Tyr254 and Tyr390 move back into the nucleotide insertion site, and the nascent base pair translocates backwards one position (post-translocation state; Berman et al., 2007). This translocation allows the 3' OH-group of the newly added nucleotide to be in a competent position to prime the following nucleotide insertion event. Direct observation of translocation in individual  $\Phi$ 29 DNAP complexes monitored with single nucleotide resolution and using the hemolysin nanopore, has allowed to conclude that  $\Phi$ 29 DNAP translocation occurs discretely from the pre-translocation state to the post-translocation state, driven by Brownian thermal motion (Dahl et al., 2012). Although nucleotide does not drive translocation, the fluctuation of the binary complexes between the pre-translocation and post-translocation states is rectified to the post-translocation state by the binding of complementary dNTP. The movement from the open, post-translocation state, to the closed pre-translocation state most probably reflects an equilibrium between the fingers-open and fingers-closed states to relieve the steric clash of the primer-terminus with residues Tyr254 and Tyr390 (see above), which occlude the nucleotide insertion site when the fingers are open (Dahl et al., 2012).





## The Degradative Reactions Catalyzed by $\Phi$ 29 DNA Polymerase: the Pyrophosphorolysis and the 3′–5′ Exonuclease Activity

In addition to the synthetic activities described above,  $\Phi$ 29 DNAP catalyzes two degradative reactions:

1. **Pyrophosphorolysis.**  $\Phi$ 29 DNAP possesses an inorganic pyrophosphate-dependent degradative activity, pyrophosphorolysis (Blasco et al., 1991). This activity, whose optimal substrate is a duplex DNA with a protruding 5′ single strand, can be considered as the reversal of polymerization as it acts in the 3′–5′ direction releasing free dNTPs by addition of PPi as substrate, in the presence of divalent metal ions, probably playing some role in fidelity (Blasco et al., 1991). The fact that  $\Phi$ 29 DNAP mutants at the catalytic amino acid residues involved in the DNA polymerization activity were also deficient in the pyrophosphorolytic activity indicated that both activities share a common polymerization active site (Blasco et al., 1991; Santos et al., 2014).
2. **3′–5′ exonuclease.** A reaction found in the N-terminal domain of the polymerase, and that requires two divalent metal ions to release dNMP units from the 3′ end of a DNA strand at a catalytic rate of  $500 \text{ s}^{-1}$  (Esteban et al., 1994). As in other replicases, the 3′–5′ exonuclease of  $\Phi$ 29 DNAP proofreads DNA insertion errors, as it degrades

preferentially mismatched primer termini (Blanco and Salas, 1985a; Garmendia et al., 1992).

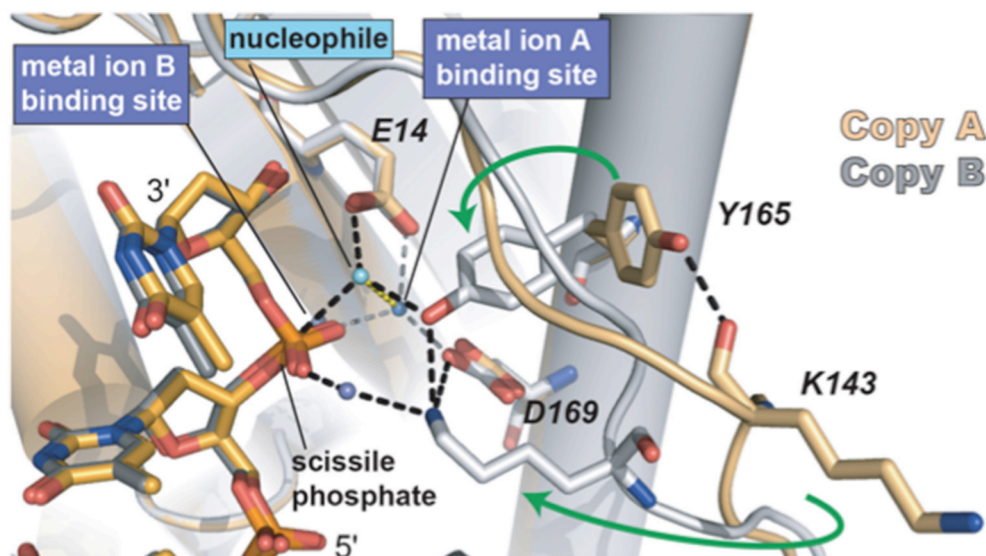
Sequence alignments and extensive site directed mutagenesis studies carried out during the last three decades in  $\Phi$ 29 DNAP have been pioneer in the identification and role of the catalytic and ssDNA ligand residues responsible for the 3′–5′ exonuclease (reviewed in Blanco and Salas, 1995, 1996). The presence of homologous residues among distantly related DNA-dependent DNA polymerases allowed us to propose the evolutionary conservation of 3′–5′ exonuclease active site (Bernad et al., 1989) in the proofreading DNA polymerases. Thus, the exonuclease active site, located at the N-terminal domain (residues 1–189; exonuclease domain, see **Figure 3**), is formed by three conserved N-terminal amino acid motifs, ExoI, ExoII, and ExoIII, that contain four carboxylate groups (Asp12, Glu14, Asp66, and Asp169 in  $\Phi$ 29 DNAP) that coordinate two metal ions, and one tyrosine residue (Tyr165 in  $\Phi$ 29 DNAP) that orients the attacking water molecule (Bernad et al., 1989). Moreover, these analyses allowed the identification of a new motif (Kx<sub>2</sub>hxA), specifically conserved in family B DNA polymerases and whose lysine residue ( $\Phi$ 29 DNAP Lys143) plays an auxiliary role in catalysis (de Vega et al., 1997), stabilizing the catalytic Asp169 of the Exo III motif (Berman et al., 2007). Crystallographic resolution of  $\Phi$ 29 DNAP with a ssDNA at the exonuclease active site demonstrated the existence of two stable conformations at the exonuclease active site of family B DNA polymerases



(see **Figure 5**), as previously suggested from comparisons of T4 and RB69 DNA polymerase exonuclease structures with the *E. coli* DNA polymerase I Klenow fragment exonuclease structure (Beese and Steitz, 1991; Wang et al., 1996, 2004). In one conformation, the tyrosine from the Exo III motif (Tyr165 in  $\Phi$ 29 DNAP) is solvent exposed, whereas in the other conformation, it contacts with the scissile phosphate through the nucleophile while conserved lysine from motif  $Kx_2hx_A$  ( $\Phi$ 29 DNAP Lys143) stabilizes the catalytic aspartate of the Exo III motif ( $\Phi$ 29 DNAP Asp169), consistent with the previous biochemical results (de Vega et al., 1997). The latter conformation seems to be the more chemically and biologically relevant complex for exonuclease activity. The two conformations observed suggest that the movement of the conserved tyrosine and lysine residues into the active site sets up the active site for the exonucleolysis reaction in the family B DNA polymerases.

A tight and fine-tuned coordination between the polymerization and exonucleolytic cycles should take place to allow a productive and faithful replication. Previous studies showed that  $\Phi$ 29 DNAP proofreads the misinserted nucleotides intramolecularly (de Vega et al., 1999). This fact implies that the DNA polymerase transfers the mismatched 3'-terminus to the 3'-5' exonuclease active site for erroneous dNMP release without dissociating from the DNA. Comparison of DNA polymerase structures of RB69 DNA polymerase in polymerization and editing modes showed that the primer-terminus switches between both active sites by the rotation of a top microdomain of the thumb subdomain (Shamoo and Steitz, 1999; Franklin et al., 2001). However, in the 3D resolution of  $\Phi$ 29 DNAP

structure the thumb subdomain has an unusual structure mainly constituted by a static long  $\beta$ -hairpin that does not rotate upon DNA binding (Kamtekar et al., 2004; Berman et al., 2007, see **Figure 3**). In addition, the blockage of the thumb movements by introducing a disulfide bond between the tips of the TPR2 and thumb subdomains had not effect in the partitioning of the primer-terminus between the polymerization and editing active sites (Rodríguez et al., 2009), a result that led us to suggest that in  $\Phi$ 29 DNAP the primer-terminus switches between both active sites by a passive diffusion mechanism. In this sense, the recent use of single-molecule manipulation method has made possible the study of the dynamics of the partitioning mechanism by applying different tension to a processive single  $\Phi$ 29 DNAP-DNA complex (Ibarra et al., 2009). Thus, the application of mechanical force to the template causes the gradual intramolecular switch of the primer between the active sites of the protein by decreasing the affinity of the polymerization active site for the template strand with the further disruption of the dsDNA primer-template structure that provokes a fraying of 4–5 bp of dsDNA, allowing primer-terminus to reach the exonuclease active site intramolecularly (Ibarra et al., 2009), supporting the passive diffusion mechanism. The energetically unfavorable gradual melting of 4–5 bp of dsDNA should be progressively balanced by the establishment of new and specific interactions with DNA ligands of the thumb subdomain (Pérez-Arnaiz et al., 2006). Such contacts would also guide the primer-terminus to interact with ssDNA ligands of the exonuclease domain that stabilize the primer-terminus at the exonuclease site (de Vega et al., 1996, 1998b; Kamtekar et al., 2004; Pérez-Arnaiz et al., 2006; Rodríguez et al., 2009).



**FIGURE 5 | The two conformations of the exonuclease active site of  $\Phi$ 29 DNAP.** The Lys143, Tyr165, and two of the catalytic aspartates are shown in stick representation. Green arrows indicate the movement of Tyr165 and Lys143 from the open conformation to the closed conformation. The black dashed lines represent the observed hydrogen bonds between Lys143 and Tyr165 with each other and with other parts of the active site. The interactions between the waters in the metal binding sites and the protein are represented as gray dashes. Most of the interactions that the water in the metal in **B** site would be making with the protein are missing due to the D12A/D66A mutations in the polymerase used in these crystallographic studies. Reproduced with permission from Berman et al. (2007).

Recent development of a single-molecule approach using a nanoscale pore has allowed to conclude that transfer of the primer strand from the polymerase to the exonuclease site takes place before translocation, the pre-translocation state being therefore the branchpoint between the DNA synthesis and editing pathways (Dahl et al., 2014). Once the 3' terminal nucleotide is released, the primer-terminus goes back to the polymerase site and pairs with the template strand in the post-translocation state being poised to bind the incoming dNTP and resume DNA synthesis (Dahl et al., 2014).

## Biotechnological Applications of $\Phi$ 29 DNA Polymerase

The two distinctive features of  $\Phi$ 29 DNAP, high processivity, and strand displacement capacity, together with a remarkably faithful replication, contributed by a high nucleotide insertion fidelity, and an intrinsic proofreading activity, led to the development of isothermal multiple displacement amplification (MDA) currently exploited (Dean et al., 2001, 2002). These amplification methods based on  $\Phi$ 29 DNAP show two main advantages respect to classical PCR DNA amplification: first, the use of random hexamer primers eliminates the previous sequence information requirement allowing the amplification of any DNA molecule, and second, the products of the amplification reaction can be much larger than those obtained by classical PCR. In addition, the capacity displayed by  $\Phi$ 29 DNAP to use circular multiply primed ssDNA templates gave rise to the development of the multiply primed rolling circle amplification of circular DNAs of variable size (Dean et al., 2001). This technology has been successfully exploited to amplify and detect circular viral genomes (Johne et al., 2009), to genotype single nucleotide polymorphisms (Qi et al., 2001), to analyze the genome of non-cultivable viruses (Johne et al., 2009), to detect and identify circular plasmids in zoonotic pathogens (Xu et al., 2008), and to describe new metagenomes (López-Bueno et al., 2009). Recently, we have been able to improve isothermal MDA by making new variants of  $\Phi$ 29 DNAP (de Vega et al., 2010). Thus, we have fused DNA binding domains (Helix-hairpin-Helix) to the C-terminus of the polymerase increasing the DNA binding ability of the enzyme without compromising its replication rate. As a result, the new variants display an improved DNA amplification efficiency on both circular plasmids and genomic DNA and are the only  $\Phi$ 29 DNAP variants with enhanced amplification performance so far.

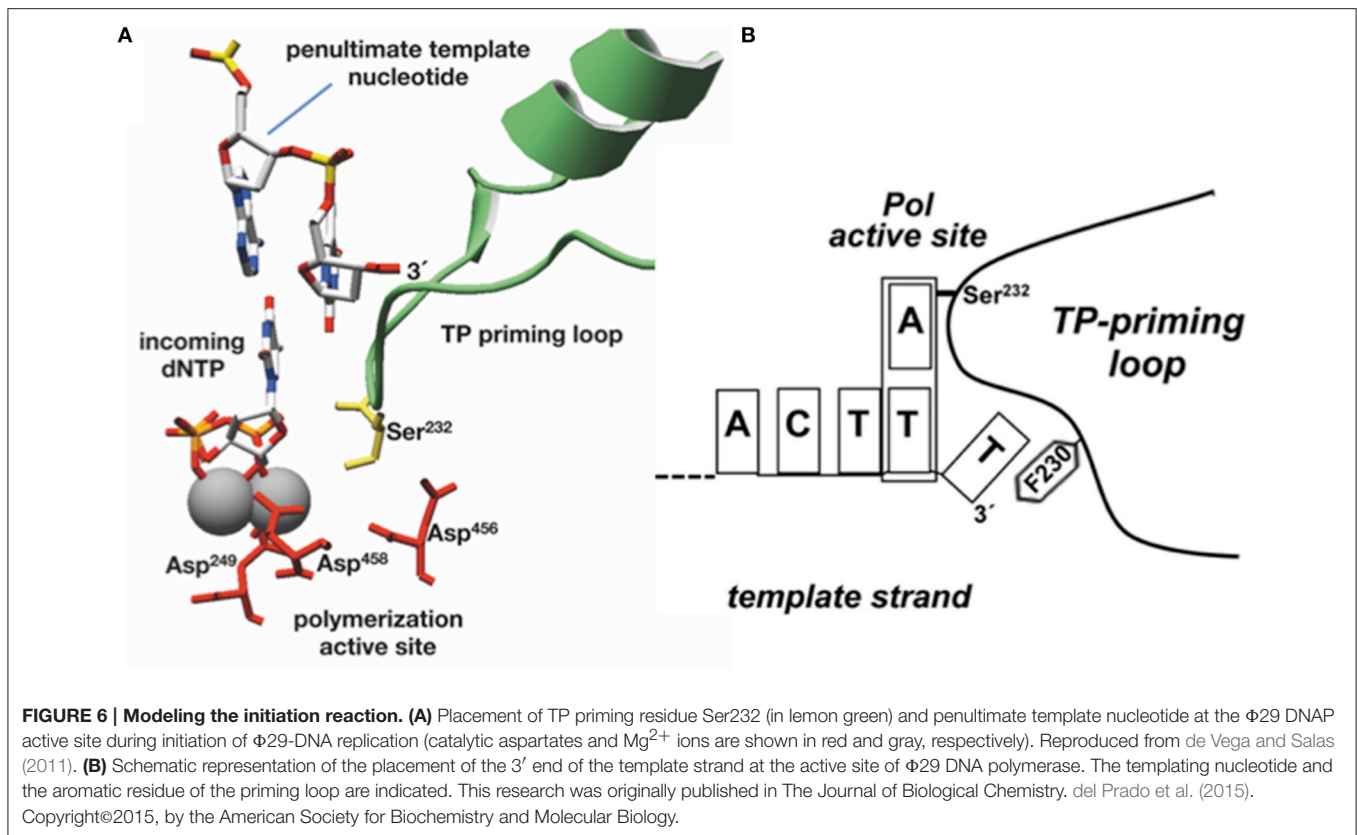
## INITIATION OPPOSITE AN INTERNAL TEMPLATING NUCLEOTIDE: A SMART SOLUTION TO PRESERVE THE FIDELITY DURING INITIATION

The  $\Phi$ 29 TP/DNAP heterodimer recognizes the replication origins at the genome ends (see **Figure 1**). Such origins are constituted by specific sequences as well as by the parental TP, the major signal in the template for recognition, a fact that suggests that the heterodimer is recruited to the origin through interactions with the parental TP. The use of heterologous systems in which DNA polymerase, primer TP, and TP-DNA

were from the  $\Phi$ 29 and Nf related phages allowed us to infer specific contacts between the DNA polymerase and the parental TP, as the initiation only occurred when the polymerase and the TP-DNA were from the same phage (González-Huici et al., 2000b; Pérez-Arnáiz et al., 2007). In addition, the presence of mutations in the intermediate domain of both the parental and primer TPs precluded DNA replication, suggesting also a role for the primer TP in the specific recognition of the replication origins (Illana et al., 1998; Serna-Rico et al., 2000; del Prado et al., 2012).

As already indicated, the DNA ends of  $\Phi$ 29 have a repetition of three nucleotides (3'-TTT...5'). Once the replication origins are specifically recognized by the TP/DNA polymerase heterodimer (Blanco et al., 1987; Freire et al., 1996; González-Huici et al., 2000a,b; Pérez-Arnáiz et al., 2007), the DNA polymerase catalyzes the formation of a phosphoester bond between the initial dAMP and the hydroxyl group of Ser232 of the TP. Therefore, during the initiation reaction, the priming Ser232 of the TP is placed at the catalytic site of the DNA polymerase to attack nucleophilically the  $\alpha$ -phosphate of the initial dAMP which is inserted opposite the 3' second nucleotide of the template strand (Méndez et al., 1992, see **Figure 6A**). This reaction is carried out by the catalytic residues responsible for canonical polymerization (Blanco and Salas, 1995, 1996). The initiation reaction implies that the 3' end of the template strand should enter deep into the catalytic site of the DNA polymerase to place the penultimate 3' dTMP of the template strand at the catalytic site (see **Figures 6A,B**). The interchanging of the priming domains of the related  $\Phi$ 29 and Nf TPs, allowed us to conclude that this domain is the one responsible for dictating the internal 3' nucleotide used as template during initiation, the 2nd and 3rd in  $\Phi$ 29 and Nf DNA, respectively (Longás et al., 2008). Recently, we have shown that the aromatic residue Phe230 of the  $\Phi$ 29 TP priming loop is the one responsible for positioning the penultimate nucleotide at the polymerization site to direct insertion of the initial dAMP during the initiation reaction, most probably by interacting with the 3' terminal base, limiting the internalization of the template strand (see del Prado et al., 2015; **Figure 6B**). To perform TP-DNA full-length synthesis, the TP-dAMP initiation product translocates backwards one position to recover the template information corresponding to the first 3'-T, the so-called sliding-back mechanism that requires a terminal repetition of 2 bp. This reiteration permits, prior to DNA elongation, the asymmetric translocation of the initiation product, TP-dAMP, to be paired with the first T residue (Méndez et al., 1992) (see scheme in **Figure 7**).

We have shown that the sliding-back, or variations of it, is a mechanism shared by the protein-priming systems to restore full-length DNA. In the case of the  $\Phi$ 29-related phage GA-1, initiation takes place at the 3' second nucleotide of the template (3'-TTT) (Illana et al., 1996). The  $\Phi$ 29-related phage Nf and the *S. pneumoniae* phage Cp-1 initiate opposite the 3' third nucleotide of their terminal repetition (3'-TTT) (Martín et al., 1996; Longás et al., 2008), whereas the *E. coli* phage PRD1 initiates at the fourth nucleotide (3'-CCCC) (Caldentey et al., 1993), being required two and three consecutive sliding-back steps, respectively, to recover the information of the DNA termini

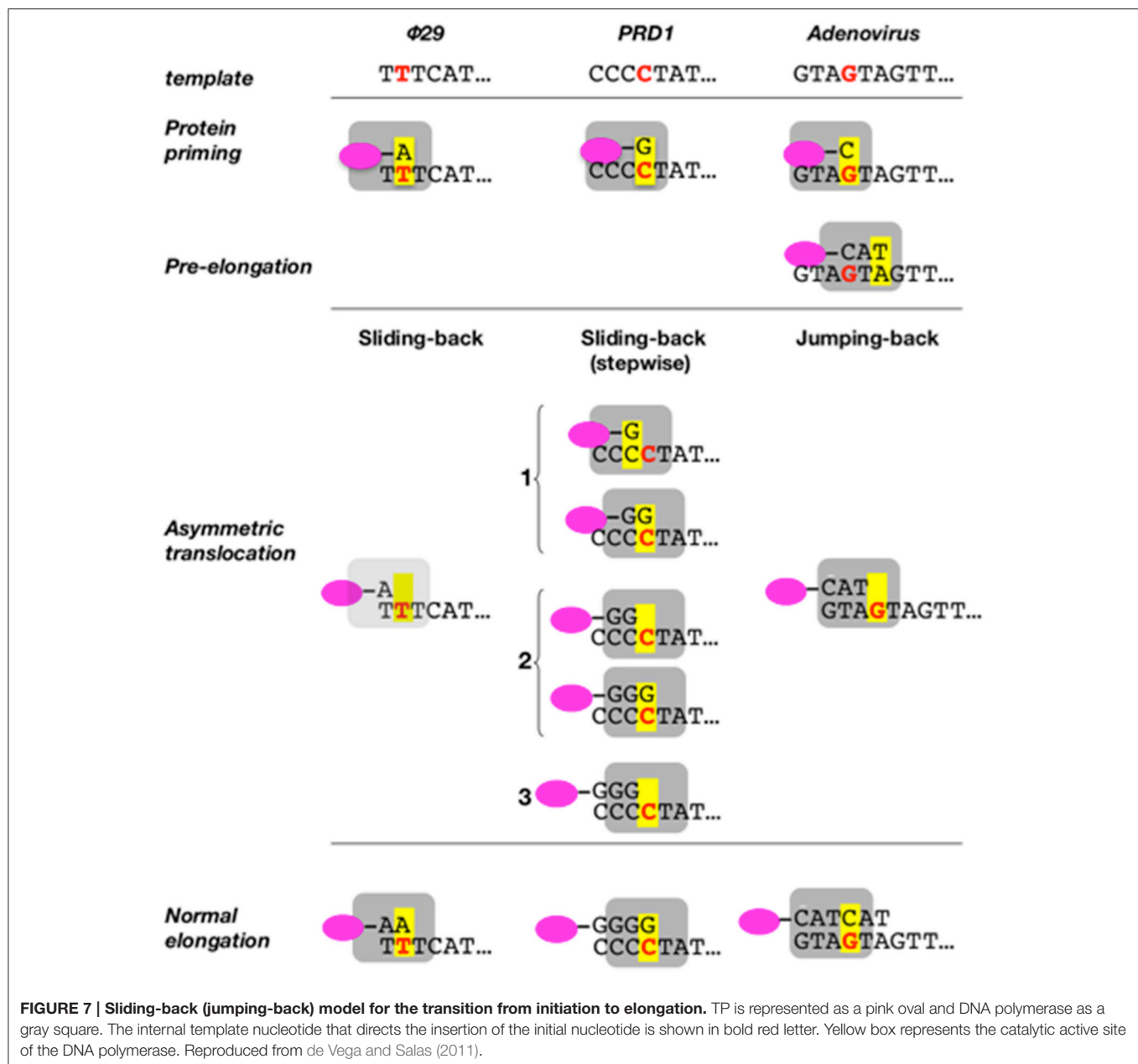


(stepwise sliding-back). The case of adenovirus is a little more complex as its genome ends have a duplication of the sequence GTA (3'-GTAGTA). In this virus, the 3' fourth to sixth template positions guides the formation of the TP-CAT initiation product that jumps back to pair with the terminal GTA, a variation of the sliding-back mechanism called jumping-back (King and van der Vliet, 1994) (see scheme in Figure 7).

What is the rationale of the sliding-back mechanism?  $\Phi$ 29 protein-primed initiation is an unfaithful reaction with a nucleotide insertion discrimination factor about  $10^2$ . In addition, the 3'-5' exonuclease activity of  $\Phi$ 29 DNAP cannot release a wrong nucleotide that had been added during the initiation reaction (Esteban et al., 1993). Therefore, the sliding-back mechanism could guarantee the fidelity during the initiation stage through different base pairing checking steps before further elongation of the TP-dNMP complex occurs (Méndez et al., 1992; King and van der Vliet, 1994). Thus, an erroneous TP-dNMP complex will not pair with the terminal 3'-T of the template after the sliding-back, hindering its further elongation. In addition, if an incorrect TP-dNMP product were elongated the resulting TP-DNA molecule could not be used as a template in the next replication round, as the 3' terminus of the template strand would not include the required nucleotide reiteration. The presence of sequence repetitions at the ends of other TP-containing genomes allows to surmise that the sliding-back type of mechanism could be a common feature of protein-primed replication systems (Méndez et al., 1992).

## TRANSITION FROM PROTEIN-PRIMED TO DNA-PRIMED REPLICATION

Previous biochemical studies showed that once the initiation reaction has taken place the polymerase incorporates the next 4 nucleotides to the TP-dAMP product while is still complexed with the primer TP (initiation mode), goes through some structural changes during insertion of the sixth to ninth nucleotide (transition mode) and finally dissociates from the primer TP once the tenth nucleotide is added to the growing strand (elongation mode) (Méndez et al., 1997). Resolution of the  $\Phi$ 29 DNAP/TP complex has given the insights on the transition mechanism, explaining how the polymerase can insert up to nine nucleotides while complexed to the TP (Kamtekar et al., 2006). The transition stage relies on a different strength interaction of the TP priming and intermediate domains with the DNA polymerase (Pérez-Arnáiz et al., 2007). On the one hand, the TP intermediate domain remains in a fixed orientation on the polymerase during insertion of 6–7 nucleotides by means of stable contacts with the TPR1 subdomain. During this stage the weakness of the interaction between the DNA polymerase and the TP priming domain allows the latter to rotate as the DNA is synthesized. The rotation of the TP priming domain with respect to the fixed TP intermediate domain is possible due to the flexibility of the hinge region that connects both domains. Once 6–7 nucleotides have been added, the proximity of the priming Ser to the hinge region would impede a further priming domain



rotation, causing heterodimer dissociation (Kamtekar et al., 2006; see Figure 8).

## $\Phi$ 29 PROTEIN P5, THE VIRAL SINGLE-STRANDED DNA BINDING PROTEIN

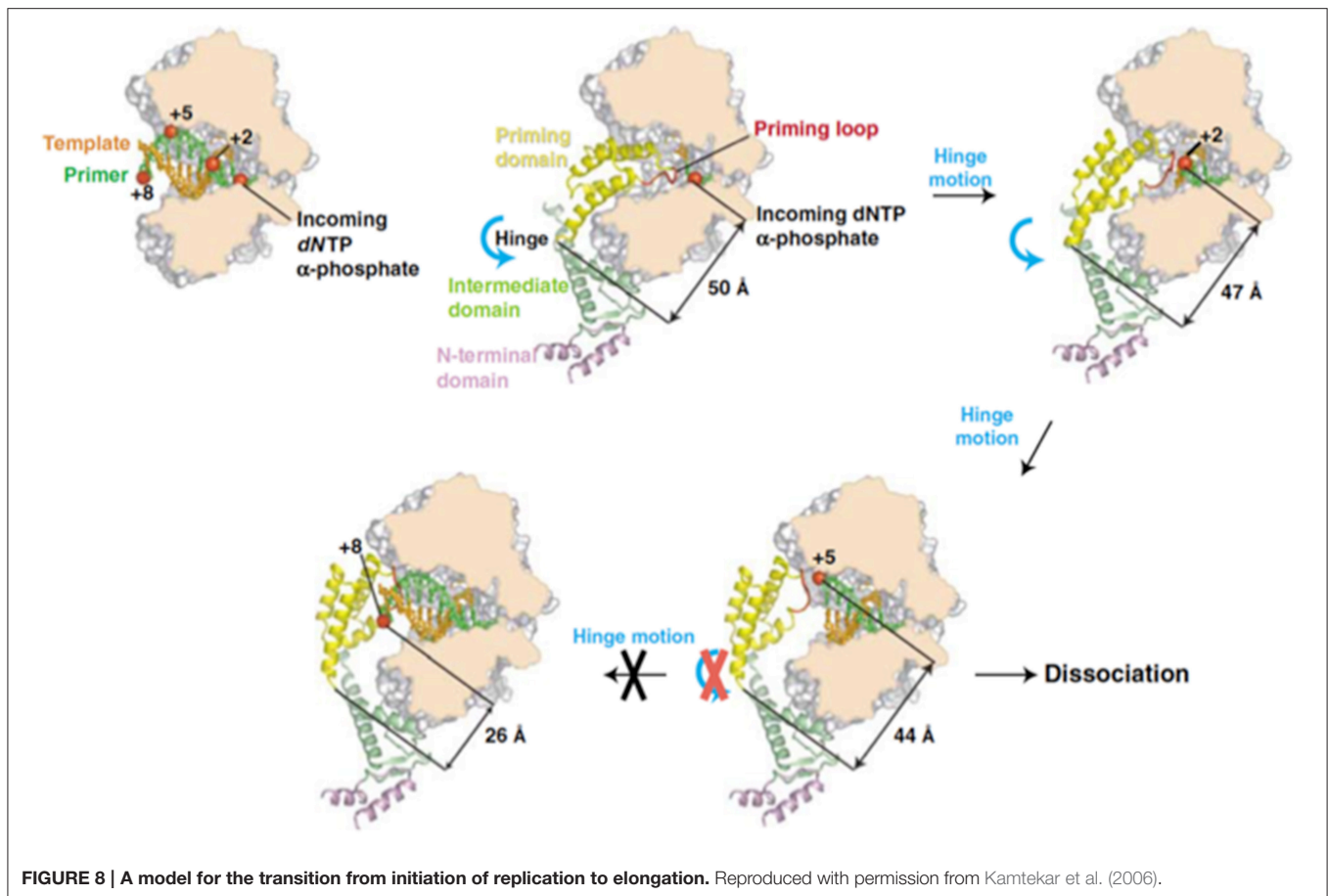
### Structural and Functional Characteristics

Single-stranded DNA-binding proteins (SSBs) are common in all three branches of organisms and in viruses and bind with high affinity to single-stranded (ss) DNA, playing essential roles as accessory proteins in DNA replication, recombination, and repair processes that entail the exposure of ssDNA. SSBs usually

bind non-specifically to DNA and can saturate long stretches of ssDNA, thus providing protection against nuclease attack, and preventing the formation of secondary structures (Chase and Williams, 1986; Kur et al., 2005). Furthermore, SSB proteins are involved in specific interactions with several proteins that play important roles in nucleic acids metabolism (Shereda et al., 2008). As a result of these properties, SSBs increase the efficiency and fidelity of a number of DNA amplification methods (Rapley, 1994; Perales et al., 2003; Inoue et al., 2006; Mikawa et al., 2009; Ducani et al., 2014).

From a structural viewpoint, SSBs exist as monomeric or multimeric proteins and, with few exceptions, they share a structural domain named OB-fold (oligonucleotide/oligosaccharide binding-fold) involved in nucleic acid





**FIGURE 8 | A model for the transition from initiation of replication to elongation.** Reproduced with permission from Kamtekar et al. (2006).

recognition (Theobald et al., 2003; Savvides et al., 2004). The OB-fold structural domain consists in a close or semi-open beta barrel made out of five-stranded  $\beta$ -strands and a  $\alpha$ -helix, commonly between the third and four strands (Murzin, 1993).

$\Phi$ 29 protein p5 is a single-stranded DNA binding protein (Martin et al., 1989) that protects DNA from nucleases (Martin et al., 1989) and prevents unproductive binding of  $\Phi$ 29 DNAP to ssDNA generated during replication (Gutiérrez et al., 1991).  $\Phi$ 29 SSB has high sequence similarity with SSBs from the related podoviruses Nf and GA-1, although  $\Phi$ 29 and Nf are monomeric in solution, whereas GA-1 SSB is hexameric (Soengas et al., 1995; Gascón et al., 2000a), by means of a N-terminal additional motif (Gascón et al., 2002). Podoviral SSBs share some key hydrophobic residues with unrelated viral SSBs (Gutiérrez et al., 1991) and, indeed, they may also share the SSBs common OB-fold, as found by secondary structure prediction and multiple sequence alignment (Figure 9). In agreement with this predicted protein folding, previous circular dichroism spectra indicated that  $\Phi$ 29 SSB is largely made up of  $\beta$ -strands (Soengas et al., 1997a).

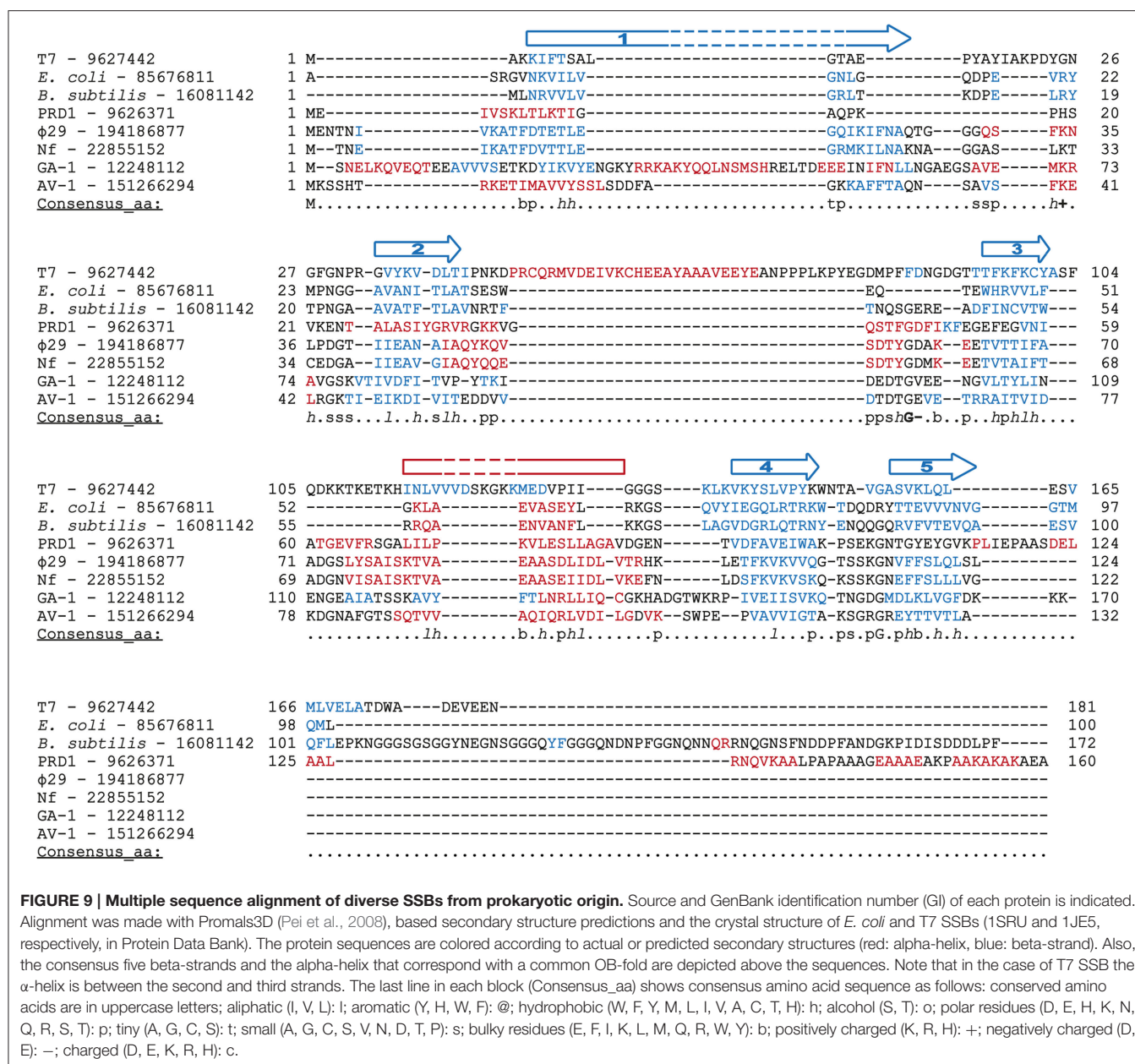
The interaction of  $\Phi$ 29 SSB with ssDNA is consistent with a moderate cooperative binding to 3–4 nt per molecule, not impaired by ionic conditions (Soengas et al., 1994). Detailed analysis of intrinsic tyrosine fluorescence quenching upon binding to ssDNA and site-directed mutagenesis indicated that

Tyr50, Tyr57, and Tyr76 play essential role in complex formation with DNA (Soengas, 1996; Soengas et al., 1997b).

As other SSBs,  $\Phi$ 29, Nf, and GA-1 SSBs, are able to unwind duplex DNA (Soengas et al., 1995; Gascón et al., 2000b), suggesting that they can favor DNA replication by unwinding the secondary structures formed in the ssDNA produced during genome replication. However, although all the three SSBs increase DNA replication efficiency (Martin et al., 1989), only  $\Phi$ 29 SSB enhances the replication rate of the DNA polymerase, especially when strand displacement is impaired, although it does not seem to have specificity for its cognate DNA polymerase (Soengas et al., 1995; Gascón et al., 2000b). Therefore, rather than the interaction of the SSB and its own DNA polymerase, improvement of the replication rate by  $\Phi$ 29 SSB is mediated by its dynamic dissociation from the nucleoprotein complexes ahead the polymerase, in agreement with its relative low intrinsic binding constant (Soengas et al., 1994; Gascón et al., 2000a).

## Biological Role

SSBs are required in stoichiometric quantities with respect to the template rather than in catalytic amounts. Accordingly,  $\Phi$ 29 SSB is required in high amounts for *in vitro*  $\Phi$ 29 genome amplification (Blanco et al., 1994) and it is an extremely abundant protein in the infected *B. subtilis* cells (~700,000 molecules per cell, Martin et al., 1989). Early



**FIGURE 9 | Multiple sequence alignment of diverse SSBs from prokaryotic origin.** Source and GenBank identification number (GI) of each protein is indicated. Alignment was made with Promals3D (Pei et al., 2008), based secondary structure predictions and the crystal structure of *E. coli* and T7 SSBs (1SRU and 1JE5, respectively, in Protein Data Bank). The protein sequences are colored according to actual or predicted secondary structures (red: alpha-helix, blue: beta-strand). Also, the consensus five beta-strands and the alpha-helix that correspond with a common OB-fold are depicted above the sequences. Note that in the case of T7 SSB the  $\alpha$ -helix is between the second and third strands. The last line in each block (Consensus\_aa) shows consensus amino acid sequence as follows: conserved amino acids are in uppercase letters; aliphatic (I, V, L): I; aromatic (Y, H, W, F): @; hydrophobic (W, F, Y, M, L, I, V, A, C, T, H): h; alcohol (S, T): o; polar residues (D, E, H, K, N, Q, R, S, T): p; tiny (A, G, C, S): t; small (A, G, C, S, V, N, D, T, P): s; bulky residues (E, F, I, K, L, M, Q, R, W, Y): b; positively charged (K, R, H): +; negatively charged (D, E): -; charged (D, E, K, R, H): c.

genetic characterization of  $\Phi$ 29 mutants allowed the mapping of temperature-sensitive mutants in gene 5 (Mellado et al., 1976). Those mutants had a strong impairment in DNA synthesis (Talavera et al., 1972), indicating an essential role in replication of the viral genome that, as mentioned above, was subsequently demonstrated thanks to the *in vitro* characterization of TP-DNA replication.

Strikingly, recent isolation of a non-sense mutant in gene 5 that only had a 20% reduction in viral yield (Tone et al., 2012), suggested that  $\Phi$ 29 SSB might be dispensable for viral replication, although it seems to be required in a temperature dependent fashion. These results led the authors to speculate that a host SSB could be able to partially complement the absence of viral SSB at permissive temperatures. However, the molecular mechanism

of this possible temperature-dependent role of  $\Phi$ 29 SB remains unclear.

## A HISTONE-LIKE PROTEIN ENCODED BY BACTERIOPHAGE $\Phi$ 29

### Structural Characteristics and DNA Binding Mechanism

The viral protein p6 is a DNA binding protein (DBP) involved both in DNA replication, activating the protein-primed initiation step, and transcriptional control, modulating the early-late switch (for a detailed review see González-Huici et al., 2004a). This pleiotropic effect is consequence

of its role as an architectural protein that organizes and compacts the viral genome, analogously to eukaryotic histones (Serrano et al., 1994).

Structure-function work on p6 indicated that the N-terminal region of the protein plays a role in both DNA binding (Otero et al., 1990) and dimer formation (Abril et al., 2000). By site-directed mutagenesis, it could be disclosed that residues Ile8 and Val44 are directly involved in protein dimer formation (Abril et al., 2000, 2002), whereas Lys2, Lys10 and, especially, Arg6, are essential for DNA binding *in vitro* and viral DNA synthesis *in vivo* (Bravo et al., 1994b; Freire et al., 1994).

According to footprinting assays (Prieto et al., 1988; Serrano et al., 1990), p6 binding to DNA gives rise to a nucleoprotein complex formed by a repeated motif of p6 dimers bound to a 24 bp DNA segment. Thus, a protein monomer would contact and bend the DNA every 12 bp, suggesting a model in which the DNA would wrap around a multimeric core of protein p6, forming a right-handed superhelix that comprises around 63 bp per turn (Serrano et al., 1990, 1993a,b). As a consequence of DNA wrapping, this nucleoprotein complexes show a remarkable reduction in length with respect to naked DNA, between 4.2- to 6.5-fold (Serrano et al., 1993a; Gutiérrez et al., 1994).

*In vivo*, p6 is able to discriminate between bacterial and viral DNA by their different superhelicity (González-Huici et al., 2004b). Thus, p6 is able to restrain positive supercoiling of the DNA *in vitro* (Prieto et al., 1988; Serrano et al., 1993b) and binds all along  $\Phi$ 29 DNA *in vivo* with a much higher affinity than for plasmid DNA, although binding to plasmid DNA is enhanced by decreasing the negative supercoiling (González-Huici et al., 2004c). Thus, the presumably lower negative superhelicity of  $\Phi$ 29 DNA respect to host chromosome likely makes the viral genome an appropriate target for the binding of p6 (Serrano et al., 1994; González-Huici et al., 2004b). Interestingly, the preferential binding of  $\Phi$ 29 p6 to the lower negatively supercoiled viral genome seems to be quite specific, since GA-1 p6, which has a highly conserved sequence (58% similarity, 39% identity), does not show this binding pattern (Freire et al., 1996) and accordingly, GA-1 p6 complex with  $\Phi$ 29 DNA is not functional (Alcorlo et al., 2007).

Moreover, p6 has a binding specificity to the ends of the  $\Phi$ 29 linear genome, which has a key role in the initiation step of replication (see below). This binding occurs at recognition regions that were mapped between positions 62–125 from the right end, and between positions 46–68 from the left one (Serrano et al., 1989). However, p6 does not recognize a sequence signal, but rather a sequence-dependent bendability pattern present in the recognition sites that act as a nucleation site for protein p6/DNA complex formation (Serrano et al., 1993a; González-Huici et al., 2004b).

## Functional Implications of p6 Nucleoprotein Complex

Protein p6 is essential for  $\Phi$ 29 DNA replication *in vivo* (Carrascosa et al., 1976; Bravo et al., 1994b). *In vitro*, p6 stimulates initiation as well as the transition to elongation (Pastrana et al., 1985; Blanco et al., 1986, 1988). Initiation

activation requires the formation of the protein p6 complex with  $\Phi$ 29 DNA terminal fragments (Serrano et al., 1989) and it was suggested to undergo through transient unwinding of DNA at the p6 specific binding sites that would favor interaction of the TP/DNAP complex with the template strand (Serrano et al., 1993b). In line with this hypothesis,  $\Phi$ 29 origins with partially unpaired ends showed increased utilization (up to 30-fold respect to wild type origins) (Gella et al., 2014). Initiation of these modified origins was still stimulated by p6, although to a lesser extent (around 1.5-fold) than the wild type origin (2.8-fold).

As mentioned above, p6 is also important for the control of transcription, either by itself or together with the viral transcriptional regulator p4 (Camacho and Salas, 2000, 2001). Thus, protein p6 switches off very early transcription from promoter C2, as shown by *in vivo* and *in vitro* studies, impairing the RNA polymerase complex access to the nucleoprotein complex at the promoter region (Serrano et al., 1989; Camacho and Salas, 2001). Moreover, formation of p6 nucleoprotein complex promotes p4-mediated repression of promoters A2b and A2c and activation of the A3 promoter (Calles et al., 2002).

In the context of the infected cell, p6 is highly abundant, which would favor oligomerization and formation of p6 nucleocomplex (Abril et al., 1997), which might be even more favored under the crowded intracellular environment (Alcorlo et al., 2009). This high density is in agreement with a histone-like function that would complex with the entire genome (Serrano et al., 1994; Holguera et al., 2012), analogously to cellular histones. At early infection stages, p6 localizes mainly in a peripheral helix-like configuration (Holguera et al., 2012), whereas the viral genome and the replication machinery associates with the host nucleoid (Muñoz-Espín et al., 2010). Since protein p6 is essential to initiate  $\Phi$ 29 DNA replication, it was suggested that a small amount of protein p6 (undetectable by immunofluorescence) would be recruited early at the bacterial nucleoid, establishing the appropriate conditions at the phage DNA ends to achieve the first rounds of replication. Then, and as viral DNA replication progresses, p6 is recruited to the bacterial nucleoid and, by topological recognition of the  $\Phi$ 29 DNA, avoids its sequestration by the higher volume of the bacterial DNA. Under this scenario,  $\Phi$ 29 p6 constitutes a histone-like protein specific for the viral genome, whose temporal and spatial subcellular localization is determined by its essential roles in genome replication and transcription (Holguera et al., 2012).

## ROLE OF HOST DNA-BINDING PROTEINS IN $\Phi$ 29 DNA REPLICATION

Bacteriophages have developed different strategies to inactivate or take advantage of cellular enzymes in their own benefit (Roucourt and Lavigne, 2009). During  $\Phi$ 29 infection several *B. subtilis* DNA-binding proteins have been shown to play a role in the development of the infective cycle.

### DNA Gyrase

Chromosomal DNA topology is controlled by various host-encoded topoisomerases, such as DNA gyrase (topoisomerase



II) (Drlica, 1992). Despite containing a TP covalently linked to the 5' ends and therefore not being covalently closed,  $\Phi$ 29 DNA is topologically constrained *in vivo* (González-Huici et al., 2004c). In this sense, it has been shown that the gyrase inhibitor novobiocin but not nalidixic acid, which also inhibits DNA gyrase but does not have topological effects on DNA, increases the binding of protein p6 to the viral genome *in vivo*. In addition, both novobiocin and nalidixic acid impair viral DNA replication *in vivo*, suggesting that *B. subtilis* gyrase is involved in viral DNA replication (González-Huici et al., 2004c). A topologically constrained DNA should be allowed to rotate freely during the replication process, explaining the necessity of the bacterial DNA gyrase (Muñoz-Espín et al., 2012). Moreover,  $\Phi$ 29 genome possesses two convergently oriented transcription units encompassing genes 7–16 and 17–16.5. Hence, without the action of DNA gyrase, a highly positive supercoiled region would be generated between the two convergently oriented transcription units, preventing the advance of the RNA polymerase, and/or inducing DNA polymerase template switching when encountering this blockage. In fact, during  $\Phi$ 29 infection subgenomic viral DNA molecules ranging from 1 to 8 Kb are accumulated, originated mainly from the right end of the genome and that these kind of molecules do not accumulate when *B. subtilis* cells are infected with the transcription deficient mutant  $\Phi$ 29 *sus4*(56), which does not express protein p4 (Murthy et al., 1998).

The topological constraint of bacteriophage  $\Phi$ 29 genome could be achieved by binding of the parental TPs either directly to the nucleoid or to the bacterial membrane through the interaction with other viral proteins such as p1 or p16.7 (see above) (Bravo et al., 2000; Serna-Rico et al., 2003; Muñoz-Espín et al., 2010).

## Uracil-DNA Glycosylase

A potential threat to genome integrity is the presence of uracil residues in DNA. Uracil is eliminated from DNA genomes by the base excision repair pathway (BER), which is initiated with the enzymatic activity of a uracil-DNA glycosylase (UDG). These enzymes (Family 1) selectively remove uracil bases from both single- and double-stranded DNA, cleaving the N-glycosidic bond between the base and the deoxyribose, leaving therefore an abasic site (Savva et al., 1995). This abasic site is then repaired through the sequential action of an apurinic/apyrimidinic endonuclease, DNA polymerase and DNA ligase. Most eukaryotic and prokaryotic cells encode a UDG to maintain the integrity of DNA genomes. However, there are some cases in which the presence of uracil in DNA could be desirable. For instance, *B. subtilis* phage PBS1 (and its clear-plaque isotype PBS2) genome contains uracil instead of thymine and, consequently, encodes a UDG inhibitor (called Ugi) to assure an efficient viral genome replication (Katz et al., 1976; Cone et al., 1980; Savva and Pearl, 1995). Additionally, phage T5 infection induces an inhibitor of *E. coli* UDG that

has not yet been identified (Warner et al., 1980). Interestingly, despite having a non-uracil containing genome, phage  $\Phi$ 29 encodes a UDG inhibitor, a small acidic protein of 56 amino acids called p56 (Serrano-Heras et al., 2006). Protein p56 is expressed early after infection and interacts with *B. subtilis* UDG, inhibiting its activity (Serrano-Heras et al., 2006). *In vitro* experiments showed that protein p56 blocks the DNA-binding ability of UDG, and structural data suggest that it does it by mimicking the structure of DNA (Serrano-Heras et al., 2007; Asensio et al., 2011; Baños-Sanz et al., 2013; Cole et al., 2013). As mentioned above, the mechanism of  $\Phi$ 29 DNA replication involves the generation of replicative intermediates that contain large stretches of ssDNA (Harding and Ito, 1980; Inciarte et al., 1980; see Figure 1). If uracil residues were present in these stretches of ssDNA by either the misincorporation of deoxyuridine monophosphate (dUMP) during the replication process or by the spontaneous deamination of cytosine in DNA, the elimination of these lesions by the BER pathway would give rise to the loss of terminal viral DNA regions. In fact, it has been shown that  $\Phi$ 29 DNA polymerase can incorporate dUMP during DNA synthesis with a catalytic efficiency of only 2-fold lower than dTMP, and perform the extension of base-paired uracil residues to give full-length DNA *in vitro* (Serrano-Heras et al., 2008). Hence, by encoding an UDG inhibitor,  $\Phi$ 29 prevents the elimination of uracil residues that could be present in the ssDNA portions of the genome replicative intermediates and that would compromise viral genome integrity (Serrano-Heras et al., 2006; Muñoz-Espín et al., 2012).

It is worth mentioning that  $\Phi$ 29-related phages PZA, B103, Nf, and GA-1 encode homologs of p56. The product of GA-1 gene 56 was purified and shown to inhibit UDG activity in extracts of both *B. subtilis* and *B. pumilus*, which is the natural host of GA-1 (Pérez-Lago et al., 2011).

The elucidation of the function of several  $\Phi$ 29 proteins yet to be characterized and the improvement of *in vivo* techniques for both protein-protein and protein-DNA interactions detection will lead to a better understanding of the virus-host interactome in the future.

## AUTHOR CONTRIBUTIONS

All authors listed have made substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

This work has been supported by grants from the Spanish Ministry of Economy and Competitiveness (BFU2014-52656-P to MS) and (BFU2014-53791-P to MV), ComFuturo Grant from Fundación General CSIC (to MR) and by an Institutional grant from Fundación Ramón Areces to the Centro de Biología Molecular “Severo Ochoa.”



## REFERENCES

- Abril, A. M., Salas, M., Andreu, J. M., Hermoso, J. M., and Rivas, G. (1997). Phage  $\Phi$ 29 protein p6 is in a monomer-dimer equilibrium that shifts to higher association states at the millimolar concentrations found *in vivo*. *Biochemistry* 36, 11901–11908. doi: 10.1021/bi970994e
- Abril, A. M., Salas, M., and Hermoso, J. M. (2000). Identification of residues within two regions involved in self-association of viral histone-like protein p6 from phage  $\Phi$ 29. *J. Biol. Chem.* 275, 26404–26410. doi: 10.1074/jbc.M002739200
- Abril, A., Salas, M., and Hermoso, J. M. (2002). The *in vivo* function of phage  $\Phi$ 29 nucleoid-associated protein p6 requires formation of dimers. *Gene* 296, 187–194. doi: 10.1016/S0378-1119(02)00857-0
- Ackermann, H. W. (1998). Tailed bacteriophages: the order Caudovirales. *Adv. Virus Res.* 51, 135–201. doi: 10.1016/S0065-3527(08)60785-X
- Ackermann, H. W. (2003). Bacteriophage observations and evolution. *Res. Microbiol.* 154, 245–251. doi: 10.1016/S0923-2508(03)00067-6
- Alcorlo, M., Jiménez, M., Ortega, A., Hermoso, J. M., Salas, M., Minton, A. P., et al. (2009). Analytical ultracentrifugation studies of phage  $\Phi$ 29 protein p6 binding to DNA. *J. Mol. Biol.* 385, 1616–1629. doi: 10.1016/j.jmb.2008.11.044
- Alcorlo, M., Salas, M., and Hermoso, J. M. (2007). *In vivo* DNA binding of bacteriophage GA-1 protein p6. *J. Bacteriol.* 189, 8024–8033. doi: 10.1128/JB.01047-07
- Anderson, D. L., and Reilly, B. E. (1993). “Morphogenesis of bacteriophage  $\Phi$ 29,” in *Bacillus subtilis and other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*, eds A. L. Sonenshein, J. A. Hoch, and R. Losick (Washington, DC: American Society for Microbiology), 859–867.
- Asensio, J. L., Pérez-Lago, L., Lázaro, J. M., González, C., Serrano-Heras, G., and Salas, M. (2011). Novel dimeric structure of phage  $\Phi$ 29-encoded protein p56: insights into uracil-DNA glycosylase inhibition. *Nucleic Acids Res.* 39, 9779–9788. doi: 10.1093/nar/gkr667
- Baños-Sanz, J. I., Mojardín, L., Sanz-Aparicio, J., Lázaro, J. M., Villar, L., Serrano-Heras, G., et al. (2013). Crystal structure and functional insights into uracil-DNA glycosylase inhibition by phage  $\Phi$ 29 DNA mimic protein p56. *Nucleic Acids Res.* 41, 6761–6773. doi: 10.1093/nar/gkt395
- Bao, K., and Cohen, S. N. (2001). Terminal proteins essential for the replication of linear plasmids and chromosomes in *Streptomyces*. *Genes Dev.* 15, 1518–1527. doi: 10.1101/gad.896201
- Bath, C., Cukalac, T., Porter, K., and Dyall-Smith, M. L. (2006). His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, Salterprovirus. *Virology* 350, 228–239. doi: 10.1016/j.virol.2006.02.005
- Beese, L. S., and Steitz, T. A. (1991). Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* 10, 25–33.
- Berman, A. J., Kamtekar, S., Goodman, J. L., Lázaro, J. M., de Vega, M., Blanco, L., et al. (2007). Structures of phi29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases. *EMBO J.* 26, 3494–3505. doi: 10.1038/sj.emboj.7601780
- Bernad, A., Blanco, L., Lázaro, J. M., Martín, G., and Salas, M. (1989). A conserved 3'-5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. *Cell* 59, 219–228. doi: 10.1016/0092-8674(89)90883-0
- Bernad, A., Zaballos, A., Salas, M., and Blanco, L. (1987). Structural and functional relationships between prokaryotic and eukaryotic DNA polymerases. *EMBO J.* 6, 4219–4225.
- Bjornsti, M. A., Reilly, B. E., and Anderson, D. L. (1982). Morphogenesis of bacteriophage  $\Phi$ 29 of *Bacillus subtilis*: DNA-gp3 intermediate in *in vivo* and *in vitro* assembly. *J. Virol.* 41, 508–517.
- Bjornsti, M. A., Reilly, B. E., and Anderson, D. L. (1983). Morphogenesis of bacteriophage  $\Phi$ 29 of *Bacillus subtilis*: oriented and quantized *in vitro* packaging of DNA protein gp3. *J. Virol.* 45, 383–396.
- Blanco, L., Bernad, A., Lázaro, J. M., Martín, G., Garmendia, C., and Salas, M. (1989). Highly efficient DNA synthesis by the phage  $\Phi$ 29 DNA polymerase. Symmetrical mode of DNA replication. *J. Biol. Chem.* 264, 8935–8940.
- Blanco, L., Bernad, A., and Salas, M. (1988). Transition from initiation to elongation in protein-primed  $\Phi$ 29 DNA replication: salt-dependent stimulation by the viral protein p6. *J. Virol.* 62, 4167–4172.
- Blanco, L., Gutiérrez, J., Lázaro, J. M., Bernad, A., and Salas, M. (1986). Replication of phage  $\Phi$ 29 DNA *in vitro*: role of the viral protein p6 in initiation and elongation. *Nucleic Acids Res.* 14, 4923–4937. doi: 10.1093/nar/14.12.4923
- Blanco, L., Lázaro, J. M., de Vega, M., Bonnin, A., and Salas, M. (1994). Terminal protein-primed DNA amplification. *Proc. Natl. Acad. Sci. U.S.A.* 91, 12198–12202. doi: 10.1073/pnas.91.25.12198
- Blanco, L., Prieto, I., Gutiérrez, J., Bernad, A., Lázaro, J. M., Hermoso, J. M., et al. (1987). Effect of NH<sub>4</sub><sup>+</sup> ions on  $\Phi$ 29 DNA-protein p3 replication: formation of a complex between the terminal protein and the DNA polymerase. *J. Virol.* 61, 3983–3991.
- Blanco, L., and Salas, M. (1984). Characterization and purification of a phage  $\Phi$ 29-encoded DNA polymerase required for the initiation of replication. *Proc. Natl. Acad. Sci. U.S.A.* 81, 5325–5329. doi: 10.1073/pnas.81.17.5325
- Blanco, L., and Salas, M. (1985a). Characterization of a 3'-5' exonuclease activity in the phage  $\Phi$ 29-encoded DNA polymerase. *Nucleic Acids Res.* 13, 1239–1249. doi: 10.1093/nar/13.4.1239
- Blanco, L., and Salas, M. (1985b). Replication of phage  $\Phi$ 29 DNA with purified terminal protein and DNA polymerase: synthesis of full-length  $\Phi$ 29 DNA. *Proc. Natl. Acad. Sci. U.S.A.* 82, 6404–6408. doi: 10.1073/pnas.82.19.6404
- Blanco, L., and Salas, M. (1986). Effect of aphidicolin and nucleotide analogs on the phage  $\Phi$ 29 DNA polymerase. *Virology* 153, 179–187. doi: 10.1016/0042-6822(86)90021-8
- Blanco, L., and Salas, M. (1995). Mutational analysis of bacteriophage  $\Phi$ 29 DNA polymerase. *Methods Enzymol.* 262, 283–294. doi: 10.1016/0076-6879(95)62024-9
- Blanco, L., and Salas, M. (1996). Relating structure to function in  $\Phi$ 29 DNA polymerase. *J. Biol. Chem.* 271, 8509–8512. doi: 10.1074/jbc.271.15.8509
- Blasco, M. A., Bernad, A., Blanco, L., and Salas, M. (1991). Characterization and mapping of the pyrophosphorolytic activity of the phage  $\Phi$ 29 DNA polymerase. Involvement of amino acid motifs highly conserved in alpha-like DNA polymerases. *J. Biol. Chem.* 266, 7904–7909.
- Blasco, M. A., Blanco, L., Parés, E., Salas, M., and Bernad, A. (1990). Structural and functional analysis of temperature-sensitive mutants of the phage  $\Phi$ 29 DNA polymerase. *Nucleic Acids Res.* 18, 4763–4770.
- Bravo, A., Hermoso, J. M., and Salas, M. (1994a). *In vivo* functional relationships among terminal proteins of *Bacillus subtilis*  $\Phi$ 29-related phages. *Gene* 148, 107–112. doi: 10.1016/0378-1119(94)90242-9
- Bravo, A., Hermoso, J. M., and Salas, M. (1994b). A genetic approach to the identification of functional amino acids in protein p6 of *Bacillus subtilis* phage  $\Phi$ 29. *Mol. Gen. Genet.* 245, 529–536. doi: 10.1007/BF00282215
- Bravo, A., Illana, B., and Salas, M. (2000). Compartmentalization of phage  $\Phi$ 29 DNA replication: interaction between the primer terminal protein and the membrane-associated protein p1. *EMBO J.* 19, 5575–5584. doi: 10.1093/emboj/19.20.5575
- Brüssow, H., and Hendrix, R. W. (2002). Phage genomics: small is beautiful. *Cell* 108, 13–16. doi: 10.1016/S0092-8674(01)00637-7
- Caldentey, J., Blanco, L., Bamford, D. H., and Salas, M. (1993). *In vitro* replication of bacteriophage PRD1 DNA. Characterization of the protein-primed initiation site. *Nucleic Acids Res.* 21, 3725–3730. doi: 10.1093/nar/21.16.3725
- Calles, B., Salas, M., and Rojo, F. (2002). The  $\Phi$ 29 transcriptional regulator contacts the nucleoid protein p6 to organize a repression complex. *EMBO J.* 21, 6185–6194. doi: 10.1093/emboj/cdf623
- Camacho, A., and Salas, M. (2000). Pleiotropic effect of protein p6 on the viral cycle of bacteriophage  $\Phi$ 29. *J. Bacteriol.* 182, 6927–6932. doi: 10.1128/JB.182.24.6927-6932.2000
- Camacho, A., and Salas, M. (2001). Mechanism for the switch of  $\Phi$ 29 DNA early to late transcription by regulatory protein p4 and histone-like protein p6. *EMBO J.* 20, 6060–6070. doi: 10.1093/emboj/20.21.6060
- Carrascosa, J. L., Camacho, A., Moreno, F., Jiménez, F., Mellado, R. P., Viñuela, E., et al. (1976). *Bacillus subtilis* phage  $\Phi$ 29. Characterization of gene products and functions. *Eur. J. Biochem.* 66, 229–241.
- Chaconas, G., and Chen, C. W. (2005). “Linear chromosomes in bacteria: no longer going around in circles,” in *The Bacterial Chromosome*, ed N. P. Higgins (Washington, DC: American Society for Microbiology), 525–539.
- Chang, P. C., and Cohen, S. N. (1994). Bidirectional replication from an internal origin in a linear *Streptomyces* plasmid. *Science* 265, 952–954. doi: 10.1126/science.8052852
- Chase, J. W., and Williams, K. R. (1986). Single-stranded DNA binding proteins required for DNA replication. *Annu. Rev. Biochem.* 55, 103–136. doi: 10.1146/annurev.bi.55.070186.000535

- Cole, A. R., Ofer, S., Ryzhenkova, K., Baltulionis, G., Hornyak, P., and Savva, R. (2013). Architecturally diverse proteins converge on an analogous mechanism to inactivate Uracil-DNA glycosylase. *Nucleic Acids Res.* 41, 8760–8775. doi: 10.1093/nar/gkt633
- Cone, R., Bonura, T., and Friedberg, E. C. (1980). Inhibitor of uracil-DNA glycosylase induced by bacteriophage PBS2. Purification and preliminary characterization. *J. Biol. Chem.* 255, 10354–10358.
- Dahl, J. M., Mai, A. H., Cherf, G. M., Jetha, N. N., Garalde, D. R., Marziali, A., et al. (2012). Direct observation of translocation in individual DNA polymerase complexes. *J. Biol. Chem.* 287, 13407–13421. doi: 10.1074/jbc.M111.338418
- Dahl, J. M., Wang, H., Lázaro, J. M., Salas, M., and Lieberman, K. R. (2014). Kinetic mechanisms governing stable ribonucleotide incorporation in individual DNA polymerase complexes. *Biochemistry* 53, 8061–8076. doi: 10.1021/bi501216a
- Dean, F. B., Hosono, S., Fang, L., Wu, X., Faruqi, A. F., Bray-Ward, P., et al. (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. U.S.A.* 99, 5261–5266. doi: 10.1073/pnas.082089499
- Dean, F. B., Nelson, J. R., Giesler, T. L., and Lasken, R. S. (2001). Rapid amplification of plasmid and phage DNA using  $\Phi$ 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* 11, 1095–1099. doi: 10.1101/gr.180501
- del Prado, A., Lázaro, J. M., Longás, E., Villar, L., de Vega, M., and Salas, M. (2015). Insights into the determination of the templating nucleotide at the initiation of  $\Phi$ 29 DNA replication. *J. Biol. Chem.* 290, 27138–27145. doi: 10.1074/jbc.M115.682278
- del Prado, A., Lázaro, J. M., Villar, L., Salas, M., and de Vega, M. (2013). Dual role of  $\Phi$ 29 DNA polymerase Lys529 in stabilisation of the DNA priming-terminus and the terminal protein-priming residue at the polymerisation site. *PLoS ONE* 8:e72765. doi: 10.1371/journal.pone.0072765
- del Prado, A., Villar, L., de Vega, M., and Salas, M. (2012). Involvement of residues of the  $\Phi$ 29 terminal protein intermediate and priming domains in the formation of a stable and functional heterodimer with the replicative DNA polymerase. *Nucleic Acids Res.* 40, 3886–3897. doi: 10.1093/nar/gkr1283
- de Vega, M., Blanco, L., and Salas, M. (1998a).  $\Phi$ 29 DNA polymerase residue Ser122, a single-stranded DNA ligand for 3'-5' exonucleolysis, is required to interact with the terminal protein. *J. Biol. Chem.* 273, 28966–28977. doi: 10.1074/jbc.273.44.28966
- de Vega, M., Blanco, L., and Salas, M. (1999). Processive proofreading and the spatial relationship between polymerase and exonuclease active sites of bacteriophage  $\Phi$ 29 DNA polymerase. *J. Mol. Biol.* 292, 39–51. doi: 10.1006/jmbi.1999.3052
- de Vega, M., Ilyina, T., Lázaro, J. M., Salas, M., and Blanco, L. (1997). An invariant lysine residue is involved in catalysis at the 3'-5' exonuclease active site of eukaryotic-type DNA polymerases. *J. Mol. Biol.* 270, 65–78. doi: 10.1006/jmbi.1997.1093
- de Vega, M., Lázaro, J. M., Mencia, M., Blanco, L., and Salas, M. (2010). Improvement of  $\Phi$ 29 DNA polymerase amplification performance by fusion of DNA binding motifs. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16506–16511. doi: 10.1073/pnas.1011428107
- de Vega, M., Lázaro, J. M., Salas, M., and Blanco, L. (1996). Primer-terminus stabilization at the 3'-5' exonuclease active site of  $\Phi$ 29 DNA polymerase. Involvement of two amino acid residues highly conserved in proofreading DNA polymerases. *EMBO J.* 15, 1182–1192.
- de Vega, M., Lázaro, J. M., Salas, M., and Blanco, L. (1998b). Mutational analysis of  $\Phi$ 29 DNA polymerase residues acting as ssDNA ligands for 3'-5' exonucleolysis. *J. Mol. Biol.* 279, 807–822. doi: 10.1006/jmbi.1998.1805
- de Vega, M., and Salas, M. (2011). "Protein-primed replication of bacteriophage  $\Phi$ 29 DNA," in *DNA Replication and Related Cellular Processes*, ed J. Kusic-Tima (Rijeka: InTech), 179–206.
- Drlica, K. (1992). Control of bacterial DNA supercoiling. *Mol. Microbiol.* 6, 425–433. doi: 10.1111/j.1365-2958.1992.tb01486.x
- Ducani, C., Bernardinelli, G., and Högberg, B. (2014). Rolling circle replication requires single-stranded DNA binding protein to avoid termination and production of double-stranded DNA. *Nucleic Acids Res.* 42, 10596–10604. doi: 10.1093/nar/gku737
- Dufour, E., Méndez, J., Lázaro, J. M., de Vega, M., Blanco, L., and Salas, M. (2000). An aspartic acid residue in TPR-1, a specific region of protein-priming DNA polymerases, is required for the functional interaction with primer terminal protein. *J. Mol. Biol.* 304, 289–300. doi: 10.1006/jmbi.200.04216
- Esteban, J. A., Salas, M., and Blanco, L. (1993). Fidelity of  $\Phi$ 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *J. Biol. Chem.* 268, 2719–2726.
- Esteban, J. A., Soengas, M. S., Salas, M., and Blanco, L. (1994). 3'-5' exonuclease active site of  $\Phi$ 29 DNA polymerase. Evidence favoring a metal ion-assisted reaction mechanism. *J. Biol. Chem.* 269, 31946–31954.
- Franklin, M. C., Wang, J., and Steitz, T. A. (2001). Structure of the replicating complex of a pol  $\alpha$  family DNA polymerase. *Cell* 105, 657–667. doi: 10.1016/S0092-8674(01)00367-1
- Freire, R., Salas, M., and Hermoso, J. M. (1994). A new protein domain for binding to DNA through the minor groove. *EMBO J.* 13, 4353–4360.
- Freire, R., Serrano, M., Salas, M., and Hermoso, J. M. (1996). Activation of replication origins in  $\Phi$ 29-related phages requires the recognition of initiation proteins to specific nucleoprotein complexes. *J. Biol. Chem.* 271, 31000–31007. doi: 10.1074/jbc.271.48.31000
- Fricova, D., Valach, M., Farkas, Z., Pfeiffer, I., Kucsera, J., Tomaska, L., et al. (2010). The mitochondrial genome of the pathogenic yeast *Candida subhashii*: GC-rich linear DNA with a protein covalently attached to the 5' termini. *Microbiology* 156, 2153–2163. doi: 10.1099/mic.0.038646-0
- García, J. A., Peñalva, M. A., Blanco, L., and Salas, M. (1984). Template requirements for initiation of phage  $\Phi$ 29 DNA replication *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* 81, 80–84.
- Garmendia, C., Bernad, A., Esteban, J. A., Blanco, L., and Salas, M. (1992). The bacteriophage  $\Phi$ 29 DNA polymerase, a proofreading enzyme. *J. Biol. Chem.* 267, 2594–2599.
- Garmendia, C., Hermoso, J. M., and Salas, M. (1990). Functional domain for priming activity in the phage  $\Phi$ 29 terminal protein. *Gene* 88, 73–79. doi: 10.1016/0378-1119(90)90061-U
- Garmendia, C., Salas, M., and Hermoso, J. M. (1988). Site-directed mutagenesis in the DNA linking site of bacteriophage  $\Phi$ 29 terminal protein: isolation and characterization of a Ser232Thr mutant. *Nucleic Acids Res.* 16, 5727–5740. doi: 10.1093/nar/16.13.5727
- Gascón, I., Carrascosa, J. L., Villar, L., Lázaro, J. M., and Salas, M. (2002). Importance of the N-terminal region of the phage GA-1 single-stranded DNA-binding protein for its self-interaction ability and functionality. *J. Biol. Chem.* 277, 22534–22540. doi: 10.1074/jbc.M202430200
- Gascón, I., Gutiérrez, C., and Salas, M. (2000a). Structural and functional comparative study of the complexes formed by viral  $\Phi$ 29, Nf and GA-1 SSB proteins with DNA. *J. Mol. Biol.* 296, 989–999. doi: 10.1006/jmbi.200.03521
- Gascón, I., Lázaro, J. M., and Salas, M. (2000b). Differential functional behavior of viral  $\Phi$ 29, Nf and GA-1 SSB proteins. *Nucleic Acids Res.* 28, 2034–2042. doi: 10.1093/nar/28.10.2034
- Gella, P., Salas, M., and Mencia, M. (2014). Improved artificial origins for phage  $\Phi$ 29 terminal protein-primed replication. Insights into early replication events. *Nucleic Acids Res.* 42, 9792–9806. doi: 10.1093/nar/gku660
- González-Huici, V., Alcorlo, M., Salas, M., and Hermoso, J. M. (2004a). Bacteriophage  $\Phi$ 29 protein p6: an architectural protein involved in genome organization, replication and control of transcription. *J. Mol. Recognit.* 17, 390–396. doi: 10.1002/jmr.701
- González-Huici, V., Alcorlo, M., Salas, M., and Hermoso, J. M. (2004c). Binding of phage  $\Phi$ 29 architectural protein p6 to the viral genome: evidence for topological restriction of the phage linear DNA. *Nucl. Acids Res.* 32, 3493–3502. doi: 10.1093/nar/gkh668
- González-Huici, V., Lázaro, J. M., Salas, M., and Hermoso, J. M. (2000a). Specific recognition of parental terminal protein by DNA polymerase for initiation of protein-primed DNA replication. *J. Biol. Chem.* 275, 14678–14683. doi: 10.1074/jbc.M910058199
- González-Huici, V., Salas, M., and Hermoso, J. M. (2000b). Sequence requirements for protein-primed initiation and elongation of phage  $\Phi$ 29 DNA replication. *J. Biol. Chem.* 275, 40547–40553. doi: 10.1074/jbc.M007170200
- González-Huici, V., Salas, M., and Hermoso, J. M. (2004b). Genome wide, supercoiling-dependent *in vivo* binding of a viral protein involved in DNA replication and transcriptional control. *Nucleic Acids Res.* 32, 2306–2314. doi: 10.1093/nar/gkh565

- Gutiérrez, C., Freire, R., Salas, M., and Hermoso, J. M. (1994). Assembly of phage  $\Phi$ 29 genome with viral protein p6 into a compact complex. *EMBO J.* 13, 269–276.
- Gutiérrez, C., Martín, G., Sogo, J. M., and Salas, M. (1991). Mechanism of stimulation of DNA replication by bacteriophage  $\Phi$ 29 single-stranded DNA-binding protein p5. *J. Biol. Chem.* 266, 2104–2111.
- Gutiérrez, J., García, J. A., Blanco, L., and Salas, M. (1986a). Cloning and template activity of the origins of replication of phage  $\Phi$ 29 DNA. *Gene* 43, 1–11. doi: 10.1016/0378-1119(86)90002-8
- Gutiérrez, J., Vinós, J., Prieto, I., Méndez, E., Hermoso, J. M., and Salas, M. (1986b). Signals in the  $\Phi$ 29 DNA-terminal protein template for the initiation of phage  $\Phi$ 29 DNA replication. *Virology* 155, 474–483. doi: 10.1016/0042-6822(86)90209-6
- Harding, N. E., and Ito, J. (1980). DNA replication of bacteriophage  $\Phi$ 29: characterization of the intermediates and location of the termini of replication. *Virology* 104, 323–338. doi: 10.1016/0042-6822(80)90337-2
- Hermoso, J. M., Méndez, E., Soriano, F., and Salas, M. (1985). Location of the serine residue involved in the linkage between the terminal protein and the DNA of phage  $\Phi$ 29. *Nucleic Acids Res.* 13, 7715–7728. doi: 10.1093/nar/13.21.7715
- Hirokawa, H. (1972). Transfecting deoxyribonucleic acid of Bacillus bacteriophage  $\Phi$ 29 that is protease sensitive. *Proc. Natl. Acad. Sci. U.S.A.* 69, 1555–1559. doi: 10.1073/pnas.69.6.1555
- Holguera, I., Ballesteros-Plaza, D., Muñoz-Espín, D., and Salas, M. (2012). Disclosing the *in vivo* organization of a viral histone-like protein in *Bacillus subtilis* mediated by its capacity to recognize the viral genome. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5723–5728. doi: 10.1073/pnas.1203824109
- Holguera, I., Muñoz-Espín, D., and Salas, M. (2015). Dissecting the role of the  $\Phi$ 29 terminal protein DNA binding residues in viral DNA replication. *Nucleic Acids Res.* 43, 2790–2801. doi: 10.1093/nar/gkv127
- Holguera, I., Redrejo-Rodríguez, M., Salas, M., and Muñoz-Espín, D. (2014). New insights in the  $\Phi$ 29 terminal protein DNA-binding and host nucleoid localization functions. *Mol. Microbiol.* 91, 232–241. doi: 10.1111/mmi.12456
- Ibarra, B., Chemla, Y. R., Pylasunov, S., Smith, S. B., Lázaro, J. M., Salas, M., et al. (2009). Proofreading dynamics of a processive DNA polymerase. *EMBO J.* 28, 2794–2802. doi: 10.1038/emboj.2009.219
- Illana, B., Blanco, L., and Salas, M. (1996). Functional characterization of the genes coding for the terminal protein and DNA polymerase from bacteriophage GA-1. Evidence for a sliding-back mechanism during protein-primed GA-1 DNA replication. *J. Mol. Biol.* 264, 453–464. doi: 10.1006/jmbi.1996.0653
- Illana, B., Lázaro, J. M., Gutiérrez, C., Meijer, W. J., Blanco, L., and Salas, M. (1999). Phage  $\Phi$ 29 terminal protein residues Asn80 and Tyr82 are recognition elements of the replication origins. *J. Biol. Chem.* 274, 15073–15079. doi: 10.1074/jbc.274.21.15073
- Illana, B., Zaballos, A., Blanco, L., and Salas, M. (1998). The RGD sequence in phage  $\Phi$ 29 terminal protein is required for interaction with  $\Phi$ 29 DNA polymerase. *Virology* 248, 12–19. doi: 10.1006/viro.1998.9276
- Inciarte, M. R., Salas, M., and Sogo, J. M. (1980). Structure of replicating DNA molecules of *Bacillus subtilis* bacteriophage  $\Phi$ 29. *J. Virol.* 34, 187–199.
- Inoue, J., Shigemori, Y., and Mikawa, T. (2006). Improvements of rolling circle amplification (RCA) efficiency and accuracy using *Thermus thermophilus* SSB mutant protein. *Nucleic Acids Res.* 34, e69. doi: 10.1093/nar/gkl350
- Johne, R., Müller, H., Rector, A., van Ranst, M., and Stevens, H. (2009). Rolling-circle amplification of viral DNA genomes using  $\Phi$ 29 polymerase. *Trends Microbiol.* 17, 205–211. doi: 10.1016/j.tim.2009.02.004
- Kamtekar, S., Berman, A. J., Wang, J., Lázaro, J. M., de Vega, M., Blanco, L., et al. (2004). Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage  $\Phi$ 29. *Mol. Cell* 16, 609–618. doi: 10.1016/j.molcel.2004.10.019
- Kamtekar, S., Berman, A. J., Wang, J., Lázaro, J. M., de Vega, M., Blanco, L., et al. (2006). The  $\phi$ 29 DNA polymerase:protein-primer structure suggests a model for the initiation to elongation transition. *EMBO J.* 25, 1335–1343. doi: 10.1038/sj.emboj.7601027
- Kapitonov, V. V., and Jurka, J. (2006). Self-synthesizing DNA transposons in eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 103, 4540–4545. doi: 10.1073/pnas.0600833103
- Katz, G. E., Price, A. R., and Pomerantz, M. J. (1976). Bacteriophage PBS2-induced inhibition of uracil-containing DNA degradation. *J. Virol.* 20, 535–538.
- King, A. J., and van der Vliet, P. C. (1994). A precursor terminal protein-trinucleotide intermediate during initiation of adenovirus DNA replication: regeneration of molecular ends *in vitro* by a jumping back mechanism. *EMBO J.* 13, 5786–5792.
- Kur, J., Olszewski, M., Dlugolecka, A., and Filipkowski, P. (2005). Single-stranded DNA-binding proteins (SSBs) – sources and applications in molecular biology. *Acta Biochim. Pol.* 52, 569–574.
- Lázaro, J. M., Blanco, L., and Salas, M. (1995). Purification of bacteriophage  $\Phi$ 29 DNA polymerase. *Methods Enzymol.* 262, 42–49. doi: 10.1016/0076-6879(95)62007-9
- Leavitt, M. C., and Ito, J. (1987). Nucleotide sequence of Bacillus phage Nf terminal protein gene. *Nucleic Acids Res.* 15, 5251–5259. doi: 10.1093/nar/15.13.5251
- Longás, E., Villar, L., Lázaro, J. M., de Vega, M., and Salas, M. (2008). Phage  $\Phi$ 29 and Nf terminal protein-priming domain specifies the internal template nucleotide to initiate DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18290–18295. doi: 10.1073/pnas.0809882105
- López-Bueno, A., Tamames, J., Velázquez, D., Moya, A., Quesada, A., and Alcami, A. (2009). High diversity of the viral community from an Antarctic lake. *Science* 326, 858–861. doi: 10.1126/science.1179287
- Martín, A. C., Blanco, L., García, P., Salas, M., and Méndez, J. (1996). *In vitro* protein-primed initiation of pneumococcal phage Cp-1 DNA replication occurs at the third 3' nucleotide of the linear template: a stepwise sliding-back mechanism. *J. Mol. Biol.* 260, 369–377. doi: 10.1006/jmbi.1996.0407
- Martin, G., Lázaro, J. M., Méndez, E., and Salas, M. (1989). Characterization of the phage  $\Phi$ 29 protein p5 as a single-stranded DNA binding protein. Function in  $\phi$ 29 DNA-protein p3 replication. *Nucleic Acids Res.* 17, 3663–3672. doi: 10.1093/nar/17.10.3663
- Meijer, W. J., Horcajadas, J. A., and Salas, M. (2001).  $\Phi$ 29 family of phages. *Microbiol. Mol. Biol. Rev.* 65, 261–287. doi: 10.1128/MMBR.65.2.261-287.2001
- Meinhardt, F., Schaffrath, R., and Larsen, M. (1997). Microbial linear plasmids. *Appl. Microbiol. Biotechnol.* 47, 329–336. doi: 10.1007/s002530050936
- Mellado, R. P., Moreno, F., Viñuela, E., Salas, M., Reilly, B. E., and Anderson, D. L. (1976). Genetic analysis of bacteriophage  $\Phi$ 29 of *Bacillus subtilis*: integration and mapping of reference mutants of two collections. *J. Virol.* 19, 495–500.
- Méndez, J., Blanco, L., Esteban, J. A., Bernad, A., and Salas, M. (1992). Initiation of  $\Phi$ 29 DNA replication occurs at the second 3' nucleotide of the linear template: a sliding-back mechanism for protein-primed DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* 89, 9579–9583. doi: 10.1073/pnas.89.20.9579
- Méndez, J., Blanco, L., and Salas, M. (1997). Protein-primed DNA replication: a transition between two modes of priming by a unique DNA polymerase. *EMBO J.* 16, 2519–2527. doi: 10.1093/emboj/16.9.2519
- Mikawa, T., Inoue, J., and Shigemori, Y. (2009). Single-stranded DNA binding protein facilitates specific enrichment of circular DNA molecules using rolling circle amplification. *Anal. Biochem.* 391, 81–84. doi: 10.1016/j.ab.2009.05.013
- Moak, M., and Molineux, I. J. (2004). Peptidoglycan hydrolytic activities associated with bacteriophage virions. *Mol. Microbiol.* 51, 1169–1183. doi: 10.1046/j.1365-2958.2003.03894.x
- Morin, J. A., Cao, F. J., Lázaro, J. M., Arias-Gonzalez, J. R., Valpuesta, J. M., Carrascosa, J. L., et al. (2012). Active DNA unwinding dynamics during processive DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* 109, 8115–8120. doi: 10.1073/pnas.1204759109
- Muñoz-Espín, D., Holguera, I., Ballesteros-Plaza, D., Carballido-López, R., and Salas, M. (2010). Viral terminal protein directs early organization of phage DNA replication at the bacterial nucleoid. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16548–16553. doi: 10.1073/pnas.1010530107
- Muñoz-Espín, D., Serrano-Heras, G., and Salas, M. (2012). Role of host factors in bacteriophage  $\Phi$ 29 DNA replication. *Adv. Virus Res.* 82, 351–383. doi: 10.1016/B978-0-12-394621-8.00020-0
- Murthy, V., Meijer, W. J., Blanco, L., and Salas, M. (1998). DNA polymerase template switching at specific sites on the  $\Phi$ 29 genome causes the *in vivo* accumulation of subgenomic  $\Phi$ 29 DNA molecules. *Mol. Microbiol.* 29, 787–798. doi: 10.1046/j.1365-2958.1998.00972.x
- Murzin, A. G. (1993). OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.* 12, 861–867.



- Otero, M. J., Lázaro, J. M., and Salas, M. (1990). Deletions at the N terminus of bacteriophage  $\Phi$ 29 protein p6: DNA binding and activity in  $\Phi$ 29 DNA replication. *Gene* 95, 25–30. doi: 10.1016/0378-1119(90)90409-K
- Paces, V., Vlcek, C., Urbánek, P., and Hostomský, Z. (1985). Nucleotide sequence of the major early region of *Bacillus subtilis* phage PZA, a close relative of  $\Phi$ 29. *Gene* 38, 45–56. doi: 10.1016/0378-1119(85)90202-1
- Pastrana, R., Lázaro, J. M., Blanco, L., García, J. A., Méndez, E., and Salas, M. (1985). Overproduction and purification of protein P6 of *Bacillus subtilis* phage  $\Phi$ 29: role in the initiation of DNA replication. *Nucleic Acids Res.* 13, 3083–3100. doi: 10.1093/nar/13.9.3083
- Pecenková, T., Benes, V., Paces, J., Vlcek, C., and Paces, V. (1997). Bacteriophage B103: complete DNA sequence of its genome and relationship to other *Bacillus* phages. *Gene* 199, 157–163. doi: 10.1016/S0378-1119(97)00363-6
- Pecenková, T., and V., Paces (1999). Molecular phylogeny of  $\Phi$ 29-like phages and their evolutionary relatedness to other protein-primed replicating phages and other phages hosted by Gram-positive bacteria. *J. Mol. Evol.* 48, 197–208. doi: 10.1007/PL00006458
- Pei, J., Kim, B. H., and Grishin, N. V. (2008). PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* 36, 2295–2300. doi: 10.1093/nar/gkn072
- Peng, X., Basta, T., Häring, M., Garrett, R. A., and Prangishvili, D. (2007). Genome of the Acidianus bottle-shaped virus and insights into the replication and packaging mechanisms. *Virology* 364, 237–243. doi: 10.1016/j.virol.2007.03.005
- Perales, C., Cava, F., Meijer, W. J., and Berenguer, J. (2003). Enhancement of DNA, cDNA synthesis and fidelity at high temperatures by a dimeric single-stranded DNA-binding protein. *Nucleic Acids Res.* 31, 6473–6480. doi: 10.1093/nar/gkg865
- Pérez-Arnaiz, P., Lázaro, J. M., Salas, M., and de Vega, M. (2006). Involvement of  $\Phi$ 29 DNA polymerase thumb subdomain in the proper coordination of synthesis and degradation during DNA replication. *Nucleic Acids Res.* 34, 3107–3115. doi: 10.1093/nar/gkl402
- Pérez-Arnaiz, P., Lázaro, J. M., Salas, M., and de Vega, M. (2009). Functional importance of bacteriophage  $\Phi$ 29 DNA polymerase residue Tyr148 in primer-terminus stabilisation at the 3′-5′ exonuclease active site. *J. Mol. Biol.* 391, 797–807. doi: 10.1016/j.jmb.2009.06.068
- Pérez-Arnaiz, P., Lázaro, J. M., Salas, M., and de Vega, M. (2010).  $\Phi$ 29 DNA polymerase active site: role of residue Val250 as metal-dNTP complex ligand and in protein-primed initiation. *J. Mol. Biol.* 395, 223–233. doi: 10.1016/j.jmb.2009.10.061
- Pérez-Arnaiz, P., Longás, E., Villar, L., Lázaro, J. M., Salas, M., and de Vega, M. (2007). Involvement of phage  $\Phi$ 29 DNA polymerase and terminal protein subdomains in conferring specificity during initiation of protein-primed DNA replication. *Nucleic Acids Res.* 35, 7061–7073. doi: 10.1093/nar/gkm749
- Pérez-Lago, L., Serrano-Heras, G., Baños, B., Lázaro, J. M., Alcorlo, M., Villar, L., et al. (2011). Characterization of *Bacillus subtilis* uracil-DNA glycosylase and its inhibition by phage  $\Phi$ 29 protein p56. *Mol. Microbiol.* 80, 1657–1666. doi: 10.1111/j.1365-2958.2011.07675.x
- Porter, K., and Dyall-Smith, M. L. (2008). Transfection of haloarchaea by the DNAs of spindle and round haloviruses and the use of transposon mutagenesis to identify non-essential regions. *Mol. Microbiol.* 70, 1236–1245. doi: 10.1111/j.1365-2958.2008.06478.x
- Prieto, I. (1986). *Mecanismo de la Iniciación de la Replicación del DNA del Bacteriófago  $\Phi$ 29*. Ph.D. thesis, Universidad Autónoma de Madrid.
- Prieto, I., Serrano, M., Lázaro, J. M., Salas, M., and Hermoso, J. M. (1988). Interaction of the bacteriophage  $\Phi$ 29 protein p6 with double-stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.* 85, 314–318. doi: 10.1073/pnas.85.2.314
- Qi, X., Bakht, S., Devos, K. M., Gale, M. D., and Osbourn, A. (2001). L-RCA (ligation-rolling circle amplification): a general method for genotyping of single nucleotide polymorphisms (SNPs). *Nucleic Acids Res.* 29:E116. doi: 10.1093/nar/29.22.e116
- Rapley, R. (1994). Enhancing PCR amplification and sequencing using DNA-binding proteins. *Mol. Biotechnol.* 2, 295–298. doi: 10.1007/BF02745882
- Redrejo-Rodríguez, M., Muñoz-Espín, D., Holguera, I., Mencía, M., and Salas, M. (2012). Functional eukaryotic nuclear localization signals are widespread in terminal proteins of bacteriophages. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18482–18487. doi: 10.1073/pnas.1216635109
- Redrejo-Rodríguez, M., Muñoz-Espín, D., Holguera, I., Mencía, M., and Salas, M. (2013). Nuclear and nucleoid localization are independently conserved functions in bacteriophage terminal proteins. *Mol. Microbiol.* 90, 858–868. doi: 10.1111/mmi.12404
- Rodríguez, I., Lázaro, J. M., Blanco, L., Kamtekar, S., Berman, A. J., Wang, J., et al. (2005). A specific subdomain in  $\Phi$ 29 DNA polymerase confers both processivity and strand-displacement capacity. *Proc. Natl. Acad. Sci. U.S.A.* 102, 6407–6412. doi: 10.1073/pnas.0500597102
- Rodríguez, I., Lázaro, J. M., Salas, M., and de Vega, M. (2004).  $\Phi$ 29 DNA polymerase-terminal protein interaction. Involvement of residues specifically conserved among protein-primed DNA polymerase. *J. Mol. Biol.* 337, 829–841. doi: 10.1016/j.jmb.2004.02.018
- Rodríguez, I., Lázaro, J. M., Salas, M., and de Vega, M. (2009). Involvement of the TPR2 subdomain movement in the activities of  $\Phi$ 29 DNA polymerase. *Nucleic Acids Res.* 37, 193–203. doi: 10.1093/nar/gkn928
- Ronda, C., López, R., Gómez, A., and García, E. (1983). Protease-sensitive transfection of *Streptococcus pneumoniae* with bacteriophage Cp-1 DNA. *J. Virol.* 48, 721–730.
- Roucourt, B., and Lavigne, R. (2009). The role of interactions between phage and animal proteins within the infected cell: a diverse and puzzling interactome. *Environ. Microbiol.* 11, 2789–2805. doi: 10.1111/j.1462-2920.2009.02029.x
- Salas, M. (1991). Protein-priming of DNA replication. *Annu. Rev. Biochem.* 60, 39–71. doi: 10.1146/annurev.bi.60.070191.000351
- Salas, M., and de Vega, M. (2006). “Bacteriophage protein-primed DNA replication,” in *Recent Advances in DNA Virus Replication by Research Signpost Transworld Research Network*, ed K. L. Hefferon (Kerala: Research Signpost), 259–288.
- Santos, E., Lázaro, J. M., Pérez-Arnaiz, P., Salas, M., and de Vega, M. (2014). Role of the LEXE motif of protein-primed DNA polymerases in the interaction with the incoming nucleotide. *J. Biol. Chem.* 289, 2888–2898. doi: 10.1074/jbc.M113.530980
- Savva, R., and Pearl, L. H. (1995). Cloning and expression of the uracil-DNA glycosylase inhibitor (UGI) from bacteriophage PBS-1 and crystallization of a uracil-DNA glycosylase-UGI complex. *Proteins* 22, 287–289. doi: 10.1002/prot.340220310
- Savva, R., McAuley-Hecht, K., Brown, T., and Pearl, L. (1995). The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature* 373, 487–493. doi: 10.1038/373487a0
- Savvides, S. N., Raghunathan, S., Fütterer, K., Kozlov, A. G., Lohman, T. M., and Waksman, G. (2004). The C-terminal domain of full-length *E. coli* SSB is disordered even when bound to DNA. *Protein Sci.* 13, 1942–1947. doi: 10.1110/ps.04661904
- Schaack, J., Ho, W. Y., Freimuth, P., and Shenk, T. (1990). Adenovirus terminal protein mediates both nuclear matrix association and efficient transcription of adenovirus DNA. *Genes Dev.* 4, 1197–1208. doi: 10.1101/gad.4.7.1197
- Serna-Rico, A., Illana, B., Salas, M., and Meijer, W. J. (2000). The putative coiled coil domain of the  $\Phi$ 29 terminal protein is a major determinant involved in recognition of the origin of replication. *J. Biol. Chem.* 275, 40529–40538. doi: 10.1074/jbc.M007855200
- Serna-Rico, A., Muñoz-Espín, D., Villar, L., Salas, M., and Meijer, W. J. (2003). The integral membrane protein p16.7 organizes *in vivo*  $\Phi$ 29 DNA replication through interaction with both the terminal protein and ssDNA. *EMBO J.* 22, 2297–2306. doi: 10.1093/emboj/cdg221
- Serrano, M., Gutiérrez, C., Freire, R., Bravo, A., Salas, M., and Hermoso, J. M. (1994). Phage  $\Phi$ 29 protein p6: a viral histone-like protein. *Biochimie* 76, 981–991. doi: 10.1016/0300-9084(94)90023-X
- Serrano, M., Gutiérrez, C., Salas, M., and Hermoso, J. M. (1993a). Superhelical path of the DNA in the nucleoprotein complex that activates the initiation of phage  $\Phi$ 29 DNA replication. *J. Mol. Biol.* 230, 248–259. doi: 10.1006/jmbi.1993.1140
- Serrano, M., Gutierrez, J., Prieto, I., Hermoso, J. M., and Salas, M. (1989). Signals at the bacteriophage  $\Phi$ 29 DNA replication origins required for protein p6 binding and activity. *EMBO J.* 8, 1879–1885.
- Serrano, M., Salas, M., and Hermoso, J. M. (1990). A novel nucleoprotein complex at a replication origin. *Science* 248, 1012–1016. doi: 10.1126/science.2111580
- Serrano, M., Salas, M., and Hermoso, J. M. (1993b). Multimeric complexes formed by DNA-binding proteins of low sequence specificity. *Trends Biochem. Sci.* 18, 202–206. doi: 10.1016/0968-0004(93)90187-R
- Serrano-Heras, G., Bravo, A., and Salas, M. (2008). Phage  $\Phi$ 29 protein p56 prevents viral DNA replication impairment caused by uracil excision activity



- of uracil-DNA glycosylase. *Proc. Natl. Acad. Sci. U.S.A.* 105, 19044–19049. doi: 10.1073/pnas.0808797105
- Serrano-Heras, G., Ruiz-Maso, J. A., del Solar, G., Espinosa, M., Bravo, A., and Salas, M. (2007). Protein p56 from the *Bacillus subtilis* phage  $\Phi$ 29 inhibits DNA-binding ability of uracil-DNA glycosylase. *Nucleic Acids Res.* 35, 5393–5401. doi: 10.1093/nar/gkm584
- Serrano-Heras, G., Salas, M., and Bravo, A. (2006). A uracil-DNA glycosylase inhibitor encoded by a non-uracil containing viral DNA. *J. Biol. Chem.* 281, 7068–7074. doi: 10.1074/jbc.M511152200
- Shamoo, Y., and Steitz, T. A. (1999). Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell* 99, 155–166. doi: 10.1016/S0092-8674(00)81647-5
- Shereda, R. D., Kozlov, A. G., Lohman, T. M., Cox, M. M., and Keck, J. L. (2008). SSB as an organizer/mobilizer of genome maintenance complexes. *Crit. Rev. Biochem. Mol. Biol.* 43, 289–318. doi: 10.1080/10409230802341296
- Soengas, M. S. (1996). *Caracterización Estructural y Funcional de la Proteína de Unión a DNA de Cadena Sencilla del Bacteriófago  $\Phi$ 29*. Ph.D. thesis, Universidad Autónoma de Madrid.
- Soengas, M. S., Esteban, J. A., Salas, M., and Gutiérrez, C. (1994). Complex formation between phage  $\Phi$ 29 single-stranded DNA binding protein and DNA. *J. Mol. Biol.* 239, 213–226. doi: 10.1006/jmbi.1994.1364
- Soengas, M. S., Gutiérrez, C., and Salas, M. (1995). Helix-destabilizing activity of  $\Phi$ 29 single-stranded DNA binding protein: effect on the elongation rate during strand displacement DNA replication. *J. Mol. Biol.* 253, 517–529. doi: 10.1006/jmbi.1995.0570
- Soengas, M. S., Mateo, C. R., Rivas, G., Salas, M., Acuña, A. U., and Gutiérrez, C. (1997a). Structural features of  $\Phi$ 29 single-stranded DNA-binding protein. II. Global conformation of  $\Phi$ 29 single-stranded DNA-binding protein and the effects of complex formation on the protein and the single-stranded DNA. *J. Biol. Chem.* 272, 303–310.
- Soengas, M. S., Mateo, C. R., Salas, M., Acuña, A. U., and Gutiérrez, C. (1997b). Structural features of  $\Phi$ 29 single-stranded DNA-binding protein. I. Environment of tyrosines in terms of complex formation with DNA. *J. Biol. Chem.* 272, 295–302.
- Talavera, A., Salas, M., and Viñuela, E. (1972). Temperature-sensitive mutants affected in DNA synthesis in phage  $\Phi$ 29 of *Bacillus subtilis*. *Eur. J. Biochem.* 31, 367–371.
- Theobald, D. L., Mitton-Fry, R. M., and Wuttke, D. S. (2003). Nucleic acid recognition by OB-fold proteins. *Annu. Rev. Biophys. Biomol. Struct.* 32, 115–133. doi: 10.1146/annurev.biophys.32.110601.142506
- Tone, T., Takeuchi, A., and Makino, O. (2012). Single-stranded DNA binding protein Gp5 of *Bacillus subtilis* phage  $\Phi$ 29 is required for viral DNA replication in growth-temperature dependent fashion. *Biosci. Biotechnol. Biochem.* 76, 2351–2353. doi: 10.1271/bbb.120587
- Tsai, H. H., Huang, C. H., Lin, A. M., and Chen, C. W. (2008). Terminal proteins of *Streptomyces* chromosome can target DNA into eukaryotic nuclei. *Nucleic Acids Res.* 36, e62. doi: 10.1093/nar/gkm1170
- Wang, C. X., Zakharova, E., Li, J., Joyce, C. M., Wang, J., and Konigsberg, W. (2004). Pre-steady-state kinetics of RB69 DNA polymerase and its exo domain mutants: effect of pH and thiophosphoryl linkages on 3'-5' exonuclease activity. *Biochemistry* 43, 3853–3861. doi: 10.1021/bi0302292
- Wang, J., Yu, P., Lin, T. C., Konigsberg, W. H., and Steitz, T. A. (1996). Crystal structures of an NH2-terminal fragment of T4 DNA polymerase and its complexes with single-stranded DNA and with divalent metal ions. *Biochemistry* 35, 8110–8119. doi: 10.1021/bi960178r
- Warner, H. R., Johnson, L. K., and Snustad, D. P. (1980). Early events after infection of *Escherichia coli* by bacteriophage T5. III. Inhibition of uracil-DNA glycosylase activity. *J. Virol.* 33, 535–538.
- Xu, Y., Gao, S., Bruno, J. F., Luft, B. J., and Dunn, J. J. (2008). Rapid detection and identification of a pathogen's DNA using  $\Phi$ 29 DNA polymerase. *Biochem. Biophys. Res. Commun.* 375, 522–525. doi: 10.1016/j.bbrc.2008.08.082
- Yoshikawa, H., Elder, J. H., and Ito, J. (1986). Comparative studies on the small *Bacillus* bacteriophages. *J. Gen. Appl. Microbiol.* 33, 39–49. doi: 10.2323/jgam.32.39
- Yoshikawa, H., Garvey, K. J., and Ito, J. (1985). Nucleotide sequence analysis of DNA replication origins of the small bacteriophages: evolutionary relationships. *Gene* 37, 125–130. doi: 10.1016/0378-1119(85)90264-1
- Zaballos, A., and Salas, M. (1989). Functional domains in the bacteriophage  $\Phi$ 29 terminal protein for interaction with the  $\Phi$ 29 DNA polymerase and with DNA. *Nucleic Acids Res.* 17, 10353–10366. doi: 10.1093/nar/17.24.10353

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Salas, Holguera, Redrejo-Rodríguez and de Vega. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Replisome Assembly at Bacterial Chromosomes and Iteron Plasmids

Katarzyna E. Wegrzyn <sup>\*†</sup>, Marta Gross <sup>†</sup>, Urszula Uciechowska and Igor Konieczny <sup>\*</sup>

Department of Molecular and Cellular Biology, Intercollegiate Faculty of Biotechnology of University of Gdansk and Medical University of Gdansk, Gdansk, Poland

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
The University of Texas Medical  
Branch at Galveston, USA

### Reviewed by:

Ramon Diaz Orejas,  
Spanish National Research Council,  
Spain  
Jose Angel Ruiz-Masó,  
Centro de Investigaciones Biológicas  
(Consejo Superior de Investigaciones  
Científicas), Spain

### \*Correspondence:

Katarzyna E. Wegrzyn  
katarzyna.wegrzyn@biotech.ug.edu.pl  
Igor Konieczny  
igor.konieczny@biotech.ug.edu.pl

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 21 June 2016

**Accepted:** 25 July 2016

**Published:** 11 August 2016

### Citation:

Wegrzyn KE, Gross M,  
Uciechowska U and Konieczny I  
(2016) Replisome Assembly at  
Bacterial Chromosomes and Iteron  
Plasmids. *Front. Mol. Biosci.* 3:39.  
doi: 10.3389/fmolb.2016.00039

The proper initiation and occurrence of DNA synthesis depends on the formation and rearrangements of nucleoprotein complexes within the *origin* of DNA replication. In this review article, we present the current knowledge on the molecular mechanism of replication complex assembly at the *origin* of bacterial chromosome and plasmid replicon containing direct repeats (iterons) within the *origin* sequence. We describe recent findings on chromosomal and plasmid replication initiators, DnaA and Rep proteins, respectively, and their sequence-specific interactions with double- and single-stranded DNA. Also, we discuss the current understanding of the activities of DnaA and Rep proteins required for replisome assembly that is fundamental to the duplication and stability of genetic information in bacterial cells.

**Keywords:** replication initiation, DnaA, Rep, iteron plasmids, replisome assembly

## INTRODUCTION

The replication of genetic material is one of the most fundamental processes that influence the proper functioning of each living cell. The synthesis of new DNA molecule, in case of both bacterial chromosomes and plasmids, starts at a well-defined place called *origin* and can be divided into the following steps: (1) *origin* recognition by replication initiation proteins and open complex formation (2) helicase loading, activation and primer synthesis (3) replisome assembly and DNA synthesis. Although these main steps during the DNA replication process are common, when considering replication of bacterial chromosomes and iteron plasmids replicated by theta mechanism, some differences can be observed (Table 1).

A DNA replication process of chromosome and plasmid DNA starts when Origin Binding Proteins (OBP) recognize and bind specific motifs located within *origin* region. Despite the differences in structure of bacterial and plasmid initiators, DnaA and Rep proteins, respectively, they have the same function. Binding of initiators results in a modulation of nearby DNA topology and opening of double-stranded helix structure in DNA unwinding element (DUE). A single-stranded DUE region becomes a place where helicase is loaded. In the next step the replisome is assembled and holoenzyme of DNA Polymerase III can play its role during DNA synthesis.

Despite many years of research on DNA replication, new aspects of this process are still being discovered. Recently, the novel activities of replication initiator proteins have been shown. However, especially in case of plasmid DNA replication, there are many questions concerning the replication initiation and replisome assembly that still need to be answered.

**TABLE 1 | Comparison of general features of iteron plasmid and chromosomal DNA replication initiation.**

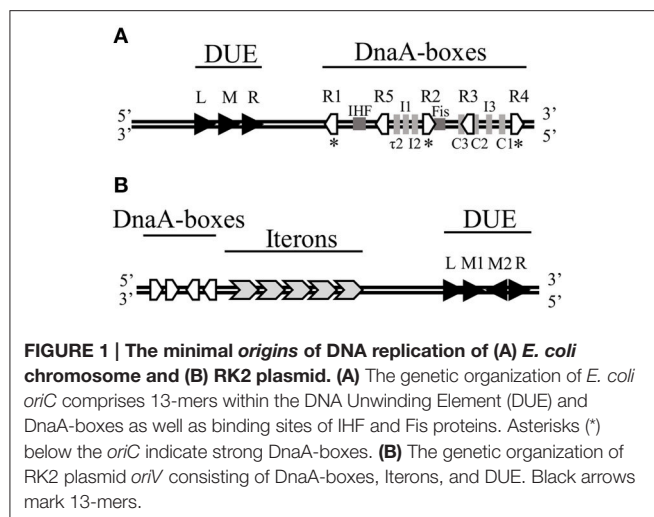
	Bacterial chromosome	Iteron plasmids
Replication initiator	DnaA	Rep, DnaA
Binding sites for initiator	Strong and weak DnaA-boxes	DnaA-boxes for DnaA Iterons for Rep Weak binding sites ?
DNA binding domain of initiator	DNA binding domain (DBD) AAA+ domain	Winged Helix domain (WH)
Nucleotide binding by initiator	+	Rep protein ? DnaA protein ?
Oligomer formation by initiator protein	+	Rep oligomers ? Rep-DnaA oligomers ?
Binding of initiator to dsDNA	+	Rep protein + DnaA protein + (in <i>Pseudomonas</i> spp. DnaA is dispensable)
Binding of initiator to ssDNA	+	Rep protein + DnaA protein ?
Assistance of architectural proteins: IHF, HU	More efficient <i>oriC</i> -dependent DNA replication	More efficient plasmid <i>origin</i> -dependent DNA replication HU is required for replication of some plasmids
Interaction of initiator with helicase DnaB	+	Rep + DnaA +

# ORIGIN RECOGNITION AND OPEN COMPLEX FORMATION BY REPLICATION INITIATION PROTEINS

## Origin Recognition and Open Complex Formation by Chromosomal Initiator at Chromosomal *Origin*

The very first step of replication initiation process is the recognition of specific motifs located within the *origin* region of DNA molecule (**Figure 1**) by replication initiation proteins (**Figure 2**, Stage I). The bacterial chromosome replication initiator DnaA protein consists of four domains, which play distinct roles (Sutton and Kaguni, 1997, **Figure 3A**). The best characterized DnaA is the *Escherichia coli* protein (*EcDnaA*), although structural data is limited only to domain I (resolved by NMR-analysis; Abe et al., 2007b) and IV (resolved in a nucleoprotein complex by crystallography; Fujikawa et al., 2003). Information concerning the structure of DnaA initiator is supplemented by structure of domains I and II of *Mycoplasma genitalium* DnaA (*MgDnaA*; Lowery et al., 2007), domains I and II of *Helicobacter pylori* DnaA (*HpDnaA*) in a complex with HobA protein (Natrajan et al., 2009), domains III and IV of *Aquifex aeolicus* DnaA (*AaDnaA*; Erzberger et al., 2002, 2006), domain III of *Thermatoga maritima* DnaA (*TmDnaA*; Ozaki et al., 2008), and domain IV of *Mycobacterium tuberculosis* (*MtDnaA*; Tsodikov and Biswas, 2011). Domain I of *EcDnaA*, located at the N-terminus of the protein, was shown to be involved in oligomerization of DnaA (Weigel et al., 1999;

Simmons et al., 2003; Abe et al., 2007a), helicase loading (Sutton et al., 1998; Seitz et al., 2000), and interaction with DiaA (Keyamura et al., 2007), HU (Chodavarapu et al., 2008a), Dps (Chodavarapu et al., 2008b), and ribosomal protein L2 (Chodavarapu et al., 2011). The interaction with DiaA homologue, HobA protein, was shown for domains I and II of *HpDnaA* (Natrajan et al., 2007, 2009; Zawilak-Pawlik et al., 2007). In *Bacillus subtilis*, domain I of DnaA (*BsDnaA*) interacts with SirA, the sporulation-related protein (Rahn-Lee et al., 2011). However, the binding partner proteins can vary among DnaA orthologs, and replication initiator from one bacterium can interact with different partners compared to other orthologs, e.g., interaction of *Thermoanaerobacter tengcongensis* DnaA with NusG protein, is not observed for *BsDnaA* (Liu et al., 2008). The second domain, forming a flexible linker, although it is not essential (Messer et al., 1999; Nozaki and Ogawa, 2008), was proposed to be involved in optimal helicase DnaB recruitment (Molt et al., 2009). The domain II, links domain I with domain III, which contains a common core structure of AAA+ proteins family members (Neuwald et al., 1999). Recent data showed that residues within this domain are engaged in interaction of DnaA (*TmDnaA*, *EcDnaA*, *AaDnaA*) with single-stranded DNA (ssDNA; Ozaki et al., 2008; Duderstadt et al., 2011). At the C-terminus of DnaA, domain IV (DNA Binding Domain, DBD) can be distinguished, which is responsible, via a helix-turn-helix motif (HTH), for interaction with double-stranded DNA (dsDNA) containing specific motifs named DnaA-boxes (Roth and Messer, 1995; Fujikawa et al., 2003). Interaction with these sequences is the very first step of the replication initiation process.

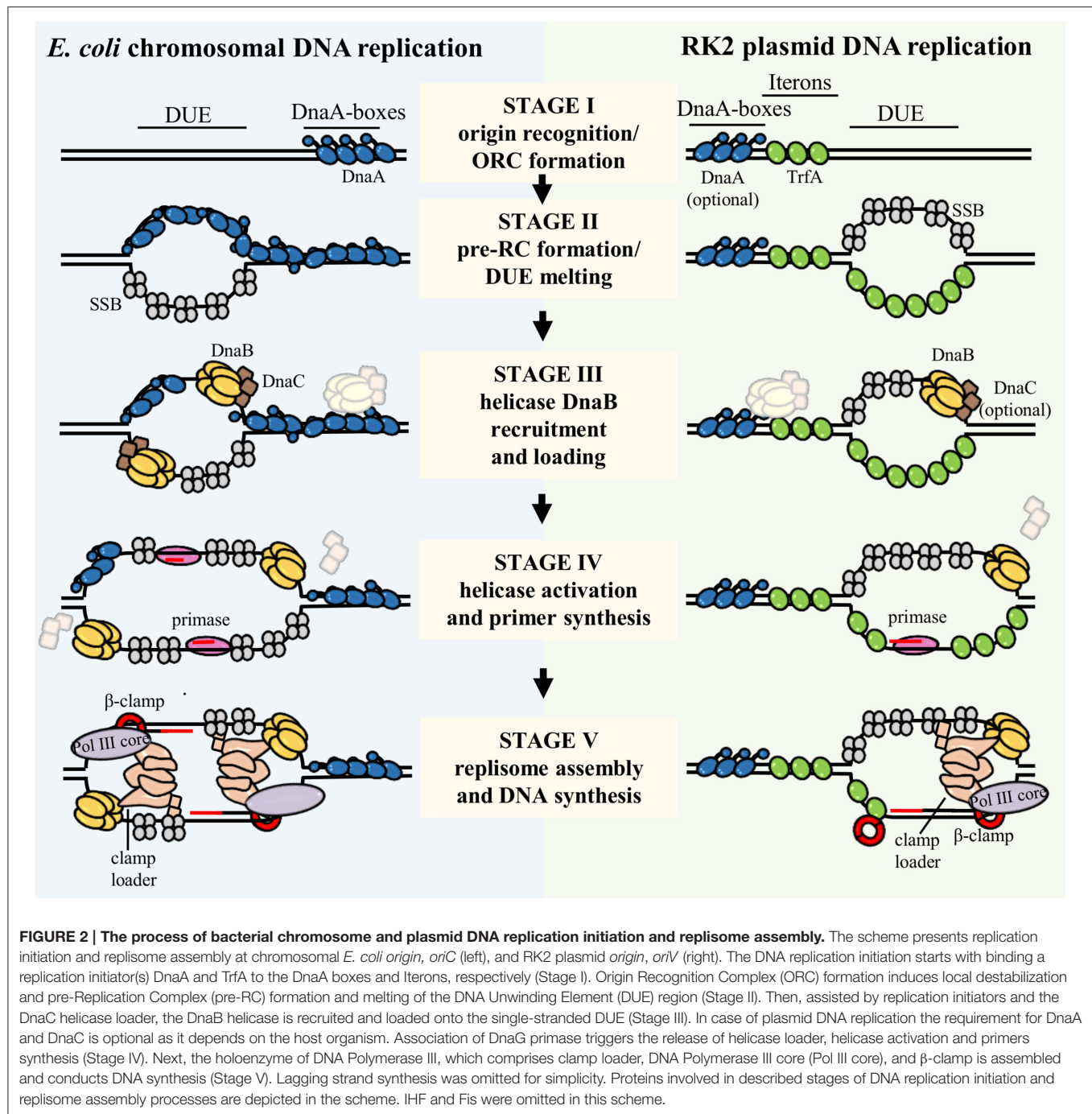


In bacterial chromosome *origin*, regions that are composed of a variable number of DnaA-boxes, can be identified (Ozaki and Katayama, 2009; Rajewska et al., 2012; Wolanski et al., 2014; Leonard and Grimwade, 2015). In the *origin* of *E. coli* chromosome (*oriC*), five 9-bp in length DnaA-boxes (R1–R5) were originally identified (Fuller et al., 1984; Matsui et al., 1985); in contrast, the *origin* of *Caulobacter crescentus* chromosome (*Cori*) possesses only two DnaA-boxes (named G-boxes; Shaheen et al., 2009). The studies with the use of *in vivo* and *in vitro* dimethylsulphate (DMS) footprinting as well as DNase I footprinting method showed that other, non-R DnaA binding sites are present in *oriC*, i.e., I (Grimwade et al., 2000; McGarry et al., 2004), C (Rozgaja et al., 2011), and  $\tau$  sites (Kawakami et al., 2005). Such non-canonical sequences recognized by bacterial initiator were also found in *oriC* of *C. crescentus* (termed W-boxes; Taylor et al., 2011). The affinity of DnaA binding to R-boxes and non-R DnaA binding sites is different. Interestingly, binding of initiator to the DnaA-boxes in *Cori* of *C. crescentus*, both G-boxes and W-boxes, is lower compared to DnaA binding to the R-boxes in *oriC* of *E. coli* (Taylor et al., 2011), which might be characteristic for bacteria with a complex regulation of development. The DnaA binding sites, bound by initiator with affinity comparable only to interaction between DnaA and weak DnaA-boxes in *E. coli* *oriC*, were found in the *origin* of *H. pylori* (Zawilak-Pawlik et al., 2007; Charbon and Løbner-Olesen, 2011). In *E. coli* *oriC* three (named R1, R2, and R4) out of five DnaA-boxes are the widely separated, high affinity DnaA-boxes. They were found to be almost constantly bound by *EcDnaA* protein (Samitt et al., 1989; Nievera et al., 2006). The occupancy of only these three sites is insufficient for spontaneous *origin* opening and it was proposed that interaction of *EcDnaA* protein at high affinity binding sites may regulate conformation of the *origin* DNA (Kaur et al., 2014). Between the peripheral R1 and R4 sites, there are two arrays of low affinity binding sites,  $\tau$ 1 R5  $\tau$ 2 I1 I2 and C3 C2 I3 C1, separated by one of high affinity—R2 (Rozgaja et al., 2011). *EcDnaA* molecules bound to the high affinity DnaA-boxes, termed bacterial Origin Recognition Complex (bORC), act as anchors and are required to assist in occupying weak

sites by the *EcDnaA* protomers (Rozgaja et al., 2011; Kaur et al., 2014), and formation of replication-active pre-replication complex (pre-RC; Figure 2, Stage II). The binding affinity to particular sequences and replication activity of *EcDnaA* protein depend on nucleotide-bound state of protein. Although ADP-*EcDnaA* binds the high affinity DnaA-boxes and also R5 and C1 low affinity ones, the ATP-*EcDnaA* form is thought to be the replication-active one (Sekimizu et al., 1987; Leonard and Grimwade, 2011). ATP-*EcDnaA* form of initiator binds efficiently both high and low affinity binding sites (McGarry et al., 2004; Kawakami et al., 2005). Based on molecular docking, binding of ATP, instead of ADP, is presumed to cause changes in the *EcDnaA* protein conformation, thus leading to the formation of large oligomeric complex within the *origin* region (Saxena et al., 2015). The crystallographic data, when nonhydrolyzable ATP analog AMP-PCP was used, showed the formation of open-ended, right-handed helical filament of *AaDnaA* (Erzberger et al., 2006). Based on biochemical and genetic approaches it was found that there is an interaction between domain III (AAA + domain) of one DnaA (*EcDnaA* or *AaDnaA*) molecule and domain IV (DBD domain) of partner subunit (Duderstadt et al., 2010). It was proposed that during pORC and pre-RC complexes formation of the DBD domain is extended and the HTH motif is exposed, which results in the efficient binding of high and low affinity binding sites (Duderstadt et al., 2010). Occupation of the *EcDnaA* binding sites was shown to be sequential and polarized and DnaA protomers are released preferentially from the peripheral high affinity R1 and R4 boxes, through arrays of low affinity binding sites to the middle high affinity one—R2 (Rozgaja et al., 2011). The formation of DnaA oligomer within the *oriC* results in DNA destabilization in the DUE region (Speck and Messer, 2001; McGarry et al., 2004; Leonard and Grimwade, 2005, 2011; Duderstadt et al., 2010). Although two arrays of low affinity binding sites separated by high affinity sequences are occupied by *EcDnaA* protomers for efficient double-stranded DNA opening, binding of *EcDnaA* to a part of *origin* (containing only R1 high affinity box and  $\tau$ 1 R5  $\tau$ 2 I1 I2 low affinity binding sites array) was shown to be active in DUE unwinding (Ozaki and Katayama, 2012). It was proposed that distinct DnaA multimers are formed on the left half (containing binding sites from R1 to I2) and the right half (containing binding sites from R2 to R4) of *oriC* (Ozaki and Katayama, 2012; Ozaki et al., 2012a).

The DUE melting is the consequence of DnaA binding to arrays of DnaA-boxes (Figures 1A, 2, Stage II). The location of particular binding sites suggests that DnaA, bound to sequences of the high affinity DnaA-boxes (R1, R2, R4), could cause the bending of DNA molecule via interaction through domain I of already bound three protomers (Rozgaja et al., 2011; Kaur et al., 2014; Leonard and Grimwade, 2015). The model of constrained loop formed by *EcDnaA* bound to the high affinity binding sites was proposed (Kaur et al., 2014). The bending of *oriC* containing DNA molecule is supported by accessory histone-like proteins HU and integration host factor (IHF). A binding site for IHF was found within the *oriC* region (Polaczek, 1990) and it was shown that IHF can enhance the unwinding of DNA by DnaA (Hwang and Kornberg, 1992; Ryan et al., 2002). It was demonstrated that HU has the same effect on DUE destabilization (Hwang and

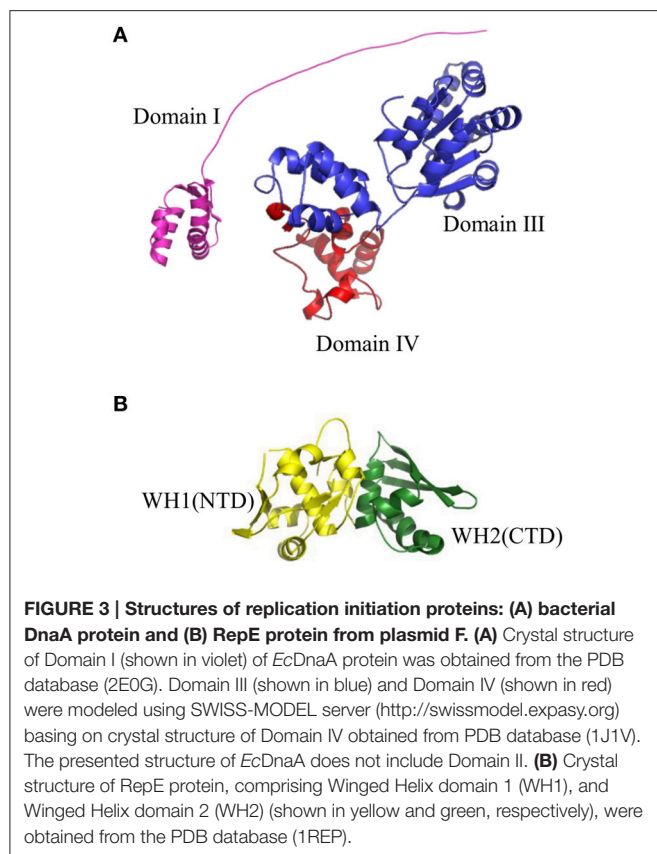




Kornberg, 1992), although its mechanism of action is different (Ryan et al., 2002). Data obtained with ELISA (Enzyme Linked Immunosorbent Assay) showed that HU interacts with domain I of *EcDnaA*, which was proposed as an interaction which stabilizes the DnaA oligomer (Chodavarapu et al., 2008a). The Fis protein, identified originally as factor for inversion stimulation in site-specific DNA recombination, was also shown to have an influence on DNA unwinding (Wold et al., 1996). Specific binding sites for Fis were identified in *oriC* (Gille et al., 1991). Although Fis, in

contrast to IHF, negatively regulates DNA replication initiation, when the *origin* lacks some DnaA binding sites resulting in altered non-functional conformation of *origin*, both Fis and IHF can work together to correct these alterations (Kaur et al., 2014). This joint action is achieved by inducing bends in *oriC* and establishing functional *origin* conformation (Kaur et al., 2014).

The formation of DnaA oligomer with synergistic action of architectural proteins can introduce torsional strain into DUE,



facilitating the melting of the double-stranded DNA helix. The binding of DnaA to DUE region was also thought to introduce DNA melting, and ATP-DnaA-boxes were distinguished within the *oriC* DUE sequence (Speck and Messer, 2001). Recent studies showed direct binding of *EcDnaA* and *AaDnaA* protein to formed single-stranded DNA within the DUE (Ozaki et al., 2008; Duderstadt et al., 2010, 2011; Cheng et al., 2015). Studies with DnaA mutants (Ozaki et al., 2008; Duderstadt et al., 2010), as well as crystallography (Duderstadt et al., 2011), showed that this interaction occurs through residues located within the AAA+ domain III of bacterial initiator. The *AaDnaA* protomers form a helical filament on ssDNA (Duderstadt et al., 2011), however, it differs from the filament formed on the dsDNA (Erzberger et al., 2006; Duderstadt et al., 2010). It was proposed that protomers in this oligomer are more compact when compared to the extended DnaA molecules in dsDNA-DnaA complex (Duderstadt et al., 2010, 2011). The binding of ssDNA concerns just one T-rich strand of DUE and depends on sequence of 13-nucleotide sequences, which can be distinguished within the DUE. In *oriC* three 13-mers are present (Bramhill and Kornberg, 1988a) and the binding of *EcDnaA* occurs at least two 13-mers. *EcDnaA* does not form a complex with ssDNA containing just one 13-mer (Ozaki et al., 2008). Formation of this nucleoprotein complex is achieved only by ATP-DnaA protein (Ozaki et al., 2008) and one *AaDnaA* protomer binds three nucleotides of ssDNA (Duderstadt et al., 2011; Cheng et al., 2015). Studies with the use of single-molecule fluorescence assays showed that

the formation of this nucleoprotein complex is highly dynamic and that *AaDnaA* molecules assemble on ssDNA in the 3' to 5' direction (Cheng et al., 2015). The presence of dsDNA region containing DnaA-boxes, adjacent to ssDNA DUE, stabilizes the DnaA (*EcDnaA* and *AaDnaA*) filament on ssDNA (Ozaki and Katayama, 2012; Cheng et al., 2015). Recently published data revealed presence of a new origin element, termed DnaA-trio, composed of repeated trinucleotide motif that stabilizes DnaA filaments on the ssDNA (Richardson et al., 2016). What is important, binding single strand of DUE region is required for *origin* activity (Ozaki et al., 2008, 2012a,b; Duderstadt et al., 2011).

### Origin Recognition and Open Complex Formation by Plasmid Initiator at *Origin* of Iteron Plasmids

Similarly as during bacterial chromosome replication, the first step in open complex formation in many theta-replicating plasmids, especially in iteron-containing plasmids, is the binding of plasmid replication initiator, Rep protein, to specific sequences within *origin* region (Figure 2, Stage I). Rep proteins are structurally different from bacterial DnaA protein and consist of winged-helix (WH) domains (Figure 3B, Komori et al., 1999; Díaz-López et al., 2003; Sharma et al., 2004; Swan et al., 2006; Nakamura et al., 2007a,b; Pierechod et al., 2009). The crystal structures of nucleoprotein complexes of  $\pi$  protein from plasmid R6K (Swan et al., 2006), RepE protein from plasmid F (Komori et al., 1999; Nakamura et al., 2007b), and a DNA binding domain of Rep protein from ColE2-P9 plasmid (Itou et al., 2015) as well as N-terminal domain of RepA protein from plasmid pPS10 (Giraldo et al., 2003) were obtained. Furthermore, homological models for plasmid Rep proteins: RepA from P1 (Sharma et al., 2004), RepA from pSC101 (Sharma et al., 2004), and TrfA from RK2 (Pierechod et al., 2009) were shown. Plasmid Reps are composed of two WH domains, of which one is responsible for oligomerization and the role of a second one is the protein's interaction with DNA (Giraldo et al., 1998; Nakamura et al., 2004; Pierechod et al., 2009). Plasmid replication initiators are present as dimers in solution, however, an exception is known i.e., RepE protein from pAM $\beta$ 1 plasmid is present as a monomer (Le Chatelier et al., 2001). Although the Rep dimers interact with DNA (Filutowicz et al., 1985; Ingmer et al., 1995; Komori et al., 1999), they are replication-active in the monomeric form (Kawasaki et al., 1990; Wickner et al., 1992; Sozhamannan and Chattoraj, 1993; Konieczny and Helinski, 1997). Conformational activation of plasmid replication initiators is carried out by chaperon proteins (Kawasaki et al., 1990; Wickner et al., 1992, 1994; Sozhamannan and Chattoraj, 1993; Konieczny and Helinski, 1997). In contrast to bacterial replication initiator DnaA, the domain responsible for binding of nucleotide was not identified in Reps' structures. There is also no evidence showing if Rep proteins can form helical filaments on DNA similar to that formed by the *AaDnaA* protein. For some Reps, e.g., TrfA protein from RK2 plasmid, two forms of protein, different in length, occur: the shorter 33 kDa (TrfA-33) and longer 44 kDa (TrfA-44). There are different requirements for each particular form

depending on the host bacterium. In *E. coli* both forms of TrfA can initiate the plasmid replication, whereas in *Pseudomonas aeruginosa* only the longer form is active (Caspi et al., 2001; Jiang et al., 2003; Konieczny, 2003; Yano et al., 2013, 2016).

During plasmid replication initiation, monomers of Reps bind to specific repeated sequences, named iterons, present within *origin* region (Figures 1B, 2, Stage I). The number of iterons varies among plasmid *origins*, from two iterons in plasmids ColE2 and ColE3 (Yasueda et al., 1989), three iterons in pSC101 (Churchward et al., 1984), and some plasmids from IncQ incompatibility group (Loftie-Eaton and Rawlings, 2012), four iterons in *origin* of plasmid F and pPS10, up to five (*origin* of plasmids RK2 and P1) or even seven such sequences in *oriY* of plasmid R6K (Rajewska et al., 2012). Iterons are short sequences, in length ranging from 17-bp in RK2 plasmid (Stalker et al., 1981), 19-bp in plasmids F (Murotsu et al., 1981), and P1 (Abeles et al., 1984), to 22-bp in R6K (Filutowicz et al., 1987), and pPS10 (Nieto et al., 1992). But in some plasmids the iteron sequences which are present in one *origin* can differ in length and apart from short sequences, significantly longer iterons [up to even 76-bp in plasmid R478 from IncHI2 incompatibility group (Page et al., 2001)] can be present. The binding of Rep proteins to iterons is sequence-specific and mutations in these motifs disrupt binding of plasmid initiation protein. Changes in a sequence of iterons abolished binding of  $\pi$  protein within the *oriY* of plasmid R6K and thus replication activity *in vivo* (McEachern et al., 1985). Negative effects on replication was also observed for mutants in a sequence of P1 plasmid iterons (Brendler et al., 1997). The sequences separating particular iterons are also important for Rep nucleoprotein complexes formation and proper replication activity of *origin*. It was shown in case of the RK2 plasmid that *in vitro* the TrfA protein has a high preference for binding to DNA containing at least two out of five binding sites, when compared to the formation of nucleoprotein complex with DNA containing just one iteron (Perri et al., 1991). The requirement for the presence of more than just one iteron sequence for TrfA binding was also shown *in vivo* (Perri and Helinski, 1993). Rep proteins bind to iterons in a cooperative manner (Perri and Helinski, 1993; Xia et al., 1993; Bowers et al., 2007) and the cooperativity of binding depends on the spatial location of iterons, since separation of two iterons by a half helical turn abolished cooperativity (Bowers et al., 2007). These results suggest the possibility of formation of higher order nucleoprotein structure on plasmid iterons bound by Reps. It was shown that WH domains of Reps contact three nucleotides in DNA. In  $\pi$  protein from R6K plasmid, WH1 domain contacts wGwCnT motif, and WH2 domain contacts GAG sequence (Swan et al., 2006). Similarly, the WH2 domain of RepE monomer also contacts three nucleotides of top (GTG sequence) and three nucleotide of bottom strand (GtCA sequence) of double-stranded molecule containing iteron sequence (Nakamura et al., 2007b). However, unlike for the bacterial DnaA protein, to date there are no evidence showing that strong and weak binding sites for Reps are present within plasmid *origins*. There were just predictions of potential binding sites, other than iterons, for  $\pi$  protein in R6K plasmid and suggestions on potential role of such sites (Rakowski and Filutowicz, 2013). Certainly like DnaA, Rep proteins can

bind within single-stranded region of melted DUE, and this binding is sequence-specific, since binding concerns a particular strand. Nucleoprotein complexes formation with the ssDNA DUE was detected for TrfA (bound with A-rich strand) and RepE (bound with T-rich strand) proteins (Wegrzyn et al., 2014). Within the DUE of plasmid *origins*, repeated sequence, similar to 13-mers distinguishable in *oriC*, can be found. There are four 13-nucleotide sequences in plasmid RK2 DUE region (Doran et al., 1998) and all of them are required for TrfA-ssDNA DUE complex formation. Lack of even one 13-mer hinders plasmid replication (Wegrzyn et al., 2014). Also, even a point mutation within this region affects plasmid replication since the lack of DUE melting was observed for some of the changed sequences (Kowalczyk et al., 2005; Rajewska et al., 2008).

The Rep protein encoded by plasmids, can be accompanied by host DnaA initiator during open complex formation and DUE melting within a plasmid *origin* (Figure 2, Stage II). DnaA binding sites have been found in replication *origin* of many plasmids including plasmids P1 (Abeles et al., 1984, 1990; Abeles, 1986), F (Kline et al., 1986; Murakami et al., 1987; Kawasaki et al., 1996), RK2 (Doran et al., 1998; Caspi et al., 2000), pSC101 (Sutton and Kaguni, 1995). The number of DnaA-box sequences differs among plasmid *origins*, the position and orientation of these binding sites are as important as position and orientation of the iterons (Doran et al., 1998, 1999). The inversion of one out of four DnaA boxes in *origin* of RK2 plasmid abolished plasmid DNA replication, despite the fact that three remaining DnaA boxes were bound by the host initiator (Doran et al., 1999). Although the DnaA protein is not required for replication initiation for some plasmids, e.g., R1, binding of DnaA increased the plasmid replication efficiency (Bernander et al., 1991, 1992) and mutations within a binding site for DnaA decreased the R1 plasmid replication (Ortega-Jiménez et al., 1992). In bacteria, ATP-DnaA form is essential for chromosomal DNA replication (Sekimizu et al., 1987; Leonard and Grimwade, 2005, 2011). Interestingly, studies with ATP-binding mutant of DnaA, which was inactive in *oriC* replication, showed that bacterial initiator lacking an ability to bind a nucleotide was effective in open complex formation within plasmid R6K *oriY* (Lu et al., 1998). Also in the presence of ATP $\gamma$ S, a non-hydrolyzable analog of ATP, the pattern of bands in KMnO<sub>4</sub> footprinting assay with DnaA and TrfA proteins and plasmid RK2 DNA showed no significant differences, when compared to opening reaction containing ATP (Konieczny et al., 1997). Thus, the DnaA is suspected to play a different role in plasmid replication initiation, compared to its role in chromosome replication. A direct interaction between plasmid and host replication initiators was shown (Lu et al., 1998; Maestro et al., 2003) and the interaction was detected in the N-terminus of  $\pi$  (between 1 and 116 aa) protein of R6K plasmid (Lu et al., 1998) and RepA protein of pSC101 (Sharma et al., 2001) and domain I and IV of host initiator (Sharma et al., 2001). The mutations in RepA protein from pPS10 plasmid were introduced, which enhanced the interaction of RepA with DnaA protein and resulted in changes in host range of pPS10 plasmid (Maestro et al., 2003).

Similarly to bacterial chromosome replication initiation, the binding of DnaA protein to DnaA-boxes within plasmid *origins*



can be enhanced by the presence of architectural proteins IHF, and HU (Shah et al., 1995; Fekete et al., 2006). The binding of IHF to its binding site in *oriY* region significantly enhanced binding of bacterial DnaA to R6K plasmid *origin* (Lu et al., 1998). In pSC101 plasmid binding IHF to its cognate binding site is required for plasmid replication initiation and mutations within this sequence disrupts plasmid replication (Stenzel et al., 1987). For plasmid P1 the binding of IHF to its site, located downstream of one out of two arrays of DnaA-boxes (the second array is located upstream of DUE) is required only when the nearby DUE array of DnaA-boxes is not active and the second DnaA-boxes array serves as a secondary origin compensating the function of the first one (Fekete et al., 2006). The P1-mini derivative was just slightly unstable in IHF *E. coli* mutant (Ogura et al., 1990). The mutations in gene for IHF protein did not affect plasmids F (Ogura et al., 1990) and RK2 (Shah et al., 1995) replication. In contrast, the lack of HU protein *in vitro* results in significant decrease in mini-F plasmid DNA synthesis (Zzaman et al., 2004) and *in vivo* KMnO<sub>4</sub> reactivity of P1 plasmid *origin* (Park et al., 1998) as well as abolishment of plasmid F replication *in vivo* (Ogura et al., 1990). During plasmid RK2 replication initiation, HU could functionally replace DnaA protein, although it could not enhance DUE melting as efficiently as DnaA (Konieczny et al., 1997). It was proposed that one of the DnaA functions could be the stabilization of *origin* melting induced by Rep protein. The other DnaA role during replication initiation is its function in helicase loading. Interestingly, for some plasmids, e.g., RK2, DnaA assists Rep during plasmid replication initiation only in particular hosts, while in others DnaA is dispensable [DnaA *P. aeruginosa* is dispensable for RK2 plasmid replication initiation, but required in *E. coli* (Caspi et al., 2001; Konieczny, 2003)].

## HELICASE LOADING, ACTIVATION, AND DNA UNWINDING

In bacteria the loading of DnaB helicase onto ssDNA of DUE is achieved by the action of replication initiation protein, DnaA, as well as the helicase loading factor, DnaC protein (Figure 2, Stage III). DnaB helicase is a two-tiered ring-shaped hexamer (Bailey et al., 2007b; Wang et al., 2008; Lo et al., 2009). Each monomer consists of N-terminal and C-terminal domain connected via linker helix (LH) region (Miron et al., 1992; Ingmer and Cohen, 1993; Komori et al., 1999). The N-terminal domain of helicase's monomers were shown to interact with ssDNA (observed in a crystal structure of *Geobacillus kaustophilus* helicase in a complex with ssDNA; Lo et al., 2009) which stabilizes the hexameric structure of DnaB (Biswas et al., 1994). The C-terminal domain, that contains RecA-like fold, is responsible for ATP binding and hydrolysis, interaction with DNA (Bailey et al., 2007a), and binding of DnaC loader factor (Lu et al., 1996). The helicase is positioned onto the ssDNA DUE in a single orientation with respect to the polarity of the sugar-phosphate backbone of DNA and the nucleic acid, bound primarily to one DnaB monomer (Jezewska et al., 1998a,b), passes through the cross-channel of helicase hexamer (Jezewska et al., 1998a). The hexamer of DnaB, when no ATP hydrolysis occurs, is bound to 20 ( $\pm$ 3) nucleotides (Jezewska et al., 1996).

The binding of nucleotide as well as particular partner protein and DNA promotes helicase to adopt specific conformation. The X-ray crystal structure of *A. aeolicus* helicase revealed large conformational rearrangements, observed in N-terminal domain and the presence of at least two highly-distinct conformations: widened with broad central channel and a highly-constricted with a narrow pore (Strycharska et al., 2013). These conformations were also observed for *E. coli* DnaB, when analyzed in solution with the use of small-angle X-ray scattering (SAXS; Strycharska et al., 2013). Structural analysis with the use of negative-stain electron microscopy (EM) and SAXS of DnaB protein in complex with its loader, DnaC, showed that the hexamer of helicase interacts with helical arrangement of six DnaC monomers (Kobori and Kornberg, 1982; Arias-Palomo et al., 2013). However, it was argued that the active form of the DnaB-DnaC complex exists in 6:3 stoichiometry, which was studied by quantitative analysis of pre-priming complex (Makowska-Grzyska and Kaguni, 2010). Furthermore, the imbalance in level of DnaB and DnaC was shown to impair DNA replication (Allen and Kornberg, 1991; Brüning et al., 2016).

The concept of DnaC as a protein that loads DnaB helicase onto ssDNA of DUE, has been early established (Wickner and Hurwitz, 1975; Funnell et al., 1987; Bell and Kaguni, 2013). To further explain its exact function, the following models have been proposed: (1) DnaC breaks the helicase ring (Davey and O'Donnell, 2003; Arias-Palomo et al., 2013), (2) DnaC traps DnaB helicase as an open ring (Chodavarapu et al., 2016). Those hypotheses were tested by the SAXS method and deuterium exchange coupled to mass spectrometry, respectively. The ATPase activity of DnaC, a member of AAA+ proteins family, is not required for helicase hexamer opening and its loading by DnaC, hence the DnaB-binding domain of loader is sufficient for this process (Arias-Palomo et al., 2013). Yet the ATP hydrolysis by DnaC was proposed to occur during DnaB helicase activation, which results in DNA unwinding (Felczak et al., 2016).

Regarding the DnaC key contribution to helicase loading and activation in *E. coli*, it is particularly interesting to discuss replicons that are independent of helicase loader. The helicase loaders were identified only in few species and it is possible that in some bacteria the yet unidentified helicase loaders are present. The lack of DnaC orthologs can also arise from ability of self-loading by helicase (Costa et al., 2013) or it is possible that another protein of already assigned role, substitutes the DnaC function. Those hypotheses can be supported by complementation of *dnaC* temperature-sensitive mutant of *E. coli* by helicase from *H. pylori* (Soni et al., 2003). The dispensability for helicase loader was also shown during RK2 plasmid replication in *Pseudomonas* species (Jiang et al., 2003). In *Pseudomonas* sp. the helicase loading at plasmid RK2 *origin* is performed by the longer form of plasmid Rep protein, TrfA-44, which interacts with *Pseudomonas* helicase (Caspi et al., 2001; Jiang et al., 2003; Zhong et al., 2003). The shorter form of this plasmid initiator, TrfA-33, is not sufficient for helicase loading in *P. aeruginosa*. In *Pseudomonas putida* TrfA-33 can load helicase but only in the presence of DnaA (Caspi et al., 2001; Jiang et al., 2003). On the contrary, the DnaC helicase loader, together with DnaA, and Rep protein (either short or long form), is absolutely



required for helicase loading at plasmid RK2 *origin* in *E. coli* (Caspi et al., 2001). It was shown that via interaction of DnaA with DnaBC, the helicase is first localized in DnaA-boxes and then via DnaA-DnaB and Rep-DnaB interactions translocated to ssDNA DUE (Pacek et al., 2001; Rajewska et al., 2008). Probably the Rep-DnaA interaction is also important in these processes. Apart from the proper protein-protein interaction, an efficiency of helicase translocation from DnaA-box position to DUE depends on the sequence of DUE region. It was shown via electron microscopy and *in vitro* experiments that even point mutations within the DUE of RK2 plasmid *origin* results in a decrease in helicase translocation and thus helicase DNA unwinding activity (Rajewska et al., 2008).

It was proposed that, upon DnaB-DnaC binding to ssDNA, DnaC dissociates, thus allowing DnaB to unwind double helix, and further to bind DnaG primase (Wahle et al., 1989, **Figure 2**, Stage IV). However, Makowska-Grzyska and Kaguni demonstrated, by performing molecular filtration of pre-priming complex at *E. coli oriC*, that the DnaG primase binds DnaB, synthesizes primer and in consequence, induces the release of DnaC from DnaB (Makowska-Grzyska and Kaguni, 2010). In *E. coli*, in further steps DnaG primase is associated with DnaB helicase and synthesizes primers on lagging strand (McHenry, 2011). Plasmid ColE2-P9 does not require DnaG primase in replication initiation (Takechi et al., 1995). Itoh group demonstrated that ColE2 *origin* and Rep protein as well as *E. coli* host DNA Polymerase I and SSB are sufficient for *in vitro* DNA synthesis (Itoh and Horii, 1989). Further studies revealed that the ColE2-Rep protein has joined functions, i.e., replication initiator and plasmid-specific primase (Takechi and Itoh, 1995).

Once activated, DnaB unwinds one nucleotide per one catalytic step in ATP-dependent manner (Lohman and Bjornson, 1996, **Figure 2**, Stage IV). It was shown that at 25°C the DnaB unwinds around 291 bp per second (Galletto et al., 2004) and it moves from 5' to 3' direction along the ssDNA (LeBowitz and McMacken, 1986). Because the replication of bacterial chromosome is bidirectional two helicases are loaded: one is loaded by DnaC on the top strand invaded by DnaA molecules and the other on the bottom strand. It was proposed that the helicase delivery to ssDNA DUE bottom A-rich strand occurs by direct interaction between DnaB and DnaA proteins (Mott et al., 2008; Soultanas, 2012). The Phe-46 of DnaA was shown to be important for this interaction (Keyamura et al., 2009). The order of helicase loading to a particular strand of DUE is not random but defined; first helicase is loaded onto the bottom/lower strand then the second onto the top/upper one (Weigel and Seitz, 2002). Such order of helicase loading probably supplies head-to-head orientation of unwound region of *oriC* and prevents back-to-back loading of the helicase. The basal level of DnaB activity in *oriC* is achieved when DnaA forms an oligomer in ssDNA DUE and dsDNA containing DnaA-boxes from R1 to I2 (called DAR-DF and DAR-LL). For the full activity of helicase the formation of DnaA filament on other DnaA-boxes (from R2 to R4; called DAR-RL and DAR-RE) is needed (Ozaki and Katayama, 2012).

The interaction between plasmid initiator Rep and helicase is an important factor for helicase activity on plasmid *origin* (**Figure 2**, Stage IV). It was shown for *E. coli* F plasmid that its initiator, RepE protein, cannot form a stable complex with

*Pseudomonas* helicase and thus it does not replicate efficiently in *Pseudomonas* cells (Zhong et al., 2005). Interaction between plasmid Rep and host DnaB was also detected via ELISA and protein affinity chromatography for  $\pi$  protein of R6K (Ratnakar et al., 1996) and mutations within  $\pi$  were identified which decreased helicase binding and resulted in impaired plasmid DNA replication (Swan et al., 2006). A similar effect was observed for mutants of RepA protein from plasmid pSC101, invalid in the interaction with helicase (Datta et al., 1999). Although the Rep-DnaB interaction is required for helicase loading, the right balance in the strength of the interaction must be maintained. It was shown that too tight binding of Rep to DnaB is undesirable and the mutations within Rep, acquired by adaptation under antibiotic selection that decreased binding to helicase, result in the decrease in fitness cost and increase in plasmid copy number (Yano et al., 2016).

## REPLISOME ASSEMBLY

Once DnaB helicase is loaded, DNA is unwound and primer is synthesized, the contribution of replication initiators becomes more enigmatic. Subsequent stages of DNA replication require building the replisome, i.e., the multiprotein replication machinery that synthesizes DNA (O'Donnell et al., 2013, **Figure 2**, Stage V). The replisome in bacteria is composed of DnaB helicase, DnaG primase, single-stranded DNA binding protein (SSB), and the holoenzyme of DNA Polymerase III (hPol III) (divided in three subcomplexes: Pol III core, clamp loader and  $\beta$ -clamp; O'Donnell, 2006). Reyes-Lamothe et al., suggested that both DnaA replication initiator and DnaC helicase loader are crucial for replisome assembly in *E. coli* (Reyes-Lamothe et al., 2008). This conclusion was drawn from studies that tracked the replisome components in living cells during the stages of DNA replication. However, it does not exclude the possibility that the role of replication initiator is limited to DUE destabilization and helicase loading, hence, indirect effects may be observed. Most studies regarding the mechanism of replisome assembly are performed using simplified experimental setup, e.g., primed ssDNA and replisome components, where replication initiators are omitted (Yuzhakov et al., 1996, 1999; Downey and McHenry, 2010; Cho et al., 2014).

### Clamp Loader and Its Activity

Following the primer synthesis, clamp loader complex loads the ring-shaped  $\beta$ -clamp (discussed below in details), that encircles dsDNA, tethers DNA polymerase, and slides along dsDNA, thus significantly increasing speed (up to 100-fold), and processivity (up to 5000-fold) of DNA replication (Kelch et al., 2012). The *E. coli* clamp loader is composed of  $\gamma$ ,  $\tau$ ,  $\delta$ ,  $\delta'$ ,  $\chi$ , and  $\psi$  subunits, albeit only  $\gamma$ ,  $\delta$ ,  $\delta'$  are crucial for  $\beta$ -clamp loading (reviewed in details in Kelch, 2016). The  $\gamma$  subunit is a truncated version of  $\tau$  subunit, encoded by *dnaX* gene, and arises from translational frameshift (Flower and McHenry, 1990). Both  $\gamma$  and  $\tau$  subunits have AAA<sup>+</sup> domain, however,  $\gamma$  subunit lacks  $\tau$  subunit domain responsible for DnaB helicase and Pol III core binding (Tsuchihashi and Kornberg, 1989; O'Donnell and Studwell, 1990; Flowers et al., 2003). Before clamp loader binds DNA, it adopts appropriate, ATP-driven conformational state

that increases its affinity for the  $\beta$ -clamp (Podobnik et al., 2003). It is under debate whether the ring structure of  $\beta$ -clamp is actively opened or captured in open conformation. The T4 bacteriophage trimeric clamp is the least stable sliding clamp and it was found to dissociate from DNA by monomerization, thus no force in opening of the ring is required (Soumilion et al., 1998). The dimeric clamps (bacterial, e.g., *E. coli*) are regarded as stable, hence more active ring-opening mechanism is expected to be in demand. On the basis of a crystal structure of single subunit of *E. coli* clamp loader (namely  $\delta$  subunit) in complex with  $\beta$ -clamp, it was proposed that  $\delta$  subunit is a molecular wrench, that induces rearrangements of  $\beta$ -clamp at dimerization interface, albeit without ATP hydrolysis (Jeruzalmi et al., 2001). With the use of real-time fluorescence-based clamp binding and opening assays, it was shown that clamp loader binds closed  $\beta$ -clamp in solution, prior to  $\beta$ -clamp opening (Paschall et al., 2011). Yet, the deuterium exchange coupled to Mass Spectrometry experiments revealed that most sliding clamps are dynamic at their monomers' interfaces (Fang et al., 2011, 2014). Therefore, it is also probable that  $\beta$ -clamp is trapped in an open conformation by clamp loader.

The crystal structure of clamp loader complex was solved from T4 bacteriophage (Kelch et al., 2011), *E. coli* (Simonetta et al., 2009), and its eukaryotic homolog, Replication Factor C (RFC), from *Saccharomyces cerevisiae* (Bowman et al., 2004). Each of the clamp loader complex reveals pentameric structure. Since AAA + ATPases usually adopt circular hexamers, it was proposed that sixth subunit was lost during the evolution (Indiani and O'Donnell, 2006). Indeed, the gap between the first and fifth clamp loader subunits is favorable, because it provides the mechanism of specific accommodation of the primer-template junction structure (Kelch, 2016). It was suggested that clamp loader recognizes minor groove and thus it binds at the 3' primer-template junction specifically. However, the crystal structure of the clamp loader:DNA complex revealed that clamp loader contacts template DNA exclusively (Bowman et al., 2005; Simonetta et al., 2009). Despite the fact that the DNA synthesis may be initiated only from 3' OH primer end, the clamp loader can assemble *in vitro* at either 3' or 5' primer terminus forming a stable complex (Park and O'Donnell, 2009). While clamp loader binds only DNA template,  $\beta$ -clamp interacts with both RNA primer, and DNA template within the RNA-DNA hybrid and it was shown that the  $\beta$ -clamp distinguishes between the 5' and 3' primer end (Park and O'Donnell, 2009). Consistently, it was demonstrated that SSB hampers clamp loading on the 5' end of primer (Hayner et al., 2014). The ATPase activity of the clamp loader is lower when it is assembled at the 5' terminus, comparing to the ATPase activity of clamp loader located at 3' terminus (Park and O'Donnell, 2009). ATP hydrolysis triggers  $\beta$ -clamp closing on DNA and the release of clamp loader from  $\beta$ -clamp:DNA nucleoprotein complex (Pietroni and von Hippel, 2008). Thereby, the 3' primed end loading preference also arises from the higher rate of clamp closure and clamp loader dissociation (Park and O'Donnell, 2009). The  $\beta$ -clamp must be closed in the ATP hydrolysis-dependent manner, to release clamp loader (Hayner et al., 2014). Clamp loader must free the  $\beta$ -clamp to allow the Pol III core to bind, since they accommodate the

same binding site within the  $\beta$ -clamp, namely the hydrophobic cleft.

## $\beta$ -Clamp—Hub for Protein Interactions

$\beta$ -clamp crystal structures were obtained from various organisms i.e., *E. coli* (Oakley et al., 2003; Burnouf et al., 2004), *P. aeruginosa* (Wolff et al., 2014), *Streptococcus pyogenes* (Argiriadi et al., 2006), *M. tuberculosis* (Gui et al., 2011; Kukshal et al., 2012; Wolff et al., 2014), *B. subtilis* (Wolff et al., 2014), *T. maritima* (structure 1VPK), *Eubacterium rectale* (structure 3T0P), *Streptococcus pneumoniae* (Argiriadi et al., 2006). The crystal structures of  $\beta$ -clamp homologs—Proliferating Cell Nuclear Antigen (PCNA)—are also available from *Eukaryotes* and *Archea* (to name a few: *Homo sapiens* (Punchihewa et al., 2012), *S. cerevisiae* (Krishna et al., 1994), *Sulfolobus solfataricus* (Williams et al., 2006). All of them adopt ring shaped homodimer (e.g., *E. coli*) or homotrimer (human PCNA, *Pyrococcus furiosus* PCNA), albeit the exception is PCNA of *Archea*, *S. solfataricus*, which exists as a heterotrimer (Dionne et al., 2003).  $\beta$ -clamp monomers bind in a head to tail manner (Kelman and O'Donnell, 1995). The  $\beta$ -clamp and PCNA structure is conserved among all kingdoms of life, in contrast to amino acid sequence (Jeruzalmi et al., 2001). However, the amino acid sequence of region termed hydrophobic cleft was found to be highly conserved (Jeruzalmi et al., 2001). The hydrophobic cleft is a site for interaction with  $\beta$ -clamp binding partners (Jeruzalmi et al., 2001).  $\beta$ -clamp forms a protein interaction hub and serves as a platform for multiple protein interactions crucial in various cellular processes, i.e., DNA elongation in every living organism (Hedglin et al., 2013), regulation of DNA replication in *E. coli*, *B. subtilis*, *C. crescentus* (Katayama et al., 2010), DNA repair in *E. coli* (Rangarajan et al., 1999), toxin-mediated replication fork collapse in *C. crescentus* (Aakre et al., 2013). All described  $\beta$ -clamp interaction partner proteins share similar motif, the Clamp Binding Motif (CBM; Dalrymple et al., 2001).

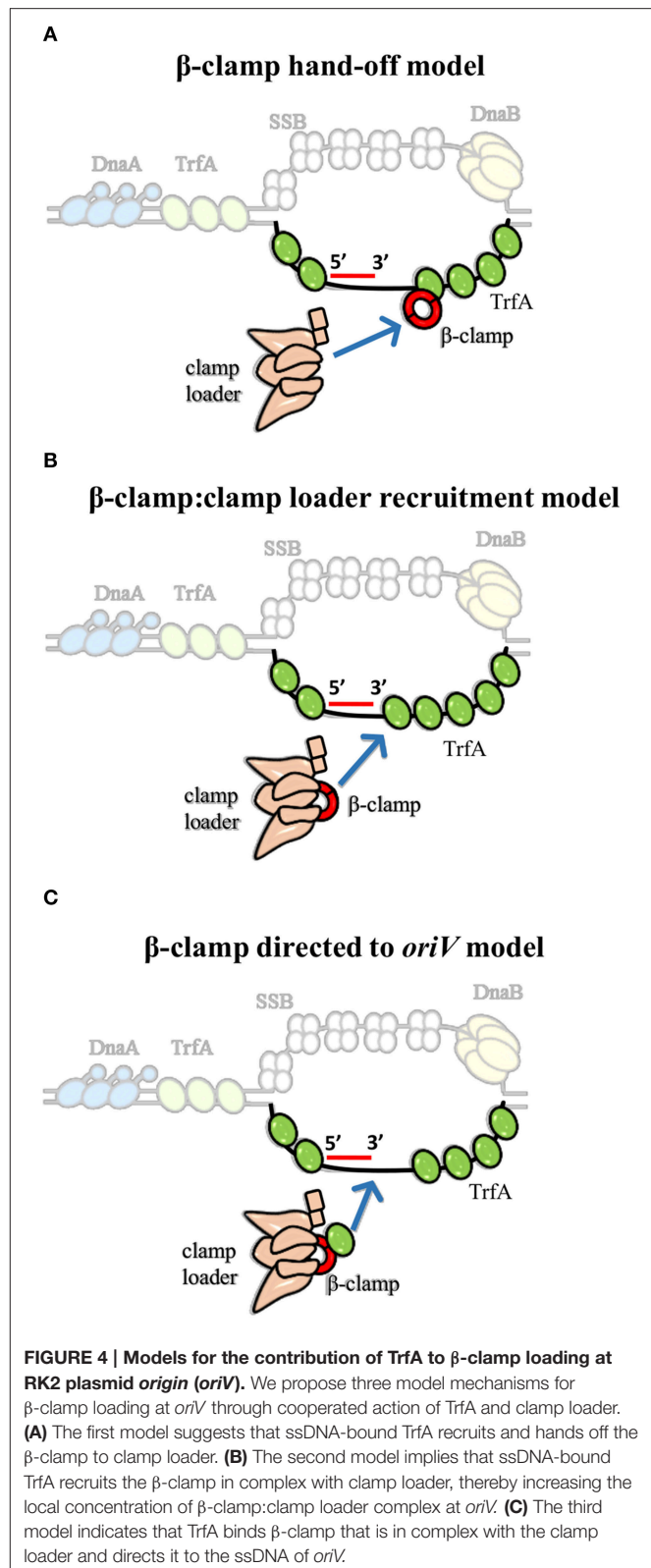
## $\beta$ -Clamp Loading at Origin of Iiteron Plasmid

Interestingly, clamp binding motif was also identified in plasmid replication initiators, including RK2 plasmid initiator-TrfA (Kongsuwan et al., 2006; Dalrymple et al., 2007). It was shown that TrfA protein lacking the leucine 137 and phenylalanine 138 within the clamp binding motif is unable to bind  $\beta$ -clamp (Kongsuwan et al., 2006). The TrfA  $\Delta$ LF mutant facilitated the determination of biological relevance of this interaction. The complex of TrfA and  $\beta$ -clamp was found to be the key feature for replisome assembly and thereby for *oriV*-dependent DNA replication of both supercoiled dsDNA plasmid and ssDNA plasmid *in vitro*, albeit the clamp loader complex is still crucial (Wawrzycka et al., 2015). Hence, the question arises—how do the Rep and clamp loader cooperate to load the  $\beta$ -clamp at plasmid origin? Three hypothetical models to explain the mechanism of Rep-mediated  $\beta$ -clamp loading could be considered (Figure 4). In the “ $\beta$ -clamp hand-off model” TrfA binds to the bottom strand of ssDNA close to the 3' end of synthesized primer, recruits  $\beta$ -clamp, and hands it off to the clamp loader complex (Figure 4A). Then, the  $\delta$  subunit of clamp loader opens the

$\beta$ -clamp and clamp loader positions it onto primer-template junction, as it is thought to occur during replisome assembly at *E. coli oriC*. This model is consistent with the results of the *in vitro* DNA replication experiments performed with the use of ssDNA, containing sequence of RK2 plasmid *oriV* (Wawrzyska et al., 2015). It was demonstrated that TrfA interacts with specific strand of ssDNA of DUE, i.e., the bottom strand, which serves as the site for replisome assembly (Wegrzyn et al., 2014; Wawrzyska et al., 2015). It can be further speculated that TrfA may assist the clamp loader in recognition of the 3' end of primer-template junction within the *oriV*. Another possible role of TrfA is illustrated in **Figure 4B** (second model, “ $\beta$ -clamp:clamp loader recruitment model”). Once TrfA is bound to bottom single strand of DUE, it recruits the  $\beta$ -clamp, which is in complex with clamp loader, to the RK2 plasmid *origin*. Thus, the local concentration of  $\beta$ -clamp:clamp loader complex increases, the clamp loader can assemble  $\beta$ -clamp onto the 3' end of a primer within the plasmid *origin*. Because TrfA- $\beta$ -clamp interaction was shown in the absence of DNA [using both ELISA and SPR (Surface Plasmon Resonance) technique (Kongsuwan et al., 2006; Wawrzyska et al., 2015)], the third model may also be justified (**Figure 4C**, “ $\beta$ -clamp directed to *oriV* model”). In the third model the TrfA that is not bound to DNA forms complex with  $\beta$ -clamp associated with the clamp loader, then directs it to the plasmid *origin*, *oriV*. Next, the clamp loader: $\beta$ -clamp:TrfA complex binds to the bottom strand of DUE via TrfA. TrfA passes the  $\beta$ -clamp bound to clamp loader on the primer-template junction. Although ATP binding to clamp loader (namely  $\gamma$  and  $\tau$  subunit) is required for  $\beta$ -clamp opening, it cannot be excluded that TrfA—whose ATPase activity has not been revealed—substitutes the clamp loader's function at this stage. TrfA may capture  $\beta$ -clamp in open conformation and load it onto primed DNA. Since ATP hydrolysis is required for  $\beta$ -clamp closing (Trakselis et al., 2001), here may participate the clamp loader.

## DNA SYNTHESIS AND THE ROLE OF SSB

After the  $\beta$ -clamp closes around primer-template junction and clamp loader dissociates, the final replisome component arrives—the Pol III core, that is composed of three subunits:  $\alpha$  (DNA polymerase),  $\epsilon$  (3'–5' proofreading exonuclease) and  $\theta$  ( $\epsilon$  subunit stabilizer; Kelman and O'Donnell, 1995; Taft-Benz and Schaaper, 2004). The number of Pol III cores within the replisome strictly depends on the clamp loader composition, since Pol III core is connected only through  $\tau$  subunit to the clamp loader. Various clamp loader complexes were widely studied in the light of processivity of DNA replication and it was established that three ATPases ( $\tau$  or  $\gamma$  subunit) must be included with  $\delta$  and  $\delta'$  subunits to form active pentameric structure (Kelch, 2016). Initially, it was thought that clamp loader contains two  $\tau$  subunits ( $\tau_2\gamma\delta\delta'\chi\psi$ ), so that two Pol III cores could constitute the replisome and synthesize the leading strand and lagging strand at the same time (Maki et al., 1988). However, further reports have argued on the stoichiometry of hPol III subunits (McInerney et al., 2007; Reyes-Lamothe et al., 2010; Dohrmann et al., 2016). The millisecond single molecule fluorescence microscopy as well as *in vitro* biochemical experiments showed that active *E. coli* replisome



contains three molecules of polymerase that are functional at replication fork (McInerney et al., 2007; Reyes-Lamothe et al., 2010). Both these studies assumed that trimeric polymerase is



associated with three molecules of  $\tau$  subunit. However, the very recent data indicated that in a bacterial cell there is predominately present Pol III<sub>2</sub>  $\tau_2\gamma\delta\delta'\chi\psi$  complex (Dohrmann et al., 2016). Since plasmids do not encode all essential proteins required for a plasmid replication, it is implied that the stage of DNA synthesis is similar during chromosomal DNA replication.

The DNA synthesis is facilitated by SSB, especially on the lagging strand (where the DNA synthesis is performed discontinuously) and is present in organisms from all domains of life (Shereda et al., 2008). Primary function of SSB is to protect ssDNA against degradation and melting secondary structures (Mackay and Linn, 1976; Meyer et al., 1979). SSB is linked to the clamp loader via  $\chi$  subunit (Glover and McHenry, 1998), which was shown to be important for DnaG primase displacement (Yuzhakov et al., 1999). Yet, SSB was termed the organizer of genome maintenance complexes and was shown to interact with at least 14 proteins, thus implying its diverse functions (reviewed in details in Shereda et al., 2008). The SSB interactions with proteins involves the C-terminal region of SSB, that is highly conserved among eubacterial SSB proteins. Some plasmids also encode SSB-like proteins, i.e., plasmid F, ColIb-P9, and RK2 (Chase et al., 1983; Howland et al., 1989; Thomas and Sherlock, 1990). While SSB of plasmid F and ColIb-P9 have similar structural domains, the RK2 SSB, termed P116, is smaller and contains only the N-terminal domain, which is responsible for DNA binding. P116 lacks the C-terminal protein binding-tail (Curth et al., 1996; Naue et al., 2013; Su et al., 2014), which may suggest that the role of P116 limits to ssDNA protection against nucleases.

## CONCLUSIONS AND PERSPECTIVES

The ground-breaking model of DNA replication initiation, introduced by Bramhill and Kornberg is still valid today (Bramhill and Kornberg, 1988b). They proposed that first the DnaA binds to DnaA-boxes to form an initial complex, then DnaA melts the AT-rich region (DUE) to form an open complex. Finally, DnaA directs the DnaB:DnaC complex into the open complex, thus forming a pre-priming complex, which marks the future forks of DNA replication (Bramhill

and Kornberg, 1988a,b). In this concept the chromosomal replication initiator, DnaA triggers the DNA replication initiation and is further required at each stage of the replication initiation process. Iteron plasmids also encode replication initiators that drive their replication initiation machinery. Despite the fact that plasmid and chromosomal replicons use overlapping set of proteins, there seems to be some subtle differences that may largely affect the whole process. Recent reports describe novel functions of replication initiators, both plasmid and chromosomal, that outreach the replication initiation process. The contribution of plasmid Rep protein to replisome assembly by providing direct Rep- $\beta$ -clamp interaction, shed a new light on how far-reaching activities replication initiators have i.e., determination of direction of DNA replication (Wawrzyczna et al., 2015). DnaA is also involved in a regulation of DNA replication initiation by a process termed RIDA (Regulatory Inactivation of DnaA; Katayama et al., 2010). One may ask if there is any other unanticipated activity of replication initiators to be discovered? What other processes are influenced by replication initiators? Described model mechanisms and unsolved questions of the structure-function relation of replication initiators in DNA replication and beyond this process await to be experimentally challenged.

## AUTHOR CONTRIBUTIONS

KW and MG wrote the manuscript and prepared figures, UU prepared the model of *E. coli* DnaA and figures, IK discussed and corrected the text of the manuscript.

## ACKNOWLEDGMENTS

We thank Dr. Magdalena Rajewska for critical reading of the manuscript. This work was supported by the Polish National Science Centre (Grant2012/04/A/NZ1/ 00048) and Polish Ministry of Science and Higher Education (DS/530-M040-D094-16). UU was supported by the European Commission from the FP7 Project Centre of Molecular Biotechnology for Healthy Life (MOBI4Health).

## REFERENCES

- Aakre, C. D., Phung, T. N., Huang, D., and Laub, M. T. (2013). A bacterial toxin inhibits DNA replication elongation through a direct interaction with the beta sliding clamp. *Mol. Cell* 52, 617–628. doi: 10.1016/j.molcel.2013.10.014
- Abe, Y., Jo, T., Matsuda, Y., Matsunaga, C., Katayama, T., and Ueda, T. (2007a). Structure and function of DnaA N-terminal domains: specific sites and mechanisms in inter-DnaA interaction and in DnaB helicase loading on oriC. *J. Biol. Chem.* 282, 17816–17827. doi: 10.1074/jbc.M701841200
- Abe, Y., Watanabe, N., Yoshida, Y., Ebata, F., Katayama, T., and Ueda, T. (2007b). Assignment of 1H, 13C and 15N resonances of N-terminal domain of DnaA protein. *Biomol. NMR Assign.* 1, 57–59. doi: 10.1007/s12104-007-9015-2
- Abeles, A. L. (1986). P1 plasmid replication. Purification and DNA-binding activity of the replication protein RepA. *J. Biol. Chem.* 261, 3548–3555.
- Abeles, A. L., Reaves, L. D., and Austin, S. J. (1990). A single DnaA box is sufficient for initiation from the P1 plasmid origin. *J. Bacteriol.* 172, 4386–4391.
- Abeles, A. L., Snyder, K. M., and Chatteraj, D. K. (1984). P1 plasmid replication: replicon structure. *J. Mol. Biol.* 173, 307–324. doi: 10.1016/0022-2836(84)90123-2
- Allen, G. C. Jr., and Kornberg, A. (1991). Fine balance in the regulation of DnaB helicase by DnaC protein in replication in *Escherichia coli*. *J. Biol. Chem.* 266, 22096–22101.
- Argiriadi, M. A., Goedken, E. R., Bruck, I., O'Donnell, M., and Kuriyan, J. (2006). Crystal structure of a DNA polymerase sliding clamp from a Gram-positive bacterium. *BMC Struct. Biol.* 6:2. doi: 10.1186/1472-6807-6-2
- Arias-Palomo, E., O'Shea, V. L., Hood, I. V., and Berger, J. M. (2013). The bacterial DnaC helicase loader is a DnaB ring breaker. *Cell* 153, 438–448. doi: 10.1016/j.cell.2013.03.006
- Bailey, S., Eliason, W. K., and Steitz, T. A. (2007a). The crystal structure of the *Thermus aquaticus* DnaB helicase monomer. *Nucleic Acids Res.* 35, 4728–4736. doi: 10.1093/nar/gkm507
- Bailey, S., Eliason, W. K., and Steitz, T. A. (2007b). Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase. *Science* 318, 459–463. doi: 10.1126/science.1147353



- Bell, S. P., and Kaguni, J. M. (2013). Helicase loading at chromosomal origins of replication. *Cold Spring Harb Perspect. Biol.* 5:a010124. doi: 10.1101/cshperspect.a010124
- Bernander, R., Dasgupta, S., and Nordström, K. (1991). The *E. coli* cell cycle and the plasmid R1 replication cycle in the absence of the DnaA protein. *Cell* 64, 1145–1153. doi: 10.1016/0092-8674(91)90269-5
- Bernander, R., Krabbe, M., and Nordström, K. (1992). Mapping of the *in vivo* start site for leading strand DNA synthesis in plasmid R1. *EMBO J.* 11, 4481–4487.
- Biswas, S. B., Chen, P. H., and Biswas, E. E. (1994). Structure and function of *Escherichia coli* DnaB protein: role of the N-terminal domain in helicase activity. *Biochemistry* 33, 11307–11314. doi: 10.1021/bi00203a028
- Bowers, L. M., Krüger, R., and Filutowicz, M. (2007). Mechanism of origin activation by monomers of R6K-encoded  $\pi$  protein. *J. Mol. Biol.* 368, 928–938. doi: 10.1016/j.jmb.2007.02.074
- Bowman, G. D., Goedken, E. R., Kazmirski, S. L., O'Donnell, M., and Kuriyan, J. (2005). DNA polymerase clamp loaders and DNA recognition. *FEBS Lett.* 579, 863–867. doi: 10.1016/j.febslet.2004.11.038
- Bowman, G. D., O'Donnell, M., and Kuriyan, J. (2004). Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. *Nature* 429, 724–730. doi: 10.1038/nature02585
- Bramhill, D., and Kornberg, A. (1988a). Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the *E. coli* chromosome. *Cell* 52, 743–755. doi: 10.1016/0092-8674(88)90412-6
- Bramhill, D., and Kornberg, A. (1988b). A model for initiation at origins of DNA replication. *Cell* 54, 915–918. doi: 10.1016/0092-8674(88)90102-X
- Brendler, T. G., Abeles, A. L., Reaves, L. D., and Austin, S. J. (1997). The iteron bases and spacers of the P1 replication origin contain information that specifies the formation of a complex structure involved in initiation. *Mol. Microbiol.* 23, 559–567. doi: 10.1046/j.1365-2958.1997.d01-1869.x
- Brüning, J. G., Myka, K. K., and McGlynn, P. (2016). Overexpression of the replicative helicase in *Escherichia coli* inhibits replication initiation and replication fork reloading. *J. Mol. Biol.* 428, 1068–1079. doi: 10.1016/j.jmb.2016.01.018
- Burnouf, D. Y., Olieric, V., Wagner, J., Fujii, S., Reinbolt, J., Fuchs, R. P., et al. (2004). Structural and biochemical analysis of sliding clamp/ligand interactions suggest a competition between replicative and translesion DNA polymerases. *J. Mol. Biol.* 335, 1187–1197. doi: 10.1016/j.jmb.2003.11.049
- Caspi, R., Helinski, D. R., Pacek, M., and Konieczny, I. (2000). Interactions of DnaA proteins from distantly related bacteria with the replication origin of the broad host range plasmid RK2. *J. Biol. Chem.* 275, 18454–18461. doi: 10.1074/jbc.M000552200
- Caspi, R., Pacek, M., Consiglieri, G., Helinski, D. R., Toukdarian, A., and Konieczny, I. (2001). A broad host range replicon with different requirements for replication initiation in three bacterial species. *EMBO J.* 20, 3262–3271. doi: 10.1093/emboj/20.12.3262
- Charbon, G., and Løbner-Olesen, A. (2011). A role for the weak DnaA binding sites in bacterial replication origins. *Mol. Microbiol.* 82, 272–274. doi: 10.1111/j.1365-2958.2011.07840.x
- Chase, J. W., Merrill, B. M., and Williams, K. R. (1983). F sex factor encodes a single-stranded DNA binding protein (SSB) with extensive sequence homology to *Escherichia coli* SSB. *Proc. Natl. Acad. Sci. U.S.A.* 80, 5480–5484. doi: 10.1073/pnas.80.18.5480
- Cheng, H. M., Gröger, P., Hartmann, A., and Schlierf, M. (2015). Bacterial initiators form dynamic filaments on single-stranded DNA monomer by monomer. *Nucleic Acids Res.* 43, 396–405. doi: 10.1093/nar/gku1284
- Cho, W. K., Jergic, S., Kim, D., Dixon, N. E., and Lee, J. B. (2014). Loading dynamics of a sliding DNA clamp. *Angew. Chem. Int. Ed Engl.* 53, 6768–6771. doi: 10.1002/anie.201403063
- Chodavarapu, S., Felczak, M. M., and Kaguni, J. M. (2011). Two forms of ribosomal protein L2 of *Escherichia coli* that inhibit DnaA in DNA replication. *Nucleic Acids Res.* 39, 4180–4191. doi: 10.1093/nar/gkq1203
- Chodavarapu, S., Felczak, M. M., Yaniv, J. R., and Kaguni, J. M. (2008a). *Escherichia coli* DnaA interacts with HU in initiation at the *E. coli* replication origin. *Mol. Microbiol.* 67, 781–792. doi: 10.1111/j.1365-2958.2007.06094.x
- Chodavarapu, S., Gomez, R., Vicente, M., and Kaguni, J. M. (2008b). *Escherichia coli* Dps interacts with DnaA protein to impede initiation: a model of adaptive mutation. *Mol. Microbiol.* 67, 1331–1346. doi: 10.1111/j.1365-2958.2008.06127.x
- Chodavarapu, S., Jones, A. D., Feig, M., and Kaguni, J. M. (2016). DnaC traps DnaB as an open ring and remodels the domain that binds primase. *Nucleic Acids Res.* 44, 210–220. doi: 10.1093/nar/gkv961
- Churchward, G., Linder, P., and Caro, L. (1984). Replication functions encoded by the plasmid pSC101. *Adv. Exp. Med. Biol.* 179, 209–214. doi: 10.1007/978-1-4684-8730-5\_21
- Costa, A., Hood, I. V., and Berger, J. M. (2013). Mechanisms for initiating cellular DNA replication. *Annu. Rev. Biochem.* 82, 25–54. doi: 10.1146/annurev-biochem-052610-094414
- Curth, U., Genschel, J., Urbanke, C., and Greipel, J. (1996). *In vitro* and *in vivo* function of the C-terminus of *Escherichia coli* single-stranded DNA binding protein. *Nucleic Acids Res.* 24, 2706–2711. doi: 10.1093/nar/24.14.2706
- Dalrymple, B. P., Kongsuwan, K., and Wijffels, G. (2007). Identification of putative DnaN-binding motifs in plasmid replication initiation proteins. *Plasmid* 57, 82–88. doi: 10.1016/j.plasmid.2006.07.005
- Dalrymple, B. P., Kongsuwan, K., Wijffels, G., Dixon, N. E., and Jennings, P. A. (2001). A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11627–11632. doi: 10.1073/pnas.191384398
- Datta, H. J., Khatri, G. S., and Bastia, D. (1999). Mechanism of recruitment of DnaB helicase to the replication origin of the plasmid pSC101. *Proc. Natl. Acad. Sci. U.S.A.* 96, 73–78. doi: 10.1073/pnas.96.1.73
- Davey, M. J., and O'Donnell, M. (2003). Replicative helicase loaders: ring breakers and ring makers. *Curr. Biol.* 13, R594–R596. doi: 10.1016/S0960-9822(03)00523-2
- Díaz-López, T., Lages-Gonzalo, M., Serrano-López, A., Alfonso, C., Rivas, G., Diaz-Orejias, R., et al. (2003). Structural changes in RepA, a plasmid replication initiator, upon binding to origin DNA. *J. Biol. Chem.* 278, 18606–18616. doi: 10.1074/jbc.M212024200
- Dionne, I., Nookala, R. K., Jackson, S. P., Doherty, A. J., and Bell, S. D. (2003). A heterotrimeric PCNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol. Cell* 11, 275–282. doi: 10.1016/S1097-2765(02)00824-9
- Dohrmann, P. R., Correa, R., Frisch, R. L., Rosenberg, S. M., and McHenry, C. S. (2016). The DNA polymerase III holoenzyme contains gamma and is not a trimeric polymerase. *Nucleic Acids Res.* 44, 1285–1297. doi: 10.1093/nar/gkv1510
- Doran, K. S., Helinski, D. R., and Konieczny, I. (1999). A critical DnaA box directs the cooperative binding of the *Escherichia coli* DnaA protein to the plasmid RK2 replication origin. *J. Biol. Chem.* 274, 17918–17923. doi: 10.1074/jbc.274.25.17918
- Doran, K. S., Konieczny, I., and Helinski, D. R. (1998). Replication origin of the broad host range plasmid RK2. *Positioning of various motifs is critical for initiation of replication.* *J. Biol. Chem.* 273, 8447–8453. doi: 10.1074/jbc.273.14.8447
- Downey, C. D., and McHenry, C. S. (2010). Chaperoning of a replicative polymerase onto a newly assembled DNA-bound sliding clamp by the clamp loader. *Mol. Cell* 37, 481–491. doi: 10.1016/j.molcel.2010.01.013
- Duderstadt, K. E., Chuang, K., and Berger, J. M. (2011). DNA stretching by bacterial initiators promotes replication origin opening. *Nature* 478, 209–213. doi: 10.1038/nature10455
- Duderstadt, K. E., Mott, M. L., Crisona, N. J., Chuang, K., Yang, H., and Berger, J. M. (2010). Origin remodeling and opening in bacteria rely on distinct assembly states of the DnaA initiator. *J. Biol. Chem.* 285, 28229–28239. doi: 10.1074/jbc.M110.147975
- Erzberger, J. P., Mott, M. L., and Berger, J. M. (2006). Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat. Struct. Mol. Biol.* 13, 676–683. doi: 10.1038/nsmb1115
- Erzberger, J. P., Pirruccello, M. M., and Berger, J. M. (2002). The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. *EMBO J.* 21, 4763–4773. doi: 10.1093/emboj/cdf496
- Fang, J., Engen, J. R., and Beuning, P. J. (2011). *Escherichia coli* processivity clamp beta from DNA polymerase III is dynamic in solution. *Biochemistry* 50, 5958–5968. doi: 10.1021/bi200580b
- Fang, J., Nevin, P., Kairys, V., Venclovas, C., Engen, J. R., and Beuning, P. J. (2014). Conformational analysis of processivity clamps in solution demonstrates that tertiary structure does not correlate with protein dynamics. *Structure* 22, 572–581. doi: 10.1016/j.str.2014.02.001

- Fekete, R. A., Venkova-Canova, T., Park, K., and Chattoraj, D. K. (2006). IHF-dependent activation of P1 plasmid origin by dnaA. *Mol. Microbiol.* 62, 1739–1751. doi: 10.1111/j.1365-2958.2006.05479.x
- Felczak, M. M., Sage, J. M., Hupert-Kocurek, K., Aykul, S., and Kaguni, J. M. (2016). Substitutions of conserved residues in the C-terminal region of DnaC cause thermolability in helicase loading. *J. Biol. Chem.* 291, 4803–4812. doi: 10.1074/jbc.M115.708586
- Filutowicz, M., Davis, G., Greener, A., and Helinski, D. R. (1985). Autorepressor properties of the pi-initiation protein encoded by plasmid R6K. *Nucleic Acids Res.* 13, 103–114. doi: 10.1093/nar/13.1.103
- Filutowicz, M., McEachern, M. J., Mukhopadhyay, P., Greener, A., Yang, S. L., and Helinski, D. R. (1987). DNA and protein interactions in the regulation of plasmid replication. *J. Cell Sci. Suppl.* 7, 15–31. doi: 10.1242/jcs.1987.Supplement\_7.2
- Flower, A. M., and McHenry, C. S. (1990). The gamma subunit of DNA polymerase III holoenzyme of *Escherichia coli* is produced by ribosomal frameshifting. *Proc. Natl. Acad. Sci. U.S.A.* 87, 3713–3717. doi: 10.1073/pnas.87.10.3713
- Flowers, S., Biswas, E. E., and Biswas, S. B. (2003). Conformational dynamics of DnaB helicase upon DNA and nucleotide binding: analysis by intrinsic tryptophan fluorescence quenching. *Biochemistry* 42, 1910–1921. doi: 10.1021/bi025992v
- Fujikawa, N., Kurumizaka, H., Nureki, O., Terada, T., Shirouzu, M., Katayama, T., et al. (2003). Structural basis of replication origin recognition by the DnaA protein. *Nucleic Acids Res.* 31, 2077–2086. doi: 10.1093/nar/gkg309
- Fuller, R. S., Funnell, B. E., and Kornberg, A. (1984). The dnaA protein complex with the *E. coli* chromosomal replication origin (*oriC*) and other DNA sites. *Cell* 38, 889–900. doi: 10.1016/0092-8674(84)90284-8
- Funnell, B. E., Baker, T. A., and Kornberg, A. (1987). *In vitro* assembly of a prepriming complex at the origin of the *Escherichia coli* chromosome. *J. Biol. Chem.* 262, 10327–10334.
- Galletto, R., Jezewska, M. J., and Bujalowski, W. (2004). Unzipping mechanism of the double-stranded DNA unwinding by a hexameric helicase: the effect of the 3' arm and the stability of the dsDNA on the unwinding activity of the *Escherichia coli* DnaB helicase. *J. Mol. Biol.* 343, 101–114. doi: 10.1016/j.jmb.2004.07.056
- Gille, H., Egan, J. B., Roth, A., and Messer, W. (1991). The FIS protein binds and bends the origin of chromosomal DNA replication, *oriC*, of *Escherichia coli*. *Nucleic Acids Res.* 19, 4167–4172. doi: 10.1093/nar/19.15.4167
- Giraldo, R., Andreu, J. M., and Díaz-Orejas, R. (1998). Protein domains and conformational changes in the activation of RepA, a DNA replication initiator. *EMBO J.* 17, 4511–4526. doi: 10.1093/emboj/17.15.4511
- Giraldo, R., Fernandez-Tornero, C., Evans, P. R., Diaz-Orejas, R., and Romero, A. (2003). A conformational switch between transcriptional repression and replication initiation in the RepA dimerization domain. *Nat. Struct. Biol.* 10, 565–571. doi: 10.1038/nsb937
- Glover, B. P., and McHenry, C. S. (1998). The chi psi subunits of DNA polymerase III holoenzyme bind to single-stranded DNA-binding protein (SSB) and facilitate replication of an SSB-coated template. *J. Biol. Chem.* 273, 23476–23484. doi: 10.1074/jbc.273.36.23476
- Grimwade, J. E., Ryan, V. T., and Leonard, A. C. (2000). IHF redistributes bound initiator protein, DnaA, on supercoiled *oriC* of *Escherichia coli*. *Mol. Microbiol.* 35, 835–844. doi: 10.1046/j.1365-2958.2000.01755.x
- Gui, W. J., Lin, S. Q., Chen, Y. Y., Zhang, X. E., Bi, L. J., and Jiang, T. (2011). Crystal structure of DNA polymerase III beta sliding clamp from *Mycobacterium tuberculosis*. *Biochem. Biophys. Res. Commun.* 405, 272–277. doi: 10.1016/j.bbrc.2011.01.027
- Hayner, J. N., Douma, L. G., and Bloom, L. B. (2014). The interplay of primer-template DNA phosphorylation status and single-stranded DNA binding proteins in directing clamp loaders to the appropriate polarity of DNA. *Nucleic Acids Res.* 42, 10655–10667. doi: 10.1093/nar/gku774
- Hedglin, M., Kumar, R., and Benkovic, S. J. (2013). Replication clamps and clamp loaders. *Cold Spring Harb. Perspect. Biol.* 5:a010165. doi: 10.1101/cshperspect.a010165
- Howland, C. J., Rees, C. E., Barth, P. T., and Wilkins, B. M. (1989). The ssb gene of plasmid ColIb-P9. *J. Bacteriol.* 171, 2466–2473.
- Hwang, D. S., and Kornberg, A. (1992). Opening of the replication origin of *Escherichia coli* by DnaA protein with protein HU or IHF. *J. Biol. Chem.* 267, 23083–23086.
- Indiani, C., and O'Donnell, M. (2006). The replication clamp-loading machine at work in the three domains of life. *Nat. Rev. Mol. Cell Biol.* 7, 751–761. doi: 10.1038/nrm2022
- Ingmer, H., and Cohen, S. N. (1993). Excess intracellular concentration of the pSC101 RepA protein interferes with both plasmid DNA replication and partitioning. *J. Bacteriol.* 175, 7834–7841.
- Ingmer, H., Fong, E. L., and Cohen, S. N. (1995). Monomer-dimer equilibrium of the pSC101 RepA protein. *J. Mol. Biol.* 250, 309–314. doi: 10.1006/jmbi.1995.0378
- Itoh, T., and Horii, T. (1989). Replication of ColE2 and ColE3 plasmids: *in vitro* replication dependent on plasmid-coded proteins. *Mol. Gen. Genet.* 219, 249–255. doi: 10.1007/BF00261184
- Ito, H., Yagura, M., Shirakihara, Y., and Itoh, T. (2015). Structural basis for replication origin unwinding by an initiator primase of plasmid ColE2-P9: duplex DNA unwinding by a single protein. *J. Biol. Chem.* 290, 3601–3611. doi: 10.1074/jbc.M114.595645
- Jeruzalmi, D., O'Donnell, M., and Kuriyan, J. (2001). Crystal structure of the processivity clamp loader gamma (gamma) complex of *E. coli* DNA polymerase III. *Cell* 106, 429–441. doi: 10.1016/S0092-8674(01)00463-9
- Jezewska, M. J., Kim, U. S., and Bujalowski, W. (1996). Binding of *Escherichia coli* primary replicative helicase DnaB protein to single-stranded DNA. Long-range allosteric conformational changes within the protein hexamer. *Biochemistry* 35, 2129–2145. doi: 10.1021/bi952345d
- Jezewska, M. J., Rajendran, S., Bujalowski, D., and Bujalowski, W. (1998a). Does single-stranded DNA pass through the inner channel of the protein hexamer in the complex with the *Escherichia coli* DnaB Helicase? *Fluorescence energy transfer studies*. *J. Biol. Chem.* 273, 10515–10529. doi: 10.1074/jbc.273.17.10515
- Jezewska, M. J., Rajendran, S., and Bujalowski, W. (1998b). Functional and structural heterogeneity of the DNA binding site of the *Escherichia coli* primary replicative helicase DnaB protein. *J. Biol. Chem.* 273, 9058–9069. doi: 10.1074/jbc.273.15.9058
- Jiang, Y., Pacek, M., Helinski, D. R., Konieczny, I., and Toukdarian, A. (2003). A multifunctional plasmid-encoded replication initiation protein both recruits and positions an active helicase at the replication origin. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8692–8697. doi: 10.1073/pnas.1532393100
- Katayama, T., Ozaki, S., Keyamura, K., and Fujimitsu, K. (2010). Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and *oriC*. *Nat. Rev. Microbiol.* 8, 163–170. doi: 10.1038/nrmicro2314
- Kaur, G., Vora, M. P., Czerwonka, C. A., Rozgaja, T. A., Grimwade, J. E., and Leonard, A. C. (2014). Building the bacterial orisome: high-affinity DnaA recognition plays a role in setting the conformation of *oriC* DNA. *Mol. Microbiol.* 91, 1148–1163. doi: 10.1111/mmi.12525
- Kawakami, H., Keyamura, K., and Katayama, T. (2005). Formation of an ATP-DnaA-specific initiation complex requires DnaA Arginine 285, a conserved motif in the AAA+ protein family. *J. Biol. Chem.* 280, 27420–27430. doi: 10.1074/jbc.M502764200
- Kawasaki, Y., Matsunaga, F., Kano, Y., Yura, T., and Wada, C. (1996). The localized melting of mini-F origin by the combined action of the mini-F initiator protein (RepE) and HU and DnaA of *Escherichia coli*. *Mol. Gen. Genet.* 253, 42–49. doi: 10.1007/s004380050294
- Kawasaki, Y., Wada, C., and Yura, T. (1990). Roles of *Escherichia coli* heat shock proteins DnaK, DnaJ and GrpE in mini-F plasmid replication. *Mol. Gen. Genet.* 220, 277–282. doi: 10.1007/BF00260494
- Kelch, B. A. (2016). Review: the lord of the rings: structure and mechanism of the sliding clamp loader. *Biopolymers* 105, 532–546. doi: 10.1002/bip.22827
- Kelch, B. A., Makino, D. L., O'Donnell, M., and Kuriyan, J. (2011). How a DNA polymerase clamp loader opens a sliding clamp. *Science* 334, 1675–1680. doi: 10.1126/science.1211884
- Kelch, B. A., Makino, D. L., O'Donnell, M., and Kuriyan, J. (2012). Clamp loader ATPases and the evolution of DNA replication machinery. *BMC Biol.* 10:34. doi: 10.1186/1741-7007-10-34
- Kelman, Z., and O'Donnell, M. (1995). DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annu. Rev. Biochem.* 64, 171–200. doi: 10.1146/annurev.bi.64.070195.001131
- Keyamura, K., Abe, Y., Higashi, M., Ueda, T., and Katayama, T. (2009). DiaA dynamics are coupled with changes in initial origin complexes

- leading to helicase loading. *J. Biol. Chem.* 284, 25038–25050. doi: 10.1074/jbc.M109.002717
- Keyamura, K., Fujikawa, N., Ishida, T., Ozaki, S., Su'etsugu, M., Fujimitsu, K., et al. (2007). The interaction of DiaA and DnaA regulates the replication cycle in *E. coli* by directly promoting ATP DnaA-specific initiation complexes. *Genes Dev.* 21, 2083–2099. doi: 10.1101/gad.1561207
- Kline, B. C., Kogoma, T., Tam, J. E., and Shields, M. S. (1986). Requirement of the *Escherichia coli* dnaA gene product for plasmid F maintenance. *J. Bacteriol.* 168, 440–443.
- Kobori, J. A., and Kornberg, A. (1982). The *Escherichia coli* dnaC gene product. II. Purification, physical properties, and role in replication. *J. Biol. Chem.* 257, 13763–13769.
- Komori, H., Matsunaga, F., Higuchi, Y., Ishiai, M., Wada, C., and Miki, K. (1999). Crystal structure of a prokaryotic replication initiator protein bound to DNA at 2.6 Å resolution. *EMBO J.* 18, 4597–4607. doi: 10.1093/emboj/18.17.4597
- Kongsuwan, K., Josh, P., Picault, M. J., Wijffels, G., and Dalrymple, B. (2006). The plasmid RK2 replication initiator protein (TrfA) binds to the sliding clamp beta subunit of DNA polymerase III: implication for the toxicity of a peptide derived from the amino-terminal portion of 33-kilodalton TrfA. *J. Bacteriol.* 188, 5501–5509. doi: 10.1128/JB.00231-06
- Konieczny, I. (2003). Strategies for helicase recruitment and loading in bacteria. *EMBO Rep.* 4, 37–41. doi: 10.1038/sj.embor.embor703
- Konieczny, I., Doran, K. S., Helinski, D. R., and Blasina, A. (1997). Role of TrfA and DnaA proteins in origin opening during initiation of DNA replication of the broad host range plasmid RK2. *J. Biol. Chem.* 272, 20173–20178. doi: 10.1074/jbc.272.32.20173
- Konieczny, I., and Helinski, D. R. (1997). The replication initiation protein of the broad-host-range plasmid RK2 is activated by the ClpX chaperone. *Proc. Natl. Acad. Sci. U.S.A.* 94, 14378–14382. doi: 10.1073/pnas.94.26.14378
- Kowalczyk, L., Rajewska, M., and Konieczny, I. (2005). Positioning and the specific sequence of each 13-mer motif are critical for activity of the plasmid RK2 replication origin. *Mol. Microbiol.* 57, 1439–1449. doi: 10.1111/j.1365-2958.2005.04770.x
- Krishna, T. S., Kong, X. P., Gary, S., Burgers, P. M., and Kuriyan, J. (1994). Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* 79, 1233–1243. doi: 10.1016/0092-8674(94)90014-0
- Kukshal, V., Khanam, T., Chopra, D., Singh, N., Sanyal, S., and Ramachandran, R. (2012). *M. tuberculosis* sliding beta-clamp does not interact directly with the NAD<sup>+</sup>-dependent DNA ligase. *PLoS ONE* 7:e35702. doi: 10.1371/journal.pone.0035702
- LeBowitz, J. H., and McMacken, R. (1986). The *Escherichia coli* dnaB replication protein is a DNA helicase. *J. Biol. Chem.* 261, 4738–4748.
- Le Chatelier, E., Janniere, L., Ehrlich, S. D., and Canceill, D. (2001). The RepE initiator is a double-stranded and single-stranded DNA-binding protein that forms an atypical open complex at the onset of replication of plasmid pAMBeta 1 from Gram-positive bacteria. *J. Biol. Chem.* 276, 10234–10246. doi: 10.1074/jbc.M010118200
- Leonard, A. C., and Grimwade, J. E. (2005). Building a bacterial orisome: emergence of new regulatory features for replication origin unwinding. *Mol. Microbiol.* 55, 978–985. doi: 10.1111/j.1365-2958.2004.04467.x
- Leonard, A. C., and Grimwade, J. E. (2011). Regulation of DnaA assembly and activity: taking directions from the genome. *Annu. Rev. Microbiol.* 65, 19–35. doi: 10.1146/annurev-micro-090110-102934
- Leonard, A. C., and Grimwade, J. E. (2015). The orisome: structure and function. *Front. Microbiol.* 6:545. doi: 10.3389/fmicb.2015.00545
- Liu, J., Pei, H., Mei, S., Li, J., Zhou, L., and Xiang, H. (2008). Replication initiator DnaA interacts with an anti-terminator NusG in *T. tengcongensis*. *Biochem. Biophys. Res. Commun.* 371, 573–577. doi: 10.1016/j.bbrc.2008.04.131
- Lo, Y. H., Tsai, K. L., Sun, Y. J., Chen, W. T., Huang, C. Y., and Hsiao, C. D. (2009). The crystal structure of a replicative hexameric helicase DnaC and its complex with single-stranded DNA. *Nucleic Acids Res.* 37, 804–814. doi: 10.1093/nar/gkn999
- Lofite-Eaton, W., and Rawlings, D. E. (2012). Diversity, biology and evolution of IncQ-family plasmids. *Plasmid* 67, 15–34. doi: 10.1016/j.plasmid.2011.10.001
- Lohman, T. M., and Bjornson, K. P. (1996). Mechanisms of helicase-catalyzed DNA unwinding. *Annu. Rev. Biochem.* 65, 169–214. doi: 10.1146/annurev.bi.65.070196.001125
- Lowery, T. J., Pelton, J. G., Chandonia, J. M., Kim, R., Yokota, H., and Wemmer, D. E. (2007). NMR structure of the N-terminal domain of the replication initiator protein DnaA. *J. Struct. Funct. Genomics* 8, 11–17. doi: 10.1007/s10969-007-9022-7
- Lu, Y. B., Datta, H. J., and Bastia, D. (1998). Mechanistic studies of initiator-initiator interaction and replication initiation. *EMBO J.* 17, 5192–5200. doi: 10.1093/emboj/17.17.5192
- Lu, Y. B., Ratnakar, P. V., Mohanty, B. K., and Bastia, D. (1996). Direct physical interaction between DnaG primase and DnaB helicase of *Escherichia coli* is necessary for optimal synthesis of primer RNA. *Proc. Natl. Acad. Sci. U.S.A.* 93, 12902–12907. doi: 10.1073/pnas.93.23.12902
- Mackay, V., and Linn, S. (1976). Selective inhibition of the dnase activity of the recBC enzyme by the DNA binding protein from *Escherichia coli*. *J. Biol. Chem.* 251, 3716–3719.
- Maestro, B., Sanz, J. M., Diaz-Orejas, R., and Fernandez-Tresguerres, E. (2003). Modulation of pPS10 host range by plasmid-encoded RepA initiator protein. *J. Bacteriol.* 185, 1367–1375. doi: 10.1128/JB.185.4.1367-1375.2003
- Maki, H., Maki, S., and Kornberg, A. (1988). DNA Polymerase III holoenzyme of *Escherichia coli*. IV. The holoenzyme is an asymmetric dimer with twin active sites. *J. Biol. Chem.* 263, 6570–6578.
- Makowska-Grzyska, M., and Kaguni, J. M. (2010). Primase directs the release of DnaC from DnaB. *Mol. Cell* 37, 90–101. doi: 10.1016/j.molcel.2009.12.031
- Matsui, M., Oka, A., Takamami, M., Yasuda, S., and Hirota, Y. (1985). Sites of dnaA protein-binding in the replication origin of the *Escherichia coli* K-12 chromosome. *J. Mol. Biol.* 184, 529–533. doi: 10.1016/0022-2836(85)90299-2
- McEachern, M. J., Filutowicz, M., and Helinski, D. R. (1985). Mutations in direct repeat sequences and in a conserved sequence adjacent to the repeats result in a defective replication origin in plasmid R6K. *Proc. Natl. Acad. Sci. U.S.A.* 82, 1480–1484. doi: 10.1073/pnas.82.5.1480
- McGarry, K. C., Ryan, V. T., Grimwade, J. E., and Leonard, A. C. (2004). Two discriminatory binding sites in the *Escherichia coli* replication origin are required for DNA strand opening by initiator DnaA-ATP. *Proc. Natl. Acad. Sci. U.S.A.* 101, 2811–2816. doi: 10.1073/pnas.0400340101
- McHenry, C. S. (2011). DNA replisomes from a bacterial perspective. *Annu. Rev. Biochem.* 80, 403–436. doi: 10.1146/annurev-biochem-061208-091655
- McInerney, P., Johnson, A., Katz, F., and O'Donnell, M. (2007). Characterization of a triple DNA polymerase replisome. *Mol. Cell* 27, 527–538. doi: 10.1016/j.molcel.2007.06.019
- Messer, W., Blaessing, F., Majka, J., Nardmann, J., Schaper, S., Schmidt, A., et al. (1999). Functional domains of DnaA proteins. *Biochimie* 81, 819–825. doi: 10.1016/S0300-9084(99)00215-1
- Meyer, R. R., Glassberg, J., and Kornberg, A. (1979). An *Escherichia coli* mutant defective in single-strand binding protein is defective in DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* 76, 1702–1705. doi: 10.1073/pnas.76.4.1702
- Miron, A., Mukherjee, S., and Bastia, D. (1992). Activation of distant replication origins *in vivo* by DNA looping as revealed by a novel mutant form of an initiator protein defective in cooperativity at a distance. *EMBO J.* 11, 1205–1216.
- Molt, K. L., Sutera, V. A. Jr., Moore, K. K., and Lovett, S. T. (2009). A role for nonessential domain II of initiator protein, DnaA, in replication control. *Genetics* 183, 39–49. doi: 10.1534/genetics.109.104760
- Mott, M. L., Erzberger, J. P., Coons, M. M., and Berger, J. M. (2008). Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators. *Cell* 135, 623–634. doi: 10.1016/j.cell.2008.09.058
- Murakami, Y., Ohmori, H., Yura, T., and Nagata, T. (1987). Requirement of the *Escherichia coli* dnaA gene function for ori-2-dependent mini-F plasmid replication. *J. Bacteriol.* 169, 1724–1730.
- Murotsu, T., Matsubara, K., Sugisaki, H., and Takamami, M. (1981). Nine unique repeating sequences in a region essential for replication and incompatibility of the mini-F plasmid. *Gene* 15, 257–271. doi: 10.1016/0378-1119(81)90135-9
- Nakamura, A., Komori, H., Kobayashi, G., Kita, A., Wada, C., and Miki, K. (2004). The N-terminal domain of the replication initiator protein RepE is a dimerization domain forming a stable dimer. *Biochem. Biophys. Res. Commun.* 315, 10–15. doi: 10.1016/j.bbrc.2004.01.018
- Nakamura, A., Wada, C., and Miki, K. (2007a). Expression and purification of F-plasmid RepE and preliminary X-ray crystallographic study of its complex with operator DNA. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 63, 346–349. doi: 10.1107/S1744309107012894



- Nakamura, A., Wada, C., and Miki, K. (2007b). Structural basis for regulation of bifunctional roles in replication initiator protein. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18484–18489. doi: 10.1073/pnas.0705623104
- Natrajan, G., Hall, D. R., Thompson, A. C., Gutsche, I., and Terradot, L. (2007). Structural similarity between the DnaA-binding proteins HobA (HP1230) from *Helicobacter pylori* and DiaA from *Escherichia coli*. *Mol. Microbiol.* 65, 995–1005. doi: 10.1111/j.1365-2958.2007.05843.x
- Natrajan, G., Noiro-Gros, M. F., Zawilak-Pawlik, A., Kapp, U., and Terradot, L. (2009). The structure of a DnaA/HobA complex from *Helicobacter pylori* provides insight into regulation of DNA replication in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21115–21120. doi: 10.1073/pnas.0908966106
- Naue, N., Beerbaum, M., Bogutzki, A., Schmieder, P., and Curth, U. (2013). The helicase-binding domain of *Escherichia coli* DnaG primase interacts with the highly conserved C-terminal region of single-stranded DNA-binding protein. *Nucleic Acids Res.* 41, 4507–4517. doi: 10.1093/nar/gkt107
- Neuwald, A. F., Aravind, L., Spouge, J. L., and Koonin, E. V. (1999). AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* 9, 27–43.
- Nieto, C., Giraldo, R., Fernández-Tresguerres, E., and Díaz, R. (1992). Genetic and functional analysis of the basic replicon of pPS10, a plasmid specific for *Pseudomonas* isolated from *Pseudomonas syringae* patovar savastanoi. *J. Mol. Biol.* 223, 415–426. doi: 10.1016/0022-2836(92)90661-3
- Nievera, C., Torgue, J. J., Grimwade, J. E., and Leonard, A. C. (2006). SeqA blocking of DnaA-oriC interactions ensures staged assembly of the *E. coli* pre-RC. *Mol. Cell* 24, 581–592. doi: 10.1016/j.molcel.2006.09.016
- Nozaki, S., and Ogawa, T. (2008). Determination of the minimum domain II size of *Escherichia coli* DnaA protein essential for cell viability. *Microbiology* 154, 3379–3384. doi: 10.1099/mic.0.2008/019745-0
- Oakley, A. J., Prosser, P., Wijffels, G., Beck, J. L., Wilce, M. C., and Dixon, N. E. (2003). Flexibility revealed by the 1.85 Å crystal structure of the beta sliding-clamp subunit of *Escherichia coli* DNA polymerase III. *Acta Crystallogr. D Biol. Crystallogr.* 59, 1192–1199. doi: 10.1107/S0907444903009958
- O'Donnell, M. (2006). Replisome architecture and dynamics in *Escherichia coli*. *J. Biol. Chem.* 281, 10653–10656. doi: 10.1074/jbc.R500028200
- O'Donnell, M., Langston, L., and Stillman, B. (2013). Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb. Perspect. Biol.* 5:a010108. doi: 10.1101/cshperspect.a010108
- O'Donnell, M., and Studwell, P. S. (1990). Total reconstitution of DNA polymerase III holoenzyme reveals dual accessory protein clamps. *J. Biol. Chem.* 265, 1179–1187.
- Ogura, T., Niki, H., Kano, Y., Imamoto, F., and Hiraga, S. (1990). Maintenance of plasmids in HU and IHF mutants of *Escherichia coli*. *Mol. Gen. Genet.* 220, 197–203. doi: 10.1007/BF00260482
- Ortega-Jiménez, S., Giraldo-Suárez, R., Fernández-Tresguerres, M. E., Berzal-Herranz, A., and Díaz-Orejas, R. (1992). DnaA dependent replication of plasmid R1 occurs in the presence of point mutations that disrupt the dnaA box of oriR. *Nucleic Acids Res.* 20, 2547–2551. doi: 10.1093/nar/20.10.2547
- Ozaki, S., and Katayama, T. (2009). DnaA structure, function, and dynamics in the initiation at the chromosomal origin. *Plasmid* 62, 71–82. doi: 10.1016/j.plasmid.2009.06.003
- Ozaki, S., and Katayama, T. (2012). Highly organized DnaA-oriC complexes recruit the single-stranded DNA for replication initiation. *Nucleic Acids Res.* 40, 1648–1665. doi: 10.1093/nar/gkr832
- Ozaki, S., Kawakami, H., Nakamura, K., Fujikawa, N., Kagawa, W., Park, S. Y., et al. (2008). A common mechanism for the ATP-DnaA-dependent formation of open complexes at the replication origin. *J. Biol. Chem.* 283, 8351–8362. doi: 10.1074/jbc.M708684200
- Ozaki, S., Noguchi, Y., Hayashi, Y., Miyazaki, E., and Katayama, T. (2012a). Differentiation of the DnaA-oriC subcomplex for DNA unwinding in a replication initiation complex. *J. Biol. Chem.* 287, 37458–37471. doi: 10.1074/jbc.M112.372052
- Ozaki, S., Noguchi, Y., Nishimura, M., and Katayama, T. (2012b). Stable nucleotide binding to DnaA requires a specific glutamic acid residue within the AAA+ box II motif. *J. Struct. Biol.* 179, 242–250. doi: 10.1016/j.jsb.2012.05.001
- Pacek, M., Konopa, G., and Konieczny, I. (2001). DnaA box sequences as the site for helicase delivery during plasmid RK2 replication initiation in *Escherichia coli*. *J. Biol. Chem.* 276, 23639–23644. doi: 10.1074/jbc.M100255200
- Page, D. T., Whelan, K. F., and Colleran, E. (2001). Characterization of two autoreplicative regions of the IncHI2 plasmid R478: RepHI2A and RepHI1A((R478)). *Microbiology* 147, 1591–1598. doi: 10.1099/00221287-147-6-1591
- Park, K., Mukhopadhyay, S., and Chatteraj, D. K. (1998). Requirements for and regulation of origin opening of plasmid P1. *J. Biol. Chem.* 273, 24906–24911. doi: 10.1074/jbc.273.38.24906
- Park, M. S., and O'Donnell, M. (2009). The clamp loader assembles the beta clamp onto either a 3' or 5' primer terminus: the underlying basis favoring 3' loading. *J. Biol. Chem.* 284, 31473–31483. doi: 10.1074/jbc.M109.050310
- Paschall, C. O., Thompson, J. A., Marzahn, M. R., Chiraniya, A., Hayner, J. N., O'Donnell, M., et al. (2011). The *Escherichia coli* clamp loader can actively pry open the beta-sliding clamp. *J. Biol. Chem.* 286, 42704–42714. doi: 10.1074/jbc.M111.268169
- Perri, S., and Helinski, D. R. (1993). DNA sequence requirements for interaction of the RK2 replication initiation protein with plasmid origin repeats. *J. Biol. Chem.* 268, 3662–3669.
- Perri, S., Helinski, D. R., and Toukdarian, A. (1991). Interactions of plasmid-encoded replication initiation proteins with the origin of DNA replication in the broad host range plasmid RK2. *J. Biol. Chem.* 266, 12536–12543.
- Pierrehod, M., Nowak, A., Saari, A., Purta, E., Bujnicki, J. M., and Konieczny, I. (2009). Conformation of a plasmid replication initiator protein affects its proteolysis by ClpXP system. *Protein Sci.* 18, 637–649. doi: 10.1002/pro.68
- Pietroni, P., and von Hippel, P. H. (2008). Multiple ATP binding is required to stabilize the “activated” (clamp open) clamp loader of the T4 DNA replication complex. *J. Biol. Chem.* 283, 28338–28353. doi: 10.1074/jbc.M804371200
- Podobnik, M., Weitze, T. F., O'Donnell, M., and Kuriyan, J. (2003). Nucleotide-induced conformational changes in an isolated *Escherichia coli* DNA polymerase III clamp loader subunit. *Structure* 11, 253–263. doi: 10.1016/S0969-2126(03)00027-3
- Polaczek, P. (1990). Bending of the origin of replication of *E. coli* by binding of IHF at a specific site. *New Biol.* 2, 265–271.
- Punchihewa, C., Inoue, A., Hishiki, A., Fujikawa, Y., Connelly, M., Evison, B., et al. (2012). Identification of small molecule proliferating cell nuclear antigen (PCNA) inhibitor that disrupts interactions with PIP-box proteins and inhibits DNA replication. *J. Biol. Chem.* 287, 14289–14300. doi: 10.1074/jbc.M112.353201
- Rahn-Lee, L., Merrikh, H., Grossman, A. D., and Losick, R. (2011). The sporulation protein SirA inhibits the binding of DnaA to the origin of replication by contacting a patch of clustered amino acids. *J. Bacteriol.* 193, 1302–1307. doi: 10.1128/JB.01390-10
- Rajewska, M., Kowalczyk, L., Konopa, G., and Konieczny, I. (2008). Specific mutations within the AT-rich region of a plasmid replication origin affect either origin opening or helicase loading. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11134–11139. doi: 10.1073/pnas.0805662105
- Rajewska, M., Wegrzyn, K., and Konieczny, I. (2012). AT-rich region and repeated sequences - the essential elements of replication origins of bacterial replicons. *FEMS Microbiol. Rev.* 36, 408–434. doi: 10.1111/j.1574-6976.2011.03000.x
- Rakowski, S. A., and Filutowicz, M. (2013). Plasmid R6K replication control. *Plasmid* 69, 231–242. doi: 10.1016/j.plasmid.2013.02.003
- Rangarajan, S., Woodgate, R., and Goodman, M. F. (1999). A phenotype for enigmatic DNA polymerase II: a pivotal role for pol II in replication restart in UV-irradiated *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9224–9229. doi: 10.1073/pnas.96.16.9224
- Ratnakar, P. V., Mohanty, B. K., Lobert, M., and Bastia, D. (1996). The replication initiator protein pi of the plasmid R6K specifically interacts with the host-encoded helicase DnaB. *Proc. Natl. Acad. Sci. U.S.A.* 93, 5522–5526. doi: 10.1073/pnas.93.11.5522
- Reyes-Lamothe, R., Possoz, C., Danilova, O., and Sherratt, D. J. (2008). Independent positioning and action of *Escherichia coli* replisomes in live cells. *Cell* 133, 90–102. doi: 10.1016/j.cell.2008.01.044
- Reyes-Lamothe, R., Sherratt, D. J., and Leake, M. C. (2010). Stoichiometry and architecture of active DNA replication machinery in *Escherichia coli*. *Science* 328, 498–501. doi: 10.1126/science.1185757
- Richardson, T. T., Harran, O., and Murray, H. (2016). The bacterial DnaA-trio replication origin element specifies single-stranded DNA initiator binding. *Nature* 534, 412–416. doi: 10.1038/nature17962



- Roth, A., and Messer, W. (1995). The DNA binding domain of the initiator protein DnaA. *EMBO J.* 14, 2106–2111.
- Rozgaja, T. A., Grimwade, J. E., Iqbal, M., Czerwonka, C., Vora, M., and Leonard, A. C. (2011). Two oppositely oriented arrays of low-affinity recognition sites in *oriC* guide progressive binding of DnaA during *Escherichia coli* pre-RC assembly. *Mol. Microbiol.* 82, 475–488. doi: 10.1111/j.1365-2958.2011.07827.x
- Ryan, V. T., Grimwade, J. E., Nievera, C. J., and Leonard, A. C. (2002). IHF and HU stimulate assembly of pre-replication complexes at *Escherichia coli* *oriC* by two different mechanisms. *Mol. Microbiol.* 46, 113–124. doi: 10.1046/j.1365-2958.2002.03129.x
- Samitt, C. E., Hansen, F. G., Miller, J. F., and Schaechter, M. (1989). *In vivo* studies of DnaA binding to the origin of replication of *Escherichia coli*. *EMBO J.* 8, 989–993.
- Saxena, R., Vasudevan, S., Patil, D., Ashoura, N., Grimwade, J. E., and Crooke, E. (2015). Nucleotide-induced conformational changes in *Escherichia coli* DnaA protein are required for bacterial ORC to Pre-RC conversion at the chromosomal origin. *Int. J. Mol. Sci.* 16, 27897–27911. doi: 10.3390/ijms161126064
- Seitz, H., Weigel, C., and Messer, W. (2000). The interaction domains of the DnaA and DnaB replication proteins of *Escherichia coli*. *Mol. Microbiol.* 37, 1270–1279. doi: 10.1046/j.1365-2958.2000.02096.x
- Sekimizu, K., Bramhill, D., and Kornberg, A. (1987). ATP activates dnaA protein in initiating replication of plasmids bearing the origin of the *E. coli* chromosome. *Cell* 50, 259–265. doi: 10.1016/0092-8674(87)90221-2
- Shah, D. S., Cross, M. A., Porter, D., and Thomas, C. M. (1995). Dissection of the core and auxiliary sequences in the vegetative replication origin of promiscuous plasmid RK2. *J. Mol. Biol.* 254, 608–622. doi: 10.1006/jmbi.1995.0642
- Shaheen, S. M., Ouimet, M. C., and Marczyński, G. T. (2009). Comparative analysis of *Caulobacter* chromosome replication origins. *Microbiology* 155, 1215–1225. doi: 10.1099/mic.0.025528-0
- Sharma, R., Kachroo, A., and Bastia, D. (2001). Mechanistic aspects of DnaA-RepA interaction as revealed by yeast forward and reverse two-hybrid analysis. *EMBO J.* 20, 4577–4587. doi: 10.1093/emboj/20.16.4577
- Sharma, S., Sathyanarayana, B. K., Bird, J. G., Hoskins, J. R., Lee, B., and Wickner, S. (2004). Plasmid P1 RepA is homologous to the F plasmid RepE class of initiators. *J. Biol. Chem.* 279, 6027–6034. doi: 10.1074/jbc.M310917200
- Shereda, R. D., Kozlov, A. G., Lohman, T. M., Cox, M. M., and Keck, J. L. (2008). SSB as an organizer/mobilizer of genome maintenance complexes. *Crit. Rev. Biochem. Mol. Biol.* 43, 289–318. doi: 10.1080/10409230802341296
- Simmons, L. A., Felczak, M., and Kaguni, J. M. (2003). DnaA Protein of *Escherichia coli*: oligomerization at the *E. coli* chromosomal origin is required for initiation and involves specific N-terminal amino acids. *Mol. Microbiol.* 49, 849–858. doi: 10.1046/j.1365-2958.2003.03603.x
- Simonetta, K. R., Kazmirski, S. L., Goedken, E. R., Cantor, A. J., Kelch, B. A., McNally, R., et al. (2009). The mechanism of ATP-dependent primer-template recognition by a clamp loader complex. *Cell* 137, 659–671. doi: 10.1016/j.cell.2009.03.044
- Soni, R. K., Mehra, P., Choudhury, N. R., Mukhopadhyay, G., and Dhar, S. K. (2003). Functional characterization of *Helicobacter pylori* DnaB helicase. *Nucleic Acids Res.* 31, 6828–6840. doi: 10.1093/nar/gkg895
- Soultanas, P. (2012). Loading mechanisms of ring helicases at replication origins. *Mol. Microbiol.* 84, 6–16. doi: 10.1111/j.1365-2958.2012.08012.x
- Soumillion, P., Sexton, D. J., and Benkovic, S. J. (1998). Clamp subunit dissociation dictates bacteriophage T4 DNA polymerase holoenzyme disassembly. *Biochemistry* 37, 1819–1827. doi: 10.1021/bi972526a
- Sozhamannan, S., and Chatteraj, D. K. (1993). Heat shock proteins DnaJ, DnaK, and GrpE stimulate P1 plasmid replication by promoting initiator binding to the origin. *J. Bacteriol.* 175, 3546–3555.
- Speck, C., and Messer, W. (2001). Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. *EMBO J.* 20, 1469–1476. doi: 10.1093/emboj/20.6.1469
- Stalker, D. M., Thomas, C. M., and Helinski, D. R. (1981). Nucleotide sequence of the region of the origin of replication of the broad host range plasmid RK2. *Mol. Gen. Genet.* 181, 8–12. doi: 10.1007/BF00338997
- Stenzel, T. T., Patel, P., and Bastia, D. (1987). The integration host factor of *Escherichia coli* binds to bent DNA at the origin of replication of the plasmid pSC101. *Cell* 49, 709–717. doi: 10.1016/0092-8674(87)90547-2
- Strycharska, M. S., Arias-Palomo, E., Lyubimov, A. Y., Erzberger, J. P., O'Shea, V. L., Bustamante, C. J., et al. (2013). Nucleotide and partner-protein control of bacterial replicative helicase structure and function. *Mol. Cell* 52, 844–854. doi: 10.1016/j.molcel.2013.11.016
- Su, X. C., Wang, Y., Yagi, H., Shishmarev, D., Mason, C. E., Smith, P. J., et al. (2014). Bound or free: interaction of the C-terminal domain of *Escherichia coli* single-stranded DNA-binding protein (SSB) with the tetrameric core of SSB. *Biochemistry* 53, 1925–1934. doi: 10.1021/bi5001867
- Sutton, M. D., Carr, K. M., Vicente, M., and Kaguni, J. M. (1998). *Escherichia coli* DnaA protein. The N-terminal domain and loading of DnaB helicase at the *E. coli* chromosomal origin. *J. Biol. Chem.* 273, 34255–34262. doi: 10.1074/jbc.273.51.34255
- Sutton, M. D., and Kaguni, J. M. (1995). Novel alleles of the *Escherichia coli* dnaA gene are defective in replication of pSC101 but not of *oriC*. *J. Bacteriol.* 177, 6657–6665.
- Sutton, M. D., and Kaguni, J. M. (1997). Novel alleles of the *Escherichia coli* dnaA gene. *J. Mol. Biol.* 271, 693–703. doi: 10.1006/jmbi.1997.1209
- Swan, M. K., Bastia, D., and Davies, C. (2006). Crystal structure of pi initiator protein-iteron complex of plasmid R6K: implications for initiation of plasmid DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* 103, 18481–18486. doi: 10.1073/pnas.0609046103
- Taft-Benz, S. A., and Schaaper, R. M. (2004). The theta subunit of *Escherichia coli* DNA polymerase III: a role in stabilizing the epsilon proofreading subunit. *J. Bacteriol.* 186, 2774–2780. doi: 10.1128/JB.186.9.2774-2780.2004
- Takechi, S., and Itoh, T. (1995). Initiation of unidirectional ColE2 DNA replication by a unique priming mechanism. *Nucleic Acids Res.* 23, 4196–4201. doi: 10.1093/nar/23.20.4196
- Takechi, S., Matsui, H., and Itoh, T. (1995). Primer RNA synthesis by plasmid-specified Rep protein for initiation of ColE2 DNA replication. *EMBO J.* 14, 5141–5147.
- Taylor, J. A., Ouimet, M. C., Wargachuk, R., and Marczyński, G. T. (2011). The *Caulobacter crescentus* chromosome replication origin evolved two classes of weak DnaA binding sites. *Mol. Microbiol.* 82, 312–326. doi: 10.1111/j.1365-2958.2011.07785.x
- Thomas, C. M., and Sherlock, S. (1990). Broad host range plasmid RK2 encodes a polypeptide related to single-stranded DNA binding protein (SSB) of *Escherichia coli*. *Nucleic Acids Res.* 18, 2812. doi: 10.1093/nar/18.9.2812
- Trakselis, M. A., Alley, S. C., Abel-Santos, E., and Benkovic, S. J. (2001). Creating a dynamic picture of the sliding clamp during T4 DNA polymerase holoenzyme assembly by using fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. U.S.A.* 98, 8368–8375. doi: 10.1073/pnas.111006698
- Tsodikov, O. V., and Biswas, T. (2011). Structural and thermodynamic signatures of DNA recognition by *Mycobacterium tuberculosis* DnaA. *J. Mol. Biol.* 410, 461–476. doi: 10.1016/j.jmb.2011.05.007
- Tsuchihashi, Z., and Kornberg, A. (1989). ATP interactions of the tau and gamma subunits of DNA polymerase III holoenzyme of *Escherichia coli*. *J. Biol. Chem.* 264, 17790–17795.
- Wahle, E., Lasken, R. S., and Kornberg, A. (1989). The dnaB-dnaC replication protein complex of *Escherichia coli*. II. Role of the complex in mobilizing dnaB functions. *J. Biol. Chem.* 264, 2469–2475.
- Wang, G., Klein, M. G., Tokonzaba, E., Zhang, Y., Holden, L. G., and Chen, X. S. (2008). The structure of a DnaB-family replicative helicase and its interactions with primase. *Nat. Struct. Mol. Biol.* 15, 94–100. doi: 10.1038/nsmb1356
- Wawrzycka, A., Gross, M., Wasaznik, A., and Konieczny, I. (2015). Plasmid replication initiator interactions with origin 13-mers and polymerase subunits contribute to strand-specific replisome assembly. *Proc. Natl. Acad. Sci. U.S.A.* 112, E4188–E4196. doi: 10.1073/pnas.1504926112
- Wegrzyn, K., Fuentes-Perez, M. E., Bury, K., Rajewska, M., Moreno-Herrero, F., and Konieczny, I. (2014). Sequence-specific interactions of Rep proteins with ssDNA in the AT-rich region of the plasmid replication origin. *Nucleic Acids Res.* 42, 7807–7818. doi: 10.1093/nar/gku453
- Weigel, C., Schmidt, A., Seitz, H., Tunger, D., Welzeck, M., and Messer, W. (1999). The N-terminus promotes oligomerization of the *Escherichia coli* initiator protein DnaA. *Mol. Microbiol.* 34, 53–66. doi: 10.1046/j.1365-2958.1999.01568.x
- Weigel, C., and Seitz, H. (2002). Strand-specific loading of DnaB helicase by DnaA to a substrate mimicking unwound *oriC*. *Mol. Microbiol.* 46, 1149–1156. doi: 10.1046/j.1365-2958.2002.03232.x

- Wickner, S., Gottesman, S., Skowrya, D., Hoskins, J., McKenney, K., and Maurizi, M. R. (1994). A molecular chaperone, ClpA, functions like DnaK and DnaJ. *Proc. Natl. Acad. Sci. U.S.A.* 91, 12218–12222. doi: 10.1073/pnas.91.25.12218
- Wickner, S., and Hurwitz, J. (1975). Interaction of *Escherichia coli* dnaB and dnaC(D) gene products *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* 72, 921–925. doi: 10.1073/pnas.72.3.921
- Wickner, S., Skowrya, D., Hoskins, J., and McKenney, K. (1992). DnaJ, DnaK, and GrpE heat shock proteins are required in oriP1 DNA replication solely at the RepA monomerization step. *Proc. Natl. Acad. Sci. U.S.A.* 89, 10345–10349. doi: 10.1073/pnas.89.21.10345
- Williams, G. J., Johnson, K., Rudolf, J., McMahon, S. A., Carter, L., Oke, M., et al. (2006). Structure of the heterotrimeric PCNA from *Sulfolobus solfataricus*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 62, 944–948. doi: 10.1107/S1744309106034075
- Wolanski, M., Donczew, R., Zawilak-Pawlik, A., and Zakrzewska-Czerwinska, J. (2014). oriC-encoded instructions for the initiation of bacterial chromosome replication. *Front. Microbiol.* 5:735. doi: 10.3389/fmicb.2014.00735
- Wold, S., Crooke, E., and Skarstad, K. (1996). The *Escherichia coli* Fis protein prevents initiation of DNA replication from oriC *in vitro*. *Nucleic Acids Res.* 24, 3527–3532. doi: 10.1093/nar/24.18.3527
- Wolff, P., Amal, I., Oliéric, V., Chaloin, O., Gygli, G., Ennifar, E., et al. (2014). Differential modes of peptide binding onto replicative sliding clamps from various bacterial origins. *J. Med. Chem.* 57, 7565–7576. doi: 10.1021/jm500467a
- Xia, G., Manen, D., Yu, Y., and Caro, L. (1993). *In vivo* and *in vitro* studies of a copy number mutation of the RepA replication protein of plasmid pSC101. *J. Bacteriol.* 175, 4165–4175.
- Yano, H., Rogers, L. M., Knox, M. G., Heuer, H., Smalla, K., Brown, C. J., et al. (2013). Host range diversification within the IncP-1 plasmid group. *Microbiology* 159, 2303–2315. doi: 10.1099/mic.0.068387-0
- Yano, H., Wegrzyn, K., Loftie-Eaton, W., Johnson, J., Deckert, G. E., Rogers, L. M., et al. (2016). Evolved plasmid-host interactions reduce plasmid interference cost. *Mol. Microbiol.* doi: 10.1111/mmi.13407. [Epub ahead of print].
- Yasueda, H., Horii, T., and Itoh, T. (1989). Structural and functional organization of ColE2 and ColE3 replicons. *Mol. Gen. Genet.* 215, 209–216. doi: 10.1007/BF00339719
- Yuzhakov, A., Kelman, Z., and O'Donnell, M. (1999). Trading places on DNA—a three-point switch underlies primer handoff from primase to the replicative DNA polymerase. *Cell* 96, 153–163. doi: 10.1016/S0092-8674(00)80968-X
- Yuzhakov, A., Turner, J., and O'Donnell, M. (1996). Replisome assembly reveals the basis for asymmetric function in leading and lagging strand replication. *Cell* 86, 877–886. doi: 10.1016/S0092-8674(00)80163-4
- Zawilak-Pawlik, A., Kois, A., Stingl, K., Boneca, I. G., Skrobuk, P., Piotr, J., et al. (2007). HobA—a novel protein involved in initiation of chromosomal replication in *Helicobacter pylori*. *Mol. Microbiol.* 65, 979–994. doi: 10.1111/j.1365-2958.2007.05853.x
- Zhong, Z., Helinski, D., and Toukdarian, A. (2003). A specific region in the N terminus of a replication initiation protein of plasmid RK2 is required for recruitment of *Pseudomonas aeruginosa* DnaB helicase to the plasmid origin. *J. Biol. Chem.* 278, 45305–45310. doi: 10.1074/jbc.M306058200
- Zhong, Z., Helinski, D., and Toukdarian, A. (2005). Plasmid host-range: restrictions to F replication in *Pseudomonas*. *Plasmid* 54, 48–56. doi: 10.1016/j.plasmid.2004.11.001
- Zzaman, S., Abhyankar, M. M., and Bastia, D. (2004). Reconstitution of F factor DNA replication *in vitro* with purified proteins. *J. Biol. Chem.* 279, 17404–17410. doi: 10.1074/jbc.M400021200

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Wegrzyn, Gross, Uciechowska and Konieczny. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# ParB Partition Proteins: Complex Formation and Spreading at Bacterial and Plasmid Centromeres

Barbara E. Funnell \*

Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Spanish National Research Council -  
Centro de Investigaciones Biológicas,  
Spain

### Reviewed by:

Christopher Morton Thomas,  
University of Birmingham, UK  
Jean-Yves Bouet,  
Centre National de la Recherche  
Scientifique, France

### \*Correspondence:

Barbara Funnell  
b.funnell@utoronto.ca

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 27 June 2016

**Accepted:** 15 August 2016

**Published:** 29 August 2016

### Citation:

Funnell BE (2016) ParB Partition  
Proteins: Complex Formation and  
Spreading at Bacterial and Plasmid  
Centromeres. *Front. Mol. Biosci.* 3:44.  
doi: 10.3389/fmolb.2016.00044

In bacteria, active partition systems contribute to the faithful segregation of both chromosomes and low-copy-number plasmids. Each system depends on a site-specific DNA binding protein to recognize and assemble a partition complex at a centromere-like site, commonly called *parS*. Many plasmid, and all chromosomal centromere-binding proteins are dimeric helix-turn-helix DNA binding proteins, which are commonly named ParB. Although the overall sequence conservation among ParBs is not high, the proteins share similar domain and functional organization, and they assemble into similar higher-order complexes. *In vivo*, ParBs “spread,” that is, DNA binding extends away from the *parS* site into the surrounding non-specific DNA, a feature that reflects higher-order complex assembly. ParBs bridge and pair DNA at *parS* and non-specific DNA sites. ParB dimers interact with each other via flexible conformations of an N-terminal region. This review will focus on the properties of the HTH centromere-binding protein, in light of recent experimental evidence and models that are adding to our understanding of how these proteins assemble into large and dynamic partition complexes at and around their specific DNA sites.

**Keywords:** chromosome dynamics, segregation, ParABS, DNA-binding, bridging

In bacteria, the segregation, or partition, of low-copy-number plasmids and cellular chromosomes depends on the activity of site-specific DNA binding proteins to recognize one or more copies of a centromere-like DNA site. These “centromere-binding proteins” generally work in concert with an ATPase or GTPase, resulting in dynamic movement and positioning of plasmids or chromosomal domains during the cell cycle (reviewed in Wang et al., 2013; Baxter and Funnell, 2014; Bouet et al., 2014). Centromere-binding proteins fall into one of two structural classes, as helix-turn-helix (HTH), or ribbon-helix-helix site-specific DNA binding proteins. In bacteria, the proteins of all chromosomal, and many plasmid partition systems are members of the HTH class. They share similar properties *in vivo* and *in vitro*, including similar domain organization and DNA-binding properties, although there are also interesting differences. They all form large partition complexes *in vivo* that can be visualized as foci using fluorescence approaches. This review will focus on the properties of the HTH centromere-binding proteins, and in particular, how they assemble into large partition complexes. I will discuss the contribution of protein domains, the “spreading” phenomenon that has been reported as a general property of HTH centromere-binding proteins, and how flexibility in the protein allows multiple conformations and binding modes in complex assembly.

The components of partition systems are commonly named ParA (the partition ATPase), ParB (the centromere-binding protein), and *parS* (the centromere or partition site). Plasmids typically

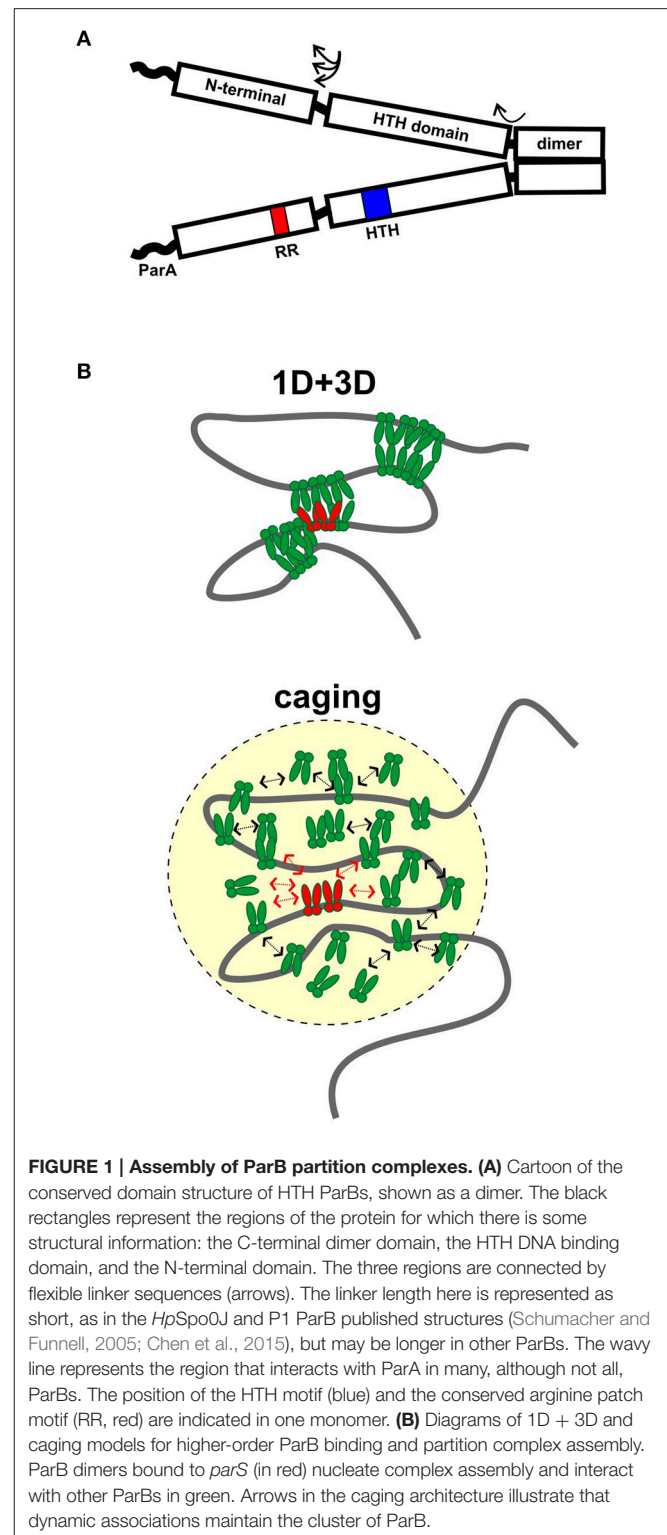
contain one *parS* located near the *parA* and *parB* genes, although some have multiple *parS* sites. Bacterial chromosomes contain several *parS* sites, which are primarily located in the chromosomal domain that contains the replication origin. Bacterial ParA and ParB are also called Soj and Spo0J, respectively, because their genes were first defined by roles in sporulation of *Bacillus subtilis* (Iretton et al., 1994). For simplicity, I will use the ParABS nomenclature, with specific names for some of the discussion to be consistent with published literature.

HTH ParBs share a similar domain organization although the primary sequence conservation is not high among members of this family. In general, the protein is divided into three regions: A central HTH DNA binding domain is flanked by a C-terminal dimer domain and an N-terminal region necessary for protein oligomerization (Figure 1A). Flexible linkers connect the domains, and flexibility in domain organization, orientation, and folding have been observed in biochemical experiments and crystal structures. The most highly conserved sequence among plasmid and chromosomal ParBs is a short arginine-rich motif in the N-terminus, which is often called an arginine patch (Yamaichi and Niki, 2000). ParA interactions are often specified by residues near the N-terminus of ParB (Radnedge et al., 1998; Figge et al., 2003; Leonard et al., 2005; Ah-Seng et al., 2009), although there are exceptions. For example, the ParA interactions for RK2 KorB and *Pseudomonas aeruginosa* ParB map to the center and dimer domain, respectively (Lukaszewicz et al., 2002; Bartosik et al., 2004). There are also added complexities to the general arrangement. For example, P1 ParB and its relatives contain an additional site-specific DNA binding activity within the dimer domain. These ParBs recognize both an inverted repeat and a second DNA motif in their *parS* sites, via their HTH domain and dimer domains, respectively (Schumacher and Funnell, 2005).

## PARTITION COMPLEX ASSEMBLY, SPREADING, AND BRIDGING

Partition complex assembly begins with the recognition of *parS* by a dimer of ParB, followed by loading of multiple ParB dimers to form a very large protein-DNA complex (Baxter and Funnell, 2014). These higher-order complexes are necessary as both the substrates and the activators of the mechanisms of partition. The number of ParB protein foci observed inside cells is usually lower than the number of *parS* sites, leading to the idea that inter and intra-molecular pairing of *parS* sites occurs in plasmids and chromosomal domains.

Spreading is an unusual feature for site-specific DNA binding proteins that is common to HTH ParBs, and it reflects how ParB assembles into higher-order complexes (Rodionov et al., 1999; Murray et al., 2006; Breier and Grossman, 2007; Sanchez et al., 2015). Measured by ChIP approaches, *in vivo* ParB binding extends beyond *parS* into the surrounding non-specific DNA, often many kb away from the site. Binding is maximal at *parS*, and diminishes non-linearly as a function of distance from *parS*. Spreading can be impeded by “roadblocks,” which are strong binding sites for other proteins (Rodionov et al., 1999; Murray et al., 2006). Spreading has also been inferred from the ability of



some ParBs, especially when overexpressed, to silence expression of nearby genes (Lynch and Wang, 1995; Rodionov et al., 1999; Bartosik et al., 2004; Bingle et al., 2005; Kusiak et al., 2011). Silencing is likely a consequence of protein overexpression and is not necessary for partition (Rodionov and Yarmolinsky, 2004).



Spreading ability is required however, as ParB mutants that do not spread are defective in partition (Rodionov et al., 1999; Autret et al., 2001; Breier and Grossman, 2007; Kusiak et al., 2011; Graham et al., 2014). In particular, the arginine patch in ParB is essential for spreading, focus formation, and partition activity *in vivo*.

The first and simplest model described spreading as lateral protein-protein association along the DNA as a one-dimensional filament (Rodionov et al., 1999). However some properties of ParBs lead investigators to question this idea. Studies with plasmid KorB and with *B. subtilis* Spo0J (*BsSpo0J*) argued that the intracellular concentration of ParB was insufficient to account for the amount of spreading observed *in vivo* if arranged as a one-dimensional filament (Bingle et al., 2005; Graham et al., 2014). It was also difficult to demonstrate biochemically that ParB binding to *parS* increased the affinity of ParB for adjacent non-specific DNA.

Two other models have emerged recently, based on sophisticated microscope and ChIP-seq technologies as well as computer modeling and traditional biochemistry (Broedersz et al., 2014; Graham et al., 2014; Sanchez et al., 2015). The first proposes that limited lateral ParB-ParB interactions (1D) in combination with inter and intra-molecular looping and bridging (3D) act to build large complexes and coalesce many ParB molecules into foci (Broedersz et al., 2014; Graham et al., 2014; **Figure 1B**). Computer modeling was used to argue that neither 1D nor 3D interactions alone could generate ParB foci; that the 1D + 3D arrangement allows focus formation because it creates a surface tension on the ParB cluster that counteracts the tendency for entropy to disperse the protein on DNA (Broedersz et al., 2014). Elegant *in vitro* TIRF microscopy experiments provided experimental support for the bridging activity: flow-stretched DNA was condensed by *BsSpo0J* in a manner that is most consistent with ParB bridging across loops within the same or across different DNA molecules (Graham et al., 2014). The experiments were however unable to demonstrate any sequence-specificity for *parS*, leading to the suggestion that experiments on flow-stretched DNA were not recapitulating an undefined aspect of critical nucleation properties of ParB bound to *parS*. For example, one factor missing from these experiments is DNA supercoiling, which affects chromosome compaction and may strongly influence the DNA binding properties of ParB in higher-order complexes.

Mutations in conserved arginine residues of the arginine patch motif eliminated bridging in the TIRF assay, consistent with the requirement for this motif for spreading and partition *in vivo* (Graham et al., 2014). Interestingly, *BsSpo0J* mutated at one residue in the arginine patch (G77S), which is unable to form foci or spread *in vivo*, was still able to bridge DNA, and with slightly higher stability than that of wild-type *BsSpo0J*. Therefore the bridging activity of ParB is necessary but not sufficient for complex formation. These observations lead to the suggestion that the G77S mutation may promote inappropriate bridging and/or alter the dynamics of bridging necessary for proper complex assembly *in vivo*. Modeling the spreading/bridging behavior also predicted that roadblocks would decrease the

probability of loops forming in their vicinity, interfering with complex assembly beyond the roadblock (Broedersz et al., 2014).

In contrast, a second model proposes that a network of stochastic binding of ParB explains the clustering of ParB molecules around *parS* (Sanchez et al., 2015). In essence, the nucleation of ParB by *parS* creates, and maintains a very high localized concentration of ParB in a “cage” by many weak but dynamic interactions with itself (dimer-dimer interactions via the N-terminal domains), as well as with non-specific DNA around *parS* (**Figure 1B**). In caging, these interactions do not need to occur simultaneously or to bridge DNA (**Figure 1B**). Computer modeling of the patterns of ParB occupancy around *parS* measured by ChIP was used to argue that they are not consistent with either 1D lateral spreading or a combination of 1D spreading and 3D bridging. Biochemical examination showed no evidence that binding of one ParB could stabilize binding of an adjacent ParB. The caging model neither requires nor excludes bridging, although bridging interactions are intuitively attractive as part of the dynamic glue. The model does depend on other properties of the DNA chromosome, such as topology or organization by other nucleoid-binding proteins *in vivo* to help restrict the DNA within the cage. Roadblocks could alter local DNA organization and reduce the proximity of *parS* to the rest of the DNA in three-dimensional space; that is, place this DNA outside the cage.

Both models agree that ParB binding to *parS* must nucleate the formation of higher-order complexes to explain ParB clustering and foci *in vivo* (Broedersz et al., 2014; Sanchez et al., 2015). This is most simple to envision in plasmid systems such as P1, in which ParB's affinity for *parS* is at least 10,000-fold higher than that for non-specific DNA (Funnell, 1991). However this affinity difference is small for some ParBs (Broedersz et al., 2014; Taylor et al., 2015), and it was suggested that a conformational change is induced in ParB by *parS*-specific binding to effectively anchor the focus at the site. Both models agree that multiple ParB-ParB interaction interfaces must be involved in assembly of higher-order partition complexes, and that the dynamics of these interactions are critical for proper assembly and function in partition. How bridging activities detected *in vitro* contribute to ParB activity *in vivo* remains to be resolved. Further refinement of these models will depend on the ability to completely reconstitute the *parS*-dependent complex assembly *in vitro*, which will in turn depend on identifying the other factors necessary for caging or bridging, and on the nature of ParB-ParB and ParB-DNA interactions at the molecular level. The influence of ParA on complex architecture and dynamics has also yet to be defined (see below).

## ParB-ParB AND ParB-DNA INTERACTIONS IN HIGHER-ORDER-COMPLEX ASSEMBLY

Site-specific DNA binding of ParBs to cognate *parS* sites has been examined directly and in detail in many different partition systems, but the *parS*-dependent formation of higher order, large partition complexes has been difficult to reconstitute *in vitro*. However, there are insights arising from structural biology of

several plasmid and chromosomal ParBs, which are leading to a preliminary, albeit incomplete, picture of partition complex assembly.

Although there are no structures of a full length ParB, those of individual domains or combinations of domains, with and without *parS* DNA, have provided clues concerning the three dimensional organization of the protein with respect to DNA and to itself. There are structures of the HTH domains of three plasmid ParBs (P1 ParB, F SopB, and RP4 KorB) in complex with their specific DNA sites, and of their dimer domains (SopB and KorB dimer domains solved separately from the HTH; Delbrück et al., 2002; Khare et al., 2004; Schumacher and Funnell, 2005; Schumacher et al., 2010). Structures of two chromosomal ParB fragments, each containing the N-terminal region and adjacent HTH domain, have visualized the oligomerization interactions of the proteins. These ParBs are *Thermus thermophilus* Spo0J (*TtSpo0J*) and *Helicobacter pylori* Spo0J (*HpSpo0J*); structure of the latter was solved bound to *parS* (Leonard et al., 2004; Chen et al., 2015).

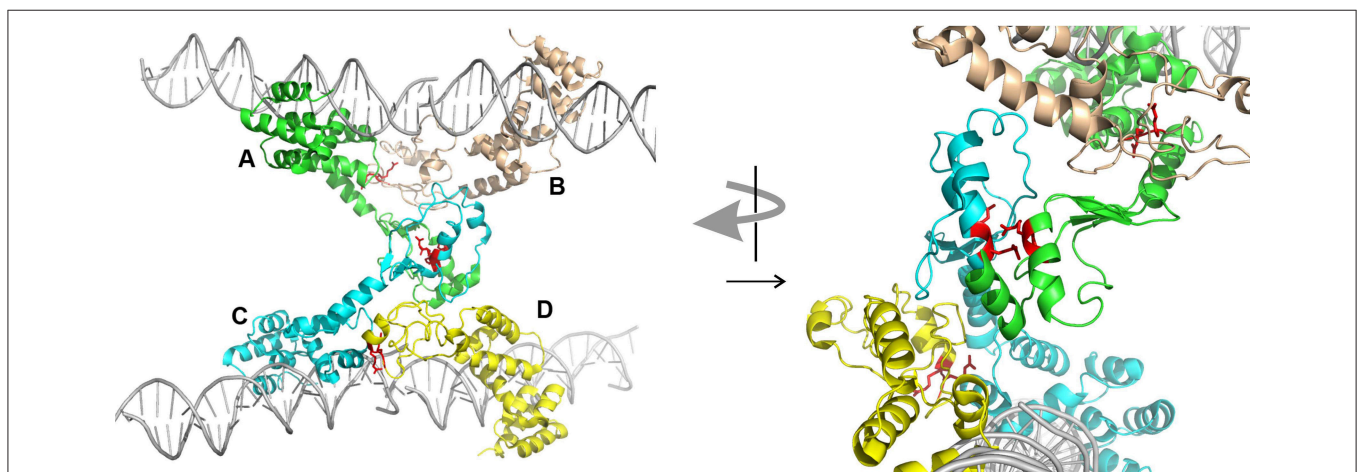
One of the first themes to highlight is that of flexibility. Taken together, the structures indicate that these three regions of the protein are connected by flexible linkers, and that their orientation with respect to each other can vary. The conformation of the N-terminus is particularly flexible (Chen et al., 2015).

Second, ParBs are bona-fide HTH site-specific DNA binding proteins, but with a twist. As expected, the HTH domains contact inverted repeat sequences within *parS* via helix insertion into the major groove of DNA. However, unexpected features emerged from the structures. First, residues outside of the recognition helix also contribute to specificity for *parS* in SopB, KorB, and *HpSpo0J* (Khare et al., 2004; Schumacher et al., 2010; Sanchez et al., 2013; Chen et al., 2015). Second, both the P1 ParB and *HpSpo0J* structures demonstrated bridging across *parS* sites mediated by the dimer and N-terminal domains,

respectively (Schumacher and Funnell, 2005; Chen et al., 2015). Each monomer of a P1 ParB dimer interacts with a half-site on a different DNA molecule, effectively pairing two *parS* sites. *HpSpo0J-parS* bridging interactions are mediated by the N-terminal oligomerization regions of the protein (Figure 2). Four monomers of *HpSpo0J* (monomeric because it lacks the C-terminal dimer domain) interact with two *parS* oligos and with each other in a cross-bridge arrangement across the DNA molecules (molecules A–D in Figure 2). The monomers share a common HTH domain, but show different conformations in the extended N-terminal regions as well as different interactions with each other. There are two adjacent (AB and CD) and one transverse (AC) sets of protein-protein interactions, which are distinct. For example, the conserved arginine patch motifs are close to each other at the AC interface, but not at the AB or CD interfaces (Figure 2). In the structure, there is no BD interaction, leaving these surfaces available, perhaps for interactions with different conformations of ParB or with different partners.

The overall fold of *HpSpo0J* is similar to that of *TtSpo0J*, except for a bend in the linker between the N-terminal and HTH domains (Leonard et al., 2004; Chen et al., 2015). It was suggested that the *TtSpo0J* structure may represent a closed conformation that opens up following DNA binding to the *HpSpo0J* architecture.

How do these structures inform us of higher-order complex assembly, particularly when ParB binds, bridges, and spreads on non-specific DNA adjacent to and away from *parS*? The simplest model is that the HTH is responsible for both specific and non-specific DNA interactions, which is supported by the observation that a triple substitution in the HTH domain of F SopB impairs both DNA binding activities (Ah-Seng et al., 2009). In this case the *HpSpo0J*-DNA structure may represent ParB-ParB and ParB-DNA interactions during spreading at any DNA site. The requirement for the N-terminus in higher-order complex formation *in vivo* is also consistent with this picture.



**FIGURE 2 | Structure of *HpSpo0J* monomers (lacking C-terminal dimer domain) bound to and across *parS* DNA (PDB 4UMK, Chen et al., 2015).** Four monomers (A to D) make adjacent (AB and CD) and transverse (AC) interactions. The arginine patch motif is illustrated by two red arginines from each monomer. The arrangement on the left is rotated approximately 90° and magnified on the right to illustrate the environments of these arginines in the different interactions. The images were generated using the PyMOL Molecular Graphics System, Version 1.8.2.1 Schrödinger, LLC.

The flexibility of and variation in the cross-monomer interactions and interfaces seen in the *HpSpo0J* structure make it an attractive model for the ability of ParB to make multiple and flexible interactions during complex assembly and maintenance. However, one recent study suggests that the specific and non-specific DNA binding activities may be distinct (Taylor et al., 2015). The specific and non-specific DNA binding activities of *BsSpo0J* showed different properties *in vitro*, including different abilities to protect the HTH domain from proteolysis. The results lead to the proposal that the N-terminus contains a DNA binding region that is distinct from the HTH. These observations may also reflect differences between plasmid and chromosomal ParBs. We must await identification of the protein-DNA contacts at non-specific sites before we can confirm the organization of ParB during higher-order partition complex assembly.

## THE ROLE OF ParA IN ParB-DNA COMPLEXES

During partition, ParAs form patterns on the surface of the bacterial nucleoid due to dynamic interactions with ParB bound to *parS* (Hatano et al., 2007; Ringgaard et al., 2009; Hatano and Niki, 2010; Ah-Seng et al., 2013). The patterning is necessary for the segregation of plasmids and chromosomal domains, and the molecular mechanisms involved are still being defined. ParA is not necessary to form the large ParB-DNA complexes seen *in vivo* and *in vitro*, but ParA can influence or modulate these complexes. For example, the behavior of several *parA* and *parB* mutants supports a proposal that ParA is necessary to separate pairs or groups of plasmids during segregation (Fung et al., 2001; Ah-Seng et al., 2013). For the ParBs that interact with ParA via N-terminal regions that are adjacent to the flexible ParB-ParB oligomerization interface, one attractive idea is that ParA-ParB interactions at the N-terminus may influence the available conformations for ParB-ParB interactions next door. For example, specific interference with the transverse ParB-ParB interactions in the *HpSpo0J* structure might favor intramolecular associations over intermolecular ones (Figure 2).

## REFERENCES

- Ah-Seng, Y., Lopez, F., Pasta, F., Lane, D., and Bouet, J.-Y. (2009). Dual role of DNA in regulating ATP hydrolysis by the SopA partition protein. *J. Biol. Chem.* 284, 30067–30075. doi: 10.1074/jbc.M109.044800
- Ah-Seng, Y., Rech, J., Lane, D., and Bouet, J. Y. (2013). Defining the role of ATP hydrolysis in mitotic segregation of bacterial plasmids. *PLoS Genet* 9:e1003956. doi: 10.1371/journal.pgen.1003956
- Autret, S., Nair, R., and Errington, J. (2001). Genetic analysis of the chromosome segregation protein Spo0J of *Bacillus subtilis*: evidence for separate domains involved in DNA binding and interactions with Soj protein. *Mol. Microbiol.* 41, 743–755. doi: 10.1046/j.1365-2958.2001.02551.x
- Bartosik, A. A., Lasocki, K., Mierzejewska, J., Thomas, C. M., and Jagura-Burdzy, G. (2004). ParB of *Pseudomonas aeruginosa*: interactions with its partner ParA and its target *parS* and specific effects on bacterial growth. *J. Bacteriol.* 186, 6983–6998. doi: 10.1128/jb.186.20.6983-6998.2004
- Baxter, J. C., and Funnell, B. E. (2014). Plasmid partition mechanisms. *Microbiol. Spectr.* 2, 6. doi: 10.1128/microbiolspec.PLAS-0023-2014

## FLEXIBILITY AND ORDER IN PARTITION COMPLEX ASSEMBLY

Taken together, the biochemistry and structural biology of the N-terminal regions of ParBs imply that the folding and structures are flexible, dynamic, and fluid. Recent experiments using magnetic tweezers support the idea that the complexes are not highly ordered (Taylor et al., 2015). ParB-ParB (dimer-dimer) interactions via the N-terminal domain must contribute to higher-order complex assembly and function. It is attractive to consider that the flexibility of the N-terminus, in folding and conformation, is important for the dynamics and architecture of the large higher-order partition complex *in vivo*. This conformational flexibility resembles the properties of so-called “intrinsically disordered” proteins and domains whose unstructured properties allow proteins to sample and bind to multiple targets and in multiple ways (Wright and Dyson, 1999; Uversky, 2016). The same region in ParB could be involved, directly or indirectly, in different binding scenarios with different partners, including ParA, and potentially non-specific DNA. Why are flexibility, dynamics, and size important for these partition complexes? ParBs are engaging in multiple and constantly changing interactions during partition. Clustering creates a condensed, organized DNA substrate and provides a high density of ParBs available to ParA. Defining the molecular nature of these interactions continues to be an essential step toward the understanding of these intriguing DNA binding proteins.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## FUNDING

This work was supported by Canadian Institutes of Health Research grant 133613.

- Bingle, L. E. H., Macartney, D. P., Fantozzi, A., Manzoor, S. E., and Thomas, C. M. (2005). Flexibility in repression and cooperativity by KorB of broad host range IncP-1 plasmid RK2. *J. Mol. Biol.* 349, 302–316. doi: 10.1016/j.jmb.2005.03.062
- Bouet, J. Y., Stouf, M., Lebailly, E., and Cornet, F. (2014). Mechanisms for chromosome segregation. *Curr. Opin. Microbiol.* 22, 60–65. doi: 10.1016/j.mib.2014.09.013
- Breier, A. M., and Grossman, A. D. (2007). Whole-genome analysis of the chromosome partitioning and sporulation protein Spo0J (ParB) reveals spreading and origin-distal sites on the *Bacillus subtilis* chromosome. *Mol. Microbiol.* 64, 703–718. doi: 10.1111/j.1365-2958.2007.05690.x
- Broedersz, C. P., Wang, X. D., Meir, Y., Loparo, J. J., Rudner, D. Z., and Wingreen, N. S. (2014). Condensation and localization of the partitioning protein ParB on the bacterial chromosome. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8809–8814. doi: 10.1073/pnas.1402529111
- Chen, B.-W., Lin, M.-H., Chu, C.-H., Hsu, C.-E., and Sun, Y.-J. (2015). Insights into ParB spreading from the complex structure of Spo0J and *parS*. *Proc. Natl. Acad. Sci. U.S.A.* 112, 6613–6618. doi: 10.1073/pnas.1421927112

- Delbrück, H., Ziegelin, G., Lanka, E., and Heinemann, U. (2002). An Src Homology 3-like domain is responsible for dimerization of the repressor protein KorB encoded by the promiscuous IncP plasmid RP4. *J. Biol. Chem.* 277, 4191–4198. doi: 10.1074/jbc.M110103200
- Figge, R. M., Easter, J., and Gober, J. W. (2003). Productive interaction between the chromosome partitioning proteins, ParA and ParB, is required for the progression of the cell cycle in *Caulobacter crescentus*. *Mol. Microbiol.* 47, 1225–1237. doi: 10.1046/j.1365-2958.2003.03367.x
- Fung, E., Bouet, J.-Y., and Funnell, B. E. (2001). Probing the ATP-binding site of P1 ParA: partition and repression have different requirements for ATP binding and hydrolysis. *EMBO J.* 20, 4901–4911. doi: 10.1093/emboj/20.17.4901
- Funnell, B. E. (1991). The P1 partition complex at *parS*: the influence of *Escherichia coli* integration host factor and of substrate topology. *J. Biol. Chem.* 266, 14328–14337.
- Graham, T. G. W., Wang, X., Song, D., Eton, C. M., van Oijen, A. M., Rudner, D. Z., et al. (2014). ParB spreading requires DNA bridging. *Genes Dev.* 28, 1228–1238. doi: 10.1101/gad.242206.114
- Hatano, T., and Niki, H. (2010). Partitioning of P1 plasmids by gradual distribution of the ATPase ParA. *Mol. Microbiol.* 78, 1182–1198. doi: 10.1111/j.1365-2958.2010.07398.x
- Hatano, T., Yamaichi, Y., and Niki, H. (2007). Oscillating focus of SopA associated with filamentous structure guides partitioning of F plasmid. *Mol. Microbiol.* 64, 1198–1213. doi: 10.1111/j.1365-2958.2007.05728.x
- Ireton, K., Gunther, N. W., and Grossman, A. D. (1994). *spoJ* is required for normal chromosome segregation as well as the initiation of sporulation in *Bacillus subtilis*. *J. Bacteriol.* 176, 5320–5329.
- Khare, D., Ziegelin, G., Lanka, E., and Heinemann, U. (2004). Sequence-specific DNA binding determined by contacts outside the helix-turn-helix motif of the ParB homolog KorB. *Nat. Struct. Mol. Biol.* 11, 656–663. doi: 10.1038/nsmb773
- Kusiak, M., Gapczyńska, A., Plochocka, D., Thomas, C. M., and Jagura-Burdzy, G. (2011). Binding and spreading of ParB on DNA determine its biological function in *Pseudomonas aeruginosa*. *J. Bacteriol.* 193, 3342–3355. doi: 10.1128/jb.00328-11
- Leonard, T. A., Butler, P. J. G., and Löwe, J. (2004). Structural analysis of the chromosome segregation protein SpoJ from *Thermus thermophilus*. *Mol. Microbiol.* 53, 419–432. doi: 10.1111/j.1365-2958.2004.04133.x
- Leonard, T. A., Butler, P. J., and Löwe, J. (2005). Bacterial chromosome segregation: structure and DNA binding of the Soj dimer – a conserved biological switch. *EMBO J.* 24, 270–282. doi: 10.1038/sj.emboj.7600530
- Lukaszewicz, M., Kostelidou, K., Bartosik, A. A., Cooke, G. D., Thomas, C. M., and Jagura-Burdzy, G. (2002). Functional dissection of the ParB homologue (KorB) from IncP-1 plasmid RK2. *Nucl. Acids Res.* 30, 1046–1055. doi: 10.1093/nar/30.4.1046
- Lynch, A. S., and Wang, J. C. (1995). SopB protein-mediated silencing of genes linked to the *sopC* locus of *Escherichia coli* F plasmid. *Proc. Natl. Acad. Sci. U.S.A.* 92, 1896–1900.
- Murray, H., Ferreira, H., and Errington, J. (2006). The bacterial chromosome segregation protein SpoJ spreads along DNA from *parS* nucleation sites. *Mol. Microbiol.* 61, 1352–1361. doi: 10.1111/j.1365-2958.2006.05316.x
- Radnedge, L., Youngren, B., Davis, M., and Austin, S. (1998). Probing the structure of complex macromolecular interactions by homolog specificity scanning: the P1 and P7 plasmid partition systems. *EMBO J.* 17, 6076–6085.
- Ringgaard, S., van Zon, J., Howard, M., and Gerdes, K. (2009). Movement and equipositioning of plasmids by ParA filament disassembly. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19369–19374. doi: 10.1073/pnas.0908347106
- Rodionov, O., Lobocka, M., and Yarmolinsky, M. (1999). Silencing of genes flanking the P1 plasmid centromere. *Science* 283, 546–549.
- Rodionov, O., and Yarmolinsky, M. (2004). Plasmid partitioning and the spreading of P1 partition protein ParB. *Mol. Microbiol.* 52, 1215–1223. doi: 10.1111/j.1365-2958.2004.04055.x
- Sanchez, A., Cattoni, D. I., Walter, J.-C., Rech, J., Parmeggiani, A., Nollmann, M., et al. (2015). Stochastic self-assembly of ParB proteins builds the bacterial DNA segregation apparatus. *Cell Syst.* 1, 163–173. doi: 10.1016/j.cels.2015.07.013
- Sanchez, A., Rech, J., Gasc, C., and Bouet, J. Y. (2013). Insight into centromere-binding properties of ParB proteins: a secondary binding motif is essential for bacterial genome maintenance. *Nucl. Acids Res.* 41, 3094–3103. doi: 10.1093/nar/gkt018
- Schumacher, M. A., and Funnell, B. E. (2005). ParB-DNA structures reveal DNA-binding mechanism of partition complex formation. *Nature* 438, 516–519. doi: 10.1038/nature04149
- Schumacher, M. A., Piro, K. M., and Xu, W. J. (2010). Insight into F plasmid DNA segregation revealed by structures of SopB and SopB-DNA complexes. *Nucl. Acids Res.* 38, 4514–4526. doi: 10.1093/nar/gkq161
- Taylor, J. A., Pastrana, C. L., Butterer, A., Pernstich, C., Gwynn, E. J., Sobott, F., et al. (2015). Specific and non-specific interactions of ParB with DNA: implications for chromosome segregation. *Nucl. Acids Res.* 43, 719–731. doi: 10.1093/nar/gku1295
- Uversky, V. N. (2016). Dancing protein clouds: the strange biology and chaotic physics of intrinsically disordered proteins. *J. Biol. Chem.* 291, 6681–6688. doi: 10.1074/jbc.R115.685859
- Wang, X. D., Llopis, P. M., and Rudner, D. Z. (2013). Organization and segregation of bacterial chromosomes. *Nat. Rev. Genet.* 14, 191–203. doi: 10.1038/nrg3375
- Wright, P. E., and Dyson, H. J. (1999). Intrinsically unstructured proteins: reassessing the protein structure-function paradigm. *J. Mol. Biol.* 293, 321–331. doi: 10.1006/jmbi.1999.3110
- Yamaichi, Y., and Niki, H. (2000). Active segregation by the *Bacillus subtilis* partitioning system in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 14656–14661. doi: 10.1073/pnas.97.26.14656

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Funnell. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Segrosome Complex Formation during DNA Trafficking in Bacterial Cell Division

**María A. Oliva \***

Department of Chemical and Physical Biology, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Spanish National Research  
Council - Centre for Biological  
Research, Spain

### Reviewed by:

Kurt Henry Plepenbrink,  
The University of Maryland School of  
Medicine, USA  
Jan Löwe,  
Medical Research Council, UK

### \*Correspondence:

María A. Oliva  
marian@cib.csic.es

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 28 June 2016

**Accepted:** 24 August 2016

**Published:** 09 September 2016

### Citation:

Oliva MA (2016) Segrosome Complex  
Formation during DNA Trafficking in  
Bacterial Cell Division.  
Front. Mol. Biosci. 3:51.  
doi: 10.3389/fmolb.2016.00051

Bacterial extrachromosomal DNAs often contribute to virulence in pathogenic organisms or facilitate adaptation to particular environments. The transmission of genetic information from one generation to the next requires sufficient partitioning of DNA molecules to ensure that at least one copy reaches each side of the division plane and is inherited by the daughter cells. Segregation of the bacterial chromosome occurs during or after replication and probably involves a strategy in which several protein complexes participate to modify the folding pattern and distribution first of the origin domain and then of the rest of the chromosome. Low-copy number plasmids rely on specialized partitioning systems, which in some cases use a mechanism that show striking similarity to eukaryotic DNA segregation. Overall, there have been multiple systems implicated in the dynamic transport of DNA cargo to a new cellular position during the cell cycle but most seem to share a common initial DNA partitioning step, involving the formation of a nucleoprotein complex called the segrosome. The particular features and complex topologies of individual segrosomes depend on both the nature of the DNA binding protein involved and on the recognized centromeric DNA sequence, both of which vary across systems. The combination of *in vivo* and *in vitro* approaches, with structural biology has significantly furthered our understanding of the mechanisms underlying DNA trafficking in bacteria. Here, I discuss recent advances and the molecular details of the DNA segregation machinery, focusing on the formation of the segrosome complex.

**Keywords:** DNA segregation, partitioning systems, segrosome, partitioning complex, nucleoprotein complex, ParB, ParR, TubR

## DNA MAINTENANCE DURING BACTERIAL CELL DIVISION

The process of DNA segregation is a crucial stage of the bacterial cell cycle and it depends on the precise coordination with other cellular events. The faithful inheritance of genetic information during cell division ensures that each daughter cell receives a copy of the newly replicated DNA. In many organisms, the DNA-encoded genome consists of a core genome (the chromosome) and accessory genomes (extra-chromosomal, mobile genetic elements, MGEs). MGEs (plasmids, phages, conjugative transposons, etc.) often confer evolutionary advantages to the host bacteria, including the adaptation to different environmental niches. Many, if not most, naturally occurring MGEs are in low or unique copy number and thus bring their own post-replication survival apparatus encoded in stability determinants (partitioning systems, toxin-antitoxin systems, and multimer-resolution systems).

Partitioning (*par*) systems help to reliably segregate sister DNAs via a process that could be seen as functionally analogous to the mitotic segregation of chromosomes in eukaryotic cells. The best studied and probably the most common partitioning systems constitute a compact genetic module that is tightly auto-regulated by one of the gene products and consists of only three elements: a cis-acting DNA sequence and two trans-acting proteins. The DNA sequence denotes a *par* site or centromere-like region, and can be located at a single site (upstream or downstream of the operon) or at multiple positions within the MGE. The trans-acting proteins consist of a centromere-binding protein (CBP) that binds to the centromere and forms a nucleoprotein complex (partition complex or segrosome), and a motor protein (an NTPase), that sometimes is a cytomotive filament, which effectively moves the MGE inside the bacteria through direct interaction with the segrosome. Initially, these systems were classified as follows, based on the molecular nature of the NTPase: type I (Walker A-type ATPase), which further divided into Ia and Ib based on differences in the trans-acting proteins and the position of the centromere in the operon; and type II (cytomotive, actin-like ATPase) (Gerdes et al., 2000). Recently, an additional type III system has emerged, in which a cytomotive, tubulin-like GTPase serves as the motor protein (Larsen et al., 2007). Further, there may exist a type IV partitioning system, in which only a cis-acting DNA site and a DNA binding protein seem required for plasmid maintenance (Simpson et al., 2003; Guynet and de la Cruz, 2011). Hence, they may use a host bacteria's motor protein to track the DNA, or may even segregate passively by establishing an association with the chromosome (Guynet and de la Cruz, 2011).

It seems that partitioning systems share a common initial step that involves the specific recognition of the centromeric DNA region by the CBP. This step is crucial for assembly of the segrosome and subsequent events during DNA segregation. However, there is a considerable divergence among *par* sites and CBPs display different domain folds and organization (Hayes and Barilla, 2006; Baxter and Funnell, 2014), indicating differences in the segrosome assembly process and by extension, the corresponding partitioning mechanism. Here, I review the molecular mechanisms underlying segrosome formation in the partitioning systems that have been studied, focusing on those where structural information is available. Despite these variations in centromere sequences and the natures of the CBPs, common to all systems is the formation of the nucleoprotein complex that I propose may be categorized into two classes: those that mediate DNA segregation via *bridging* and those that do so via *wrapping*.

## DNA BRIDGING IN TYPE IA PARTITIONING SYSTEMS

Many plasmids, phages and chromosomes encode type Ia partitioning systems (Martin et al., 1987; Balzer et al., 1992; Lewis and Errington, 1997; Grigoriev and Lobočka, 2001). No single, common segrosome assembly mechanism has been described for these systems, probably owing to the wide diversity of centromeres and variations on the CBPs (below). The exact

nature of the partition complex is unknown, if it even exists in only one particular conformation, but the CBP bridges distant regions of DNA via both specific and non-specific binding (Rodionov et al., 1999; Bingle et al., 2005; Murray et al., 2006; Schumacher et al., 2007b; Graham et al., 2014), enabling the formation of a nucleoprotein complex linking and/or spanning thousands of base pairs with a small number of CBPs. Furthermore, spreading of the CBP has a masking effect on the function of the covered DNA, preventing interaction between the motor protein and the DNA and favoring the interaction with the segrosome (Bouet et al., 2007).

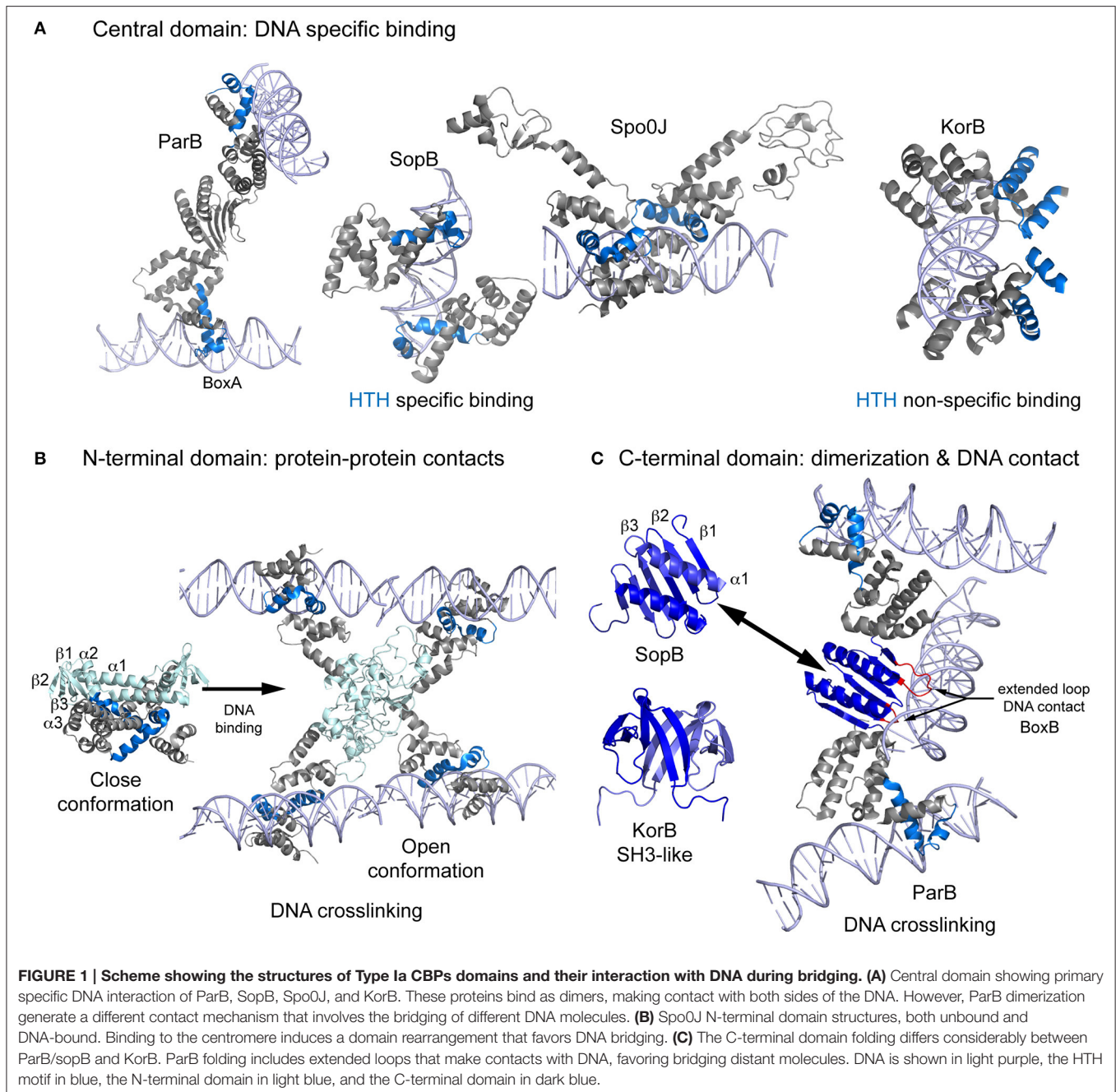
While the sequences of cis-acting sites (*parS*, *sopC* or *O<sub>B</sub>*) vary, the sites always contain inverted repeats. The *parS* site contains two different repeats asymmetrically arranged around a binding site for the IHF protein (Davis and Austin, 1988; Funnell, 1988b). One of the motifs is a heptamer (A-box) and the other a hexamer (B-box). Binding of IHF bends the DNA by 180°, thus strongly promoting ParB binding (Funnell, 1988a; Funnell and Gagnier, 1993; Rice et al., 1996; Bouet et al., 2000; Surtees and Funnell, 2001). Some chromosomes contain several *parS* sites dispersed over ~15% of the DNA, surrounding the replication origin (Lin and Grossman, 1998; Livny et al., 2007). However, the chromosomal *parS* sites consist exclusively of palindromic A-box motifs. *SopC* and *O<sub>B</sub>* sites comprise only one type of short inverted repeats contained within longer iterons that can be found either at a single locus (Mori et al., 1986) or scattered across the genome (Balzer et al., 1992; Ravin and Lane, 1999). The function of the regions flanking the inverted repeats is puzzling, as their presence is not conserved (Ravin and Lane, 1999). Similarly, the need for more than one iteron remains unclear, as in almost all cases a single copy is sufficient for segregation (Martin et al., 1987; Williams et al., 1998; Yates et al., 1999). However, given that the full-length centromere maximizes partitioning efficiency (Martin et al., 1987), the architecture of each segrosome may reflect evolutionary pressure on how well the systems work.

Type I CBPs are members of the ParB protein superfamily but show low sequence conservation. ParB, Spo0J, SopB, and KorB share the same domain organization, consisting of three flexibly linked domains (Schumacher et al., 2010): N-terminal, central (with a DNA-binding helix-turn-helix, HTH, motif), and C-terminal domains, which have been seen in various inter-domain conformations (Chen et al., 2015). The central domain is responsible for the primary CBP-DNA interaction, and the N- and C-terminal domains contribute to CBPs spreading around the centromere DNA. ParB proteins show high structural conservation only in the central domain, probably due to the presence of the HTH motif. For DNA binding, the HTH recognition helix inserts into the major groove, but there are differences between CBPs. In ParB, the HTH motif binds the *parS* box-A exclusively via the recognition helix (Schumacher and Funnell, 2005). SopB uses the recognition helix and an Arg outside the HTH (Schumacher et al., 2010; Sanchez et al., 2013). Spo0J binding is very similar to that observed for SopB but uses a Lys instead of an Arg and form additional specific contacts via another Arg and a Glu (Chen et al., 2015). Surprisingly, the HTH motif of KorB mediates only non-specific interactions, and DNA

binding specificity depends on contacts formed via a Thr and an Arg located outside the HTH (Khare et al., 2004). All these proteins bind DNA as dimers, whereby each molecule generally interacts with opposite sides of the inverted repeat (**Figure 1A**). However, in the crystal structure of ParB, the monomers of each dimer bind to box elements of different DNA molecules, suggesting a possible DNA crosslinking function or, that crystal packing occluded correct binding (Schumacher and Funnell, 2005).

ParBs' flexible N-terminal domain is responsible for the binding to the motor protein, oligomerization of the CBP around

the centromere, and also loading of bacterial condensin (Gruber and Errington, 2009; Sullivan et al., 2009; Minnen et al., 2011; Havey et al., 2012; Graham et al., 2014). **Figure 1B** shows the domain topology of Spo0J ( $\alpha 1$ - $\beta 1$ - $\beta 2$ - $\alpha 2$ - $\beta 3$ - $\alpha 3$ ), in which  $\beta$ -strands fold to form a  $\beta$ -sheet (Leonard et al., 2004). The two conserved motifs, box 1 and box 2 [with an "arginine patch," (Yamaichi and Niki, 2000)] are located between  $\alpha 1$  and  $\beta 1$  and between  $\beta 2$  and  $\alpha 2$ , respectively (Chen et al., 2015). Upon DNA binding the protein opens into an elongated, 78Å long structure, leaving the N-terminal domain exposed and available for protein-protein interactions (Chen et al., 2015). These interactions are





very flexible, but always include box 1 and 2 (Kusiak et al., 2011). Through this arrangement, the N-terminal domain is able to assist CBP spreading (Kusiak et al., 2011; Graham et al., 2014). Due to a lack of structural data, it remains unclear how the flexibility of the domains and their binding to DNA enables simultaneous or alternative interactions with the condensin and motor proteins.

The C-terminal domain is the most divergent, but in all these proteins shares the ability to dimerize (Leonard et al., 2004; Chen et al., 2015). The domain topology of ParB/SopB is  $\beta 1$ - $\beta 2$ - $\beta 3$ - $\alpha 1$ , where the  $\beta 3$ s of each monomer combine to form a continuous 6-stranded  $\beta$ -sheet and the helices interact to form an antiparallel coiled-coil (**Figure 1C**, Schumacher and Funnell, 2005; Schumacher et al., 2010). ParB contains extended loops between  $\beta 1$ - $\beta 2$  and  $\beta 2$ - $\beta 3$  that form highly specific contacts with the *parS* B-box (Schumacher and Funnell, 2005), generating a secondary DNA binding domain and contributing to DNA bridging during segrosome formation. By contrast, the C-terminal domain of KorB displays a completely different folding pattern, resembling an SH3 protein and consisting of a 5-stranded antiparallel  $\beta$  sheet (Delbruck et al., 2002). However, crosslinking studies suggest that this domain also facilitates DNA binding (Delbruck et al., 2002).

## SEGROSOME ASSEMBLY VIA WRAPPING

This strategy involves the formation of a filamentous nucleoprotein complex, in which the CBP wraps the centromere (type Ia partition systems) or the DNA wraps around a CBP oligomer (type II and III partition systems). The resulting segrosome is a single and discrete structure.

### Type Ib Systems

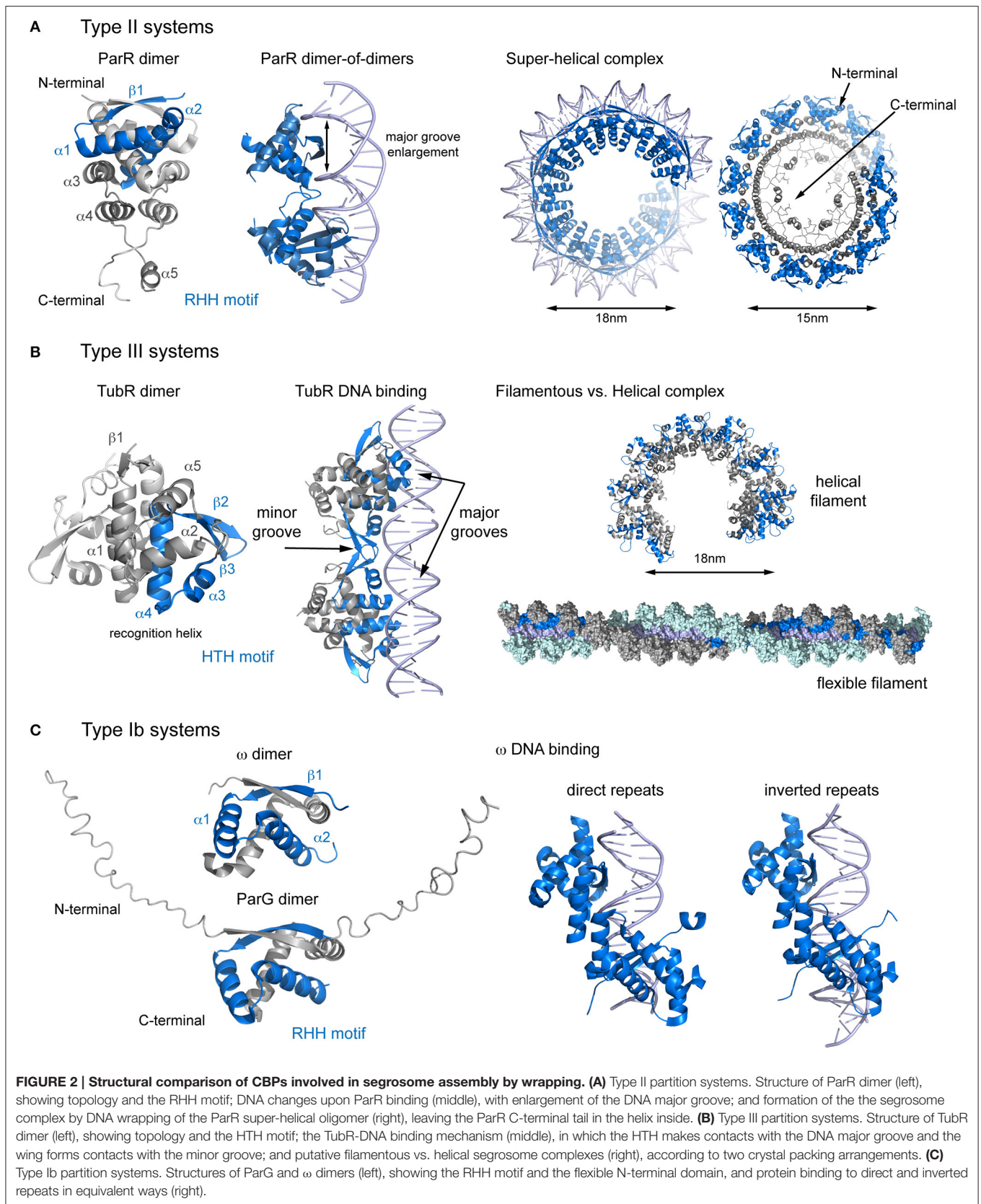
Surprisingly, the arrangement of the components in Type Ib systems is the only common aspect shared with the aforementioned systems. The interactions between their main components are different, and so may be the segregation mechanism. The centromere site localizes upstream of the *par* operon and consists of direct and inverted repeats. However, in plasmid pCXC100 the centromeric site contains only direct repeats (Yin et al., 2006; Huang et al., 2011). The CBPs, which also functions as repressors (Carmelo et al., 2005; Weihofen et al., 2006) are small proteins that share the arrangement into N- and C-terminal domains (**Figure 2C**). The N-terminal domain, which shows a highly divergent sequence, is flexible and unstructured, and includes a conserved arginine finger that has been implicated in the activation of ATP hydrolysis in the motor protein (Barilla et al., 2007). The C-terminal domain topology is  $\beta 1$ - $\alpha 1$ - $\alpha 2$  and includes a ribbon-helix-helix (RHH) DNA-binding motif (Murayama et al., 2001; Golovanov et al., 2003; Huang et al., 2011). The  $\beta 1$  strand from two different molecules pairs into an antiparallel  $\beta$ -ribbon, meaning that these CBPs are also present as dimers in solution (Barilla and Hayes, 2003; Golovanov et al., 2003).

Plasmid pSM19035 harbors a unique partitioning system: rather than being encoded in a single operon, each gene is transcribed separately from different promoters. The full

centromere contains 3 separate *parS* sites, consisting of 9, 7, and 10 iterons that occur twice in the plasmid genome (*parS1*, *parS1'*, *parS2*, *parS2'*, *parS3*, *parS3'*, de la Hoz et al., 2000, 2004; Dmowski et al., 2006). However, *parS2* appears to be the main centromeric sequence (Dmowski and Kern-Zdanowicz, 2016). Interestingly, each *parS* overlaps with the promoters of genes involved in plasmid copy number and maintenance: *parS1* with P $\delta$ , *parS2* with P $\omega$  and *parS3* with P $\text{copS}$  (de la Hoz et al., 2000). The CBP,  $\omega$ , binds to each *parS* with different affinities, depending on the number of iterons (de la Hoz et al., 2004). This feature may be crucial to fine-tune repressor affinity for different promoters (Weihofen et al., 2006). The nucleoprotein complex is a left-handed protein helix that wraps the DNA (Weihofen et al., 2006) covering only the *parS* site (Pratto et al., 2009). Protein binding to both direct and inverted repeats involves comparable interactions, due to the pseudo-symmetry of the dimer (Weihofen et al., 2006, **Figure 2C**). Binding induces minor structural changes mainly affecting the loop connecting  $\alpha 1$  and  $\alpha 2$ . In contrast to other RHH DNA-binding proteins, there is no DNA bending (Pratto et al., 2009). Because the DNA is not curved,  $\omega$  first makes contact with the DNA major groove via base specific interactions with residues on the  $\beta$ -sheet and then the N-termini of the  $\alpha 2$  helices clamp the phosphate backbones (Weihofen et al., 2006). Assuming nearly straight DNA, the number and orientation of repetitions will affect the distances between helices  $\alpha 1$  of adjacent  $\omega$  dimers, thereby modulating the cooperativity. The motor protein,  $\delta$ , binds non-specifically to DNA but is recruited to the location of the segrosome (Pratto et al., 2009) to form a ternary complex, giving rise to intermolecular pairing of *parS* regions (Pratto et al., 2008, 2009). This bridging may increase the local concentration of  $\omega$ , in turn increasing the ATPase activity of the motor protein and thus inducing detachment of this protein and promoting mobility (Pratto et al., 2009). This system may combine both DNA wrapping mechanisms (during segrosome formation) and DNA bridging mechanisms (when the motor protein participates in segregation).

In plasmid TP228, the centromere (*parH*) is continuous and consists of direct and inverted repeats separated by AT-rich regions. A DNA region between the operon genes and the centromere ( $O_F$ ) comprises more repeats that play important roles in partitioning and transcription regulation (Zampini et al., 2009; Wu et al., 2011). Binding of the CBP, ParG, to *parH* occurs via the RHH motif, but unlike in  $\omega$ , ParG is also dependent on the protein's N-terminal tail, which modulates binding affinity (Golovanov et al., 2003). Apparently, the AT-enriched spacers may increase the binding cooperativity of ParG to DNA (Wu et al., 2011). Like plasmid pSM19035, the centromere site is not curved and ParG binding does not induce DNA bending. The motor protein, ParF polymerizes into filaments and does not bind to DNA (Barilla et al., 2005; Schumacher et al., 2012). The N-terminal domain of ParG is not only important for the activation of the ATPase but also facilitates ParF filament nucleation and bundling (Barilla et al., 2007). Furthermore, in contrast to all other described systems, the ParF-ParG interaction is not dependent on the formation of the segrosome. This





suggests that pSM19035 and TP228 despite sharing the same type Ib partitioning system employ distinct segregation mechanisms.

## Type II Systems

The centromeric site (*parC*) consists of tandem repeats localized in a single locus upstream of the operon. The arrangement can be continuous (plasmid pSK41, Schumacher et al., 2007a) or split into two regions (plasmid R1), with the *par* cassette promoter in the middle (Dam and Gerdes, 1994). However, the resulting segregation complexes are very similar. The CBP, ParR, contains two domains, an N-terminal domain with a RHH DNA binding motif (as seen in type Ib systems), and a C-terminal domain that is involved in the interaction with the motor protein. The domain topology of the ParR N-terminal domain is  $\beta 1-\alpha 1-\alpha 2-\alpha 3-\alpha 4-\alpha 5$  (Figure 2A). The  $\beta 1$ -strands from two monomers combine in an antiparallel fashion and the  $\alpha 1-\alpha 2$  helices come together to form an extensive dimer (Moller-Jensen et al., 2007; Schumacher et al., 2007a). The C-terminal domain includes a 3-helix cap that reinforces the tight dimerization of the N-terminal domain and an unstructured C-terminal tail with a high degree of sequence conservation (Moller-Jensen et al., 2007).

The nucleoprotein complex forms a discrete helical arrangement with a diameter of 15-nm (Moller-Jensen et al., 2007; Hoischen et al., 2008). The structure of the nucleoprotein complex (Schumacher et al., 2007a) reveals a continuous helical array in the crystal packing (Figure 2A). Each turn consists of 6 symmetrical pairs of dimers (involving the assembly of 12 ParR dimers), producing distinct negative and positive electrostatic on the inner and outer surfaces of the helix. The DNA wraps ParR by interacting with the outer, positively charged surface of the super helix, with each dimer binding one *parC* iteron. When the centromere is split in two, the promoter region forms a DNA loop that protrudes out of the ParR-*parC* ring structure (Hoischen et al., 2008; Salje and Lowe, 2008), repressing the promoter (Jensen et al., 1994; Breuner et al., 1996), and regulating transcription of the partition genes (Salje and Lowe, 2008). The DNA is bent by  $46^\circ$  and widened so that the major groove grows from 11 to 14 Å (Schumacher et al., 2007a). The groove enlargement allows the insertion of the RHH motif, as described for other DNA-binding RHH proteins (Somers and Phillips, 1992; Raumann et al., 1994; Gomis-Ruth et al., 1998). Interestingly, the phosphate contacts cluster at the 5' ends of the 10-bp repeats, creating the closest physical associations between ParR and the DNA. Full-length ParR from plasmid pB171 crystallized in a helical superstructure in the absence of DNA, with a diameter very similar to that measured in the nucleoprotein complex (15 vs. 18 nm) (Figure 2A). Moreover, the protein arrangement into dimers and the electrostatic distribution are also similar (Moller-Jensen et al., 2007). These observations lead to the question; which event occurs first? If ParR assembly into a super-helical structure occurs first, then the macromolecular complex may recruit *parC*. Otherwise, the centromere might function as a scaffold for ParR oligomerization.

For ParR, the segrosome structure positions the conserved C-tails clustered on the inside surface of the helix, where they mediate binding to the motor protein, ParM (Schumacher

et al., 2007a; Salje and Lowe, 2008). The ParR tail binds to a hydrophobic pocket in ParM in an interaction resembling that described for actin polymer modulators and the barbed end of actin filaments (Gayathri et al., 2012). Furthermore, the segrosome binds only at the growing end of the polar ParM double helical filament favoring filament growth via a formin-like mechanism (Gayathri et al., 2012). Why does the ParR-ParM interaction require the clustering of so many ParR tails? It is possible that several tails bind to a single ParM molecule with distinct affinities, regulating ParM filament dynamics. Alternatively, the presence of free tails may be necessary to explore the space around the filament end and to facilitate the addition of ParM molecules to the growing filament while remaining attached at all times.

## Type III Systems

The type III system were the most recently discovered partitioning systems (Larsen et al., 2007). For TubZRC, the centromeric site (*tubC*) is localized upstream of the operon and contains several direct repeats in a single locus that can be split into two (pBtoxis) or three (pBsph) blocks, resembling discontinuous *parC* sites (Aylett and Lowe, 2012; Ge et al., 2014a). During partitioning, the CBP (TubR in this case) mediates the assembly of the segrosome nucleoprotein complex and acts as a repressor of *tubRZ* transcription (Tang et al., 2006; Larsen et al., 2007; Ge et al., 2014a). TubR is a small winged-helix DNA-binding protein with a high degree of structure conservation. The topology is  $\beta 1-\alpha 1-\alpha 2-\alpha 3-\alpha 4-\beta 2-\beta 3-\alpha 5$ , where the  $\alpha 3-\alpha 4$  helices form the HTH motif ( $\alpha 4$  is the “recognition helix”) and the loop between  $\beta 2-\beta 3$  forms the wing (Figure 2B, Ni et al., 2010). Interestingly, TubR forms a highly intertwined dimer involving the canonical HTH motif, resulting in an atypical protein-DNA binding (Aylett and Lowe, 2012). The N-termini of both recognition helices in a dimer protrude into the major groove of the DNA, while the acidic patch in the wing complements the DNA backbone phosphate in the minor groove. The nucleoprotein complex takes the shape of a flexible filament, with TubR wrapping helically around both sides of *tubC* (Aylett and Lowe, 2012, Figure 2B). The filamentous complex closes to form 18-nm wide ring-like structures (Aylett and Lowe, 2012). However, the structure of plasmid pBM400 TubR, with no DNA bound, reveals a helical arrangement, resembling the ParR super-helical complex (Figures 2A,B). It thus remains unclear whether the segrosome complex is formed by TubR wrapping of the DNA or by DNA wrapping of the TubR oligomer, which could lead to different interacting mechanism with the motor protein (TubZ).

TubR binds to TubZ C-terminal tail (Ni et al., 2010). However, the interaction is only possible following formation of the filamentous segrosome. Neither TubR alone (Oliva et al., 2012) nor TubR bound to either of the two-iteron clusters are capable of interacting with TubZ (Aylett and Lowe, 2012; Fink and Lowe, 2015). Therefore, the clustering of TubR may generate the binding site for TubZ. Differently to type II systems, the segrosome tracks the shrinking minus end of the TubZ filament, suggesting a pulling segregation mechanism (Fink and Lowe, 2015).

Type III partition systems involve a third protein with a predicted HTH DNA-binding motif and a long coiled-coil domain (TubY), located downstream of the partition operon (Oliva et al., 2012). TubY seems to be a regulator protein that modulates TubZ assembly (Oliva et al., 2012) and also acts as a transcriptional activator (Ge et al., 2014b) but the exact molecular mechanisms remain elusive.

It is still common for new partitioning systems to be discovered in plasmids, phages, and on chromosomes. Together with a growing body of molecular insights these will help to broaden our understanding of DNA trafficking during bacterial cell division and in particular how DNA is attached to the CBP during segrosome

formation and then to the motor protein through the segrosome.

## AUTHOR CONTRIBUTIONS

MO conceived and wrote this mini-review.

## ACKNOWLEDGMENTS

This work was supported by the Ministerio de Ciencia e Innovación (grant RYC-2011-07900) and by co-funding grant BFU2013-47014-P from the Ministerio de Economía y Competitividad and European Regional Development Fund.

## REFERENCES

- Aylett, C. H., and Lowe, J. (2012). Superstructure of the centromeric complex of TubZRC plasmid partitioning systems. *Proc. Natl. Acad. Sci. U.S.A.* 109, 16522–16527. doi: 10.1073/pnas.1210899109
- Balzer, D., Ziegelin, G., Pansegrau, W., Kruft, V., and Lanka, E. (1992). KorB protein of promiscuous plasmid RP4 recognizes inverted sequence repetitions in regions essential for conjugative plasmid transfer. *Nucleic Acids Res.* 20, 1851–1858. doi: 10.1093/nar/20.8.1851
- Barillà, D., Carmelo, E., and Hayes, F. (2007). The tail of the ParG DNA segregation protein remodels ParF polymers and enhances ATP hydrolysis via an arginine finger-like motif. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1811–1816. doi: 10.1073/pnas.0607216104
- Barillà, D., and Hayes, F. (2003). Architecture of the ParF\*ParG protein complex involved in prokaryotic DNA segregation. *Mol. Microbiol.* 49, 487–499. doi: 10.1046/j.1365-2958.2003.03564.x
- Barillà, D., Rosenberg, M. F., Nobbmann, U., and Hayes, F. (2005). Bacterial DNA segregation dynamics mediated by the polymerizing protein ParF. *EMBO J.* 24, 1453–1464. doi: 10.1038/sj.emboj.7600619
- Baxter, J. C., and Funnell, B. E. (2014). Plasmid Partition Mechanisms. *Microbiol. Spectr.* 2:PLAS-0023-2014. doi: 10.1128/microbiolspec.PLAS-0023-2014
- Bingle, L. E., Macartney, D. P., Fantozzi, A., Manzo, S. E., and Thomas, C. M. (2005). Flexibility in repression and cooperativity by KorB of broad host range IncP-1 plasmid RK2. *J. Mol. Biol.* 349, 302–316. doi: 10.1016/j.jmb.2005.03.062
- Bouet, J. Y., Ah-Seng, Y., Benmeradi, N., and Lane, D. (2007). Polymerization of SopA partition ATPase: regulation by DNA binding and SopB. *Mol. Microbiol.* 63, 468–481. doi: 10.1111/j.1365-2958.2006.05537.x
- Bouet, J. Y., Surtees, J. A., and Funnell, B. E. (2000). Stoichiometry of P1 plasmid partition complexes. *J. Biol. Chem.* 275, 8213–8219. doi: 10.1074/jbc.275.11.8213
- Breüner, A., Jensen, R. B., Dam, M., Pedersen, S., and Gerdes, K. (1996). The centromere-like parC locus of plasmid R1. *Mol. Microbiol.* 20, 581–592. doi: 10.1046/j.1365-2958.1996.5351063.x
- Carmelo, E., Barillà, D., Golovanov, A. P., Lian, L. Y., Derome, A., and Hayes, F. (2005). The unstructured N-terminal tail of ParG modulates assembly of a quaternary nucleoprotein complex in transcription repression. *J. Biol. Chem.* 280, 28683–28691. doi: 10.1074/jbc.M501173200
- Chen, B. W., Lin, M. H., Chu, C. H., Hsu, C. E., and Sun, Y. J. (2015). Insights into ParB spreading from the complex structure of Spo0J and parS. *Proc. Natl. Acad. Sci. U.S.A.* 112, 6613–6618. doi: 10.1073/pnas.1421927112
- Dam, M., and Gerdes, K. (1994). Partitioning of plasmid R1. Ten direct repeats flanking the parA promoter constitute a centromere-like partition site parC, that expresses incompatibility. *J. Mol. Biol.* 236, 1289–1298. doi: 10.1016/0022-2836(94)90058-2
- Davis, M. A., and Austin, S. J. (1988). Recognition of the P1 plasmid centromere analog involves binding of the ParB protein and is modified by a specific host factor. *EMBO J.* 7, 1881–1888.
- de la Hoz, A. B., Ayora, S., Sitkiewicz, I., Fernández, S., Pankiewicz, R., Alonso, J. C., et al. (2000). Plasmid copy-number control and better-than-random segregation genes of pSM19035 share a common regulator. *Proc. Natl. Acad. Sci. U.S.A.* 97, 728–733. doi: 10.1073/pnas.97.2.728
- de la Hoz, A. B., Pratto, F., Misselwitz, R., Speck, C., Weihofen, W., Welfle, K., et al. (2004). Recognition of DNA by  $\omega$  protein from the broad-host range *Streptococcus pyogenes* plasmid pSM19035: analysis of binding to operator DNA with one to four heptad repeats. *Nucleic Acids Res.* 32, 3136–3147. doi: 10.1093/nar/gkh633
- Delbrück, H., Ziegelin, G., Lanka, E., and Heinemann, U. (2002). An Src homology 3-like domain is responsible for dimerization of the repressor protein KorB encoded by the promiscuous IncP plasmid RP4. *J. Biol. Chem.* 277, 4191–4198. doi: 10.1074/jbc.M110103200
- Dmowski, M., and Kern-Zdanowicz, I. (2016). The Omega (ParB) binding sites together with the RNA polymerase-recognized sequence are essential for centromeric functions of the  $P_{\omega}$  region in the partition system of pSM19035. *Microbiology* 162, 1114–1124. doi: 10.1099/mic.0.000308
- Dmowski, M., Sitkiewicz, I., and Ceglowski, P. (2006). Characterization of a novel partition system encoded by the  $\delta$  and  $\omega$  genes from the streptococcal plasmid pSM19035. *J. Bacteriol.* 188, 4362–4372. doi: 10.1128/JB.01922-05
- Fink, G., and Löwe, J. (2015). Reconstitution of a prokaryotic minus end-tracking system using TubRC centromeric complexes and tubulin-like protein TubZ filaments. *Proc. Natl. Acad. Sci. U.S.A.* 112, E1845–E1850. doi: 10.1073/pnas.1423746112
- Funnell, B. E. (1988a). Mini-P1 plasmid partitioning: excess ParB protein destabilizes plasmids containing the centromere parS. *J. Bacteriol.* 170, 954–960.
- Funnell, B. E. (1988b). Participation of *Escherichia coli* integration host factor in the P1 plasmid partition system. *Proc. Natl. Acad. Sci. U.S.A.* 85, 6657–6661.
- Funnell, B. E., and Gagnier, L. (1993). The P1 plasmid partition complex at parS. II. Analysis of ParB protein binding activity and specificity. *J. Biol. Chem.* 268, 3616–3624.
- Gayathri, P., Fujii, T., Møller-Jensen, J., van den Ent, F., Namba, K., and Löwe, J. (2012). A bipolar spindle of antiparallel ParM filaments drives bacterial plasmid segregation. *Science* 338, 1334–1337. doi: 10.1126/science.1229091
- Ge, Y., Hu, X., Zhao, N., Shi, T., Cai, Q., and Yuan, Z. (2014a). A new tubRZ operon involved in the maintenance of the *Bacillus sphaericus* mosquitoicidal plasmid pBsph. *Microbiology* 160, 1112–1124. doi: 10.1099/mic.0.075465-0
- Ge, Y., Zhao, N., Hu, X., Shi, T., Cai, Q., and Yuan, Z. (2014b). A novel transcriptional activator, tubX, is required for the stability of *Bacillus sphaericus* mosquitoicidal plasmid pBsph. *J. Bacteriol.* 196, 4304–4314. doi: 10.1128/JB.01855-14
- Gerdes, K., Møller-Jensen, J., and Bugge Jensen, R. (2000). Plasmid and chromosome partitioning: surprises from phylogeny. *Mol. Microbiol.* 37, 455–466. doi: 10.1046/j.1365-2958.2000.01975.x
- Golovanov, A. P., Barillà, D., Golovanova, M., Hayes, F., and Lian, L. Y. (2003). ParG, a protein required for active partition of bacterial plasmids, has a dimeric ribbon-helix-helix structure. *Mol. Microbiol.* 50, 1141–1153. doi: 10.1046/j.1365-2958.2003.03750.x
- Gomis-Rüth, F. X., Solá, M., Acebo, P., Parraga, A., Guasch, A., Eritja, R., et al. (1998). The structure of plasmid-encoded transcriptional repressor



- CopG unliganded and bound to its operator. *EMBO J.* 17, 7404–7415. doi: 10.1093/emboj/17.24.7404
- Graham, T. G., Wang, X., Song, D., Etson, C. M., van Oijen, A. M., Rudner, D. Z., et al. (2014). ParB spreading requires DNA bridging. *Genes Dev.* 28, 1228–1238. doi: 10.1101/gad.242206.114
- Grigoriev, P. S., and Lobočka, M. B. (2001). Determinants of segregational stability of the linear plasmid-prophage N15 of *Escherichia coli*. *Mol. Microbiol.* 42, 355–368. doi: 10.1046/j.1365-2958.2001.02632.x
- Gruber, S., and Errington, J. (2009). Recruitment of condensin to replication origin regions by ParB/Spo0J promotes chromosome segregation in *B. subtilis*. *Cell* 137, 685–696. doi: 10.1016/j.cell.2009.02.035
- Guynet, C., and de la Cruz, F. (2011). Plasmid segregation without partition. *Mob. Genet. Elements* 1, 236–241. doi: 10.4161/mge.1.3.18229
- Havey, J. C., Vecchiarelli, A. G., and Funnell, B. E. (2012). ATP-regulated interactions between P1 ParA, ParB and non-specific DNA that are stabilized by the plasmid partition site, parS. *Nucleic Acids Res.* 40, 801–812. doi: 10.1093/nar/gkr747
- Hayes, F., and Barillà, D. (2006). The bacterial segrosome: a dynamic nucleoprotein machine for DNA trafficking and segregation. *Nat. Rev. Microbiol.* 4, 133–143. doi: 10.1038/nrmicro1342
- Hoischen, C., Bussiek, M., Langowski, J., and Diekmann, S. (2008). *Escherichia coli* low-copy-number plasmid R1 centromere parC forms a U-shaped complex with its binding protein ParR. *Nucleic Acids Res.* 36, 607–615. doi: 10.1093/nar/gkm672
- Huang, L., Yin, P., Zhu, X., Zhang, Y., and Ye, K. (2011). Crystal structure and centromere binding of the plasmid segregation protein ParB from pCX100. *Nucleic Acids Res.* 39, 2954–2968. doi: 10.1093/nar/gkq915
- Jensen, R. B., Dam, M., and Gerdes, K. (1994). Partitioning of plasmid R1. The parA operon is autoregulated by ParR and its transcription is highly stimulated by a downstream activating element. *J. Mol. Biol.* 236, 1299–1309. doi: 10.1016/0022-2836(94)90059-0
- Khare, D., Ziegler, G., Lanka, E., and Heinemann, U. (2004). Sequence-specific DNA binding determined by contacts outside the helix-turn-helix motif of the ParB homolog KorB. *Nat. Struct. Mol. Biol.* 11, 656–663. doi: 10.1038/nsmb773
- Kusiak, M., Gapczynska, A., Plochocka, D., Thomas, C. M., and Jagura-Burdzy, G. (2011). Binding and spreading of ParB on DNA determine its biological function in *Pseudomonas aeruginosa*. *J. Bacteriol.* 193, 3342–3355. doi: 10.1128/JB.00328-11
- Larsen, R. A., Cusumano, C., Fujioka, A., Lim-Fong, G., Patterson, P., and Pogliano, J. (2007). Treadmilling of a prokaryotic tubulin-like protein, TubZ, required for plasmid stability in *Bacillus thuringiensis*. *Genes Dev.* 21, 1340–1352. doi: 10.1101/gad.1546107
- Leonard, T. A., Butler, P. J., and Löwe, J. (2004). Structural analysis of the chromosome segregation protein Spo0J from *Thermus thermophilus*. *Mol. Microbiol.* 53, 419–432. doi: 10.1111/j.1365-2958.2004.04133.x
- Lewis, P. J., and Errington, J. (1997). Direct evidence for active segregation of oriC regions of the *Bacillus subtilis* chromosome and co-localization with the Spo0J partitioning protein. *Mol. Microbiol.* 25, 945–954. doi: 10.1111/j.1365-2958.1997.mmi530.x
- Lin, D. C., and Grossman, A. D. (1998). Identification and characterization of a bacterial chromosome partitioning site. *Cell* 92, 675–685. doi: 10.1016/S0092-8674(00)81135-6
- Livny, J., Yamaichi, Y., and Waldor, M. K. (2007). Distribution of centromere-like parS sites in bacteria: insights from comparative genomics. *J. Bacteriol.* 189, 8693–8703. doi: 10.1128/JB.01239-07
- Martin, K. A., Friedman, S. A., and Austin, S. J. (1987). Partition site of the P1 plasmid. *Proc. Natl. Acad. Sci. U.S.A.* 84, 8544–8547. doi: 10.1073/pnas.84.23.8544
- Minnen, A., Attaiech, L., Thon, M., Gruber, S., and Veening, J. W. (2011). SMC is recruited to oriC by ParB and promotes chromosome segregation in *Streptococcus pneumoniae*. *Mol. Microbiol.* 81, 676–688. doi: 10.1111/j.1365-2958.2011.07722.x
- Møller-Jensen, J., Ringgaard, S., Mercogliano, C. P., Gerdes, K., and Löwe, J. (2007). Structural analysis of the ParR/parC plasmid partition complex. *EMBO J.* 26, 4413–4422. doi: 10.1038/sj.emboj.7601864
- Mori, H., Kondo, A., Ohshima, A., Ogura, T., and Hiraga, S. (1986). Structure and function of the F plasmid genes essential for partitioning. *J. Mol. Biol.* 192, 1–15. doi: 10.1016/0022-2836(86)90459-6
- Murayama, K., Orth, P., de la Hoz, A. B., Alonso, J. C., and Saenger, W. (2001). Crystal structure of  $\omega$  transcriptional repressor encoded by *Streptococcus pyogenes* plasmid pSM19035 at 1.5 Å resolution. *J. Mol. Biol.* 314, 789–796. doi: 10.1006/jmbi.2001.5157
- Murray, H., Ferreira, H., and Errington, J. (2006). The bacterial chromosome segregation protein Spo0J spreads along DNA from parS nucleation sites. *Mol. Microbiol.* 61, 1352–1361. doi: 10.1111/j.1365-2958.2006.05316.x
- Ni, L., Xu, W., Kumaraswami, M., and Schumacher, M. A. (2010). Plasmid protein TubR uses a distinct mode of HTH-DNA binding and recruits the prokaryotic tubulin homolog TubZ to effect DNA partition. *Proc. Natl. Acad. Sci. U.S.A.* 107, 11763–11768. doi: 10.1073/pnas.1003817107
- Oliva, M. A., Martin-Galiano, A. J., Sakaguchi, Y., and Andreu, J. M. (2012). Tubulin homolog TubZ in a phage-encoded partition system. *Proc. Natl. Acad. Sci. U.S.A.* 109, 7711–7716. doi: 10.1073/pnas.1121546109
- Pratto, F., Cicek, A., Weihofen, W. A., Lurz, R., Saenger, W., and Alonso, J. C. (2008). *Streptococcus pyogenes* pSM19035 requires dynamic assembly of ATP-bound ParA and ParB on parS DNA during plasmid segregation. *Nucleic Acids Res.* 36, 3676–3689. doi: 10.1093/nar/gkn170
- Pratto, F., Suzuki, Y., Takeyasu, K., and Alonso, J. C. (2009). Single-molecule analysis of protein-DNA complexes formed during partition of newly replicated plasmid molecules in *Streptococcus pyogenes*. *J. Biol. Chem.* 284, 30298–30306. doi: 10.1074/jbc.M109.035410
- Raumann, B. E., Rould, M. A., Pabo, C. O., and Sauer, R. T. (1994). DNA recognition by  $\beta$ -sheets in the Arc repressor-operator crystal structure. *Nature* 367, 754–757. doi: 10.1038/367754a0
- Ravin, N., and Lane, D. (1999). Partition of the linear plasmid N15: interactions of N15 partition functions with the sop locus of the F plasmid. *J. Bacteriol.* 181, 6898–6906.
- Rice, P. A., Yang, S., Mizuuchi, K., and Nash, H. A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* 87, 1295–1306. doi: 10.1016/S0092-8674(00)81824-3
- Rodionov, O., Lobočka, M., and Yarmolinsky, M. (1999). Silencing of genes flanking the P1 plasmid centromere. *Science* 283, 546–549. doi: 10.1126/science.283.5401.546
- Salje, J., and Löwe, J. (2008). Bacterial actin: architecture of the ParMRC plasmid DNA partitioning complex. *EMBO J.* 27, 2230–2238. doi: 10.1038/emboj.2008.152
- Sanchez, A., Rech, J., Gasc, C., and Bouet, J. Y. (2013). Insight into centromere-binding properties of ParB proteins: a secondary binding motif is essential for bacterial genome maintenance. *Nucleic Acids Res.* 41, 3094–3103. doi: 10.1093/nar/gkt018
- Schumacher, M. A., and Funnell, B. E. (2005). Structures of ParB bound to DNA reveal mechanism of partition complex formation. *Nature* 438, 516–519. doi: 10.1038/nature04149
- Schumacher, M. A., Glover, T. C., Brzoska, A. J., Jensen, S. O., Dunham, T. D., Skurray, R. A., et al. (2007a). Segrosome structure revealed by a complex of ParR with centromere DNA. *Nature* 450, 1268–1271. doi: 10.1038/nature06392
- Schumacher, M. A., Mansoor, A., and Funnell, B. E. (2007b). Structure of a four-way bridged ParB-DNA complex provides insight into P1 segrosome assembly. *J. Biol. Chem.* 282, 10456–10464. doi: 10.1074/jbc.M610603200
- Schumacher, M. A., Piro, K. M., and Xu, W. (2010). Insight into F plasmid DNA segregation revealed by structures of SopB and SopB-DNA complexes. *Nucleic Acids Res.* 38, 4514–4526. doi: 10.1093/nar/gkq161
- Schumacher, M. A., Ye, Q., Barge, M. T., Zampini, M., Barillà, D., and Hayes, F. (2012). Structural mechanism of ATP-induced polymerization of the partition factor ParF: implications for DNA segregation. *J. Biol. Chem.* 287, 26146–26154. doi: 10.1074/jbc.M112.373696
- Simpson, A. E., Skurray, R. A., and Firth, N. (2003). A single gene on the staphylococcal multiresistance plasmid pSK1 encodes a novel partitioning system. *J. Bacteriol.* 185, 2143–2152. doi: 10.1128/JB.185.7.2143-2152.2003
- Somers, W. S., and Phillips, S. E. (1992). Crystal structure of the met repressor-operator complex at 2.8 Å resolution reveals DNA recognition by  $\beta$ -strands. *Nature* 359, 387–393. doi: 10.1038/359387a0



- Sullivan, N. L., Marquis, K. A., and Rudner, D. Z. (2009). Recruitment of SMC by ParB-parS organizes the origin region and promotes efficient chromosome segregation. *Cell* 137, 697–707. doi: 10.1016/j.cell.2009.04.044
- Surtees, J. A., and Funnell, B. E. (2001). The DNA binding domains of P1 ParB and the architecture of the P1 plasmid partition complex. *J. Biol. Chem.* 276, 12385–12394. doi: 10.1074/jbc.M009370200
- Tang, M., Bideshi, D. K., Park, H. W., and Federici, B. A. (2006). Minireplicon from pBtoxis of *Bacillus thuringiensis* subsp. israelensis. *Appl. Environ. Microbiol.* 72, 6948–6954. doi: 10.1128/AEM.00976-06
- Weihofen, W. A., Cicek, A., Pratto, F., Alonso, J. C., and Saenger, W. (2006). Structures of  $\omega$  repressors bound to direct and inverted DNA repeats explain modulation of transcription. *Nucleic Acids Res.* 34, 1450–1458. doi: 10.1093/nar/gkl015
- Williams, D. R., Macartney, D. P., and Thomas, C. M. (1998). The partitioning activity of the RK2 central control region requires only incC, korB and KorB-binding site O(B)3 but other KorB-binding sites form destabilizing complexes in the absence of O(B)3. *Microbiology* 144(Pt 12), 3369–3378. doi: 10.1099/00221287-144-12-3369
- Wu, M., Zampini, M., Bussiek, M., Hoischen, C., Diekmann, S., and Hayes, F. (2011). Segrosome assembly at the pliable parH centromere. *Nucleic Acids Res.* 39, 5082–5097. doi: 10.1093/nar/gkr115
- Yamaichi, Y., and Niki, H. (2000). Active segregation by the *Bacillus subtilis* partitioning system in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 14656–14661. doi: 10.1073/pnas.97.26.14656
- Yates, P., Lane, D., and Biek, D. P. (1999). The F plasmid centromere, sopC, is required for full repression of the sopAB operon. *J. Mol. Biol.* 290, 627–638. doi: 10.1006/jmbi.1999.2909
- Yin, P., Li, T. Y., Xie, M. H., Jiang, L., and Zhang, Y. (2006). A Type Ib ParB protein involved in plasmid partitioning in a gram-positive bacterium. *J. Bacteriol.* 188, 8103–8108. doi: 10.1128/JB.01232-06
- Zampini, M., Derome, A., Bailey, S. E., Barillà, D., and Hayes, F. (2009). Recruitment of the ParG segregation protein to different affinity DNA sites. *J. Bacteriol.* 191, 3832–3841. doi: 10.1128/JB.01630-08

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a shared affiliation, though no other collaboration, with the author and states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2016 Oliva. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Modulating *Salmonella* Typhimurium's Response to a Changing Environment through Bacterial Enhancer-Binding Proteins and the RpoN Regulon

Christine E. Hartman<sup>†</sup>, David J. Samuels<sup>†</sup> and Anna C. Karls<sup>\*</sup>

Department of Microbiology, University of Georgia, Athens, GA, USA

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas Medical Branch at  
Galveston, USA

### Reviewed by:

David John Studholme,  
University of Exeter, UK  
Larry Reitzer,  
University of Texas at Dallas, USA

### \*Correspondence:

Anna C. Karls  
akarls@uga.edu

### <sup>†</sup>Present Address:

Christine E. Hartman,  
Office for Teaching and Learning,  
Wayne State University, Detroit,  
MI, USA  
David J. Samuels,  
Department of Biology, Georgetown  
University, Washington, DC, USA

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 27 June 2016

**Accepted:** 28 July 2016

**Published:** 17 August 2016

### Citation:

Hartman CE, Samuels DJ and  
Karls AC (2016) Modulating  
*Salmonella* Typhimurium's Response  
to a Changing Environment through  
Bacterial Enhancer-Binding Proteins  
and the RpoN Regulon.  
Front. Mol. Biosci. 3:41.  
doi: 10.3389/fmolb.2016.00041

Transcription sigma factors direct the selective binding of RNA polymerase holoenzyme ( $E\sigma$ ) to specific promoters. Two families of sigma factors determine promoter specificity, the  $\sigma^{70}$  (RpoD) family and the  $\sigma^{54}$  (RpoN) family. In transcription controlled by  $\sigma^{54}$ , the  $E\sigma^{54}$ -promoter closed complex requires ATP hydrolysis by an associated bacterial enhancer-binding protein (bEBP) for the transition to open complex and transcription initiation. Given the wide host range of *Salmonella enterica* serovar Typhimurium, it is an excellent model system for investigating the roles of RpoN and its bEBPs in modulating the lifestyle of bacteria. The genome of *S. Typhimurium* encodes 13 known or predicted bEBPs, each responding to a unique intracellular or extracellular signal. While the regulons of most alternative sigma factors respond to a specific environmental or developmental signal, the RpoN regulon is very diverse, controlling genes for response to nitrogen limitation, nitric oxide stress, availability of alternative carbon sources, phage shock/envelope stress, toxic levels of zinc, nucleic acid damage, and other stressors. This review explores how bEBPs respond to environmental changes encountered by *S. Typhimurium* during transmission/infection and influence adaptation through control of transcription of different components of the *S. Typhimurium* RpoN regulon.

**Keywords:** *Salmonella* RpoN regulon, sigma 54, bacterial enhancer-binding protein, bEBP, transcription activation, stress adaptation

## INTRODUCTION

*Salmonella enterica* subsp. *enterica* serovar Typhimurium is the most common serotype of *Salmonella enterica* subspecies, which causes tens of millions of cases of salmonellosis and more than 100,000 deaths worldwide each year (Majowicz et al., 2010). *S. Typhimurium* has been extensively studied to reveal the virulence factors and strategies that lead to morbidity and mortality, defining novel mechanisms of bacterial transmission and pathogenesis (reviewed in Fàbrega and Vila, 2013). The response of *S. Typhimurium* to the stresses it encounters in its infectious pathway—from the external environment to the host's intestines—is controlled largely by overlapping transcriptional regulatory systems (reviewed in Runkel et al., 2013).

Transcription in bacteria is carried out by the RNA polymerase core enzyme (RNAP;  $\alpha_2\beta\beta'\omega$ ). However, the core enzyme alone cannot recognize specific promoter sequences; the variable sigma

( $\sigma$ ) subunit confers DNA-binding specificity to ensure that transcription starts at the appropriate promoter sequence (reviewed in Feklistov et al., 2014). RNAP and  $\sigma$  together make up the holoenzyme ( $E\sigma$ ). There are two families of sigma factors: the  $\sigma^{70}$  (RpoD) family and the  $\sigma^{54}$  (RpoN) family. The  $\sigma^{70}$  family includes the housekeeping sigma factor ( $\sigma^{70/D}$ ) and all of the alternative sigma factors, except  $\sigma^{54}$ . These  $\sigma^{70}$ -type sigma factors, which in *Salmonella* include  $\sigma^{70/D}$ ,  $\sigma^{24/E}$ ,  $\sigma^{32/H}$ ,  $\sigma^{38/S}$ , and  $\sigma^{28}$ , exhibit similar structure and recognize promoter sequences with  $-35$  (TTGACA) and  $-10$  (TATAAT) promoter elements that are conserved to varying extents. When  $E\sigma^{70}$  binds to promoter sequences, it initially forms a closed complex, where no DNA melting has occurred. Free energy from specific interactions of  $E\sigma^{70}$  with promoter DNA activate conformational changes in both  $E\sigma^{70}$  and DNA to form a stable open complex in which duplex DNA is opened at the  $+1$  transcription start site and the template strand moves into the active site of RNAP (reviewed in Saecker et al., 2011).

$\sigma^{54}$  is structurally distinct from the  $\sigma^{70}$ -type sigma factors (Yang et al., 2015), thus  $E\sigma^{54}$  recognizes very different promoter elements located at  $-24$  (GC) and  $-12$  (GG) upstream of the transcription start site (Morett and Buck, 1989). When  $E\sigma^{54}$  binds to a promoter, it forms a stable closed complex due to direct interaction of  $E\sigma^{54}$  with two bases within a DNA distortion immediately downstream of the  $-12$  element (Morris et al., 1994). Open complex formation by  $E\sigma^{54}$  requires an activator protein (bacterial enhancer-binding protein; bEBP; Yang et al., 2015). bEBPs are typically found as dimers in the cell but, upon receiving the appropriate cellular signal, they oligomerize into complexes that are competent to bind ATP and interact with enhancer sequences usually located 80–150 bp upstream of the promoter (Figure 1A). A DNA-looping event, often facilitated by integration host factor, brings the bEBP oligomer in contact with  $E\sigma^{54}$  at the promoter (Wedel et al., 1990); bEBP then hydrolyzes ATP, which causes conformational changes in bEBP that trigger remodeling of  $E\sigma^{54}$  and stimulate open complex formation (Chen et al., 2010). Bacteria often have multiple bEBPs that are responsive to different environmental signals and activate transcription of different sets of genes (Francke et al., 2011).

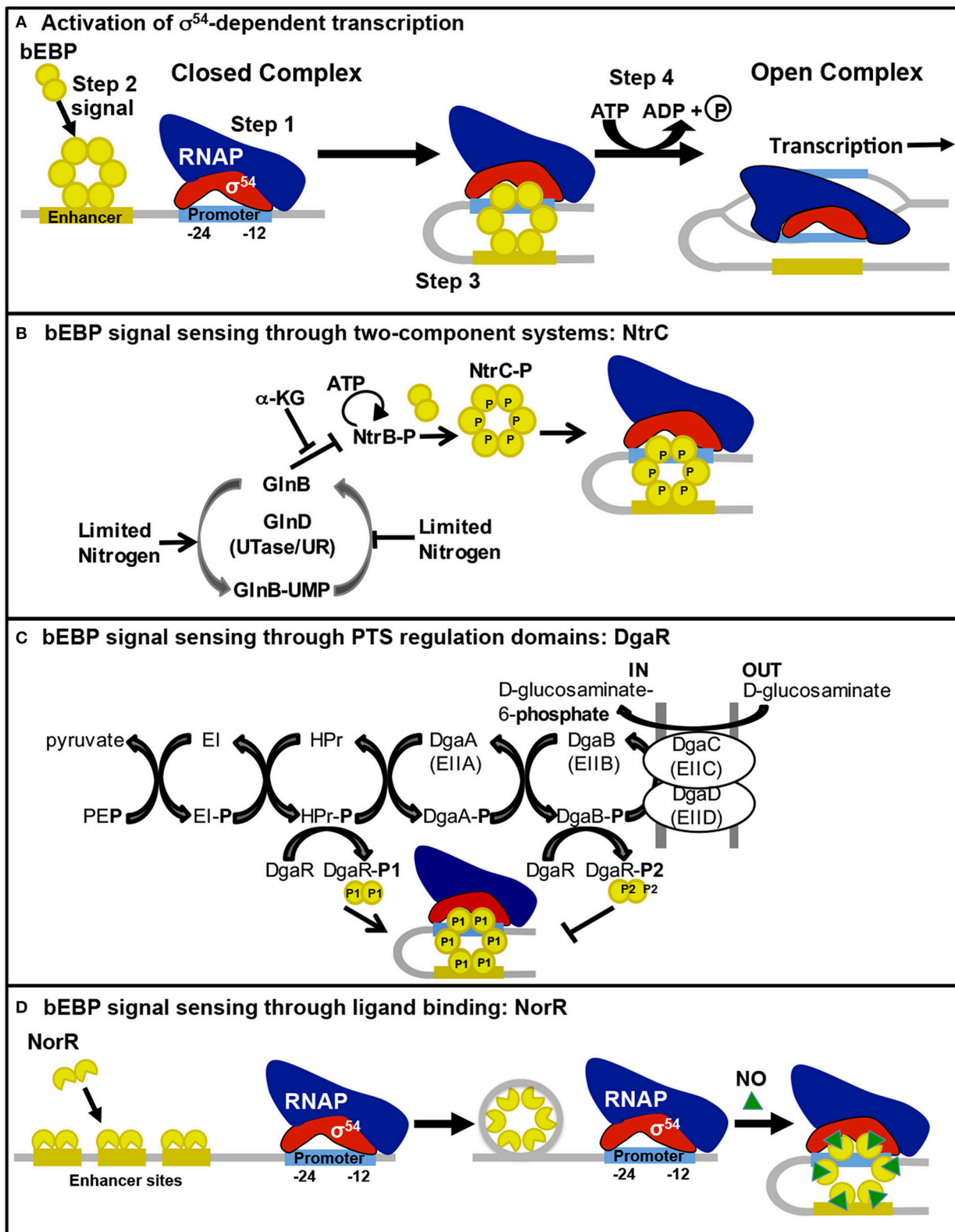
The global RpoN regulon of *S. Typhimurium*, including  $\sigma^{54}$ -dependent transcripts and  $E\sigma^{54}$  chromosomal DNA-binding sites, was characterized in the presence of a promiscuous, constitutively-active bEBP using microarray and ChIP-chip analyses (Samuels et al., 2013). Promoters of this extensive and diverse RpoN regulon in *S. Typhimurium* respond to 1 of 13 known or predicted bEBPs (Table 1; Studholme, 2002). The target promoters and activating environmental stimuli for most of these bEBPs have been demonstrated experimentally or inferred from studies with orthologs in *E. coli* (Table 1). RpoN regulons of *S. Typhimurium* (Samuels et al., 2013) and *E. coli* (Zhao et al., 2010; Bonocora et al., 2015) share many genes/operons (see Table 1); significant differences include the absence in *Salmonella* of *nac*, the LysR-type regulator of multiple operons involved in nitrogen assimilation (Zimmer et al., 2000), and the absence in *E. coli* of the *gfr* operon and *rsr-yrkBA* of the *Salmonella* RNA repair operon (see below). Cellular processes

regulated by  $\sigma^{54}$ -dependent bEBPs in *S. Typhimurium* include nitrogen metabolism in response to limiting nitrogen conditions [NtrC (GlnG, NRI), Keener and Kustu, 1988; Zimmer et al., 2000], transport and catabolism of D-glucosamine (DgaR, Miller et al., 2013) and glucoselysine/fructoselysine (GfrR, Miller et al., 2015), regulation of cytoplasmic pH homeostasis during fermentative growth by the formate-hydrogen lyase system (FhlA, Hopper and Böck, 1995; Lamichhane-Khadka et al., 2015), response to assaults to the cell envelope (PspF, Karlinsey et al., 2010; Flores-Kim and Darwin, 2015 and zinc-dependent ZraR, Appia-Ayme et al., 2012), reduction of nitric oxide under anaerobic conditions (NorR, Hutchings et al., 2002; Mills et al., 2005), propionate catabolism (PrpR, Palacios and Escalante-Semerena, 2000), regulation of amino-sugar synthesis by sRNAs (GlrR; Gopel et al., 2011), and RNA repair/processing (RtcR, Samuels, 2014; Engl et al., 2016). A comprehensive study of the genes that are required for infection of animal hosts by *S. Typhimurium* identified RpoN as important in colonization of chicks, pigs, cattle and mice; transposon mutants in bEBP genes *ntrC* (*glnG*) and *prpR* were attenuated in at least two animal hosts and RpoN-regulated genes *argT*, *glnA*, *glnL*, and *gfrACDEF* (SL1344\_4466, 4468–4471) were attenuated in at least two animal hosts (Chaudhuri et al., 2013).

## BACTERIAL ENHANCER-BINDING PROTEINS OF *S. TYPHIMURIUM* SENSE AND RESPOND TO SIGNALS FOR ADAPTATION IN A CHANGING ENVIRONMENT

bEBPs typically consist of three domains: an N-terminal regulatory domain, a central AAA+ ATPase/transcriptional activation domain, and a C-terminal DNA-binding domain. The N-terminal regulatory domain responds to cellular signals and negatively or positively controls AAA+ domain oligomerization, ATPase activity, and/or interaction with  $\sigma^{54}$ . The central AAA+ ATPase domain is responsible for bEBP oligomerization; association of two AAA+ domains within the bEBP oligomer forms the ATP hydrolysis site. This domain also includes the highly conserved GAFTGA motif that mediates the interaction with  $\sigma^{54}$ . The C-terminal DNA-binding domain contains a helix-turn-helix DNA-binding motif, which determines bEBP specificity for an enhancer. For some bEBPs binding to the enhancer facilitates or stabilizes oligomerization. Consensus enhancer sequences for bEBPs found in *S. Typhimurium* are given in Table 1. Further details on bEBP structure and function are reviewed in (Bush and Dixon, 2012).

The regulatory domains of *S. Typhimurium* bEBPs can function as response-regulator domains of two-component systems (TCS), phosphotransferase regulation domains (PRDs) or ligand-binding domains. One bEBP, PspF, lacks a regulatory domain, but a separate protein, PspA, controls PspF activity. The PspF-PspA system, which is required for *S. Typhimurium* virulence in a mouse model (Karlinsey et al., 2010), is not further discussed in this review, but two recent studies



**FIGURE 1 | Bacterial enhancer-binding protein sensing of environmental signals and activation of  $\sigma^{54}$ -dependent transcription.** The process for bEBP activation of  $\sigma^{54}$ -dependent transcription is illustrated in (A). Step 1,  $E\sigma^{54}$  binds to the promoter in a stable closed complex. Step 2, the bEBP receives a signal from (Continued)



**FIGURE 1 | Continued**

the internal or external environment, becomes active, and binds to an enhancer sequence. Step 3, DNA looping brings the bEBP in contact with  $E\sigma^{54}$ . Step 4, the bEBP hydrolyzes ATP to promote open complex formation. The mechanism for bEBP sensing of environmental signals through **(B)** two-component systems, **(C)** PTS regulatory domains, and **(D)** ligand binding are illustrated here and described in the text.

**TABLE 1 |  $\sigma^{54}$ -dependent genes and associated bEBPs in *S. Typhimurium*<sup>a</sup>.**

Locus tag <sup>b</sup>	Gene symbol	Function	bEBP <sup>c</sup>	bEBP enhancer sequence <sup>d</sup>	Activating signal/condition <sup>e</sup>
STM0368-71	<i>prpBCDE</i>	Propionate catabolism	PrpR	CGTTTCATGAAACG	2-methylcitrate
STM0462	<i>glnK amtB</i>	Regulator of N metabolism; NH <sub>3</sub> transporter	NtrC	TGCACC(A/T) <sub>4</sub> TGGTGCA	Low intracellular glutamine
STM0665-62 <sup>f</sup>	<i>gltJKL</i>	Glutamate/aspartate transporter	NtrC	TGCACC(A/T) <sub>4</sub> TGGTGCA	Low intracellular glutamine
STM0830-28	<i>glnHPQ</i>	Glutamine high-affinity transporter	NtrC	TGCACC(A/T) <sub>4</sub> TGGTGCA	Low intracellular glutamine
STM1285-84 <sup>f</sup>	<i>yeaGH</i>	Serine protein kinase	NtrC	TGCACC(A/T) <sub>4</sub> TGGTGCA	Low intracellular glutamine
STM1303-07	<i>astCABDE</i>	Arginine/ornithine/glutamine metabolism	NtrC	TGCACC(A/T) <sub>4</sub> TGGTGCA	Low intracellular glutamine
STM2355	<i>argT</i>	Lysine/arginine/ornithine transport protein	NtrC	TGCACC(A/T) <sub>4</sub> TGGTGCA	Low intracellular glutamine
STM4007-05	<i>glnALG</i>	Glutamine synthetase	NtrC	TGCACC(A/T) <sub>4</sub> TGGTGCA	Low intracellular glutamine
STM0577-72	-----	Putative PTS	STM0571	NK	NK
STM0649-51	-----	Putative hydrolase, 2-keto-3-deoxygluconate permease	STM0652	NK	NK
STM1690-86	<i>pspABCDE</i>	Phage shock proteins	PspF	TAGTGAATTCGCTAACT	Cell envelope stress
STM4244 <sup>f</sup>	<i>pspG</i>	Phage shock protein	PspF	TAGTGAATTCGCTAACT	Cell envelope stress
STM2360-56	----- <i>ubiX</i>	Amino acid transport	STM2361	NK	NK
STM2840-41	<i>norV ygbD</i>	Nitric oxide reductase	NorR	GT(N) <sub>7</sub> AC	Nitric oxide
STM2843-42	<i>hydN hypF</i>	Putative hydrogenase maturation proteins	FhlA	CATTCGTACGAAATG	Formate
STM2853-44	<i>hycABCDEFGH-I</i>	Hydrogenase 3	FhlA	CATTCGTACGAAATG	Formate
STM2854-58	<i>hypABCDE</i>	Hydrogenase maturation proteins	FhlA	CATTCGTACGAAATG	Formate
STM3521-18	<i>rsr yrfBA rtcBA</i>	Nucleic acid repair/processing	RtcR	NK	Nucleic acid damage
STM3568 <sup>f</sup>	<i>rpoH</i>	Heat shock sigma factor ( $\sigma^{32}$ )	NK	NK	NK
STM3772-66	<i>dgaABCDEF</i>	D-glucosamine utilization	DgaR	NK	D-glucosamine
STM4172	<i>zraP</i>	Zinc-dependent chaperone	ZraR	NK	[Zinc] and cell envelope stress
STM4173-74	<i>zraSR</i>	Zinc-responsive two component system	ZraR	NK	[Zinc] and cell envelope stress
STM4285	<i>fdhF</i>	Formate dehydrogenase	FhlA	CATTCGTACGAAATG	Formate
STM4535-40	<i>gfrABCDEF</i>	Glucoselysine and fructoselysine utilization	GfrR	NK	Glucoselysine, fructoselysine
STM_R0152 <sup>g</sup>	<i>glmY</i>	GlmY sRNA	GlrR	TGTC(N) <sub>10</sub> GACA	NK
STM_R0167 <sup>g</sup>	<i>glmZ</i>	GlmZ sRNA	GlrR	TGTC(N) <sub>10</sub> GACA	NK

<sup>a</sup> $E\sigma^{54}$  binding to promoters for all indicated operons was confirmed in *S. Typhimurium* by ChIP-chip (Samuels et al., 2013);  $\sigma^{54}$ -dependent expression of all genes in *S. Typhimurium* was confirmed by microarray (Samuels et al., 2013), with the few exceptions that are footnoted.

<sup>b</sup>Locus tags for  $\sigma^{54}$ -dependent genes in *S. Typhimurium* LT2 are underlined if found in *E. coli* (solid line if found in most sequenced strains; dashed line if found in few *E. coli* strains; dotted line if only part of the operon is found in *E. coli*).

<sup>c</sup>Known or predicted bacterial enhancer-binding protein (bEBP) that activates the  $\sigma^{54}$ -dependent gene or operon (see text for references). NK, not known.

<sup>d</sup>Consensus enhancer sequence given for each bEBP is based on enhancers associated with one or more of the target promoters in one or more bacterial genus; references: PrpR (Palacios and Escalante-Semerena, 2004), NtrC (Ferro-Luzzi Ames and Nikaïdo, 1985), PspF (Lloyd et al., 2004), NorR (Tucker et al., 2004), FhlA (Leonhartsberger et al., 2000), and GlrR (Gopel et al., 2011). NK, not known.

<sup>e</sup>Specific signal or condition that results in activation of the bEBP (see text for references). NK, not known.

<sup>f</sup>Evidence for expression from the  $\sigma^{54}$ -dependent promoter in *Salmonella* has not been published.

<sup>g</sup> $\sigma^{54}$ -dependent expression in *Salmonella* was shown in Gopel et al. (2011).

provide insight into this anti-activator mechanism for regulating bEBP activity (Flores-Kim and Darwin, 2015; Osadnik et al., 2015). Representative examples for the different mechanisms by which the regulatory domains of bEBPs from *S. Typhimurium* respond to extracellular or intracellular signals are considered here.

## Signal Sensing through Two-Component Systems

*S. Typhimurium* has three bEBPs (NtrC, ZraR, and GlrR) that are response regulators of TCSs, in which a sensor kinase protein recognizes the cellular signal, autophosphorylates, and transfers the phosphate to a conserved aspartate residue of the

response regulator. Phosphorylation of the regulatory domain stimulates the bEBP to interact with enhancer sequence(s) and the  $E\sigma^{54}$  closed complex, activating open complex formation (Figure 1B).

### NtrC (GlnG)

The NtrB-NtrC TCS is activated in response to limited nitrogen conditions. NtrB is the sensor kinase of the TCS. Nitrogen limitation is perceived by the cell as low intracellular levels of glutamine (Ikeda et al., 1996), which stimulates the uridylyltransferase GlnD to uridylylate the P<sub>II</sub> protein GlnB (Jiang et al., 1998). Unmodified GlnB inhibits NtrB kinase activity but GlnB-UMP cannot interact with NtrB, thus allowing autophosphorylation of NtrB and transfer of the phosphate to NtrC (Reitzer, 2003). GlnB also responds to  $\alpha$ -ketoglutarate. During nitrogen limitation, the level of  $\alpha$ -ketoglutarate is high and inhibits GlnB interaction with NtrB, thereby increasing NtrC phosphorylation (Schumacher et al., 2013). Phosphorylation of NtrC dimers results in oligomerization and enhancer binding (Weiss et al., 1991). NtrC-dependent transcription of target genes (Table 1) allows the cell to assimilate low levels of ammonia and utilize alternative nitrogen sources in nutrient-limited environments; NtrC-regulated *glnA* (glutamine synthetase) and *glnHQ* (glutamine transport) together contribute to *S. Typhimurium* virulence in a mouse model and increased survival in macrophages (Klose and Mekalanos, 1997).

### ZraR (HydG)

In *S. Typhimurium*, ZraR is a response regulator, activated by its sensor kinase ZraS in a zinc-dependent response to envelope stress (Leonhartsberger et al., 2001; Appia-Ayme et al., 2012). ZraR controls expression from divergent  $\sigma^{54}$ -dependent promoters for *zraSR* and *zraP*. ZraP encodes a zinc-binding periplasmic protein that acts as a zinc-dependent chaperone in both *S. Typhimurium* and *E. coli*; ZraP responds to misfolding of periplasmic and outer membrane proteins due to envelope stress, such as disruption of the outer membrane by antimicrobial cationic peptides that may be encountered in the environment and/or the host (Appia-Ayme et al., 2012; Petit-Härtlein et al., 2015).

## Signal Sensing through Phosphotransferase Regulation Domains

The bEBPs DgaR, GfrR, and STM0571 of *S. Typhimurium* are members of the family of LevR-like regulators, which previously have only been described in Gram-positive bacteria controlling transcription of the genes for permease components of phosphotransferase systems (PTSs) and enzymes required for utilization of the imported sugar/amino sugar (reviewed in Deutscher et al., 2014). PTSs import and phosphorylate sugars through the Enzyme II complex (EII) membrane-bound components that are linked to a cascade of phosphoryl transfer, beginning with phosphoenolpyruvate as the donor and continuing through Enzyme I (EI), HPr, and finally the EII complex (Figure 1C). These PTS enzymes control the activity of the LevR-like bEBPs through phosphorylation of the regulatory domain. In contrast to most bEBPs, the regulatory

domains of LevR-like bEBPs are found at the C-terminus. These regulatory domains contain two PTS regulation domains (PRDs) with competing activities. HPr-mediated phosphorylation of a conserved histidine residue adjacent to PRD1 leads to activation while EII-mediated phosphorylation of a conserved histidine residue within PRD2 is inhibitory (Martin-Verstraete et al., 1998).

### DgaR

The LevR-like bEBP DgaR is phosphorylated by PTS HPr~P (DgaR-P1), resulting in expression of *dgaABCDEF*, which encodes the permease and catabolic enzymes for D-glucosamine (Miller et al., 2013). When D-glucosamine is present, EII preferentially phosphorylates the sugar, instead of DgaR, to complete the PTS cascade; but in the absence of D-glucosamine, DgaR is phosphorylated by EII (DgaR-P2), which inhibits DgaR activation (Figure 1C; Miller et al., 2013).

*S. Typhimurium* can utilize D-glucosamine as both a carbon and nitrogen source (Miller et al., 2013), so it is likely that this PTS system gives *S. Typhimurium* a competitive advantage over competing microbes under nutrient-limited conditions; the source of D-glucosamine in the environment/host is likely to be other bacteria containing D-glucosamine in lipid A or glucose oxidase that effectively oxidizes D-glucosamine (Miller et al., 2013).

### GfrR

GfrR activates  $\sigma^{54}$ -dependent transcription of the *gfrABCDEF* operon (Miller et al., 2015). GfrR differs from DgaR and other LevR-like bEBPs in its regulatory domain by substitution with tyrosine of the conserved histidine that is normally phosphorylated by HPr~P. By analogy to another LevR-like bEBP, MtlR (Joyet et al., 2015), GfrR is likely controlled solely by the repressive EII-mediated phosphorylation of PRD2; this results in GfrR being insensitive to the catabolite repression observed for DgaR (Miller et al., 2013), in which EI and HPr phosphorylation activity is directed to the uptake of another primary carbon source (glucose) instead of phosphorylation of the bEBP. Thus, *S. Typhimurium* is able to utilize glucose and fructoselysine (or glucoselysine) simultaneously (Miller et al., 2015).

Enzymes encoded by the *gfrABCDEF* operon enable glucoselysine and fructoselysine uptake and catabolism. Glucoselysine and fructoselysine, as well as other Maillard reaction products, are found at varying levels in the gut of human and animal hosts depending on the diet and microbiota (reviewed in Tuohy et al., 2006). The PTS permease and dual deglycases encoded by *gfrABCDEF* give *S. Typhimurium* flexibility in carbon and nitrogen sources, improving persistence in animal hosts (Chaudhuri et al., 2013).

## Signal Sensing through Ligand Binding

In *S. Typhimurium* there are four bEBPs that are known, or predicted, to be regulated by the binding of an effector molecule to the regulatory domain: NorR, FhlA, PrpR, and RtcR. Although the regulatory domain structure is different for each of these bEBPs, in each case ligand binding alters the bEBP structure

such that repression of AAA+ domain oligomerization, ATPase activity, and/or interaction with  $\sigma^{54}$  by the regulatory domain is relieved (**Figure 1D**).

## NorR

NorR stimulates expression of nitric oxide (NO) reductase genes, *norVW*, in response to NO under anaerobic conditions (Gardner et al., 2003). The N-terminal region of NorR contains a GAF (cyclic GMP-specific and stimulated phosphodiesterases, *Anabaena* adenylate cyclases, and *E. coli* FhlA) domain with a non-heme iron center that recognizes NO (D'Autr aux et al., 2005). Binding of NO to the GAF domain relieves repression of the ATPase activity of the AAA+ domain, allowing activation of transcription from the  $\sigma^{54}$ -dependent promoter for *norVW* (D'Autr aux et al., 2005). NorR recognizes three enhancer sequences upstream of the *norVW* operon, all of which are required for transcriptional activation (Tucker et al., 2010). As illustrated in **Figure 1D**, unlike many bEBPs, NorR is able to multimerize in the absence of the activating ligand, forming hexamers through assembly of dimers that are bound to the enhancer sequences. The hexamer-enhancer complex is unable to hydrolyze ATP until activated by NO binding (Bush et al., 2015). It has been suggested that this "pre-activated" complex may exist to enable rapid response to the presence of NO (Bush et al., 2015). NO and other reactive nitrogen species are generated by macrophages during the immune response to infection and have bactericidal and bacteriostatic effects on *Salmonella* (Vazquez-Torres et al., 2000). Transient increased sensitivity of a *norV* mutant to NO suggests that the NorR-regulated NO reductase is part of a multiple enzyme response to NO stress during the infection process (Mills et al., 2005).

## RtcR

RtcR controls  $\sigma^{54}$ -dependent transcription of putative RNA repair operons of *S. Typhimurium* (*rsr-yrlBA-rtcBA*; Chen et al., 2013; Samuels, 2014) and *E. coli* (*rtcBA*; Genschik et al., 1998; Engl et al., 2016). *rtcB* and *rtcA* encode homologs of the metazoan and archaeal RNA ligase and RNA 3'-phosphate cyclase, respectively (Das and Shuman, 2013). *rsr* and *yrlBA* of *Salmonella* encode homologs of metazoan Ro60 and Y-RNAs that form ribonucleoprotein complexes involved in noncoding-RNA quality control (Chen et al., 2013; Wolin et al., 2013). The regulatory domain of RtcR exhibits significant sequence similarity with the CRISPR-associated Rossmann fold (CARF) domains (Makarova et al., 2014). CARF domains are predicted to bind nucleotides, but the RtcR regulatory domain lacks a positively-charged residue involved in nucleotide binding (Makarova et al., 2014). The lack of this residue suggests that RtcR utilizes a different ligand, possibly a nucleoside or modified nucleotide (Makarova et al., 2014).

Metazoan RtcB functions in repair of *xbp-1* mRNA, which is required for the unfolded protein response (Jurkin et al., 2014), as well as tRNA splicing (Popow et al., 2011). RtcA repairs 3'-phosphate or 2'-phosphate ends of cleaved RNA to 2',3'-cyclic phosphates, which can serve as substrates for RtcB-mediated ligation (Remus and Shuman, 2013). RtcB and RtcA from *E. coli*

exhibit the same biochemical activities as the metazoan homologs *in vitro* (Genschik et al., 1998; Tanaka et al., 2011). In addition, *E. coli* RtcB and RtcA utilize DNA substrates; RtcB adds a guanylyl "cap" to a 3'-phosphate end of nicked DNA (Das et al., 2013, 2014), and RtcA adenylates DNA 5'-phosphate ends (Chakravarty and Shuman, 2011). In *S. Typhimurium*, the Rsr-YrlA complex associates with PNPase (polynucleotide phosphorylase; Chen et al., 2013); this is consistent with the activity of Rsr in *Deinococcus radiodurans*, where Rsr forms a ribonucleoprotein complex with YrlA and PNPase and is involved in starvation-induced rRNA degradation (Wurtmann and Wolin, 2010). Additionally, Rsr works with RNase PH and RNase II to fully process 23S rRNAs during growth at elevated temperature (37°C; Chen et al., 2007).

RtcR is activated in *S. Typhimurium* upon exposure to the antibiotic mitomycin C (MMC), stimulating transcription of the *rsr-yrlBA-rtcBA* operon (Samuels, 2014). MMC is an alkylating agent that causes intra- and inter-strand crosslinking in nucleic acids, and results in the formation of DNA-MMC (Bizanek et al., 1992) and RNA-MMC adducts (Snodgrass et al., 2010). MMC induces the SOS response (Kenyon and Walker, 1980), and RtcR activation by MMC is RecA-dependent, suggesting involvement of the SOS response in the activation of RtcR (Samuels, 2014). In *E. coli* RtcR is activated by conditions that disrupt translation, including VapC-mediated cleavage of tRNA<sup>fmet</sup> and treatment with tetracycline (Engl et al., 2016). The signal that is recognized by RtcR in either bacterium is unknown, but candidate signal molecules include: alkylated bases or DNA-MMC adducts removed by nucleotide excision repair in the SOS response (reviewed in Kisker et al., 2013); MMC-modified nucleotides from rRNA or increased free nucleotide/nucleoside intracellular pools upon MMC-induced rRNA degradation (Suzuki and Kilgore, 1967a,b); 2',3'-cyclic NMPs released from RNAs cleaved by toxins of toxin-antitoxin systems, which leave 2',3'-cyclic phosphate at the 3'-end of cleaved RNA (reviewed in Sofos et al., 2015); or modified nucleotides of tRNAs (reviewed in Motorin and Helm, 2010) released by cleavage/degradation. The substrates for RtcA, RtcB, and Rsr-YrlA/B are unidentified in both *S. Typhimurium* and *E. coli*, although ribosome analysis in an *E. coli* *rtcB* mutant suggests a role in 16S rRNA stability (Engl et al., 2016).

## CONCLUSION

The  $\sigma^{54}$  regulon of *S. Typhimurium* is involved in a range of potential stress responses, including nitrogen/carbon limitation, cell envelope stress, nitric oxide stress, and nucleic acid damage/turnover. As summarized in this mini-review, the response to these stresses and the resulting modulation of the *S. Typhimurium* lifestyle are often mediated through bEBPs, which receive signals from the environment through a variety of mechanisms and activate the appropriate components of the  $\sigma^{54}$  regulon. Further characterization of RtcR activation by nucleic acid damage/modification and of the three currently uncharacterized bEBPs (STM0571, STM0652, and STM2361) will give a clearer picture of how bEBPs can

alter the lifestyle of *S. Typhimurium* and other pathogens to improve their chances of survival during the infection process.

## AUTHOR CONTRIBUTIONS

CH, AK, and DS each made substantial intellectual contributions to the work, participated in the writing of the mini-review, and approved it for publication.

## REFERENCES

- Appia-Ayme, C., Hall, A., Patrick, E., Rajadurai, S., Clarke, T. A., and Rowley, G. (2012). ZraP is a periplasmic molecular chaperone and a repressor of the zinc-responsive two-component regulator ZraSR. *Biochem. J.* 442, 85–93. doi: 10.1042/BJ20111639
- Bizaneck, R., McGuinness, B. F., Nakanishi, K., and Tomasz, M. (1992). Isolation and structure of an intrastrand cross-link adduct of mitomycin C and DNA. *Biochemistry* 31, 3084–3091. doi: 10.1021/bi00127a008
- Bonocora, R. P., Smith, C., Lapierre, P., and Wade, J. T. (2015). Genome-Scale mapping of *Escherichia coli*  $\sigma^{54}$  reveals widespread, conserved intragenic binding. *PLoS Genet.* 11:e1005552. doi: 10.1371/journal.pgen.1005552
- Bush, M., and Dixon, R. (2012). The role of bacterial enhancer binding proteins as specialized activators of sigma54-dependent transcription. *Microbiol. Mol. Biol. Rev.* 76, 497–529. doi: 10.1128/MMBR.00006-12
- Bush, M., Ghosh, T., Sawicka, M., Moal, I. H., Bates, P. A., Dixon, R., et al. (2015). The structural basis for enhancer-dependent assembly and activation of the AAA transcriptional activator NorR. *Mol. Microbiol.* 95, 17–30. doi: 10.1111/mmi.12844
- Chakravarty, A. K., and Shuman, S. (2011). RNA 3'-phosphate cyclase (RtcA) catalyzes ligase-like adenylation of DNA and RNA 5'-monophosphate ends. *J. Biol. Chem.* 286, 4117–4122. doi: 10.1074/jbc.M110.196766
- Chaudhuri, R. R., Morgan, E., Peters, S. E., Pleasance, S. J., Hudson, D. L., Davies, H. M., et al. (2013). Comprehensive assignment of roles for *Salmonella* Typhimurium genes in intestinal colonization of food-producing animals. *PLoS Genet.* 9:e1003456. doi: 10.1371/journal.pgen.1003456
- Chen, B., Sysoeva, T. A., Chowdhury, S., Guo, L., De Carlo, S., Hanson, J. A., et al. (2010). Engagement of arginine finger to ATP triggers large conformational changes in NtrC1 AAA+ ATPase for remodeling bacterial RNA polymerase. *Structure* 18, 1420–1430. doi: 10.1016/j.str.2010.08.018
- Chen, X., Taylor, D. W., Fowler, C. C., Galan, J. E., Wang, H. W., and Wolin, S. L. (2013). An RNA degradation machine sculpted by Ro autoantigen and noncoding RNA. *Cell* 153, 166–177. doi: 10.1016/j.cell.2013.02.037
- Chen, X., Wurtmann, E. J., Van Batavia, J., Zybilov, B., Washburn, M. P., and Wolin, S. L. (2007). An ortholog of the Ro autoantigen functions in 23S rRNA maturation in *D. radiodurans*. *Genes Dev.* 21, 1328–1339. doi: 10.1101/gad.1548207
- Das, U., Chakravarty, A. K., Remus, B. S., and Shuman, S. (2013). Rewriting the rules for end joining via enzymatic splicing of DNA 3'-PO4 and 5'-OH ends. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20437–20442. doi: 10.1073/pnas.1314289110
- Das, U., Chauleau, M., Ordonez, H., and Shuman, S. (2014). Impact of DNA3'pp5'G capping on repair reactions at DNA 3' ends. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11317–11322. doi: 10.1073/pnas.1409203111
- Das, U., and Shuman, S. (2013). 2'-Phosphate cyclase activity of RtcA: a potential rationale for the operon organization of RtcA with an RNA repair ligase RtcB in *Escherichia coli* and other bacterial taxa. *RNA* 19, 1355–1362. doi: 10.1261/rna.039917.113
- D'Autréaux, B., Tucker, N. P., Dixon, R., and Spiro, S. (2005). A non-haem iron centre in the transcription factor NorR senses nitric oxide. *Nature* 437, 769–772. doi: 10.1038/nature03953
- Deutscher, J., Aké, F. M., Derkaoui, M., Zébré, A. C., Cao, T. N., Bouraoui, H., et al. (2014). The bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system: regulation by protein phosphorylation and phosphorylation-dependent

## FUNDING

NSF (MCB-1051175) and NIH (R21 AI117102-01A1) grants (to ACK) funded the study of the *Salmonella* RpoN regulon and putative RNA repair operon, respectively.

## ACKNOWLEDGMENTS

We thank Tim Hoover and Maureen Powers for their edits and suggestions for this review.

- protein-protein interactions. *Microbiol. Mol. Biol. Rev.* 78, 231–256. doi: 10.1128/MMBR.00001-14
- Engl, C., Schaefer, J., Kotta-Loizou, I., and Buck, M. (2016). Cellular and molecular phenotypes depending upon the RNA repair system RtcAB of *Escherichia coli*. *Nucleic Acids Res.* doi: 10.1093/nar/gkw628. [Epub ahead of print].
- Fàbrega, A., and Vila, J. (2013). *Salmonella enterica* serovar Typhimurium skills to succeed in the host: virulence and regulation. *Clin. Microbiol. Rev.* 26, 308–341. doi: 10.1128/CMR.00066-12
- Feklistov, A., Sharon, B. D., Darst, S. A., and Gross, C. A. (2014). Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu. Rev. Microbiol.* 68, 357–376. doi: 10.1146/annurev-micro-092412-155737
- Ferro-Luzzi Ames, G., and Nikaido, K. (1985). Nitrogen regulation in *Salmonella typhimurium*. Identification of an ntrC protein-binding site and definition of a consensus binding sequence. *EMBO J.* 4, 539–547.
- Flores-Kim, J., and Darwin, A. J. (2015). Activity of a bacterial cell envelope stress response is controlled by the interaction of a protein binding domain with different partners. *J. Biol. Chem.* 290, 11417–11430. doi: 10.1074/jbc.M114.614107
- Francke, C., Groot Kormelink, T., Hagemeijer, Y., Overmars, L., Sluijter, V., Moezelaar, R., et al. (2011). Comparative analyses imply that the enigmatic Sigma factor 54 is a central controller of the bacterial exterior. *BMC Genomics* 12:385. doi: 10.1186/1471-2164-12-385
- Gardner, A. M., Gessner, C. R., and Gardner, P. R. (2003). Regulation of the nitric oxide reduction operon (norRVW) in *Escherichia coli*. Role of NorR and sigma54 in the nitric oxide stress response. *J. Biol. Chem.* 278, 10081–10086. doi: 10.1074/jbc.M212462200
- Genshik, P., Drabikowski, K., and Filipowicz, W. (1998). Characterization of the *Escherichia coli* RNA 3'-terminal phosphate cyclase and its sigma54-regulated operon. *J. Biol. Chem.* 273, 25516–25526. doi: 10.1074/jbc.273.39.25516
- Göpel, Y., Lüttmann, D., Heroven, A. K., Reichenbach, B., Dersch, P., and Görke, B. (2011). Common and divergent features in transcriptional control of the homologous small RNAs GlnY and GlnZ in *Enterobacteriaceae*. *Nucleic Acids Res.* 39, 1294–1309. doi: 10.1093/nar/gkq986
- Hopper, S., and Böck, A. (1995). Effector-mediated stimulation of ATPase activity by the sigma 54-dependent transcriptional activator FHLA from *Escherichia coli*. *J. Bacteriol.* 177, 2798–2803.
- Hutchings, M. I., Mandhana, N., and Spiro, S. (2002). The NorR protein of *Escherichia coli* activates expression of the flavorubredoxin gene *norV* in response to reactive nitrogen species. *J. Bacteriol.* 184, 4640–4643. doi: 10.1128/JB.184.16.4640-4643.2002
- Ikeda, T. P., Shauger, A. E., and Kustu, S. (1996). *Salmonella* Typhimurium apparently perceives external nitrogen limitation as internal glutamine limitation. *J. Mol. Biol.* 259, 589–607. doi: 10.1006/jmbi.1996.0342
- Jiang, P., Peliska, J. A., and Ninfa, A. J. (1998). Enzymological characterization of the signal-transducing uridylyltransferase/uridylyl-removing enzyme (EC 2.7.7.59) of *Escherichia coli* and its interaction with the PII protein. *Biochemistry* 37, 12782–12794. doi: 10.1021/bi980667m
- Joyet, P., Derkaoui, M., Bouraoui, H., and Deutscher, J. (2015). PTS-Mediated regulation of the transcription activator mtlr from different species: surprising differences despite strong sequence conservation. *J. Mol. Microbiol. Biotechnol.* 25, 94–105. doi: 10.1159/000369619
- Jurkin, J., Henkel, T., Nielsen, A. F., Minnich, M., Popow, J., Kaufmann, T., et al. (2014). The mammalian tRNA ligase complex mediates splicing of XBP1



- mRNA and controls antibody secretion in plasma cells. *EMBO J.* 33, 2922–2936. doi: 10.15252/embj.201490332
- Karlinsky, J. E., Maguire, M. E., Becker, L. A., Crouch, M. L., and Fang, F. C. (2010). The phage shock protein PspA facilitates divalent metal transport and is required for virulence of *Salmonella enterica* sv. Typhimurium. *Mol. Microbiol.* 78, 669–685. doi: 10.1111/j.1365-2958.2010.07357.x
- Keener, J., and Kustu, S. (1988). Protein kinase and phosphoprotein phosphatase activities of nitrogen regulatory proteins NtrB and NtrC of enteric bacteria: roles of the conserved amino-terminal domain of NtrC. *Proc. Natl. Acad. Sci. U.S.A.* 85, 4976–4980. doi: 10.1073/pnas.85.14.4976
- Kenyon, C. J., and Walker, G. C. (1980). DNA-damaging agents stimulate gene expression at specific loci in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 77, 2819–2823. doi: 10.1073/pnas.77.5.2819
- Kisker, C., Kuper, J., and Van Houten, B. (2013). Prokaryotic nucleotide excision repair. *Cold Spring Harb. Perspect. Biol.* 5:a012591. doi: 10.1101/cshperspect.a012591
- Klose, K. E., and Mekalanos, J. J. (1997). Simultaneous prevention of glutamine synthesis and high-affinity transport attenuates *Salmonella* Typhimurium virulence. *Infect. Immun.* 65, 587–596.
- Lamichane-Khadka, R., Benoit, S. L., Miller-Parks, E. F., and Maier, R. J. (2015). Host hydrogen rather than that produced by the pathogen is important for *Salmonella enterica* serovar Typhimurium virulence. *Infect. Immun.* 83, 311–316. doi: 10.1128/IAI.02611-14
- Leonhartsberger, S., Ehrenreich, A., and Böck, A. (2000). Analysis of the domain structure and the DNA binding site of the transcriptional activator FhlA. *Eur. J. Biochem.* 267, 3672–3684. doi: 10.1046/j.1432-1327.2000.01399.x
- Leonhartsberger, S., Huber, A., Lottspeich, F., and Böck, A. (2001). The *hydH/G* Genes from *Escherichia coli* code for a zinc and lead responsive two-component regulatory system. *J. Mol. Biol.* 307, 93–105. doi: 10.1006/jmbi.2000.4451
- Lloyd, L. J., Jones, S. E., Jovanovic, G., Gyaneshwar, P., Rolfe, M. D., Thompson, A., et al. (2004). Identification of a new member of the phage shock protein response in *Escherichia coli*, the phage shock protein G (PspG). *J. Biol. Chem.* 279, 55707–55714. doi: 10.1074/jbc.M408994200
- Majowicz, S. E., Musto, J., Scallan, E., Angulo, F. J., Kirk, M., O'Brien, S. J., et al. (2010). The global burden of nontyphoidal *Salmonella* gastroenteritis. *Clin. Infect. Dis.* 50, 882–889. doi: 10.1086/650733
- Makarova, K. S., Anantharaman, V., Grishin, N. V., Koonin, E. V., and Aravind, L. (2014). CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. *Front. Genet.* 5:102. doi: 10.3389/fgene.2014.00102
- Martin-Verstraete, I., Charrier, V., Stülke, J., Galinier, A., Erni, B., Rapoport, G., et al. (1998). Antagonistic effects of dual PTS-catalysed phosphorylation on the *Bacillus subtilis* transcriptional activator LevR. *Mol. Microbiol.* 28, 293–303. doi: 10.1046/j.1365-2958.1998.00781.x
- Miller, K. A., Phillips, R. S., Kilgore, P. B., Smith, G. L., and Hoover, T. R. (2015). A mannose family phosphotransferase system permease and associated enzymes are required for utilization of Fructoselysine and Glucoselysine in *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.* 197, 2831–2839. doi: 10.1128/JB.00339-15
- Miller, K. A., Phillips, R. S., Mrázek, J., and Hoover, T. R. (2013). *Salmonella* utilizes D-glucosaminatase via a mannose family phosphotransferase system permease and associated enzymes. *J. Bacteriol.* 195, 4057–4066. doi: 10.1128/JB.00290-13
- Mills, P. C., Richardson, D. J., Hinton, J. C., and Spiro, S. (2005). Detoxification of nitric oxide by the flavorebuxin of *Salmonella enterica* serovar Typhimurium. *Biochem. Soc. Trans.* 33(Pt 1), 198–199. doi: 10.1042/BST0330198
- Moret, E., and Buck, M. (1989). *In vivo* studies on the interaction of RNA polymerase-sigma 54 with the *Klebsiella pneumoniae* and *Rhizobium meliloti* *nifH* promoters. The role of NifA in the formation of an open promoter complex. *J. Mol. Biol.* 210, 65–77. doi: 10.1016/0022-2836(89)90291-X
- Morris, L., Cannon, W., Claverie-Martin, F., Austin, S., and Buck, M. (1994). DNA distortion and nucleation of local DNA unwinding within sigma-54 (sigma N) holoenzyme closed promoter complexes. *J. Biol. Chem.* 269, 11563–11571.
- Motorin, Y., and Helm, M. (2010). tRNA stabilization by modified nucleotides. *Biochemistry* 49, 4934–4944. doi: 10.1021/bi100408z
- Osadnik, H., Schöpfel, M., Heidrich, E., Mehner, D., Lilie, H., Parthier, C., et al. (2015). PspF-binding domain PspA1-144 and the PspA.F complex: new insights into the coiled-coil-dependent regulation of AAA+ proteins. *Mol. Microbiol.* 98, 743–759. doi: 10.1111/mmi.13154
- Palacios, S., and Escalante-Semerena, J. C. (2000). *prpR*, *ntrA*, and *ihf* functions are required for expression of the *prpBCDE* operon, encoding enzymes that catabolize propionate in *Salmonella enterica* serovar Typhimurium LT2. *J. Bacteriol.* 182, 905–910. doi: 10.1128/JB.182.4.905-910.2000
- Palacios, S., and Escalante-Semerena, J. C. (2004). 2-Methylcitrate-dependent activation of the propionate catabolic operon (*prpBCDE*) of *Salmonella enterica* by the PrpR protein. *Microbiology* 150(Pt 11), 3877–3887. doi: 10.1099/mic.0.27299-0
- Petit-Härtlein, I., Rome, K., de Rosny, E., Molton, F., Duboc, C., Gueguen, E., et al. (2015). Biophysical and physiological characterization of ZraP from *Escherichia coli*, the periplasmic accessory protein of the atypical ZraSR two-component system. *Biochem. J.* 472, 205–216. doi: 10.1042/BJ20150827
- Popow, J., Englert, M., Weitzer, S., Schleiffer, A., Mierzwa, B., Mechtler, K., et al. (2011). HSPC117 is the essential subunit of a human tRNA splicing ligase complex. *Science* 331, 760–764. doi: 10.1126/science.1197847
- Reitzer, L. (2003). Nitrogen assimilation and global regulation in *Escherichia coli*. *Annu. Rev. Microbiol.* 57, 155–176. doi: 10.1146/annurev.micro.57.030502.090820
- Remus, B. S., and Shuman, S. (2013). A kinetic framework for tRNA ligase and enforcement of a 2'-phosphate requirement for ligation highlights the design logic of an RNA repair machine. *RNA* 19, 659–669. doi: 10.1261/rna.038406.113
- Runkel, S., Wells, H. C., and Rowley, G. (2013). Living with stress: a lesson from the enteric pathogen *Salmonella enterica*. *Adv. Appl. Microbiol.* 83, 87–144. doi: 10.1016/B978-0-12-407678-5.00003-9
- Saecker, R. M., Record, M. T. Jr., and Dehaseth, P. L. (2011). Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J. Mol. Biol.* 412, 754–771. doi: 10.1016/j.jmb.2011.01.018
- Samuels, D. J. (2014). *The RpoN Regulon of Salmonella and its Component RNA Repair System*. Dissertation, University of Georgia, Athens, GA.
- Samuels, D. J., Frye, J. G., Porwollik, S., McClelland, M., Mrázek, J., Hoover, T. R., et al. (2013). Use of a promiscuous, constitutively-active bacterial enhancer-binding protein to define the sigma(54) (RpoN) regulon of *Salmonella* Typhimurium LT2. *BMC Genomics* 14:602. doi: 10.1186/1471-2164-14-602
- Schumacher, J., Behrends, V., Pan, Z., Brown, D. R., Heydenreich, F., Lewis, M. R., et al. (2013). Nitrogen and carbon status are integrated at the transcriptional level by the nitrogen regulator NtrC *in vivo*. *MBio* 4, e00881–e00813. doi: 10.1128/mBio.00881-13
- Snodgrass, R. G., Collier, A. C., Coon, A. E., and Pritsos, C. A. (2010). Mitomycin C inhibits ribosomal RNA: a novel cytotoxic mechanism for bio-reductive drugs. *J. Biol. Chem.* 285, 19068–19075. doi: 10.1074/jbc.M109.040477
- Sofos, N., Xu, K., Dedic, E., and Brodersen, D. E. (2015). Cut to the chase—Regulating translation through RNA cleavage. *Biochimie* 114, 10–17. doi: 10.1016/j.biochi.2015.01.009
- Studholme, D. J. (2002). Enhancer-dependent transcription in *Salmonella enterica* Typhimurium: new members of the sigmaN regulon inferred from protein sequence homology and predicted promoter sites. *J. Mol. Microbiol. Biotechnol.* 4, 367–374.
- Suzuki, H., and Kilgore, W. W. (1967a). Decomposition of ribosomal particles in *Escherichia coli* treated with mitomycin C. *J. Bacteriol.* 94, 666–676.
- Suzuki, H., and Kilgore, W. W. (1967b). Effects of mitomycin C on macromolecular synthesis in *Escherichia coli*. *J. Bacteriol.* 93, 675–682.
- Tanaka, N., Chakravarty, A. K., Maughan, B., and Shuman, S. (2011). Novel mechanism of RNA repair by RtcB via sequential 2',3'-cyclic phosphodiesterase and 3'-Phosphate/5'-hydroxyl ligation reactions. *J. Biol. Chem.* 286, 43134–43143. doi: 10.1074/jbc.M111.302133
- Tucker, N. P., D'Autréaux, B., Studholme, D. J., Spiro, S., and Dixon, R. (2004). DNA binding activity of the *Escherichia coli* nitric oxide sensor NorR suggests a conserved target sequence in diverse proteobacteria. *J. Bacteriol.* 186, 6656–6660. doi: 10.1128/JB.186.19.6656-6660.2004
- Tucker, N. P., Ghosh, T., Bush, M., Zhang, X., and Dixon, R. (2010). Essential roles of three enhancer sites in sigma54-dependent transcription by the nitric oxide sensing regulatory protein NorR. *Nucleic Acids Res.* 38, 1182–1194. doi: 10.1093/nar/gkp1065
- Tuohy, K. M., Hinton, D. J., Davies, S. J., Crabbe, M. J., Gibson, G. R., and Ames, J. M. (2006). Metabolism of Maillard reaction products by the human gut microbiota—implications for health. *Mol. Nutr. Food Res.* 50, 847–857. doi: 10.1002/mnfr.200500126

- Vazquez-Torres, A., Jones-Carson, J., Mastroeni, P., Ischiropoulos, H., and Fang, F. C. (2000). Antimicrobial actions of the NADPH phagocyte oxidase and inducible nitric oxide synthase in experimental salmonellosis. I. Effects on microbial killing by activated peritoneal macrophages *in vitro*. *J. Exp. Med.* 192, 227–236. doi: 10.1084/jem.192.2.227
- Wedel, A., Weiss, D. S., Popham, D., Dröge, P., and Kustu, S. (1990). A bacterial enhancer functions to tether a transcriptional activator near a promoter. *Science* 248, 486–490. doi: 10.1126/science.1970441
- Weiss, D. S., Batut, J., Klose, K. E., Keener, J., and Kustu, S. (1991). The phosphorylated form of the enhancer-binding protein NtrC has an ATPase activity that is essential for activation of transcription. *Cell* 67, 155–167. doi: 10.1016/0092-8674(91)90579-N
- Wolin, S. L., Belair, C., Boccitto, M., Chen, X., Sim, S., Taylor, D. W., et al. (2013). Non-coding Y RNAs as tethers and gates: insights from bacteria. *RNA Biol.* 10, 1602–1608. doi: 10.4161/rna.26166
- Wurtmann, E. J., and Wolin, S. L. (2010). A role for a bacterial ortholog of the Ro autoantigen in starvation-induced rRNA degradation. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4022–4027. doi: 10.1073/pnas.1000307107
- Yang, Y., Darbari, V. C., Zhang, N., Lu, D., Glyde, R., Wang, Y. P., et al. (2015). Structures of the RNA polymerase-sigma54 reveal new and conserved regulatory strategies. *Science* 349, 882–885. doi: 10.1126/science.aab1478
- Zhao, K., Liu, M., and Burgess, R. R. (2010). Promoter and regulon analysis of nitrogen assimilation factor, sigma54, reveal alternative strategy for *E. coli* MG1655 flagellar biosynthesis. *Nucleic Acids Res.* 38, 1273–1283. doi: 10.1093/nar/gkp1123
- Zimmer, D. P., Soupene, E., Lee, H. L., Wendisch, V. F., Khodursky, A. B., Peter, B. J., et al. (2000). Nitrogen regulatory protein C-controlled genes of *Escherichia coli*: scavenging as a defense against nitrogen limitation. *Proc. Natl. Acad. Sci. U.S.A.* 97, 14674–14679. doi: 10.1073/pnas.97.26.14674

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Hartman, Samuels and Karls. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Bacterial RNA Polymerase-DNA Interaction—The Driving Force of Gene Expression and the Target for Drug Action

Jookyung Lee and Sergei Borukhov\*

Department of Cell Biology, Rowan University School of Osteopathic Medicine, Stratford, NJ, USA

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas System located in  
Galveston, USA

### Reviewed by:

Elizabeth Campbell,  
The Rockefeller University, USA  
Bibhusita Pani,  
NYU School of Medicine, USA

### \*Correspondence:

Sergei Borukhov  
borukhse@rowan.edu

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 31 August 2016

**Accepted:** 24 October 2016

**Published:** 09 November 2016

### Citation:

Lee J and Borukhov S (2016) Bacterial  
RNA Polymerase-DNA  
Interaction—The Driving Force of  
Gene Expression and the Target for  
Drug Action. *Front. Mol. Biosci.* 3:73.  
doi: 10.3389/fmolb.2016.00073

DNA-dependent multisubunit RNA polymerase (RNAP) is the key enzyme of gene expression and a target of regulation in all kingdoms of life. It is a complex multifunctional molecular machine which, unlike other DNA-binding proteins, engages in extensive and dynamic interactions (both specific and nonspecific) with DNA, and maintains them over a distance. These interactions are controlled by DNA sequences, DNA topology, and a host of regulatory factors. Here, we summarize key recent structural and biochemical studies that elucidate the fine details of RNAP-DNA interactions during initiation. The findings of these studies help unravel the molecular mechanisms of promoter recognition and open complex formation, initiation of transcript synthesis and promoter escape. We also discuss most current advances in the studies of drugs that specifically target RNAP-DNA interactions during transcription initiation and elongation.

**Keywords:** RNA polymerase, transcription, promoter recognition, transcription initiation, transcription factors

## INTRODUCTION

Bacterial multisubunit DNA-dependent RNA polymerase (RNAP) is the key enzyme of gene expression and a target of regulation. It is responsible for the synthesis of all RNAs in the cell using ribonucleoside triphosphates (NTPs) substrates. The core enzyme consists of five evolutionarily conserved subunits ( $\alpha_2\beta\beta'\omega$ ) with a total molecular weight of  $\sim 380$  kDa (Borukhov and Nudler, 2008). Although catalytically active, the core enzyme alone is unable to recognize specific promoter sequences, or melt the DNA and initiate transcription. For this, it associates with one of several specificity factors,  $\sigma$  (20–70 kDa), to form RNAP holoenzyme ( $\alpha_2\beta\beta'\omega\sigma$ ) (Murakami and Darst, 2003; Decker and Hinton, 2013). In all bacterial species, most housekeeping genes are transcribed by holoenzyme of one major sigma factor, such as  $\sigma 70$  in *E. coli* or  $\sigma A$  in *Thermus thermophilus*. Binding of alternative  $\sigma$  factors generates multiple forms of holoenzyme that can utilize different classes of promoters under various growth conditions and in response to environmental cues. The number of  $\sigma$  factors in different bacterial species varies widely from 1 in *Mycoplasma genitalium* to 7 in *E. coli* to 63 in *Streptomyces coelicolor*. Many regulatory factors besides  $\sigma$  can modulate RNAP's ability to recognize promoters and initiate transcription, modify its enzymatic functions and properties (Gruber and Gross, 2003).

## Overview of RNAP Structure

In the last 16 years, a wealth of structural information on bacterial RNAPs core and holoenzymes was made available. Initially, the high resolution (2.5–4.5 Å) X-ray crystal structures of RNAP and RNAP complexes with nucleic acids, regulatory factors, and small-molecule inhibitors were

obtained using thermophilic organisms, *T. thermophilus* (*Tth*) and *T. aquaticus* (*Taq*). Beginning in 2013, high-resolution structures of *E. coli* (*Eco*) RNAP holoenzyme and its complexes with nucleic acid and inhibitors began to emerge from several groups. Altogether more than 24 high resolution structures of RNAP/RNAP complexes to date have been deposited to database (Zhang et al., 2012, 2014; Bae et al., 2013, 2015a,b), (Bae et al., 2015c; Molodtsov et al., 2013, 2015; Murakami, 2013; Sarkar et al., 2013), (Zuo et al., 2013; Basu et al., 2014; Degen et al., 2014; Feng et al., 2015, 2016; Liu et al., 2015, 2016; Yang et al., 2015b; Zuo and Steitz, 2015). These include: the structures of *Taq* core; *Eco* core with transcription factor, RapA; *Taq*, *Tth*, and *Eco* holoenzymes and open promoter complexes; *Tth* ternary elongation complexes with and without transcription factors (Gfh1, GreA); *Taq* and *Tth* RNAP complexes with small-molecule inhibitors and antibiotics. Structural data compilation was also aided by high-resolution structures (1.8–2.9 Å) of sub-domains of *Eco* RNAP subunits  $\beta'$ ,  $\alpha$ , and  $\sigma$ , and their complexes with DNA and regulatory factors. For a comprehensive list of currently available bacterial RNAP structures, see the recent review (Murakami, 2015). Together with information gained from a wide range of biochemical, biophysical, and genetic studies, these data refine our understanding of bacterial RNAP structure-function and provide a broad view of transcription process and its regulation.

The overall structure of a bacterial RNAP core enzyme resembles a crab claw, with the two clamps representing  $\beta$  and  $\beta'$  subunits (Figure 1). The clamps are joined at the base by the N-terminal domains of  $\sigma$ -dimer ( $\alpha$ NTDs) serving as a platform for RNAP assembly.  $\sigma$ I-NTD and  $\alpha$ II-NTD contact mostly  $\beta$  and  $\beta'$  subunits, respectively. The C-terminal domains of  $\alpha$ -dimer ( $\alpha$ CTD), each tethered to NTD through a flexible linker, project out from the side of RNAP facing upstream DNA. The large internal cleft between  $\beta$  and  $\beta'$  clamps is partitioned into the main “primary channel” that accommodates downstream dsDNA and RNA-DNA hybrid; the “secondary channel,” which serves as the site for NTP entry; and the “RNA exit channel” which is involved in RNA/DNA hybrid strand separation and interactions with RNA hairpins during pausing and termination. The active center is located on the back wall of the primary channel, at the center of the claw, where the catalytic loop with three aspartates holding essential  $Mg^{2+}$  ion resides. A long  $\alpha$ -helical “bridge” (bridge helix, BH) connecting the  $\beta$  and  $\beta'$  clamps, the two flexible  $\alpha$ -helices of the “trigger” loop (TL), and an extended loop (F-loop), together with the catalytic loop, comprise the active center (reviewed in Nudler, 2009). The  $\omega$  subunit is bound near the  $\beta'$  C-terminus at the bottom pincer, serving as a  $\beta'$  chaperone.

In the structure of  $\sigma^{70}$ -holoenzyme, the bulk of the  $\sigma$  subunit (domains  $\sigma 1$ – $\sigma 3$ ) is bound on the core surface at the entrance to the major cleft, except for the linker connecting  $\sigma$  domains 3 and 4 ( $\sigma 3$ –4 linker containing conserved region  $\sigma 3.2$ ), which threads through the primary channel, reaches the catalytic pocket with its hairpin loop ( $\sigma$  finger), and comes out from the RNA exit channel, almost completely blocking it (Figure 2). The rest of  $\sigma$  is wedged between the  $\beta$  and  $\beta'$  clamps at the upstream side of the core enzyme, creating a wall that partially blocks the opening of the primary channel. Transition from core to

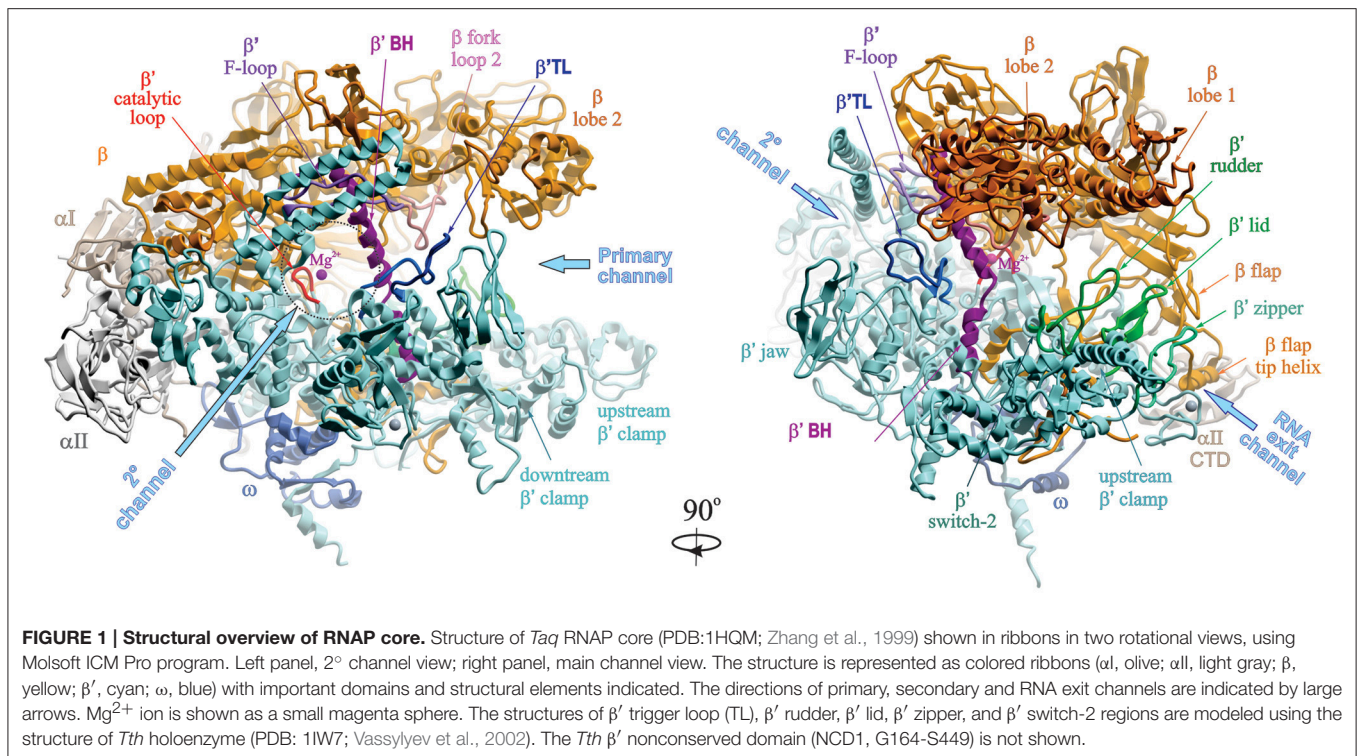
holoenzyme is accompanied by partial closing of the  $\beta, \beta'$  clamps by  $\sim 5$  Å and movement of the flap domain (tip helix) induced by  $\sigma 4$  by  $\sim 12$  Å (Vassylyev et al., 2002). The  $\sigma 2$ ,  $\sigma 3$ , and  $\sigma 4$  domains are optimally positioned to contact the  $-10$ , extended  $-10$ , and  $-35$  elements of the promoter DNA, respectively. In the crystal structures of *Eco* holoenzyme, consistent with previous biophysical studies (Mekler et al., 2002),  $\sigma$  region 1.1 is located in the downstream dsDNA binding region, blocking the access to DNA (Bae et al., 2013; Murakami, 2013). This location of  $\sigma 1.1$  explains why nonspecific transcription initiation by  $\sigma^{70}$ -holoenzyme at promoterless DNA sequences is very low (Shorenstein and Losick, 1973). The  $\sigma$ -core interface is extensive with multiple cooperative contacts (Sharp et al., 1999; Gruber et al., 2001; Murakami and Darst, 2003), explaining the high stability of the  $\sigma$ -core association ( $K_D \sim 0.3$  nM; Maeda et al., 2000). However, most of these contacts appear to be relatively weak (Vassylyev et al., 2002; Borukhov and Nudler, 2003), which allows alternate  $\sigma$  factors to successfully compete for binding to core. The conserved regions 2.1 and 2.2 of  $\sigma 2$  make the most stable contacts with the upstream  $\beta'$  clamp helices—the major  $\sigma$  docking site (Figure 2).  $\sigma 4$  interacts with the  $\beta$ -flap domain, with the C-terminus of  $\sigma$  contacting the  $\beta$ -flap tip. In the presence of specific activators,  $\sigma 4$  also interacts with  $\sigma$ I-CTD. In the recent structure of  $\sigma^S$ -initiation complex,  $\sigma^S$  regions from 1.2 to 4.2 display the same fold as  $\sigma^{70}$ , including the linker 3.2 that inserts into the active site pocket (Liu et al., 2016).  $\sigma^S$  lacks the non-conserved domain present in  $\sigma^{70}$ , which may explain its lower binding affinity to core ( $K_D \sim 4$  nM) (Maeda et al., 2000).

## Overview of the Transcription Cycle

Transcription process consists of three major stages: initiation, elongation, and termination. In bacteria, initiation occurs through five steps (Figure 3; reviewed in Murakami and Darst, 2003; Saeccker et al., 2011).

*First*, RNAP core enzyme, composed of five subunits ( $\alpha_2\beta\beta'\omega$ ), binds one of several specificity factors,  $\sigma$  (such as  $\sigma^{70}$  for transcription of housekeeping genes in *Escherichia coli*) to form holoenzyme ( $E\sigma^{70}$ ). *Second*,  $E\sigma^{70}$  recognizes and binds promoter DNA, a pair of conserved hexameric sequences present at positions  $-35$  and  $-10$  relative to the transcription start site (TSS), where it forms a closed promoter complex (RPC). Sequences immediately upstream and downstream of the  $-10$  element including “ $-15TG^{-14}$  extended  $-10$ ” (Keilty and Rosenberg, 1987), “ $-15$  enhancer” (or “ $-17/-18$  zipper-binding”; Liu et al., 2004; Yuzenkova et al., 2011) and “ $-6GGGA^{-3}$  discriminator” (Feklistov et al., 2006; Haugen et al., 2006) regions, and A/T-rich regions upstream of the  $-35$  element [“UP-element” at  $-45$ ;  $-65$  (Ross et al., 1993)], also contribute to specific recognition by the  $E\sigma^{70}$  (reviewed in Decker and Hinton, 2013; Feklistov et al., 2014). *Third*, RPC undergoes a series of conformational changes (isomerization) through several transition states (such as intermediate complex,  $RP_i$ ), to form an open promoter complex ( $RP_o$ ). Isomerization results in unwinding of DNA duplex around the  $-10$  region (typically, between nt  $-11$  and  $+2$ – $+4$ ) and creates a 12–15 nt long transcription bubble, a hallmark of  $RP_o$ . *Fourth*, in the presence of rNTPs,  $RP_o$  converts to an initial transcribing



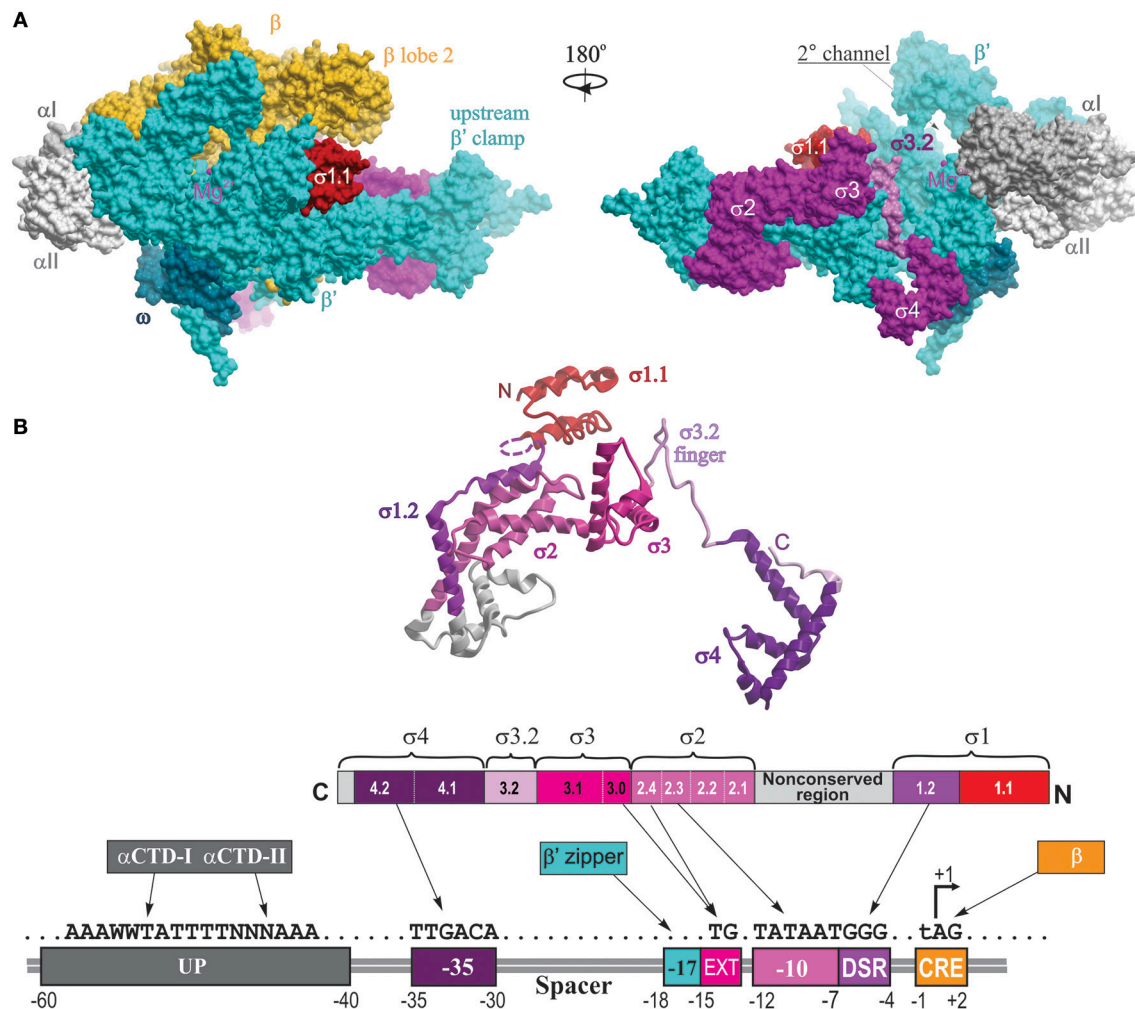


complex ( $RP_{init}$ ), forms the first phosphodiester bond between rNTPs positioned at +1 and +2 sites and then begins the RNA synthesis. During the synthesis beyond dinucleotide,  $RP_{init}$  undergoes “scrunching” whereby the downstream DNA (from +2 to +15) is pulled into the enzyme to be transcribed, resulting in bubble expansion (up to  $\sim 25$  nt), while the upstream DNA-RNAP contacts remain intact (Kapanidis et al., 2006; Revyakin et al., 2006). At the same time, growth of nascent RNA beyond 6-mer is obstructed by the presence of  $\sigma 3-4$  linker in the RNA exit channel (Basu et al., 2014; Bae et al., 2015b). Biochemical data suggest that stress associated with DNA scrunching, and more importantly the steric clash between RNA 5'-end and  $\sigma 3-4$  ( $\sigma$  finger), together cause RNAP to repetitively synthesize and release short RNAs without leaving the promoter (abortive initiation; Sen et al., 1998; Murakami et al., 2002b; Kulbachinskiy and Mustaev, 2006; Samanta and Martin, 2013; Winkelman et al., 2016b). In the final step, the enzyme synthesizes an RNA of a critical length (typically 11–15 nt, of which  $\sim 9$  nt are in the transcription bubble as RNA–DNA hybrid), removes the exit channel blockage, and escapes from promoter, entering the elongation stage of transcription. During this transition, RNAP undergoes global conformational change, which leads to the loss of RNAP-promoter DNA contacts, gradual  $\sigma$  dissociation, and formation of a highly stable and processive ternary elongation complex (EC) (Murakami and Darst, 2003).

Throughout elongation stage, the size of the transcription bubble in the EC remains constant at  $\sim 12 \pm 1$  nt, and the size of RNA/DNA hybrid is maintained at  $\sim 9-10$  bp (Svetlov and Nudler, 2009; Kireeva et al., 2010). Elongating RNAP can transcribe DNA over long distances ( $>10,000$  bp) without

dissociation and release of RNA product. However, elongation does not proceed at a uniform rate; monotonous movement of RNAP can be interrupted by various roadblocks imposed by certain DNA sequences, DNA topology, lesions in transcribed DNA template, RNA secondary structures, DNA-binding proteins, DNA replication and repair machineries, ribosomes, other transcription complexes, RNAP-binding transcription factors (including  $\sigma 70$  that can be recruited back to EC upon encountering promoter-like sequences; Goldman et al., 2015; Sengupta et al., 2015), and small-molecule effectors (ppGpp) (Landick, 2006; Perdue and Roberts, 2011; Nudler, 2012; Imashimizu et al., 2014; Belogurov and Artsimovitch, 2015; Kamarthapu et al., 2016). Eventually, RNAP encounters a termination signal—a 20–35 nt-long G/C-rich RNA sequence of dyad symmetry that forms a hairpin structure immediately followed by a 7–9 nt-long stretch of Us (Yarnell and Roberts, 1999). During termination, RNAP releases the nascent transcript and dissociates from the DNA template, after which it can rebind a  $\sigma$  factor and start a new round of transcription. Under certain conditions, transcription termination can also be induced by termination factors  $\rho$  and Mfd (Roberts and Park, 2004; Kriner et al., 2016).

In this review, we discuss recent findings that elucidate molecular details of RNAP-DNA interactions during initiation the transcription cycle. Specifically, we will describe most current advances in the structural and biochemical studies of the molecular mechanisms underlying promoter recognition and RPo formation, activation of initiation, and promoter escape. Finally, we will review the mechanisms of action of known antibacterial drugs that specifically target RNAP.



**FIGURE 2 | Structural overview of RNAP holoenzyme and  $\sigma$ -DNA interactions.** (A) The structure of *Tth*  $\sigma^A$ -holoenzyme is shown in molecular surface views with color coding as follows:  $\alpha I$ , slate gray;  $\alpha II$ , light gray;  $\beta$ , yellow;  $\beta'$ , cyan,  $\sigma$ , magenta,  $\omega$ , dark cyan. Locations of conserved  $\sigma$  domains are indicated. The catalytic  $Mg^{2+}$  ion is shown as a small magenta sphere. The N-terminal domain of *Eco*  $\sigma^{70}$  carrying region 1.1 is modeled from the structure of *Eco* holoenzyme (PDB: 4YG2, Murakami, 2013) and shown as a red surface. Left panel, 2° channel view (as in Figure 1); right panel is obtained by rotation of the left panel view by 180° around the vertical axis, with the  $\beta$  subunit removed to reveal the location of  $\sigma 3.2$  finger region (colored light magenta) and  $\sigma 4$  occupying the RNA exit channel. (B) The structural and functional organization of  $\sigma$ . Top, a ribbon view of  $\sigma^A$  from *Tth* holoenzyme structure (PDB: 1IW7; Vassilyev et al., 2002) shown on the right panel in (A). Colored regions correspond to the evolutionarily conserved domains of  $\sigma$  as shown in the functional map of  $\sigma 70$  below. Bottom diagram, a linear representation of  $\sigma$  polypeptide with structural domains and conserved regions shown as numbered and color-coded boxes. Underneath is a diagram of DNA promoter regions showing interactions made by DNA-binding domains of  $\sigma$ ,  $\alpha$ CTD  $\beta'$  zipper, and CRE-binding  $\beta$  lobe-2 elements.

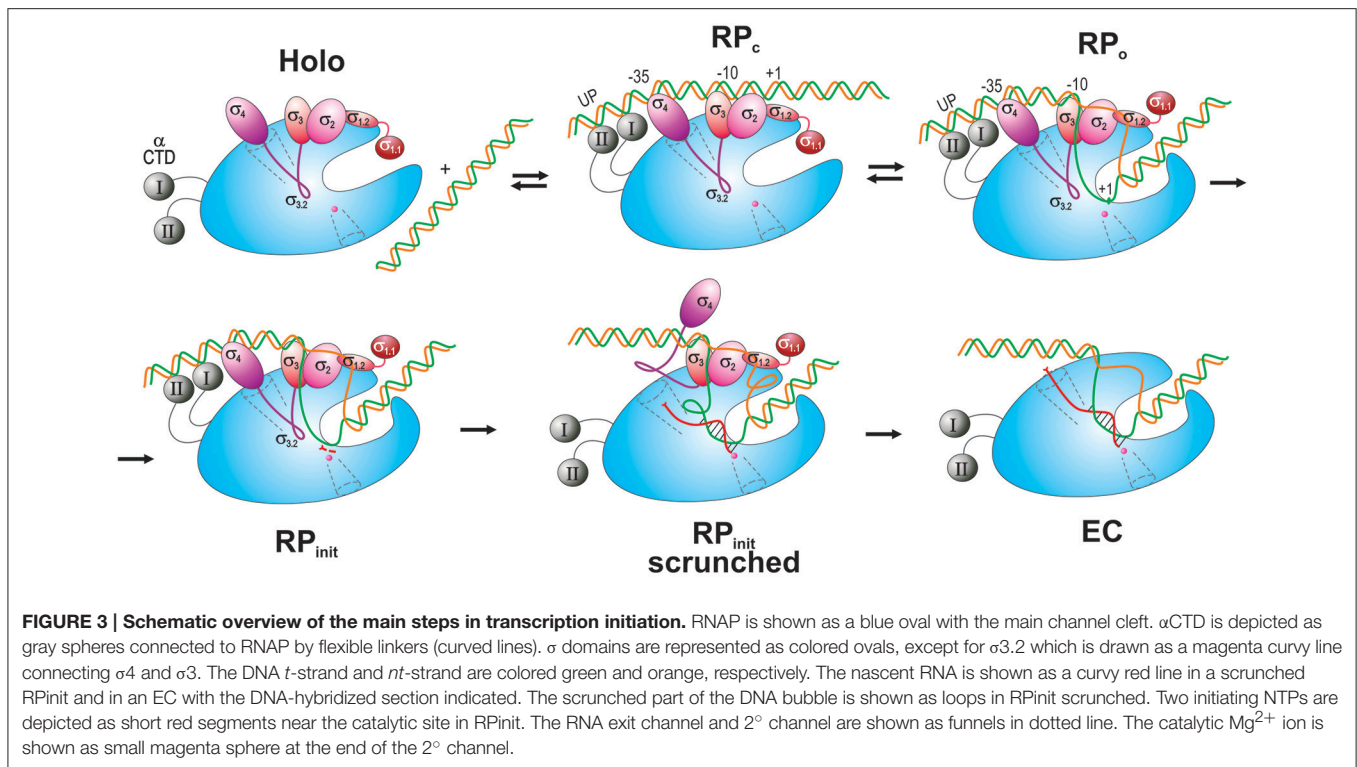
## MOLECULAR DETAILS OF TRANSCRIPTION INITIATION AND ITS ACTIVATION

### Structure of Holoenzyme-Promoter DNA Complexes

#### Promoter Recognition and Binding: Closed Promoter Complex (RP<sub>C</sub>)

Currently, there are no high resolution structure of the RP<sub>C</sub>, but a model of *Taq* RP<sub>C</sub> based on the existing structural, biochemical, biophysical, and genetic data has been proposed (Murakami et al., 2002a; Murakami and Darst, 2003). In the

model, promoter dsDNA rests on the outer surface of the RNAP main channel, bound mostly by  $\sigma$  (Figure 3). RNAP contacts with  $-10$ , extended  $-10$ , and  $-35$  elements of the promoter are established by  $\sigma 2$ ,  $\sigma 3$ , and  $\sigma 4$  (regions 2.2–2.4, 3.0, and 4.2, respectively) through polar and van der Waals contacts. Additionally, residues of two  $\alpha$ CTD helix-hairpin-helix motifs (*Eco* R265, N294, and K298) may contact A/T-rich sequences in the minor groove of the UP element at positions from  $-40$  to  $-60$  and up to  $-90$  (Ross et al., 2001; Benoff et al., 2002). These weak but specific UP/ $\alpha$ CTD interactions contribute to RP<sub>C</sub> formation (Haugen et al., 2008). However, with the exception of the  $-35$  element/ $\sigma 4$  and the  $-12$  bp of  $-10$  element/ $\sigma 2$  contacts, other



upstream dsDNA-RNAP interactions are mostly non-specific and weak, which makes  $RP_c$  intrinsically unstable. Nonetheless, these interactions may provide initial promoter recognition and increase its occupancy by RNAP. They may also induce local distortion in DNA structure facilitating subsequent steps in transcription initiation: DNA melting, strand separation and template strand insertion into the active site cleft. For instance, the RNAP-bound DNA in the  $RP_c$  appears to be bent or kinked at three places: around  $-25$ , to accommodate variable spacer length, at  $-35$ , induced by insertion of  $\sigma_4$  helix-turn-helix motif into the major groove, and at  $-45$ , induced by  $\alpha$ CTD-DNA minor groove interactions (Benoff et al., 2002). The DNA bending at  $-35$  aids in the proper binding of upstream DNA by  $\alpha$ CTD and upstream transcription activators. Additionally, recent structural and biochemical data on  $RPO$  (see below) implicate conserved residues of  $\sigma_3$  region 3.0 (*Tth* H278 and R274) and  $\beta'$  zipper (*Tth* Y34 and R35, T36, and L37) in the recognition of non-canonical  $-17/-18$  “Z-element” which contribute to promoter recognition and binding in  $RP_c$  (Yuzenkova et al., 2011; Bae et al., 2015b). Notably,  $\sigma_{1.1}$  blocks the access to downstream (ds and ss) DNA in the main channel by binding with its negatively charged face (mimicking DNA) to the basic surface of the  $\beta$  lobe-2 and the downstream  $\beta'$  clamp (Murakami, 2013). However, the opposite, positively charged face of  $\sigma_{1.1}$  is positioned to interact with the downstream DNA which may further stabilize  $RP_c$ . Subsequent steps leading to displacement of  $\sigma_{1.1}$  by downstream DNA during transition to  $RPO$  are poorly understood, but are thought to involve  $\beta'$ -clamp opening triggered by upstream promoter DNA binding and initial DNA unwinding around  $-11$  (see below).

Recently an alternative view on initial promoter recognition and  $RP_c$  formation was proposed based on structural and biochemical data (Feklistov and Darst, 2011; Zhang et al., 2012; reviewed in Hook-Barnard and Hinton, 2007; Decker and Hinton, 2013; Feklistov, 2013). In this view, except for direct RNAP recruitment by transcription activators that provide sequence-specific DNA recognition, the upstream DNA-RNAP interactions (involving UP and  $-35$  elements) play only an auxiliary role in  $RP_c$  formation. Instead, initial promoter binding/recognition is accomplished through (i) indirect readout of DNA shape (a distinctive conformational patterns of stacked bases in dsDNA) by RNAP, and (ii) by direct readout of an indispensable  $-10$  element by  $\sigma_2$ -specific interaction with two flipped-out consensus nucleotides,  $-11(A)$  and  $-7(T)$ , of nt-strand DNA (see below). These two views on the mechanism of promoter recognition are not mutually exclusive and could be eventually addressed when the structure of  $RP_c$  becomes available.

### Advances in Structural Studies of Open Promoter Complex ( $RPO$ )

In the last 3 years several high resolution structures of *Tth*, *Taq*, and *Eco*  $RPO$  with different DNA scaffolds have been solved. These include upstream fork and downstream fork promoter DNA (Murakami et al., 2002a; Zhang et al., 2012, 2014; Basu et al., 2014), and complete transcription bubble promoter template with upstream and downstream dsDNA (Bae et al., 2015b; Zuo and Steitz, 2015; Liu et al., 2016). The structures of  $RPO$  correlate well and complement each other. Taken together, they reveal the positions of ds and ssDNA (from  $-36$  to  $+12$ ) in the complex



and the key residues of RNAP that make critical interactions with DNA and RNA. Unlike R<sub>Pc</sub>, in R<sub>Po</sub> both strands of downstream dsDNA up to +12 are fully enclosed inside the RNAP main channel (**Figure 4**). In R<sub>Po</sub>, RNAP makes tight contacts with DNA from −36 to −30 and −18 to +9, in agreement with DNA footprinting and crosslinking data (Ross and Gourse, 2009; Winkelman et al., 2015). The RNAP interactions with the upstream portion of dsDNA (from −36 to −12) are similar to that shown in R<sub>Pc</sub> model, however, at −13/−12 the DNA bends sharply by ~90° toward the RNAP. At position −11, the t- and nt-strands separate, and enter different paths for ~13 downstream nucleotides until they form dsDNA at position +3, thus creating the “transcription bubble.”

#### Upstream DNA (−35, −17/−18, and extended −10 elements)

In two R<sub>Po</sub> structures that contain a full bubble DNA (Bae et al., 2015b; Zuo and Steitz, 2015), contacts made at −35 region are the same as in isolated  $\sigma 4$  domain/−35 element (**Figure 4A**) (Campbell et al., 2002). As proposed for R<sub>Pc</sub>, R<sub>Po</sub> structure demonstrates that duplex DNA upstream of the −10 element (−18 to −12) makes functionally important contacts with conserved residues of  $\beta'$  zipper,  $\sigma 3$ , and  $\sigma 2$  (**Figure 4B**), mostly through phosphate backbone of the nt-strand (Yuzenkova et al., 2011; Bae et al., 2015b; Feng et al., 2016). These contacts were not visible in the low resolution structures of RNAP with nucleic acids. The sequence specific recognition of extended −10 element (−15T:A, −14G:C) by conserved residues of  $\sigma 2/\sigma 3$ , E281 (*Eco* E458), R264 (*R441*), and V277 (*V454*), stabilizes R<sub>Po</sub> and can substitute for a poor or absent −35 element (Keilty and Rosenberg, 1987). Mutational analysis shows that all three residues are essential for promoter recognition (Daniels et al., 1990).

#### −10 element

The initial DNA melting starts at the A:T bp at position −11 (Chen and Helmann, 1997; Lim et al., 2001), when −11A flips out from the duplex DNA, and continues downstream to +1. Two groups of aromatic residues are involved in the initial stages of DNA melting (**Figure 4C**). First, W256 and W257 forming a chair-like structure interact with base-paired −12T at the upstream edge of the bubble replacing the flipped-out −11A. The second group of aromatic residues Y253, F242, and F248, together with two polar residues R246, E243 form a pocket that captures the flipped-out −11A base. Additionally, in the context of a true promoter, the −11T on the t-strand orphaned by the flipping of −11A, may be stabilized by stacking interaction with another highly conserved aromatic residue Y217 (Bae et al., 2015b). Neither the W256A nor Y217A substitution affected the promoter binding, but rather decreased the rate of R<sub>Pc</sub>→R<sub>Po</sub> isomerization. Based on this, it is proposed that these residues maintain the ds-ss junction at the upstream edge of the bubble, preventing bubble collapse and R<sub>Po</sub> dissociation. In another structure of R<sub>Po</sub> with full bubble DNA and activator protein, R258 (*R436*) stacks on −12A of t-DNA, facilitating flipping of the t-strand −11 base (Feng et al., 2016). The second canonical nucleotide of the −10 element, −7T of nt-strand, is flipped out

and captured in a pocket made of five  $\sigma 2$  and  $\sigma 1.2$  residues: E114, N206, L209, K249, and S251 (*Eco* E116, N383, R385, L386, S428; **Figure 4C**), all of which are functionally important (Zhang et al., 2012). Other nucleotides in −10 element make mostly nonspecific contacts with  $\sigma 2$ .

#### Discriminator

The “discriminator” (DSR) region of nt-strand (consensus sequence −6G, −5G, −4G) interacts with eight  $\sigma 1.2$  residues, of which *Tth* L100 (*Eco* M102) and *Tth* H101 (*Eco* R103) provide the most functionally important contacts (**Figure 4D**). The purine-rich DSR contributes to the high stability of R<sub>Po</sub>, whereas pyrimidine-rich sequence in this region destabilizes it (Haugen et al., 2006). This is due to change in the interaction made by the key nucleotides of DSR (−5) with  $\sigma$ ; when it is G, it forms and maintains an ordered, stacked conformation of the nt-strand, but when it is C, it flips and is captured by a pocket in  $\sigma 2$ , resulting in unstacking and compaction of the downstream ssDNA. Importantly, the presence of C in the middle of DSR in rRNA promoters is also one of the determinants of the transcription start site (TSS) selection (Haugen et al., 2006; Winkelman et al., 2016b). The DSR- $\sigma 1.2$  interaction is a major determinant of the susceptibility of rRNA and other promoters to negative regulation by ppGpp and DksA (Haugen et al., 2006).

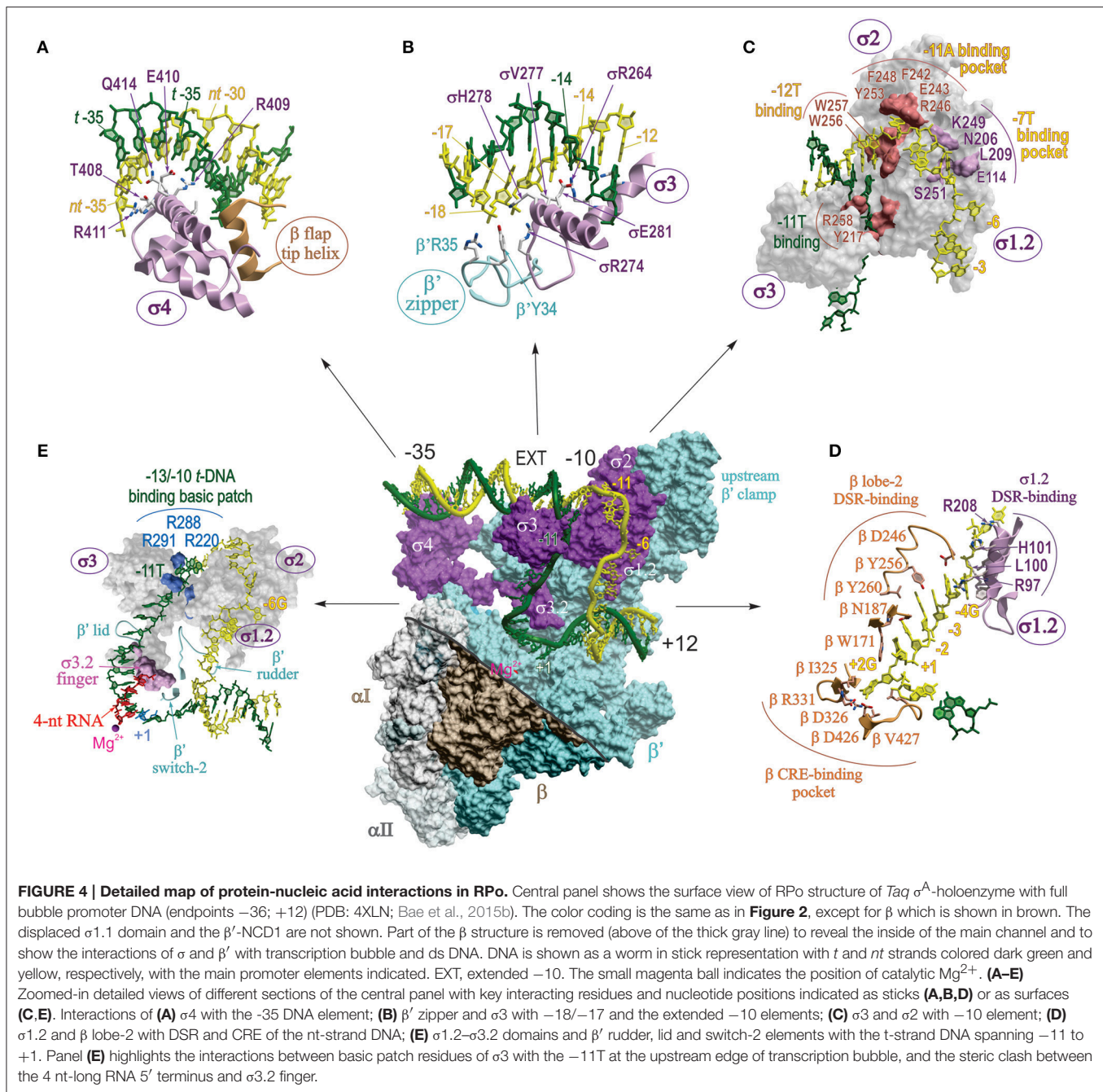
#### CRE

Nucleotides at positions −3 to +2 on nt-strand constitute “core recognition element” (CRE) that interacts specifically with 10 residues of RNAP  $\beta$ -subunit. Six of these residues form a pocket that captures the flipped-out +2G of CRE at the downstream edge of the bubble, reminiscent of the capturing of flipped-out −11A by  $\sigma 2$ . In the pocket, *Tth*  $\beta$ D326 (*Eco* D446) makes a hydrogen bond with +2G, which proved to be most critical (Zhang et al., 2012). Adjacent residue *Tth*  $\beta$ W171 (*W183*) unstacks the +1T away from the +2G, facilitating its placement into  $\beta$  pocket. In addition to stabilizing and maintaining transcription bubble in R<sub>Po</sub>, CRE-core interaction affects sequence-specific pausing, and determines TSS selection (Vvedenskaya et al., 2016; Winkelman et al., 2016b). Moreover, it is predicted that CRE-RNAP interactions will affect all stages of transcription where unwound transcription bubble is involved, e.g., slippage, abortive synthesis, promoter escape, factor dependent pausing, termination.

#### T-strand

A cluster of conserved, positively charged residues of  $\sigma 2.4$  and  $\sigma 3.0$  (*Taq* R288, and R291, R220) pulls the DNA t-strand from −13 to −10, bending it by 90° through electrostatic interaction with the phosphate backbone, into a groove formed by  $\sigma 3$  linker,  $\beta'$  lid, and the  $\beta'$  rudder (**Figure 4E**; Bae et al., 2015b; Zuo and Steitz, 2015). The t-strand DNA (−9 to −5) is then placed into the main channel between the active site wall, mostly of  $\beta$ , and  $\sigma 3.2$  hairpin loop ( $\sigma$  finger), which participates in juxtaposing DNA +1 position to the catalytic center (**Figure 4E**). Simultaneously, the dsDNA downstream of +3 to +12 is brought inside the downstream DNA binding clamp between the  $\beta'$  jaw,  $\beta$  lobe-2 and the downstream  $\beta'$  clamp. The  $\beta'$  switch-2, a small flexible





loop residing in the upstream  $\beta'$  clamp in the middle of the main channel cavity, controls the binding of the downstream part of the unwound DNA t-strand in the active site cleft.  $\beta'$  switch-2 functions as a hinge mediating opening and closing of the  $\beta'$  clamp, and plays a critical role in downstream propagation of transcription bubble during formation of RPo (Mukhopadhyay et al., 2008; Belogurov et al., 2009; Bae et al., 2015b).

#### Role of $\sigma 3.2$ finger in $RP_{init}$

Recent structural studies showed that during initial transcript synthesis,  $\sigma 3.2$  loop physically occupies the path of nascent

RNA and sterically blocks its extension beyond 4~5 nucleotides (Zhang et al., 2012; Basu et al., 2014; Bae et al., 2015b; Zuo and Steitz, 2015). Consistent with its position in the structure of RPo and  $RP_{init}$ , biochemical data show that  $\sigma 3.2$  finger positively affects the binding of the first two initiating NTPs, abortive RNA synthesis, and promoter escape (Murakami et al., 2002b; Kulbachinskiy and Mustaev, 2006). More recent studies revealed that  $\sigma 3.2$  contributes to promoter opening (Morichaud et al., 2016), suppresses  $\sigma$ -dependent promoter proximal pausing, and accelerates  $\sigma$  dissociation during transition from initiation to elongation (Pupov et al., 2014).

### Transcription start site selection (role of DNA scrunching in RPo)

Transcription typically initiates 7 or 8 bp downstream from the  $-10$  element, with a strong bias for purine (R) over pyrimidine (Y) as the initiating nucleotide (+1 position) (Shultzaberger et al., 2007). To identify the determinants for TSS selection, Nickels and coworkers combined a high throughput sequencing (MASTER, “massively systematic transcript end readout”) with multiplexed site-specific RNAP-DNA crosslinking and X-ray structural analyses, to dissect and characterize the role of sequence variation within  $-6$  to  $+4$  positions of the promoter on TSS selection (Winkelman et al., 2016b). The studies identified DSR and CRE as sequence elements that significantly influence TSS selection. G-rich DSR (GGG) and  $+2$ G CRE shortens the distance between TSS and  $-10$  element (6-/7-nt from the edge of  $-10$  element), whereas pyrimidine-rich DSR (CCC) and lack of CRE shifts TSS further downstream (8-/9-nt from the edge of  $-10$  element). Disrupting the DSR- $\sigma 1.2$  and/or CRE- $\beta$  pocket interactions results in downstream shift of TSS. The changes in TSS correlate with the corresponding shift in the downstream edge of the transcription bubble (in  $+$  or  $-$  direction), while the upstream edge of the transcription bubble ( $-10$  element) remains constant, demonstrating TSS selection involves transcription-bubble expansion (“scrunching”) and transcription-bubble contraction (“anti-scrunching”; Vvedenskaya et al., 2015; Winkelman et al., 2016b).

Importantly, the unique features of ribosomal promoter sequence (short suboptimal 16 bp spacer, absence of extended  $-10$ , and the lack of interactions of DSR/ $\sigma 1.2$ , and CRE/ $\beta$  pocket) lead to RPo pre-scrunching and downstream shift of TSS to an unusual position 9 nt from the  $-10$  element. These features reduce abortive synthesis and facilitate promoter escape during initiation, contributing to the high transcriptional activity of rRNA promoters (Winkelman et al., 2016a). At the same time, they destabilize RPo and increase its sensitivity to initiating NTP concentrations, providing a mechanism for rapid down-regulation during starvation.

Besides the TSS region sequences, the negative DNA supercoiling that increases the size of the transcription bubble in RPo also causes downstream shift in the TSS position (Vvedenskaya et al., 2015). These results are consistent with biophysical data correlating bubble expansion with TSS selection (Robb et al., 2013).

### Initiation of RNA synthesis

Structures of RP<sub>init</sub> were obtained by stabilizing the crystal structures of RPo with RNA oligos complementary to t-strand in the bubble from positions  $-4$  to  $+1$ . However, these structures do not reflect the natural state of nascent transcript and DNA during transcription initiation. Recently, more functionally relevant structural data were obtained for RP<sub>init</sub> by soaking the crystals of RPo with two initiating substrates, ATP and a non-hydrolyzable analog of CTP, CMPcPP, occupying  $i$  and  $i+1$  sites, respectively. The structure revealed the location of initiating substrate, ATP, in the catalytic center. However, CMPcPP does not occupy the catalytically reactive site since the position of its  $\alpha$ -phosphate and the second Mg<sup>2+</sup> ion coordinated by its  $\beta$ - and  $\gamma$ -phosphates are

too far for catalysis. Also, the trigger helix is partly disordered and does not interact with phosphates of the substrate. Therefore, it is proposed that this structure captures RP<sub>init</sub> in a transient state where,  $i+1$  nucleotide is in pre-catalytic conformation (Basu et al., 2014; Zhang et al., 2014; Zuo and Steitz, 2015). *De novo* synthesis of 6-nt long transcript *in crystallo* generated a structure of RP<sub>init</sub> with RNA-DNA hybrid and the scrunched downstream dsDNA. In the structure,  $\sigma 3.2$  finger is displaced from its position near the active site by the RNA 5' end, signaling the beginning of  $\sigma$  release from RNAP (Basu et al., 2014).

During scrunching, the pulled-in portions of t- and nt-strands must bulge out of the transcription bubble. Because the X-ray structures of the scrunched ssDNA in RP<sub>init</sub> are disordered, their paths have been recently assessed by site-specific DNA-protein crosslinking (Winkelman et al., 2015), exploiting the unusual stability of RP<sub>init</sub> formed at ribosomal rrnBP1 promoter (Borukhov et al., 1993). In RP<sub>init</sub> containing 5-mer RNA, the nt-strand bulge is extruded through the space between lobe 1 and 2 of  $\beta$  clamp into solvent, whereas the t-strand bulge remains inside the RNAP main channel restricted by the  $\beta$  flap,  $\beta$  lid,  $\beta'$  clamp, and  $\sigma 3.2$  finger. Mapping results indicate that the t-strand bulge moves toward RNA exit channel, but its exact position is unclear. Extension of RNA beyond 5–6 nt will lead to further bulge expansion resulting in stress build-up, which can be relieved by displacement of  $\sigma 3.2$  finger and/or by opening of the  $\beta$  flap and  $\beta'$  clamp domains. Further stress accumulation may cause the t-strand bulge to extrude outside either through expansion of the RNA exit channel or through the space opened up between  $\beta$  lobe 1 and  $\sigma 3$ . Eventually, the growing 5' end of nascent RNA will occupy the exit channel displacing  $\sigma 3$  and 4, commencing promoter escape. Another way to relieve the stress caused by t-strand bulge expansion is to reverse the scrunching by releasing the abortive RNA products through the 2°-channel and repeat initiation (Kapanidis et al., 2006; Revyakin et al., 2006).

### Activation of Transcription Initiation

Many bacterial promoters contain suboptimal sequences that require binding of specific factors for efficient transcription initiation. Various classes of activators act by facilitating RNAP recruitment to promoters and by accelerating isomerization steps in initiation pathway: R<sub>Pc</sub>  $\rightarrow$  R<sub>Pi</sub>  $\rightarrow$  R<sub>Po</sub>  $\rightarrow$  RP<sub>init</sub>  $\rightarrow$  promoter escape (reviewed in Roy et al., 1998; Lee et al., 2012; Decker and Hinton, 2013). Below, we present two examples of transcription activation systems that have been recently characterized with structural, genetic, and biochemical studies.

### Transcription Activation by Class II Initiation Factors: CAP/TAP

Two well-characterized classes of transcriptional activators (class I and II) act through simple RNAP recruitment to promoters with missing or inefficient core promoter elements. Class I activators, exemplified by *E. coli* CAP (catabolite activator protein) binding at the  $-61.5$  DNA site upstream of the *lac* promoter, stimulate R<sub>Pc</sub> formation through direct interactions with the RNAP  $\alpha$ CTD. Class II activators, such as *E. coli* CAP binding at the  $-41.5$  site overlapping the  $-35$  element of *gal* promoter, facilitate formation

of R<sub>Pc</sub> and its isomerization into transcriptionally active R<sub>Po</sub> through multiple contacts (activation regions AR1–AR3) with RNAP  $\alpha$ NTD,  $\alpha$ CTD, and  $\sigma$ 4 domains (Lawson et al., 2004). A structural model of an *E. coli* class I transcription activation complex of CAP-R<sub>Po</sub> on a modified *lac* promoter based on a low resolution electron microscopy data has been generated (Lawson et al., 2004; Hudson et al., 2009) providing information on CAP/DNA,  $\alpha$ CTD-DNA, and  $\alpha$ CTD- $\sigma$ 4 interactions.

Recently, a 4.4 Å-resolution crystal structure of class II transcription activation complex was reported. It shows *Tth* activator protein TAP, a homolog of *E. coli* CAP, in complex with *Tth* R<sub>Po</sub> assembled on a TAP-dependent *Tth crtB* promoter, with a full transcription bubble and a 4 nt-RNA primer (Feng et al., 2016). In the structure, *Tth* TAP homodimer is bound to DNA at position –41.5 from transcription start site. As expected based on their sequence homology, the structures of *Tth* TAP-DNA and *Eco* CAP-DNA are very similar. The structure of R<sub>Po</sub> is mostly unchanged in TAP-R<sub>Po</sub>, except that DNA upstream of –35 is slightly distorted by TAP, resulting in reduced interaction between –35 region and  $\sigma$ 4. Not surprisingly, biochemical data indicate that specificity of –35 recognition by  $\sigma$ 4 does not play a major role in Class II CAP-mediated activation (Rhodius et al., 1997). Yet, intriguingly, mutations of the two  $\sigma$ 4 residues (R584, E585) contacting bases at –32 and –33 strongly inhibited R<sub>Po</sub> formation (Feng et al., 2016). Apparently, the few remaining specific interactions between  $\sigma$ 4 and –35 still play an important role in TAP mediated activation.

#### Role of $\alpha$ CTD in activation

In TAP-R<sub>Po</sub> structure, the distal subunit of TAP homodimer interacts with one  $\alpha$ CTD through an interface that includes ~8 pairs of partnering residues. This interaction (mediated by activation region AR4, a unique feature of TAP) is essential for R<sub>Pc</sub> formation, as demonstrated by the loss of promoter binding following substitutions of AR4/ $\alpha$ CTD interface residues. In addition, unlike *Eco* CAP-R<sub>Po</sub>,  $\alpha$ CTD in TAP-R<sub>Po</sub> does not interact with DNA. DNA footprinting data show that *Tth*  $\alpha$ CTD does not contribute to DNA binding irrespective of TAP or promoter sequences (Feng et al., 2016), indicating that  $\alpha$ CTD is used by TAP only as an RNAP-tether. Indeed, CAP and TAP use different regions to contact  $\alpha$ CTD (AR1 and AR4, respectively) that may play a different role in activation. Whereas *Eco* CAP-AR1/ $\alpha$ CTD interaction serves to recruit RNAP to the promoter by DNA-bound CAP, TAP-AR4/ $\alpha$ CTD interaction facilitates the association of free RNAP and TAP prior to promoter DNA binding. Because the RNAP-TAP binding constant (6  $\mu$ M) is comparable to the intracellular RNAP concentration [ $>5 \mu$ M (Patrick et al., 2015)], it is proposed that in addition to classic recruitment pathway, TAP may activate transcription via a pre-recruitment pathway, similar to the mechanism of eukaryotic transcription activation.

#### The role of TAP-AR2 and -AR3 in activation

TAP-AR2 interacts with  $\beta$  flap domain while TAP-AR3 interacts with  $\sigma$ 4 and the  $\beta$ -flap tip helix. Most of these interactions are mediated through polar contacts and salt bridges, and are conserved between *Eco* CAP and *Tth* TAP. Mutation of residues

in the AR2/ $\beta$ -flap, AR3/ $\beta$ -flap tip, and AR3/ $\sigma$ 4 interfaces lead to defects in transcription activation by TAP. Kinetic analysis indicate that similar to CAP (Niu et al., 1996; Rhodius et al., 1997; Rhodius and Busby, 2000), interactions of TAP AR2- and AR3- with RNAP accelerate the transition of R<sub>Pc</sub> to R<sub>Po</sub> but do not play a significant role in the initial DNA binding (R<sub>Pc</sub> formation).

From the observation that the R<sub>Po</sub> structure does not change upon interaction with TAP, it is inferred that the mechanism of TAP/CAP Class II activation entails sequential stabilization of intermediate complexes (between R<sub>Pc</sub> and R<sub>Po</sub>) through simple contacts between ARs and RNAP without inducing any conformational perturbations in RNAP. It should be noted, however, that the reported TAP-R<sub>Po</sub> structure presents the complex in the final activated state, whereas the path to this state is poorly understood. An alternative to the “activation by adhesion” mechanism can be envisaged which entails allosteric/conformational changes in the intermediate initiation complexes.

#### Transcription Activation by R<sub>Po</sub> Stabilization: CarD

Unlike RNAP of the model organism *Eco*, RNAPs of many bacteria form intrinsically unstable R<sub>Po</sub> even at consensus promoters, and require additional factors to stabilize R<sub>Po</sub>. One such factor is CarD, a global regulator which is an essential factor in *Mtb*. CarD is widely distributed in at least ten bacterial phyla, including Firmicutes, Cyanobacteria, Actinobacteria, and Deinococcus-Thermus (Bae et al., 2015a), but is absent in  $\gamma$ -Proteobacteria such as *Eco*. The structure and function of *Mtb* and *Tth* CarD have been characterized (Stallings et al., 2009; Gulten and Sacchettini, 2013; Srivastava et al., 2013), and the molecular mechanism of its action was recently proposed based on the 3-D structure of *Tth* CarD in complex with R<sub>Po</sub> assembled on consensus promoter DNA with full transcription bubble (Bae et al., 2015a).

CarD consists of N-terminal, RNAP-interacting domain (CarD-RID) and  $\alpha$  helical C-terminal domain (CarD-CTD). In CarD-R<sub>Po</sub> structure, CarD-RID binds RNAP  $\beta$  lobe-1 domain, orienting CarD-CTD toward the upstream ds/ss junction of the transcription bubble near the –10 element. One of the  $\alpha$ -helices of CarD-CTD inserts a conserved W86 into the DNA minor groove near positions –12/–13, and acts as a wedge to maintain the distorted conformation of the minor groove immediately upstream of the fork junction. This action is proposed to prevent the reannealing of t- and nt-strands and collapse of transcription bubble, thus stabilizing the R<sub>Po</sub>. The proposed mechanism of CarD action is strongly supported by the experimental evidence. First, CarD did not affect the conformation of R<sub>Po</sub> in the structure or alter the size of the transcription bubble during R<sub>Po</sub> formation on native promoters. Second, kinetic data showed that CarD increased the resistance of R<sub>Po</sub> to competitor challenge in *in vitro* assays on native promoters, although it had no effect on R<sub>Po</sub> assembled on artificial bubble templates (Davis et al., 2015). Finally, mutational analysis indicated that specific interaction between W86 and –12T of nt-strand plays a critical role in R<sub>Po</sub> stabilization (Bae et al., 2015a). Thus, transcription activation



by CarD entails stabilizing the RPo, and prolonging its lifetime sufficient for successful initiation of RNA synthesis.

Analysis of CarD chromosomal distribution using ChIPSeq revealed that CarD is associated with RNAP predominantly at promoter regions, co-localizing with  $\sigma^A$  (Stallings et al., 2009; Srivastava et al., 2013), suggesting that CarD dissociates during early stages of elongation. It's unclear what causes its dissociation. Since the CarD-RID is homologous to the NTD of Mfd, transcription coupled repair factor, which also binds to the  $\beta$  lobe-1, it is possible that CarD is displaced from elongation complex by Mfd. Also, because CarD stabilizes the RPo, it would be expected to negatively affect the rate of promoter escape. Additional biochemical experiments would be needed to address this hypothesis.

## TRANSCRIPTION INHIBITORS THAT TARGET RNAP

Because bacterial RNAP performs essential functions in the cell, and because it differs sufficiently from eukaryotic RNAPs, it is an attractive target for antibiotics (for comprehensive review on the subject, see Ma et al., 2016). Currently, known drugs targeting RNAP can be divided into three groups based on their modes of action (**Table 1**): (i) those that disrupt RNAP interactions with DNA, RNA or NTPs; (ii) those that interfere with the movement of RNAP mobile elements during nucleotide addition cycle (NAC); and (iii) those that disrupt RNAP interactions with the housekeeping initiation factor,  $\sigma^{70}$ . Although many of these drugs were discovered decades ago, and have been extensively characterized biochemically and genetically since then, it is only with the recent avalanche of structural data obtained for bacterial RNAP in complex many inhibitors that their mechanism of action began to be truly revealed at the molecular level (Murakami, 2015). Below, we summarize briefly the current understanding of how these drugs interact with RNAP from a structural point of view.

The first group comprises inhibitors that bind in the primary channel, the secondary channel, or to the  $\beta'$  switch-2 region of RNAP. Rifamycins (RIF) and sorangicin bind in the primary channel near the active site, directly contacting  $\beta$  subunit, and sterically block the path of growing RNA beyond 2-3 nucleotides in length, effectively locking the abortive initiation complex at the promoter (Campbell et al., 2001, 2005; Molodtsov et al., 2013). GE23077 binds to the *i* and *i*+1 sites of the active center, immediately adjacent to the catalytic  $Mg^{2+}$  ion, and sterically occludes natural substrates from binding to these sites, inhibiting RNAP from initiating transcription *de novo* (Zhang et al., 2014). Using the structural information and modeling, a bipartite drug that binds to adjoining (but not overlapping) sites near the active center of RNAP was created by covalently linking GE and rifamycin SV (RIF derivative). The resulting compound, RifaGE-3, was active against both Rif<sup>R</sup> and GE<sup>R</sup> RNAPs (Zhang et al., 2014), suggesting that bipartite drugs could represent a new class of antibiotics to combat pathogenic bacteria that are increasingly drug-resistant. However, their large size and complexity may cause reduced permeability and increased cytotoxicity.

Microcin binds in the 2° channel and competitively prevents NTP uptake or binding, thereby inhibiting abortive initiation and elongation (Adelman et al., 2004; Mukhopadhyay et al., 2004). Compounds like myxopyronins, coralopyronin (Mukhopadhyay et al., 2008), ripostatin (Mukhopadhyay et al., 2008; Belogurov et al., 2009), and squaramides (Buurman et al., 2012; Molodtsov et al., 2015) bind to RNAP  $\beta'$  switch-2 region that controls the hinged, swinging motion of  $\beta'$  clamp, which in turn is responsible for the opening and closing of the primary channel (Srivastava et al., 2011). Binding of these compounds prevents the  $\beta'$  clamp from opening, stabilizes the  $\beta'$  clamp/switch regions in a partly closed/fully closed conformation, and prevents template DNA from reaching the active site. In particular, squaramide, in its co-crystal structure with RNAP, is shown to displace  $\beta'$  switch-2 into the DNA binding main channel of RNAP (Molodtsov et al., 2015), which would interfere with proper placement of the melted template DNA (Bae et al., 2015b). Fidaxomicin and lipiarmycin (Tupin et al., 2010; Artsimovitch et al., 2012; Morichaud et al., 2016) are structurally closely related natural compounds that also bind to the  $\beta'$  switch-2 region and prevent t-strand DNA from accessing the RNAP active-site cleft. Interestingly, the sensitivity of RNAP to lipiarmycin is aggravated in the presence of specific mutations in  $\sigma^{70}$  that are known to destabilize RPo (Morichaud et al., 2016). This result supports the assertion that lipiarmycin, and likely fidoxymicin, competes with t-strand DNA for the same binding site on RNAP  $\beta'$  switch-2 region during RPo formation, and effectively inhibits promoter melting and RPo formation.

The second group comprises inhibitors that interact with, or bind near the catalytically important mobile elements of RNAP,  $\beta'$  BH,  $\beta'$  TL,  $\beta$ -link, and F-loop. These mobile elements are located in the immediate vicinity of the active site of RNAP, and are proposed to undergo conformational changes in concert with NAC (Malinen et al., 2012). Notably, the  $\beta'$  TL alternates between “open” (unfolded) and “closed” (folded) states, while the adjacent  $\beta'$  BH alternates between bent and straight conformations (Wang et al., 2006; Jovanovic et al., 2011). These structural changes are thought to accompany each NAC during transcription. In the “closed” conformation,  $\beta'$  TL forms a three-helix bundle with  $\beta'$  BH, and is directly contacted by the residues of F-loop and the  $\beta$ -link, which leads to further stabilization of the folded conformation of the  $\beta'$  TL. NTP loading and catalysis occur in this state. In the “open” conformation, the three-helix bundle collapses:  $\beta'$  BH bends toward RNA-DNA hybrid;  $\beta'$  TL becomes unfolded. This state presumably permits RNA and DNA translocation following catalysis. Streptolydigin (Temiaikov et al., 2005; Tuske et al., 2005; Vassilyev et al., 2007), salinamide (Degen et al., 2014), CBR (Artsimovitch et al., 2003, 2011; Malinen et al., 2012, 2014; Yuzenkova et al., 2013; Bae et al., 2015c; Feng et al., 2015), and, most likely targetitoxin (Artsimovitch et al., 2011; Malinen et al., 2012; Yuzenkova et al., 2013), are all inhibitors that bind RNAP and interact intimately with non-overlapping residues of  $\beta'$  BH,  $\beta'$  TL,  $\beta'$  F-loop, and/or  $\beta$ -link in the active site milieu. Extensive structural and biochemical studies support the mechanistic model that these inhibitors stabilize an intermediate complex formed during NAC by immobilizing one or more mobile elements in a fixed conformation, thereby halting



TABLE 1 | Small molecule inhibitors that target RNAP.

Classification	Name	Source	Binding site	Comment	References
Inhibitors that disrupt RNAP—DNA/RNA/NTPs interactions	Rifamycins	<i>Amycolatopsis mediterranei</i>	$\beta$ subunit active site cleft	High resistance spectrum	Campbell et al., 2001; Molodtsov et al., 2013
	Sorangicin	<i>Sorangium cellulosum</i>	i/t-1 NTP binding site	Low resistance spectrum but poor membrane permeability	Campbell et al., 2005 Zhang et al., 2014
	GE23077	<i>Actinomadura</i> sp	composite of Rif- and GE-binding sites		Zhang et al., 2014
	GE-Rif	semi-synthetic			
	Microcin	<i>E. coli</i> AY25	2° channel (based on modeling)	High resistance spectrum, active against Gram(–) only	Adelman et al., 2004; Mukhopadhyay et al., 2004
	Myxopyronins	<i>Myxococcus fulvus</i>	$\beta'$ Switch-2	broad-spectrum antibacterial (Gram+ and Gram–)	Mukhopadhyay et al., 2008; Belogurov et al., 2009; Srivastava et al., 2011
	Corallopyronin	<i>Corallococcus coralloides</i>			
	Ripostatin	<i>Sorangium cellulosum</i>			
	Squaramides	synthetic			
	Flaxomicin	<i>Dactyloporangium aurantiacum</i>			
Inhibitors that constrain the mobile elements of RNAP active center	Lipiamycin	<i>Actinoplanes deccanensis</i>	$\beta'$ Switch-2 ( $\sigma^{70}$ /S R3.2)*	active against Gram (+)	Buuman et al., 2012; Molodtsov et al., 2015 Artsimovitch et al., 2012 Tupin et al., 2010; Morichaud et al., 2016
	Streptolydigin	<i>Streptomyces lydicus</i>	$\beta'$ BH, $\beta'$ TL, $\beta$ fork loop-2 (downstream DNA)*	High resistance spectrum	Temiakov et al., 2005; Tuske et al., 2005; Vassilyev et al., 2007
	Salinamide	marine and soil <i>Streptomyces</i> spp	$\beta'$ BH hinge region (near 2°-channel), $\beta'$ F-loop, $\beta$ -link, ( $\beta'$ TL)*	Poor membrane permeability	Degen et al., 2014
	CBR compounds	chemical compound library	$\beta$ element between fork loop 2 and $\beta$ DII, $\beta'$ F-loop, $\beta'$ BH hinge region	Low resistance spectrum but tested only in <i>ToIc</i> strain; cytotoxic	Artsimovitch et al., 2003, 2011; Malinen et al., 2012, 2014; Yuzenkova et al., 2013; Bae et al., 2015c; Feng et al., 2015
	Tagetitoxin	<i>Pseudomonas syringae</i> pv. <i>tagetis</i>	2°-channel; $\beta'$ catalytic loop and adjacent residues, $\beta'$ clip; $\beta$ -link, $\beta$ active site loop, $Mg^{2+}$ (II) ( $\beta'$ TL)*	Low membrane permeability	Artsimovitch et al., 2011; Malinen et al., 2012; Yuzenkova et al., 2013
Inhibitors that disrupt RNAP- $\sigma^{70}$ interactions	GKL003	chemical compound library	$\beta'$ clamp helices (CH)	High affinity ( $K_i$ ~6 nM) but low solubility and permeability	Ma et al., 2013; Yang et al., 2015a
	DSHS00507	drug-like compound library	$\beta'$ clamp helices (CH)	Effective against Gram-positive bacteria	
	SB series	compound library	undefined	IC <sub>50</sub> : 2~15 $\mu$ M but low target specificity	André et al., 2004, 2006

\*Indicates possible auxiliary binding site.

the iterative catalytic process. Binding sites of salinamide and CBR on *Eco* RNAP identified from analysis of X-ray co-crystal structures are consistent with their genetically-mapped binding sites. Interestingly, it was reported that two CBR<sup>R</sup> mutations, P750→L in  $\beta'$  F-loop and F773→V in  $\beta'$  BH N-terminus, also conferred CBR dependence for cell growth (Bae et al., 2015c). One possible explanation for this observation is that the mutations made these mobile elements too flexible to support NAC, and that binding of CBR compensated for this defect. This interpretation would be consistent with the proposed mechanism of action of CBR.

The third group of inhibitors include compounds of GKL-, DSHS-, and SB-series, all derived from chemical compound libraries, that are predicted to directly inhibit RNAP- $\sigma^{70}$  interaction. The SB-series compounds were discovered by screening the library using ELISA-based assay, and while they inhibited RNAP- $\sigma^{70}$  association with IC<sub>50</sub> ranging from 2 to 15  $\mu$ M, many showed nonspecific binding to unrelated targets *in vivo* (André et al., 2004, 2006). GKL- and DSHS-series were screened *in silico* based on the strategy of structure-based drug design, and subsequently tested *in vitro* for validation (Ma et al., 2013; Yang et al., 2015a). As predicted by pharmacophore modeling, select compounds of GKL- and DSHS-series were shown to compete with  $\sigma^{70}$  for binding to RNAP core to form holoenzyme. More analysis and characterization is required to determine if these compounds can be further developed as potential antibacterial drugs. In theory, it should be possible to screen for compounds that inhibit  $\sigma$  dissociation from RPo using the same pharmacophore model (Ma et al., 2013), or one based on the RPo complex structure. An attractive target interface for such an inhibitor would be the RNA exit channel, where blocking of  $\sigma^{3.2}$  release would cause essentially the same inhibitory effect as RIF.

A survey of currently available RNAP-specific inhibitors reveals that for many lead compounds, the predicted *in vivo* effectiveness is often low due to their poor permeability, cytotoxicity, and broad resistance spectrum. Therefore, future drug designs will need to include strategies for incorporating effective delivery mechanisms, such as nanoparticles or functional conjugates that can be cleaved/unloaded inside the cell. Design for new drugs should also aim to improve solubility and reduce nonspecific, aggregate-forming properties of drugs associated with cytotoxicity. The construction of bipartite molecules, in principle, offers a highly promising approach

to achieve increased potency and low resistance spectrum. It remains to be seen if, with the right combination of linked inhibitors and modifications, an effective bipartite drug can be constructed that is permeable with negligible toxicity. Finally, to design inhibitors of RNAP with narrow resistance spectrum, it is instructive to note that Sorangicin, which resembles RIF and binds to the same site on RNAP, exhibits a narrower resistance spectrum than RIF. This is attributed to its presumed greater conformational flexibility, enabling it to accommodate mutations in the RIF-binding pocket (Campbell et al., 2005). This suggests that a built-in structural flexibility of the compound may be an important factor in smart drug design.

## CONCLUDING REMARKS

The last 15 years saw a remarkable progress in our understanding of the structure-function relationship of bacterial RNAP thanks to the advances in structural studies of this enzyme. It is hoped that in the near future, structural studies will continue to reveal fine details of transcription, especially of the events during formation of R<sub>Pc</sub>, scrunching of R<sub>Pinit</sub> and termination processes. Additionally, an exciting new direction in RNAP research is emerging with the advent of high throughput sequencing and screening techniques, which will help to shed new light on the function of RNAP in the context of true physiological environment.

## AUTHOR CONTRIBUTIONS

JL and SB equally contributed to the preparation and writing of the manuscript.

## FUNDING

Research in SB lab is supported by the Department of Cell Biology and the Graduate School of Biomedical Studies at Rowan University.

## ACKNOWLEDGMENTS

Authors are grateful to our colleagues R. Gourse, E. Nudler, A. Kulbachinskiy, and M. Cashel for helpful discussions and advices. We apologize to those researchers whose work has not been cited due to space limitations.

## REFERENCES

- Adelman, K., Yuzenkova, J., La Porta, A., Zenkin, N., Lee, J., Lis, J. T., et al. (2004). Molecular mechanism of transcription inhibition by peptide antibiotic Microcin J25. *Mol. Cell* 14, 753–762. doi: 10.1016/j.molcel.2004.05.017
- André, E., Bastide, L., Michaux-Charachon, S., Gouby, A., Villain-Guillot, P., Latouche, J., et al. (2006). Novel synthetic molecules targeting the bacterial RNA polymerase assembly. *J. Antimicrob. Chemother.* 57, 245–251. doi: 10.1093/jac/dki426
- André, E., Bastide, L., Villain-Guillot, P., Latouche, J., Roubly, J., and Leonetti, J. P. (2004). A multiwell assay to isolate compounds inhibiting the assembly of the prokaryotic RNA polymerase. *Assay Drug Dev. Technol.* 2, 629–635. doi: 10.1089/adt.2004.2.629
- Artsimovitch, I., Chu, C., Lynch, A. S., and Landick, R. (2003). A new class of bacterial RNA polymerase inhibitor affects nucleotide addition. *Science* 302, 650–654. doi: 10.1126/science.1087526
- Artsimovitch, I., Seddon, J., and Sears, P. (2012). Fidaxomicin is an inhibitor of the initiation of bacterial RNA synthesis. *Clin. Infect. Dis.* 55(Suppl. 2), S127–S131. doi: 10.1093/cid/cis358
- Artsimovitch, I., Svetlov, V., Nemetski, S. M., Epshtein, V., Cardozo, T., and Nudler, E. (2011). Tagetitoxin inhibits RNA polymerase through trapping of the trigger loop. *J. Biol. Chem.* 286, 40395–40400. doi: 10.1074/jbc.M111.300889

- Bae, B., Chen, J., Davis, E., Leon, K., Darst, S. A., and Campbell, E. A. (2015a). CarD uses a minor groove wedge mechanism to stabilize the RNA polymerase open promoter complex. *Elife* 4:e08505. doi: 10.7554/eLife.08505
- Bae, B., Davis, E., Brown, D., Campbell, E. A., Wigneshweraraj, S., and Darst, S. A. (2013). Phage T7 Gp2 inhibition of *Escherichia coli* RNA polymerase involves misappropriation of sigma70 domain 1.1. *Proc. Natl. Acad. Sci. U.S.A.* 110, 19772–19777. doi: 10.1073/pnas.1314576110
- Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R., and Darst, S. A. (2015b). Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *Elife* 4:e08504. doi: 10.7554/eLife.08504
- Bae, B., Nayak, D., Ray, A., Mustaev, A., Landick, R., and Darst, S. A. (2015c). CBR antimicrobials inhibit RNA polymerase via at least two bridge-helix cap-mediated effects on nucleotide addition. *Proc. Natl. Acad. Sci. U.S.A.* 112, E4178–E4187. doi: 10.1073/pnas.1502368112
- Basu, R. S., Warner, B. A., Molodtsov, V., Pupov, D., Esysunina, D., Fernández-Tornero, C., et al. (2014). Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J. Biol. Chem.* 289, 24549–24559. doi: 10.1074/jbc.M114.584037
- Belogurov, G. A., and Artsimovitch, I. (2015). Regulation of Transcript Elongation. *Annu. Rev. Microbiol.* 69, 49–69. doi: 10.1146/annurev-micro-091014-104047
- Belogurov, G. A., Vassilyeva, M. N., Sevostyanova, A., Appleman, J. R., Xiang, A. X., Lira, R., et al. (2009). Transcription inactivation through local refolding of the RNA polymerase structure. *Nature* 457, 332–335. doi: 10.1038/nature07510
- Benoff, B., Yang, H., Lawson, C. L., Parkinson, G., Liu, J., Blatter, E., et al. (2002). Structural basis of transcription activation: the CAP- $\alpha$  CTD-DNA complex. *Science* 297, 1562–1566. doi: 10.1126/science.1076376
- Borukhov, S., and Nudler, E. (2003). RNA polymerase holoenzyme: structure, function and biological implications. *Curr. Opin. Microbiol.* 6, 93–100. doi: 10.1016/S1369-5274(03)00036-5
- Borukhov, S., and Nudler, E. (2008). RNA polymerase: the vehicle of transcription. *Trends Microbiol.* 16, 126–134. doi: 10.1016/j.tim.2007.12.006
- Borukhov, S., Sagitov, V., Josaitis, C. A., Gourse, R. L., and Goldfarb, A. (1993). Two modes of transcription initiation *in vitro* at the *rrnB* P1 promoter of *Escherichia coli*. *J. Biol. Chem.* 268, 23477–23482.
- Buurman, E. T., Foulk, M. A., Gao, N., Laganas, V. A., McKinney, D. C., Moustakas, D. T., et al. (2012). Novel rapidly diversifiable antimicrobial RNA polymerase switch region inhibitors with confirmed mode of action in *Haemophilus influenzae*. *J. Bacteriol.* 194, 5504–5512. doi: 10.1128/JB.01103-12
- Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., et al. (2001). Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* 104, 901–912. doi: 10.1016/S0092-8674(01)00286-0
- Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., et al. (2002). Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol. Cell* 9, 527–539. doi: 10.1016/S1097-2765(02)00470-7
- Campbell, E. A., Pavlova, O., Zenkin, N., Leon, F., Irschik, H., Jansen, R., et al. (2005). Structural, functional, and genetic analysis of sorangicin inhibition of bacterial RNA polymerase. *EMBO J.* 24, 674–682. doi: 10.1038/sj.emboj.7600499
- Chen, Y. F., and Helmann, J. D. (1997). DNA-melting at the *Bacillus subtilis* flagellin promoter nucleates near -10 and expands unidirectionally. *J. Mol. Biol.* 267, 47–59. doi: 10.1006/jmbi.1996.0853
- Daniels, D., Zuber, P., and Losick, R. (1990). Two amino acids in an RNA polymerase sigma factor involved in the recognition of adjacent base pairs in the -10 region of a cognate promoter. *Proc. Natl. Acad. Sci. U.S.A.* 87, 8075–8079. doi: 10.1073/pnas.87.20.8075
- Davis, E., Chen, J., Leon, K., Darst, S. A., and Campbell, E. A. (2015). Mycobacterial RNA polymerase forms unstable open promoter complexes that are stabilized by CarD. *Nucleic Acids Res.* 43, 433–445. doi: 10.1093/nar/gku1231
- Decker, K. B., and Hinton, D. M. (2013). Transcription regulation at the core: similarities among bacterial, archaeal, and eukaryotic RNA polymerases. *Annu. Rev. Microbiol.* 67, 113–139. doi: 10.1146/annurev-micro-092412-155756
- Degen, D., Feng, Y., Zhang, Y., Ebright, K. Y., Ebright, Y. W., Gigliotti, M., et al. (2014). Transcription inhibition by the depsipeptide antibiotic salinamide A. *Elife* 3:e02451. doi: 10.7554/eLife.02451
- Feklistov, A. (2013). RNA polymerase: in search of promoters. *Ann. N.Y. Acad. Sci.* 1293, 25–32. doi: 10.1111/nyas.12197
- Feklistov, A., Barinova, N., Sevostyanova, A., Heyduk, E., Bass, I., Vvedenskaya, I., et al. (2006). A basal promoter element recognized by free RNA polymerase sigma subunit determines promoter recognition by RNA polymerase holoenzyme. *Mol. Cell* 23, 97–107. doi: 10.1016/j.molcel.2006.06.010
- Feklistov, A., and Darst, S. A. (2011). Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell* 147, 1257–1269. doi: 10.1016/j.cell.2011.10.041
- Feklistov, A., Sharon, B. D., Darst, S. A., and Gross, C. A. (2014). Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu. Rev. Microbiol.* 68, 357–376. doi: 10.1146/annurev-micro-092412-155737
- Feng, Y., Degen, D., Wang, X., Gigliotti, M., Liu, S., Zhang, Y., et al. (2015). Structural Basis of Transcription Inhibition by CBR Hydroxamides and CBR Pyrazoles. *Structure* 23, 1470–1481. doi: 10.1016/j.str.2015.06.009
- Feng, Y., Zhang, Y., and Ebright, R. H. (2016). Structural basis of transcription activation. *Science* 352, 1330–1333. doi: 10.1126/science.aaf4417
- Goldman, S. R., Nair, N. U., Wells, C. D., Nickels, B. E., and Hochschild, A. (2015). The primary sigma factor in *Escherichia coli* can access the transcription elongation complex from solution *in vivo*. *Elife* 4:e10514. doi: 10.7554/eLife.10514
- Gruber, T. M., and Gross, C. A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.* 57, 441–466. doi: 10.1146/annurev-micro.57.030502.090913
- Gruber, T. M., Markov, D., Sharp, M. M., Young, B. A., Lu, C. Z., Zhong, H. J., et al. (2001). Binding of the initiation factor sigma to core RNA polymerase is a multistep process. *Mol. Cell* 8, 21–31. doi: 10.1016/S1097-2765(01)00292-1
- Gulten, G., and Sacchettini, J. C. (2013). Structure of the Mtb CarD/RNAP beta-lobes complex reveals the molecular basis of interaction and presents a distinct DNA-binding domain for Mtb CarD. *Structure* 21, 1859–1869. doi: 10.1016/j.str.2013.08.014
- Haugen, S. P., Berkmen, M. B., Ross, W., Gaal, T., Ward, C., and Gourse, R. L. (2006). rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. *Cell* 125, 1069–1082. doi: 10.1016/j.cell.2006.04.034
- Haugen, S. P., Ross, W., and Gourse, R. L. (2008). Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat. Rev. Microbiol.* 6, 507–519. doi: 10.1038/nrmicro1912
- Hook-Barnard, I. G., and Hinton, D. M. (2007). Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. *Gene Regul. Syst. Bio.* 1, 275–293.
- Hudson, B. P., Quispe, J., Lara-González, S., Kim, Y., Berman, H. M., Arnold, E., et al. (2009). Three-dimensional EM structure of an intact activator-dependent transcription initiation complex. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19830–19835. doi: 10.1073/pnas.0908782106
- Imashimizu, M., Shimamoto, N., Oshima, T., and Kashlev, M. (2014). Transcription elongation: heterogeneous tracking of RNA polymerase and its biological implications. *Transcription* 5:e28285. doi: 10.4161/trns.28285
- Jovanovic, M., Burrows, P. C., Bose, D., Cámara, B., Wiesler, S., Zhang, X., et al. (2011). Activity map of the *Escherichia coli* RNA polymerase bridge helix. *J. Biol. Chem.* 286, 14469–14479. doi: 10.1074/jbc.M110.212902
- Kamarthapu, V., Epshtein, V., Benjamin, B., Proshkin, S., Mironov, A., Cashel, M., et al. (2016). ppGpp couples transcription to DNA repair in *E. coli*. *Science* 352, 993–996. doi: 10.1126/science.aad6945
- Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonja, E., Weiss, S., and Ebright, R. H. (2006). Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* 314, 1144–1147. doi: 10.1126/science.1131399
- Keilty, S., and Rosenberg, M. (1987). Constitutive function of a positively regulated promoter reveals new sequences essential for activity. *J. Biol. Chem.* 262, 6389–6395.
- Kireeva, M., Kashlev, M., and Burton, Z. F. (2010). Translocation by multi-subunit RNA polymerases. *Biochim. Biophys. Acta* 1799, 389–401. doi: 10.1016/j.bbagr.2010.01.007
- Kriner, M. A., Sevostyanova, A., and Groisman, E. A. (2016). Learning from the Leaders: gene regulation by the transcription termination factor Rho. *Trends Biochem. Sci.* 41, 690–699. doi: 10.1016/j.tibs.2016.05.012
- Kulbachinskiy, A., and Mustaev, A. (2006). Region 3.2 of the sigma subunit contributes to the binding of the 3'-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *J. Biol. Chem.* 281, 18273–18276. doi: 10.1074/jbc.C600060200
- Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. *Biochem. Soc. Trans.* 34(Pt 6), 1062–1066. doi: 10.1042/BST0341062

- Lawson, C. L., Swigon, D., Murakami, K. S., Darst, S. A., Berman, H. M., and Ebright, R. H. (2004). Catabolite activator protein: DNA binding and transcription activation. *Curr. Opin. Struct. Biol.* 14, 10–20. doi: 10.1016/j.sbi.2004.01.012
- Lee, D. J., Minchin, S. D., and Busby, S. J. (2012). Activating transcription in bacteria. *Annu. Rev. Microbiol.* 66, 125–152. doi: 10.1146/annurev-micro-092611-150012
- Lim, H. M., Lee, H. J., Roy, S., and Adhya, S. (2001). A “master” in base unpairing during isomerization of a promoter upon RNA polymerase binding. *Proc. Natl. Acad. Sci. U.S.A.* 98, 14849–14852. doi: 10.1073/pnas.261517398
- Liu, B., Zuo, Y., and Steitz, T. A. (2015). Structural basis for transcription reactivation by RapA. *Proc. Natl. Acad. Sci. U.S.A.* 112, 2006–2010. doi: 10.1073/pnas.1417152112
- Liu, B., Zuo, Y., and Steitz, T. A. (2016). Structures of *E. coli* sigmaS-transcription initiation complexes provide new insights into polymerase mechanism. *Proc. Natl. Acad. Sci. U.S.A.* 113, 4051–4056. doi: 10.1073/pnas.1520555113
- Liu, M., Tolstorukov, M., Zhurkin, V., Garg, S., and Adhya, S. (2004). A mutant spacer sequence between –35 and –10 elements makes the Plac promoter hyperactive and cAMP receptor protein-independent. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6911–6916. doi: 10.1073/pnas.0401929101
- Ma, C., Yang, X., Kandemir, H., Mielczarek, M., Johnston, E. B., Griffith, R., et al. (2013). Inhibitors of bacterial transcription initiation complex formation. *ACS Chem. Biol.* 8, 1972–1980. doi: 10.1021/cb400231p
- Ma, C., Yang, X., and Lewis, P. J. (2016). Bacterial transcription as a target for antibacterial drug development. *Microbiol. Mol. Biol. Rev.* 80, 139–160. doi: 10.1128/MMBR.00055-15
- Maeda, H., Fujita, N., and Ishihama, A. (2000). Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res.* 28, 3497–3503. doi: 10.1093/nar/28.18.3497
- Malinen, A. M., Nandymazumdar, M., Turtola, M., Malmi, H., Grocholski, T., Artsimovitch, I., et al. (2014). CBR antimicrobials alter coupling between the bridge helix and the beta subunit in RNA polymerase. *Nat. Commun.* 5, 3408. doi:10.1038/ncomms4408
- Malinen, A. M., Turtola, M., Parthiban, M., Vainonen, L., Johnson, M. S., and Belogurov, G. A. (2012). Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic Acids Res.* 40, 7442–7451. doi: 10.1093/nar/gks383
- Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A. N., et al. (2002). Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* 108, 599–614. doi: 10.1016/S0092-8674(02)00667-0
- Molodtsov, V., Fleming, P. R., Eyermann, C. J., Ferguson, A. D., Foulk, M. A., McKinney, D. C., et al. (2015). X-ray crystal structures of *Escherichia coli* RNA polymerase with switch region binding inhibitors enable rational design of squaramides with an improved fraction unbound to human plasma protein. *J. Med. Chem.* 58, 3156–3171. doi: 10.1021/acs.jmedchem.5b00050
- Molodtsov, V., Nawarathne, I. N., Scharf, N. T., Kirchhoff, P. D., Showalter, H. D., Garcia, G. A., et al. (2013). X-ray crystal structures of the *Escherichia coli* RNA polymerase in complex with benzoxazinorifamycins. *J. Med. Chem.* 56, 4758–4763. doi: 10.1021/jm4004889
- Morichaud, Z., Chaloin, L., and Brodolin, K. (2016). Regions 1.2 and 3.2 of the RNA Polymerase sigma subunit promote DNA melting and attenuate action of the antibiotic lipiarmycin. *J. Mol. Biol.* 428, 463–476. doi: 10.1016/j.jmb.2015.12.017
- Mukhopadhyay, J., Das, K., Ismail, S., Koppstein, D., Jang, M., Hudson, B., et al. (2008). The RNA polymerase “switch region” is a target for inhibitors. *Cell* 135, 295–307. doi: 10.1016/j.cell.2008.09.033
- Mukhopadhyay, J., Sineva, E., Knight, J., Levy, R. M., and Ebright, R. H. (2004). Antibacterial peptide microcin J25 inhibits transcription by binding within and obstructing the RNA polymerase secondary channel. *Mol. Cell* 14, 739–751. doi: 10.1016/j.molcel.2004.06.010
- Murakami, K. S. (2013). X-ray crystal structure of *Escherichia coli* RNA polymerase sigma70 holoenzyme. *J. Biol. Chem.* 288, 9126–9134. doi: 10.1074/jbc.M112.430900
- Murakami, K. S. (2015). Structural biology of bacterial RNA polymerase. *Biomolecules* 5, 848–864. doi: 10.3390/biom5020848
- Murakami, K. S., and Darst, S. A. (2003). Bacterial RNA polymerases: the whole story. *Curr. Opin. Struct. Biol.* 13, 31–39. doi: 10.1016/S0959-440X(02)00005-2
- Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., and Darst, S. A. (2002a). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* 296, 1285–1290. doi: 10.1126/science.1069595
- Murakami, K. S., Masuda, S., and Darst, S. A. (2002b). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* 296, 1280–1284. doi: 10.1126/science.1069594
- Niu, W., Kim, Y., Tau, G., Heyduk, T., and Ebright, R. H. (1996). Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. *Cell* 87, 1123–1134. doi: 10.1016/S0092-8674(00)81806-1
- Nudler, E. (2009). RNA polymerase active center: the molecular engine of transcription. *Annu. Rev. Biochem.* 78, 335–361. doi: 10.1146/annurev.biochem.76.052705.164655
- Nudler, E. (2012). RNA polymerase backtracking in gene regulation and genome instability. *Cell* 149, 1438–1445. doi: 10.1016/j.cell.2012.06.003
- Patrick, M., Dennis, P. P., Ehrenberg, M., and Bremer, H. (2015). Free RNA polymerase in *Escherichia coli*. *Biochimie* 119, 80–91. doi: 10.1016/j.biochi.2015.10.015
- Perdue, S. A., and Roberts, J. W. (2011). Sigma(70)-dependent transcription pausing in *Escherichia coli*. *J. Mol. Biol.* 412, 782–792. doi: 10.1016/j.jmb.2011.02.011
- Pupov, D., Kuzin, I., Bass, I., and Kulbachinskiy, A. (2014). Distinct functions of the RNA polymerase sigma subunit region 3.2 in RNA priming and promoter escape. *Nucleic Acids Res.* 42, 4494–4504. doi: 10.1093/nar/gkt1384
- Revyakin, A., Liu, C., Ebright, R. H., and Strick, T. R. (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* 314, 1139–1143. doi: 10.1126/science.1131398
- Rhodius, V. A., and Busby, S. J. (2000). Transcription activation by the *Escherichia coli* cyclic AMP receptor protein: determinants within activating region 3. *J. Mol. Biol.* 299, 295–310. doi: 10.1006/jmbi.2000.3736
- Rhodius, V. A., West, D. M., Webster, C. L., Busby, S. J., and Savery, N. J. (1997). Transcription activation at class II CRP-dependent promoters: the role of different activating regions. *Nucleic Acids Res.* 25, 326–332. doi: 10.1093/nar/25.2.326
- Robb, N. C., Cordes, T., Hwang, L. C., Gryte, K., Duchi, D., Craggs, T. D., et al. (2013). The transcription bubble of the RNA polymerase-promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: implications for transcription start-site selection. *J. Mol. Biol.* 425, 875–885. doi: 10.1016/j.jmb.2012.12.015
- Roberts, J., and Park, J. S. (2004). Mfd, the bacterial transcription repair coupling factor: translocation, repair and termination. *Curr. Opin. Microbiol.* 7, 120–125. doi: 10.1016/j.mib.2004.02.014
- Ross, W., Ernst, A., and Gourse, R. L. (2001). Fine structure of *E. coli* RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. *Genes Dev.* 15, 491–506. doi: 10.1101/gad.870001
- Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., et al. (1993). A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* 262, 1407–1413. doi: 10.1126/science.8248780
- Ross, W., and Gourse, R. L. (2009). Analysis of RNA polymerase-promoter complex formation. *Methods* 47, 13–24. doi: 10.1016/j.ymeth.2008.10.018
- Roy, S., Garg, S., and Adhya, S. (1998). Activation and repression of transcription by differential contact: two sides of a coin. *J. Biol. Chem.* 273, 14059–14062. doi: 10.1074/jbc.273.23.14059
- Saecker, R. M., Record, M. T. Jr., and Dehaseth, P. L. (2011). Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J. Mol. Biol.* 412, 754–771. doi: 10.1016/j.jmb.2011.01.018
- Samanta, S., and Martin, C. T. (2013). Insights into the mechanism of initial transcription in *Escherichia coli* RNA polymerase. *J. Biol. Chem.* 288, 31993–32003. doi: 10.1074/jbc.M113.497669
- Sarkar, P., Sardesai, A. A., Murakami, K. S., and Chatterji, D. (2013). Inactivation of the bacterial RNA polymerase due to acquisition of secondary structure by the omega subunit. *J. Biol. Chem.* 288, 25076–25087. doi: 10.1074/jbc.M113.468520
- Sen, R., Nagai, H., Hernandez, V. J., and Shimamoto, N. (1998). Reduction in abortive transcription from the lambdaPR promoter by mutations in region 3



- of the sigma70 subunit of *Escherichia coli* RNA polymerase. *J. Biol. Chem.* 273, 9872–9877. doi: 10.1074/jbc.273.16.9872
- Sengupta, S., Prajapati, R. K., and Mukhopadhyay, J. (2015). Promoter Escape with bacterial two-component sigma factor suggests retention of sigma region 2 in the elongation complex. *J. Biol. Chem.* 290, 28575–28583. doi: 10.1074/jbc.M115.666008
- Sharp, M. M., Chan, C. L., Lu, C. Z., Marr, M. T., Nechaev, S., Merritt, E. W., et al. (1999). The interface of sigma with core RNA polymerase is extensive, conserved, and functionally specialized. *Genes Dev.* 13, 3015–3026. doi: 10.1101/gad.13.22.3015
- Shorenstein, R. G., and Losick, R. (1973). Purification and properties of the sigma subunit of ribonucleic acid polymerase from vegetative *Bacillus subtilis*. *J. Biol. Chem.* 248, 6163–6169.
- Shultzaberger, R. K., Chen, Z., Lewis, K. A., and Schneider, T. D. (2007). Anatomy of *Escherichia coli* sigma70 promoters. *Nucleic Acids Res.* 35, 771–788. doi: 10.1093/nar/gkl956
- Srivastava, A., Talaue, M., Liu, S., Degen, D., Ebricht, R. Y., Sineva, E., et al. (2011). New target for inhibition of bacterial RNA polymerase: 'switch region.' *Curr. Opin. Microbiol.* 14, 532–543. doi: 10.1016/j.mib.2011.07.030
- Srivastava, D. B., Leon, K., Osmundson, J., Garner, A. L., Weiss, L. A., Westblade, L. F., et al. (2013). Structure and function of CarD, an essential mycobacterial transcription factor. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12619–12624. doi: 10.1073/pnas.1308270110
- Stallings, C. L., Stephanou, N. C., Chu, L., Hochschild, A., Nickels, B. E., and Glickman, M. S. (2009). CarD is an essential regulator of rRNA transcription required for *Mycobacterium tuberculosis* persistence. *Cell* 138, 146–159. doi: 10.1016/j.cell.2009.04.041
- Svetlov, V., and Nudler, E. (2009). Macromolecular micromovements: how RNA polymerase translocates. *Curr. Opin. Struct. Biol.* 19, 701–707. doi: 10.1016/j.sbi.2009.10.002
- Temiaikov, D., Zenkin, N., Vassilyeva, M. N., Perederina, A., Tahirov, T. H., Kashkina, E., et al. (2005). Structural basis of transcription inhibition by antibiotic streptolydigin. *Mol. Cell* 19, 655–666. doi: 10.1016/j.molcel.2005.07.020
- Tupin, A., Gualtieri, M., Leonetti, J. P., and Brodolin, K. (2010). The transcription inhibitor lipiarmycin blocks DNA fitting into the RNA polymerase catalytic site. *EMBO J.* 29, 2527–2537. doi: 10.1038/emboj.2010.135
- Tuske, S., Sarafianos, S. G., Wang, X., Hudson, B., Sineva, E., Mukhopadhyay, J., et al. (2005). Inhibition of bacterial RNA polymerase by streptolydigin: stabilization of a straight-bridge-helix active-center conformation. *Cell* 122, 541–552. doi: 10.1016/j.cell.2005.07.017
- Vassilyev, D. G., Sekine, S., Laptchenko, O., Lee, J., Vassilyeva, M. N., Borukhov, S., et al. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* 417, 712–719. doi: 10.1038/nature752
- Vassilyev, D. G., Vassilyeva, M. N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007). Structural basis for substrate loading in bacterial RNA polymerase. *Nature* 448, 163–168. doi: 10.1038/nature05931
- Vvedenskaya, I. O., Vahedian-Movahed, H., Zhang, Y., Taylor, D. M., Ebricht, R. H., and Nickels, B. E. (2016). Interactions between RNA polymerase and the core recognition element are a determinant of transcription start site selection. *Proc. Natl. Acad. Sci. U.S.A.* 113, E2899–E2905. doi: 10.1073/pnas.1603271113
- Vvedenskaya, I. O., Zhang, Y., Goldman, S. R., Valenti, A., Visone, V., Taylor, D. M., et al. (2015). Massively systematic transcript end readout, "MASTER": transcription start site selection, transcriptional slippage, and transcript yields. *Mol. Cell* 60, 953–965. doi: 10.1016/j.molcel.2015.10.029
- Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D., and Kornberg, R. D. (2006). Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* 127, 941–954. doi: 10.1016/j.cell.2006.11.023
- Winkelman, J. T., Chandransu, P., Ross, W., and Gourse, R. L. (2016a). Open complex scrunching before nucleotide addition accounts for the unusual transcription start site of *E. coli* ribosomal RNA promoters. *Proc. Natl. Acad. Sci. U.S.A.* 113, E1787–E1795. doi: 10.1073/pnas.1522159113
- Winkelman, J. T., Vvedenskaya, I. O., Zhang, Y., Zhang, Y., Bird, J. G., Taylor, D. M., et al. (2016b). Multiplexed protein-DNA cross-linking: Scrunching in transcription start site selection. *Science* 351, 1090–1093. doi: 10.1126/science.aad6881
- Winkelman, J. T., Winkelman, B. T., Boyce, J., Maloney, M. F., Chen, A. Y., Ross, W., et al. (2015). Crosslink mapping at amino acid-base resolution reveals the path of scrunched DNA in initial transcribing complexes. *Mol. Cell* 59, 768–780. doi: 10.1016/j.molcel.2015.06.037
- Yang, X., Ma, C., and Lewis, P. J. (2015a). Identification of inhibitors of bacterial RNA polymerase. *Methods* 86, 45–50. doi: 10.1016/j.ymeth.2015.05.005
- Yang, Y., Darbari, V. C., Zhang, N., Lu, D., Glyde, R., Wang, Y. P., et al. (2015b). TRANSCRIPTION. Structures of the RNA polymerase-sigma54 reveal new and conserved regulatory strategies. *Science* 349, 882–885. doi: 10.1126/science.aab1478
- Yarnell, W. S., and Roberts, J. W. (1999). Mechanism of intrinsic transcription termination and antitermination. *Science* 284, 611–615. doi: 10.1126/science.284.5414.611
- Yuzenkova, Y., Roghanian, M., Bochkareva, A., and Zenkin, N. (2013). Tagetitoxin inhibits transcription by stabilizing pre-translocated state of the elongation complex. *Nucleic Acids Res.* 41, 9257–9265. doi: 10.1093/nar/gkt708
- Yuzenkova, Y., Tadigotla, V. R., Severinov, K., and Zenkin, N. (2011). A new basal promoter element recognized by RNA polymerase core enzyme. *EMBO J.* 30, 3766–3775. doi: 10.1038/emboj.2011.252
- Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K., and Darst, S. A. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* 98, 811–824. doi: 10.1016/S0092-8674(00)81515-9
- Zhang, Y., Degen, D., Ho, M. X., Sineva, E., Ebricht, K. Y., Ebricht, Y. W., et al. (2014). GE23077 binds to the RNA polymerase 'i' and 'i+1' sites and prevents the binding of initiating nucleotides. *Elife* 3:e02450. doi: 10.7554/eLife.02450
- Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M. X., Arnold, E., et al. (2012). Structural basis of transcription initiation. *Science* 338, 1076–1080. doi: 10.1126/science.1227786
- Zuo, Y., and Steitz, T. A. (2015). Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. *Mol. Cell* 58, 534–540. doi: 10.1016/j.molcel.2015.03.010
- Zuo, Y., Wang, Y., and Steitz, T. A. (2013). The mechanism of *E. coli* RNA polymerase regulation by ppGpp is suggested by the structure of their complex. *Mol. Cell* 50, 430–436. doi: 10.1016/j.molcel.2013.03.020

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Lee and Borukhov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Genome-Wide Transcriptional Regulation and Chromosome Structural Arrangement by GalR in *E. coli*

Zhong Qian<sup>1</sup>, Andrei Trostel<sup>1</sup>, Dale E. A. Lewis<sup>1</sup>, Sang Jun Lee<sup>2</sup>, Ximiao He<sup>3</sup>, Anne M. Stringer<sup>4</sup>, Joseph T. Wade<sup>4,5</sup>, Thomas D. Schneider<sup>6</sup>, Tim Durfee<sup>7</sup> and Sankar Adhya<sup>1\*</sup>

<sup>1</sup> Laboratory of Molecular Biology, National Institutes of Health, National Cancer Institute, Bethesda, MD, USA, <sup>2</sup> Microbiomics and Immunity Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, Korea, <sup>3</sup> Laboratory of Metabolism, National Institutes of Health, National Cancer Institute, Bethesda, MD, USA, <sup>4</sup> Wadsworth Center, New York State Department of Health, Albany, NY, USA, <sup>5</sup> Department of Biomedical Sciences, School of Public Health, University of Albany, Albany, NY, USA, <sup>6</sup> Gene Regulation and Chromosome Biology Laboratory, National Institutes of Health, National Cancer Institute, Center for Cancer Research, Frederick, MD, USA, <sup>7</sup> DNASTAR, Inc., Madison, WI, USA

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas Medical Branch,  
USA

### Reviewed by:

Maria A. Miteva,  
Paris Diderot University, France  
Lucia B. Rothman-Denes,  
University of Chicago, USA  
Agnieszka Bera,  
University of Tübingen, Germany

### \*Correspondence:

Sankar Adhya  
sadhya@helix.nih.gov

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 12 August 2016

**Accepted:** 26 October 2016

**Published:** 16 November 2016

### Citation:

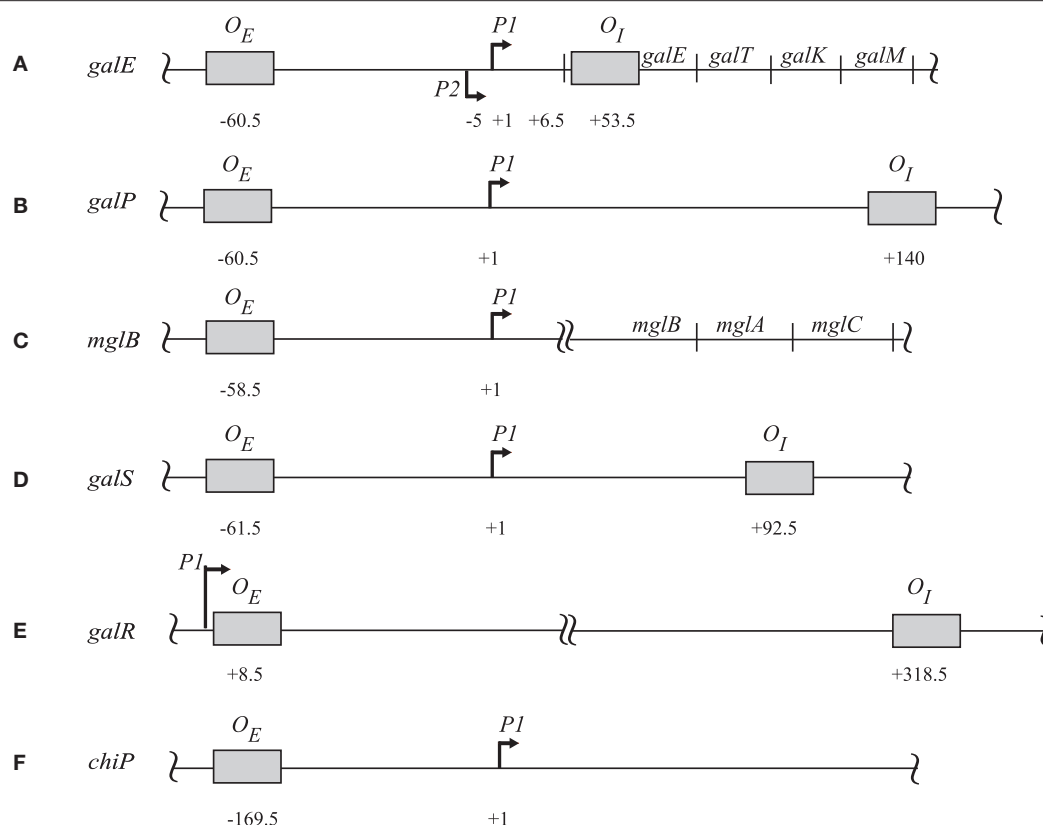
Qian Z, Trostel A, Lewis DEA, Lee SJ, He X, Stringer AM, Wade JT, Schneider TD, Durfee T and Adhya S (2016) Genome-Wide Transcriptional Regulation and Chromosome Structural Arrangement by GalR in *E. coli*. *Front. Mol. Biosci.* 3:74. doi: 10.3389/fmolb.2016.00074

The regulatory protein, GalR, is known for controlling transcription of genes related to D-galactose metabolism in *Escherichia coli*. Here, using a combination of experimental and bioinformatic approaches, we identify novel GalR binding sites upstream of several genes whose function is not directly related to D-galactose metabolism. Moreover, we do not observe regulation of these genes by GalR under standard growth conditions. Thus, our data indicate a broader regulatory role for GalR, and suggest that regulation by GalR is modulated by other factors. Surprisingly, we detect regulation of 158 transcripts by GalR, with few regulated genes being associated with a nearby GalR binding site. Based on our earlier observation of long-range interactions between distally bound GalR dimers, we propose that GalR indirectly regulates the transcription of many genes by inducing large-scale restructuring of the chromosome.

**Keywords:** GalR regulon, mega-loop, ChIP-chip, nucleoid, DNA superhelicity

## INTRODUCTION

The 4.6 Mb *Escherichia coli* chromosomal DNA is packaged into a small volume (0.2–0.5  $\mu\text{m}^3$ ) for residing inside a cell volume of 0.5–5  $\mu\text{m}^3$  (Loferer-Krossbacher et al., 1998; Skoko et al., 2006; Luijsterburg et al., 2008). It has been suggested that a bacterial chromosome has a 3-D structure that dictates the entire chromosome's gene expression pattern (Kar et al., 2005; Macvanin and Adhya, 2012). The chromosome structure and the associated volume are defined and environment-dependent. The compaction of the DNA into a structured chromosome (nucleoid) is facilitated by several architectural proteins, often called “nucleoid-associated proteins” (NAPs). NAPs are well-characterized bacterial histone-like proteins such as HU, H-NS, Fis, and Dps (Ishihama, 2009). For example, deletion of the gene encoding the NAP HU leads to substantial changes in cell volume and in the global transcription profile, presumably due to changes in chromosome architecture (Kar et al., 2005; Oberto et al., 2009; Priyadarshini et al., 2013). A recent and surprising addition to the list of NAPs in *E. coli* is the sequence-specific DNA-binding transcription regulatory protein,



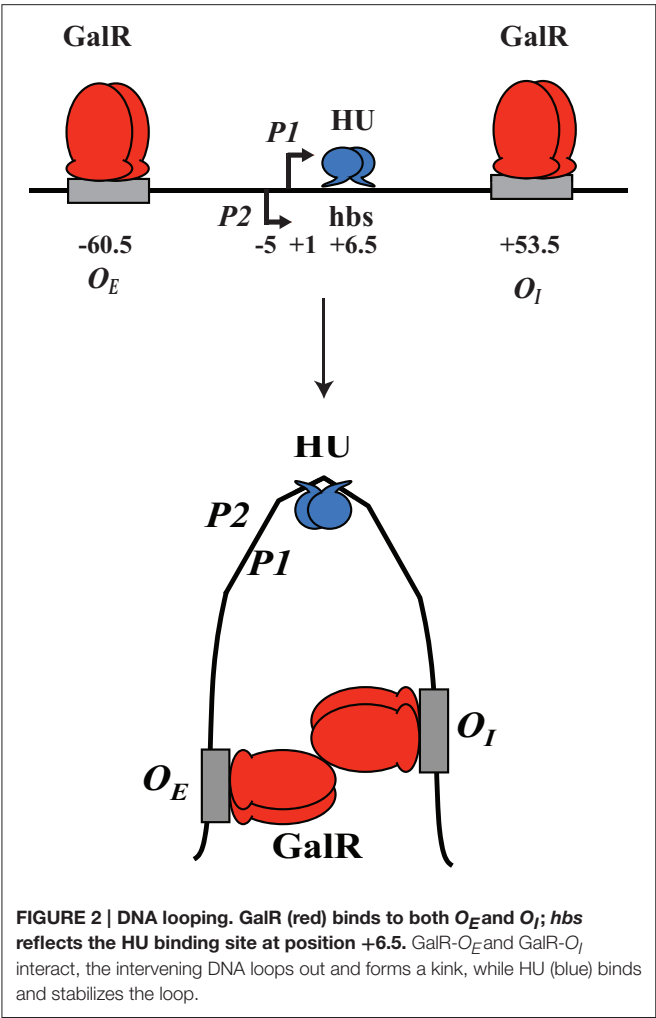
**FIGURE 1 | The regulatory sites of (A) *galE*, (B) *galP*, (C) *mglB*, (D) *galS*, (E) *galR*, and (F) *chiP* operons.** The operator sites in each operon are in shaded boxes with their locations relative to the corresponding *tsp*. For *galE* the *tsp*s are indicated as (+1) for *P1* and (−5) for *P2*. The amino terminus of the first protein in each operon is indicated by a double-arrow. The diagram is not drawn to scale (Hogg et al., 1991; Weickert and Adhya, 1993a,b; Plumbridge et al., 2014).

GalR (Qian et al., 2012). In contrast, related DNA-binding proteins PurR, MalT, FruR, and TyrR do not appear to affect the chromosome structure (Qian et al., 2012). Here, we discuss experimental results that led us to explore the idea that GalR also regulates transcription at a global scale through DNA architectural changes.

GalR regulates transcription of the *galETKM*, *galP*, *galR*, *galS*, and *mglBAC* transcripts (Figure 1). These genes all encode proteins involved in the transport and metabolism of D-galactose. Moreover, GalR controls expression of the *chiPQ* operon, which encodes genes involved in the transport of chitosugar. The *galETKM* operon (Figure 1) is transcribed as a polycistronic mRNA from two overlapping promoters, *P1* (+1) and *P2* (−5) (Musso et al., 1977; Aiba et al., 1981). GalR regulates *P1* and *P2* promoters differentially. GalR binds two operators, *O<sub>E</sub>*, located at position −60.5, and *O<sub>I</sub>*, located at +53.5 (Irani et al., 1983; Majumdar and Adhya, 1984, 1987). Binding of GalR to *O<sub>E</sub>* represses *P1* and activates *P2* by arresting RNA polymerase, and facilitating the step of RNA polymerase isomerization, respectively (Roy et al., 2004). When GalR binds to both *O<sub>E</sub>* and *O<sub>I</sub>*, which are 113 bp apart and do not overlap with the two promoters, it prevents transcription initiation from both *P1* and *P2* (Aki et al., 1996; Aki and Adhya, 1997; Semsey et al., 2002; Roy et al., 2005). Mechanistically, two

DNA-bound GalR dimers transiently associate, creating a loop in the intervening promoter DNA segment. Kinking at the apex of the loop facilitates binding of HU, which in turn stabilizes the loop (Figure 2; Kar and Adhya, 2001). The DNA structure in the looped form is topologically closed and binds RNA polymerase, but does not allow isomerization into an actively transcribing complex (Choy et al., 1995).

Following the example of GalR-mediated DNA loop formation by interaction of GalR bound to two operators in the *galE* operon, and considering the fact that GalR operators in the *galP*, *mglB*, *galS*, *galR*, and *chiP* promoters are scattered around the chromosome, we hypothesized that GalR may oligomerize while bound to distal sites, thereby forming much larger DNA loops (“mega-loops”). We employed the Chromosome Conformation Capture (3C) method to investigate interactions between distal GalR operators (Dekker et al., 2002). Thus, we showed that GalR does indeed oligomerize over long distances, resulting in the formation of mega-loops. Moreover, our data suggested the existence of other unidentified GalR binding sites around the chromosome, with these novel sites also participating in long-distance interactions (Qian et al., 2012). Figure 3 shows in a cartoon from the demonstrable GalR-mediated DNA-DNA connections as listed in Table 1. Although, we originally proposed that DNA-bound GalR-mediated mega-loops may



serve to increase the local concentrations of GalR around their binding sites for regulation of the adjacent promoters (Oehler and Muller-Hill, 2010), global regulation of gene expression due to change in chromosome structure may be another consequence of mega-loop formation. We propose that GalR-mediated mega-loop formation results in the formation of topologically independent DNA domains, with the level of superhelicity in each domain influencing transcription of the local promoters.

MATERIALS AND METHODS

Bacterial and Bacteriophage Strains

Bacteriophage P1 lysates of *galR::kanR* (from Keio collection; (Baba et al., 2006)) were made and *E. coli* K-12 MG1655 *galR* deletion strains were constructed from MG655 by bacteriophage P1 transduction using the lysate. Cells were then grown in 125 ml corning flasks (Corning® 430421) containing 30 ml of M63 minimal medium plus D-fructose (final concentration 0.3%) at 37°C with 230 rpm shaking. At OD600 0.6, cell cultures were separated into two flasks. Subsequently, D-galactose (final

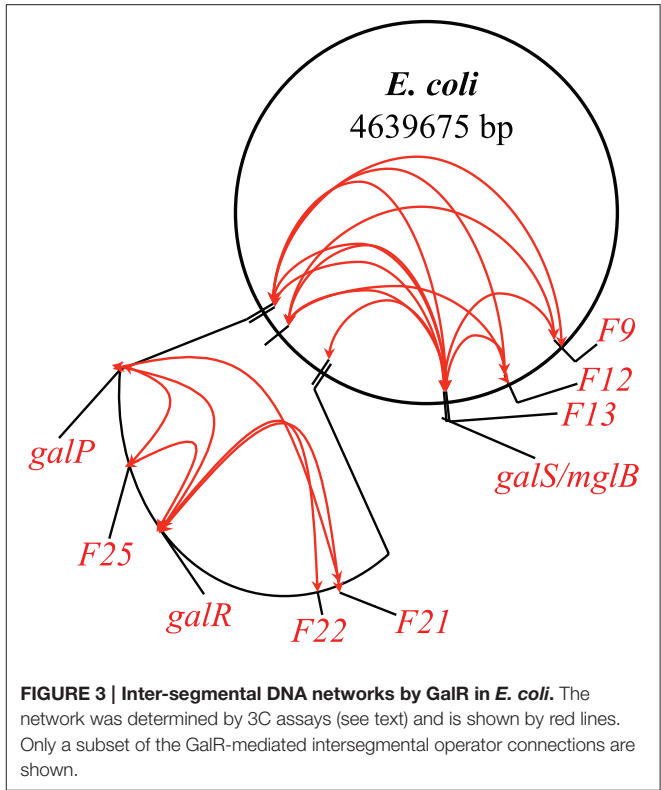


TABLE 1 | List of GalR operators identified by 3C method.

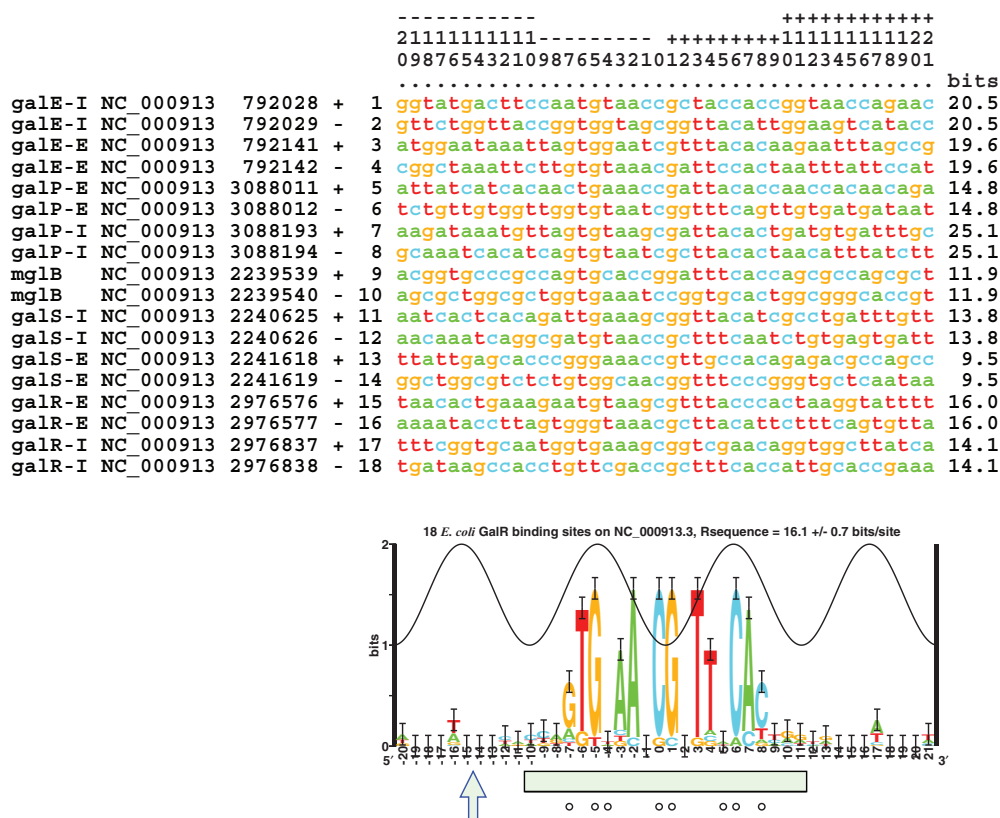
Chromosome	Coordinates	Operators	DNA sequence(s)
3088004	3088019	$O_E$ ( <i>galP</i> )	CTGAAACCGATTACAC
3088186	3088201	$O_I$ ( <i>galP</i> )	GTGTAATCGCTTACAC
2976569	2976584	$O_E$ ( <i>galR</i> )	ATGTAAGCGTTTACCC
2976830	2976845	$O_I$ ( <i>galR</i> )	GTTCGACCGCTTTTCAC
2240618	2240633	$O_E$ ( <i>galS</i> )	TTGAAAGCGGTTACAT
2241611	2241626	$O_I$ ( <i>galS</i> )	GGGAAACCGTTGCCAC
2239532	2239547	$O$ ( <i>mglB</i> )	GTGCACCGGATTTCAC
1737872	1737887	$O$ (F9)	GTGGAAACGTTTGCTC
1990112	1990127	$O$ (F12)	ATTTAACCGTTTTCTG
2246944	2246959	$O$ (F13)	TTGTTATCGTTTGCAT
2738456	2738471	$O$ (F21)	ATGGAAGGTTGCAC
2783816	2783831	$O$ (F22-1)	GCGAAACCGTTTAAG
2784177	2784192	$O$ (F22-2)	CTGCAAGCTTTTCCA
2786317	2786332	$O$ (F22-3)	TTGCAATTACTTTCAC
3072949	3072964	$O$ (F25-1)	CTTAAATCGATTGCCG
3072989	3073004	$O$ (F25-2)	TTTGAAGCGATTGCGG
3073430	3073445	$O$ (F25-3)	CTGCAATCGCTCCCCT

Connections were detected among these sites except *galE<sub>E</sub>* and *galE<sub>I</sub>* by 3C assays. The first seven operators that showed connections by 3C were known before. The ones named as F were discovered during the 3C studies (Qian et al., 2012).

concentration 0.3%) or water was added and cells were cultivated for an additional 1.5 h at 37°C.

*E. coli* MG1655 *galR*-TAP (AMD032) was constructed by bacteriophage P1 transduction of the *kanR*-linked TAP tag cassette from DY330 *galR*-TAP (Butland et al., 2005). The *kanR*





**FIGURE 4 | Sequence alignment of 9 studied GalR binding sites (NC\_000913.3) and corresponding sequence logo (Schneider and Stephens, 1990).** The numbers on the top are to be read vertically. The peak of the sine wave indicates the location where the major groove of the DNA faces GalR (Papp et al., 1993). The information content of each site is indicated in the right hand column, they average to the area under the sequence logo, 16.1 ± 0.7 bits/site (Schneider et al., 1986; Schneider, 1997). A DNase I-hypersensitive site is indicated by an arrow and DNase I protection by a green box (Majumdar and Adhya, 1987). Bases protected from DMS by GalR are shown by circles (Majumdar and Adhya, 1989). The aligned listing and logo were created using Delila programs (Schneider et al., 1982, 1984). In this logo, base pair 9 in the consensus sequence mentioned elsewhere is marked as zero.

cassette was removed using pCP20, as described previously (Datsenko and Wanner, 2000). *E. coli* MG1655 *galR*-FLAG<sub>3</sub> (AMD188) was constructed using FRUIT (Stringer et al., 2012).

## RNA Isolation

Cell cultures were placed on ice and RNaprotect™ Bacteria Reagent (Qiagen® 76506) was added to stabilize the RNA (Lee et al., 2014). Cells were harvested for RNA purification by RNeasy® Mini Kit (Qiagen® 74104) following the manufacturer's recommendations. RNA concentrations and purity were measured using a Thermo Scientific NanoDrop™ 1000. Further sample processing was performed according to the Affymetrix GeneChip® Expression Analysis Technical Manual, Section 3: Prokaryotic Sample and Array Processing (701029 Rev.4).

## cDNA Synthesis

Isolated RNA (10 µg) was used for Random Primer cDNA synthesis using SuperScript II™ Reverse Transcriptase (Invitrogen Life Technologies 18064-071). The reaction mixture was treated with 1N NaOH to degrade any remaining RNA

and treated with 1N HCl to neutralize the NaOH. Synthesized cDNA was then purified using MinElute® PCR Purification columns (Qiagen® 28004). Purified cDNA concentration and purity were measured using a Thermo Scientific NanoDrop™ 1000.

## cDNA Fragmentation

Purified cDNA was fragmented to between 50 and 200 bp by 0.6 U/µg of DNase I (Amersham Biosciences 27-0514-01) for 10 min at 37°C in 1X One-Phor-All buffer (Amersham Biosciences 27-0901-02). Heat inactivation of the DNase I enzyme was performed at 98°C for 10 min.

## cDNA Labeling

Fragmented cDNA was then 3' termini biotin labeled using the GeneChip® DNA Labeling Reagent (Affymetrix 900542) and 60 U of Terminal Deoxynucleotidyl Transferase (Promega M1875) at 37°C for 60 min. The labeling reaction was then stopped by the addition of 0.5 M EDTA.

TABLE 2 | GalR binding sites predicted by information theory.

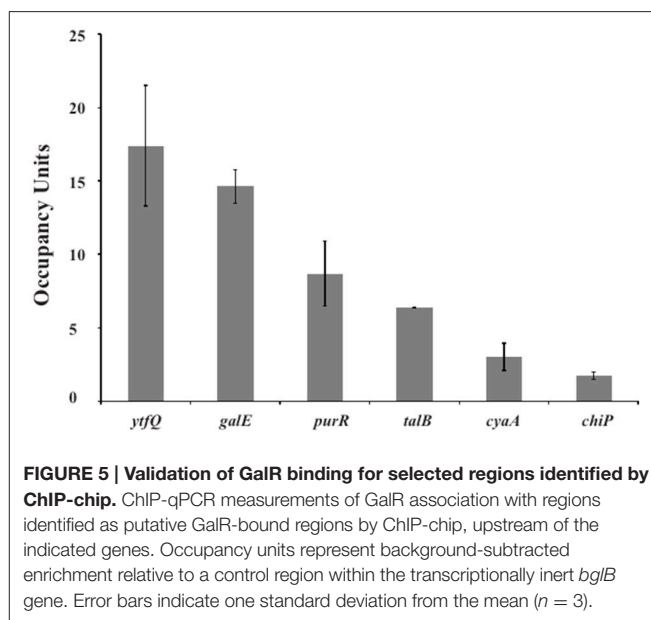
Chromosome coordinate	GalR binding bits	Cognate gene
37821	10.03857	<i>calC</i>
43400	9.814199	<i>fixB</i>
<b>74447</b>	<b>11.21879</b>	<b><i>thiQ-thiB</i></b>
89735	11.40643	<i>mraZ</i>
<b>103352</b>	<b>9.751084</b>	<b><i>ftsQ</i></b>
<b>161073</b>	<b>9.635808</b>	<b><i>sfsA</i></b>
167231	9.801062	<i>mrcB</i>
<b>234579</b>	<b>10.76336</b>	<b><i>gloB</i></b>
<b>306553</b>	<b>10.13984</b>	<b><i>ecpE-ecpC</i></b>
<b>390979</b>	<b>9.664174</b>	-
<b>741888</b>	<b>11.21261</b>	<b><i>dtpD</i></b>
787535	12.30803	<i>gpmA</i>
791362	10.91125	<i>galE</i>
<b>792028</b>	<b>20.50652</b>	<b><i>galE</i></b>
<b>792141</b>	<b>19.60408</b>	<b><i>galE-modF</i></b>
914977	9.870273	<i>ybjE-aqpZ</i>
986589	10.22057	<i>ompF</i>
<b>1109183</b>	<b>10.44796</b>	<b><i>opgC-opgG</i></b>
1191794	9.613439	<i>purB</i>
1253804	9.870957	<i>ycgV</i>
1307943	10.76489	<i>clsA</i>
<b>1347064</b>	<b>9.781962</b>	<b><i>mb</i></b>
1353246	10.53007	<i>sapD</i>
<b>1466984</b>	<b>9.647879</b>	-
1539818	12.88323	<i>narY-narU</i>
1572923	11.06094	<i>pqqL</i>
1712019	10.58487	<i>rsx-dtpA</i>
1798963	9.813867	<i>pheS-pheM</i>
1803105	9.629957	-
<b>1857739</b>	<b>13.26586</b>	<b><i>ydiI</i></b>
1958198	11.84791	<i>torY-cutC</i>
2012349	10.87027	-
<b>2076502</b>	<b>11.88679</b>	<b><i>yeeR-yeeT</i></b>
2188111	11.89384	<i>yehA</i>
<b>2239539</b>	<b>11.87394</b>	<b><i>mgIB</i></b>
<b>2240625</b>	<b>13.75395</b>	<b><i>mgIB-galS</i></b>
<b>2241618</b>	<b>9.470329</b>	<b><i>galS</i></b>
<b>2241771</b>	<b>10.95166</b>	<b><i>galS-yeiB</i></b>
2390045	11.86583	<i>yfbP-nuoN</i>
2585453	9.76911	<i>aegA</i>
<b>2738463</b>	<b>10.28531</b>	<b><i>pheA</i></b>
2751444	11.95696	<i>nadK</i>
2783823	12.3442	<i>ypjC</i>
2839356	12.65853	<i>ascG-ascF</i>
<b>2976576</b>	<b>16.03935</b>	<b><i>omrB-galR</i></b>
<b>2976837</b>	<b>14.07919</b>	<b><i>galR</i></b>
3069624	11.39436	<i>mscS</i>
<b>3088011</b>	<b>14.78781</b>	<b><i>metK-galP</i></b>
<b>3088193</b>	<b>25.1484</b>	<b><i>metK-galP</i></b>
3115470	9.845925	<i>sslE</i>
3236977	10.89744	<i>ygiQ</i>

(Continued)

TABLE 2 | Continued

Chromosome coordinate	GalR binding bits	Cognate gene
3287545	12.34189	<i>yraH</i>
3288641	10.85878	<i>yral</i>
3492468	9.731226	<i>ppiA-tsgA</i>
<b>3656067</b>	<b>10.56542</b>	<b><i>hdeB</i></b>
3665637	9.573374	<i>gadX</i>
3700787	10.21665	<i>yhjV</i>
<b>4124542</b>	<b>13.28446</b>	<b><i>cytR-priA</i></b>
4155030	10.04465	<i>argC</i>
4573916	11.67178	-

The bold are also present in Table S1.



## Microarray Hybridization

Labeled cDNA fragments (3  $\mu$ g) were then hybridized for 16 h (60 rpms) at 45°C to tiling array chips (Ecoli\_Tab520346F) purchased from Affymetrix (Santa Clara, CA). The chips have 1,159,908 probes in 1.4 cm  $\times$  1.4 cm and a 25-mer probe every 8 bps in both strands of whole *E. coli* genome. In addition, the probes are also overlapped by 4 bps with other strand probes. Each 25-mer DNA probe in the tiling array chip are 8 bp apart from the next probe. Probes are designed to cover the whole *E. coli* genome.

## Microarray: Washing and Staining

The chips were then washed with Wash Buffer A: Non-Stringent Wash Buffer (6X SSPE, 0.01% Tween-20). Wash Buffer B: (100 mM MES, 0.1M [Na<sup>+</sup>] and 0.01% Tween-20) and stained with Streptavidin Phycoerythrin (Molecular Probes S-866) and anti-streptavidin antibody (goat), biotinylated (Vector Laboratories BA-0500) on a Genechip Fluidics Station 450 (Affymetrix) according to washing and staining protocol, ProkGE-WS2\_450.

**TABLE 3 | GalR-bound regions identified by ChIP-chip assays.**

Peak probe position <sup>a</sup>	Binding score <sup>b</sup>	Nearby gene(s) <sup>c</sup>	Inferred binding site(s) <sup>d</sup>
<b>8005</b>	<b>15.4</b>	<b>yaaJ/talB</b>	<b>TTGGTAACGTTTACA</b>
708069	4.9	chiP	ATGAAAGCGGTTACA
767563	4.5	(mngA)	GTGGAAGCGGTTACG
<b>792029</b>	<b>78.1</b>	<b>galE</b>	<b>GTGGTAGCGGTTACA</b>
792441			GTGGAATCGTTTACA
1627951	3.9	(ydfG)	GTGGTAACGTTTACG
<b>1737880</b>	<b>9.5</b>	<b>ynhF/purR</b>	<b>GTGGAAACGTTTGCT</b>
1737791			AGGC AACGTTTACC
<b>2240625</b>	<b>9.5</b>	<b>mgIB</b>	<b>TTGAAAGCGGTTACA</b>
<b>2241771</b>	<b>10.9</b>	<b>galS</b>	<b>ATGAAAGCGGTTACA</b>
<b>2241619</b>			<b>GTGGCAACGGTTTCC</b>
<b>2976577</b>	<b>21.2</b>	<b>omrB/galR</b>	<b>GGGTAAACGCTTACA</b>
<b>3088194</b>	<b>115.4</b>	<b>galP</b>	<b>GTGTAATCGCTTACA</b>
3532886	4.6	yhgE/pck	ATGATATCGTTTACA
3991055	12.4	hemC/cyaA	GTGGTAACGGTTACC
4124542	3.5	cytR	GTGAAAGCGGTTACA
4338179	6.9	adiY	ATGGCAACGTTTCA
4338257			GTGGTTACGCTTTCA
<b>4449971</b>	<b>22</b>	<b>ppa/ytfQ</b>	<b>GTGGAAACGCTTACT</b>

The bold labeled motifs are the GRS as defined in text.

<sup>a</sup>Genome coordinate corresponding to the center of the microarray probe in the associated GalR-bound region.

<sup>b</sup>Ratio of ChIP-chip signal for the ChIP and input control samples, for the peak probe (i.e., the microarray probe with the highest ratio in the GalR-bound region).

<sup>c</sup>Genes in parentheses correspond to peak probes whose genomic location does not overlap with an intergenic region upstream of a gene. All other genes listed begin immediately downstream of intergenic regions that overlap the peak probe.

<sup>d</sup>Putative GalR binding site(s) identified using MEME.

## Microarray: Scanning and Data Analysis

Hybridized, washed, and stained microarrays were scanned using a Genechip Scanner 3000 (Affymetrix). Standardized signals, for each probe in the arrays, were generated using the MAT analysis software, which provides a model-based, sequence-specific, background correction for each sample (Johnson et al., 2006). A gene specific score was then calculated for each gene by averaging all MAT scores (natural log) for all probes under the annotated gene coordinates. Gene annotation was from the ASAP database at the University of Wisconsin-Madison, for *E. coli* K-12 MG1655 version m56 (Glasner et al., 2003). Data were graphed with ArrayStar<sup>®</sup>, version 2.1. DNASTAR, Madison, WI. The tiling array data was submitted to NCBI Gene Expression Omnibus. The accession number is GSE85334.

## ChIP-Chip Assays

MG1655 *galR*-TAP (AMD032) cells were grown in LB at 37°C to an OD<sub>600</sub> of ~0.6. ChIP-chip was performed as described previously (Stringer et al., 2014). Data analysis was performed as described previously except that probes were ignored only if they had a score of <100 pixels, indicating regions that are likely missing from the genome (Stringer

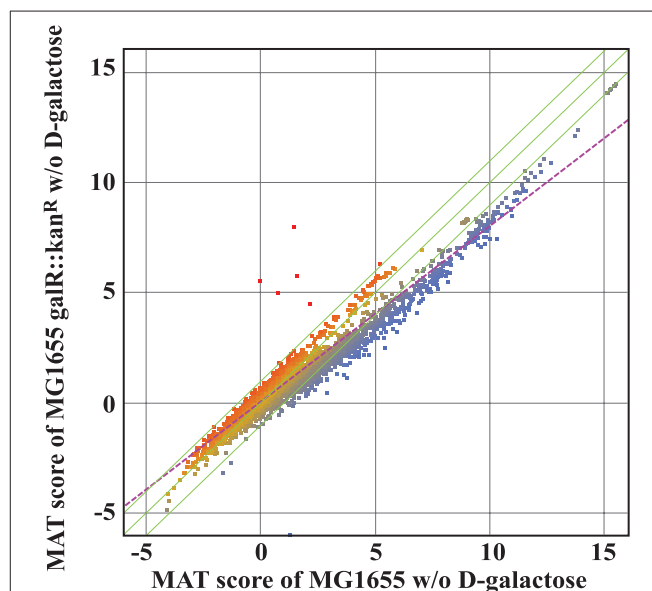
et al., 2014). Adjacent probes scoring above the threshold for being called as being in GalR-bound regions were merged, and the highest-scoring probe was selected as the “peak position.” The closely spaced peaks upstream of *mgIB* and *galS* were manually separated. The ChIP-chip data was submitted to the EBI Array Express repository. The accession number is E-MTAB-4903.

## Identification of an Enriched Sequence Motif from ChIP-Seq Data

For each peak position, we extracted genomic DNA sequence using the following formulae to determine the upstream and downstream coordinates: upstream coordinate:  $U_p - ((U_p - U_{p-1}) * (S_{p-1} / S_p))$ ; downstream coordinate:  $D_p - ((D_{p+1} - D_p) * (S_{p+1} / S_p))$ ; where  $S$  = probe score,  $U$  = genome coordinate corresponding to the upstream end of a probe,  $D$  = genome coordinate corresponding to the downstream end of a probe,  $p$  = peak probe,  $p-1$  = probe upstream of peak, and  $p+1$  = probe downstream of peak. We used MEME (version 4.11.2, default parameters except any number of motif repetitions was allowed) to identify an enriched sequence motif (Bailey and Elkan, 1994).

## ChIP-qPCR

MG1655 *galR*-FLAG<sub>3</sub> (AMD188) cells were grown in LB at 37°C to an OD<sub>600</sub> of 0.6–0.8. ChIP-qPCR was performed as described previously (Stringer et al., 2014).



**FIGURE 6 | MAT analysis of the transcriptome of wild type and  $\Delta galR$  cells grown in M63 minimal medium.** Green lines represent the mean  $\pm$  2SD, while the purple dotted line represents the regression line. The red represents the up genes and the blue represents the down genes. There is a marked down-regulation of many genes in the absence of GalR.

TABLE 4 | Gene Regulatory Sites discovered by sequence analysis.

Operator sequence	Strand	Chromosome	Coordinate	Cognate gene
TTGGTAACGTTTACAC	—	7997	8012	<i>yaaJ/b0007</i>
GTTTAACCGCATTACAC	—	10286	10301	<i>satP/b0010</i>
GCGCAACCGCTACCAC	+	74440	74455	<i>thiP/b0067</i>
GTGGTGGCGCTTACAC	+	74460	74475	<i>thiP/b0067</i>
GTGAACGCCATTACAC	+	103345	103360	<i>ftsQ/b0093</i>
CTGAAAGGGTTTGCAC	+	118258	118273	<i>nadC/b0109</i>
GTGAACCTCTTTCCAC	+	151361	151376	<i>yadK/b0136</i>
TTGCAAATGGTTCCAC	—	151873	151888	<i>yadL/b0137</i>
GTGAAAATGATTGCAC	—	155580	155595	<i>yadV/b0140</i>
GTGGAAGGTTTCCAC	—	234572	234587	<i>gloB/b0212</i>
GTAGAAACGCTGCAC	+	329207	329222	<i>betI/b0313</i>
GTGGCATCGCTTCCAC	—	397495	397510	<i>sbmA/0377</i>
GTTTAAGCACTTTCAC	+	684934	684949	<i>gltL/b0652</i>
ATGAAATCGATGCCAC	—	720300	720315	<i>speF/b0693</i>
<b>ATGTAACCGCTACCAC</b>	<b>+</b>	<b>792021</b>	<b>792036</b>	<b><i>galE/b0759</i></b>
<b>GTGGAATCGTTTACAC</b>	<b>+</b>	<b>792134</b>	<b>792149</b>	<b><i>galE/b0759</i></b>
GTGAAGCGCTGTCAC	—	863678	863693	<i>fsaA/b0825</i>
GTAACCCCGTTTCAC	—	957266	957281	<i>ycaP/b0906</i>
GTCAAAACAGTTGCAC	+	1074131	1074146	<i>rutR/b1013</i>
TTGCAACCGTTTTCAC	—	1109176	1109191	<i>opgG/b1048</i>
GTGACATCGCGTCCAC	—	1110864	1113406	<i>opgH/b1049</i>
GAGGAACCGGTAGCAC	—	1156743	1156758	<i>holB/b1099</i>
GTGCAACCGCTATCAG	—	1176651	1176666	<i>lolD/b1117</i>
CTGAAATGGCTTTCAC	+	1413858	1413873	<i>ralR/b1348</i>
CTGCAAGCGCTTGAAC	+	1674225	1674240	<i>tqsA/b1601</i>
GAGCAACGCTTTCAC	+	1737872	1737887	<i>purR/b1658</i>
ATGGAAGCTTTTCCAC	+	1771431	1771446	<i>ydiM/b1690</i>
GAGTAACCGTCTACAC	—	1791263	1791278	<i>ydiU/b1706</i>
GGGAAAACGATGCCAC	—	1857732	1857747	<i>ydiJ/b1773</i>
GTGTCATCGACTGCAC	—	1896369	1896384	<i>nudL/b1813</i>
GTGCAGGAGATTGCAC	+	2005842	2005857	<i>flit/b1926</i>
ATGGAACATTTCAC	+	2012342	2012357	<i>yedN/b1932</i>
GTGAAGAGGGTTTCAC	—	2076495	2076510	<i>yeeS/b2002</i>
ATGCAACCGGTTACCC	—	2077222	2077237	<i>cbeA/b2004</i>
GTGTACGCATTTCAC	+	2108205	2108220	<i>glf/b2036</i>
GTGCACCGGATTTCAC	+	2239532	2239547	<i>mgIB/b2150</i>
TTGAAAGCGTTTACAT	+	2240618	2240633	<i>galS/b2151</i>
GGGAAAACGTTGCCAC	+	2241611	2241626	<i>galS/b2151</i>
GCGGAATCGGTTCAAC	+	2278144	2278159	<i>yefG/b2181</i>
GTGCGAACTCTTCCAC	+	2414919	2414934	<i>pta/b2297</i>
<b>CTGCATCCGTTTGCAC</b>	<b>+</b>	<b>2427600</b>	<b>2427615</b>	<b><i>argT/b2310</i></b>
CTGCAATCGCCTTCAC	+	2527286	2527301	<i>yfeH/b2410</i>
ATGCAATCGGTTACGC	—	2634124	2634139	<i>guaB/b2508</i>
GTGTACTCTATTACAC	—	2637479	2637494	<i>bamB/b2512</i>
<b>GTAAGACGATTTCAC</b>	<b>+</b>	<b>2661317</b>	<b>2661332</b>	<b><i>iscS/b2530</i></b>
GTGTCGCCGTTTTCAC	+	2796513	2796528	<i>ygaU/b2665</i>
GAGGAAGCGGTTTCAC	+	2817230	2817245	<i>yqaB/b2690</i>
CTGGAAGCGATTGCC	—	2832047	2832064	<i>norR/b2709</i>
GTGTGAACATTTCAC	—	2837945	2837960	<i>hydN/b2713</i>
AAGAAACCGGTTTCAC	—	2839425	2839440	<i>ascF/b2715</i>
CTGCAAGCGTTGCAC	+	2848912	2848927	<i>hycG/b2723</i>

(Continued)

TABLE 4 | Continued

Operator sequence	Strand	Chromosome	Coordinate	Cognate gene
ATGTAAGCGTTTACCC	+	2976569	2976584	<i>galR/b2837</i>
GTTGCACCGCTTTCAC	—	2976830	2976845	<i>galR/b2837</i>
GTTAAAGCATTTACAC	—	2995106	2995121	<i>ygeK/b2856</i>
ATGCAAGTGCTTTCAC	—	3041236	3041251	<i>ygfZ/b2898</i>
<b>CTGAAACCGATTACAC</b>	<b>+</b>	<b>3088004</b>	<b>3088019</b>	<b><i>galP/b2943</i></b>
<b>GTGTAAGCGATTACAC</b>	<b>+</b>	<b>3088186</b>	<b>3088201</b>	<b><i>galP/b2943</i></b>
GTTGCAGCGATTTCAC	+	3133074	3133089	<i>yghR/b2984</i>
GAGGAAGTGATTGCAC	—	3320107	3320122	<i>yhbX/b3173</i>
CTGGAACCGTATTTCAC	+	3372189	3372204	<i>nanT/b3224</i>
GTGGGATCGAGTACAC	—	3375005	3375020	<i>dcuD/b3227</i>
<b>GTAAGAACGGTTACAC</b>	<b>—</b>	<b>3453341</b>	<b>3453356</b>	<b><i>rpsJ/b3321</i></b>
AGGAAACCGCTTCCAC	—	3540370	3540385	<i>feoA/b3408</i>
CAGGAAGCGCTTTCAC	—	3552425	3552440	<i>malP/b3417</i>
ATCAAATCGATTACAC	—	3710451	3710466	<i>eptB/b3546</i>
GCGCAACCGCTTCCAC	+	3760922	3760937	<i>selA/b3591</i>
GCGAAATTGATTACAC	+	3824831	3824846	<i>trmH/b3651</i>
<b>GCGCAACCGTTCTCAC</b>	<b>+</b>	<b>3884368</b>	<b>3884383</b>	<b><i>rpmH/b3703</i></b>
GGGTAATCGCGTCCAC	—	4256787	4256802	<i>dgkA/b4042</i>
GTGCAAAAGATTGCAC	—	4281671	4281686	<i>yjcE/b4065</i>
GGGTAATCGGTTTTAC	—	4330520	4330535	<i>proP/b4111</i>
GAGAAAACGCTTCAAC	—	4378149	4378164	<i>ampC/b4150</i>
CTGCGCATCGTTTACAC	—	4433627	4433642	<i>qorB/b4211</i>
AAGTAAGCGTTTCCAC	—	4449964	4449979	<i>ytfQ/b4227</i>
TTGCCACCGCTTTCAC	—	4483949	4483964	<i>holC/b4259</i>

The motifs in bold letters are also present in Table S2.

## RESULTS

### *In silico* Identification of Novel GalR Target Genes in *E. coli*

A consensus sequence of GalR binding sites from the previously known functional 9 operators in the *gal* regulon (*galE*, *galP*, *mgIB*, *galS*, and *galR* promoters; **Figure 1**) appears to be a 16-bp hyphenated dyad symmetry sequence with the center between positions 8 and 9: <sup>1</sup>GTGNAANC.GNTTNCAC<sup>16</sup> (with N being any nucleotide; Weickert and Adhya, 1993a). Genetic analysis showed that mutations at any of the positions 3, 5, 9, and 15 (labeled in bold) create a functionally defective operator (Adhya and Miller, 1979). Therefore, we used a motif in which nucleotides at positions 3, 5, 9, and 15 were fixed to search through the whole genome of *E. coli* (NC\_000193.3) (Baba et al., 2006) for putative GalR operators, allowing two mismatches at other non-N positions as described (Qian et al., 2012). Thus, we found 165 potential GalR operators distributed across the genome (Table S1).

Further analysis of the original 9 GalR-target operators sequences with critical information content was conducted (**Figure 1**; Schneider and Mastronarde, 1996). A unique alignment of 42 bp length was obtained; the information content of the optimally aligned sites was  $R_{\text{sequence}} = 16.1 \pm 0.7$  bits/site for the 42 bp sequence range (Shannon, 1948; Pierce, 1980; Schneider et al., 1986). The information content



needed to find these 9 sites in the 4,641,652 bp *E. coli* genome (NC\_000913.3) is  $R_{\text{frequency}} = 18.98$  bits/site; the information content in the sites is not sufficient for them to be found in the genome,  $R_{\text{sequence}}/R_{\text{frequency}} = 0.85 \pm 0.04$ , so the binding sites do not have enough information content for them to be located in the genome (Schneider et al., 1986; Schneider, 2000). This result implies that there could be  $66 \pm 32$  sites in the genome. As shown in Figure 4, the sequence logo of the binding sites covers the DNase I protection segment (Majumdar and Adhya, 1987; Schneider and Stephens, 1990). There may be additional conservation near a DNase I-hypersensitive site in a major groove one helical turn from the central two major grooves bound by GalR ( $-16$  and  $+17$ ; Figure 4). The sequence conservation in the center of the site at bases 0 and 1 exceeds the sine wave, indicating that GalR binds to non-B-form DNA (Schneider, 2001) as was previously suggested (Majumdar and Adhya, 1989). An individual information weight matrix corresponding to positions  $-20$  to  $+21$  of the logo in Figure 4 was created and scanned across the *E. coli* genome (Schneider, 1997). Sixty sites were identified that contain more than 9.4 bits,

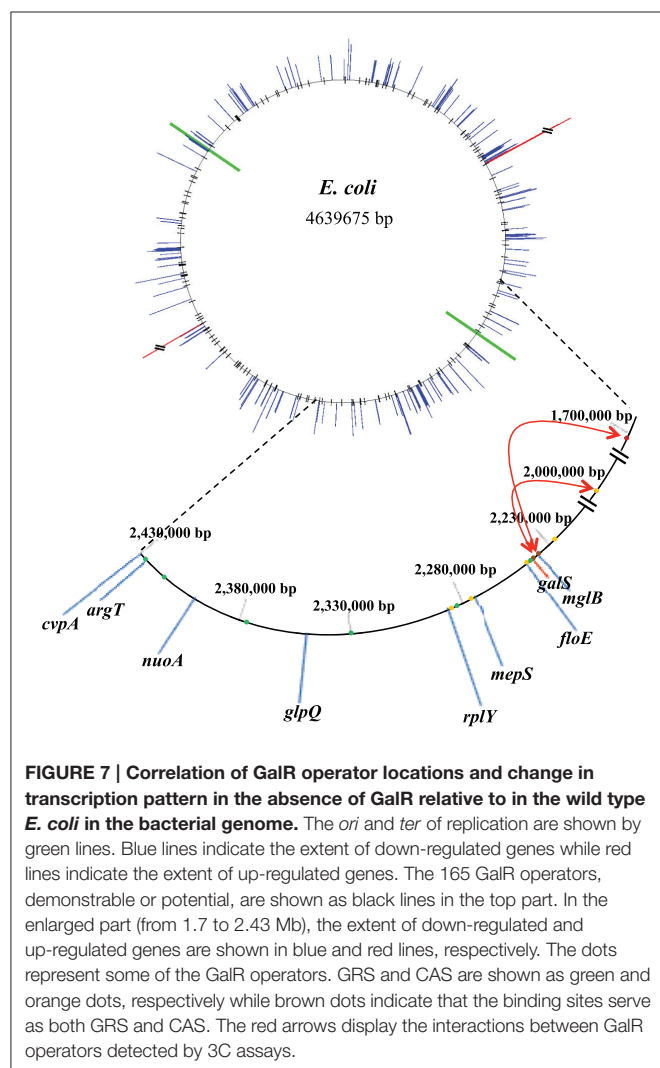
the lowest information content of the biochemically proven sites. The sequences of novel GalR predicted sites corresponding to the logo are summarized in Table 2.  $R_{\text{frequency}}$  for these sites in the genome is 16.24 bits/site, which is close to the observed  $16.3 \pm 0.1$  bits/site from all the predicted genomic sites.

## Functional Analysis of the Putative GalR Binding Sites Using ChIP-chip Assays

For the functional analysis of the putative binding sites, a ChIP-chip assay was performed to detect GalR target sequences genome-wide *in vivo* (Collas, 2010; Wade, 2015). In this ChIP-chip assay the binding of C-terminally TAP (tandem affinity purification) -tagged GalR (tagged at its native locus in an unmarked strain) was mapped across the *E. coli* genome. The experimental data resulting from ChIP-chip analysis were validated by quantitative real-time PCR (ChIP/qPCR). To demonstrate that the ChIP signal was not an artifact of the TAP tag, we constructed an unmarked derivative of *E. coli* MG1655 that expressed a C-terminally FLAG<sub>3</sub>-tagged GalR from its native locus. We selected six (*ytfQ*, *galE*, *purR*, *talB*, *cyaA*, and *chiP*) sites for validation, including *ytfQ*, *talB*, and *cyaA* that had not been described or predicted previously. In all cases, we detected significant signal of GalR binding indicating that these are genuine sites of GalR binding (Figure 5). The inferred binding sites from ChIP-chip assays are listed in Table 3. We identified 15 GalR-bound regions, four of which contain two operators. These include 8 known operators (in *galE*, *galP*, *galS*, *galR*, *chiP*, and *mglB*; Weickert and Adhya, 1993b; Plumbridge et al., 2014). Thirteen of the 15 putative GalR-bound regions overlap an intergenic region upstream of a gene start. This is a strong enrichment over the number expected by chance (only  $\sim 12\%$  of the genome is intergenic).

## Global Transcription Profile in the Presence and Absence of GalR

Since both *in silico* investigation and ChIP-chip assays suggested that the regulatory role of GalR goes beyond D-galactose metabolism, we used transcriptome profiling to gain further insight into the impact of GalR on genome-wide transcription. To evaluate the effect of *galR* deletion on global gene expression patterns, we compared the ratio of RNA isolated from a  $\Delta galR$  mutant to that isolated from wild-type cells, using DNA tiling microarrays (Tokeson et al., 1991). The results of the transcriptional analysis are displayed in the MAT plot shown in Figure 6. For all analysis, we arbitrarily selected a stringent ratio cut-off of 3. We identified 238 genes with values exceeding this cut-off (Table S2). These 238 genes are transcribed from 158 promoters. Three transcripts (5 genes) of the 158 promoters are up-regulated (GalR acting as a repressor) and 155 transcripts (233 genes) are down-regulated (GalR acting as an activator; Table S2). Interestingly, several genes including *mglB* are dys-regulated by GalR but fall outside of the cut-off range. All three (*galP*, *galP1*, and *galP2*) of the up-regulated promoters have adjacent operators. Of the 155 down-regulated promoters, 4 promoters contain adjacent operators and the remaining 151 do not.



## DISCUSSION

Using a combination of bioinformatic and experimental approaches we identified many putative novel GalR operators in the *E. coli* genome. As expected, several of these putative operators were identified by both information theory and ChIP-chip assays, demonstrating that they represent genuine GalR binding sites. Thus, we have substantially expanded the known GalR regulon. Surprisingly, our data suggest that GalR, a regulator of D-galactose metabolism, also regulates the expression of genes involved in other cellular processes. Interestingly, three of the putative novel GalR target genes—*cytR*, *purR*, and *adiY*—encode transcription factors, suggesting that GalR may be part of a more complex regulatory network. Moreover, putative GalR operators upstream of *cytR* and *purR* overlap with operators for CytR and PurR, respectively, indicating combinatorial regulation of these genes (Meng et al., 1990; Rolfes and Zalkin, 1990; Mengeritsky et al., 1993). Despite our identification of GalR operators with high confidence upstream of genes mentioned above, our expression microarray data show little or no regulation of these genes by GalR. We propose that regulation of these genes by GalR is condition-specific, requiring input from additional regulatory factors.

### Role of GalR in Gene Regulation

DNA tiling array analysis revealed that the transcription of a surprisingly large number of promoters (158) in *E. coli* is dysregulated by deletion of the *galR* gene. On the other hand, we identified 165 established or potential GalR operators in the chromosome, 76 of which are located between −200 to +400 bp from the *tsp* of promoters (cognate), and the other 89 operators are not (Table S1). We called the former group of operators, “Gene Regulatory Sites” (GRS, listed in Table 4). Consistent with a previous proposal (Macvanin and Adhya, 2012), we believe that 89 non-cognate operators around the chromosome are playing an architectural role in chromosome organization. The unattached operators would be referred to as “Chromosome Anchoring Sites” (CAS). Some of the sites may serve as both GRS and CAS. The 76 (46%) GRS and 89 (54%) CAS are shown in Table S1. Seventy-six GRS include 9 previously known operators of the *gal* regulon (see Figure 1); the other 67, which control promoters, were not known previously. The discovery of new GRS indicates that GalR, a well-known regulator of D-galactose metabolism, also regulates the expression of other genes. Among the new GRS, 3 (in *yaaJ*, *purR*, and *ytfQ* promoters) were confirmed by *in vivo* DNA-binding (ChIP-chip assays) as shown in Table 3. The salient features of our findings presented in this paper are shown schematically in Figure 7.

Although we identified 158 transcripts whose expression was regulated by GalR, very few of these are associated with a putative GalR operator identified *in silico* and/or ChIP-chip assays, strongly suggesting that the majority of regulation by GalR occurs indirectly. Based on our earlier observation that GalR mediates mega-loop formation, we propose that long-range oligomerization of GalR indirectly regulates transcription by altering chromosome structure. There are at least three possible

mechanisms for such regulation: indirect control, enhancer activity, and modulation of DNA superhelicity. In the indirect control model, GalR directly regulates another regulator, such as PurR or CytR, and the downstream regulator directly regulates other genes. The regulation by GalR is indirect, but occurs by a classical regulatory mechanism. In the enhancer activity model, GalR stimulates transcription of some target genes by binding to a distal site and forming an enhancer-loop with a protein bound to the promoter region. Examples of enhancer activity have been described before for some prokaryotic and many eukaryotic promoters (Rombel et al., 1998; Schaffner, 2015). In the DNA superhelicity modulation model, GalR creates DNA topological domains by mega-loop formation and defines local chromosomal superhelicity by GalR-GalR interactions between distally bound dimers. The strength of a promoter is usually defined by superhelical nature of the DNA (Pruss and Drlica, 1989; Lim et al., 2003). We propose that GalR entraps different amount of superhelicity in different topological domains and thus controls transcription of the constituent promoters. In the absence of GalR such domains are not formed resulting in a change in local DNA superhelicity, and thus a change in the strength of the constituent promoters. In this model, GalR protein indirectly regulates gene transcription as an architectural protein. We are currently studying the regional superhelicities in the entire chromosome in the presence and absence of GalR as well as the implication of genes affected by GalR, but independent of D-galactose metabolism (Lal et al., 2016).

## AUTHOR CONTRIBUTIONS

ZQ: designed genome-wide sequence analysis, interpreted sequence analysis data and tiling array data; AT and SL: executed tiling array experiments and data analysis; XH: executed genome-wide sequence analysis; TD: integrated tiling array and genome-wide sequence data; AS and JW: executed ChIP-chip and ChIP-qPCR experiments and data analysis; DL: data analysis; TS: executed Information Theory and data analysis; SA: organized and designed experiments, and data analysis. All authors contributed to the manuscript preparation.

## ACKNOWLEDGMENTS

This work was supported by the Intramural Research Program of the National Institutes of Health, the National Cancer Institute, and the Center for Cancer Research. The authors have no conflict of interest to declare. We thank the Wadsworth Center Applied Genomic Technologies Core Facility for assistance with microarrays for ChIP-chip assays.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmolb.2016.00074/full#supplementary-material>

## REFERENCES

- Adhya, S., and Miller, W. (1979). Modulation of the two promoters of the galactose operon of *Escherichia coli*. *Nature* 279, 492–494. doi: 10.1038/279492a0
- Aiba, H., Adhya, S., and de Crombrughe, B. (1981). Evidence for two functional gal promoters in intact *Escherichia coli* cells. *J. Biol. Chem.* 256, 11905–11910.
- Aki, T., and Adhya, S. (1997). Repressor induced site-specific binding of HU for transcriptional regulation. *EMBO J.* 16, 3666–3674. doi: 10.1093/emboj/16.12.3666
- Aki, T., Choy, H. E., and Adhya, S. (1996). Histone-like protein HU as a specific transcriptional regulator: co-factor role in repression of gal transcription by GAL repressor. *Genes Cells* 1, 179–188. doi: 10.1046/j.1365-2443.1996.d01-236.x
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2, 2006 02008. doi: 10.1038/msb4100050
- Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.
- Butland, G., Peregrin-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., et al. (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* 433, 531–537. doi: 10.1038/nature03239
- Choy, H. E., Park, S. W., Aki, T., Parrack, P., Fujita, N., Ishihama, A., et al. (1995). Repression and activation of transcription by Gal and Lac repressors: involvement of alpha subunit of RNA polymerase. *EMBO J.* 14, 4523–4529.
- Collas, P. (2010). The current state of chromatin immunoprecipitation. *Mol. Biotechnol.* 45, 87–100. doi: 10.1007/s12033-009-9239-8
- Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6640–6645. doi: 10.1073/pnas.120163297
- Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311. doi: 10.1126/science.1067799
- Glasner, J. D., Liss, P., Plunkett, G. III, Darling, A., Prasad, T., et al. (2003). ASAP, a systematic annotation package for community analysis of genomes. *Nucleic Acids Res.* 31, 147–151. doi: 10.1093/nar/gkg125
- Hogg, R. W., Voelker, C., and Von Carlowitz, I. (1991). Nucleotide sequence and analysis of the mgl operon of *Escherichia coli* K12. *Mol. Gen. Genet.* 229, 453–459. doi: 10.1007/BF00267469
- Irani, M. H., Orosz, L., and Adhya, S. (1983). A control element within a structural gene: the gal operon of *Escherichia coli*. *Cell* 32, 783–788. doi: 10.1016/0092-8674(83)90064-8
- Ishihama, A. (2009). The nucleoid: an overview. *EcoSal Plus* 3, 1–47. doi: 10.1128/ecosal.2.6
- Johnson, W. E., Li, W., Meyer, C. A., Gottardo, R., Carroll, J. S., Brown, M., et al. (2006). Model-based analysis of tiling-arrays for ChIP-chip. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12457–12462. doi: 10.1073/pnas.0601180103
- Kar, S., and Adhya, S. (2001). Recruitment of HU by piggyback: a special role of GalR in repressosome assembly. *Genes Dev.* 15, 2273–2281. doi: 10.1101/gad.920301
- Kar, S., Edgar, R., and Adhya, S. (2005). Nucleoid remodeling by an altered HU protein: reorganization of the transcription program. *Proc. Natl. Acad. Sci. U.S.A.* 102, 16397–16402. doi: 10.1073/pnas.0508032102
- Lal, A., Dhar, A., Trostel, A., Kouzine, F., Seshasayee, A. S., and Adhya, S. (2016). Genome scale patterns of supercoiling in a bacterial chromosome. *Nat. Commun.* 7:11055. doi: 10.1038/ncomms11055
- Lee, S. J., Trostel, A., and Adhya, S. (2014). Metabolite changes signal genetic regulatory mechanisms for robust cell behavior. *MBio* 5:e00972-13. doi: 10.1128/mBio.00972-13
- Lim, H. M., Lewis, D. E., Lee, H. J., Liu, M., and Adhya, S. (2003). Effect of varying the supercoiling of DNA on transcription and its regulation. *Biochemistry* 42, 10718–10725. doi: 10.1021/bi030110t
- Loferer-Krössbacher, M., Klima, J., and Psenner, R. (1998). Determination of bacterial cell dry mass by transmission electron microscopy and densitometric image analysis. *Appl. Environ. Microbiol.* 64, 688–694.
- Luijsterburg, M. S., White, M. F., van Driel, R., and Dame, R. T. (2008). The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Crit. Rev. Biochem. Mol. Biol.* 43, 393–418. doi: 10.1080/10409230802528488
- Macvanin, M., and Adhya, S. (2012). Architectural organization in *E. coli* nucleoid. *Biochim. Biophys. Acta* 1819, 830–835. doi: 10.1016/j.bbarm.2012.02.012
- Majumdar, A., and Adhya, S. (1984). Demonstration of two operator elements in gal: *in vitro* repressor binding studies. *Proc. Natl. Acad. Sci. U.S.A.* 81, 6100–6104. doi: 10.1073/pnas.81.19.6100
- Majumdar, A., and Adhya, S. (1987). Probing the structure of gal operator-repressor complexes. conformation change in DNA. *J. Biol. Chem.* 262, 13258–13262.
- Majumdar, A., and Adhya, S. (1989). Effect of ethylation of operator-phosphates on Gal repressor binding. DNA contortion by repressor. *J. Mol. Biol.* 208, 217–223. doi: 10.1016/0022-2836(89)90383-5
- Meng, L. M., Kilstrup, M., and Nygaard, P. (1990). Autoregulation of PurR repressor synthesis and involvement of purR in the regulation of purB, purC, purL, purMN and guaBA expression in *Escherichia coli*. *Eur. J. Biochem.* 187, 373–379. doi: 10.1111/j.1432-1033.1990.tb15314.x
- Mengeritsky, G., Goldenberg, D., Mendelson, I., Giladi, H., and Oppenheim, A. B. (1993). Genetic and biochemical analysis of the integration host factor of *Escherichia coli*. *J. Mol. Biol.* 231, 646–657. doi: 10.1006/jmbi.1993.1316
- Musso, R. E., Di Lauro, R., Adhya, S., and de Crombrughe, B. (1977). Dual control for transcription of the galactose operon by cyclic AMP and its receptor protein at two interspersed promoters. *Cell* 12, 847–854. doi: 10.1016/0092-8674(77)90283-5
- Oberto, J., Nabti, S., Jooste, V., Mignot, H., and Rouviere-Yaniv, J. (2009). The HU regulon is composed of genes responding to anaerobiosis, acid stress, high osmolarity and SOS induction. *PLoS ONE* 4:e4367. doi: 10.1371/journal.pone.0004367
- Oehler, S., and Müller-Hill, B. (2010). High local concentration: a fundamental strategy of life. *J. Mol. Biol.* 395, 242–253. doi: 10.1016/j.jmb.2009.10.056
- Papp, P. P., Chatteraj, D. K., and Schneider, T. D. (1993). Information analysis of sequences that bind the replication initiator repA. *J. Mol. Biol.* 233, 219–230. doi: 10.1006/jmbi.1993.1501
- Pierce, J. R. (1980). *An Introduction to Information Theory: Symbols, Signals and Noise*, 2nd Edn. New York, NY: Dover Publications, Inc.
- Plumbridge, J., Bossi, L., Oberto, J., Wade, J. T., and Figueroa-Bossi, N. (2014). Interplay of transcriptional and small RNA-dependent control mechanisms regulates chitosugar uptake in *Escherichia coli* and *Salmonella*. *Mol. Microbiol.* 92, 648–658. doi: 10.1111/mmi.12573
- Priyadarshini, R., Cugini, C., Arndt, A., Chen, T., Tjokro, N. O., Goodman, S. D., et al. (2013). The nucleoid-associated protein HUBeta affects global gene expression in *Porphyromonas gingivalis*. *Microbiology* 159, 219–229. doi: 10.1099/mic.0.061002-0
- Pruss, G. J., and Drlaca, K. (1989). DNA supercoiling and prokaryotic transcription. *Cell* 56, 521–523. doi: 10.1016/0092-8674(89)90574-6
- Qian, Z., Dimitriadis, E. K., Edgar, R., Eswaramoorthy, P., and Adhya, S. (2012). Galactose repressor mediated intersegmental chromosomal connections in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11336–11341. doi: 10.1073/pnas.1208595109
- Rolfes, R. J., and Zalkin, H. (1990). Autoregulation of *Escherichia coli* purR requires two control sites downstream of the promoter. *J. Bacteriol.* 172, 5758–5766.
- Rombel, I., North, A., Hwang, I., Wyman, C., and Kustu, S. (1998). The bacterial enhancer-binding protein NtrC as a molecular machine. *Cold Spring Harb. Symp. Quant. Biol.* 63, 157–166. doi: 10.1101/sqb.1998.63.157
- Roy, S., Dimitriadis, E. K., Kar, S., Geanakopoulos, M., Lewis, M. S., and Adhya, S. (2005). Gal repressor-operator-HU ternary complex: pathway of repressosome formation. *Biochemistry* 44, 5373–5380. doi: 10.1021/bi047720t
- Roy, S., Semsey, S., Liu, M., Gussin, G. N., and Adhya, S. (2004). GalR represses galP1 by inhibiting the rate-determining open complex formation through RNA polymerase contact: a GalR negative control mutant. *J. Mol. Biol.* 344, 609–618. doi: 10.1016/j.jmb.2004.09.070
- Schaffner, W. (2015). Enhancers, enhancers - from their discovery to today's universe of transcription enhancers. *Biol. Chem.* 396, 311–327. doi: 10.1515/hsz-2014-0303
- Schneider, T. D. (1997). Information content of individual genetic sequences. *J. Theor. Biol.* 189, 427–441. doi: 10.1006/jtbi.1997.0540
- Schneider, T. D. (2000). Evolution of biological information. *Nucleic Acids Res.* 28, 2794–2799. doi: 10.1093/nar/28.14.2794

- Schneider, T. D. (2001). Strong minor groove base conservation in sequence logos implies DNA distortion or base flipping during replication and transcription initiation. *Nucleic Acids Res.* 29, 4881–4891. doi: 10.1093/nar/29.23.4881
- Schneider, T. D., and Mastronarde, D. N. (1996). Fast multiple alignment of ungapped DNA sequences using information theory and a relaxation method. *Discrete Appl. Math.* 71, 259–268. doi: 10.1016/S0166-218X(96)00068-6
- Schneider, T. D., and Stephens, R. M. (1990). Sequence logos - a new way to display consensus sequences. *Nucleic Acids Res.* 18, 6097–6100. doi: 10.1093/nar/18.20.6097
- Schneider, T. D., Stormo, G. D., Gold, L., and Ehrenfeucht, A. (1986). Information-content of binding-sites on nucleotide-sequences. *J. Mol. Biol.* 188, 415–431. doi: 10.1016/0022-2836(86)90165-8
- Schneider, T. D., Stormo, G. D., Haemer, J. S., and Gold, L. (1982). A design for computer nucleic-acid-sequence storage, retrieval, and manipulation. *Nucleic Acids Res.* 10, 3013–3024. doi: 10.1093/nar/10.9.3013
- Schneider, T. D., Stormo, G. D., Yarus, M. A., and Gold, L. (1984). Delila system tools. *Nucleic Acids Res.* 12, 129–140. doi: 10.1093/nar/12.1Part1.129
- Semsey, S., Geanakopoulou, M., Lewis, D. E., and Adhya, S. (2002). Operator-bound GalR dimers close DNA loops by direct interaction: tetramerization and inducer binding. *EMBO J.* 21, 4349–4356. doi: 10.1093/emboj/cdf431
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x
- Skoko, D., Yoo, D., Bai, H., Schnurr, B., Yan, J., McLeod, S. M., et al. (2006). Mechanism of chromosome compaction and looping by the *Escherichia coli* nucleoid protein Fis. *J. Mol. Biol.* 364, 777–798. doi: 10.1016/j.jmb.2006.09.043
- Stringer, A. M., Currenti, S. A., Bonocora, R. P., Petrone, B. L., Palumbo, M. J., Reilly, A. E., et al. (2014). Genome-scale analyses of *Escherichia coli* and *Salmonella enterica* AraC reveal non-canonical targets and an expanded core regulon. *J. Bacteriol.* 196, 660–671. doi: 10.1128/JB.01007-13
- Stringer, A. M., Singh, N., Yermakova, A., Petrone, B. L., Amarasinghe, J. J., Reyes-Diaz, L., et al. (2012). FRUIT, a scar-free system for targeted chromosomal mutagenesis, epitope tagging, and promoter replacement in *Escherichia coli* and *Salmonella enterica*. *PLoS ONE* 7:e44841. doi: 10.1371/journal.pone.0044841
- Tokeson, J. P., Garges, S., and Adhya, S. (1991). Further inducibility of a constitutive system: ultrainduction of the gal operon. *J. Bacteriol.* 173, 2319–2327.
- Wade, J. T. (2015). Mapping transcription regulatory networks with ChIP-seq and RNA-seq. *Adv. Exp. Med. Biol.* 883, 119–134. doi: 10.1007/978-3-319-23603-2\_7
- Weickert, M. J., and Adhya, S. (1993a). Control of transcription of gal repressor and isorepressor genes in *Escherichia coli*. *J. Bacteriol.* 175, 251–258.
- Weickert, M. J., and Adhya, S. (1993b). The galactose regulon of *Escherichia coli*. *Mol. Microbiol.* 10, 245–251.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Qian, Trostel, Lewis, Lee, He, Stringer, Wade, Schneider, Durfee and Adhya. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Post-translational Serine/Threonine Phosphorylation and Lysine Acetylation: A Novel Regulatory Aspect of the Global Nitrogen Response Regulator GlnR in *S. coelicolor* M145

Rafat Amin<sup>1</sup>, Mirita Franz-Wachtel<sup>2</sup>, Yvonne Tiffert<sup>3</sup>, Martin Heberer<sup>4</sup>, Mohamed Meky<sup>4</sup>, Yousra Ahmed<sup>4,5</sup>, Arne Matthews<sup>4</sup>, Sergii Krysenko<sup>4</sup>, Marco Jakobi<sup>4</sup>, Markus Hinder<sup>4</sup>, Jane Moore<sup>6</sup>, Nicole Okoniewski<sup>4</sup>, Boris Maček<sup>2</sup>, Wolfgang Wohlleben<sup>4</sup> and Agnieszka Bera<sup>4\*</sup>

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Centro de Investigaciones Biológicas -  
CSIC, Spain

### Reviewed by:

Sébastien Rigali,  
University of Liège, Belgium  
Marie-Joelle Virolle,  
Center National de la Recherche  
Scientifique, France

### \*Correspondence:

Agnieszka Bera  
agnieszka.bera@  
biotech.uni-tuebingen.de

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 26 May 2016

**Accepted:** 25 July 2016

**Published:** 09 August 2016

### Citation:

Amin R, Franz-Wachtel M, Tiffert Y,  
Heberer M, Meky M, Ahmed Y,  
Matthews A, Krysenko S, Jakobi M,  
Hinder M, Moore J, Okoniewski N,  
Maček B, Wohlleben W and Bera A  
(2016) Post-translational  
Serine/Threonine Phosphorylation and  
Lysine Acetylation: A Novel Regulatory  
Aspect of the Global Nitrogen  
Response Regulator GlnR in  
*S. coelicolor* M145.  
Front. Mol. Biosci. 3:38.  
doi: 10.3389/fmolb.2016.00038

<sup>1</sup> Department of Pathology, Dow International Medical College, Dow Research Institute of Biotechnology and Biomedical Sciences, Dow University of Health Sciences, Karachi, Pakistan, <sup>2</sup> Proteome Center Tübingen, Interdepartmental Institute for Cell Biology (IFIZ), University of Tübingen, Tübingen, Germany, <sup>3</sup> B.R.A.I.N. Biotechnology Research and Information Network AG, Zwingenberg, Germany, <sup>4</sup> Microbiology and Biotechnology, Interfaculty Institute of Microbiology and Infection Medicine, University of Tübingen, Tübingen, Germany, <sup>5</sup> Department of Pharmaceutical Biotechnology, Helmholtz Institute for Pharmaceutical Research Saarland, Saarland University Campus, Saarbrücken, Germany, <sup>6</sup> John Innes Center, Norwich Research Park, Norwich, UK

Soil-dwelling *Streptomyces* bacteria such as *S. coelicolor* have to constantly adapt to the nitrogen (N) availability in their habitat. Thus, strict transcriptional and post-translational control of the N-assimilation is fundamental for survival of this species. GlnR is a global response regulator that controls transcription of the genes related to the N-assimilation in *S. coelicolor* and other members of the *Actinomycetales*. GlnR represents an atypical orphan response regulator that is not activated by the phosphorylation of the conserved aspartate residue (Asp 50). We have applied transcriptional analysis, LC-MS/MS analysis and electrophoretic mobility shift assays (EMSAs) to understand the regulation of GlnR in *S. coelicolor* M145. The expression of *glnR* and GlnR-target genes was revisited under four different N-defined conditions and a complex N-rich condition. Although, the expression of selected GlnR-target genes was strongly responsive to changing N-concentrations, the *glnR* expression itself was independent of the N-availability. Using LC-MS/MS analysis we demonstrated that GlnR was post-translationally modified. The post-translational modifications of GlnR comprise phosphorylation of the serine/threonine residues and acetylation of lysine residues. In the complex N-rich medium GlnR was phosphorylated on six serine/threonine residues and acetylated on one lysine residue. Under defined N-excess conditions only two phosphorylated residues were detected whereas under defined N-limiting conditions no phosphorylation was observed. GlnR phosphorylation is thus clearly correlated with N-rich conditions. Furthermore, GlnR was acetylated on four lysine residues independently of the N-concentration in the defined media and on only one lysine residue in the complex N-rich medium. Using EMSAs we demonstrated that phosphorylation inhibited the binding of GlnR to its targets genes,

whereas acetylation had little influence on the formation of GlnR-DNA complex. This study clearly demonstrated that GlnR DNA-binding affinity is modulated by post-translational modifications in response to changing *N*-conditions in order to elicit a proper transcriptional response to the latter.

**Keywords:** nitrogen assimilation, *Streptomyces coelicolor*, GlnR, regulation, post-translational modifications, acetylation, phosphorylation

## INTRODUCTION

Streptomycetes, like other microorganisms, need to accurately modulate their regulatory network according to their developmental stage while simultaneously responding to environmental changes, such as the continuous variation in nutrient availability, including nitrogen (*N*), in the soil habitat. Therefore, strict transcriptional and post-translational control of the *N*-assimilation is fundamental for survival of streptomycetes. Transcriptional regulation of the *N*-assimilation in streptomycetes is accomplished by GlnR (global nitrogen response regulator; Tiffert et al., 2008; Pullan et al., 2011). However, control of the *N*-metabolism by GlnR is not only restricted to streptomycetes, since conserved GlnR homologs were also found in other actinomycetes such as: *Mycobacterium* sp., *Amycolatopsis* sp., *Saccharopolyspora* sp., *Bifidobacterium* sp., *Frankia* sp., *Nocardia* sp., *Propionibacterium* sp., and *Rhodococcus* sp. (Amon et al., 2009), signifying evolutionary importance of this regulator. The GlnR regulator in *S. coelicolor* controls at least 10 genes which are directly involved in *N*-metabolism and seven additional genes encoding proteins of unknown function (Reuther and Wohlleben, 2007; Tiffert et al., 2008; Wang and Zhao, 2009; Amin et al., 2012). Proteomic analysis demonstrated a more comprehensive regulatory role of GlnR in connection with central carbon metabolic pathways in *S. coelicolor* M145 (Tiffert et al., 2011). Over 50 proteins associated with amino acid biosynthesis and carbon metabolism were shown to be differentially expressed between *S. coelicolor* M145 and the *glnR* mutant (Tiffert et al., 2011). GlnR-mediated control of carbohydrate transport (Liao et al., 2015) and regulation of ectoin (Shao et al., 2015) as well as validomycin A production (Qu et al., 2015) extended the GlnR role beyond the regulation of the *N*-assimilation. Regulation of the *N*-metabolism in *S. coelicolor* is very complex and depending on the conditions, involves additional control by other regulators such as: GlnRII (Fink et al., 2002; Reuther and Wohlleben, 2007), PhoP (Rodríguez-García et al., 2009; Martín et al., 2011; Sola-Landa et al., 2013), Crp (Gao et al., 2012), and AfsQ1 (Wang R. et al., 2013).

Although, the expression of the GlnR target genes in *S. coelicolor* (Tiffert et al., 2008) and other actinomycetes was extensively studied (Pullan et al., 2011; Jenkins et al., 2013; Yao et al., 2014; Williams et al., 2015), little is known on how GlnR controls expression of its target genes according to changing *N*-conditions and thus how the DNA-binding activity of GlnR is modulated. The GlnR regulator belongs to the OmpR-family of transcriptional response regulators, commonly existing as a two component systems with a cognate histidine kinase. Usually the histidine kinase autophosphorylates upon reception

of an unknown signal from the environment and subsequently phosphorylates a conserved aspartic acid residue in the receiver domain, of the cognate response regulator, generating an appropriate adaptive response. The typical “phosphorylation pocket” of the response regulator OmpR is composed of six essential residues: the phosphor-accepting aspartate (Asp 55), three catalytic residues (Asp 11, Asp12, and Lys 105) and two conformational switch residues (Thr 87 and Tyr 106; Brissette et al., 1991). GlnR possess only two out of six conserved residues, namely Asp 50 and Thr 79 equivalent to Asp 55 and Thr 87 in OmpR, respectively. The analysis of the partial crystal structures of GlnR from *A. mediterranei* and *M. smegmatis* as well as the structure-based sequence alignment of GlnR from *S. coelicolor*, demonstrated that GlnR not only lacks the typical “phosphorylation pocket” but it is also not phosphorylated at the conserved Asp 50 residue (Lin et al., 2014). However, the conserved Asp50 residue is critical for GlnR homodimerization via its charge interactions with the surrounding residues and is essential for the physiological function of GlnR as shown by *in vitro* and *in vivo* studies (Lin et al., 2014). Furthermore, GlnR is an “orphan” response regulator since no associated sensor kinase gene could be found in its close proximity in the *S. coelicolor* M145 genome. So, since GlnR is not activated by the classical phosphorylation observed for canonical OmpR/PhoP—family members, important question remains still unanswered: how this regulator is activated? How does *S. coelicolor* sense the availability of different *N*-sources and how does GlnR elicit the proper transcriptional response according to changing *N*-conditions in the environment? In this study we report for the first time the phosphorylation of the global nitrogen regulator GlnR on serine/threonine residues as well as its acetylation on lysine residues. We also demonstrated that such unusual post-translational modifications play a crucial role in the regulation of the GlnR DNA-binding activity.

## MATERIALS AND METHODS

### Bacterial Strains, Plasmids, and Growth Conditions

Strains and plasmids used in this study are listed in **Table 1**. *E. coli* strains were cultivated either on a solid or in a liquid Luria-Bertani (LB) medium at 37°C (Sambrook et al., 1989). *Streptomyces coelicolor* M145 was cultivated at 30°C on R2YE agar or Mannitol Soy flour (MS) agar (Kieser et al., 2000). For growth in liquid medium, complex S-medium (Okanishi et al., 1974), and defined Evans medium (Evans et al., 1970) was used. Carbon to nitrogen ratio was set as follows: for

**TABLE 1 | Strains and plasmids used in this study.**

Strains	Genotype	References
<i>E. coli</i> BL21 (DE3)	F <sup>-</sup> , <i>dcm</i> , <i>ompT</i> , <i>hsdS</i> (r <sub>B</sub> -m <sub>B</sub> -), <i>gal</i> , (DE3)	Studier and Moffatt, 1986
<i>E. coli</i> BL21 pET15b-His-CobB1	His-CobB1 overexpression strain Cm <sup>R</sup> , Amp <sup>R</sup>	This work
<i>E. coli</i> BL21 pET15b-His-CobB2	His-CobB2 overexpression strain Cm <sup>R</sup> , Amp <sup>R</sup>	This work
<i>S. coelicolor</i> M145	<i>S. coelicolor</i> A3(2) <i>spc1</i> <sup>-</sup> and <i>spc2</i> <sup>-</sup>	Kieser et al., 2000
<i>S. coelicolor</i> M145 <i>glnR</i>	<i>glnR</i> mutant strain of <i>S. coelicolor</i> M145; <i>glnR</i> replaced by an <i>aac(3)/IV</i> cassette, Apr <sup>R</sup>	Tiffert et al., 2011
<i>S. coelicolor</i> M145 pGMStrep- <i>glnR</i>	<i>S. coelicolor</i> M154 with pGM-Strep- <i>glnR</i> , P <sub>tipA</sub> , Km <sup>R</sup>	Tiffert et al., 2008
pGM190	<i>Streptomyces-E.coli</i> shuttle vector, Ts <sup>R</sup> , Km <sup>R</sup> , pSG5 derivative, P <sub>tipA</sub>	Wohlleben et al., 2009
pET15b	His-(N-term), Amp <sup>R</sup> , P <sub>tipA</sub>	Novagen

*N*-limitation in Evans medium C:N of 2 and for *N*-excess in Evans medium C:N of 60. Media was supplemented when appropriate with: ampicillin (150 µg/ml), kanamycin (50 µg/ml), chloramphenicol (25 µg/ml), nalidixic acid (255 µg/ml) or thiostrepton (12.5 µg/ml), unless otherwise stated. Genetic manipulation of *S. coelicolor* M145 and *E. coli* was performed as described by (Kieser et al., 2000) and (Sambrook et al., 1989), respectively.

## RT-PCR

For the transcriptional analysis experiments, the *S. coelicolor* M145 wild type and the *glnR* mutant were grown in the complex S-medium for 4 days at 30°C. After 4 days, cells were harvested and washed twice with the defined Evans medium without *N*-source to remove traces of the S-medium. The biomass was subsequently transferred into the defined Evans medium supplemented with variable concentrations of either ammonium (low: 5 mM, high: 100 mM) or nitrate (low: 5 mM, high: 100 mM) and glucose (low: 2.5 g/l or high: 25 g/l). RT-PCR was conducted using total RNA isolated from *S. coelicolor* M145 and the *glnR* mutant after 24 h of growth in defined Evans medium. The RNA isolation was performed with an RNeasy kit (Qiagen). All RNA preparations were treated twice with DNase (Fermentas). First, an on-column digestion was carried out for 30 min at 24°C, and afterwards RNA samples were treated with DNase for 1.5 h at 37°C. RNA concentrations and quality were checked using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific). The cDNA from 3 µg RNA was generated with random nonamer primers (Sigma), reverse transcriptase and cofactors (Fermentas). The reverse transcription products (1 µl) were then used as template for PCR amplification. A standard PCR protocol using Taq DNA polymerase (GENAXXON bioscience) and primers annealing to internal parts of the various genes was used. Primers targeting *hrdB* were used as positive controls for RNA quality. Annealing temperatures were optimized for each primer combination. PCR reactions were performed with the primers listed in Table 2. The PCR conditions were as follows: 95°C for 5 min; 35 cycles of 95°C for 15 s, 55–60°C for 30 s and 72°C for 30 s, and 72°C for 10 min. Negative controls containing nuclease free water and total RNA were performed to exclude any DNA contamination. Positive controls containing total genomic DNA from *S. coelicolor* M145 were performed to ensure specific

**TABLE 2 | Primers used in this study.**

Oligonucleotide	Sequence 5'-3'	References
rt_hrdB1 (as control)	GAGTCCGTCTCTGTCATGGCG	Tiffert et al., 2008
rt_hrdB2 (as control)	TCGTCTCGTCGGACAGCAGC	
rt_glnA1	GGGACAAGACCCTCAACATC	Amin et al., 2012
rt_glnA2	CTTGATAGCGACCTTGTAAC	
rt_amtB1	TCCTGGTCTTCCAGCTGATG	
rt_amtB2	TTGCCGATGACGAGGATCAC	
rt_glnI1	ACCTGGAGAACTGCCTGAAG	
rt_glnI2	TGATGATCGCGTCGTAACCC	
rt_glnR1	GACGACGTACTGCTCGACAC	
rt_glnR2	TCGGCCTTCTCGGACTTATC	
RT_nirBfw	GTGCTCGCCAGCAGTCCGAGC	
RT_nirBrev	CCAGCCCCGCCTCCCGCGCCAG	
cobB1fw_Ndel	CATATGGCGCATGCGCCCCACTCTGAG	This work
cobB1rev_BamHI	GGATCCGGCCGTCGCCGCGTCCCCAC	
cobB2fw_Ndel	CATATGACCGGCAAGCCTCTCGTCGCC	
cobB2rev_BamHI	GGATCCGCCAGCCCCGCGCAGCAGCG	

Restriction enzymes sites are underlined.

amplification of the PCR product. The PCR products were separated during electrophoresis on 2% agarose gels. All reverse transcription/PCR reactions were carried out in triplicate using RNA isolated from three independent cultivations.

## Strep-Tagged GlnR Overexpression in *S. coelicolor* M145 and Purification

For Strep-GlnR purification, *S. coelicolor* M145 carrying pGM-Strep-*glnR* overexpression strain (Table 1) was grown for 4 days in complex S-medium at 30°C. Subsequently, cells were harvested and washed twice with nitrogen-free Evans medium. Washed cell biomass was transferred into Evans medium supplemented with 5 or 100 mM NH<sub>4</sub>Cl and 5 or 100 mM NaNO<sub>3</sub> as a sole nitrogen source, respectively. The expression of Strep-GlnR was induced with 12.5 µg/ml thiostrepton for 36 h. After cultivation cells were harvested and washed with a solution of 100 mM Tris and 150 mM NaCl (pH 8.0). In

order to prevent phosphatase and protease activity, 5 mM sodium fluoride and 5 mM orthovanadate and the EDTA-free cOmplete protease inhibitor cocktail (Roche) were added to the buffer. Cell lysis was performed by Emulsifex (Avestin, Ottawa, Canada) with three consecutive passages. Cell debris and insoluble proteins were separated from the soluble fraction by centrifugation (60 min, 14800 g, 4°C). Soluble proteins were loaded onto a pre-equilibrated Gravity flow Strep-Tactin® Sepharose® column for one-step purification of recombinant Strep-tag® proteins (1-ml bed volume; IBA, Germany). Strep-GlnR was competitively eluted using elution buffer supplemented with 2.5 mM desthiobiotin. Concentrated fractions containing the pure Strep-GlnR were stored at 4°C.

## Cloning, Overexpression, and Purification of the His-Tagged CobB1 and CobB2 in *E. coli* BL21

Oligonucleotide primers (Table 2) were designed to incorporate an NdeI and BamHI restriction sites into PCR-fragments containing *cobB1* (SCO0452) and *cobB2* (SCO6464) genes. The PCR products were cloned into pET15b (Novagen, UK) to generate plasmids pET15b-CobB1 and pET15b-CobB2 used to transform *E. coli* BL21 (DE3). The over-night cultures of *E. coli* BL21 pET15b-CobB1 and *E. coli* BL21 pET15b-CobB2 were used to inoculate 500 ml fresh LB containing 12.5 µg/ml chloramphenicol and 50 µg/ml ampicillin. At an OD<sub>578</sub> of 0.3, the cells were induced for the expression of His-CobB1 and His-CobB2 with 0.1 mM IPTG and further incubated at 30°C for 5–10 h. All subsequent procedures were performed at 4°C. Cells were harvested by centrifugation at 5000 rpm for 30 min and resuspended in lysis buffer [50 mM Tris-HCl buffer; pH 7.5; supplemented with the EDTA-free cOmplete protease inhibitor cocktail (Roche)]. Cell lysis was performed by Emulsifex (Avestin, Ottawa, Canada). Cell debris and insoluble proteins were separated from the soluble fraction by centrifugation (30 min, 14,800 g, 4°C). Soluble proteins were loaded onto a pre-equilibrated Ni<sup>2+</sup>-nitrilotriacetic acid (NTA)-agarose column (1-ml bed volume; IBA, Germany), and His-CobB1 and CobB2 were competitively eluted using elution buffer supplemented with 200 mM imidazole. Fractions containing the pure proteins were pooled and desalted using a HiTrap desalting column (GE Healthcare) with 20 mM Tris-HCl buffer. Concentrated fractions His-CobB1 and His-CobB2 were stored at 4°C.

## Western Blot Analysis

*S. coelicolor* M145 was grown, harvested and lysed as described above. The protein concentration was determined using a Nanodrop (PEQLAB, Erlangen, Germany). Proteins were separated by SDS-PAGE (Laemmli, 1970) and transferred to a nitrocellulose membrane (Roth, Karlsruhe, Germany) by semidry electroblotting (PEQLAB) using transfer buffer (25 mM Tris, 150 mM glycine, 20% methanol, pH 9.2) for 30 min at 400 mA. The membrane was blocked with TBST buffer with 5% BSA at room temperature for 1 h. Subsequently the membrane was incubated at room temperature for 1 h with rabbit anti-GlnR polyclonal antibodies (SEQLAB, Göttingen, Germany) or

overnight at 4°C with rabbit anti-N-acetyl lysine polyclonal antibodies (Cell Signaling TECHNOLOGY) in TBST buffer for GlnR detection (10 mM Tris pH 8, 150 mM NaCl, 0.05% Tween 20 supplemented with 2.5% of milk powder) or TBST buffer for N-acetyl-lysine detection (25 mM Tris pH 8.0, 125 mM NaCl, 0.1% Tween 20 supplemented with 3% of BSA), respectively. Membranes were washed with TBST (four times) for 5 min and the binding of the primary antibodies against GlnR protein or N-acetylated lysine was detected using anti-rabbit IgG horse-radish-peroxidase-conjugated antiserum (BIORAD, München, Germany) solved in TBST buffer. After 2 h of incubation at ambient temperature the binding of secondary antibodies against rabbit antibodies was detected using the ECL Western blotting detection system (GE Healthcare, München, Germany).

## Electrophoretic Mobility Shift Assay (EMSA)

DNA fragments containing promoter regions of GlnR target genes were amplified with Taq polymerase (GENAXXON bioscience, Germany) using genomic DNA from *S. coelicolor* M145. For this, genomic DNA was isolated with the NucleoSpin® Tissue Kit (Macherey-Nagel, Düren, Germany). Primer sequences, PCR and labeling conditions were carried out as reported in Tiffert et al. (2008). All EMSA reactions were carried out in EMSA buffer (100 mM Tris, 150 mM NaCl, 10 mM β-mercaptoethanol, pH 8) containing an excess of unlabeled nonspecific salmon sperm DNA. The Cy5-labeled target DNA (2 ng) and 4 µg of Strep-GlnR protein or 50 µg of the cleared cell lysate from *glnR* mutant were dissolved in EMSA buffer and incubated for 10 min at 24°C. After incubation, loading buffer (0.25 × TBE buffer and 60% glycerol) was added and the fragments were separated using gel electrophoresis on 2% TAE agarose gels. DNA bands were visualized by the fluorescence imaging using a Typhoon Trio+ Variable Mode Imager (GE Healthcare).

## In vitro Deacetylation of Strep-GlnR

Deacetylation of Strep-GlnR was performed in the presence of NAD<sup>+</sup> as a substrate for the deacetylase. For this reaction, 100 µg of Strep-GlnR in 50 mM Tris-HCl pH 8.3, were incubated with 1 mM NAD<sup>+</sup> for 6 h at 30°C in the presence and absence of 20 µg of His-CobB1 or His-CobB2. Acetylated and deacetylated Strep-GlnR samples were analyzed by immunoblotting using rabbit polyclonal anti-N-acetyl lysine antibodies (Cell Signaling TECHNOLOGY).

## Nano LC-MS/MS Analysis of the Purified Strep-GlnR

Purified Strep-GlnR was—either in solution or in gel—digested with trypsin (1:100 w/w) as described previously (Borchert et al., 2010). Ten percentage of the peptide mixtures in the resulting in-solution digests were directly analyzed by LC-MS/MS. Additionally, the tryptic peptides were subjected to titanium dioxide chromatography to enrich detection of the phosphorylated peptides. For phosphopeptide enrichment acetonitrile was added to the peptide mixture to a final concentration of 30% and the pH was adjusted to



2–3. Enrichment of phosphopeptides by titanium dioxide chromatography was done as described previously (Olsen and Macek, 2009) with the following modifications: phosphopeptide elution from the beads was performed three times with 100 ml of 40% ammonia hydroxide solution in 60% acetonitrile at a pH > 10.5.

For peptide analysis a Proxeon Easy-LC system (Proxeon Biosystems, Odense, Denmark) coupled to a LTQ-Orbitrap-XL (Thermo Fisher Scientific, Bremen, Germany) equipped with a nanoelectrospray ion source (Proxeon Biosystems) as described previously (Koch et al., 2011) was used. The five most intense precursor ions were fragmented by activation of neutral loss ions at −98, −49, and −32.6 relative to the precursor ion (multistage activation). Mass spectra were analyzed using the software suite MaxQuant, version 1.0.14.3 (Cox et al., 2009). The data were searched against a target-decoy *Streptomyces coelicolor* database containing 8154 forward protein sequences and 262 frequently observed contaminants. Trypsin was set as protease in which two missed cleavage sites were allowed. Beside acetylation at the N-terminus of lysine and oxidation of methionine, phosphorylation of serine, threonine, and tyrosine were set as variable modifications. Carbamidomethylation of cysteine was set as fixed modification. Initial precursor mass tolerance was set to 7 parts per million (ppm) at the precursor ion and 0.5 Da at the fragment ion level. Identified peptides were parsed using the identify module of MaxQuant and further processed for statistical validation of identified peptides, modified sites and protein groups. False discovery rates were set to 1% at peptide, modified site, and the protein group level. To assign a phosphorylation and acetylation site, respectively, to a specific residue a minimal reported localization probability of 0.75 was set as a threshold. The fragmentation spectra of potential modified peptides were manually validated for presence of phosphorylation and acetylation sites.

## RESULTS

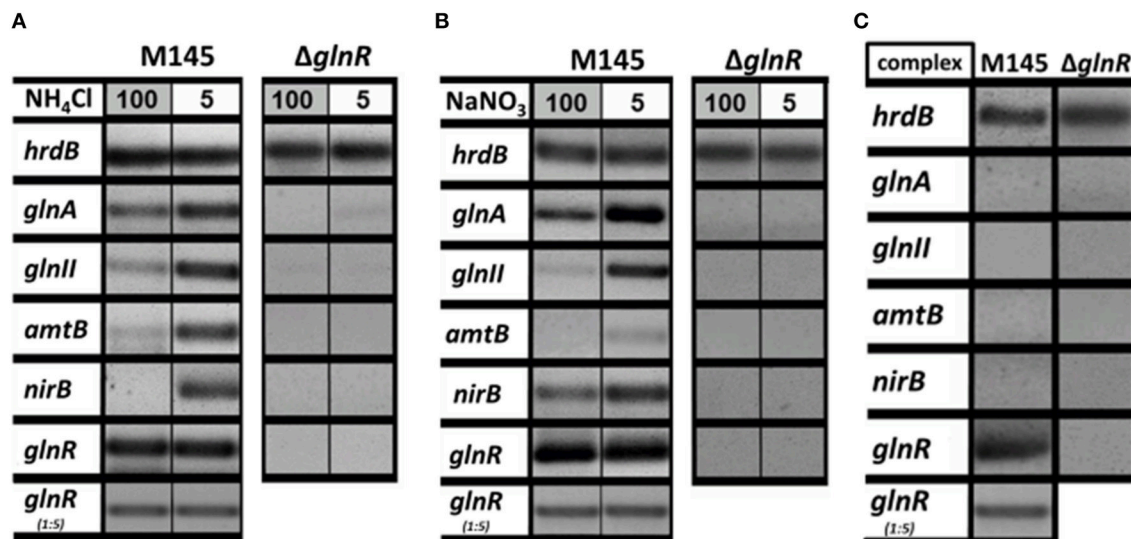
### Revisiting the Transcriptional Regulation of *glnR* And GlnR-Target Genes under Defined and Complex Nitrogen Conditions

The GlnR regulator controls genes related to N-assimilation including genes involved in the ammonium uptake (operon *amtB-glnK-glnD*), nitrate and nitrite assimilation (*nnar*, *nasA*, *nirB*, *narK*) and synthesis of the central metabolic nitrogen donors glutamine and glutamate (*glnA*, *glnII*, and *gdhA*; Tiffert et al., 2008; Amin et al., 2012). Although many genetic studies on the regulation of the N-assimilatory genes in *S. coelicolor* and other actinobacteria have been performed, the regulation of the GlnR activity itself is still enigmatic. In a first step, we aimed to show how *glnR* itself and selected GlnR target genes are regulated at the transcriptional level upon growth in four defined N-conditions as well as in a complex medium. For this purpose, RT-PCR was performed using total RNA isolated from *S. coelicolor* M145 and the *glnR* mutant. The strains were grown for 4 days in the complex S-medium to obtain high biomass. Cells were harvested by centrifugation and

washed with Evans medium to remove traces of S-medium. Subsequently the biomass was transferred into defined Evans medium containing different N-sources (ammonium chloride or sodium nitrate) at 100 mM (N-excess) or 5 mM (N-limitation) or into complex S-medium. All cultures were further cultivated for 24 h, at 30°C. Total RNA isolated from *S. coelicolor* M145 and the *glnR* mutant was used to generate cDNA. Subsequently, RT-PCR analysis using internal primers for *glnR* and selected GlnR target genes was performed. The *hrdB* (encoding the essential principal sigma factor of RNA polymerase) was used as an internal standard due to its relatively constant levels of expression throughout the growth (Buttner et al., 1990). All reverse transcription/PCR reactions were carried out in triplicate using RNA isolated from three independent cultures (for details see Section RT-PCR). For the transcriptional analysis the following GlnR target genes encoding proteins involved in ammonium and nitrate assimilation were selected: *glnA*, *glnII* (encoding glutamine synthetase GSI and GSII, respectively), *amtB* (encoding an ammonium transporter), and *nirB* (encoding a nitrate reductase). The transcriptional analysis confirmed that GlnR enhanced the expression of *glnA*, *glnII*, *amtB*, and *nirB* under low concentration of ammonium chloride, whereas high ammonium chloride concentration inhibited expression of these genes (Figure 1A). To verify whether this regulatory effect was ammonium chloride dependent, transcript levels for the selected GlnR target genes were also analyzed in the presence of sodium nitrate. Our results showed that expression of all tested genes *glnA*, *glnII*, *amtB*, and *nirB* was strongly induced in condition of low nitrate concentration in *S. coelicolor* M145 (Figure 1B). At last, transcriptional analysis revealed that the *glnA*, *glnII*, *amtB*, and *nirB* were not expressed in *S. coelicolor* M145 grown in S-medium (Figure 1C). Expression of *glnA*, *glnII*, *amtB*, and *nirB* was totally abolished in the *glnR* mutant under all tested conditions, indicating that GlnR was necessary for their expression. These results indicate that expression of the selected GlnR target genes was strictly regulated by GlnR whose activity seemed to be modulated according to the N-status of the cell. Transcriptional analysis showed that the *glnR* transcript was present under all tested conditions, demonstrating that *glnR* is not regulated at the transcriptional level in the response to changing N-concentrations. These analyses led us to assume that GlnR regulatory activity might be modulated by post-translational modifications.

### The GlnR Protein is Present in the Cell under Both N-Limited and N-Proficient Conditions

In order to verify whether the GlnR protein was present (as its transcript) in all tested conditions, Western blot analysis was performed. As a control cell lysates from the *glnR* mutant as well as purified Strep-GlnR protein were used. For this, *S. coelicolor* M145 was grown in S-medium for 4 days at 30°C; cells were washed twice with Evans medium to remove traces of the complex S-medium. The biomass was transferred into defined Evans medium and further cultivated for 36 h at 30°C.

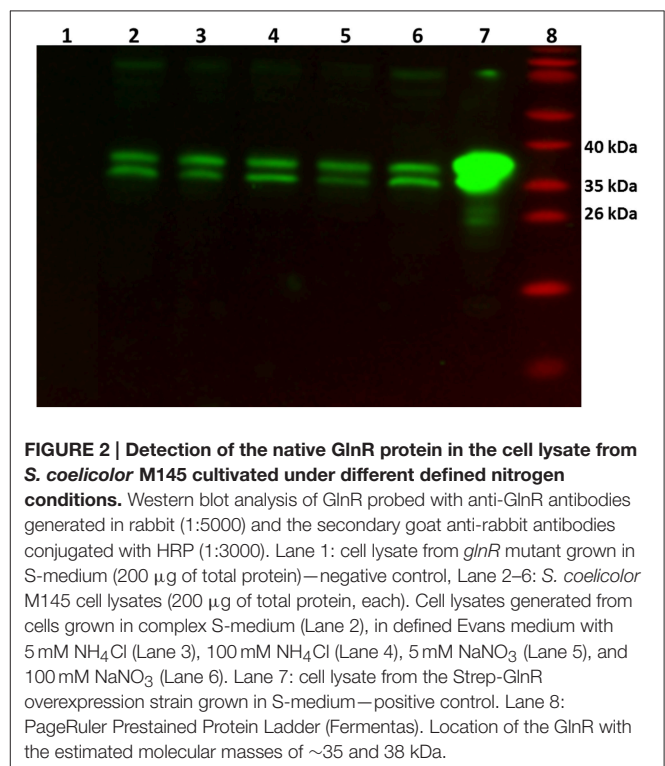


**FIGURE 1 | Transcriptional analysis of the selected GlnR target genes under variable nitrogen and carbon conditions. (A)** RT-PCR of *glnA*, *glnII*, *amtB*, *nirB*, and *glnR* in *S. coelicolor* M145, and *glnR* mutant cultivated in defined Evans medium with low (5 mM) or high (100 mM) concentrations of ammonium chloride as sole nitrogen source. **(B)** RT-PCR of *glnA*, *glnII*, *amtB*, *nirB*, and *glnR* in *S. coelicolor* M145 and *glnR* mutant cultivated in defined Evans medium with low (5 mM) or high (100 mM) sodium nitrate as sole nitrogen source. **(C)** RT-PCR of *glnA*, *glnII*, *amtB*, *nirB*, and *glnR* in *S. coelicolor* M145 and *glnR* mutant cultivated under complex nitrogen rich conditions (S-medium). Total RNA was isolated from mycelium harvested after 24 h of cultivation.

*S. coelicolor* M145 cells from S-medium and Evans medium were harvested and disrupted. As a negative control cell lysate generated from the *glnR* mutant grown in S-medium was used. As a positive control Strep-GlnR overexpressed and purified from the *S. coelicolor* M145 grown in S-medium was used. Clarified cell lysates (200  $\mu$ g of total protein) as well as isolated Strep-GlnR protein (20  $\mu$ g) were run on a 12.5% SDS polyacrylamide gel and subsequently transferred onto a nitrocellulose membrane. Signals for the native GlnR and Strep-GlnR were detected by using anti-GlnR antibodies. Western blot analysis revealed the presence of native GlnR in *S. coelicolor* M145 under all tested conditions, whereas no signal was detected in the cell lysate from the *glnR* mutant. No sign of GlnR proteolysis could be detected under N-limitation or proficiency. The predicted size of the GlnR protein calculated from the amino acid sequence is 29.8 kDa. Interestingly, GlnR appeared as a double band with the estimated size of  $\sim$ 35 and 38 kDa (Figure 2). This indicated that this regulator may undergo post-translational modification.

### Only GlnR Interacts with GlnR-Target Genes

In order to determine whether other regulators besides GlnR other regulators are able to recognize and interact with selected promoters of the GlnR-target genes, comparative EMSAs were performed using cell lysates from the *glnR* mutant and Cy5-labeled promoter regions of  $P_{glnA}$ ,  $P_{amtB}$ , and  $P_{glnII}$ . To do so, the *glnR* mutant was grown in S-medium for 4 days at 30°C; cells were washed twice with Evans medium, transferred into Evans medium or directly transferred in S-medium and further cultivated for 36 h at 30°C. Cells were harvested, disrupted and the cell lysates (50  $\mu$ g of total protein) were



**FIGURE 2 | Detection of the native GlnR protein in the cell lysate from *S. coelicolor* M145 cultivated under different defined nitrogen conditions.** Western blot analysis of GlnR probed with anti-GlnR antibodies generated in rabbit (1:5000) and the secondary goat anti-rabbit antibodies conjugated with HRP (1:3000). Lane 1: cell lysate from *glnR* mutant grown in S-medium (200  $\mu$ g of total protein)—negative control, Lane 2–6: *S. coelicolor* M145 cell lysates (200  $\mu$ g of total protein, each). Cell lysates generated from cells grown in complex S-medium (Lane 2), in defined Evans medium with 5 mM  $\text{NH}_4\text{Cl}$  (Lane 3), 100 mM  $\text{NH}_4\text{Cl}$  (Lane 4), 5 mM  $\text{NaNO}_3$  (Lane 5), and 100 mM  $\text{NaNO}_3$  (Lane 6). Lane 7: cell lysate from the Strep-GlnR overexpression strain grown in S-medium—positive control. Lane 8: PageRuler Prestained Protein Ladder (Fermentas). Location of the GlnR with the estimated molecular masses of  $\sim$ 35 and 38 kDa.

used for EMSAs. No shifts were observed with cell lysates generated from the *glnR* mutant (Figure S1), indicating that solely GlnR was able to recognize tested promoters under studied conditions.

## Different GlnR Post-translational Modification Patterns Were Detected under Complex *N*-Rich Conditions and *N*-Defined Conditions

Post-translational modifications of GlnR were identified by LC-MS/MS using Strep-GlnR overexpressed and isolated from *S. coelicolor* M145 grown in the complex and *N*-rich S-medium. The purified Strep-GlnR samples were run on a 12.5% SDS polyacrylamide gel and subsequently stained with Coomassie blue. Finally, the band corresponding to Strep-GlnR was cut out from the gel and subjected to proteolytic digestion with trypsin. In addition to a direct measurement of tryptic Strep-GlnR peptides, phosphopeptides were selectively enriched using titanium dioxide affinity chromatography prior to LC-MS/MS analysis. To specifically detect peptides carrying phosphorylated and acetylated residues, the LC-MS/MS spectra of the modified peptides were compared to the intensities of spectra of the corresponding non-modified peptides. Phosphorylation and acetylation sites were considered as resulting from high confidence phosphorylation and acetylation events only if the LP was higher to 0.75 (or equal), the PEP score was lower than 0.01 (or equal) and the Mascot score higher than 39 (or equal; for details of this evaluation see Materials and Methods, in the Section Nano LC-MS/MS Analysis of the Purified Strep-GlnR). Mapping of the post-translational modifications on the Strep-GlnR purified from *S. coelicolor* M145 grown in this medium revealed peptides carrying phosphorylated serine and threonine residues: Ser 133, Thr 138, Ser 207, Thr 211, Thr 256, Ser 264/265, and only one acetylated residue, Lys 142 (Table 3 and Supplementary Material Data Sheets 1–4).

Post-translational modifications of GlnR under defined *N*-conditions were identified by LC-MS/MS using Strep-GlnR isolated from *S. coelicolor* M145 grown in Evans media supplemented with either 100 mM NaNO<sub>3</sub> (nitrate proficiency) or 5 mM NaNO<sub>3</sub> (nitrate limitation) as a sole *N*-source. Preparation of the Strep-GlnR samples, LC-MS/MS analysis and data processing were performed as stated in the Section Different GlnR Post-translational Modification Patterns Were Detected under Complex *N*-Rich Conditions and *N*-Defined Conditions. Strep-GlnR isolated from cells grown under nitrate proficient conditions revealed only two phosphorylated residues: Ser 140 and Thr 256, whereas no serine/threonine phosphorylation was detected in Strep-GlnR isolated from cells grown under nitrate limitation. Interestingly, Strep-GlnR isolated from *S. coelicolor* M145 cultivated in condition of nitrate proficiency or limitation exhibited identical acetylation pattern (Table 3 and Supplementary Material Data Sheets 1–4). The following lysine residues were acetylated in Strep-GlnR independent of the nitrate concentration in the Evans medium: Lys 142, Lys 153, Lys 159, and Lys 200, whereas only Lys142 was acetylated when Strep-GlnR originated from culture in S-medium. The Strep-GlnR isolated from cells grown in defined Evans medium supplemented with nitrate as a sole *N*-source revealed higher acetylation level than the Strep-GlnR from complex S-medium. Most of the acetylated and phosphorylated GlnR residues

were localized within the helix turn helix motif involved in the DNA recognition and binding. These modifications are thus likely to have an impact on the GlnR DNA-binding activity.

## Serine/Threonine Phosphorylation Influenced the DNA-Binding Activity of GlnR

GlnR was previously shown to specifically bind numerous promoters of genes that encode proteins involved in *N*-assimilation in *S. coelicolor* M145. GlnR binding boxes were defined and localized in the promoter regions of *glnA*, *glnII*, *amtB*, *nirB*, and other genes encoding proteins involved in *N*-metabolism (Tiffert et al., 2008; Amin et al., 2012). In order to determine whether the differently modified GlnR protein isolated from *S. coelicolor* M145 could recognize and interact with the selected target promoters comparative electrophoretic mobility shift assays (EMSAs) were performed using the Cy5-labeled promoter regions P<sub>glnA</sub>, P<sub>amtB</sub>, P<sub>glnII</sub>, P<sub>nirB</sub>. Three different modified forms of Strep-GlnR isolated from *S. coelicolor* M145 grown either under complex nitrogen rich conditions (Strep-GlnR<sub>N++</sub>) or nitrogen defined conditions (nitrate limited—Strep-GlnR<sub>N–</sub>; nitrate excess—Strep-GlnR<sub>N+</sub>) were used. Post-translational modifications of Strep-GlnR<sub>N++</sub>, Strep-GlnR<sub>N–</sub>, and Strep-GlnR<sub>N+</sub> were confirmed by LC-MS/MS prior to EMSAs analysis. Multiply phosphorylated Strep-GlnR<sub>N++</sub> was not able to interact with the promoter regions P<sub>glnA</sub>, P<sub>amtB</sub>, P<sub>glnII</sub>, and P<sub>nirB</sub> indicating that phosphorylation inhibits the binding of Strep-GlnR<sub>N++</sub> to the target DNA. Interestingly, multiply acetylated GlnR (Strep-GlnR<sub>N–</sub>) was able to interact with and to shift all tested promoter regions. Finally, the multiply acetylated Strep-GlnR<sub>N+</sub> also phosphorylated on the Ser 140 and Thr 256 residues, generated diffuse shifts with P<sub>glnA</sub>, P<sub>amtB</sub>, P<sub>glnII</sub>, and P<sub>nirB</sub>, suggesting different binding of the GlnR or instability of the GlnR-DNA complex (Figure 3). The Ser/Thr phosphorylations altered the *in vitro* DNA-binding activity of GlnR purified from cells grown under *N*-excess conditions while the multiple acetylations did not inhibit the formation of GlnR-DNA complexes under *N*-limiting conditions. These results are consistent with the transcriptional analysis of the GlnR-target genes under *N*-limited and excess conditions reported in *Results Revisiting the Transcriptional Regulation of glnR and GlnR-Target Genes under Defined and Complex Nitrogen Conditions*. High induction of the expression of GlnR target genes under nitrogen limitation is thus achieved by the unphosphorylated GlnR as shown by the LC-MS/MS analysis.

## Impact of the Acetylation of GlnR on Its DNA-Binding Activity

Interestingly, the Ser/Thr phosphorylation of GlnR occurs in condition of *N*-proficiency whereas acetylation is independent of the *N*-status of the cell. To study the impact of the acetylation on the GlnR DNA-binding activity, we attempted to remove the acetyl groups from GlnR by an enzymatic deacetylation. *S. coelicolor* M145 possess two genes (SCO0452 and SCO6464) encoding enzymes annotated as a sirtuin-like (deacetylase-like).



**TABLE 3 |** Phosphorylated and acetylated GlnR peptides detected by LC-MS/MS under complex and defined *N*-conditions.

Modified GlnR peptides (nitrogen rich, complex S-medium)	Phosphosite (ph) and acetylation site (ac)	LP	PEP	MASCOT
NGDLSVDEATY <b>SAK</b>	(ph) Ser 133	0.999	0.000	49.52
NGDLSVDEATY <b>SAK</b>	(ph) Thr 138	0.999	0.000	49.52
LGPEHESLIGTV <b>R</b>	(ph) Ser 207	0.970	0.000	44.29
LGPEHESLIGTV <b>R</b>	(ph) Thr 211	0.999	0.000	44.29
AAAETNEAAGAR	(ph) Thr 256	1.000	0.000	37.80
AAAETNEAAGAR <b>SSKV</b>	(ph) Ser 264/265	0.979/0.999	0.000	37.80
NGDLSVDEATY <b>SAK</b>	(ac) Lys 142	1.000	0.000	60.27
<b>MODIFIED GlnR PEPTIDES (DEFINED EVANS MEDIUM WITH 100 mM NaNO<sub>3</sub>)</b>				
NGDLSVDEATY <b>SAK</b>	(ph) Ser 140	0.999	0.000	54.70
AAAETNEAAGAR	(ph) Thr 256	1.000	0.001	42.16
NGDLSVDEATY <b>SAK</b>	(ac) Lys 142	0.999	0.000	66.18
VLDLTF <b>K</b> EFELLK	(ac) Lys 153	1.000	0.000	86.48
VLDLTFKEFELL <b>K</b> YLAQHGPGR	(ac) Lys 159	1.000	0.000	68.96
AKLGPEHESLIGTV <b>R</b>	(ac) Lys 200	1.000	0.000	86.58
<b>MODIFIED GlnR PEPTIDES (DEFINED EVANS MEDIUM WITH 5 mM NaNO<sub>3</sub>)</b>				
NGDLSVDEATY <b>SAK</b>	(ac) Lys 142	0.999	0.000	103.18
VLDLTF <b>K</b> EFELLK	(ac) Lys 153	1.000	0.000	76.10
VLDLTFKEFELL <b>K</b> YLAQHGPGR	(ac) Lys 159	1.000	0.000	76.10
AKLGPEHESLIGTV <b>R</b>	(ac) Lys 200	1.000	0.000	86.56

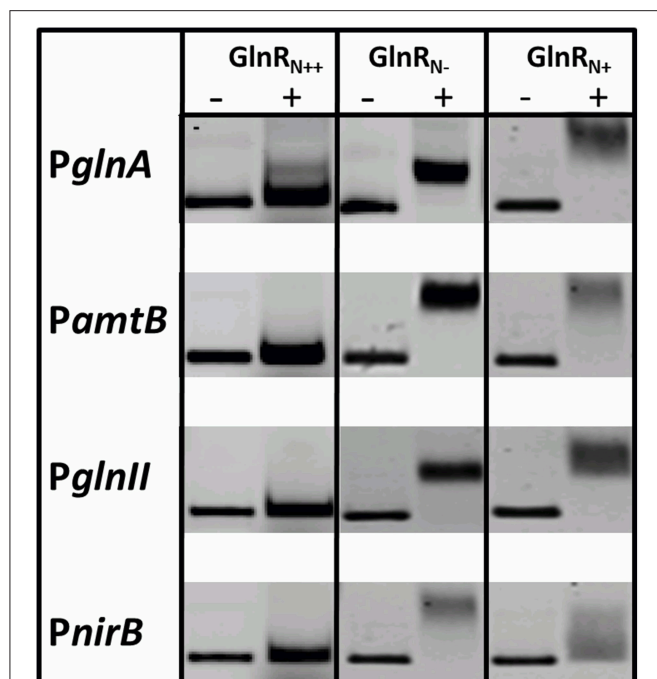
Acetylated or phosphorylated residues are in bold.

SCO0452 named CobB1 is most similar to human sirtuin SIRT4 and was functionally characterized as NAD<sup>+</sup>-dependent deacetylase from *S. coelicolor* (Mikulik et al., 2012). The SCO6464 designated as CobB2, shares significant homology with human SIRT5 (Moore et al., 2012). The CobB2 homolog in *S. erythraea* (with 68% similarity on the protein level) was shown to catalyze deacetylation of acetyl-CoA synthetase AcsA *in vitro* (You et al., 2014). Bacterial sirtuins are able to deacetylate a large number of target proteins (Castaño-Cerezo et al., 2014) and shows no preference for enzymatic and nonenzymatic lysine acetylation substrate sites (AbouElfetouh et al., 2014). We therefore assumed that the CobB1 and CobB2 deacetylases from *S. coelicolor* might be potentially able to deacetylate GlnR. Overexpression of *N*-terminally His<sub>6</sub>-tagged CobB1 and CobB2 (His-CobB1 and His-CobB2) was achieved with the IPTG-inducible system in *E. coli* BL21. Acetylated Strep-GlnR, isolated from *S. coelicolor* M145 grown under *N*-limiting conditions was used as a substrate for the *in vitro* deacetylation assays. A successful GlnR deacetylation was proved by Western blot analysis. Signals for acetylated Strep-GlnR (Ac+) protein and deacetylated Strep-GlnR (Ac-) were detected by anti-GlnR antibodies and anti-*N*-acetyl lysine antibodies. Western blot analysis revealed that the deacetylase CobB2 was able to remove the acetyl groups from Strep-GlnR in the *in vitro* assay (Figure S2). EMSAs performed with double-stranded Cy5-labeled selected promoter regions of *P<sub>glnA</sub>*, *P<sub>amtB</sub>*, *P<sub>glnII</sub>*, and *P<sub>nirB</sub>* showed that both the Strep-GlnR (Ac-) as well as the Strep-GlnR (Ac+) were able to interact with the target DNA (Figure 4). However, the Strep-GlnR (Ac+) generated more retarded shift than did the deacetylated Strep-GlnR (Ac-), suggesting that acetylation changes the DNA-binding affinity of GlnR.

## DISCUSSION

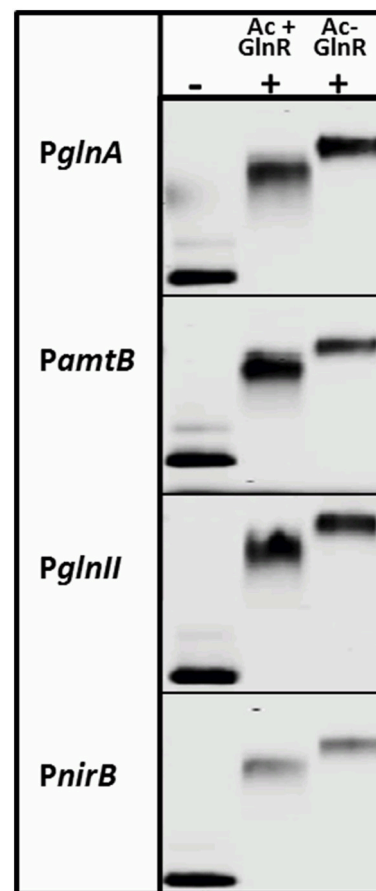
Members of the family of *Streptomyetaceae* are well-known antibiotic producers. The production of antimicrobial compounds ensured the adaptation of these bacteria to a wide variety of ecological niches and successful competition with other microorganisms for space and resources in their soil habitat. Depending on the soil type and seasonal changes, soil may exhibit a high diversity in nutrients availability ranging from nutrient-poor to nutrient-rich conditions. *Streptomyces* are able to adjust rapidly to these changing conditions. To do so, sensing, and responding to changes in a nutrient availability and subsequently coordination of the metabolic switch is necessary. The global nitrogen response regulator GlnR in *S. coelicolor* controls genes related to *N*-metabolism and ensures a dynamic and fast response under fluctuating *N*-conditions (Tiffert et al., 2008; Amin et al., 2012). Our transcriptional analysis revealed strong expression of *glnR* under all tested conditions. This indicates that *glnR* expression is not regulated at the transcriptional level in the response to the *N*-availability in *S. coelicolor* M145 as in other members of *Actinomycetales* such as: *M. smegmatis* (Amin et al., 2008; Petridis et al., 2015), *A. mediterranei* U32 (Wang et al., 2014) and *Microbispora* ATCC-PTA-5024 (to be published). The transcriptional regulation of the *glnR* in *S. coelicolor* and *M. smegmatis* is probably not achieved by GlnR itself, since GlnR binding boxes were not detected in the *glnR* promoter region (Tiffert et al., 2008; Jenkins et al., 2013). In contrast, GlnR-self-regulation at the transcriptional level was reported in *S. erythraea* (Yao et al., 2014). Despite its unchanged expression, GlnR controls the expression of its target genes in response to *N*-availability in *S. coelicolor* as well as in





**FIGURE 3 | EMSAs demonstrating different binding behavior of Strep-GlnR depending on differential modification patterns.** Lane (–), control, promoter regions of the known GlnR target genes (Cy5—labeled probe without Strep-GlnR protein). Lane (+), Cy5—labeled probe with 4  $\mu$ g of the Strep-GlnR<sub>N++</sub> from complex, rich condition (phosphorylated on the six Ser/Thr residues and acetylated on the one Lys residue), Strep-GlnR<sub>N–</sub> from nitrate limited conditions (acetylated on four Lys residues), and Strep-GlnR<sub>N+</sub> from the nitrate excess conditions (acetylated on four Lys residues and additionally phosphorylated on two Ser/Thr residues). The differential modification pattern of Strep-GlnR was confirmed by LC-MS/MS prior EMSAs. EMSAs were performed in the presence of the 300  $\times$  excess of the unlabeled, unspecific salmon sperm DNA.

*M. smegmatis*, *A. mediterranei*, and *S. erythraea* (Tiffert et al., 2008; Jenkins et al., 2013; Yao et al., 2014). Indeed, in *S. coelicolor* expression of *glnA*, *glnII*, *amtB*, and *nirB* was totally abolished in the *glnR* mutant but it was enhanced in condition of nitrate and ammonium limitation and reduced or completely abolished in condition of *N*-proficiency. The *nirB*, that is expressed at the similar level in the presence of high and low *N*, escapes this general rule likely because GlnR cooperate with NnaR (transcriptional regulator for nitrate/nitrite assimilatory genes, GlnR-target) to regulate *nirB* expression in the presence of nitrate (Amin et al., 2012). These findings led us to assume that GlnR undergoes post-translational modifications that might alter its DNA-binding ability in a response to *N*-availability. Our LC-MS/MS analysis revealed that GlnR was post-translational modified by Ser/Thr phosphorylation and acetylation on Lys residues. Such modifications have not been reported for GlnR in *Actinomycetales* so far. Phosphorylated GlnR was detected in cells grown in the *N*-rich Evans medium and complex S-medium, demonstrating that the phosphorylation of the GlnR regulator is associated with *N*-excess. In contrast, the unphosphorylated form of GlnR was detected by the LC-MS/MS analysis only under *N*-limited conditions. Lack of the phosphorylation on



**FIGURE 4 | EMSAs demonstrating different binding behavior of the acetylated Strep-GlnR and *in vitro* deacetylated Strep-GlnR.** Lane (–), control, promoter regions of the known GlnR target genes (Cy5—labeled probe without Strep-GlnR protein). Lane (+), Cy5—labeled probe with 4  $\mu$ g of the acetylated Strep-GlnR (Ac+) or deacetylated Strep-GlnR (Ac–). The loss of the acetylation on Strep-GlnR was confirmed by Western blot analysis prior EMSAs. EMSAs were performed in the presence of the 300  $\times$  excess of the unlabeled, unspecific salmon sperm DNA.

the Asp 50 and lack of any Ser/Thr phosphorylation in GlnR isolated from *S. coelicolor* grown under *N*-limiting conditions was also demonstrated by Lin et al. (2014), in agreement with our results.

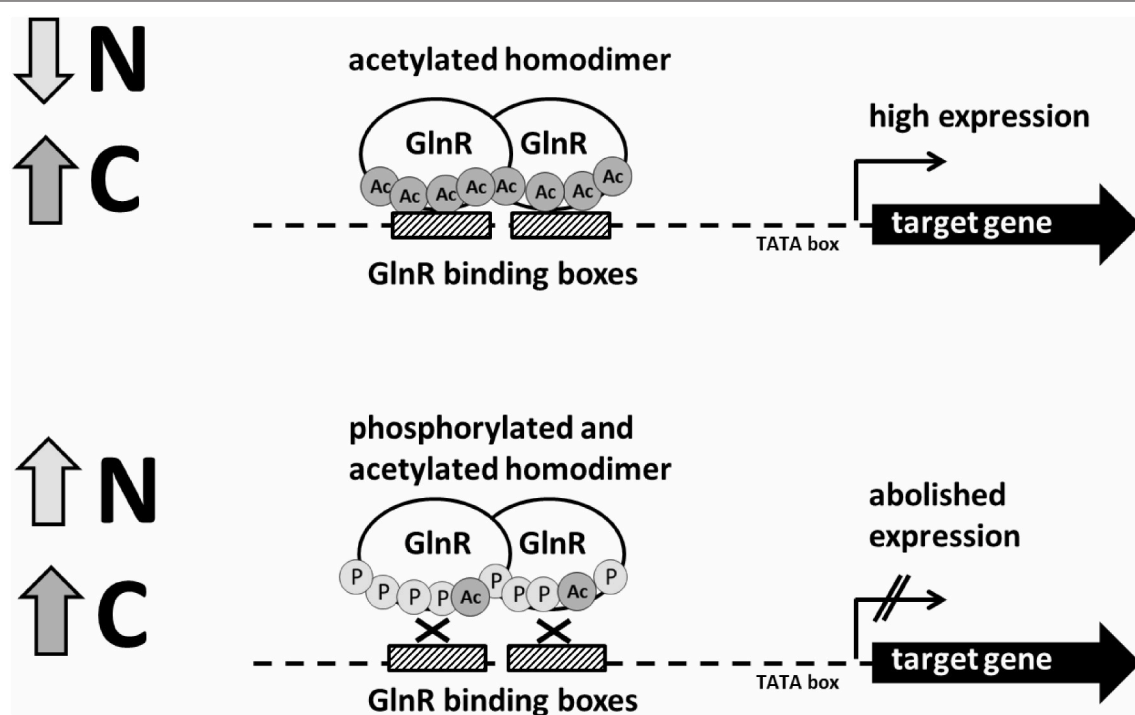
It has long been thought that signal transduction systems in bacteria relied solely on histidine/aspartate phosphorylation, while signal transduction systems based on serine/threonine phosphorylation and *N*-lysine acetylation were restricted to eukaryotes. However, with the increasing availability of phosphoproteomic and acetylproteomic data, a number of proteins modified by acetylation and phosphorylation have been detected in bacteria (Soufi et al., 2012; Cain et al., 2014). These post-translational modifications (PTMs) are known to influence changes in the profile of the bacterial transcriptome and proteome. PTMs have significant influence on the protein charge, size, hydrophobicity, and conformation. Therefore, PTMs can alter the activity, stability or cellular location and/or

affinity of the modified protein for its binding partners (Hu et al., 2010; Jones and O'Connor, 2011). Large number of proteins phosphorylated on serine and/or threonine residues was identified in *S. coelicolor* (Parker et al., 2010; Manteca et al., 2011), *M. tuberculosis* (Prisic et al., 2010), and other bacteria (see Cain et al., 2014 for review). Some bacterial transcription regulators were reported to be phosphorylated on serine and threonine residues (see Kalantari, 2015 for review). For instance, the post-translational serine/threonine phosphorylation was reported for EmrR, the transcriptional activator of arabinan biosynthesis genes in *Mycobacterium tuberculosis* (Molle et al., 2003; Sharma et al., 2006) and AfsR, the transcriptional activator of *afsS* involved in the regulation of secondary metabolism in *Streptomyces coelicolor* (Sawai et al., 2004).

Alignment of the amino acid sequences of GlnR from *S. coelicolor*, *A. mediterranei*, and *M. smegmatis* showed that residues corresponding to the phosphorylated Thr 138, Ser 140, and Thr 211 were conserved (Figure S3). Therefore, one can predict that these GlnR residues could be also phosphorylated under *N*-excess conditions in *M. smegmatis* and *A. mediterranei*. Since the GlnR structure was only partially elucidated (only *N*-terminal receiver domain), superimposition of the C-terminal GlnR response domain model on the crystal structure of the DNA-bound response regulator PhoP from *M. tuberculosis* (PDB: 5ed4; He et al., 2016) was performed. This comparative analysis revealed that Thr 138 and Ser 140 correspond to the Thr162 and

Glu164 residues in the PhoP crystal structure (He et al., 2016), respectively.

These residues are involved in the hydrophobic interactions stabilizing the PhoP dimer (He et al., 2016). Thus, one can assume that phosphorylation of the corresponding residues in GlnR could influence GlnR dimer formation. The Thr 211 residue from GlnR corresponds to the Thr235 residue in the PhoP structure. The side chain of the Thr 235 residue forms hydrogen bonds with the phosphate in the DNA thereby supporting the binding of the PhoP to the minor groove of the DNA (He et al., 2016). Again one can assume that phosphorylation of the corresponding Thr 211 residue in GlnR could influence its DNA binding affinity. However, solving the full GlnR crystal structure is necessary to achieve a detailed analysis of the influence of the modified residues on GlnR conformation DNA binding ability. Even though, the Ser/Thr phosphorylation has never been reported for GlnR, the post-translational serine/threonine phosphorylation is not uncommon in *S. coelicolor*. Forty proteins involved in gene regulation, central metabolism, protein biosynthesis, membrane transport, cell division, sporulation, and morphological differentiation were reported to be phosphorylated on serine and threonine residues in *S. coelicolor* (Parker et al., 2010; Manteca et al., 2011; Ladwig et al., 2015). The phosphoproteomic studies did not report phosphorylation of GlnR, presumably due to different cultivation conditions used (R5 medium and solid GYM medium Parker et al., 2010; Manteca et al., 2011, respectively).



**FIGURE 5 |** Schematic model of the regulation of the GlnR-target genes under *N*-rich and *N*-limited conditions by the differentially modified GlnR regulator.

Lysine acetylation of proteins was first discovered in eukaryotes and best characterized for histones (Waterborg, 2001) and eukaryotic transcription factors (Boyes et al., 1998; Marzio et al., 2000; Furia et al., 2002). Detailed analysis of acetylomes of *E. coli*, *S. enterica*, and *M. pneumoniae* gave a first wide view on this post-translational modification in bacteria (Yu and Auwerx, 2009; Zhang et al., 2009; Wang et al., 2010; Noort et al., 2012). Recently, the lysine acetylation was also reported for actinobacteria such as: *S. roseosporus* (Liao et al., 2015), *M. tuberculosis* (Liu et al., 2014), and *S. erythraea* (Huang et al., 2015). However, the physiological meaning of lysine acetylation was described only for a few prokaryotic proteins. For example, acetylation blocks the enzymatic activity of the acetyl-coenzyme A synthetase in *Salmonella enterica* (Starai et al., 2002), *Bacillus subtilis* (Gardner et al., 2006), *Rhodopseudomonas palustris* (Crosby et al., 2010), *Mycobacterium smegmatis* (Xu et al., 2011), and *S. coelicolor* (Mikulik et al., 2012), demonstrating conserved regulatory mechanism. The acetylation dependent modulation of the DNA-binding activity of the transcriptional regulator was reported for the capsule and flagellum biosynthesis regulator RcsA in *E. coli* (Thao et al., 2010). Multiply acetylation was also shown to inhibit the protein-protein interactions between the CheY (chemotaxis response regulator) and its target proteins in *E. coli* (Liarzi et al., 2010).

Both serine/threonine phosphorylation and lysine acetylation are conserved throughout evolution in all three life kingdoms: prokaryotes, archaea, and eukaryotes (Kennelly, 2002; Choudhary et al., 2009; Thao and Escalante-Semerena, 2011). Many proteins involved in central metabolic processes (synthesis of acetyl-CoA, glycolysis, gluconeogenesis, the TCA cycle and the glyoxylate bypass, glycogen biosynthesis, amino acid biosynthesis, fatty acid metabolism, and urea detoxification) were reported to be *N*-acetylated on lysine residues (Kim et al., 2006; Yu and Auwerx, 2009; Zhang et al., 2009; Wang et al., 2010). Recent reports on acetylation of proteins in bacteria demonstrated its role in the physiological adaptations to changes in carbon nutrient availability as reported for *B. subtilis* (Kosono et al., 2015), *S. enterica* (Wang et al., 2010), and *E. coli* (Weinert et al., 2013). The similar acetylation pattern of GlnR under defined conditions detected by LC-MS/MS was independent of the nitrogen source concentration, demonstrating that acetylation apparently does not reflect the nitrogen status of the cell but might signal carbon source availability as in other microorganisms. Comparative analysis of the amino acid sequences of 37 GlnR homologs from *Streptomyces* sp., revealed overall highly conserved sequence except last 30 residues from the C-terminus. Four acetyltable lysine residues Lys 142, Lys 153, Lys 159, and Lys 200 residues were conserved in all GlnR homologs (Figure S4). Superimposing of the GlnR response domain on the DNA-bound PhoP revealed that residues corresponding to acetylated Lys 153 and Lys200 residues are located within the conserved  $\alpha$ 8 helix involved in the recognition and binding of the target DNA. For instance, the Lys 153 residue in GlnR corresponds to the Thr 177 residue in PhoP (He et al., 2016). The Thr 177 forms a hydrogen bond with the phosphate from the major groove of the PhoP target DNA (He et al., 2016). This interaction (including other interactions resulting from  $\alpha$ 8 helix) contributes to the binding

affinity and influence sequence-specific interactions by changing the conformation of the PhoP, target-DNA, or both (He et al., 2016). Therefore, one could imagine that the acetylation of the corresponding residue Lys 153 in GlnR changes the positive charge of Lys 153 into neutral charge and thus may influence the GlnR DNA-binding affinity. The Lys 200 residue in GlnR corresponds to the Lys 224 residue in PhoP that is located in the C-terminus of the  $\alpha$ 8 helix close to the Arg 222 and Arg 223 both involved in the interactions with phosphates in the major groove of the DNA. However, these structural comparisons could be seen as speculative and solving the DNA-bound GlnR structure is necessary to determine how modified/unmodified residues participate in the GlnR dimer formation and its DNA-binding ability.

So far, EMSAs studies were carried out with GlnR from *A. mediterranei* (Wang Y. et al., 2013), *S. erythraea* (Yao et al., 2014), *M. smegmatis* (Jenkins et al., 2013), and *M. tuberculosis* (Malm et al., 2009) overexpressed and purified from *E. coli* and not from the original host. Our study is the first report on changes in the GlnR post-translational modifications and DNA-binding ability in its original host grown under biologically relevant conditions. We could demonstrate that both Ser/Thr phosphorylation and Lys acetylation influenced the DNA-binding activity of GlnR *in vitro*. The influence of the post-translational modifications on the GlnR regulatory function is summarized in the GlnR regulatory model (Figure 5). Bacterial protein modification by acetylation appears to be as common as phosphorylation and often both modifications can be detected on the same protein, suggesting possible cross-talk between these post-translational modifications that might signal C and N availability (Soufi et al., 2012). Different combinations in the acetylation and phosphorylation events open multitude of possibilities for the regulation of the GlnR regulon depending on the concentration of N- but also likely C-source in the environment. Our work provides a basis for further studies and detailed analysis of the mechanisms by which *S. coelicolor* senses and responds to changes in nutrient availability and how does GlnR coordinate the regulation of the N-related genes to govern the metabolic switch thereby guaranteeing cell homeostasis. In this respect, future work will focus on the identification of the Ser/Thr kinase(s) and acetyltransferase(s) involved in the post-translational modifications of GlnR.

## AUTHOR CONTRIBUTIONS

MHE and MM performed the RT-PCR analysis. RA, YT and MHE performed the EMSAs. YA and SK performed Western blot. RA and MJ overexpressed and purified Strep-GlnR for the LC-MS/MS analysis. MF performed the LC-MS/MS analysis, collected and processed the LC-MS/MS data. JM constructed the His-CobB1 and His-CobB2 overexpression strains and established protein purification method. AM overexpressed and purified His-CobB1 and His-CobB2 necessary for the deacetylation assay and MHE and MM performed deacetylation assay. MHI performed the bioinformatic analysis. NO was involved in the technical assistance. AB formulated the original

problem and provided direction and guidance as well as designed the study and developed the methodology. WW and BM provided helpful feedback on an early draft of the paper and assisted with data analysis. AB contributed to the writing of the manuscript and resolved final approval of the version to be published.

## FUNDING

RA was funded by a scholarship provided by the Higher Education Commission Pakistan in collaborations with Dow University of Health Sciences Karachi, Pakistan. YT was funded by “Studienstiftung des deutschen Volkes.” YA was jointly funded by the Higher Education Ministry of Damascus University, Syria and the German Academic Exchange Service (DAAD). This work

was supported by the University of Tuebingen “Projektförderung für NachwuchswissenschaftlerInnen 2012–2013.” SK is member of the DFG Research Training Group GRK1708. MHI was funded by the grant from the SYSTERACT project (ERASysAPP).

## ACKNOWLEDGMENTS

We wish to thank Johannes Madlung (Proteom Center, Tübingen) for assistance with the LC-ESI MS/MS.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmolb.2016.00038>

## REFERENCES

- Abouelfetouh, A., Kuhn, M. L., Hu, L. I., Scholle, M. D., Sorensen, D. J., Sahu, A. K., et al. (2014). The *E. coli* sirtuin CobB shows no preference for enzymatic and nonenzymatic lysine acetylation substrate sites. *Microbiologyopen* 4, 66–83. doi: 10.1002/mbio.3.223
- Amin, R., Reuther, J., Bera, A., Wohlleben, W., and Mast, Y. (2012). A novel GlnR target gene, *nnar*, is involved in nitrate/nitrite assimilation in *Streptomyces coelicolor*. *Microbiology* 158, 1172–1182. doi: 10.1099/mic.0.054817-0
- Amon, J., Bräu, T., Grimrath, A., Hanssler, E., Hasselt, K., Höller, M., et al. (2008). Nitrogen control in *Mycobacterium smegmatis*: nitrogen-dependent expression of ammonium transport and assimilation proteins depends on the OmpR-type regulator GlnR. *J. Bacteriol.* 190, 7108–7116. doi: 10.1128/JB.00855-08
- Amon, J., Titgemeyer, F., and Burkovski, A. (2009). A genomic view on nitrogen metabolism and nitrogen control in *Mycobacteria*. *J. Mol. Microbiol. Biotechnol.* 17, 20–29. doi: 10.1159/000159195
- Borchert, N., Dieterich, C., Krug, K., Schütz, W., Jung, S., Nordheim, et al. (2010). Proteogenomics of *Pristionchus pacificus* reveals distinct proteome structure of nematode models. *Genome Res.* 20, 837–846. doi: 10.1101/gr.103119.109
- Boyes, J., Byfield, P., Nakatani, Y., and Ogryzko, V. (1998). Regulation of activity of the transcription factor GATA-1 by acetylation. *Nature* 396, 594–598. doi: 10.1038/25166
- Brissette, R. E., Tsung, K. L., and Inouye, M. (1991). Suppression of a mutation in OmpR at the putative phosphorylation center by a mutant EnvZ protein in *Escherichia coli*. *J. Bacteriol.* 173, 601–608.
- Buttner, M. J., Chater, K. F., and Bibb, M. J. (1990). Cloning, disruption, and transcriptional analysis of three RNA polymerase sigma factor genes of *Streptomyces coelicolor* A3(2). *J. Bacteriol.* 172, 3367–3378.
- Cain, J. A., Solis, N., and Cordwell, S. J. (2014). Beyond gene expression: the impact of protein post-translational modifications in bacteria. *J. Proteomics* 97, 265–286. doi: 10.1016/j.jprot.2013.08.012
- Castaño-Cerezo, S., Bernal, V., Post, H., Fuhrer, T., Cappadona, S., Sánchez-Díaz, N. C., et al. (2014). Protein acetylation affects acetate metabolism, motility and acid stress response in *Escherichia coli*. *Mol. Syst. Biol.* 10:762. doi: 10.15252/msb.20145227
- Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., et al. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325, 834–840. doi: 10.1126/science.1175371
- Cox, J., Matic, I., Hilger, M., Nagaraj, N., Selbach, M., Olsen, J. V., et al. (2009). A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat. Protoc.* 4, 698–705. doi: 10.1038/nprot.2009.36
- Crosby, H. A., Heiniger, E. K., Harwood, C. S., and Escalante-Semerena, J. C. (2010). Reversible N-lysine acetylation regulates the activity of acyl-CoA synthetases involved in anaerobic benzoate catabolism in *Rhodospseudomonas palustris*. *Mol. Microbiol.* 76, 874–888. doi: 10.1111/j.1365-2958.2010.07127.x
- Evans, C. G., Herbert, D., and Tempest, D. W. (1970). The continuous cultivation of micro-organisms. II. Construction of a chemostat. *Methods Microbiol.* 2, 277–327. doi: 10.1016/S0580-9517(08)70227-7
- Fink, D., Weissschuh, N., Reuther, J., Wohlleben, W., and Engels, A. (2002). Two transcriptional regulators GlnR and GlnRII are involved in regulation of nitrogen metabolism in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* 46, 331–347. doi: 10.1046/j.1365-2958.2002.03150.x
- Furia, B., Deng, L., Wu, K., Baylor, S., Kehn, K., Li, H., et al. (2002). Enhancement of nuclear factor-kappa B acetylation by coactivator p300 and HIV-1 Tat proteins. *J. Biol. Chem.* 277, 4973–4980. doi: 10.1074/jbc.M107848200
- Gao, C., Hindra-Mulder, D., Yin, C., and Elliot, M. A. (2012). Crp is a global regulator of antibiotic production in streptomyces. *mBio* 3, e00407–e00412. doi: 10.1128/mBio.00407-12
- Gardner, J. G., Grundy, F. J., Henkin, T. M., and Escalante-Semerena, J. C. (2006). Control of acetyl-coenzyme A synthetase (AcsA) activity by acetylation/deacetylation without NAD(+) involvement in *Bacillus subtilis*. *J. Bacteriol.* 188, 5460–5468. doi: 10.1128/JB.00215-06
- He, X., Wang, L., and Wang, S. (2016). Structural basis of DNA sequence recognition by the response regulator PhoP in *Mycobacterium tuberculosis*. *Sci. Rep.* 6, 24442–24442. doi: 10.1038/srep24442
- Hu, L. I., Lima, B. P., and Wolfe, A. J. (2010). Bacterial protein acetylation: the dawning of a new age. *Mol. Microbiol.* 77, 15–21. doi: 10.1111/j.1365-2958.2010.07204.x
- Huang, D., Li, Z. H., Di, Y., Zhou, Y., and Ye, B.-C. (2015). Lysine acetylproteome analysis suggests its roles in primary and secondary metabolism in *Saccharopolyspora erythraea*. *Appl. Microbiol. Biotechnol.* 99, 1399–1413. doi: 10.1007/s00253-014-6144-2
- Jenkins, V. A., Barton, G. R., Robertson, B. D., and Williams, K. J. (2013). Genome wide analysis of the complete GlnR nitrogen-response regulon in *Mycobacterium smegmatis*. *BMC Genomics* 14:301. doi: 10.1186/1471-2164-14-301
- Jones, J. D., and O'Connor, C. D. (2011). Protein acetylation in prokaryotes. *Proteomics* 11, 3012–3022. doi: 10.1002/pmic.201000812
- Kalantari, A., Derouiche, A., Shi, L., and Mijakovic, I. (2015). Serine/threonine/tyrosine phosphorylation regulates DNA binding of bacterial transcriptional regulators. *Microbiology* 161, 1720–1729. doi: 10.1099/mic.0.000148
- Kennelly, P. J. (2002). Protein kinases and protein phosphatases in prokaryotes: a genomic perspective. *FEMS Microbiol. Lett.* 206, 1–8. doi: 10.1111/j.1574-6968.2002.tb10978.x
- Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., and Hopwood, D. A. (2000). *Practical Streptomyces Genetics*. Norwich: The John Innes Foundation.
- Kim, S. C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., et al. (2006). Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol. Cell.* 23, 607–618. doi: 10.1016/j.molcel.2006.06.026
- Koch, A., Krug, K., Pengelley, S., Macek, B., and Hauf, S. (2011). Mitotic substrates of the kinase aurora with roles in chromatin regulation identified



- through quantitative phosphoproteomics of fission yeast. *Sci. Signal* 4, rs6. doi: 10.1126/scisignal.2001588
- Kosono, S., Tamura, M., Suzuki, S., Kawamura, Y., Yoshida, A., Nishiyama, M., et al. (2015). Changes in the acetylome and succinylome of *Bacillus subtilis* in response to carbon source. *PLoS ONE* 10:e0131169. doi: 10.1371/journal.pone.0131169
- Ladwig, N., Franz-Wachtel, M., Hezel, F., Soufi, B., Macek, B., Wohlleben, W., et al. (2015). Control of morphological differentiation of *Streptomyces coelicolor* A3(2) by phosphorylation of MreC and PBP2. *PLoS ONE* 10:e0125425. doi: 10.1371/journal.pone.0125425
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680. doi: 10.1038/227680a0
- Liao, C. H., Yao, L., Xu, Y., Liu, W. B., Zhou, Y., and Ye, B. C. (2015). Nitrogen regulator GlnR controls uptake and utilization of non-phosphotransferase-system carbon sources in actinomycetes. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15630–15635. doi: 10.1073/pnas.1508465112
- Liarzi, O., Barak, R., Bronner, V., Dines, M., Sagi, Y., Shainskaya, A., et al. (2010). Acetylation represses the binding of CheY to its target proteins. *Mol. Microbiol.* 76, 932–943. doi: 10.1111/j.1365-2958.2010.07148.x
- Lin, W., Wang, Y., Han, X., Zhang, Z., Wang, C., Wang, J., et al. (2014). Atypical OmpR/PhoB subfamily response regulator GlnR of actinomycetes functions as a homodimer, stabilized by the unphosphorylated conserved Asp-focused charge interactions. *J. Biol. Chem.* 289, 15413–15425. doi: 10.1074/jbc.M113.543504
- Liu, F., Yang, M., Wang, X., Yang, S., Gu, J., Zhou, J., et al. (2014). Acetylome analysis reveals diverse functions of lysine acetylation in *Mycobacterium tuberculosis*. *Mol. Cell. Proteomics* 13, 3352–3366. doi: 10.1074/mcp.M114.041962
- Malm, S., Tiffert, Y., Micklinghoff, J., Schultze, S., Joost, I., Weber, I., et al. (2009). The roles of the nitrate reductase NarGHJI, the nitrite reductase NirBD and the response regulator GlnR in nitrate assimilation of *Mycobacterium tuberculosis*. *Microbiology* 155, 1332–1339. doi: 10.1099/mic.0.023275-0
- Manteca, A., Ye, J., Sánchez, J., and Jensen, O. N. (2011). Phosphoproteome analysis of *Streptomyces* development reveals extensive protein phosphorylation accompanying bacterial differentiation. *J. Proteome Res.* 10, 5481–5492. doi: 10.1021/pr200762y
- Martín, J. F., Sola-Landa, A., Santos-Beneit, F., Fernández-Martínez, L. T., Prieto, C., and Rodríguez-García, A. (2011). Cross-talk of global nutritional regulators in the control of primary and secondary metabolism in *Streptomyces*. *Microb. Biotechnol.* 4, 165–174. doi: 10.1111/j.1751-7915.2010.00235.x
- Marzio, G., Wagener, C., Gutierrez, M. I., Cartwright, P., Helin, K., and Giacca, M. (2000). E2F family members are differentially regulated by reversible acetylation. *J. Biol. Chem.* 275, 10887–10892. doi: 10.1074/jbc.275.15.10887
- Mikulik, K., Felsenberg, J., Kudračková, E., Bezouskova, S., Setinova, D., Stodulkova, E., et al. (2012). CobB1 deacetylase activity in *Streptomyces coelicolor*. *Biochem. Cell Biol.* 90, 179–187. doi: 10.1139/o11-086
- Molle, V., Kremer, L., Girard-Blanc, C., Besra, G. S., Cozzzone, A. J., and Prost, J. F. (2003). An FHA phosphoprotein recognition domain mediates protein EmrB phosphorylation by PknH, a Ser/Thr protein kinase from *Mycobacterium tuberculosis*. *Biochemistry* 42, 15300–15309. doi: 10.1021/bi035150b
- Moore, J. M., Bradshaw, E., Seipke, R. F., Hutchings, M. I., and McArthur, M. (2012). Use and discovery of chemical elicitors that stimulate *Streptomyces* bacteria. *Methods Enzymol.* 517, 367–385. doi: 10.1016/B978-0-12-404634-4.00018-8
- Noort, V., Seebacher, J., Bader, S., Mohammed, S., Vonkova, I., Betts, M. J., et al. (2012). Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium. *Mol. Syst. Biol.* 8, 571. doi: 10.1038/msb.2012.4
- Okanishi, W. L., Suzuki, K., and Umezawa, H. (1974). Formation and reversion of *Streptomyces* protoplasts: cultural conditions and morphological study. *J. Gen. Microbiol.* 80, 389–400. doi: 10.1099/00221287-80-2-389
- Olsen, J. V., and Macek, B. (2009). High accuracy mass spectrometry in large-scale analysis of protein phosphorylation. *Meth. Mol. Biol.* 492, 131–142. doi: 10.1007/978-1-59745-493-3\_7
- Parker, J. L., Jones, A. M., Serazetdinova, L., Saalbach, G., Bibb, M. J., and Naldrett, M. J. (2010). Analysis of the phosphoproteome of the multicellular bacterium *Streptomyces coelicolor* A3(2) by protein/peptide fractionation, phosphopeptide enrichment and high accuracy mass spectrometry. *Proteomics* 10, 2486–2497. doi: 10.1002/pmic.201000090
- Petridis, M., Benjak, A., and Cook, G. M. (2015). Defining the nitrogen regulated transcriptome of *Mycobacterium smegmatis* using continuous culture. *BMC Genomics* 16:821. doi: 10.1186/s12864-015-2051-x
- Prisic, S., Dankwa, S., Schwartz, D., Chou, M. F., Locasale, J. W., Kang, C. M., et al. (2010). Extensive phosphorylation with overlapping specificity by *Mycobacterium tuberculosis* serine/threonine protein kinases. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7521–7526. doi: 10.1073/pnas.0913482107
- Pullan, S. T., Chandra, G., Bibb, M. J., and Merrick, M. (2011). Genome-wide analysis of the role of GlnR in *Streptomyces venezuelae* provides new insights into global nitrogen regulation in actinomycetes. *BMC Genomics* 12:175. doi: 10.1186/1471-2164-12-175
- Qu, S., Kang, Q., Wu, H., Wang, L., and Bai, L. (2015). Positive and negative regulation of GlnR in validamycin A biosynthesis by binding to different loci in promoter region. *Appl. Microbiol. Biotechnol.* 99, 4771–4783. doi: 10.1007/s00253-015-6437-0
- Reuther, J., and Wohlleben, W. (2007). Nitrogen metabolism in *Streptomyces coelicolor*: transcriptional and post-translational regulation. *J. Mol. Microbiol. Biotechnol.* 12, 139–146. doi: 10.1159/000096469
- Rodríguez-García, A., Sola-Landa, A., Apel, K., Santos-Beneit, F., and Martín, J. F. (2009). Phosphate control over nitrogen metabolism in *Streptomyces coelicolor*: direct and indirect negative control of *glnR*, *glnA*, *glnII* and *amtB* expression by the response regulator PhoP. *Nucleic Acids Res.* 37, 3230–3242. doi: 10.1093/nar/gkp162
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual, 2nd Edn.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Sawai, R., Suzuki, A., Takano, Y., Lee, P. C., and Horinouchi, S. (2004). Phosphorylation of AfsR by multiple serine/threonine kinases in *Streptomyces coelicolor* A3(2). *Gene* 334, 53–61. doi: 10.1016/j.gene.2004.02.046
- Shao, Z., Deng, W., Li, S., He, J., Ren, S., Huang, W., et al. (2015). GlnR-Mediated Regulation of *ectABCD* transcription expands the role of the GlnR regulon to osmotic stress management. *J. Bacteriol.* 197, 3041–3047. doi: 10.1128/JB.00185-15
- Sharma, K., Gupta, M., Gupta, N., Pathak, M., Koul, A., Sarangi, S., et al. (2006). Transcriptional control of the mycobacterial *embCAB* operon by PknH through a regulatory protein, EmrB, *in vivo*. *J. Bacteriol.* 188, 2936–2944. doi: 10.1128/JB.188.8.2936-2944.2006
- Sola-Landa, A., Rodríguez-García, A., Amin, R., Wohlleben, W., and Martín, J. F. (2013). Competition between the GlnR and PhoP regulators for the *glnA* and *amtB* promoters in *Streptomyces coelicolor*. *Nucleic Acids Res.* 41, 1767–1782. doi: 10.1093/nar/gks1203
- Soufi, B., Soares, N. C., Ravikumar, V., and Macek, B. (2012). Proteomics reveals evidence of cross-talk between protein modifications in bacteria: focus on acetylation and phosphorylation. *Curr. Opin. Microbiol.* 15, 357–363. doi: 10.1016/j.mib.2012.05.003
- Starai, V. J., Celic, I., Cole, R. N., Boeke, J. D., and Escalante-Semerena, J. C. (2002). Sir2-dependent activation of acetyl-CoA synthetase by deacetylation deacetylation of active lysine. *Science*, 298, 2390–2392. doi: 10.1126/science.1077650
- Studier, F. W., and Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189, 113–130. doi: 10.1016/0022-2836(86)90385-2
- Thao, S., Chen, C. S., Zhu, H., and Escalante-Semerena, J. C. (2010). Nε-lysine acetylation of a bacterial transcription factor inhibits its DNA-binding activity. *PLoS ONE* 5:15123. doi: 10.1371/journal.pone.0015123
- Thao, S., and Escalante-Semerena, J. C. (2011). Control of protein function by reversible Nε-lysine acetylation in bacteria. *Curr. Opin. Microbiol.* 14, 200–204. doi: 10.1016/j.mib.2010.12.013
- Tiffert, Y., Franz-Wachtel, M., Fladerer, C., Nordheim, A., Reuther, J., Wohlleben, W., et al. (2011). Proteomic analysis of the GlnR-mediated response to nitrogen limitation in *Streptomyces coelicolor* M145. *Appl. Microbiol. Biotechnol.* 89, 1149–1159. doi: 10.1007/s00253-011-3086-9
- Tiffert, Y., Supra, P., Wurm, R., Wohlleben, W., Wagner, R., and Reuther, J. (2008). The *Streptomyces coelicolor* GlnR regulon: identification of new GlnR targets and evidence for a central role of GlnR in nitrogen

- metabolism in actinomycetes. *Mol. Microbiol.* 67, 861–880. doi: 10.1111/j.1365-2958.2007.06092.x
- Wang, J., and Zhao, G.-P. (2009). GlnR positively regulates *nasA* transcription in *Streptomyces coelicolor*. *Biochem. Biophys. Res. Commun.* 386, 77–81. doi: 10.1016/j.bbrc.2009.05.147
- Wang, Q., Zhang, Y., Yang, C., Xiong, H., Lin, Y., Yao, J., et al. (2010). Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. *Science* 327, 1004–1007. doi: 10.1126/science.1179687
- Wang, R., Mast, Y., Wang, J., Zhang, W., Zhao, G., Wohlleben, W., et al. (2013). Identification of two-component system AfsQ1/Q2 regulon and its cross-regulation with GlnR in *Streptomyces coelicolor*. *Mol. Microbiol.* 87, 30–48. doi: 10.1111/mmi.12080
- Wang, Y., Li, C., Duan, N., Li, B., Ding, X.-M., Yao, Y.-F., et al. (2014). GlnR negatively regulates the transcription of the alanine dehydrogenase encoding gene *ald* in *Amycolatopsis mediterranei* U32 under nitrogen limited conditions via specific binding to its major transcription initiation site. *PLoS ONE* 9:e104811. doi: 10.1371/journal.pone.0104811
- Wang, Y., Wang, J.-Z., Shao, Z.-H., Yuan, H., Lu, Y.-H., Jiang, W.-H., et al. (2013). Three of four GlnR binding sites are essential for GlnR-mediated activation of transcription of the *Amycolatopsis mediterranei* *nas* operon. *J. Bacteriol.* 195, 2595–2602. doi: 10.1128/JB.00182-13
- Waterborg, J. H. (2001). Dynamics of histone acetylation in *Saccharomyces cerevisiae*. *Biochem.* 40, 2599–2605. doi: 10.1021/bi002480c
- Weinert, B. T., Iesmantavicius, V., Wagner, S. A., Schölz, C., Gummeson, B., Beli, P., et al. (2013). Acetyl-phosphate is a critical determinant of lysine acetylation in *E. coli*. *Mol. Cell* 51, 265–272. doi: 10.1016/j.molcel.2013.06.003
- Williams, K. J., Jenkins, V. A., Barton, G. R., Bryant, W. A., Krishnan, N., and Robertson, B. D. (2015). Deciphering the metabolic response of *Mycobacterium tuberculosis* to nitrogen stress. *Mol. Microbiol.* 97, 1142–1157. doi: 10.1111/mmi.13091
- Wohlleben, W., Stegmann, E., and Süssmuth, R. D. (2009). Chapter 18. Molecular genetic approaches to analyze glycopeptide biosynthesis. *Methods Enzymol.* 458, 459–486. doi: 10.1016/S0076-6879(09)04818-6
- Xu, H., Hegde, S. S., and Blanchard, J. S. (2011). Reversible acetylation and inactivation of *Mycobacterium tuberculosis* acetyl-CoA synthetase is dependent on cAMP. *Biochemistry* 50, 5883–5892. doi: 10.1021/bi200156t
- Yao, L. L., Liao, C. H., Huang, G., Zhou, Y., Rigali, S., Zhang, B., et al. (2014). GlnR-mediated regulation of nitrogen metabolism in the actinomycete *Saccharopolyspora erythraea*. *Appl. Microbiol. Biotechnol.* 98, 7935–7948. doi: 10.1007/s00253-014-5878-1
- Yu, J., and Auwerx, J. (2009). The role of sirtuins in the control of metabolic homeostasis. *Ann. N.Y. Acad. Sci.* 1173(Suppl. 1), E10–E19. doi: 10.1111/j.1749-6632.2009.04952.x
- You, D., Yao, L. L., Huang, D., Escalante-Semerena, J. C., and Ye, B. C. (2014). Acetyl coenzyme A synthetase is acetylated on multiple lysine residues by a protein acetyltransferase with a single Gcn5-type N-acetyltransferase (GNAT) domain in *Saccharopolyspora erythraea*. *J. Bacteriol.* 196, 3169–3178. doi: 10.1128/JB.01961-14
- Zhang, J., Sprung, R., Pei, J., Tan, X., Kim, S., Zhu, H., et al. (2009). Lysine acetylation is a highly abundant and evolutionarily conserved modification in *Escherichia coli*. *Mol. Cell. Proteomics* 8, 215–225. doi: 10.1074/mcp.M800187-MCP200

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Amin, Franz-Wachtel, Tiffert, Heberer, Meky, Ahmed, Matthews, Krysenko, Jakobi, Hinder, Moore, Okoniewski, Maček, Wohlleben and Bera. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The *Escherichia Coli* Hfq Protein: An Unattended DNA-Transactions Regulator

Grzegorz M. Cech<sup>1</sup>, Agnieszka Szalewska-Pałasz<sup>1</sup>, Krzysztof Kubiak<sup>1,2,3</sup>, Antoine Malabirade<sup>2</sup>, Wilfried Grange<sup>3,4</sup>, Veronique Arluison<sup>2,4\*</sup> and Grzegorz Węgrzyn<sup>1\*</sup>

<sup>1</sup> Department of Molecular Biology, University of Gdańsk, Gdańsk, Poland, <sup>2</sup> Laboratoire Léon Brillouin, CEA, Centre National de la Recherche Scientifique, Université Paris Saclay, CEA Saclay, Gif-sur-Yvette, France, <sup>3</sup> IPCMS/Centre National de la Recherche Scientifique, Strasbourg, France, <sup>4</sup> Université Paris Diderot, UFR Science du Vivant, Paris, France

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Spanish National Research Council,  
Spain

### Reviewed by:

Jorge Humberto Leitão,  
Universidade de Lisboa, Portugal  
Mikolaj Olejniczak,  
Adam Mickiewicz University in  
Poznań, Poland

### \*Correspondence:

Veronique Arluison  
veronique.arluison@  
univ-paris-diderot.fr  
Grzegorz Węgrzyn  
grzegorz.wegrzyn@biol.ug.edu.pl

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 27 May 2016

**Accepted:** 13 July 2016

**Published:** 28 July 2016

### Citation:

Cech GM, Szalewska-Pałasz A,  
Kubiak K, Malabirade A, Grange W,  
Arluison V and Węgrzyn G (2016) The  
*Escherichia Coli* Hfq Protein: An  
Unattended DNA-Transactions  
Regulator. *Front. Mol. Biosci.* 3:36.  
doi: 10.3389/fmolb.2016.00036

The Hfq protein was discovered in *Escherichia coli* as a host factor for bacteriophage Q $\beta$  RNA replication. Subsequent studies indicated that Hfq is a pleiotropic regulator of bacterial gene expression. The regulatory role of Hfq is ascribed mainly to its function as an RNA-chaperone, facilitating interactions between bacterial non-coding RNA and its mRNA target. Thus, it modulates mRNA translation and stability. Nevertheless, Hfq is able to interact with DNA as well. Its role in the regulation of DNA-related processes has been demonstrated. In this mini-review, it is discussed how Hfq interacts with DNA and what is the role of this protein in regulation of DNA transactions. Particularly, Hfq has been demonstrated to be involved in the control of ColE1 plasmid DNA replication, transposition, and possibly also transcription. Possible mechanisms of these Hfq-mediated regulations are described and discussed.

**Keywords:** Hfq, RNA chaperone, nucleoid associated protein, DNA replication, transposition

## INTRODUCTION

The story of the Hfq protein of *Escherichia coli* was initiated in 1968, when the roles of host factors required for replication of bacteriophage Q $\beta$  genetic material, which is RNA, have been evidenced (Franze De Fernandez et al., 1972). Subsequent studies demonstrated that there are at least two such factors (Shapiro et al., 1968). One of them was purified and identified as an RNA-binding protein (Franze De Fernandez et al., 1972). It has been named Host Factor I (HF I), as Q $\beta$  RNA synthesis *in vitro* by the phage-encoded RNA polymerase was strictly dependent on this protein (Franze De Fernandez et al., 1972). The purified protein was demonstrated to interact with single-stranded RNA, however, no binding of HF I to double-stranded RNA and to single- or double-stranded DNA was detected (Franze De Fernandez et al., 1972). The HF I name has then been replaced with Hfq (for host factor for phage Q $\beta$  replication), after cloning and sequencing the corresponding gene (Kajitani and Ishihama, 1991).

Today, the Hfq protein is known to be a major riboregulator that facilitates cellular RNA-RNA interactions. Particularly well documented is the binding of Hfq to small non-coding RNAs (sRNA) that play important roles in the regulation of gene expression at the post-transcriptional level. Enhancement of sRNA-mRNA interaction, which is facilitated by Hfq, most often inhibits translation by blocking the Shine-Dalgarno and/or start codon regions, but also affects RNA

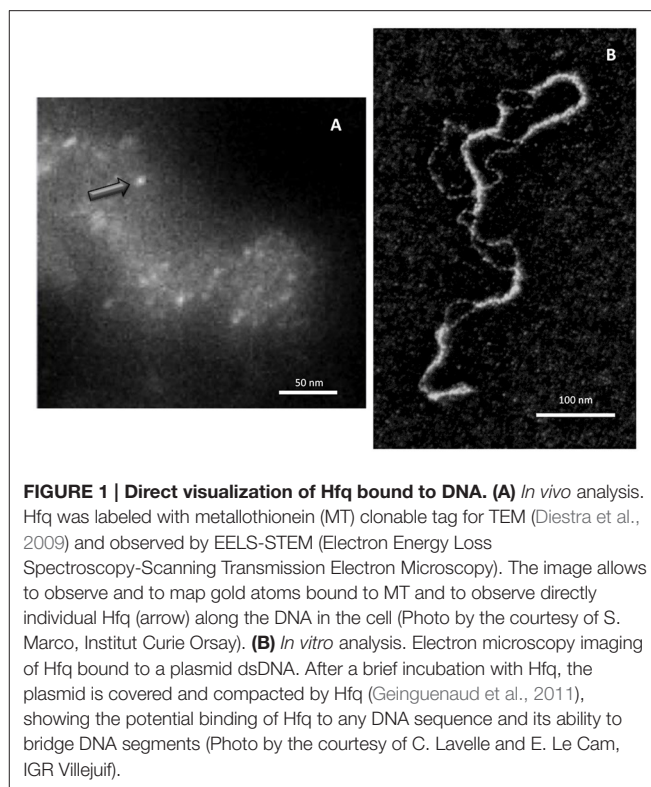
**Abbreviations:** NAP, Nucleoid Associated Protein; sRNA, Small non-coding RNA; CTR/NTR, C/N-terminal domain.

stability. This allows bacteria to adapt to their environment, especially in the case of the host infection. The multiple functions of Hfq connected to its interactions with RNA molecules have been recently reviewed in several excellent articles (Vogel and Luisi, 2011; Sobrero and Valverde, 2012; Gottesman and Storz, 2015; Updegrove et al., 2016). In this mini-review, we will focus on other Hfq activities, namely its involvement in DNA transactions.

## DIRECT INTERACTIONS BETWEEN HFQ AND DNA

Although early experiments failed to identify Hfq binding to DNA (Franze De Fernandez et al., 1972), the ability of the *hfq* gene product to interact with both supercoiled and linear plasmid DNA has been demonstrated 25 years later (Takada et al., 1997). Clearly, Hfq preferentially binds RNA molecules to DNA: while equilibrium dissociation constants ( $K_d$ ) for DNA range from nM to  $\mu$ M (Updegrove et al., 2010; Geinguenaud et al., 2011), for cellular RNA they range from tens of pM for *rpsO* polyadenylated mRNA to nM for sRNAs, such as MicA and DsrA (Folichon et al., 2003; Lease and Woodson, 2004; Fender et al., 2010). For shorter model oligonucleotides, the tightest value measured for oligoriboadenylate ( $rA_{16}$ ) was 1.4 nM, and affinity was 60 times weaker for the corresponding oligodesoxyriboadenylates ( $dA_6$ ) vs. oligoriboadenylates ( $rA_6$ ) (Link et al., 2009). Despite its apparent cellular abundance (10  $\mu$ M), Hfq low availability *in vivo* questions about its simultaneous binding to RNA and DNA (Hussein and Lim, 2011; Wagner, 2013). Nevertheless, Hfq has been shown to be one of the nucleoid-associated proteins (NAP) (Azam and Ishihama, 1999). If the presence of Hfq in the nucleoid could result from its binding to transcribed RNA, its direct binding to genomic DNA also occurs as DNA fragments are found associated with the purified protein (Updegrove et al., 2010). Note that the direct observation of Hfq in the nucleoid is possible, but difficult taking into account its abundance along the inner bacterial membrane (Azam et al., 2000; Taghbalout et al., 2014). Furthermore, Hfq nucleoid fraction represents 10–20%, while its cytoplasmic and membrane-bound fractions are about 30 and 50%, respectively (Diestra et al., 2009). This makes its observation inside the cell difficult, but its presence along the DNA *in vivo* could be confirmed by electron microscopy imaging of bacteria ultrathin-section (**Figure 1A**; Diestra et al., 2009).

*In vitro* analyses allowed measuring the  $K_d$  constants of Hfq-DNA complexes. For the complexes with single stranded  $dA_{20}$ , the  $K_d$  was  $\sim$  200 nM, while for double stranded  $dA_{20}$ - $dT_{20}$  it was  $\sim$  250 nM. This agrees with previous reports indicating that Hfq-bound DNA fragments are of curved topology (Azam and Ishihama, 1999). Moreover, low affinity was measured for  $dT_{20}$  and  $dG_{20}$  ( $>1 \mu$ M) and no complex was detected for  $dC_{20}$  or  $dC_{20}$ - $dG_{20}$  (Geinguenaud et al., 2011). On the other hand, *in vivo* analyses indicated a preferred Hfq-binding motif ( $A/T$ )T( $A/G$ )TGCCG (Updegrove et al., 2010). The affinity of Hfq to this motif was slightly lower than to A-tracts. Intriguingly, analysis of identified Hfq-binding sequences indicated that most of them derived from genes coding for membrane proteins (Updegrove et al., 2010). The identified genes did not include any cistrons, which mRNAs were known to be regulated by Hfq,



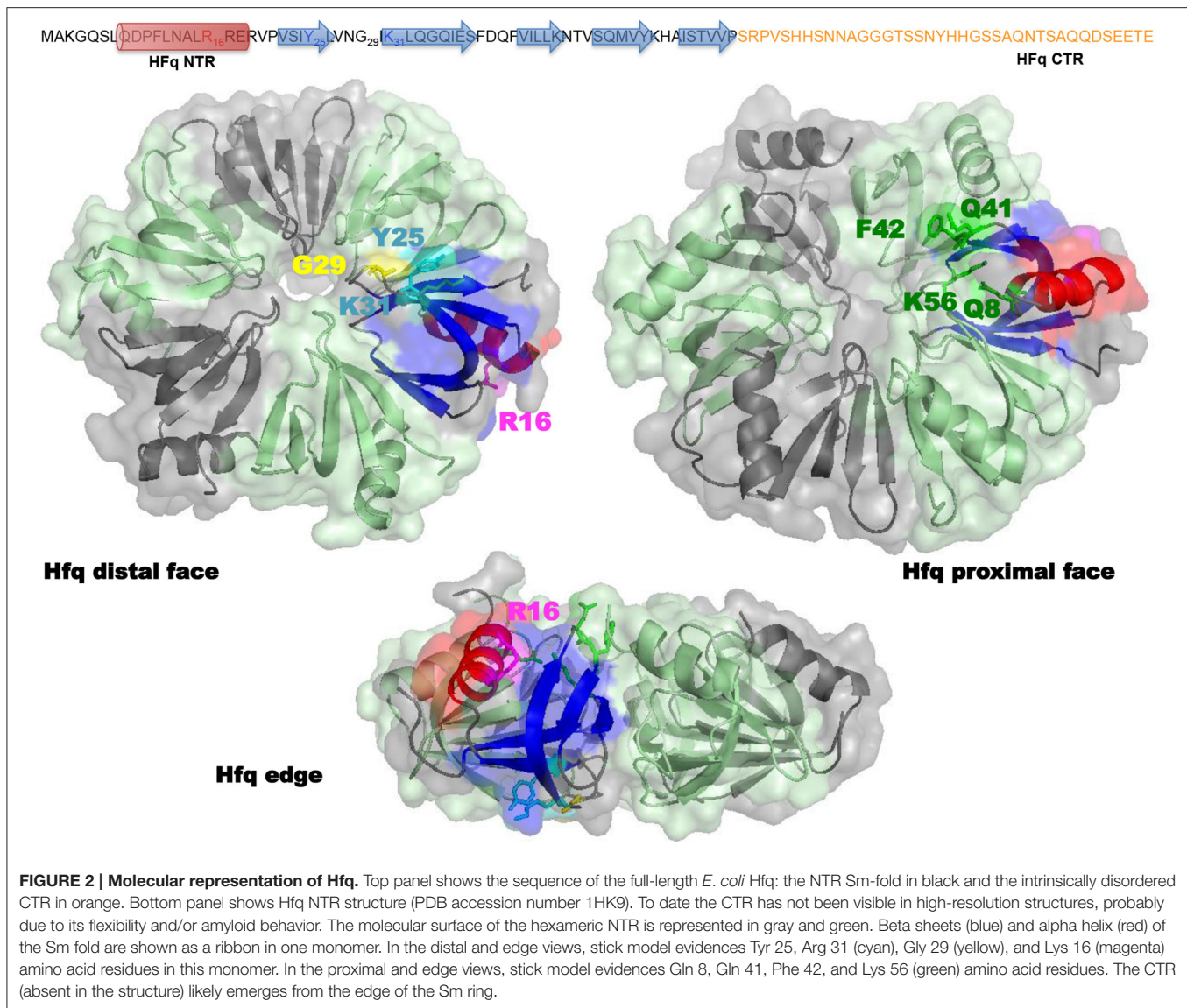
**FIGURE 1 | Direct visualization of Hfq bound to DNA. (A)** *In vivo* analysis. Hfq was labeled with metallothionein (MT) clonable tag for TEM (Diestra et al., 2009) and observed by EELS-STEM (Electron Energy Loss Spectroscopy-Scanning Transmission Electron Microscopy). The image allows to observe and to map gold atoms bound to MT and to observe directly individual Hfq (arrow) along the DNA in the cell (Photo by the courtesy of S. Marco, Institut Curie Orsay). **(B)** *In vitro* analysis. Electron microscopy imaging of Hfq bound to a plasmid dsDNA. After a brief incubation with Hfq, the plasmid is covered and compacted by Hfq (Geinguenaud et al., 2011), showing the potential binding of Hfq to any DNA sequence and its ability to bridge DNA segments (Photo by the courtesy of C. Lavelle and E. Le Cam, IGR Villejuif).

and no previously-described sRNA were encoded by these DNA fragments. This result suggests a general role for Hfq in the regulation of membrane protein expression (Guillier et al., 2006).

Despite a relative high affinity to A-tracts regions, at higher Hfq concentrations the protein interacts with DNA in a sequence-nonspecific manner, as suggested earlier (Takada et al., 1997; Azam and Ishihama, 1999). As seen by molecular imaging, Hfq binds and covers long DNA sequences (**Figure 1B**). The presence of continuous Hfq stretches, separated by naked DNA, suggests a cooperative binding mechanism by which Hfq could nucleate on high affinity sites followed by spreading along surrounding regions. This model is in agreement with previous observations showing supershifted DNA bands on gel, suggesting multiple Hfq binding to DNA when increasing Hfq concentrations (Azam and Ishihama, 1999; Updegrove et al., 2010).

Structurally, *E. coli* Hfq forms an Sm-fold in its N-terminal region ( $\sim$  65 amino acids; **Figure 2**). This fold consists of a five  $\beta$ -strands antiparallel  $\beta$ -sheet, capped by an  $\alpha$ -helix. The  $\beta$ -sheets from six monomers interact with each other to assemble in a toroidal structure with two non-equivalent faces, i.e., the proximal (on which the  $\alpha$ -helix is exposed) and distal surfaces (Link et al., 2009; **Figure 2**). It appears that the distal surface, the edge and the C-terminal region (CTR,  $\sim$  35 amino acids) of Hfq, are involved in DNA binding. Such a conclusion was made on the basis of experiments with *hfq* mutants bearing either deletions of Hfq C-terminal (CTR) domain, or point mutations, such as R16A (edge), K31A and Y25A (distal face) (Updegrove et al., 2010; **Figure 2**). While CTR is dispensable for RNA binding (Arluison et al., 2004),





the distal surface and the edge of the protein seem to be involved in both DNA and RNA binding. Oppositely, the proximal surface seems to be involved in RNA binding only (mainly Q8, Q41, F42, K56, and H57 amino acid residues, **Figure 2**; Updegrove et al., 2010; Wang et al., 2011). This makes sense taking into account adenylate binding to the distal face, while the proximal side is more dedicated to uridine-rich sequences, absent in DNA (Link et al., 2009; Schu et al., 2015). The preference of Hfq for A-rich RNA over DNA is explained by the formation of a hydrogen bond between the ribosyl 2' hydroxyl group and the carbonyl oxygen of residue Gly 29 (Link et al., 2009; **Figure 2**). While dispensing for DNA binding, it was proposed that the C-terminal domain might anchor the Hfq protein to DNA, while the distal face of the Sm-core could be required to direct interactions with the nucleic acid.

Indeed, the involvement of Hfq CTR in DNA binding has also been recently confirmed by low resolution structure of Hfq:DNA

complex (Jiang et al., 2015). Previous electron microscopy imaging evidenced that Hfq acts by bridging together two DNA molecules (**Figure 1B**; Geinguenaud et al., 2011). Such an activity was formerly documented for the H-NS (histone-like nucleoid structuring) NAP. A model for Hfq propensity for bridging has thus been proposed to be linked to its CTR-arm. This interaction results in a change of the mechanical properties of the double helix, and in a compaction of DNA into a condensed form (Jiang et al., 2015). Recently, a new unexpected property of Hfq CTR-arm has also been evidenced. Indeed, the CTR region, which was considered as intrinsically unstructured (Vogel and Luisi, 2011), contains an amyloid sequence that allows the protein to self-assemble (Arлуison et al., 2006; Fortas et al., 2015). This new property of the CTR could explain its self-assembly and spreading of Hfq on DNA, but this needs to be investigated further.

Finally, one important question remains about the precise effect of Hfq on DNA *in vivo*. Hfq has been proposed to change DNA superhelicity, precisely plasmids purified in the absence of

Hfq are less negatively supercoiled during the stationary phase (Tsui et al., 1994). Nevertheless, this property has never been analyzed in detail. Sugar re-puckering has also been described but should be confirmed by *in vivo* analyses (Geinguenaud et al., 2011).

As for the relative amount of Hfq compared to other NAP, its abundance in actively growing cells is similar to those of the most abundant NAPs, Fis (factor for inversion simulation) or HU (heat-unstable protein). Nevertheless, Hfq quantity in the nucleoid is lower than for HU and Fis, which are present in the nucleoid only. A rough estimation indicate that Hfq represents about 5% of the total nucleoid-associated proteins (vs. 20 and 40% for Fis and HU during the exponential phase of growth, respectively, Talukder and Ishihama, 2015). It is now established that Hfq concentration increases when reaching the stationary phase (Tsui et al., 1997; Diestra et al., 2009; Cech et al., 2014), even if Dps (DNA-binding protein from starved cells) is the main nucleoid-associated protein during this phase of growth. Hfq nucleoid-bound fraction remains more or less constant during the cell cycle (Talukder and Ishihama, 2015). Furthermore, it appears that Hfq can not only interact with DNA, but also: (i) interacts with some NAP, such as H-NS (Kajitani and Ishihama, 1991), (ii) cooperates in the organization of the bacterial chromosome with other proteins, like Fis, HU, H-NS, IHF (integration host factor protein) and StpA (suppressor of *td* mutant phenotype A; Ohniwa et al., 2013) and (iii) regulates the expression of other NAPs (Lease and Belfort, 2000; Lu et al., 2016). The proteins associated to the nucleoid are usually divided into two groups depending on whether they bridge or bend DNA (Gruber, 2014). While HU, IHF and Fis belong to the bending group of NAP that causes local folding of DNA, H-NS and Hfq belong to the bridging group that organize large parts of chromosomes into isolated domains (Dorman, 2009; Geinguenaud et al., 2011).

Structurally, the nucleoid structure during the exponential growth of *E. coli* is described as a chromatin-like fibers ranging from 5 to 80 nm in diameter (Ohniwa et al., 2013). Atomic force microscopic analyses indicate that over 60% of the fiber structures enter into the "thin" category (i.e., 5–20 nm), and that this chromatin-like structure is significantly condensed upon entering into the stationary phase. The absence of some NAPs can change the relative abundance of "thin" vs. "thick" fibers (Ohniwa et al., 2013). The absence of Hfq results in an increase of both "thin" and "thick" fiber populations. None of the single deletions of NAP were described to cause a significant change in the fraction of the fibers population and more likely the lack of only one nucleoid-associated proteins can be compensated by functions of other proteins from the same group (Ohniwa et al., 2013). Indeed, whether the bridging Hfq and H-NS could replace each other is still unknown.

## INDIRECT INVOLVEMENT OF HFQ IN DNA-RELATED PROCESSES

The fact that Hfq can interact with DNA and change its properties, inspired studies on the putative role of this protein in various DNA transactions. However, to date, only two processes

were investigated in more depth, replication and transposition. Some clues tend to indicate an effect in transcription, but the role of Hfq in this process is still an open question.

## Replication

The role of Hfq in the regulation of DNA replication has been investigated to date only using the models of plasmid replicons. Employing wild-type *E. coli* strain and otherwise isogenic *hfq* mutant, the efficiency of replication of ColE1-like (pMB1—and p15A) and bacteriophage  $\lambda$ -derived plasmids was investigated. In bacteria devoid of Hfq, significant differences in plasmid amount and kinetics of plasmid DNA synthesis were observed relative to wild-type cells, but only for ColE1-like plasmids, not for  $\lambda$  replicons (Cech et al., 2014).

Levels of the Hfq protein in the wild-type strain were increased at the late exponential and early stationary phases of bacterial culture growth relative to the early exponential phase. Accordingly, ColE1-like plasmids replicated in the *hfq* mutant more efficiently than in the wild-type bacteria at late exponential and early stationary phases, while less efficiently at the early exponential phase (Cech et al., 2014). Thus, the differences between wild-type and *hfq* mutant hosts were the most pronounced under conditions corresponding to the highest levels of Hfq in wild-type bacteria. Interestingly, effects of the *hfq* deletion on ColE1-like plasmid replication were impaired in the absence of the *rom* gene (Cech et al., 2014). The regulation of replication of ColE1-like plasmids is based on the action of an anti-sense RNA, called RNA I, which impairs the priming reaction by binding to pre-primer RNA, called RNA II. Since *rom* codes for a protein responsible for enhancing RNA I-RNA II interactions, it was hypothesized that Hfq might either modulate these interactions or interplay with Rom to influence the negative regulation facilitated by this protein (Cech et al., 2014).

In the  $\lambda$  replicons, which are not affected by the absence of Hfq, no RNA-RNA interactions are required for the replication regulation. Although transcription plays a crucial role in the replication initiation from *ori $\lambda$* , it appears that Hfq does not modulate the transcriptional activation of this site (Cech et al., 2014).

## Transposition

Hfq influences many cellular processes at the post-transcriptional level mostly by regulating and promoting interactions between RNAs. In addition to direct interactions of Hfq with DNA, indirect effects of the protein on the DNA-related processes have been reported. These include the regulation of transposition process in several known transposon systems in bacteria: Tn10, Tn5, and IS200. In all of these systems, Hfq is a potent inhibitor of transposition as shown in experiments with *hfq*-deficient mutant strains. However, the mode of Hfq action in transposition varies among these systems (Ellis and Haniford, 2016).

For Tn10, Hfq exerts its function typically by promoting base-pairing between transposase RNA (RNA-IN) and its antisense RNA (RNA-OUT) encoded by Tn10/IS10 antisense system (Ross et al., 2013). As a result, downregulation of RNA-IN expression *in vivo* occurs. In addition to this effect, Hfq can inhibit transposase expression independently of the antisense

system. It was proposed that Hfq binds directly to RNA-IN, blocking its translation (Ross et al., 2010; Ellis and Haniford, 2016).

Hfq-mediated regulation of transcription of the gene coding for the IS50 transposase might also explain the inhibition of Tn5 transposition (see Transcription; Haniford and Ellis, 2015). Moreover, Crp (cyclic AMP receptor protein) was proposed to mediate Hfq effect on Tn5 transposition (Ross et al., 2014). This, together with the known role of Hfq in stress response processes, supports the hypothesis of linking of transposition to the physiological status of the cell.

IS200 elements, abundant in Enterobacteriaceae, transpose with a very low frequency and are limited by the expression on the transposase gene, *tnpA*. In addition, transposase expression is downregulated by antisense RNA, *art200* (Ellis et al., 2015). However, the role of Hfq in the regulation of IS200 transposition does not involve pairing between mRNA and *art200*. It was shown that Hfq binds upstream of SD sequence of *tnpA* mRNA, inhibiting ribosome binding (Ellis et al., 2015).

## Transcription

Finally, few studies signified a role for Hfq in transcription. So far, how Hfq affects transcription remains unclear. Direct effects could exist as Fourier spectroscopy (FTIR) technique allowed to observe a partial unwinding of a DNA double helix at AT-rich regions by the Hfq protein (Geinguenaud et al., 2011). This opens the discussion on a possible role for this protein in the regulation of transcription initiation. A role for Hfq in the modulation of transcription elongation has also been proposed. According to the hypothesis, interaction of Hfq with a nascent transcript would help to overcome transcription pauses in order to prevent preliminary transcript release (Le Derout et al., 2010). Independently, indirect effects could also occur. Indeed, Hfq function may also be mediated by protein-protein contact and its interaction with RNA polymerase or Rho (a transcription termination factor) have been described (Sukhodolets and Garges, 2003; Rabhi et al., 2011). Transcriptional control by Hfq still remains largely unexplored.

## REFERENCES

- Arluison, V., Folichon, M., Marco, S., Derreumaux, P., Pellegrini, O., Seguin, J., et al. (2004). The C-terminal domain of *Escherichia coli* Hfq increases the stability of the hexamer. *Eur. J. Biochem.* 271, 1258–1265. doi: 10.1111/j.1432-1033.2004.04026.x
- Arluison, V., Mura, C., Guzman, M. R., Liquier, J., Pellegrini, O., Gingery, M., et al. (2006). Three-dimensional structures of fibrillar sm proteins: Hfq and other sm-like proteins. *J. Mol. Biol.* 356, 86–96. doi: 10.1016/j.jmb.2005.11.010
- Azam, T. A., Hiraga, S., and Ishihama, A. (2000). Two types of localization of the DNA-binding proteins within the *Escherichia coli* nucleoid. *Genes Cells* 5, 613–626. doi: 10.1046/j.1365-2443.2000.00350.x
- Azam, T. A., and Ishihama, A. (1999). Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. *J. Biol. Chem.* 274, 33105–33113. doi: 10.1074/jbc.274.46.33105

## CONCLUSIONS

Although the Hfq protein has been discovered as an RNA-interacting factor, and investigated in this light for many years, it also binds DNA and affects significantly the structure of this nucleic acid. Experimental evidence that Hfq is involved directly or indirectly in different DNA transactions exists, even if molecular mechanisms of these regulatory processes are still poorly understood. Nevertheless, Hfq-mediated control of ColE1-like plasmids appears to be specific, thus, particular mechanisms can be recognized in forthcoming studies. Regulation of transposition and transcription by Hfq might be less complicated than the function of this protein in the control of DNA replication. Regulations at both the post-transcriptional and the DNA-structuring levels allow bacteria to adapt to their environment, with important consequences for its physiology (including virulence). The physiological impacts of Hfq-DNA interactions *in vivo* thus need further investigations.

## AUTHOR CONTRIBUTIONS

GC drafted a part of chapter 2, and chapters 3.1 and 3.3, and participated in preparation of the final manuscript. AS wrote chapter 3.2. KK and AM participated in the general assembly of the text. WG participated in the biophysical description of Hfq. VA drafted chapter 2, prepared figures, and participated in preparation of the final version of the manuscript. GW elaborated the concept of the manuscript, drafted chapters 1 and 4, and prepared the final version of the paper. VA and GW contributed equally to this work.

## ACKNOWLEDGMENTS

This work was supported by the National Science Center (Poland) grant no. 2012/04/M/NZ1/00067 (GW), the POLONIUM program grant no. 33547NE, the joint research projects USPC/NUS RL2015-106 (VA) and by the Institut français de Pologne (KK). We thank PICT-Ibisa for access to TEM facilities.

- Cech, G. M., Pakula, B., Kamrowska, D., Wegrzyn, G., Arluison, V., and Szalewska-Palasz, A. (2014). Hfq protein deficiency in *Escherichia coli* affects ColE1-like but not lambda plasmid DNA replication. *Plasmid* 73, 10–15. doi: 10.1016/j.plasmid.2014.04.005
- Diestra, E., Cayrol, B., Arluison, V., and Risco, C. (2009). Cellular electron microscopy imaging reveals the localization of the Hfq protein close to the bacterial membrane. *PLoS ONE* 4:e8301. doi: 10.1371/journal.pone.0008301
- Dorman, C. J. (2009). Nucleoid-associated proteins and bacterial physiology. *Adv. Appl. Microbiol.* 67, 47–64. doi: 10.1016/S0065-2164(08)01002-2
- Ellis, M. J., and Haniford, D. B. (2016). Riboregulation of bacterial and archaeal transposition. *Wiley Interdiscip. Rev. RNA* 7, 382–398. doi: 10.1002/wrna.1341
- Ellis, M. J., Trussler, R. S., and Haniford, D. B. (2015). A cis-encoded sRNA, Hfq and mRNA secondary structure act independently to suppress IS200 transposition. *Nucleic Acids Res.* 43, 6511–6527. doi: 10.1093/nar/gkv584
- Fender, A., Elf, J., Hampel, K., Zimmermann, B., and Wagner, E. G. (2010). RNAs actively cycle on the Sm-like protein Hfq. *Genes Dev.* 24, 2621–2626. doi: 10.1101/gad.591310



- Folichon, M., Arluison, V., Pellegrini, O., Huntzinger, E., Regnier, P., and Hajnsdorf, E. (2003). The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. *Nucleic Acids Res.* 31, 7302–7310. doi: 10.1093/nar/gkg915
- Fortas, E., Piccirilli, F., Malabirade, A., Militello, V., Trépout, S., Marco, S., et al. (2015). New insight into the structure and function of Hfq C-terminus. *Biosci. Rep.* 35:e00190. doi: 10.1042/BSR20140128
- Franze De Fernandez, M. T., Hayward, W. S., and August, J. T. (1972). Bacterial proteins required for replication of phage Q $\beta$  ribonucleic acid. *J. Biol. Chem.* 247, 824–821.
- Geinguenaud, F., Calandrini, V., Teixeira, J., Mayer, C., Liquier, J., Lavelle, C., et al. (2011). Conformational transition of DNA bound to Hfq probed by infrared spectroscopy. *Phys. Chem. Chem. Phys.* 13, 1222–1229. doi: 10.1039/C0CP01084G
- Gottesman, S., and Storz, G. (2015). RNA reflections: converging on Hfq. *RNA* 21, 511–512. doi: 10.1261/rna.050047.115
- Gruber, S. (2014). Multilayer chromosome organization through DNA bending, bridging and extrusion. *Curr. Opin. Microbiol.* 22, 102–110. doi: 10.1016/j.mib.2014.09.018
- Guillier, M., Gottesman, S., and Storz, G. (2006). Modulating the outer membrane with small RNAs. *Genes Dev.* 20, 2338–2348. doi: 10.1101/gad.1457506
- Haniford, D. B., and Ellis, M. J. (2015). Transposons Tn10 and Tn5. *Microbiol. Spectr.* 3:MDNA3-0002-2014. doi: 10.1128/microbiolspec.mdna3-0002-2014
- Hussein, R., and Lim, H. N. (2011). Disruption of small RNA signaling caused by competition for Hfq. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1110–1115. doi: 10.1073/pnas.1010082108
- Jiang, K., Zhang, C., Guttula, D., Liu, F., Van Kan, J. A., Lavelle, C., et al. (2015). Effects of Hfq on the conformation and compaction of DNA. *Nucl. Acids Res.* 43, 4332–4341. doi: 10.1093/nar/gkv268
- Kajitani, M., and Ishihama, A. (1991). Identification and sequence determination of the host factor gene for bacteriophage Q $\beta$ . *Nucl. Acids Res.* 9, 1063–1066. doi: 10.1093/nar/19.5.1063
- Lease, R. A., and Belfort, M. (2000). Riboregulation by DsrA RNA: transactions for global economy. *Mol. Microbiol.* 38, 667–672. doi: 10.1046/j.1365-2958.2000.02162.x
- Lease, R. A., and Woodson, S. A. (2004). Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. *J. Mol. Biol.* 344, 1211–1223. doi: 10.1016/j.jmb.2004.10.006
- Le Derout, J., Boni, I. V., Regnier, P., and Hajnsdorf, E. (2010). Hfq affects mRNA levels independently of degradation. *BMC Mol. Biol.* 11:17. doi: 10.1186/1471-2199-11-17
- Link, T. M., Valentin-Hansen, P., and Brennan, R. G. (2009). Structure of *Escherichia coli* Hfq bound to polyribadenylate RNA. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19292–19297. doi: 10.1073/pnas.0908744106
- Lu, P., Wang, Y., Zhang, Y., Hu, Y., Thompson, K. M., and Chen, S. (2016). RpoS-dependent sRNA RgsA regulates Fis and AcpP in *Pseudomonas aeruginosa*. *Mol. Microbiol.* doi: 10.1111/mmi.13458. [Epub ahead of print].
- Ohniwa, R. L., Muchaku, H., Saito, S., Wada, C., and Morikawa, K. (2013). Atomic force microscopy analysis of the role of major DNA-binding proteins in organization of the nucleoid in *Escherichia coli*. *PLoS ONE* 8:e72954. doi: 10.1371/journal.pone.0072954
- Rabhi, M., Espeli, O., Schwartz, A., Cayrol, B., Rahmouni, A. R., Arluison, V., et al. (2011). The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *EMBO J.* 30, 2805–2816. doi: 10.1038/emboj.2011.192
- Ross, J. A., Ellis, M. J., Hossain, S., and Haniford, D. B. (2013). Hfq restructures RNA-IN and RNA-OUT and facilitates antisense pairing in the Tn10/IS10 system. *RNA* 19, 670–684. doi: 10.1261/rna.037747.112
- Ross, J. A., Trussler, R. S., Black, M. D., McEllan, C. R., and Haniford, D. B. (2014). Tn5 transposition in *Escherichia coli* is repressed by Hfq and activated by over-expression of the small non-coding RNA SgrS. *Mob. DNA* 5, 27. doi: 10.1186/s13100-014-0027-z
- Ross, J. A., Wardle, S. J., and Haniford, D. B. (2010). Tn10/IS10 transposition is downregulated at the level of transposase expression by the RNA-binding protein Hfq. *Mol. Microbiol.* 78, 607–621. doi: 10.1111/j.1365-2958.2010.07359.x
- Schu, D. J., Zhang, A., Gottesman, S., and Storz, G. (2015). Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. *EMBO J.* 34, 2557–2573. doi: 10.15252/embj.201591569
- Shapiro, L., Franze De Fernandez, M. T., and August, J. T. (1968). Resolution of two factors required in the Q $\beta$ -RNA polymerase reaction. *Nature* 220, 478–480. doi: 10.1038/220478a0
- Sobrero, P., and Valverde, C. (2012). The bacterial protein Hfq: much more than a mere RNA-binding factor. *Crit. Rev. Microbiol.* 38, 276–299. doi: 10.3109/1040841X.2012.664540
- Sukhodolets, M. V., and Garges, S. (2003). Interaction of *Escherichia coli* RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. *Biochemistry* 42, 8022–8034. doi: 10.1021/bi020638i
- Taghbalout, A., Yang, Q., and Arluison, V. (2014). The *Escherichia coli* RNA processing and degradation machinery is compartmentalized within an organized cellular network. *Biochem. J.* 458, 11–22. doi: 10.1042/BJ20131287
- Takada, A., Wachi, M., Kaidow, A., Takamura, M., and Nagai, K. (1997). DNA binding properties of the *hfq* gene product of *Escherichia coli*. *Biochem. Biophys. Res. Com.* 236, 576–579. doi: 10.1006/bbrc.1997.7013
- Talukder, A., and Ishihama, A. (2015). Growth phase dependent changes in the structure and protein composition of nucleoid in *Escherichia coli*. *Sci. China Life Sci.* 58, 902–911. doi: 10.1007/s11427-015-4898-0
- Tsui, H. C., Feng, G., and Winkler, M. E. (1997). Negative regulation of *mutS* and *mutH* repair gene expression by the Hfq and RpoS global regulators of *Escherichia coli* K-12. *J. Bacteriol.* 179, 7476–7487.
- Tsui, H. C., Leung, H. C., and Winkler, M. E. (1994). Characterization of broadly pleiotropic phenotypes caused by an *hfq* insertion mutation in *Escherichia coli* K-12. *Mol. Microbiol.* 13, 35–49. doi: 10.1111/j.1365-2958.1994.tb00400.x
- Updegrove, T. B., Correia, J. J., Galletto, R., Bujalowski, W., and Wartell, R. M. (2010). *E. coli* DNA associated with isolated Hfq interacts with Hfq's distal surface and C-terminal domain. *Biochim. Biophys. Acta* 1799, 588–596. doi: 10.1016/j.bbagr.2010.06.007
- Updegrove, T. B., Zhang, A., and Storz, G. (2016). Hfq: the flexible RNA matchmaker. *Curr. Opin. Microbiol.* 30, 133–138. doi: 10.1016/j.mib.2016.02.003
- Vogel, J., and Luisi, B. F. (2011). Hfq and its constellation of RNA. *Nat. Rev. Microbiol.* 9, 578–589. doi: 10.1038/nrmicro2615
- Wagner, E. G. (2013). Cycling of RNAs on Hfq. *RNA Biol.* 10, 619–626. doi: 10.4161/rna.24044
- Wang, W., Wang, L., Zou, Y., Zhang, J., Gong, Q., Wu, J., et al. (2011). Cooperation of *Escherichia coli* Hfq hexamers in DsrA binding. *Genes Dev.* 25, 2106–2117. doi: 10.1101/gad.16746011

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Cech, Szalewska-Palasz, Kubiak, Malabirade, Grange, Arluison and Węgrzyn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Keeping the Wolves at Bay: Antitoxins of Prokaryotic Type II Toxin-Antitoxin Systems

Wai Ting Chan<sup>1</sup>, Manuel Espinosa<sup>1\*</sup> and Chew Chieng Yeo<sup>2\*</sup>

<sup>1</sup> Molecular Microbiology and Infection Biology, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain, <sup>2</sup> Faculty of Medicine, Biomedical Research Centre, Universiti Sultan Zainal Abidin, Kuala Terengganu, Malaysia

## OPEN ACCESS

### Edited by:

Brian M. Baker,  
University of Notre Dame, USA

### Reviewed by:

Kurt Henry Piepenbrink,  
The University of Maryland, USA  
Andrew Benjamin Herr,  
Cincinnati Children's Hospital Medical  
Center, USA

### \*Correspondence:

Manuel Espinosa  
mespinosa@cib.csic.es;  
Chew Chieng Yeo  
chewchieng@gmail.com

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 21 December 2015

**Accepted:** 04 March 2016

**Published:** 22 March 2016

### Citation:

Chan WT, Espinosa M and Yeo CC  
(2016) Keeping the Wolves at Bay:  
Antitoxins of Prokaryotic Type II  
Toxin-Antitoxin Systems.  
Front. Mol. Biosci. 3:9.  
doi: 10.3389/fmolb.2016.00009

In their initial stages of discovery, prokaryotic toxin-antitoxin (TA) systems were confined to bacterial plasmids where they function to mediate the maintenance and stability of usually low- to medium-copy number plasmids through the post-segregational killing of any plasmid-free daughter cells that developed. Their eventual discovery as nearly ubiquitous and repetitive elements in bacterial chromosomes led to a wealth of knowledge and scientific debate as to their diversity and functionality in the prokaryotic lifestyle. Currently categorized into six different types designated types I–VI, type II TA systems are the best characterized. These generally comprised of two genes encoding a proteic toxin and its corresponding proteic antitoxin, respectively. Under normal growth conditions, the stable toxin is prevented from exerting its lethal effect through tight binding with the less stable antitoxin partner, forming a non-lethal TA protein complex. Besides binding with its cognate toxin, the antitoxin also plays a role in regulating the expression of the type II TA operon by binding to the operator site, thereby repressing transcription from the TA promoter. In most cases, full repression is observed in the presence of the TA complex as binding of the toxin enhances the DNA binding capability of the antitoxin. TA systems have been implicated in a gamut of prokaryotic cellular functions such as being mediators of programmed cell death as well as persistence or dormancy, biofilm formation, as defensive weapons against bacteriophage infections and as virulence factors in pathogenic bacteria. It is thus apparent that these antitoxins, as DNA-binding proteins, play an essential role in modulating the prokaryotic lifestyle whilst at the same time preventing the lethal action of the toxins under normal growth conditions, i.e., keeping the proverbial wolves at bay. In this review, we will cover the diversity and characteristics of various type II TA antitoxins. We shall also look into some interesting deviations from the canonical type II TA systems such as tripartite TA systems where the regulatory role is played by a third party protein and not the antitoxin, and a unique TA system encoding a single protein with both toxin as well as antitoxin domains.

**Keywords:** toxin-antitoxin, DNA-binding motifs, transcriptional repressor proteins, autoregulation, conditional cooperativity

**Abbreviations:** TA, toxin-antitoxin; HTH, helix-turn-helix; RHH, ribbon-helix-helix; FIS, factor for inversion stimulation.

## INTRODUCTION

The profusion of toxin-antitoxin (TA) genes among the realm of prokaryotes has sparked the interest of researchers to reveal the rationale of TA existence. One could hardly imagine the reason behind the finding that TA, which is mainly found in the genomes of bacteria and archaea, can be present up to 88 copies in *Mycobacterium tuberculosis*, although only 30 of them are functional (Ramage et al., 2009). In general (but not in all cases), a TA system is comprised of two genes, the antitoxin gene and its cognate toxin gene, which are located adjacent to each other. There are various modes of action by the toxin protein to exert its toxicity, but the most common ones involve inhibition of translation or replication, or targeting the cell wall synthesis of the host cells. TA systems, which have not been found in eukaryotes are, however, also able to poison eukaryotic cells because eukaryotes share common transcription and translation machineries with prokaryotes (Christensen et al., 2001; Pimentel et al., 2005; Nariya and Inouye, 2008; Amitai et al., 2009; Hurley and Woychik, 2009; Yamaguchi and Inouye, 2009; Agarwal et al., 2010; Dienemann et al., 2011; Castro-Roa et al., 2013; Germain et al., 2013). The product of the antitoxin gene, which can be either RNA or protein, is usually less stable compared to the toxin protein. Depending on the mechanism by which the antitoxin neutralizes the toxin, TAs have been categorized into six different types: (i) Type I, in which the antitoxin mRNA binds to its complementary toxin mRNA to prevent translation of the toxin gene; (ii) Type II, the antitoxin is a protein that forms a stable complex with the toxin protein and blocking the active site of the toxin under normal growth conditions; (iii) Type III, the antitoxin is an RNA with multiple tandem repeats that binds directly to the toxin protein rendering the toxin inactive; (iv) Type IV, the antitoxin protein does not bind to the toxin, but antagonize the toxin effect by competing for binding to the cellular target; and (v) Type V, the antitoxin protein is an RNase that cleaves directly its cognate toxin mRNA (Alonso et al., 2007; Hayes and Van Melder, 2011; Masuda et al., 2012; Cataudella et al., 2013; Unterholzner et al., 2013; Barbosa et al., 2015). A likely new type of TA system (a possible type VI) was recently discovered in *Caulobacter crescentus* where both the SocA antitoxin and SocB toxin are proteins, as in types II and IV TA systems. However, in this case, the SocA antitoxin functions as a ClpXP protease adaptor for the SocB toxin, promoting degradation of the toxin and thereby abolishing its lethality (Aakre et al., 2013; Markovski and Wickner, 2013). Thus, in a type VI TA system, the toxin is the unstable partner whereas in type II TA systems, the antitoxins are the labile partners due to their susceptibility to protease degradation. To date, TA systems belonging to types I and II are the most abundant in prokaryotes with type II TAs being the best characterized (Hayes and Van Melder, 2011; Unterholzner et al., 2013; Bertram and Schuster, 2014; Hayes and Kędzierska, 2014).

TA genes, which do not seem to be essential to the host cells (Van Melder and Saavedra De Bast, 2009; Van Melder, 2010), have been linked in countless ways to the lifestyle of the bacteria. The function of plasmid-encoded TAs has been commonly recognized as to stabilize the plasmid by a phenomenon denoted

as post-segregational killing of the daughter cells that do not inherit its parental plasmid (Jaffe et al., 1985; Gerdes et al., 1986) or “addiction,” as once the cells acquire the TA-encoded plasmid horizontally, the cells are no longer able to survive if they lost that plasmid (Lehnherr and Yarmolinsky, 1995; Hernández-Arriaga et al., 2014). Nevertheless, the chromosomally-encoded TA genes are known to have broader impact to the host cells. Since the consequences of toxin effect can be bactericidal or bacteriostatic, chromosomally-encoded TAs have been related to altruistic cell death or stress response when the cells are under unfavorable circumstances. Altruistic cell death adopted the idea of bacterial cells living as a community, and when under stressful states like scarcity in nutrition, some of the cells will “sacrifice” themselves via triggering of their TA systems, subsequently lysing and releasing nutrients for the rest of their populations’ need (Aizenman et al., 1996; Engelberg-Kulka and Glaser, 1999). Of course one could argue that instead of altruism, cannibalism (e.g., in *Bacillus subtilis*; González-Pastor, 2011) or fratricide (e.g., in *Streptococcus pneumoniae*; Eldholm et al., 2009) would more likely have happened for bacteria, which are the more primitive life forms. As activation of most of the toxins leads to cell stasis, the postulation of TAs involving in stress response is more widely accepted (Gerdes et al., 2005). The stress response mediated by TAs was well-demonstrated by the persistence phenomenon observed in *Escherichia coli* and other bacteria. Persister cells refer to a small portion of cells among isogenic antibiotic-sensitive bacterial population that stochastically switch to slow growth (or a quasi-dormant state) leading to multidrug tolerance when exposed to antibiotics (Lewis, 2010). In the persister cells, the increased levels of the signaling nucleotide (p)ppGpp (guanosine pentaphosphate/tetraphosphate) trigger slow growth by activating certain TAs through a regulatory cascade, which is dependent on Lon protease and inorganic polyphosphate (Maisonneuve et al., 2013). There are also other studies that demonstrated the involvement of chromosomally-encoded TAs in biofilm formation, increased survival rate, colonization of new niches, phage abortive infection, maintenance of bacterial mobilomes, virulence of pathogenic bacteria, and as anti-addiction modules (Christensen et al., 2001; Rowe-Magnus et al., 2003; Szekeres et al., 2007; Saavedra De Bast et al., 2008; Harrison et al., 2009; Mine et al., 2009; Hallez et al., 2010; Makarova et al., 2011; Armalyte et al., 2012; Norton and Mulvey, 2012; Cheng et al., 2014). Thus, the diversity of TA systems in prokaryotes is reflected in their diversity of cellular function.

## ANTITOXINS NEUTRALIZE THE LETHALITY OF THEIR COGNATE TOXINS AND ALSO FUNCTION AS DNA-BINDING PROTEINS THAT MODULATE THE PROKARYOTIC LIFESTYLE

Type II TA systems are so far, the best studied of the TA families. Like other systems, Type II TAs are usually comprised of two genes with the antitoxin gene preceding the toxin gene, and with both genes co-transcribed from a single promoter

located upstream of the antitoxin gene (Leplae et al., 2011). In general, the two TA genes overlap by a few nucleotides, indicating coupled translation of the two genes. Under normal conditions, the antitoxin protein binds avidly to the toxin protein to safeguard its harmfulness to the cells, as it has also been shown by determination of the three dimensional structures of TA complexes. However, because the antitoxin protein seems to be structurally partially folded (Cherny et al., 2005), it is thus more fragile and susceptible to the degradation by the host proteases (e.g., Lon or Clp); antitoxin cleavage would release the more stable toxin protein to act on its cellular target. Hence, the antitoxin protein needs to be constantly replenished in order to avoid a surfeit of toxin proteins. This explains the organization of the majority of type II TA operons: the antitoxin gene preceding the toxin gene would enable the antitoxin to be transcribed and translated before synthesis of the toxin starts.

Toxins target various cellular structures and essential molecular processes and thus, hinder cellular activities (Hayes and Van Melderen, 2011; Hayes and Kędzierska, 2014). Majority of class II toxins that have been examined to date act as endoribonucleases and thus, inhibit the translation machinery (Yamaguchi and Inouye, 2009; Yamaguchi et al., 2011). Some of these endoribonucleases, such as the MazF toxin, cleave free mRNA in a sequence-dependent manner (Zhang et al., 2003), whereas other endoribonuclease toxins, such as RelE, target mRNA associated with ribosomes (Pedersen et al., 2003). Some type II toxins interfere with the translation process by other means such as the cleavage of initiator tRNA by the VapC toxins of *Shigella flexneri* and *Salmonella enterica* serovar Typhimurium (Winther and Gerdes, 2011), or phosphorylation of elongation factor EF-Tu by the *E. coli*-encoded HipA toxin and the bacteriophage P1-encoded Doc toxin (Schumacher et al., 2009; Cruz et al., 2014). On the other hand, certain type II toxins (such as CcdB and ParE) affect DNA replication by direct inhibition of gyrase activity, which is required to relieve supercoiling that occurs ahead of the replication fork (Bernard and Couturier, 1992; Yuan et al., 2010). The  $\zeta$  and PezT toxins blocks cell wall synthesis by phosphorylating peptidoglycan precursors, thereby inhibiting the first step in peptidoglycan synthesis (Mutschler et al., 2011).

Type II antitoxins abrogate the lethality of their cognate toxins through a toxin-binding domain, which is usually natively unstructured until formation of the toxin-antitoxin complex. One of the hallmarks of toxin inactivation is a direct interaction whereby the antitoxin wraps around the toxin and inhibits toxin activity by blocking or masking the toxin active site (Blower et al., 2011; Bøggild et al., 2012; Schureck et al., 2014). For example, the *E. coli*-encoded MazE antitoxin wraps across the surface of the MazF toxin, blocking the active site as well as forcing out the S1-S2 loop that stabilizes the catalytic triad of the toxin (Kamada et al., 2003). In the case of the  $\zeta$  and PezT toxins, inactivation is due to the respective cognate  $\epsilon$  or PezA antitoxin sterically hindering ATP/GTP binding within the toxin (Meinhart et al., 2003; Khoo et al., 2007). The *Caulobacter crescentus*-encoded ParD antitoxin inhibits its cognate ParE toxin by binding as a dimer to a conserved complementary patch at the C-terminus of ParE without inducing conformational changes (Dalton and Crosson, 2010). In contrast, binding of the *E. coli*-encoded

RelB antitoxin inhibits the RelE toxin by perturbing the toxin structure, specifically through displacement of a flexible  $\alpha$ -helix at the C-terminal that contains the Tyr-87 residue essential for RelE activity (Li et al., 2009). Similarly, the VapB5 antitoxin of *M. tuberculosis* act to prevent its cognate VapC5 toxin from binding  $Mg^{2+}$  as a co-factor by reorienting the side-chain of VapC5 Arg-112, locking the VapC5 Glu-57 residue in an unfavorable conformation to bind  $Mg^{2+}$  (Miallau et al., 2009). In some TA complexes, the antitoxin binds to the toxin but does not occlude the active site of the toxin. The *E. coli*-encoded HipB antitoxin binds far from the active site of its cognate HipA toxin and functions to inhibit the toxin activity by locking the toxin in an inactive open conformation (Schumacher et al., 2009). Similarly, the HigA antitoxin of *Proteus vulgaris* only makes two regions of contact with the HigB toxin, both of which are distant from the HigB active site. The HigB toxin functions as a ribosome-dependent endoribonuclease and it was proposed that binding of HigA sterically inhibits HigB from interacting with mRNA in the A site of the ribosome (Schureck et al., 2014).

The antitoxin protein is not only the nemesis of its cognate toxin, but also the key factor that regulates transcription of the TA operon. The antitoxin is generally a DNA-binding protein that binds, albeit usually weakly, to the operator of the operon to repress its own transcription; whereas the toxin protein, which does not bind to the DNA upstream of the operon, usually serves as a co-repressor, by binding to the antitoxin protein and changing the conformation of the antitoxin-DNA complex, which lead to further repression (Bertram and Schuster, 2014; Hayes and Kędzierska, 2014; Kędzierska and Hayes, 2016). In some cases, the molar ratio of antitoxin and toxin has great impact on the formation of the TA complex in terms of stoichiometries (Gelens et al., 2013). More importantly, TA complexes with different stoichiometries have different affinity to the binding of the operator (Overgaard et al., 2008; Garcia-Pino et al., 2010; see below). Therefore, the ratio of the antitoxin and toxin is very crucial to the regulation of the transcription of the TA operon and to determine the lifestyle and fate of the bacterial host.

## DNA-BINDING DOMAINS FOUND IN ANTITOXINS: HELIX-TURN-HELIX (HTH), RIBBON-HELIX-HELIX (RHH) FOLD, AND SpoVT/AbrB-TYPE

Concerning the three-dimensional structure of the TAs, several of them have been determined, either the antitoxin alone or in complex with the cognate toxin. In most cases, the antitoxin protein appears to be divided into two domains: the N-terminal domain usually comprises the DNA binding region, whereas the C-terminal domain is generally involved in the interaction with the cognate toxin to offset its toxicity. These two domains may be interconnected by a flexible small loop or hinge-like region. Determination of the crystal structure of the TA complexes (mostly without DNA) has been achieved for an increasing number of them. In general, the structure of the DNA-binding domains of the antitoxins can be grouped into three

different types, namely helix-turn-helix (HTH), ribbon-helix-helix (RHH), and SpoVT/AbrB-type (Table 1; Figure 1). The HTH motif consists of around 20 amino acid residues distributed into two  $\alpha$ -helices separated by a short turn, generally mediated by a Gly residue (Figure 1A). The second helix of the HTH motif (also termed “the reading head”) recognizes and binds to the target DNA *via* a number of hydrogen bonds and hydrophobic interactions, which occur between specific side chains of the protein and the exposed bases and thymine methyl groups within the major groove of the DNA, whereas the first helix, and sometimes a third one, helps to stabilize the structure of the motif (Brennan and Matthews, 1989). The HTH motif has been reported in a number of prokaryotic DNA repressor proteins as well as in eukaryotes (Brennan and Matthews, 1989). Examples of the existence of the HTH motif in solved structures of antitoxins include PezA (Khoo et al., 2007) as well as HigA (Schureck et al., 2014). Another example of antitoxin harboring the HTH motif is MsqA, although in this case the motif is present at the C-terminal region of the protein; the N-terminal region having a Zn-binding domain involved in the interaction with its cognate toxin MsqR (Brown et al., 2009).

The RHH proteins have been found mostly in prokaryotes. These structures are arranged as two antiparallel  $\beta$ -strands that generate a ribbon (Figure 1B); each strand comes from one of two protein monomers and they are involved both in dimer formation and in specific interactions with the DNA bases in the antitoxin DNA target. In the simplest form, like the transcriptional repressor protein CopG (45 residues per protomer), or the *Salmonella* phage P22 Arc repressor (53 residues per protomer) the ribbon participates in DNA recognition and in the dimerization process, so that the proteins would be mostly in a disordered state if it were a monomer. However, no monomers of the protein seem to exist, and mutational analyses indicated that dimerization and folding could be considered as part of the same process, and the proteins would only exist as dimers (Milla et al., 1995; Gomis-Rüth et al., 1998). Perhaps, and lacking further information on the structure of other antitoxins, the RHH motif seems to be the most common structural motif in antitoxins, as it is present in CcdA (Madl et al., 2006), ParD (Oberer et al., 2007), RelB (Bøggild et al., 2012), DinJ (Liang et al., 2014), FitA (Mattison et al., 2006), and VapB (Min et al., 2012).

The number of antitoxins with a SpoVT/AbrB-type domain is also steadily increasing. They all share similarities to the transcriptional regulator AbrB, found in the Gram-positive bacterium *B. subtilis* and that is involved in the regulation of many genes. The structure of the DNA-binding domain of AbrB (Figure 1C) revealed the presence of a specific domain, in which two molecules (each having two  $\beta$ -hairpins) dimerizes to generate a so-called layered “ $\beta$ -sandwich.” A similar structure has been reported for the *S. flexneri* VapBC TA pair, in which four N-terminal antitoxin VapB domains generate two DNA-binding domains; each of these domains is constructed by a three-stranded antiparallel  $\beta$ -sheets, and a four-stranded antiparallel  $\beta$ -sheet. These arrangements form a strand-switched dimer interface in which the two  $\beta$ -sheets are tightly packed against each other, thus generating the DNA-binding domain (Dienemann

et al., 2011). Similar to VapB, but exhibiting a simpler structure is the MazE antitoxin (Kamada et al., 2003; Bobay et al., 2005), which, in turn has structural homology to the well-characterized Kis antitoxin (Kamphuis et al., 2007a,b).

The DNA binding targets of the antitoxin proteins are usually perfect or imperfect palindromic sequences (Khoo et al., 2007; Chan et al., 2011) that overlap with all or part of the promoter region; thus, binding of the antitoxin to its target would thwart the binding of the host RNA polymerase to the promoter resulting in transcription inhibition (see below).

## AUTOREGULATION AS A PARADIGM OF TYPE II TA LOCI: STRUCTURE AND FUNCTION OF THE MazE ANTITOXIN

MazEF is the first chromosomally-encoded TA discovered in *E. coli* (Aizenman et al., 1996). The *mazEF* operon is located in the *E. coli* *rel* locus, downstream of the *relA* gene. Expression of *mazEF* was shown to be regulated by the cellular levels of ppGpp, the product of the RelA protein. During amino acid starvation, increased levels of the alarmone guanosine tetraphosphate (ppGpp) lead to inhibition of transcription of *mazEF* and triggers programmed cell death (Aizenman et al., 1996). The MazF toxin is an endoribonuclease that cleaves cellular mRNA at the specific sequence, 5'-ACA-3' (Zhang et al., 2003). Interestingly, MazF also cleaves ACA sites that are close to the region upstream of the AUG start site of some specific mRNAs, thus generating a pool of leaderless mRNAs. In addition, MazF also targets 16S rRNA within 30S ribosomal subunits at the decoding center, therefore removing 43 nucleotides from the 3' terminus that comprises the anti-Shine-Dalgarno. As a result, a modified translation machinery is formed to selectively translate the leaderless mRNAs to adapt to the stress condition (Vesper et al., 2011). The antitoxin MazE harbors two domains: (i) the N-terminus consists of a SpoVT/AbrB-type domain with a swapped-hairpin  $\beta$ -strand motif that binds to the operator to negatively autoregulate its transcription, and (ii) the C-terminal domain is intrinsically disordered and upon binding to MazF toxin will form an extended conformation that is more stable and protected from the host protease degradation (Kamada et al., 2003; Loris et al., 2003). The C-terminal tail of MazE is not directly involved in DNA binding and remained disordered upon interaction of the N-terminal domain with the DNA (Vesper et al., 2011).

Along the same operon downstream of *mazEF* is another open reading frame called *mazG*, which is co-transcribed with *mazEF*. MazG is a pyrophosphohydrolase that hydrolyses dNTPs and thus depletes ppGpp. However, MazG activity is also inhibited by the MazEF complex (Gross et al., 2006). Therefore, during amino acid starvation, in addition to inhibition of *mazEFG* transcription due to increased ppGpp, degradation of MazE will inactivate the inhibition activity of the MazEF complex against the existing MazG. Activation of MazG will deplete ppGpp levels, which in turn causes re-transcription of *mazEF* to replenish MazE, which consequently triggers the cells to emerge from their dormant state (Gross et al., 2006).



**TABLE 1 | Solved type II toxin-antitoxin structures grouped according to the DNA-binding domain of the antitoxins.**

DNA-binding domain	Antitoxin	Toxin	Host organism	Toxin-antitoxin complex stoichiometries	References
HTH	HipB	HipA	<i>Escherichia coli</i>	HipB <sub>2</sub> HipA <sub>2</sub>	Schumacher et al., 2009
	HipB <sub>SO</sub>	HipA <sub>SO</sub>	<i>Shewanella oneidensis</i>	(HipB <sub>SO</sub> ) <sub>2</sub> (HipA <sub>SO</sub> ) <sub>2</sub>	Wen et al., 2014
	MqsA	MqsR	<i>E. coli</i>	MqsR-MqsA <sub>2</sub> -MqsR	Brown et al., 2009, 2011
	HigA	HigB	<i>E. coli</i>	unknown <sup>a</sup>	Arbing et al., 2010
	HigA	HigB	<i>Proteus vulgaris</i>	HigA <sub>2</sub> HigB <sub>2</sub>	Schureck et al., 2014
	HigA2	HigB2	<i>Vibrio cholerae</i>	Unknown	Hadži et al., 2013
	PezA	PezT	<i>Streptococcus pneumoniae</i>	PezA <sub>2</sub> PezT <sub>2</sub>	Khoo et al., 2007
RHH	RelB	RelE	<i>E. coli</i>	RelB <sub>2</sub> RelE <sub>2</sub>	Bøggild et al., 2012
	RelB	RelE	<i>Methanococcus jannaschii</i>	RelB <sub>2</sub> RelE <sub>2</sub>	Francuski and Saenger, 2009
	DinJ	YafQ	<i>E. coli</i>	DinJ <sub>2</sub> YafQ <sub>2</sub>	Liang et al., 2014; Ruangprasert et al., 2014
	ParD	ParE	<i>E. coli</i> plasmid RK2	ParD <sub>2</sub> <sup>b</sup>	Oberer et al., 2007
	ParD1	ParE1	<i>Caulobacter crescentus</i>	(ParD1) <sub>2</sub> (ParE1) <sub>2</sub>	Dalton and Crosson, 2010
	CcdA	CcdB	<i>E. coli</i> F plasmid	(CcdA <sup>37–72</sup> ) <sub>2</sub> (CcdB) <sub>2</sub> (CcdA <sup>37–72</sup> ) <sub>2</sub> (CcdB) <sub>2</sub>	Madl et al., 2006; De Jonge et al., 2009
	HicB3	HicA3	<i>Yersinia pestis</i>	(HicB3) <sub>4</sub> (HicA3) <sub>2</sub>	Bibi-Triki et al., 2014
	FitA	FitB	<i>Neisseria gonorrhoeae</i>	(FitA-FitB) <sub>4</sub>	Mattison et al., 2006
	VapB3	VapC3	<i>Mycobacterium tuberculosis</i>	(VapB3) <sub>2</sub> (VapC3) <sub>2</sub>	Min et al., 2012
	VapB5 <sup>c</sup>	VapC5	<i>M. tuberculosis</i>	(VapB5 <sup>53–86</sup> )(VapC5)	Miallau et al., 2009
	VapB30	VapC30	<i>M. tuberculosis</i>	(VapB30) <sub>2</sub> (VapC30) <sub>2</sub>	Lee et al., 2015
	ω (regulator for ε-ζ, not antitoxin)		<i>Streptococcus pyogenes</i> pSM19035 tripartite TA system ω-ε-ζ	ε <sub>2</sub> ζ <sub>2</sub> for the TA complex; ω <sub>2</sub> for the regulator protein	Murayama et al., 2001; Meinhart et al., 2003
SpoVT/AbrB	MazE	MazF	<i>E. coli</i>	MazF <sub>2</sub> -MazE <sub>2</sub> -MazF <sub>2</sub>	Kamada et al., 2003; Loris et al., 2003; Zorzini et al., 2015
	VapB2	VapC2	<i>Rickettsia felis</i>	(VapC2) <sub>4</sub> (VapB2) <sub>2</sub> (VapC2) <sub>4</sub> (VapC2) <sub>4</sub>	Maté et al., 2012
	VapB	VapC	<i>Shigella flexneri</i>	VapB <sub>4</sub> VapC <sub>4</sub>	Dienemann et al., 2011
Phd/YefM <sup>e</sup>	Phd	Doc	<i>E. coli</i> phage P1	Doc-Phd <sub>2</sub> -Doc	Arbing et al., 2010; Garcia-Pino et al., 2010
	YefM	YoeB	<i>E. coli</i>	YefM <sub>2</sub> YoeB	Kamada and Hanaoka, 2005
	YefM	YoeB	<i>M. tuberculosis</i>	unknown <sup>d</sup>	Kumar et al., 2008
Unknown <sup>f</sup>	aRelB	aRelE	<i>Pyrococcus horikoshii</i> OT3	(aRelB) <sub>2</sub> (aRelE) <sub>2</sub>	Takagi et al., 2005
Unknown <sup>g</sup>	VapB15	VapC15	<i>M. tuberculosis</i>	(VapB15) <sub>2</sub> (VapC15) <sub>2</sub> (VapB15)(VapC15) <sub>2</sub>	Das et al., 2014

<sup>a</sup>Structure was only available for the HigA antitoxin (Arbing et al., 2010).

<sup>b</sup>Structure only solved for ParD in solution by NMR (Oberer et al., 2007).

<sup>c</sup>N-terminal region of VapB5 could not be modeled but predicted to be RHH motif (Miallau et al., 2009).

<sup>d</sup>TA complex possibly YefM<sub>2</sub>YoeB; only YefM was crystalized (Kumar et al., 2008).

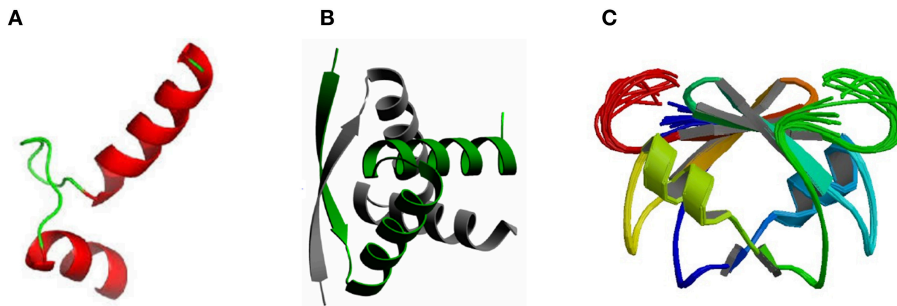
<sup>e</sup>YefM was found to share structural similarity with the Phd antitoxin with strong conservation of the N-terminal DNA-binding domain, which are thus classified as having a Phd/YefM-like fold (Arbing et al., 2010).

<sup>f</sup>DNA-binding domain unclear, potentially leucine zipper dimerization with N-terminal basic residues used for DNA recognition (Takagi et al., 2005).

<sup>g</sup>N-terminal residues of VapB15 could not be modeled into the electron density (Das et al., 2014).

Two promoters, which are located 13 nucleotides apart, have been identified upstream of the *mazEFG* operon (**Figure 2A**). The P<sub>2</sub> promoter is about 10-fold stronger than the P<sub>3</sub> promoter (Marianovsky et al., 2001). Expression of both promoters is repressed by MazE and highly repressed with the MazEF complex. Within the promoters lies an unusual fragment termed the “alternating palindrome.” This alternating palindrome, which

is the operator of *mazEFG*, could exist in one of two alternative states: its middle part designated “a,” complements either the downstream fragment “b” or upstream fragment “c” (**Figure 2A**). Binding of the MazEF complex to either arm of this alternating palindrome will strongly repress the transcription of the *mazEF* operon. The numerous mutations that were introduced into this alternating palindrome did not affect the binding efficiency of



**FIGURE 1 | Three-dimensional structures showing the most frequent DNA-binding domains found in prokaryotic type II antitoxin proteins.** The most frequent DNA-binding domains found in type II antitoxins include: **(A)** the HTH-motif (of which the smallest structural motif is shown) that has two  $\alpha$ -helices (red) connected by a small loop (green); **(B)** the RHH folding motif, in which the minimal structure (as the one depicted corresponding to the CopG transcriptional repressor) is generated by two antiparallel  $\beta$ -strands (arrows, with arrowheads pointing to the C-terminal part of the protomer) that generate a ribbon; each strand comes from one of two protein monomers and they are involved both in dimer formation and in specific interactions with the DNA bases in the antitoxin DNA target (adapted from Gomis-Rüth et al., 1998), and **(C)** the SpoVT/Abr DNA binding motif in which the dimeric molecules are constructed by three- and four-stranded antiparallel  $\beta$ -sheets (upper part of the molecule) that are tightly packed, generating the DNA-binding domain. Loops keeping the outer part of the molecules are indicated by light green and red colors.

the MazEF complex, suggesting that the secondary structure of this regulatory region is more important than its DNA sequence (Marianovsky et al., 2001). MazE has higher binding affinity for fragment “a” than “b” or “c”.

Determination of the three-dimensional structure showed that the MazE homodimer binds into the major groove of DNA fragment “a,” involving the side-chains of residues Trp-9, Asn-11, and Arg-16 for the main interactions with the oligonucleotide (**Figure 2B**; Zorzini et al., 2015). Mobility shift assay with titration of MazF showed that MazF could increase the affinity of MazE for a single operator site where the concentration of MazE itself is not sufficient to cause a band-shift. Superposition of MazE-DNA complex on the crystal structure of the MazE-MazF complex demonstrated that the interaction between DNA and protein increased through the flanking basic regions of the MazF homodimer. This indicates that the augmentation of DNA binding by MazF is due to cooperative binding of the antitoxin and toxin to the DNA instead of an allosteric effect. However, reduced band-shift corresponding to the complex was observed after a peak with increasing MazF, and the affinity of MazE for binding to the “a” fragment is abolished at very high ratio of MazF:MazE (Zorzini et al., 2015). This resembles the conditional cooperativity phenomenon that was observed in other TA systems like *ccdAB*, *phd-doc*, and *relBE* whereby the expression of the TA operon is modulated by ratios of antitoxin:toxin (Overgaard et al., 2008; De Jonge et al., 2009, 2010; Garcia-Pino et al., 2010).

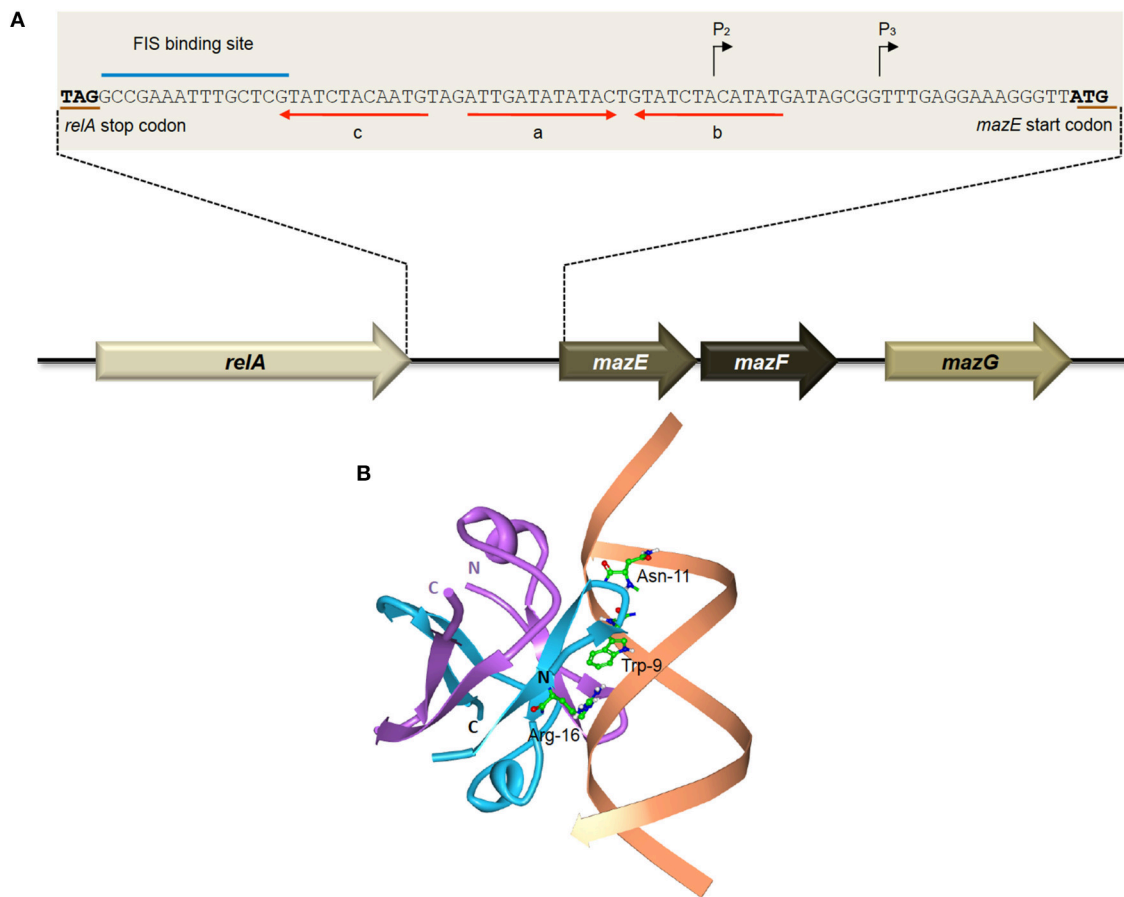
Besides having two promoters and an unusual alternating palindrome as operator site, the regulation of *mazEF* is also governed by another positive regulation mechanism. Further upstream of the alternating palindrome is a binding site for the factor for inversion stimulation (FIS), which positively regulates the transcription of *mazEF* operon (Marianovsky et al., 2001). FIS is a homodimer that binds and introduces bends in the DNA, thereby increasing the binding efficiency of RNA polymerase (Pan et al., 1996). The cellular level of FIS varies (up to 100-fold), depending upon the growth phase and nutritional conditions of

the cells. The concentrations of FIS are highly elevated in the early exponential phase but sharply declined toward the stationary phase (Marianovsky et al., 2001), indicating positive regulation of *mazEF* is maximal at rich medium during exponential phase. Thus, the complex regulatory mechanism which combines two promoters, alternating palindromes, the FIS-binding activation site, concentrations of ppGpp and MazG, as well as the ratio and the co-operative binding activities of the MazE and MazF to the operator enables the expression of *mazEF* to become more dynamic and to ensure a prompt response to cope with various stresses or changes in the environment (Marianovsky et al., 2001).

An interesting dimension to the regulation of the MazF toxin was reported recently whereby infection of *E. coli* with bacteriophage T4 led to the addition of an ADP-ribosyl group to MazF (Alawneh et al., 2016). This chemical modification of MazF was catalyzed by phage T4-encoded Alt ADP-ribosyltransferase which transfers an ADP-ribosyl group from nicotinamide adenine dinucleotide ( $\beta$ -NAD<sup>+</sup>) to the Arg-4 residue of MazF, resulting in partial reduction of MazF cleavage activity *in vitro*. This inferred that phage T4 may harbor a unique antitoxin to inactivate MazF during T4 infection and MazF could function as an anti-phage mechanism in its *E. coli* host (Alawneh et al., 2016; Otsuka, 2016). The biological significance of the T4-dependent ADP-ribosylation of MazF and its effects on the existing *mazEF* regulatory circuit awaits further investigations.

## REGULATION VIA CONDITIONAL COOPERATIVITY OF THE *phd-doc*, *relBE*, AND *kis-kid* LOCI

Phd-Doc is a TA system found on bacteriophage P1 (Lehnherr et al., 1993). The regulation of the *phd-doc* TA operon relies on the stoichiometries of the Phd antitoxin and the Doc toxin, which is a phenomenon called conditional cooperativity, as mentioned above. Like other TAs, the Phd antitoxin has an intrinsically disordered C-terminus that forms an  $\alpha$ -helix upon binding to



**FIGURE 2 | The *E. coli*-encoded MazEF TA system. (A)** Genetic organization of the *E. coli* *mazEF* operon, and the regulatory elements on the *mazEF* promoter. Black arrows denote the transcriptional start sites of promoters P2 and P3. The stop codon of RelA and the start codon of MazE are underlined in brown. The FIS binding site is indicated with a blue line. Alternating palindromic regions “c-a” or “a-b” are indicated with red arrows. Adapted and modified from Marianovsky et al. (2001). **(B)** Structure of the MazE<sup>1–50</sup> antitoxin homodimer-DNA complex (PDB accession: 2MRU). The MazE<sup>1–50</sup> homodimer is indicated in blue and purple with the operator DNA indicated in orange. The N- and C-termini of the two MazE<sup>1–50</sup> units are as labeled. The key amino acid residues of MazE that are involved in binding to the major groove of the double-stranded “a” operator DNA, i.e., Trp-9, Asn-11, and Arg-16 (Zorzini et al., 2015), are shown for one of the MazE monomers (blue).

the Doc toxin (Garcia-Pino et al., 2008). The N-terminal domain of the Phd antitoxin is a dimerization domain that binds to the DNA operator to repress *phd-doc* expression. Doc toxin, which impedes translation by phosphorylating the conserved Thr-382 residue on elongation factor EF-Tu (Castro-Roa et al., 2013; Cruz et al., 2014), can also serve as a corepressor or derepressor depending on the molar ratio of both Doc and Phd proteins (Garcia-Pino et al., 2010). A monomeric Doc toxin has two binding sites that are able to interact with two Phd dimers, with different affinities, bridging the Phd dimers to bind more avidly to the operator. However, saturation of Doc toxin will be in favor of the high-affinity sites (H sites), outcompeting the low-affinity sites (L-sites) by Phd. This results in the restructuring of the repressor-corepressor complex to an alternative non-repressing Doc-Phd<sub>2</sub>-Doc complex, which cannot bind to the operator DNA due to steric reasons (Liu et al., 2008; Arbing et al., 2010; Garcia-Pino et al., 2010). Thus, the stoichiometry of Phd:Doc complex is important in modulating the regulation of *phd-doc* operon.

The *relBE* operon is one of the most prevalent and best-characterized TA system that was originally discovered on the chromosome of *E. coli* (Gotfredsen and Gerdes, 1998). The RelE toxin does not target free mRNA but cleaves mRNA in the ribosomal A site with codon specificity (Christensen and Gerdes, 2003). The RelB antitoxin neutralizes the toxic activity of RelE by displacing the  $\alpha 4$  helix, thereby disrupting the geometry of the critical catalytic residues of the free RelE structure. RelB dimers bind to the operator through a RHH motif to autoregulate transcription (Overgaard et al., 2009). However, the affinity of RelB binding to DNA is relatively low, and addition of RelE up to a ratio of 2 RelB: 1 RelE drastically enhanced the binding affinity (Christensen-Dalsgaard et al., 2008; Overgaard et al., 2008). The RelB<sub>2</sub>RelE heterotrimer complexes bind strongly and cooperatively to the promoter to repress transcription. When RelE is in excess, an unusual V-shaped structure is formed, with two RelE bound at the distant ends of the RelB dimer. These heterotetramer complexes will clash when two RelB dimerization domains bind adjacently to the DNA, which leads to the complex

falling-off from the operator DNA and derepressing transcription (Bøggild et al., 2012). The destabilization of the RelB<sub>2</sub>RelE complex from the DNA can be due to “stripping,” in which the excessive free toxin molecules invade the RelB<sub>2</sub>RelE heterotrimer complex; or the bulk formation of RelB<sub>2</sub>RelE<sub>2</sub> heterotetramer complexes that sequester the heterotrimer complex (Cataudella et al., 2012). During normal cellular growth, *relB* has higher rate of translation than *relE*, leading to tenfold more RelB than RelE protein molecules (Overgaard et al., 2009). Binding of the RelB<sub>2</sub>RelE heterotrimer will repress transcription of *relBE* to a minimal level. When cells undergo nutritional stress, since the lifetime of RelB is 10-fold shorter than RelE (Overgaard et al., 2009), the labile RelB will be degraded more rapidly by Lon proteases and this subsequently increases the RelE:RelB ratio. Consequently, more RelB<sub>2</sub>RelE<sub>2</sub> heterotetrameric complexes are formed that eventually derepress the repression of the *relBE* operon to replenish RelB levels in the cell. This conditional cooperativity of RelBE has also been shown to facilitate the fast recovery of cells from RelE-mediated reduction in translation when the nutritional stress is removed (Cataudella et al., 2012).

In the case of the RelBE2*Spn* operon from *S. pneumoniae*, the interaction of the two proteins with the DNA target was approached by means of band-shift, analytical ultracentrifugation, and native mass spectroscopy. The results led to the conclusion that the stoichiometry of the RelBE2*Spn* antitoxin in complex with its DNA target and of the RelBE2*Spn* protein-protein complex was compatible with a heterohexamer composed of four antitoxin and two toxin protein molecules, in both conditions: protein-protein and protein-DNA complexes (Moreno-Córdoba et al., 2012).

The *parD* operon of plasmid R1 from *E. coli* encodes the Kis-Kid TA (Bravo et al., 1988). Kid toxin is a ribonuclease that preferentially cleaves single stranded RNA at the 5' of the adenosine residue of sequence 5'-UA(A/C)-3'. However, cleavage at 3' of the adenosine residue on double stranded RNA was also evident (Pimentel et al., 2005; Kamphuis et al., 2006). Besides hindering the toxic effect of Kid toxin, Kis antitoxin is also a weak repressor that binds to its own promoter to regulate the transcription of *parD*. Like other typical TA, Kid toxin does not bind to the promoter but acts as a co-repressor. There are two binding regions where Kis dimers, but not monomers, preferentially bind to region I compared to region II (Figure 3). Region I harbors a perfect palindrome that overlaps the -10 consensus region of the promoter, whereas region II is an imperfect palindrome that is located upstream of the -35 sequence (Kamphuis et al., 2007b). Differential molar ratios of Kis and Kid can result in multiple complexes of Kis-Kid with different stoichiometries and oligomeric states (Monti et al., 2007). When Kid toxin is in excess, Kid<sub>2</sub>-Kis<sub>2</sub>-Kid<sub>2</sub> hexamer, which has weak affinity to *parD* DNA, is most abundant. Conversely, when the Kis antitoxin equals or exceeds the concentration of the Kid toxin, strong cooperative effect will form between *parD* DNA and Kid<sub>2</sub>-Kis<sub>2</sub>-Kid<sub>2</sub>-Kis<sub>2</sub> octamers. The Kis-Kid octameric complex can bind to the two half-sites of *parD* DNA region I and II with two dimers of the Kis antitoxin. However, the Kis-Kid hexamer can only bind to the two half-sites using one dimer, which thus explains its weak affinity (Figure 3

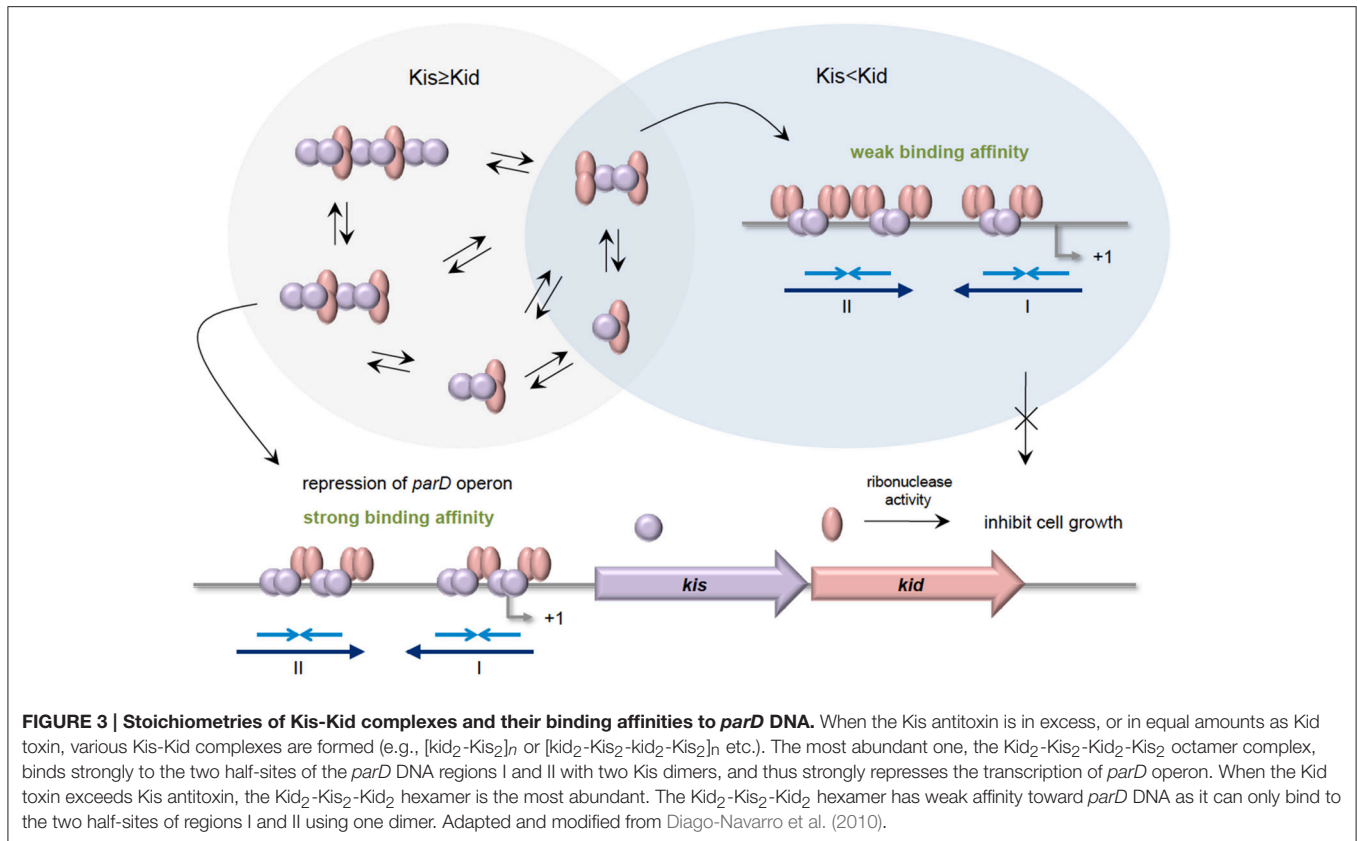
Kamphuis et al., 2007a,b; Monti et al., 2007; Diago-Navarro et al., 2010). Therefore, the cooperative binding between region I and region II with the Kis-Kid octamer plays an important role in the transcription regulation of the *parD* operon, and this is dependent critically on the molar ratios of Kis and Kid.

## THE HYBRID YefM-YoeB TA SYSTEM: FURTHER COMPLEXITIES IN TA REGULATION

Toxins of type II TA systems have been divided into 12 superfamilies whereas type II antitoxins have been classified into 20 superfamilies based on sequence similarities (Leplae et al., 2011). Toxins and antitoxins from different families can associate and form hybrid systems (Arbing et al., 2010; Leplae et al., 2011), with the *yefM-yoeB* locus of *E. coli* being one such example. The YefM antitoxin is from the Phd superfamily whereas the YoeB toxin belongs to the ParE/RelE superfamily; the canonical association would be Phd-Doc and RelB-RelE (Polom et al., 2013). YefM-YoeB was identified as a potential TA system based on sequence similarities of YefM with the Phd antitoxin of phage P1 (Pomerantsev et al., 2001) and homology with the *axe-txe* TA system of *Enterococcus faecium* plasmid pRUM (Grady and Hayes, 2003). Ectopic overexpression of YoeB was shown to be toxic to *E. coli* but YefM counteracted this toxicity (Grady and Hayes, 2003). Since then, the *yefM-yoeB* TA system has been found in diverse bacterial species including *S. pneumoniae* (Nieto et al., 2007; Chan et al., 2011), *Streptococcus suis* (Zheng et al., 2015), *M. tuberculosis* (Kumar et al., 2008), *Staphylococcus aureus* (Yoshizumi et al., 2009), *Staphylococcus equorum* (Nolle et al., 2013), *Lactobacillus rhamnosus* (Krügel et al., 2015), and *Streptomyces* (Sevillano et al., 2012).

The *E. coli*-encoded YoeB toxin binds with the 70S ribosome with both the 30S and 50S subunits participating in YoeB binding and cleaves mRNA at the second position of the A site codon, thus inhibiting translation initiation in *E. coli* (Kamada and Hanaoka, 2005; Feng et al., 2013). YefM-YoeB forms a heterotrimeric YefM<sub>2</sub>-YoeB complex where one C-terminal peptide of the YefM dimer binds with YoeB while the other projects into the solvent (Figure 4A). The YefM dimer has symmetrical N-terminal globular structure while the C-terminus of YefM appears to be structurally disordered in the absence of YoeB and undergoes a disorder-to-order transition upon YoeB binding (Kamada and Hanaoka, 2005). YoeB forms a compact globular structure with structural similarities in its active site to RelE and other microbial RNases. Binding of YoeB to YefM in the heterotrimeric complex leads to conformational rearrangement of the RNase catalytic site of YoeB and direct obstruction by YefM, thus suppressing the toxicity of YoeB (Kamada and Hanaoka, 2005). The crystal structure of the YefM antitoxin from *M. tuberculosis* also indicated an ordered N-terminal domain and a very flexible C-terminal end that adopts different conformations in different monomers. This flexibility is postulated to make YefM more prone to proteolytic degradation (Kumar et al., 2008). YoeB-dependent mRNA cleavage is indeed activated by overproduction of the Lon protease in *E.*



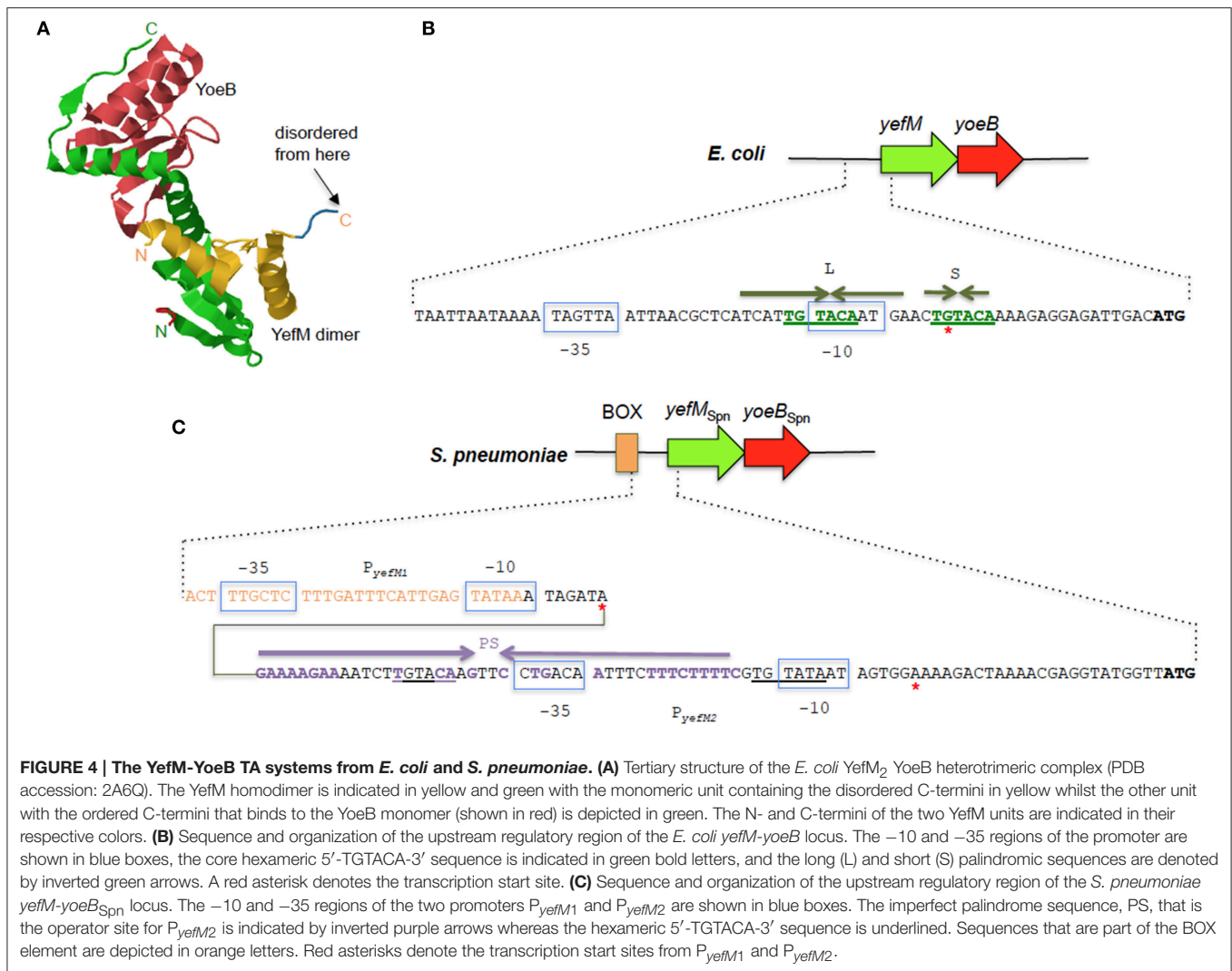


*coli*, suggesting that Lon is responsible for YefM degradation (Christensen et al., 2004).

Like most type II TA systems, the *E. coli yefM-yoeB* locus is transcriptionally autoregulated with YefM being the repressor and YoeB being a co-repressor that enhances the transcriptional repression (Kedzierska et al., 2007). There are no conventional DNA-binding motifs apparent in YefM but the N-terminal domains of the YefM dimer in the YefM<sub>2</sub> YoeB trimeric complex display conserved basic patches below the symmetrical dimer interface and this was suggested as the primary DNA anchor for operator site binding (Kamada and Hanaoka, 2005; Bailey and Hayes, 2009). Two arginine residues within this basic patch (R10 and R31) were mutated and found to be essential for DNA binding by the YefM<sub>2</sub> YoeB complex (Bailey and Hayes, 2009). Thus, a novel protein fold likely mediates operator recognition by YefM and we will have to await the elucidation of the YefM and YefM-YoeB structures bound to DNA for affirmation. The operator site in the *E. coli yefM-yoeB* locus consists of short (S) and long (L) palindromes, both of which possess a core hexameric 5'-TGTACA-3' motif. The center-to-center distance between the L and S palindromes are 12 bp with the L palindrome overlapping the -10 promoter region (Figure 4B). YefM initially binds to the L palindrome followed by the S palindrome (Kedzierska et al., 2007). Changing the spacing between the two palindromes perturbs the cooperative binding of YefM-YoeB to the repeats whereby binding to the L repeat is maintained but binding to the S repeat is disrupted (Bailey and Hayes, 2009). The L and

S palindromes appeared to be conserved in regions upstream of *yefM-yoeB* homologs from several bacterial genomes such as *Shigella boydii*, *Pseudomonas aeruginosa*, and *Erwinia carotovora* (Kedzierska et al., 2007) inferring that interaction of the YefM-YoeB homologs with these motifs could be a conserved mode of transcriptional autoregulation in these operons (Hayes and Kedzierska, 2014).

However, investigations into the regulation of the *yefM-yoeB* locus in *S. pneumoniae*, designated *yefM-yoeB*<sub>Spn</sub>, indicated a different and more complex regulatory mechanism (Chan et al., 2011). Expression of the *yefM-yoeB*<sub>Spn</sub> locus is driven by two  $\sigma^{70}$ -type promoters 30 bp apart: *P*<sub>yefM2</sub>, which is closer to the *yefM-yoeB*<sub>Spn</sub> genes and *P*<sub>yefM1</sub>, which lies further upstream to *P*<sub>yefM2</sub> (Figure 4C). The hexameric 5'-TGTACA-3' motif (Kedzierska et al., 2007) is also found within the pneumococcal *yefM-yoeB*<sub>Spn</sub> promoter region with one of the motifs being part of a longer 44 bp incomplete palindrome sequence that overlapped the -35 region of *P*<sub>yefM2</sub> (Figure 4C) and which was shown by footprinting experiments to be the operator site for the operon (Chan et al., 2011). *P*<sub>yefM2</sub> is likely the native promoter for *yefM-yoeB*<sub>Spn</sub> as its expression is autoregulated like other canonical type II TA systems, i.e., YefM<sub>Spn</sub> represses transcription from *P*<sub>yefM2</sub> while YoeB<sub>Spn</sub> exerts further repression in complex with YefM<sub>Spn</sub>. However, *P*<sub>yefM1</sub> appeared to be a constitutive, weaker promoter as compared to *P*<sub>yefM2</sub> and is not regulated by YefM-YoeB<sub>Spn</sub> (Chan et al., 2011). Interestingly, the *P*<sub>yefM1</sub> promoter came about from insertion of a BOX element upstream of *P*<sub>yefM2</sub>



and more intriguingly, transcriptional activation was observed when the BOX element, *P<sub>yefM1</sub>*, *P<sub>yefM2</sub>*, and *yefM<sub>Spn</sub>* were all in *cis* (but not when *yefM<sub>Spn</sub>* was provided in *trans*), hinting at the possible involvement of other hitherto unknown *cis*-acting factors in the regulation of the *yefM-yoeB<sub>Spn</sub>* locus (Chan et al., 2011). BOX elements are enigmatic sequences, considered to be potentially mobile and distributed randomly in numerous copies in the intergenic regions of pneumococci and related species. The occurrence and placement of the BOX element seems to be conserved in all *S. pneumoniae* strains that harbor *yefM-yoeB<sub>Spn</sub>*, suggesting its likely evolutionary importance to the biological function of *yefM-yoeB<sub>Spn</sub>* in pneumococci (Chan et al., 2011, 2012).

A BOX-like element was also found upstream of the *yefM-yoeB<sub>Lrh</sub>* locus of *L. rhamnosus* but unlike in *S. pneumoniae*, this insertion did not lead to the creation of an additional promoter (Krügel et al., 2015). Nevertheless, the regulation of the *yefM-yoeB<sub>Lrh</sub>* locus appeared to be complex as well with two transcription start sites detected within the *yefM<sub>Lrh</sub>* gene besides the main transcript that is expressed from a  $\sigma^{70}$ -type

promoter upstream of *yefM<sub>Lrh</sub>*. Furthermore, the expression levels of *yefM<sub>Lrh</sub>* and *yoeB<sub>Lrh</sub>* differed during various stages of growth and environmental stresses and appeared to respond differently in different *L. rhamnosus* strains (Krügel et al., 2015). The surprising discovery of a short transcript that is divergently transcribed and overlaps the *yoeB<sub>Lrh</sub>* gene and with similarities to several type I antitoxins also hints at further complexity of the *yefM-yoeB<sub>Lrh</sub>* operon regulation in *L. rhamnosus* (Krügel et al., 2015).

Such complex and multilayered regulatory control was also observed for the *yefM-yoeB* homolog, *axe-txe*, found in the *E. faecium* plasmid pRUM (Boss et al., 2013). The main promoter, *p<sub>at</sub>*, is autoregulated like other type II TA systems with the Axe antitoxin repressing the promoter weakly and stronger repression with the Axe-Txe TA complex. However, an internal promoter within the *axe* gene also directs the expression of the downstream *txe* toxin gene and this promoter did not appear to be regulated by Axe-Txe. Nevertheless, this internal promoter is crucial for *axe-txe* to function as a plasmid stabilization module, suggesting that it plays a role in setting the appropriate Axe:Txe ratio for

proper functioning of the system (Boss et al., 2013). The finding of a cryptic transcript that originates within the *txe* reading frame along with a putative transcription terminator-like sequence downstream of *txe* that possibly modulates production of Txe are indicative of further complexities in the regulation of the *axe-txe* operon (Boss et al., 2013; Hayes and Kędzierska, 2014).

## THE MqsA ANTITOXIN OF THE *mqsRA* TA LOCI: AN ANTITOXIN THAT ALSO REGULATES OTHER GENES

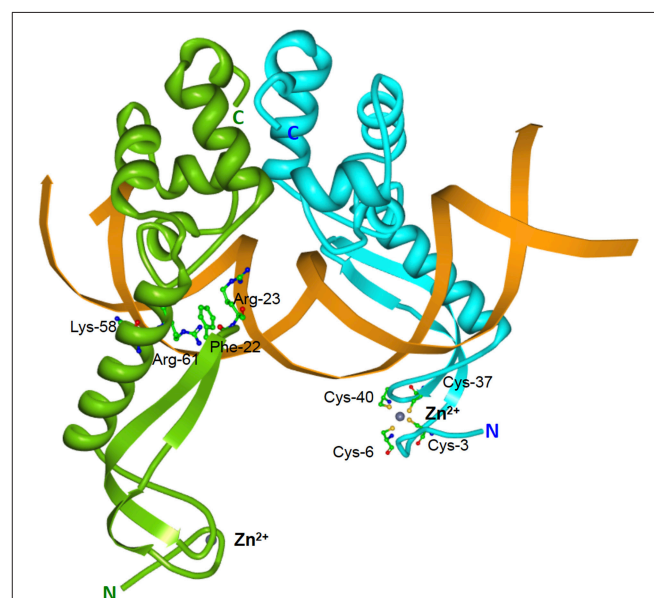
The MqsRA locus of *E. coli* K-12 is an unusual TA locus that differs from most canonical TA systems. The MqsR (motility quorum sensing regulator) toxin was initially identified as a regulator of motility and quorum sensing, influencing the development of biofilms by mediating the cellular response to autoinducer-2 (Ren et al., 2004; González Barrios et al., 2006). The *mqsR* gene was also significantly upregulated in persister cells and, along with its downstream gene, *mqsA*, were shown to be a type II TA system (Brown et al., 2009; Yamaguchi et al., 2009; Christensen-Dalsgaard et al., 2010; Kasari et al., 2010). MqsR is a ribosome-independent endoribonuclease that specifically cleaves mRNA at 5'-GCU-3' and, to a lesser extent, 5'-GCA-3' sequences (Yamaguchi et al., 2009; Christensen-Dalsgaard et al., 2010).

The MqsRA system is unique in several aspects. The *mqsR* toxin gene precedes the *mqsA* antitoxin gene, an arrangement that so far has been observed only in a few type II TA loci, namely *higBA* (Tian et al., 1996), *hicAB* (Jørgensen et al., 2009), and *mlAB* (Koga et al., 2011). The MqsA antitoxin is larger than MqsR toxin (14.7 kDa and 11.2 kDa, respectively) whereas in canonical TA systems, the toxin is larger than the antitoxin, with the exception of *HicB* (Jørgensen et al., 2009). Both MqsA and MqsR are also basic proteins whereas usually, the toxin is basic while the antitoxin is acidic (Kasari et al., 2010).

Elucidation of the MqsRA crystal structure also revealed a few surprises. The MqsRA complex is a dimer of dimers, comprising of two copies of MqsR and two copies of MqsA (Brown et al., 2009). The MqsA antitoxin monomer is well-ordered throughout its entire length and is composed of two structurally-distinct domains connected by a flexible linker which enables the two domains to rotate independently of each other (Brown et al., 2009). The N-terminal domain of MqsA binds zinc *via* coordination with four conserved cysteine residues with the bound zinc serving as a structural and not a catalytic role (Figure 5). MqsA interacts with DNA mainly through its C-terminal HTH domain that is also responsible for MqsA dimerization (Brown et al., 2009). Most antitoxins interact with DNA through their N-terminal residues with the exception of *HicB* (Hayes and Kędzierska, 2014). However, MqsA binding to DNA leads to bending of the DNA by more than 55° as well as a rotation of the N-terminal domain by more than 105°. This changes MqsA from a highly extended conformation into a narrow, elongated DNA “clamp” as a result of the formation of a DNA-binding pocket which positions several MqsA N-terminal residues (Phe-22, Arg-23, Lys-58, and Arg-61) for DNA interaction (Figure 5; Brown et al., 2011). Such

conformational change is unprecedented for a bacterial antitoxin. The neutralization of MqsR toxicity by MqsA is through steric occlusion and not by direct binding of the toxin active site as in many other antitoxins. Formation of a MqsR-MqsA-DNA complex induces substantial conformational changes (Yamaguchi et al., 2009) whereby the MqsR active site residues Lys-56, Gln-68, Tyr-81, and Lys-96 face inwards and toward the other MqsR toxin pair with a separation of only 13–15 Å (Brown et al., 2011). This severely limits the accessibility of the MqsR active sites for mRNA.

Like most other type II antitoxins, the MqsA antitoxin represses transcription of *mqsRA* but instead of acting as a co-repressor, the MqsR toxin functions as a transcriptional derepressor by disrupting the MqsA-DNA interaction. In fact, a 1:1 ratio of MqsR to MqsA ablated MqsA-DNA binding due to partial overlapping of binding sites on MqsA (in particular, the Arg-61 residue) for both MqsR and DNA (Brown et al., 2013). Another unique aspect of the MqsA antitoxin is that it serves not only as a transcriptional regulator of its own *mqsRA* locus but also of several other *E. coli* genes including *mcbR*, *cspD*, *spy* and the general stress response sigma factor *rpoS* (Brown et al., 2009; Kim et al., 2010; Wang et al., 2011). An *mqsRA*-like palindromic operator site is found upstream of *rpoS* (Wang et al., 2011) and *csgD*, which encodes a master regulator of biofilm formation through the control of curli (thin proteinaceous amyloid fibers which is a major extracellular component that promotes biofilm formation) and cellulose production (Soo and Wood, 2013).



**FIGURE 5 | Structure of the *E. coli* MqsA-DNA complex.** Tertiary structure of the *E. coli*-encoded MqsA dimer in complex with its operator DNA (PDB accession: 3O9X). The monomers of the MqsA dimer are colored either in green or in blue with their N- and C-termini indicated in their respective colors; zinc ions are shown as gray spheres; the *mqsRA* operator DNA is depicted in orange. For clarity, the MqsA amino acid residues that are crucial for interaction with operator DNA (Phe-22, Arg-23, Lys-58, and Arg-61) are shown for only one of the monomers as are the cysteine residues (Cys-3, Cys-6, Cys-37, and Cys-40) involved in coordination with the zinc ion (Brown et al., 2011).

CsgD also transcriptionally activates the gene for diguanylate cyclase (AdrA) which synthesizes the secondary messenger 3,5-cyclic diguanylic acid (c-di-GMP). Levels of c-di-GMP controls the switch from motility (low c-di-GMP) to sessility (high c-di-GMP) of *E. coli* (Soo and Wood, 2013). When nutrients are plentiful, MqsA increases motility by increasing the expression of *flhD*, the master regulator of *E. coli* motility partly through *rpoS* inhibition and partly through *csgD* inhibition, which also leads to low levels of c-di-GMP. Thus, in the absence of stress, MqsA functions to inhibit biofilm formation. When *E. coli* is under stressful conditions, Lon protease degrades MqsA, activating the MqsR toxin. Degradation of MqsA leads to derepression of *rpoS* and *csgD*, inhibition of *flhD*, high levels of c-di-GMP and subsequently, increased biofilm formation (Wang et al., 2011; Soo and Wood, 2013).

The MqsRA system was also recently shown to control the type V TA system, GhoST (Wang et al., 2013). The MqsR toxin enriches the *ghoT* toxin mRNA as the transcript lacks the MqsR cleavage site, 5'-GCU. GhoT functions as a membrane toxin that produces the phenotype known as ghost cells (lysed cells with damaged membranes; Wang et al., 2012). Under stressful conditions, MqsR is freed and the toxin degrades mRNAs primarily at 5'-GCU sites such as the 5'-end of the *ghoST* mRNA within the *ghoS* antitoxin coding sequence (which contains three 5'-GCU sites) but not *ghoT*. This leads to higher levels of GhoT toxin, which exerts its effects on the cell membrane, ultimately increasing persistence (Wang et al., 2013). Thus, there appears to be a hierarchy of TA systems in *E. coli* cells in which MqsRA controls GhoST.

## TRANSCRIPTIONAL ACTIVATORS THAT FUNCTION AS ANTITOXINS: THE TALE OF THE MrpC REGULATOR AND THE SOLITARY MazF TOXIN OF MYXOCOCCUS XANTHUS

*Myxococcus xanthus* is a Gram-negative, rod-shaped bacterium that provides a prokaryotic model for multicellular developmental processes. Under nutrient starvation conditions, cells form aggregates which mature into fruiting bodies with some cells differentiating into spores. Other cells remain outside the fruiting bodies as persister-like cells termed peripheral rods whereas the majority of cells lyse during this developmental process (Nariya and Inouye, 2008; Lee et al., 2012; Robinson et al., 2014). The *M. xanthus* genome was found to encode a solitary *mazF* toxin gene, *mazF-mx*, without a cognate *mazE*-like antitoxin gene (Nariya and Inouye, 2008). The *mazF-mx* gene was found to be developmentally regulated and deletion of *mazF-mx* in *M. xanthus* DZF1 reduced developmental cell lysis, produced a severe delay in aggregation and reduced sporulation. Interestingly, it was reported that MrpC, which is an essential developmental transcription factor, was found to regulate the expression of *mazF-mx* and also functions as an antitoxin for MazF-mx by forming a stable complex with MazF-mx (Nariya and Inouye, 2008). The *mrpC* gene is encoded 4.44 Mbp downstream from *mazF-mx* and activates expression of many

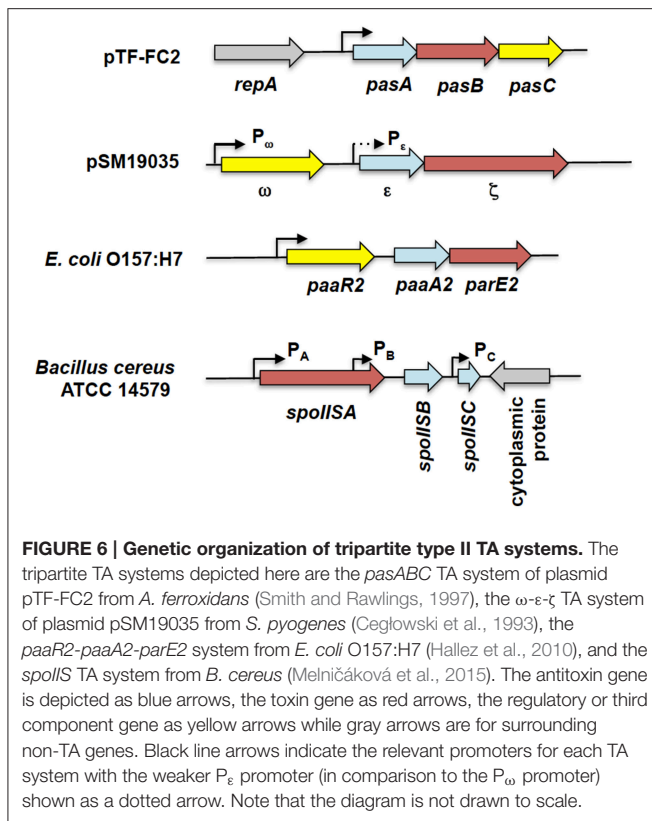
development-specific genes with strains lacking *mrpC* failing to develop and sporulate. Severe cell toxicity by MazF-mx was observed in a  $\Delta mrpC$  mutant when *mazF-mx* expression was induced (Nariya and Inouye, 2008). It was thus proposed that the orphan *mazF-mx* in *M. xanthus* was successfully integrated into the cellular developmental programme with another transcription factor unrelated to the common cognate MazE antitoxin functioning as the surrogate antitoxin.

However, some apparently conflicting data have recently emerged regarding the MazF-mx function. Lee et al. (2012) showed that deletion of *mazF-mx* from the wild-type strains DK1622 and DZ2 had minimal to no effect on developmental cell lysis and sporulation as opposed to its deletion in strain DZF1 as reported by Nariya and Inouye (2008). It was postulated that the DZF1 background contains a *pilQ1* allele bearing two missense mutations in *pilQ* (G741S and N762G) which greatly sensitizes *M. xanthus* cells and render them more susceptible to lysis (Lee et al., 2012). Indeed, it was shown that the phenotypic effects of *mazF-mx* removal in DZF1 were recreated in strain DK1622 by introducing the *pilQ1* mutation into a  $\Delta mazF$  mutant (Boynton et al., 2013). Lee et al. (2012) proposed the existence of two parallel, redundant pathways of developmental programmed cell death in DK1622 and DZ2, one of which is controlled by MazF-mx, and the other by an unknown mechanism which was disrupted in strain DZF1. However, Boynton et al. (2013) raised the possibility that the observed phenotypic differences may be artifactual, resulting from increased membrane permeability due to the *pilQ1* allele. Further, Boynton et al. (2013) reported that MrpC enhanced MazF-mx endoribonuclease activity in direct contrast to the inhibitory antitoxin behavior reported by Nariya and Inouye (2008) leading to a model in which MazF-mx was postulated to function without an antitoxin partner. Thus, MazF-mx seems to have elicited a scientific conundrum reminiscent of the cell death vs. cell stasis debate that erupted for the *E. coli*-encoded MazEF system more than a decade ago (Christensen et al., 2003; Gerdes et al., 2005; Engelberg-Kulka et al., 2006; Kolodkin-Gal et al., 2007; Van Melder and Saavedra De Bast, 2009; Van Melder, 2010). We await further experimental evidences and scientific arguments that will be presented regarding this topic.

## TRIPARTITE TYPE II TA LOCI

The *pas* (plasmid addiction system) found in the 12.2 kb broad-host range, mobilizable plasmid pTF-FC2 from *Acidithiobacillus ferrooxidans* (formerly *Thiobacillus ferrooxidans*) was a curious example of a type II TA system with three components, the PasA antitoxin, the PasB toxin, and PasC (Figure 6), a third component that appeared to enhance the ability of the PasA antitoxin to neutralize the PasB toxin (Smith and Rawlings, 1997, 1998b; Rawlings, 1999). The *pas* locus is autoregulated with PasA as the transcriptional repressor and PasB as the co-repressor. Full repression of the *pas* promoter was observed with the PasAB complex whereas PasC did not appear to play any regulatory role (Smith and Rawlings, 1998a). A similar plasmid, pTC-F14 from *Acidithiobacillus caldus*,





harbors only *pasA* and *pasB* and it was found that the two-component *pasAB* from pTC-F14 was less efficient at stabilizing a heterologous, low-copy tester plasmid pOU82 in *E. coli* when compared to the three-component *pasABC* of pTF-FC2 (Deane and Rawlings, 2004). Perhaps PasC forms a complex along with PasAB to augment the PasA antitoxin in neutralizing PasB. PasC can indeed be expressed along with PasA and PasB in *E. coli* (Smith and Rawlings, 1997) but there has yet to be any published reports on whether such a PasABC protein complex does occur. Hence, the actual function of PasC and how it helps PasA to abrogate the lethality of PasB remains unknown in the absence of further experimental results.

Another tripartite type II TA system, the  $\omega$ - $\epsilon$ - $\zeta$  system that was discovered encoded on the low-copy number plasmid pSM19035 from a clinical isolate of *Streptococcus pyogenes* (Figure 6), differed from the *pas* locus in which the regulatory role is played by a third party. In this TA system, both the  $\epsilon$  antitoxin and the  $\zeta$  toxin have no roles in transcriptional regulation, the function which is played instead by the  $\omega$  regulator (de la Hoz et al., 2000; Volante et al., 2014).  $\epsilon$ - $\zeta$  is a type II TA system in which the 10 kDa  $\epsilon$  antitoxin inactivates the 32 kDa  $\zeta$  toxin through steric occlusion. The crystal structure of the  $\epsilon$ - $\zeta$  complex indicated a heterotetrameric  $\epsilon_2\zeta_2$  arrangement whereby the N-terminal of  $\epsilon$  sterically blocks the ATP-binding active site of  $\zeta$  (Meinhart et al., 2003). The mechanism of  $\zeta$  toxicity is unique among bacterial TA toxins whereby its target is the cell wall precursor UDP-*N*-acetylglucosamine (UNAG) which is phosphorylated by  $\zeta$  to UDP-*N*-acetylglucosamine-3'-phosphate (UNAG-3P) using

ATP. UNAG is a basic unit of the peptidoglycan scaffold and phosphorylation of UNAG by  $\zeta$  converts it into a metabolite unusable for peptidoglycan synthesis (Mutschler et al., 2011). Besides that, UNAG-3P is also a competitive inhibitor of MurA, the enzyme that catalyzes the first step in peptidoglycan synthesis. Therefore,  $\zeta$  functions to inhibit bacterial cell wall synthesis (Mutschler and Meinhart, 2011; Mutschler et al., 2011). However, a recent paper reported that expression of  $\zeta$  only reduced the UNAG pool and did not totally deplete it with transient expression of  $\zeta$  (120 min) reversibly inducing a dormant state that was subsequently rescued by  $\epsilon$  expression (Lioy et al., 2012; Tabone et al., 2014). It was proposed that  $\zeta$  expression induces diverse responses to cope with stress with reduction in the UNAG levels as one of these responses rather than triggering a latent suicide program by depleting the UNAG pool (Tabone et al., 2014).

The  $\omega$ - $\epsilon$ - $\zeta$  genes form an operon with two distinct promoters,  $P_\omega$  upstream of the  $\omega$  reading frame, and  $P_\epsilon$  upstream of the  $\epsilon$  reading frame (Figure 6). Transcription mainly initiates from the  $\sigma^{70}$ -type  $P_\omega$  promoter whereas  $P_\epsilon$  appeared to be a weak, constitutive promoter that contributes marginally to transcription of the  $\epsilon$ - $\zeta$  operon (de la Hoz et al., 2000). The  $\omega$  regulator belongs to the MetJ/Arc repressor family, has an unstructured N-terminal domain followed by a RHH DNA-binding motif (Murayama et al., 2001). The binding site recognized by  $\omega$  is distinctive, comprising of both palindromic and non-palindromic heptad repeats (5'-NATCACN-3') in the operator site. A single  $\omega_2$  dimer binds to one heptad repeat and it was suggested that cooperative binding of the  $\omega_2$  dimer is achieved by polymerization of  $\omega_2$  on arrays of the repeated heptad elements (Weihofen et al., 2006).  $\omega$  also functions as a global regulator for plasmid pSM19035, controlling the expression of genes such as the copy control gene *copS* and the plasmid partitioning gene  $\delta$ , which encodes a ParA ATPase. Interestingly,  $\omega_2$  can either activate or repress  $P_\omega$  in a concentration-dependent manner with  $\delta_2$  acting as a co-activator by increasing the half-life of the  $\omega_2$ - $P_\omega$  DNA complexes (Volante et al., 2015).

Another tripartite type II TA system, the *paaR-paaA-parE* system, was identified in the genome of *E. coli* O157:H7 (Figure 6; Hallez et al., 2010). The ParE toxin is usually associated with the ParD antitoxin (Gerdes et al., 2005). PaaA is a novel antitoxin family that is associated with the ParE toxin, and *paaA-parE* forms an operon with a third component, *paaR* that functions as a transcriptional regulator. The *paaR-paaA-parE* operon is co-transcribed from a  $\sigma^{70}$ -type promoter upstream of *paaR* (Hallez et al., 2010). Unlike the  $\omega$ - $\epsilon$ - $\zeta$  system in which  $\epsilon$ - $\zeta$  did not play any role in transcriptional regulation (Mutschler and Meinhart, 2013), the PaaA antitoxin forms a complex with the ParE toxin that repress transcription from the *paaR* promoter, albeit partially. Full repression of transcription requires the PaaR regulator (Hallez et al., 2010). However, the two repressor complexes (i.e., PaaA-ParE and PaaR) probably act independently as no three-protein complexes were detected under experimental conditions (Hallez et al., 2010). Interestingly, the genome of *E. coli* O157:H7 contains two paralogous *paaR-paaA-parE* systems with the second *paaR2-paaA2-parE2* system

located in a predicted prophage. Both systems apparently coexist independently as the PaaA1 antitoxin is unable to neutralize ParE2 toxicity and vice versa (Hallez et al., 2010).

A recent report regarding the SpoIIS TA system from *Bacillus cereus* revealed a curious variation of the tripartite TA system (Melničáková et al., 2015). The *spoIIS* locus was initially identified in *B. subtilis* and was then deduced to consist of two genes, the *spoIISA* toxin-coding gene and the *spoIISB* antitoxin-coding gene, i.e., a typical type II TA locus (Adler et al., 2001). However, transcriptome analysis of *B. subtilis* had indicated the presence of a third transcriptionally active region within the *spoIIS* locus designated S458 (Nicolas et al., 2012) and which has been renamed *spoIISC* (Melničáková et al., 2015). Intriguingly, it was discovered that *spoIISC* in *B. subtilis* as well as *B. cereus* coded for an antitoxin that neutralizes the toxicity of SpoIISA. In other words, the SpoIISA toxin is neutralized by two antitoxins, SpoIISB and SpoIISC (Melničáková et al., 2015). In a departure from most type II TA systems, each gene in the *spoIIS* locus is transcribed from its own promoter (Figure 6) and each promoter is apparently transcribed under different conditions. For example, in *B. subtilis* only *spoIISA* and *spoIISB* are transcribed during nutrient deprivation, whereas during ethanol stress, only the *spoIISA* is transcribed and *spoIISC* transcribed during biofilm formation (Nicolas et al., 2012; Melničáková et al., 2015). This gives a hint at the complexity of the regulation of the *spoIIS* locus that may necessitate the need for two antitoxins, each of which could antagonize the toxicity of SpoIISA. However, at this point, there is no information as to whether the expression of the *spoIIS* genes is autoregulated.

## CAVEATS: THE EzeT AND VapC-1 TOXINS

The  $\zeta$  toxin of the tripartite  $\omega$ - $\epsilon$ - $\zeta$  system has two types of interesting chromosomally-encoded homologs. One homolog is exemplified by the PezT toxin of the *pezAT* system of *S. pneumoniae* whereby *pezAT* is a typical type II TA system in which the antitoxin PezA also plays an autoregulatory role, unlike the  $\epsilon$  antitoxin (Khoo et al., 2007). PezA contains an N-terminal HTH DNA-binding motif as its repressor domain, which is fused with the three-helix bundle domain that binds and inhibits the PezT toxin. In this instance, no homologs of the  $\omega$  regulator is evident in the *S. pneumoniae* genome (Khoo et al., 2007) and it is clear that  $\omega$  and the repressor domain of PezA have different evolutionary origins. It was postulated that *pezA* likely originated from a fusion event of an unrelated transcriptional repressor coding sequence to the 5'-end of the coding sequence of an  $\epsilon$  ortholog (Mutschler and Meinhart, 2013). Hints of involvement of PezAT in the pathogenicity of *S. pneumoniae* and its function in the pneumococcal pathogenicity island 1 (Brown et al., 2004; Harvey et al., 2011; Mutschler and Meinhart, 2011; Chan et al., 2012) as well as a pneumococcal integrative and conjugative element (ICE; Chan et al., 2014; Iannelli et al., 2014) warrants further investigations. Another interesting  $\zeta$  homolog is found in the genomes of several bacteria. These  $\zeta$  homologs are much larger than either  $\zeta$  or PezT and are found not associated with a corresponding  $\epsilon$  or PezA

antitoxins (Chan et al., 2012). The functionality of these solitary  $\zeta$  homologs was enigmatic as overexpression of a homolog from *Acinetobacter baumannii* was reportedly non-lethal (Jurenaite et al., 2013). However, the Meinhart group in a recent report has elegantly demonstrated that one of these solitary  $\zeta$  homologs in *E. coli*, designated EzeT, consisted of a toxin domain in the C-terminal and an antitoxin domain in the N-terminal in a single polypeptide chain (Rocker and Meinhart, 2015a). *E. coli* cells that expressed full-length EzeT grew normally with no UNAG-3P detected. However, in cells that expressed an EzeT variant EzeT $\Delta$ N83, that had its first 83 amino acid residues from the N-terminal deleted, a strong reduction in viability was observed in parallel with increased cell permeabilization and accumulation of UNAG-3P. Co-expression of the toxin domain (EzeT $\Delta$ N83) and the N-terminal antitoxin domain [EzeT(1-82)] from separate expression vectors led to similar growth profiles as for full-length EzeT, indicative of *trans*-complementation (Rocker and Meinhart, 2015a). Intriguingly, it was found that the toxicity of EzeT $\Delta$ N83 was only evident at low temperatures (below 30°C) and at 37°C, EzeT was non-functional (Rocker and Meinhart, 2015a) similar to what was reported for the GraTA system of the soil bacterium *Pseudomonas putida* (Tamman et al., 2014). Whether EzeT is autoregulated like other type II TA systems is still unknown and transcription is likely initiated from a weak promoter with a conventional -10 hexamer but without a -35 element (Rocker and Meinhart, 2015a). Nevertheless, a closer examination of solitary or orphan toxin homologs is clearly needed as EzeT has been demonstrated to be likely a new type of TA system in which a *cis*-acting antitoxin is tethered to the toxin within a single polypeptide. Besides that, large, possibly multi-domain  $\zeta$ -toxin homologs linked to phosphatase or peptidoglycan-binding domains have been detected along with other toxin families such as ParE, Fic/Doc, and PemK as parts of multi-domain proteins (Rocker and Meinhart, 2015b). Their characterization and biological functions await further investigations.

The VapBC TA system is by far the most numerous among TA families with many bacterial genomes containing multiple *vapBC* loci (Pandey and Gerdes, 2005; Leplae et al., 2011; Shao et al., 2011). The VapC toxins are characterized by a PIN (PiIT N-terminus) domain and display similarities to several nuclease domains. VapC from enterobacteria are tRNAses that inhibit global translation by site-specific cleavage of tRNA<sup>fMet</sup> between the anticodon stem and loop (Winther and Gerdes, 2011) whereas the VapC toxins from other bacterial species have different RNA target specificities (Ahidjo et al., 2011; McKenzie et al., 2012). As with other type II TAs, the VapBC complexes bind to operators in the promoter regions to autoregulate transcription (Robson et al., 2009; Winther and Gerdes, 2012). However, the *vapBC-1* locus of nontypeable *Haemophilus influenzae* showed notable differences as in stark contrast to other VapBC homologs and type II TA systems that have been described, the VapC-1 toxin possesses DNA binding activity whereas the VapB-1 antitoxin does not interact directly with DNA (Cline et al., 2012). However, VapB-1 increases the affinity of VapC-1 for DNA and confers specificity for the operator site for the VapBC complex. The *vapBC-1* locus is also regulated by

the FIS which is responsible for activation of *vapBC-1* during nutrient upshifts (Cline et al., 2012). During nutrient starvation conditions, VapB-1 would be degraded by endogenous proteases, releasing active VapC-1 toxins, and facilitating entry of *H. influenzae* cells into the persister state. When conditions favor cellular growth, FIS activates *vapBC-1* transcription and displaces any bound VapC-1 on the operator site (which is unstable in the absence of VapB-1). Levels of FIS decreases during early exponential growth, and this allows the VapBC-1 complex to bind and restore transcriptional equilibrium (Cline et al., 2012).

## CONCLUSIONS AND PERSPECTIVES

Our knowledge on toxin-antitoxin systems has indeed come a long way since they were coined as “addiction” modules that function to ensure the stable maintenance of plasmids in the absence of selection pressure by killing off any plasmid-free daughter cells that developed following cell division. The near ubiquity of these systems in prokaryotic genomes and their wide variety reflect their myriad functions in the prokaryotic lifestyle. The antitoxins are central to the proper functioning of these TA systems and in this review, we delved in detail on how these antitoxins usually play dual roles in regulating the expression of the TA operon as well as neutralizing the lethal action of the toxins during normal cellular growth, i.e., keeping the proverbial wolves at bay. In most cases, cellular survival hinges on maintaining the balance between the amounts of toxin and its cognate antitoxin. Hence we have seen how some TA systems have evolved beyond the basic autoregulatory circuit to incorporate additional regulatory elements. Such further complexities to the regulation of TA expression are speculated to provide additional possibilities to fine-tune and optimize the production of toxin and antitoxin under diverse environmental conditions enabling the cells to better adapt to rapid fluctuations. Such rapid fluctuations may be extreme in soil-inhabiting bacteria and related environmental niches whereas bacteria that lead a relatively “comfortable” life in a host such as pneumococci in biofilms in the human nasopharynx may have to more frequently confront changes in the host immune system. As the ectopic expression of some of the bacterial TA toxin genes leads to severe growth defects and cell death, there has been increasing interest in TAs as potential targets for novel antimicrobial agents. Several strategies have been proposed

and developed to enable toxin activation in pathogenic bacteria such as interfering with the TA complex formation or the transcription of the operon itself through ligands that block the interaction of the antitoxin with the operator site. These and the potential for exploiting TAs as antimicrobial agents have been recently reviewed (Alonso et al., 2007; Mutschler and Meinhart, 2011; Williams and Hergenrother, 2012; Tanouchi et al., 2013; Hayes and Kędzierska, 2014; Chan et al., 2015). TAs have also been harnessed as tools for biotechnology and molecular biology such as in the development of positive selection plasmid vectors (Stieber et al., 2008; Unterholzner et al., 2013). The discovery of their functionality in eukaryotic cells have opened up interesting avenues for research and development including as anticancer and antiviral gene therapies, and as a containment system for genetically modified organisms (de la Cueva-Méndez et al., 2003; Chono et al., 2011; de la Cueva-Méndez and Pimentel, 2013; Shimazu et al., 2014; Wieteska et al., 2014; Chan et al., 2015; Preston et al., 2015; Yeo et al., 2016). With more TAs being discovered and characterized in the coming months and years ahead, our understanding of their variety and complexity, particularly in the regulatory circuits of these small genetic loci, will be greatly enhanced. Additional knowledge on these systems would enable novel and improved strategies for harnessing TAs for various biomedical and biotechnological applications. These serve to underline the importance and essentiality of TA systems in modulating the prokaryotic lifestyle.

## AUTHOR CONTRIBUTIONS

WTC, ME and CCY conceived, wrote, edited and approved this review.

## ACKNOWLEDGMENTS

Work supported by Grants CSD2008/00013 (to ME and WTC), and BIO2015-69085-REDC and BIO2013-49148-C2-2-R (to ME) from the Spanish Ministry of Economy and Competitiveness; PRPUM grant CG011-2014 (to CCY) from the Malaysian Ministry of Higher Education. CCY would like to thank J. A. Harikrishna, K. L. Thong, and T. S. Cha for on-going fruitful collaborations and to Universiti Sultan Zainal Abidin for facilities and support.

## REFERENCES

- Aakre, C. D., Phung, T. N., Huang, D., and Laub, M. T. (2013). A bacterial toxin inhibits DNA replication elongation through a direct interaction with the  $\beta$  sliding clamp. *Mol. Cell* 52, 617–628. doi: 10.1016/j.molcel.2013.10.014
- Adler, E., Barák, I., and Stragier, P. (2001). *Bacillus subtilis* locus encoding a killer protein and its antidote. *J. Bacteriol.* 183, 3574–3581. doi: 10.1128/JB.183.12.3574-3581.2001
- Agarwal, S., Mishra, N. K., Bhatnagar, S., and Bhatnagar, R. (2010). PemK toxin of *Bacillus anthracis* is a ribonuclease: an insight into its active site, structure, and function. *J. Biol. Chem.* 285, 7254–7270. doi: 10.1074/jbc.M109.073387
- Ahidjo, B. A., Kuhnert, D., McKenzie, J. L., Machowski, E. E., Gordhan, B. G., Arcus, V., et al. (2011). VapC toxins from *Mycobacterium tuberculosis* are ribonucleases that differentially inhibit growth and are neutralized by cognate VapB antitoxins. *PLoS ONE* 6:e21738. doi: 10.1371/journal.pone.0021738
- Aizenman, E., Engelberg-Kulka, H., and Glaser, G. (1996). An *Escherichia coli* chromosomal “addiction module” regulated by guanosine 3',5'-bispyrophosphate: a model for programmed bacterial cell death. *Proc. Natl. Acad. Sci. U.S.A.* 93, 6059–6063. doi: 10.1073/pnas.93.12.6059
- Alawneh, A. M., Qi, D., Yonesaki, T., and Otsuka, Y. (2016). An ADP-ribosyltransferase Alt of bacteriophage T4 negatively regulates the *Escherichia coli* MazF toxin of a toxin-antitoxin module. *Mol. Microbiol.* 99, 188–198. doi: 10.1111/mmi.13225
- Alonso, J. C., Balsa, D., Cherny, I., Christensen, S. K., Espinosa, M., Francuski, D., et al. (2007). “Bacterial toxin-antitoxin systems as targets for the development of novel antibiotics,” in *Enzyme-Mediated Resistance to Antibiotics*, eds M.



- E. Tolmasky and R. A. Bonomo (Washington DC: American Society of Microbiology), 313–329.
- Amitai, S., Kolodkin-Gal, I., Hananya-Melabashi, M., Sacher, A., and Engelberg-Kulka, H. (2009). *Escherichia coli* MazF leads to the simultaneous selective synthesis of both “Death Proteins” and “Survival Proteins.” *PLoS Genet.* 5:e1000390. doi: 10.1371/journal.pgen.1000390
- Arbing, M. A., Handelman, S. K., Kuzin, A. P., Verdon, G., Wang, C., Su, M., et al. (2010). Crystal structures of Phd-Doc, HigA, and YeeU establish multiple evolutionary links between microbial growth-regulating toxin-antitoxin systems. *Structure* 18, 996–1010. doi: 10.1016/j.str.2010.04.018
- Armalyte, J., Jurenaite, M., Beinoraviciute, G., Teiserskas, J., and Suziedeliene, E. (2012). Characterization of *Escherichia coli* dinJ-yafQ toxin-antitoxin system using insights from mutagenesis data. *J. Bacteriol.* 194, 1523–1532. doi: 10.1128/JB.06104-11
- Bailey, S. E. S., and Hayes, F. (2009). Influence of operator site geometry on transcriptional control by the YefM-YoeB toxin-antitoxin complex. *J. Bacteriol.* 191, 762–772. doi: 10.1128/JB.01331-08
- Barbosa, L. C. B., Garrido, S. S., and Marchetto, R. (2015). BtoxDB: a comprehensive database of protein structural data on toxin-antitoxin systems. *Comput. Biol. Med.* 58, 146–153. doi: 10.1016/j.combiomed.2015.01.010
- Bernard, P., and Couturier, M. (1992). Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J. Mol. Biol.* 226, 735–745. doi: 10.1016/0022-2836(92)90629-X
- Bertram, R., and Schuster, C. F. (2014). Post-transcriptional regulation of gene expression in bacterial pathogens by toxin-antitoxin systems. *Front. Cell. Infect. Microbiol.* 4:6. doi: 10.3389/fcimb.2014.00006
- Bibi-Triki, S., Li de la Sierra-Gallay, I., Lazar, N., Leroy, A., Van Tilbeurgh, H., Sebbane, F., et al. (2014). Functional and structural analysis of HicA3-HicB3, a novel toxin-antitoxin system of *Yersinia pestis*. *J. Bacteriol.* 196, 3712–3723. doi: 10.1128/JB.01932-14
- Blower, T. R., Salmond, G. P. C., and Luisi, B. F. (2011). Balancing at survival's edge: The structure and adaptive benefits of prokaryotic toxin-antitoxin partners. *Curr. Opin. Struct. Biol.* 21, 109–118. doi: 10.1016/j.sbi.2010.10.009
- Bobay, B. G., Andreeva, A., Mueller, G. A., Cavanagh, J., and Murzin, A. G. (2005). Revised structure of the AbrB N-terminal domain unifies a diverse superfamily of putative DNA-binding proteins. *FEBS Lett.* 579, 5669–5674. doi: 10.1016/j.febslet.2005.09.045
- Bøggild, A., Sofos, N., Andersen, K. R., Feddersen, A., Easter, A. D., Passmore, L. A., et al. (2012). The crystal structure of the intact *E. coli* RelBE toxin-antitoxin complex provides the structural basis for conditional cooperativity. *Structure* 20, 1641–1648. doi: 10.1016/j.str.2012.08.017
- Boss, L., Labudda, Ł., Węgrzyn, G., Hayes, F., and Kędzierska, B. (2013). The axe-txe complex of *Enterococcus faecium* presents a multilayered mode of toxin-antitoxin gene expression regulation. *PLoS ONE* 8:e73569. doi: 10.1371/journal.pone.0073569
- Boynton, T. O., Mcmurry, J. L., and Shimkets, L. J. (2013). Characterization of *Myxococcus xanthus* MazF and implications for a new point of regulation. *Mol. Microbiol.* 87, 1267–1276. doi: 10.1111/mmi.12165
- Bravo, A., Ortega, S., de Torrontegui, G., and Díaz, R. (1988). Killing of *Escherichia coli* cells modulated by components of the stability system ParD of plasmid R1. *Mol. Gen. Genet.* 215, 146–151. doi: 10.1007/BF00331316
- Brennan, R. G., and Matthews, B. W. (1989). The Helix-Turn-Helix DNA binding motif. *J. Biol. Chem.* 264, 1903–1906.
- Brown, B. L., Grigoriu, S., Kim, Y., Arruda, J. M., Davenport, A., Wood, T. K., et al. (2009). Three dimensional structure of the MqsR:MqsA complex: a novel TA pair comprised of a toxin homologous to RelE and an antitoxin with unique properties. *PLoS Pathog.* 5:e1000706. doi: 10.1371/journal.ppat.1000706
- Brown, B. L., Lord, D. M., Grigoriu, S., Peti, W., and Page, R. (2013). The *Escherichia coli* toxin MqsR destabilizes the transcriptional repression complex formed between the antitoxin MqsA and the mqsRA operon promoter. *J. Biol. Chem.* 288, 1286–1294. doi: 10.1074/jbc.M112.421008
- Brown, B. L., Wood, T. K., Peti, W., and Page, R. (2011). Structure of the *Escherichia coli* antitoxin MqsA (YgiT/b3021) bound to its gene promoter reveals extensive domain rearrangements and the specificity of transcriptional regulation. *J. Biol. Chem.* 286, 2285–2296. doi: 10.1074/jbc.M110.172643
- Brown, J. S., Gilliland, S. M., Spratt, B. G., and Holden, D. W. (2004). A locus contained within a variable region of pneumococcal pathogenicity Island 1 contributes to virulence in mice. *Infect. Immun.* 72, 1587–1593. doi: 10.1128/IAI.72.3.1587-1593.2004
- Castro-Roa, D., Garcia-Pino, A., De Gieter, S., van Nuland, N. A. J., Loris, R., and Zenkin, N. (2013). The Fic protein Doc uses an inverted substrate to phosphorylate and inactivate EF-Tu. *Nat. Chem. Biol.* 9, 811–817. doi: 10.1038/nchembio.1364
- Cataudella, I., Sneppen, K., Gerdes, K., and Mitarai, N. (2013). Conditional cooperativity of toxin - antitoxin regulation can mediate bistability between growth and dormancy. *PLoS Comput. Biol.* 9:e1003174. doi: 10.1371/journal.pcbi.1003174
- Cataudella, I., Trusina, A., Sneppen, K., Gerdes, K., and Mitarai, N. (2012). Conditional cooperativity in toxin-antitoxin regulation prevents random toxin activation and promotes fast translational recovery. *Nucleic Acids Res.* 40, 6424–6434. doi: 10.1093/nar/gks297
- Ceglowski, P., Boitsov, A., Chai, S., and Alonso, J. C. (1993). Analysis of the stabilization system of pSM19035-derived plasmid pBT233 in *Bacillus subtilis*. *Gene* 136, 1–12. doi: 10.1016/0378-1119(93)90441-5
- Chan, W. T., Balsa, D., and Espinosa, M. (2015). One cannot rule them all: are bacterial toxins-antitoxins druggable? *FEMS Microbiol. Rev.* 39, 522–540. doi: 10.1093/femsre/fuv002
- Chan, W. T., Moreno-Córdoba, I., Yeo, C. C., and Espinosa, M. (2012). Toxin-antitoxin genes of the gram-positive pathogen *Streptococcus pneumoniae*: so few and yet so many. *Microbiol. Mol. Biol. Rev.* 76, 773–791. doi: 10.1128/MMBR.00030-12
- Chan, W. T., Nieto, C., Harikrishna, J. A., Khoo, S. K., Othman, R. Y., Espinosa, M., et al. (2011). Genetic regulation of the yefM-yoeB toxin-antitoxin locus of *Streptococcus pneumoniae*. *J. Bacteriol.* 193, 4612–4625. doi: 10.1128/JB.05187-11
- Chan, W. T., Yeo, C. C., Sadowy, E., and Espinosa, M. (2014). Functional validation of putative toxin-antitoxin genes from the Gram-positive pathogen *Streptococcus pneumoniae*: phd-doc is the fourth bona-fide operon. *Front. Microbiol.* 5:677. doi: 10.3389/fmicb.2014.00677
- Cheng, H.-Y., Soo, V. W. C., Islam, S., McAnulty, M. J., Benedik, M. J., and Wood, T. K. (2014). Toxin GhoT of the GhoT/GhoS toxin/antitoxin system damages the cell membrane to reduce adenosine triphosphate and to reduce growth under stress. *Environ. Microbiol.* 16, 1741–1754. doi: 10.1111/1462-2920.12373
- Cherny, I., Rockah, L., and Gazit, E. (2005). The YoeB Toxin is a folded protein that forms a physical complex with the Unfolded YefM antitoxin: implications for a structural-based differential stability of toxin-antitoxin systems. *J. Biol. Chem.* 280, 30063–30072. doi: 10.1074/jbc.M506220200
- Chono, H., Matsumoto, K., Tsuda, H., Saito, N., Lee, K., Kim, S., et al. (2011). Acquisition of HIV-1 resistance in T lymphocytes using an ACA-specific *E. coli* mRNA interferase. *Hum. Gene Ther.* 22, 35–43. doi: 10.1089/hum.2010.001
- Christensen, S. K., and Gerdes, K. (2003). RelE toxins from bacteria and Archaea cleave mRNAs on translating ribosomes, which are rescued by tmRNA. *Mol. Microbiol.* 48, 1389–1400. doi: 10.1046/j.1365-2958.2003.03512.x
- Christensen, S. K., Maenhaut-Michel, G., Mine, N., Gottesman, S., Gerdes, K., and Van Melderen, L. (2004). Overproduction of the Lon protease triggers inhibition of translation in *Escherichia coli*: involvement of the yefM-yoeB toxin-antitoxin system. *Mol. Microbiol.* 51, 1705–1717. doi: 10.1046/j.1365-2958.2003.03941.x
- Christensen, S. K., Mikkelsen, M., Pedersen, K., and Gerdes, K. (2001). RelE, a global inhibitor of translation, is activated during nutritional stress. *Proc. Natl. Acad. Sci. U.S.A.* 98, 14328–14333. doi: 10.1073/pnas.251327898
- Christensen, S. K., Pedersen, K., Hansen, F. G., and Gerdes, K. (2003). Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated mRNAs and are counteracted by tmRNA. *J. Mol. Biol.* 332, 809–819. doi: 10.1016/S0022-2836(03)00922-7
- Christensen-Dalsgaard, M., Jørgensen, M. G., and Gerdes, K. (2010). Three new RelE-homologous mRNA interferases of *Escherichia coli* differentially induced by environmental stresses. *Mol. Microbiol.* 75, 333–348. doi: 10.1111/j.1365-2958.2009.06969.x
- Christensen-Dalsgaard, M., Overgaard, M., Winther, K. S., and Gerdes, K. (2008). RNA decay by messenger RNA interferases. *Methods Enzymol.* 447, 521–535. doi: 10.1016/S0076-6879(08)02225-8
- Cline, S. D., Saleem, S., and Daines, D. A. (2012). Regulation of the vapBC-1 toxin-antitoxin locus in nontypeable *Haemophilus influenzae*. *PLoS ONE* 7:e32199. doi: 10.1371/journal.pone.0032199



- Cruz, J. W., Rothenbacher, F. P., Maehigashi, T., Lane, W. S., Dunham, C. M., and Woychik, N. A. (2014). Doc toxin is a kinase that inactivates elongation factor Tu. *J. Biol. Chem.* 289, 7788–7798. doi: 10.1074/jbc.M113.544429
- Dalton, K. M., and Crosson, S. (2010). A conserved mode of protein recognition and binding in a ParD-ParE toxin-antitoxin complex. *Biochemistry* 49, 2205–2215. doi: 10.1021/bi902133s
- Das, U., Pogenberg, V., Subhramanyam, U. K. T., Wilmanns, M., Gourinath, S., and Srinivasan, A. (2014). Crystal structure of the VapBC-15 complex from *Mycobacterium tuberculosis* reveals a two-metal ion dependent PIN-domain ribonuclease and a variable mode of toxin-antitoxin assembly. *J. Struct. Biol.* 188, 249–258. doi: 10.1016/j.jsb.2014.10.002
- De Jonge, N., Garcia-Pino, A., Buts, L., Haesaerts, S., Charlier, D., Zangger, K., et al. (2009). Rejuvenation of CcdB-poisoned gyrase by an intrinsically disordered protein domain. *Mol. Cell* 35, 154–163. doi: 10.1016/j.molcel.2009.05.025
- De Jonge, N., Hohlweg, W., Garcia-Pino, A., Respondek, M., Buts, L., Haesaerts, S., et al. (2010). Structural and thermodynamic characterization of *Vibrio fischeri* CcdB. *J. Biol. Chem.* 285, 5606–5013. doi: 10.1074/jbc.M109.068429
- de la Cueva-Méndez, G., Mills, A. D., Clay-Farrace, L., Díaz-Orejas, R., and Laskey, R. A. (2003). Regulatable killing of eukaryotic cells by the prokaryotic proteins Kid and Kis. *EMBO J.* 22, 246–251. doi: 10.1093/emboj/cdg026
- de la Cueva-Méndez, G., and Pimentel, B. (2013). “Biotechnological and medical exploitation of toxin-antitoxin genes and their components,” in *Prokaryotic Toxin-Antitoxins*, ed K. Gerdes (Berlin, Heidelberg: Springer Berlin Heidelberg), 341–360.
- de la Hoz, A. B., Ayora, S., Sitkiewicz, I., Fernández, S., Pankiewicz, R., Alonso, J. C., et al. (2000). Plasmid copy-number control and better-than-random segregation genes of pSM19035 share a common regulator. *Proc. Natl. Acad. Sci. U.S.A.* 97, 728–733. doi: 10.1073/pnas.97.2.728
- Deane, S. M., and Rawlings, D. E. (2004). Plasmid evolution and interaction between the plasmid addiction stability systems of two related broad-host-range IncQ-like plasmids. *J. Bacteriol.* 186, 2123–2133. doi: 10.1128/JB.186.7.2123-2133.2004
- Diago-Navarro, E., Hernández-Arriaga, A. M., López-Villarejo, J., Muñoz-Gómez, A. J., Kamphuis, M. B., Boelens, R., et al. (2010). ParD toxin-antitoxin system of plasmid R1 - Basic contributions, biotechnological applications and relationships with closely-related toxin-antitoxin systems. *FEBS J.* 277, 3097–3117. doi: 10.1111/j.1742-4658.2010.07722.x
- Dienemann, C., Bøggild, A., Winther, K. S., Gerdes, K., and Brodersen, D. E. (2011). Crystal structure of the VapBC toxin-antitoxin complex from *Shigella flexneri* reveals a hetero-octameric DNA-binding assembly. *J. Mol. Biol.* 414, 713–722. doi: 10.1016/j.jmb.2011.10.024
- Eldholm, V., Johnsborg, O., Haugen, K., Ohnstad, H. S., and Håvarstein, L. S. (2009). Fratricide in *Streptococcus pneumoniae*: contributions and role of the cell wall hydrolases CbpD, LytA and LytC. *Microbiology* 155, 2223–2234. doi: 10.1099/mic.0.026328-0
- Engelberg-Kulka, H., Amitai, S., Kolodkin-Gal, I., and Hazan, R. (2006). Bacterial programmed cell death and multicellular behavior in bacteria. *PLoS Genet.* 2:e135. doi: 10.1371/journal.pgen.0020135
- Engelberg-Kulka, H., and Glaser, G. (1999). Addiction modules and programmed cell death and antideath in bacterial cultures. *Annu. Rev. Microbiol.* 53, 43–70. doi: 10.1146/annurev.micro.53.1.43
- Feng, S., Chen, Y., Kamada, K., Wang, H., Tang, K., Wang, M., et al. (2013). YoeB-ribosome structure: a canonical RNase that requires the ribosome for its specific activity. *Nucleic Acids Res.* 41, 9549–9556. doi: 10.1093/nar/gkt742
- Francuski, D., and Saenger, W. (2009). Crystal structure of the antitoxin-toxin protein complex RelB-RelE from *Methanococcus jannaschii*. *J. Mol. Biol.* 393, 898–908. doi: 10.1016/j.jmb.2009.08.048
- Garcia-Pino, A., Balasubramanian, S., Wyns, L., Gazit, E., De Greve, H., Magnuson, R. D., et al. (2010). Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. *Cell* 142, 101–111. doi: 10.1016/j.cell.2010.05.039
- Garcia-Pino, A., Christensen-Dalsgaard, M., Wyns, L., Yarmolinsky, M., Magnuson, R. D., Gerdes, K., et al. (2008). Doc of Prophage P1 is inhibited by its antitoxin partner phd through fold complementation. *J. Biol. Chem.* 283, 30821–30827. doi: 10.1074/jbc.M805654200
- Gelens, L., Hill, L., Vandervelde, A., Danckaert, J., and Loris, R. (2013). A general model for toxin-antitoxin module dynamics can explain persister cell formation in *E. coli*. *PLoS Comput. Biol.* 9:e1003190. doi: 10.1371/journal.pcbi.1003190
- Gerdes, K., Bech, F. W., Jørgensen, S. T., Løbner-Olesen, A., Rasmussen, P. B., Atlung, T., et al. (1986). Mechanism of postsegregational killing by the hok gene product of the parB system of plasmid R1 and its homology with the relF gene product of the *E. coli* relB operon. *EMBO J.* 5, 2023–2029.
- Gerdes, K., Christensen, S. K., and Løbner-Olesen, A. (2005). Prokaryotic toxin-antitoxin stress response loci. *Nat. Rev. Microbiol.* 3, 371–382. doi: 10.1038/nrmicro1147
- Germain, E., Castro-Roa, D., Zenkin, N., and Gerdes, K. (2013). Molecular mechanism of bacterial persistence by HipA. *Mol. Cell* 52, 248–254. doi: 10.1016/j.molcel.2013.08.045
- Gomis-Rüth, F. X., Solá, M., Acebo, P., Parraga, A., Guasch, A., Eritja, R., et al. (1998). The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator. *EMBO J.* 17, 7404–7415. doi: 10.1093/emboj/17.24.7404
- González Barrios, A. F., Zuo, R., Hashimoto, Y., Yang, L., Bentley, W. E., and Wood, T. K. (2006). Autoinducer 2 controls biofilm formation in *Escherichia coli* through a novel motility quorum-sensing regulator (MqsR, B3022). *J. Bacteriol.* 188, 305–316. doi: 10.1128/JB.188.1.305-316.2006
- González-Pastor, J. E. (2011). Cannibalism: a social behavior in sporulating *Bacillus subtilis*. *FEMS Microbiol. Rev.* 35, 415–424. doi: 10.1111/j.1574-6976.2010.00253.x
- Gotfredsen, M., and Gerdes, K. (1998). The *Escherichia coli* relBE genes belong to a new toxin-antitoxin gene family. *Mol. Microbiol.* 29, 1065–1076. doi: 10.1046/j.1365-2958.1998.00993.x
- Grady, R., and Hayes, F. (2003). Axe-Txe, a broad-spectrum proteic toxin-antitoxin system specified by a multidrug-resistant, clinical isolate of *Enterococcus faecium*. *Mol. Microbiol.* 47, 1419–1432. doi: 10.1046/j.1365-2958.2003.03387.x
- Gross, M., Marianovsky, I., and Glaser, G. (2006). MazG - A regulator of programmed cell death in *Escherichia coli*. *Mol. Microbiol.* 59, 590–601. doi: 10.1111/j.1365-2958.2005.04956.x
- Hadži, S., Garcia-Pino, A., Martínez-Rodríguez, S., Verschuere, K., Christensen-Dalsgaard, M., Gerdes, K., et al. (2013). Crystallization of the HigBA2 toxin-antitoxin complex from *Vibrio cholerae*. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* 69, 1052–1059. doi: 10.1107/S1744309113021490
- Hallez, R., Geeraerts, D., Sterckx, Y., Mine, N., Loris, R., and Van Melderen, L. (2010). New toxins homologous to ParE belonging to three-component toxin-antitoxin systems in *Escherichia coli* O157:H7. *Mol. Microbiol.* 76, 719–732. doi: 10.1111/j.1365-2958.2010.07129.x
- Harrison, J. J., Wade, W. D., Akierman, S., Vacchi-Suzzi, C., Stremick, C. A., Turner, R. J., et al. (2009). The Chromosomal Toxin Gene yafQ is a determinant of multidrug tolerance for *Escherichia coli* growing in a biofilm. *Antimicrob. Agents Chemother.* 53, 2253–2258. doi: 10.1128/AAC.00043-09
- Harvey, R. M., Stroeh, U. H., Ogunniyi, A. D., Smith-Vaughan, H. C., Leach, A. J., and Paton, J. C. (2011). A variable region within the genome of *Streptococcus pneumoniae* contributes to strain-strain variation in virulence. *PLoS ONE* 6:e19650. doi: 10.1371/journal.pone.0019650
- Hayes, F., and Kędzierska, B. (2014). Regulating toxin-antitoxin expression: controlled detonation of intracellular molecular timebombs. *Toxins (Basel)* 6, 337–358. doi: 10.3390/toxins6010337
- Hayes, F., and Van Melderen, L. (2011). Toxins-antitoxins: diversity, evolution and function. *Crit. Rev. Biochem. Mol. Biol.* 46, 386–408. doi: 10.3109/10409238.2011.600437
- Hernández-Arriaga, A. M., Chan, W. T., Espinosa, M., and Díaz-Orejas, R. (2014). Conditional activation of toxin-antitoxin systems: postsegregational killing and beyond. *Microbiol. Spectr.* 2:PLAS-0009-2013. doi: 10.1128/microbiolspec.PLAS-0009-2013
- Hurley, J. M., and Woychik, N. A. (2009). Bacterial toxin HigB associates with ribosomes and mediates translation-dependent mRNA cleavage at A-rich sites. *J. Biol. Chem.* 284, 18605–18613. doi: 10.1074/jbc.M109.008763
- Iannelli, F., Santoro, F., Oggioni, M. R., and Pozzi, G. (2014). Nucleotide sequence analysis of integrative conjugative element Tn5253 of *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* 58, 1235–1239. doi: 10.1128/AAC.01764-13
- Jaffe, A., Ogura, T., and Hiraga, S. (1985). Effects of the ccd function of the F plasmid on bacterial growth. *J. Bacteriol.* 163, 841–849.

- Jørgensen, M. G., Pandey, D. P., Jaskolska, M., and Gerdes, K. (2009). HicA of *Escherichia coli* defines a novel family of translation-independent mRNA interferases in bacteria and archaea. *J. Bacteriol.* 191, 1191–1199. doi: 10.1128/JB.01013-08
- Jurenaite, M., Markuckas, A., and Suziedeliene, E. (2013). Identification and characterization of type II toxin-antitoxin systems in the opportunistic pathogen *Acinetobacter baumannii*. *J. Bacteriol.* 195, 3165–3172. doi: 10.1128/JB.00237-13
- Kamada, K., and Hanaoka, F. (2005). Conformational change in the catalytic site of the ribonuclease YoeB toxin by YefM antitoxin. *Mol. Cell* 19, 497–509. doi: 10.1016/j.molcel.2005.07.004
- Kamada, K., Hanaoka, F., and Burley, S. K. (2003). Crystal Structure of the MazE/MazF Complex. *Mol. Cell* 11, 875–884. doi: 10.1016/S1097-2765(03)00097-2
- Kamphuis, M., Chiara Monti, M., H., van den Heuvel, R., Lopez-Villarejo, J., Diaz-Orejas, R., and Boelens, R. (2007b). Structure and function of bacterial kid-kis and related toxin-antitoxin systems. *Protein Pept. Lett.* 14, 113–124. doi: 10.2174/092986607779816096
- Kamphuis, M. B., Bonvin, A. M., Monti, M. C., Lemonnier, M., Muñoz-Gómez, A., Van Den Heuvel, R. H., et al. (2006). Model for RNA binding and the catalytic site of the RNase Kid of the bacterial parD toxin-antitoxin system. *J. Mol. Biol.* 357, 115–126. doi: 10.1016/j.jmb.2005.12.033
- Kamphuis, M. B., Monti, M. C., van den Heuvel, R. H. H., Santos-Sierra, S., Folkers, G. E., Lemonnier, M., et al. (2007a). Interactions between the toxin Kid of the bacterial parD system and the antitoxins Kis and MazE. *Proteins* 67, 219–231. doi: 10.1002/prot.21254
- Kasari, V., Kurg, K., Margus, T., Tenson, T., and Kaldalu, N. (2010). The *Escherichia coli* mqsR and ygiT genes encode a new toxin-antitoxin pair. *J. Bacteriol.* 192, 2908–2919. doi: 10.1128/JB.01266-09
- Kedzierska, B., Lian, L.-Y., and Hayes, F. (2007). Toxin-antitoxin regulation: bimodal interaction of YefM-YoeB with paired DNA palindromes exerts transcriptional autorepression. *Nucleic Acids Res.* 35, 325–339. doi: 10.1093/nar/gkl1028
- Kędzierska, B., and Hayes, F. (2016). “Transcriptional control of toxin-antitoxin expression: keeping toxins under wraps until the time is right,” in *Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria*, ed F. J. de Bruijn (Hoboken, NJ: John Wiley & Sons, Inc.), 463–472.
- Kho, S. K., Loll, B., Chan, W. T., Shoeman, R. L., Ngoo, L., Yeo, C. C., et al. (2007). Molecular and structural characterization of the PezAT chromosomal toxin-antitoxin system of the human pathogen *Streptococcus pneumoniae*. *J. Biol. Chem.* 282, 19606–19618. doi: 10.1074/jbc.M701703200
- Kim, Y., Wang, X., Zhang, X.-S., Grigoriu, S., Page, R., Peti, W., et al. (2010). *Escherichia coli* toxin/antitoxin pair MqsR/MqsA regulate toxin CspD. *Environ. Microbiol.* 12, 1105–1121. doi: 10.1111/j.1462-2920.2009.02147.x
- Koga, M., Otsuka, Y., Lemire, S., and Yonesaki, T. (2011). *Escherichia coli* rnlA and rnlB compose a novel toxin-antitoxin system. *Genetics* 187, 123–130. doi: 10.1534/genetics.110.121798
- Kolodkin-Gal, I., Hazan, R., Gaathon, A., Carmeli, S., and Engelberg-Kulka, H. (2007). A linear pentapeptide is a quorum-sensing factor required for mazEF-mediated cell death in *Escherichia coli*. *Science* 318, 652–655. doi: 10.1126/science.1147248
- Krügel, H., Klimina, K. M., Mrotzek, G., Tretyakov, A., Schöfl, G., Saluz, H.-P., et al. (2015). Expression of the toxin-antitoxin genes yefMLrh, yoeBLrh in human *Lactobacillus rhamnosus* isolates. *J. Basic Microbiol.* 55, 982–991. doi: 10.1002/jobm.201400904
- Kumar, P., Issac, B., Dodson, E. J., Turkenburg, J. P., and Mande, S. C. (2008). Crystal structure of *Mycobacterium tuberculosis* YefM antitoxin reveals that it is not an intrinsically unstructured protein. *J. Mol. Biol.* 383, 482–493. doi: 10.1016/j.jmb.2008.08.067
- Lee, B., Holkenbrink, C., Treuner-lange, A., and Higgs, P. I. (2012). *Myxococcus xanthus* developmental cell fate production: heterogeneous accumulation of developmental regulatory proteins and reexamination of the Role of MazF in developmental lysis. *J. Bacteriol.* 194, 3058–3068. doi: 10.1128/JB.06756-11
- Lee, I.-G., Lee, S. J., Chae, S., Lee, K.-Y., Kim, J.-H., and Lee, B.-J. (2015). Structural and functional studies of the *Mycobacterium tuberculosis* VapBC30 toxin-antitoxin system: implications for the design of novel antimicrobial peptides. *Nucleic Acids Res.* 43, 7624–7637. doi: 10.1093/nar/gkv689
- Lehnherr, H., Maguin, E., Jafri, S., and Yarmolinsky, M. B. (1993). Plasmid addiction genes of bacteriophage P1: doc, which causes cell death on curing of prophage, and phd, which prevents host death when prophage is retained. *J. Mol. Biol.* 233, 414–428. doi: 10.1006/jmbi.1993.1521
- Lehnherr, H., and Yarmolinsky, M. B. (1995). Addiction protein Phd of plasmid prophage P1 is a substrate of the ClpXP serine protease of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 92, 3274–3277. doi: 10.1073/pnas.92.8.3274
- Leplae, R., Geeraerts, D., Hallez, R., Guglielmini, J., Drze, P., and Van Melderen, L. (2011). Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic Acids Res.* 39, 5513–5525. doi: 10.1093/nar/gkr131
- Lewis, K. (2010). Persister cells. *Annu. Rev. Microbiol.* 64, 357–372. doi: 10.1146/annurev.micro.112408.134306
- Li, G.-Y., Zhang, Y., Inouye, M., and Ikura, M. (2009). Inhibitory mechanism of *Escherichia coli* RelE-RelB toxin-antitoxin module involves a helix displacement near an mRNA interferase active site. *J. Biol. Chem.* 284, 14628–14636. doi: 10.1074/jbc.M809656200
- Liang, Y., Gao, Z., Wang, F., Zhang, Y., Dong, Y., and Liu, Q. (2014). Structural and functional characterization of *Escherichia coli* toxin-antitoxin complex DinJ-YafQ. *J. Biol. Chem.* 289, 21191–21202. doi: 10.1074/jbc.M114.559773
- Lioy, V. S., Machon, C., Tabone, M., Gonzalez-Pastor, J. E., Dauglavicius, R., Ayora, S., et al. (2012). The  $\zeta$  toxin induces a set of protective responses and dormancy. *PLoS ONE* 7:e30282. doi: 10.1371/journal.pone.0030282
- Liu, M., Zhang, Y., Inouye, M., and Woychik, N., a (2008). Bacterial addiction module toxin Doc inhibits translation elongation through its association with the 30S ribosomal subunit. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5885–5890. doi: 10.1073/pnas.0711949105
- Loris, R., Marianovsky, I., Lah, J., Laeremans, T., Engelberg-Kulka, H., Glaser, G., et al. (2003). Crystal structure of the intrinsically flexible addiction antitoxin MazE. *J. Biol. Chem.* 278, 28252–28257. doi: 10.1074/jbc.M302336200
- Madl, T., Van Melderen, L., Mine, N., Respondek, M., Oberer, M., Keller, W., et al. (2006). Structural basis for nucleic acid and toxin recognition of the bacterial antitoxin CcdA. *J. Mol. Biol.* 364, 170–185. doi: 10.1016/j.jmb.2006.08.082
- Maisonneuve, E., Castro-Camargo, M., and Gerdes, K. (2013). (p)ppGpp controls bacterial persistence by stochastic induction of toxin-antitoxin activity. *Cell* 154, 1140–1150. doi: 10.1016/j.cell.2013.07.048
- Makarova, K. S., Wolf, Y. I., Snir, S., and Koonin, E. V. (2011). Defense Islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.* 193, 6039–6056. doi: 10.1128/JB.05535-11
- Marianovsky, I., Aizenman, E., Engelberg-Kulka, H., and Glaser, G. (2001). The Regulation of the *Escherichia coli* mazEF promoter involves an unusual alternating palindrome. *J. Biol. Chem.* 276, 5975–5984. doi: 10.1074/jbc.M008832200
- Markovski, M., and Wickner, S. (2013). Preventing bacterial suicide: a novel toxin-antitoxin strategy. *Mol. Cell* 52, 611–612. doi: 10.1016/j.molcel.2013.1.018
- Masuda, H., Tan, Q., Awano, N., Wu, K.-P., and Inouye, M. (2012). YeeU enhances the bundling of cytoskeletal polymers of MreB and FtsZ, antagonizing the CbtA (YeeV) toxicity in *Escherichia coli*. *Mol. Microbiol.* 84, 979–989. doi: 10.1111/j.1365-2958.2012.08068.x
- Maté, M. J., Vincentelli, R., Foos, N., Raoult, D., Cambillau, C., and Ortiz-Lombardia, M. (2012). Crystal structure of the DNA-bound VapBC2 antitoxin/toxin pair from *Rickettsia felis*. *Nucleic Acids Res.* 40, 3245–3258. doi: 10.1093/nar/gkr1167
- Mattison, K., Wilbur, J. S., So, M., and Brennan, R. G. (2006). Structure of FitAB from *Neisseria gonorrhoeae* bound to DNA reveals a tetramer of toxin-antitoxin heterodimers containing pin domains and ribbon-helix-helix motifs. *J. Biol. Chem.* 281, 37942–37951. doi: 10.1074/jbc.M605198200
- McKenzie, J. L., Duyvestyn, J. M., Smith, T., Bendak, K., Mackay, J., Cursons, R., et al. (2012). Determination of ribonuclease sequence-specificity using Pentaprobates and mass spectrometry. *RNA* 18, 1267–1278. doi: 10.1261/rna.031229.111
- Meinhart, A., Alonso, J. C., Sträter, N., and Saenger, W. (2003). Crystal structure of the plasmid maintenance system  $\epsilon$  /  $\epsilon$ ? : functional mechanism of toxin  $\zeta$  and inactivation by  $\epsilon$ 2 $\zeta$  complex formation. *Proc. Natl. Acad. Sci. U.S.A.* 100, 1661–1666. doi: 10.1073/pnas.0434325100

- Melničáková, J., Bešárová, Z., Makroczyová, J., and Barák, I. (2015). Analysis of the *Bacillus cereus* SpoIIS antitoxin-toxin system reveals its three-component nature. *Front. Microbiol.* 6:808. doi: 10.3389/fmicb.2015.00808
- Miallau, L., Faller, M., Chiang, J., Arbing, M., Guo, F., Cascio, D., et al. (2009). Structure and proposed activity of a member of the VapBC family of toxin-antitoxin systems. VapBC-5 from *Mycobacterium tuberculosis*. *J. Biol. Chem.* 284, 276–83. doi: 10.1074/jbc.M805061200
- Millá, M. E., Brown, B. M., Waldburger, C. D., and Sauer, R. T. (1995). P22 Arc repressor: transition state properties inferred from mutational effects on the rates of protein unfolding and refolding. *Biochemistry* 34, 13914–13919. doi: 10.1021/bi00042a024
- Min, A. B., Miallau, L., Sawaya, M. R., Habel, J., Cascio, D., and Eisenberg, D. (2012). The crystal structure of the Rv0301-Rv0300 VapBC-3 toxin-antitoxin complex from *M. tuberculosis* reveals a  $Mg^{2+}$  ion in the active site and a putative RNA-binding site. *Protein Sci.* 21, 1754–1767. doi: 10.1002/pro.2161
- Mine, N., Guglielmini, J., Wilbaux, M., and Van Melderen, L. (2009). The decay of the chromosomally encoded ccdO157 toxin-antitoxin system in the *Escherichia coli* species. *Genetics* 181, 1557–1566. doi: 10.1534/genetics.108.095190
- Monti, M. C., Hernández-Arriaga, A. M., Kamphuis, M. B., López-Villarejo, J., Heck, A. J. R., Boelens, R., et al. (2007). Interactions of Kid-Kis toxin-antitoxin complexes with the parD operator-promoter region of plasmid R1 are piloted by the Kis antitoxin and tuned by the stoichiometry of Kid-Kis oligomers. *Nucleic Acids Res.* 35, 1737–1749. doi: 10.1093/nar/gkm073
- Moreno-Córdoba, I., Diago-Navarro, E., Barendregt, A., Heck, A. J. R., Alfonso, C., Díaz-Orejas, R., et al. (2012). The toxin-antitoxin proteins relBE2Spn of *Streptococcus pneumoniae*: characterization and association to their DNA target. *Proteins* 80, 1834–1846. doi: 10.1002/prot.24081
- Murayama, K., Orth, P., de la Hoz, A. B., Alonso, J. C., and Saenger, W. (2001). Crystal structure of omega transcriptional repressor encoded by *Streptococcus pyogenes* plasmid pSM19035 at 1.5 Å resolution. *J. Mol. Biol.* 314, 789–796. doi: 10.1006/jmbi.2001.5157
- Mutschler, H., Gebhardt, M., Shoeman, R. L., and Meinhart, A. (2011). A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis. *PLoS Biol.* 9:e1001033. doi: 10.1371/journal.pbio.1001033
- Mutschler, H., and Meinhart, A. (2011).  $\epsilon/\zeta$  systems: their role in resistance, virulence, and their potential for antibiotic development. *J. Mol. Med.* 89, 1183–1194. doi: 10.1007/s00109-011-0797-4
- Mutschler, H., and Meinhart, A. (2013). “Type II Toxin-antitoxin loci: the epsilon/zeta family,” in *Prokaryotic Toxin-Antitoxins*, ed K. Gerdes (Berlin, Heidelberg: Springer Berlin Heidelberg), 205–223.
- Nariya, H., and Inouye, M. (2008). MazF, an mRNA Interferase, mediates programmed cell death during multicellular myxococcus development. *Cell* 132, 55–66. doi: 10.1016/j.cell.2007.11.044
- Nicolas, P., Mäder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., et al. (2012). Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science* 335, 1103–1106. doi: 10.1126/science.1206848
- Nieto, C., Cherny, I., Khoo, S. K., De Lacoba, M. G., Chan, W. T., Yeo, C. C., et al. (2007). The yefM-yoeB toxin-antitoxin systems of *Escherichia coli* and *Streptococcus pneumoniae*: functional and structural correlation. *J. Bacteriol.* 189, 1266–1278. doi: 10.1128/JB.01130-06
- Nolle, N., Schuster, C. F., and Bertram, R. (2013). Two paralogous yefM-yoeB loci from *Staphylococcus equorum* encode functional toxin-antitoxin systems. *Microbiology* 159, 1575–1585. doi: 10.1099/mic.0.068049-0
- Norton, J. P., and Mulvey, M. A. (2012). Toxin-antitoxin systems are important for niche-specific colonization and stress resistance of uropathogenic *Escherichia coli*. *PLoS Pathog.* 8:e1002954. doi: 10.1371/journal.ppat.1002954
- Oberer, M., Zangger, K., Gruber, K., and Keller, W. (2007). The solution structure of ParD, the antidote of the ParDE toxin antitoxin module, provides the structural basis for DNA and toxin binding. *Protein Sci.* 16, 1676–1688. doi: 10.1110/ps.062680707
- Otsuka, Y. (2016). Prokaryotic toxin-antitoxin systems: novel regulations of the toxins. *Curr. Genet.* doi: 10.1007/s00294-015-0557-z. [Epub ahead of print].
- Overgaard, M., Borch, J., and Gerdes, K. (2009). RelB and RelE of *Escherichia coli* form a tight complex that represses transcription via the ribbon-helix-helix motif in RelB. *J. Mol. Biol.* 394, 183–196. doi: 10.1016/j.jmb.2009.09.006
- Overgaard, M., Borch, J., Jørgensen, M. G., and Gerdes, K. (2008). Messenger RNA interferase RelE controls relBE transcription by conditional cooperativity. *Mol. Microbiol.* 69, 841–857. doi: 10.1111/j.1365-2958.2008.06313.x
- Pan, C. Q., Finkel, S. E., Cramton, S. E., Feng, J. A., Sigman, D. S., and Johnson, R. C. (1996). Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. *J. Mol. Biol.* 264, 675–695. doi: 10.1006/jmbi.1996.0669
- Pandey, D. P., and Gerdes, K. (2005). Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res.* 33, 966–976. doi: 10.1093/nar/gki201
- Pedersen, K., Zavialov, A. V., Pavlov, M. Y., Elf, J., Gerdes, K., and Ehrenberg, M. (2003). The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. *Cell* 112, 131–140. doi: 10.1016/S0092-8674(02)01248-5
- Pimentel, B., Madine, M. A., and de la Cueva-Méndez, G. (2005). Kid cleaves specific mRNAs at UUACU sites to rescue the copy number of plasmid R1. *EMBO J.* 24, 3459–3469. doi: 10.1038/sj.emboj.7600815
- Połom, D., Boss, L., Węgrzyn, G., Hayes, F., and Kędzierska, B. (2013). Amino acid residues crucial for specificity of toxin-antitoxin interactions in the homologous Axe-Txe and YefM-YoeB complexes. *FEBS J.* 280, 5906–5918. doi: 10.1111/febs.12517
- Pomerantsev, A. P., Golovliov, I. R., Ohara, Y., Mokrievich, A. N., Obuchi, M., Norqvist, A., et al. (2001). Genetic organization of the Francisella plasmid pFNL10. *Plasmid* 46, 210–222. doi: 10.1006/plas.2001.1548
- Preston, M. A., Pimentel, B., Bermejo-Rodríguez, C., Dionne, I., Turnbull, A., and de la Cueva-Méndez, G. (2015). Repurposing a prokaryotic Toxin-Antitoxin system for the selective killing of oncogenically stressed human cells. *ACS Synth. Biol.* doi: 10.1021/acssynbio.5b00096. [Epub ahead of print].
- Ramage, H. R., Connolly, L. E., and Cox, J. S. (2009). Comprehensive functional analysis of *Mycobacterium tuberculosis* toxin-antitoxin systems: implications for pathogenesis, stress responses, and evolution. *PLoS Genet.* 5:e1000767. doi: 10.1371/journal.pgen.1000767
- Rawlings, D. E. (1999). Proteic toxin-antitoxin, bacterial plasmid addiction systems and their evolution with special reference to the pas system of pTF-FC2. *FEMS Microbiol. Lett.* 176, 269–277. doi: 10.1111/j.1574-6968.1999.tb13672.x
- Ren, D., Bedzyk, L. A., Ye, R. W., Thomas, S. M., and Wood, T. K. (2004). Stationary-phase quorum-sensing signals affect autoinducer-2 and gene expression in *Escherichia coli*. *Appl. Environ. Microbiol.* 70, 2038–2043. doi: 10.1128/AEM.70.4.2038-2043.2004
- Robinson, M., Son, B., Kroos, D., and Kroos, L. (2014). Transcription factor MrpC binds to promoter regions of hundreds of developmentally-regulated genes in *Myxococcus xanthus*. *BMC Genomics* 15:1123. doi: 10.1186/1471-2164-15-1123
- Robson, J., McKenzie, J. L., Cursons, R., Cook, G. M., and Arcus, V. L. (2009). The vapBC operon from *Mycobacterium smegmatis* is an autoregulated toxin-antitoxin module that controls growth via inhibition of translation. *J. Mol. Biol.* 390, 353–367. doi: 10.1016/j.jmb.2009.05.006
- Rocker, A., and Meinhart, A. (2015a). A cis-acting antitoxin domain within the chromosomal toxin-antitoxin module EzeT of *Escherichia coli* quenches toxin activity. *Mol. Microbiol.* 97, 589–604. doi: 10.1111/mmi.13051
- Rocker, A., and Meinhart, A. (2015b). Type II toxin: antitoxin systems. More than small selfish entities? *Curr. Genet.* doi: 10.1007/s00294-015-0541-7. [Epub ahead of print].
- Rowe-Magnus, D. A., Guerout, A.-M., Biskri, L., Bouige, P., and Mazel, D. (2003). Comparative analysis of superintegrons: engineering extensive genetic diversity in the Vibrionaceae. *Genome Res.* 13, 428–442. doi: 10.1101/gr.617103
- Ruangprasert, A., Maehigashi, T., Miles, S. J., Giridharan, N., Liu, J. X., and Dunham, C. M. (2014). Mechanisms of toxin inhibition and transcriptional repression by *Escherichia coli* DinJ-YafQ. *J. Biol. Chem.* 289, 20559–20569. doi: 10.1074/jbc.M114.573006
- Saavedra De Bast, M., Mine, N., and Van Melderen, L. (2008). Chromosomal toxin-antitoxin systems may act as antiaddiction modules. *J. Bacteriol.* 190, 4603–4809. doi: 10.1128/JB.00357-08
- Schumacher, M. A., Piro, K. M., Xu, W., Hansen, S., Lewis, K., and Brennan, R. G. (2009). Molecular mechanisms of HipA-mediated multidrug tolerance and its neutralization by HipB. *Science* 323, 396–401. doi: 10.1126/science.1163806
- Schureck, M. A., Maehigashi, T., Miles, S. J., Marquez, J., Cho, S. E., Erdman, R., et al. (2014). Structure of the *Proteus vulgaris* HigB-(HigA)2-HigB toxin-antitoxin complex. *J. Biol. Chem.* 289, 1060–1070. doi: 10.1074/jbc.M113.512095



- Sevillano, L., Díaz, M., Yamaguchi, Y., Inouye, M., and Santamaría, R. I. (2012). Identification of the first functional toxin-antitoxin system in *Streptomyces*. *PLoS ONE* 7:e32977. doi: 10.1371/journal.pone.0032977
- Shao, Y., Harrison, E. M., Bi, D., Tai, C., He, X., Ou, H.-Y., et al. (2011). TADB: a web-based resource for Type 2 toxin-antitoxin loci in bacteria and archaea. *Nucleic Acids Res.* 39, D606–D611. doi: 10.1093/nar/gkq908
- Shimazu, T., Mirochnitchenko, O., Phadtare, S., and Inouye, M. (2014). Regression of solid tumors by induction of MazF, a bacterial mRNA endoribonuclease. *J. Mol. Microbiol. Biotechnol.* 24, 228–233. doi: 10.1159/000365509
- Smith, A. S., and Rawlings, D. E. (1997). The poison-antidote stability system of the broad-host-range *Thiobacillus ferrooxidans* plasmid pTF-FC2. *Mol. Microbiol.* 26, 961–970. doi: 10.1046/j.1365-2958.1997.6332000.x
- Smith, A. S., and Rawlings, D. E. (1998a). Autoregulation of the pTF-FC2 proteic poison-antidote plasmid addiction system (pas) is essential for plasmid stabilization. *J. Bacteriol.* 180, 5463–5465.
- Smith, A. S., and Rawlings, D. E. (1998b). Efficiency of the pTF-FC2 pas poison-antidote stability system in *Escherichia coli* is affected by the host strain, and antidote degradation requires the lon protease. *J. Bacteriol.* 180, 5458–5462.
- Soo, V. W. C., and Wood, T. K. (2013). Antitoxin MqsA represses curli formation through the master biofilm regulator CsgD. *Sci. Rep.* 3, 3186. doi: 10.1038/srep03186
- Stieber, D., Gabant, P., and Szpirer, C. Y. (2008). The art of selective killing: plasmid toxin/antitoxin systems and their technological applications. *Biotechniques* 45, 344–346. doi: 10.2144/000112955
- Szekeres, S., Dauti, M., Wilde, C., Mazel, D., and Rowe-Magnus, D. A. (2007). Chromosomal toxin-antitoxin loci can diminish large-scale genome reductions in the absence of selection. *Mol. Microbiol.* 63, 1588–1605. doi: 10.1111/j.1365-2958.2007.05613.x
- Tabone, M., Ayora, S., and Alonso, J. C. (2014). Toxin  $\zeta$  reversibly induces dormancy and reduces the UDP-N-acetylglucosamine pool as one of the protective responses to cope with stress. *Toxins (Basel)*. 6, 2787–2803. doi: 10.3390/toxins6092787
- Takagi, H., Kakuta, Y., Okada, T., Yao, M., Tanaka, I., and Kimura, M. (2005). Crystal structure of archaeal toxin-antitoxin RelE-RelB complex with implications for toxin activity and antitoxin effects. *Nat. Struct. Mol. Biol.* 12, 327–331. doi: 10.1038/nsmb911
- Tamman, H., Ainelo, A., Ainsaar, K., and Hörak, R. (2014). A moderate toxin, GraT, modulates growth rate and stress tolerance of *Pseudomonas putida*. *J. Bacteriol.* 196, 157–169. doi: 10.1128/JB.00851-13
- Tanouchi, Y., Lee, A. J., Meredith, H., and You, L. (2013). Programmed cell death in bacteria and implications for antibiotic therapy. *Trends Microbiol.* 21, 265–270. doi: 10.1016/j.tim.2013.04.001
- Tian, Q. B., Ohnishi, M., Tabuchi, A., and Terawaki, Y. (1996). A new plasmid-encoded proteic killer gene system: cloning, sequencing, and analyzing hlg locus of plasmid Rts1. *Biochem. Biophys. Res. Commun.* 220, 280–284. doi: 10.1006/bbrc.1996.0396
- Unterholzner, S. J., Poppenberger, B., and Rozhon, W. (2013). Toxin-antitoxin systems: biology, identification, and application. *Mob. Genet. Elements* 3:e26219. doi: 10.4161/mge.26219
- Van Melder, L., and Saavedra De Bast, M. (2009). Bacterial toxin-antitoxin systems: more than selfish entities? *PLoS Genet.* 5:e1000437. doi: 10.1371/journal.pgen.1000437
- Van Melder, L. (2010). Toxin-antitoxin systems: why so many, what for? *Curr. Opin. Microbiol.* 13, 781–785. doi: 10.1016/j.mib.2010.10.006
- Vesper, O., Amitai, S., Belitsky, M., Byrgazov, K., Kaberdina, A. C., Engelberg-Kulka, H., et al. (2011). Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in *Escherichia coli*. *Cell* 147, 147–157. doi: 10.1016/j.cell.2011.07.047
- Volante, A., Carrasco, B., Tabone, M., and Alonso, J. C. (2015). The interaction of  $\omega 2$  with the RNA polymerase  $\beta'$  subunit functions as an activation to repression switch. *Nucleic Acids Res.* 43, 9249–9261. doi: 10.1093/nar/gkv788
- Volante, A., Soberón, N. E., Ayora, S., and Alonso, J. C. (2014). The interplay between different stability systems contributes to faithful segregation: *Streptococcus pyogenes* pSM19035 as a Model. *Microbiol. Spectr.* 2:PLAS-0007-2013. doi: 10.1128/microbiolspec.PLAS-0007-2013
- Wang, X., Kim, Y., Hong, S. H., Ma, Q., Brown, B. L., Pu, M., et al. (2011). Antitoxin MqsA helps mediate the bacterial general stress response. *Nat. Chem. Biol.* 7, 359–366. doi: 10.1038/nchembio.560
- Wang, X., Lord, D. M., Cheng, H.-Y., Osbourne, D. O., Hong, S. H., Sanchez-Torres, V., et al. (2012). A new type V toxin-antitoxin system where mRNA for toxin GhoT is cleaved by antitoxin GhoS. *Nat. Chem. Biol.* 8, 855–861. doi: 10.1038/nchembio.1062
- Wang, X., Lord, D. M., Hong, S. H., Peti, W., Benedik, M. J., Page, R., et al. (2013). Type II toxin/antitoxin MqsR/MqsA controls type V toxin/antitoxin GhoT/GhoS. *Environ. Microbiol.* 15, 1734–1744. doi: 10.1111/1462-2920.12063
- Weihofen, W. A., Cicek, A., Pratto, F., Alonso, J. C., and Saenger, W. (2006). Structures of omega repressors bound to direct and inverted DNA repeats explain modulation of transcription. *Nucleic Acids Res.* 34, 1450–1458. doi: 10.1093/nar/gkl015
- Wen, Y., Behiels, E., Felix, J., Elegheert, J., Vergauwen, B., Devreese, B., et al. (2014). The bacterial antitoxin HipB establishes a ternary complex with operator DNA and phosphorylated toxin HipA to regulate bacterial persistence. *Nucleic Acids Res.* 42, 10134–10147. doi: 10.1093/nar/gku665
- Wieteska, L., Skulimowski, A., Cybula, M., and Szmaj, J. (2014). Toxins vapC and pasB from prokaryotic TA modules remain active in mammalian cancer cells. *Toxins (Basel)*. 6, 2948–2961. doi: 10.3390/toxins6102948
- Williams, J. J., and Hergenrother, P. J. (2012). Artificial activation of toxin-antitoxin systems as an antibacterial strategy. *Trends Microbiol.* 20, 291–298. doi: 10.1016/j.tim.2012.02.005
- Winther, K. S., and Gerdes, K. (2011). Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7403–7407. doi: 10.1073/pnas.1019587108
- Winther, K. S., and Gerdes, K. (2012). Regulation of enteric vapBC transcription: Induction by VapC toxin dimer-breaking. *Nucleic Acids Res.* 40, 4347–4357. doi: 10.1093/nar/gks029
- Yamaguchi, Y., and Inouye, M. (2009). mRNA interferases, sequence-specific endoribonucleases from the toxin-antitoxin systems. *Prog. Mol. Biol. Transl. Sci.* 85, 467–500. doi: 10.1016/S0079-6603(08)00812-X
- Yamaguchi, Y., Park, J.-H., and Inouye, M. (2009). MqsR, a crucial regulator for quorum sensing and biofilm formation, is a GCU-specific mRNA interferase in *Escherichia coli*. *J. Biol. Chem.* 284, 28746–28753. doi: 10.1074/jbc.M109.032904
- Yamaguchi, Y., Park, J.-H., and Inouye, M. (2011). Toxin-antitoxin systems in bacteria and archaea. *Annu. Rev. Genet.* 45, 61–79. doi: 10.1146/annurev-genet-110410-132412
- Yeo, C. C., Abu Bakar, F., Chan, W. T., Espinosa, M., and Hari Krishna, J. A. (2016). Heterologous expression of toxins from bacterial toxin-antitoxin systems in eukaryotic cells: strategies and applications. *Toxins* 8:49. doi: 10.3390/toxins8020049
- Yoshizumi, S., Zhang, Y., Yamaguchi, Y., Chen, L., Kreiswirth, B. N., and Inouye, M. (2009). *Staphylococcus aureus* YoeB homologues inhibit translation initiation. *J. Bacteriol.* 191, 5868–5872. doi: 10.1128/JB.00623-09
- Yuan, J., Sterckx, Y., Mitchenall, L. A., Maxwell, A., Loris, R., and Waldor, M. K. (2010). *Vibrio cholerae* ParE2 poisons DNA gyrase via a mechanism distinct from other gyrase inhibitors. *J. Biol. Chem.* 285, 40397–40408. doi: 10.1074/jbc.M110.138776
- Zhang, Y., Zhang, J., Hoeflich, K. P., Ikura, M., Qing, G., and Inouye, M. (2003). MazF Cleaves Cellular mRNAs Specifically at ACA to Block Protein Synthesis in *Escherichia coli*. *Mol. Cell* 12, 913–923. doi: 10.1016/S1097-2765(03)00402-7
- Zheng, C., Xu, J., Ren, S., Li, J., Xia, M., Chen, H., et al. (2015). Identification and characterization of the chromosomal yefM-yoeB toxin-antitoxin system of *Streptococcus suis*. *Sci. Rep.* 5:13125. doi: 10.1038/srep13125
- Zorzini, V., Buts, L., Schrank, E., Sterckx, Y. G. J., Respondek, M., Engelberg-Kulka, H., et al. (2015). *Escherichia coli* antitoxin MazE as transcription factor: insights into MazE-DNA binding. *Nucleic Acids Res.* 43, 1241–1256. doi: 10.1093/nar/gku1352

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Chan, Espinosa and Yeo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# The *Verrucomicrobia* LexA-Binding Motif: Insights into the Evolutionary Dynamics of the SOS Response

Ivan Erill<sup>1</sup>, Susana Campoy<sup>2</sup>, Sefa Kılıç<sup>1</sup> and Jordi Barbé<sup>2\*</sup>

<sup>1</sup> Erill Lab, Department of Biological Sciences, University of Maryland Baltimore County, Baltimore, MD, USA, <sup>2</sup> Unitat de Microbiologia, Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Barcelona, Spain

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Centro de Investigaciones Biológicas,  
Spain

### Reviewed by:

Juan Carlos Alonso,  
Centro Nacional de Biotecnología,  
Spain

José R. Penadés,  
University of Glasgow, UK

### \*Correspondence:

Jordi Barbé  
jordi.barbe@uab.cat

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 27 May 2016

**Accepted:** 04 July 2016

**Published:** 20 July 2016

### Citation:

Erill I, Campoy S, Kılıç S and Barbé J  
(2016) The *Verrucomicrobia*  
LexA-Binding Motif: Insights into the  
Evolutionary Dynamics of the SOS  
Response. *Front. Mol. Biosci.* 3:33.  
doi: 10.3389/fmolb.2016.00033

The SOS response is the primary bacterial mechanism to address DNA damage, coordinating multiple cellular processes that include DNA repair, cell division, and translesion synthesis. In contrast to other regulatory systems, the composition of the SOS genetic network and the binding motif of its transcriptional repressor, LexA, have been shown to vary greatly across bacterial clades, making it an ideal system to study the co-evolution of transcription factors and their regulons. Leveraging comparative genomics approaches and prior knowledge on the core SOS regulon, here we define the binding motif of the *Verrucomicrobia*, a recently described phylum of emerging interest due to its association with eukaryotic hosts. Site directed mutagenesis of the *Verrucomicrobium spinosum recA* promoter confirms that LexA binds a 14 bp palindromic motif with consensus sequence TGTC-N4-GAACA. Computational analyses suggest that recognition of this novel motif is determined primarily by changes in base-contacting residues of the third alpha helix of the LexA helix-turn-helix DNA binding motif. In conjunction with comparative genomics analysis of the LexA regulon in the *Verrucomicrobia* phylum, electrophoretic shift assays reveal that LexA binds to operators in the promoter region of DNA repair genes and a mutagenesis cassette in this organism, and identify previously unreported components of the SOS response. The identification of tandem LexA-binding sites generating instances of other LexA-binding motifs in the *lexA* gene promoter of *Verrucomicrobia* species leads us to postulate a novel mechanism for LexA-binding motif evolution. This model, based on gene duplication, successfully addresses outstanding questions in the intricate co-evolution of the LexA protein, its binding motif and the regulatory network it controls.

**Keywords:** SOS response, LexA regulon, DNA repair, comparative genomics, binding motif, translesion synthesis, uracil-DNA glycosylase, regulatory network evolution

## INTRODUCTION

The SOS response is the primary mechanism for coordinating the response to DNA damage in *Bacteria* (Erill et al., 2007). First reported in *Escherichia coli*, (Little and Mount, 1982), the SOS response has been documented in a broad range of bacterial species (Erill et al., 2007). In *E. coli* and *Bacillus subtilis*, the SOS response has been shown to regulate between 30 and 40 genes involved in DNA repair, translesion synthesis, and cell-division arrest (Fernandez De Henestrosa et al., 2000; Walker et al., 2000; Au et al., 2005). This regulatory network is governed by the transcriptional

repressor LexA, which in *E. coli* binds as a homodimer to specific sites upstream of regulated operons and blocks transcription initiation (Thliveris et al., 1991; Walker et al., 2000). Upon DNA damage, the RecA protein binds single-stranded DNA (ssDNA) fragments originating at stalled replication forks, and forms active nucleoprotein filaments capable of promoting self-cleavage of the LexA repressor (Sassanfar and Roberts, 1990). Self-cleavage of the LexA dimer leads to de-repression of target operons, which typically include the *lexA* and *recA* genes (Little, 1991), and full induction of the system (Walker et al., 2000). In recent years, the SOS response has attracted increasing interest due to its active involvement in the regulation of mobile genetic elements, such as integrative and conjugative elements (Beaber et al., 2004), pathogenicity islands (Ubeda et al., 2007), and integron integrases (Guerin et al., 2009), as well as its induction by different types of antibiotics (Beaber et al., 2004; Ubeda et al., 2005; Maiques et al., 2006).

Beyond its clinical interest, the SOS response also constitutes a unique model for the study of the evolution of transcriptional regulatory networks. In contrast with many other transcriptional regulators, the LexA repressor displays remarkably different binding motifs across multiple phyla, changing both the specificity of the dyad region recognized by each LexA monomer as well as the dyad space (Erill et al., 2007). Reported LexA-binding motifs range from short inverted repeats (GAAC-N4-GTTC) in the *Firmicutes* and *Actinobacteria* (Davis et al., 2002; Au et al., 2005), to larger palindromic motifs (CTGT-N8-ACAG) in the *Gammaproteobacteria* (Fernandez De Henestrosa et al., 2000; Erill et al., 2003) and even direct repeat motifs (GTTC-N7-GTTC) in the *Alphaproteobacteria* (Fernandez de Henestrosa et al., 1998; Erill et al., 2004). This variability in LexA-binding motifs is matched by extreme plasticity in the size and composition of the SOS regulatory network, which can regulate from 3 to 40 genes (Fernandez De Henestrosa et al., 2000; Au et al., 2005; Campoy et al., 2005) and has been shown to broadly comprise a minimal shared SOS regulon core consisting of *lexA*, *recA*, and a mutagenesis gene cassette (*imuA-imuB-dnaE2*; Erill et al., 2006).

Named after *Verrucomicrobium spinosum*, the *Verrucomicrobia* are a recently established bacterial phylum characterized by species with distinct wart-like morphology (Garrity and Holt, 2001) and divided in three main classes (Opitutae, Spartobacteria, and *Verrucomicrobiae*; Bergmann et al., 2011). *Verrucomicrobia* possess several unusual features, like the presence of a eukaryotic-like tubulin (Schlieper et al., 2005), but interest in this phylum has grown in recent years due mainly to metagenomics analyses revealing the association of *Verrucomicrobia* with several eukaryotic hosts (Sait et al., 2011), their prominence in many soil communities (Bergmann et al., 2011) and a significant role in the adaptability of the human gut microbiome (Dubourg et al., 2013; Liou et al., 2013). *Verrucomicrobia* are clustered with the *Planctomycetes* and the *Chlamydiae* in the *Planctomycetes-Verrucomicrobia-Chlamydiae* (PVC) super-phylum, a large and diverse phylogenetic clade of clinical and biotechnological interest in which the SOS response has not been documented (Gupta et al., 2012). The genome of the representative *Verrucomicrobia* species, *V. spinosum*, reveals

the presence of orthologs for the three core constituents of the SOS response (*lexA*, *recA*, and the *imuA-imuB-dnaE2* operon), suggesting that the *Verrucomicrobia* have a functional SOS response. Here we combine *in silico* and *in vitro* approaches to characterize the LexA-binding motif of *Verrucomicrobia* and analyze the SOS regulatory network of this bacterial phylum. Our results illustrate the extraordinary plasticity of this transcriptional regulatory network and provide novel insights into the molecular mechanisms driving its evolution.

## MATERIALS AND METHODS

### Functional and Taxonomical Assignments

Orthology assignment of putative function for all reported genes was performed using the HHPRED web service (HHPred, RRID:SCR\_010276) with default options and an *e*-value threshold of  $10^{-30}$  on the TIGRFAM (JCVI TIGRFAMS, RRID:SCR\_005493), PFAM (Pfam, RRID:SCR\_004726), and COG (COG, RRID:SCR\_007139) databases (Tatusov et al., 2000; Söding et al., 2005; Haft et al., 2013; Finn et al., 2016). eggNOG identifiers, categories, and descriptions were retrieved from the eggNOG database (eggNOG, RRID:SCR\_002456) using HMMER (Hmmer, RRID:SCR\_005305; Eddy, 2011; Powell et al., 2014). Taxonomical identification numbers (TaxID) for all species analyzed in the manuscript were retrieved from the NCBI Taxonomy server (NCBI Taxonomy, RRID:SCR\_003256; Federhen, 2012).

### Ortholog Detection and Motif Discovery

Protein sequences for *V. spinosum* DSM 4136 [TaxID: 240016] LexA [WP\_009959117], RecA [WP\_009965965], and ImuA [WP\_009959308] were obtained from the NCBI RefSeq database [RefSeq, RRID:SCR\_003496; Pruitt et al., 2012]. The promoter regions ([-250, 0] with respect to the translational start site) for orthologs of these sequences in *Verrucomicrobia* genome and metagenome assemblies were downloaded from the Integrated Microbial Genomes (IMG) resource [IMG, RRID:SCR\_007733; Markowitz et al., 2014], after identifying orthologs through the IMG BLAST service using a  $10^{-20}$  *e*-value threshold and the *V. spinosum* sequences as queries (Altschul et al., 1997). Downloaded promoter sequences were used as input for the MEME motif discovery tool on the MEME Suite server [MEME Suite—Motif-based sequence analysis tools, RRID:SCR\_001783], requesting palindromic motifs between 8 and 20 bp long and otherwise default options (Bailey et al., 2006). Sequence logos were generated using the Weblogo sequence service (WEBLOGO, RRID:SCR\_010236; Crooks et al., 2004).

### Analysis Of $\alpha 3$ Helix Motifs

A sequence model for the *Verrucomicrobia*  $\alpha 3$  helix of the LexA N-terminal helix-turn-helix motif was obtained through multiple sequence alignment of available LexA protein sequences for this phylum, using the information in P0A7C2 (UniProtKB, RRID:SCR\_004426) on the *E. coli* LexA crystal structure (Zhang et al., 2010) to define penalty masks for CLUSTALW profile alignment mode (Thompson et al., 1994). The *Verrucomicrobia* LexA  $\alpha 3$  helix motif was compared to previously compiled LexA

$\alpha 3$  helix motifs for different phyla (Sanchez-Alberola et al., 2015) using the TomTom service of the MEME suite (Gupta et al., 2007). Amino acid property plots and differential analysis of the *Verrucomicrobia* LexA  $\alpha 3$  helix motif with respect to the *Betaproteobacteria*, *Actinobacteria*, and *Firmicutes*  $\alpha 3$  helix motifs were generated with the iceLogo web service (iceLogo, RRID:SCR\_012137; Colaert et al., 2009).

## LexA-Binding Motif Search and Comparative Genomics Analysis

Experimentally validated LexA-binding motifs were downloaded from the CollecTF database (CollecTF, RRID:SCR\_014405; Kiliç et al., 2014). Whole genome shotgun assemblies for *Verrucomicrobia* species with total sequence lengths larger than or equal to the smallest complete *Verrucomicrobia* genome (*Methylobacterium thermophilum* V4, TaxID: 481448; 2,287,145 bp) were obtained from the NCBI RefSeq database (RefSeq, RRID:SCR\_003496). LexA-binding motif searches on individual genomes were performed using xFITOM (xFITOM, SCR\_014445) with the sequence information content ( $R_i$ ) scoring method and default parameters (Schneider, 1997; Bhargava and Erill, 2010). Comparative genomics analyses were performed with CGB, a collection of Python scripts implementing a computational pipeline for comparative genomics of regulatory transcriptional networks in bacterial genomes. The pipeline is based on previous work (Sanchez-Alberola et al., 2015) and is available under a GPL license on GitHub (<http://www.github.com/erilllab/cgb>). Given a set of genome assemblies, a transcription factor (TF) and its known binding motif, the pipeline first searches for binding motif instances in the promoter region of all genes (−250, +50 of TLS). Genes in directons with intergenic distance below the mean intergenic distance of each genome are considered to form operons and TF-binding sites identified in the lead operon gene are assigned accordingly to all operon members. The presence of genes with high-scoring TF-binding sites within predicted operons is used to revise operon predictions. Orthologs across all analyzed species are detected as best reciprocal BLAST hits using a  $10^{-20}$  *e*-value threshold. The pipeline summarizes analysis results using a heatmap with species clustered using a distance-based TF tree and a color scheme indicating the presence/absence of orthologs and the score of detected TF-binding sites in the corresponding operon.

## Bacterial Strains and Culture Conditions

*E. coli* (DH5 $\alpha$  and BL21; Thermo Fisher Scientific, RRID:SCR\_013270) and *V. spinosum* DSM 4136 strains were grown at 37°C in LB (Green and Sambrook, 2012) and at 30°C in M13 media [DSMZ 607; German Collection of Microorganisms and Cell Cultures, RRID:SCR\_001711], respectively. Antibiotics were added to the cultures at reported concentrations (Green and Sambrook, 2012).

## Oligonucleotides and DNA Techniques

Plasmid isolation, restriction digestion, DNA ligation, transformation, DNA extraction, and PCR were carried out using standard protocols, as described elsewhere (Green and

Sambrook, 2012). Restriction enzymes, T4 DNA ligase, DNA polymerase, and the DIG-DNA labeling and detection kit were from Roche (Roche NimbleGen, RRID:SCR\_008571). The oligonucleotides used for this work are listed in Supplementary Material 1 and were purchased from Invitrogen (Molecular Probes, RRID:SCR\_013318). Mutants of the *V. spinosum* *recA* promoter (VSP\_RS32310) were obtained using oligonucleotides carrying designed substitutions (Supplementary Material 1). The DNA sequence of generated fragments was verified by sequencing (Macrogen, RRID:SCR\_014454).

## Protein Purification and Electrophoresis Mobility Shift Assays

*V. spinosum* DSM 4136 DNA was extracted from phosphate buffered (50 mM) saline (pH 8.0)-washed pellets containing cells using the easy-DNA™ DNA isolation kit (Molecular Probes, RRID:SCR\_013318). The *V. spinosum* DSM 4136 was amplified using suitable primers (Supplementary Material 1) and cloned into a pET15b vector (Millipore, RRID:SCR\_008983). The *O. terrae* PB90-1 *lexA* was obtained by chemical synthesis (GeneArt; Thermo Fisher Scientific, RRID:SCR\_013270) and cloned into a pET15b vector. Overexpression and purification of the *V. spinosum* DSM 4136 and *O. terrae* PB90-1 and *B. subtilis* LexA proteins was performed as described previously for other LexA proteins (Cambray et al., 2011; Cornish et al., 2014). DNA probes for electro-mobility shift assays (EMSA) were generated using two complementary synthetic oligonucleotides centered on the target LexA-binding sites (Supplementary Material 1). The dsDNA synthetic fragments were ligated into pGEMT vector (Roche NimbleGen, RRID:SCR\_008571) and transformed into *E. coli* DH5 $\alpha$  (Thermo Fisher Scientific, RRID:SCR\_013270). In all cases the plasmids were confirmed by sequencing and DNA probes were obtained by PCR using M13 forward and reverse digoxigenin-labeled oligos (Supplementary Material 1). EMSAs were performed as described previously (Sanchez-Alberola et al., 2012), using 20 ng of each digoxigenin-marked DNA probe in the binding mixture and adding the corresponding LexA protein (from 80 to 400 nM). Samples were loaded onto 6% non-denaturing Tris-glycine polyacrylamide gels and digoxigenin-labeled DNA-protein complexes were detected using the manufacturer's protocol (Roche NimbleGen, RRID:SCR\_008571).

## RESULTS

### LexA Targets a Novel LexA-Binding Motif in the *Verrucomicrobia*

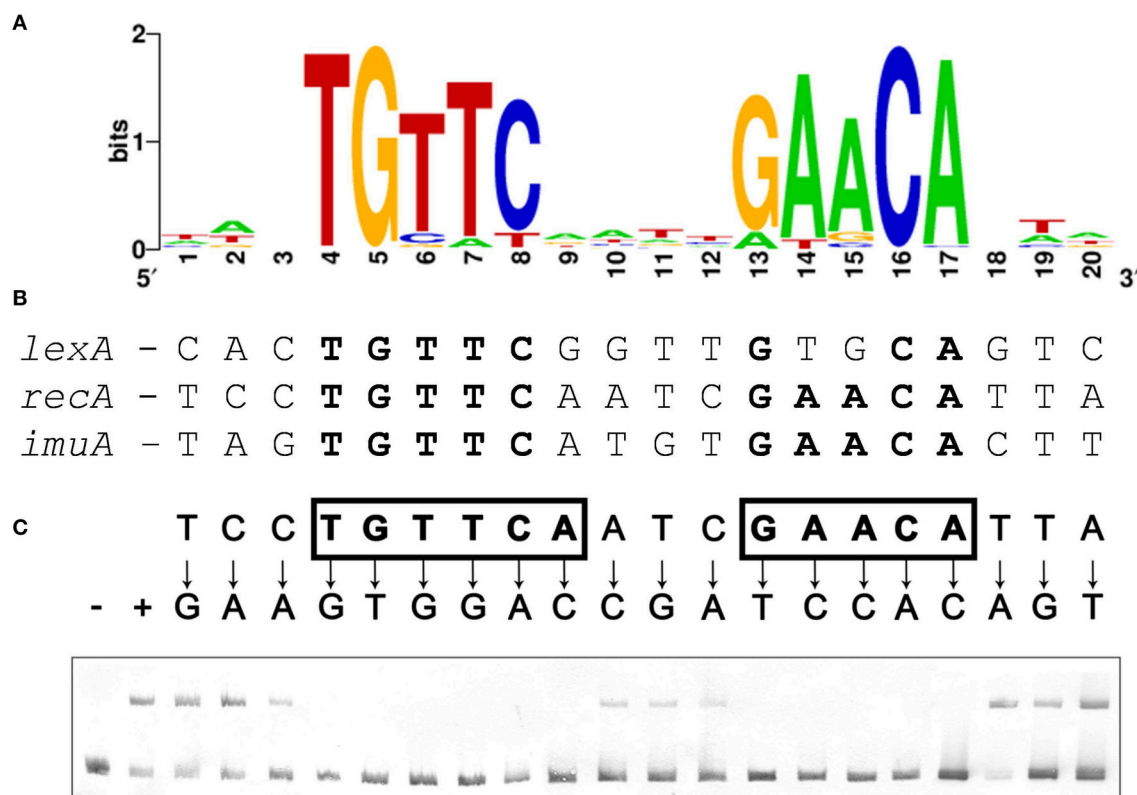
The presence of core SOS response operons (*lexA* [VSP\_RS04780], *recA* [VSP\_RS32310], and *imuA-imuB-dnaE2* [VSP\_RS05590-VSP\_RS05595-VSP\_RS05600]) in the genome of the representative *Verrucomicrobia* species *V. spinosum* DSM 4136 indicates that this phylum might possess a functional LexA regulatory network. However, computational searches using known LexA-binding motifs did not yield putative LexA-binding sites upstream of any SOS related genes in *V. spinosum*. Taking advantage of the availability of multiple



genome and metagenome assemblies for the *Verrucomicrobia* phylum, we compiled 116 promoter sequences from 59 different assemblies corresponding to orthologs of the *V. spinosum* DSM 4136 *lexA*, *recA*, and *imuA* genes through the JGI-IMG service (Supplementary Material 2). We then used MEME to identify overrepresented motifs in these sequences. The most significant motif identified by MEME (**Figure 1A**) is a 14 bp palindromic motif with consensus sequence (TGTTC-N4-GAACA). This motif was identified in the promoter region of 27 *lexA* genes, 25 *recA* genes, and 3 *imuA* genes, corresponding to 36 different genome and metagenome assemblies and spanning all three major groups of *Verrucomicrobia* (Supplementary Material 3). A computational search also identified instances of this motif in the promoter sequences of the *V. spinosum* DSM 4136 *lexA*, *recA*, and *imuA-imuB-dnaE2* operons (**Figure 1B**). The TGTTC-N4-GAACA motif is reminiscent of the LexA-binding motif (GAAC-N4-GTTC) previously reported in the *Firmicutes*, *Actinobacteria*, and *Gallionellales* (Davis et al., 2002; Au et al., 2005; Sanchez-Alberola et al., 2015). Together with its structural similarity to previously reported LexA-binding motifs, the presence of this motif in the promoter region of multiple orthologs for three core components of the SOS response

strongly suggested that the identified TGTTC-N4-GAACA motif is the LexA-binding motif of the *Verrucomicrobia*.

To validate that the palindromic motif identified *in silico* was the LexA-binding motif of *V. spinosum*, we purified the *V. spinosum* DSM 4136 LexA protein [WP\_009959117] and performed electro-mobility shift assays (EMSA) with wild-type and mutant versions of the *V. spinosum* DSM 4136 *recA* promoter containing single-nucleotide substitutions at each position of the predicted LexA-binding motif. The results of this site-directed mutagenesis analysis (**Figure 1C**) are in broad agreement with the motif predicted *in silico*, confirming that *V. spinosum* LexA targets a spaced dyad motif with consensus sequence TGTTC-N4-GAACA. Single-nucleotide mutations to the bases of the inverted repeat regions (TGTTC and GAACA) of the *V. spinosum* LexA-binding motif systematically abolish LexA binding in the *recA* promoter context, indicating that these conserved elements likely correspond to the monomer binding site and are therefore essential for LexA binding activity (Groban et al., 2005). In contrast, the 4 bp spacer region and 3 bp flanking regions tolerate single-nucleotide mutations, suggesting that they are predominantly involved in indirect readout and DNA bending (Zhang et al., 2010).



**FIGURE 1 | (A)** Sequence logo of the palindromic motif identified by MEME in the promoter region of *Verrucomicrobia* *lexA*, *recA*, and *imuA* genes. **(B)** Computationally identified LexA-binding sites matching the *Verrucomicrobia* LexA-binding motif in the promoter region of *V. spinosum* *lexA*, *recA*, and *imuA-imuB-dnaE2* operons. Bases matching the motif consensus are highlighted in bold typeface. **(C)** Electro-mobility shift assays on wild-type and single-nucleotide mutation-containing fragments of the *V. spinosum* *recA* promoter using *V. spinosum* LexA (80 nM). The “+” and “-” symbols denote, respectively, lanes for the negative control (no LexA protein) and the wild-type *recA* promoter fragment. For all other lanes, arrows designate the introduced single-nucleotide mutations. Positions on which single-nucleotide mutations abolish binding are shown in bold typeface and boxed.

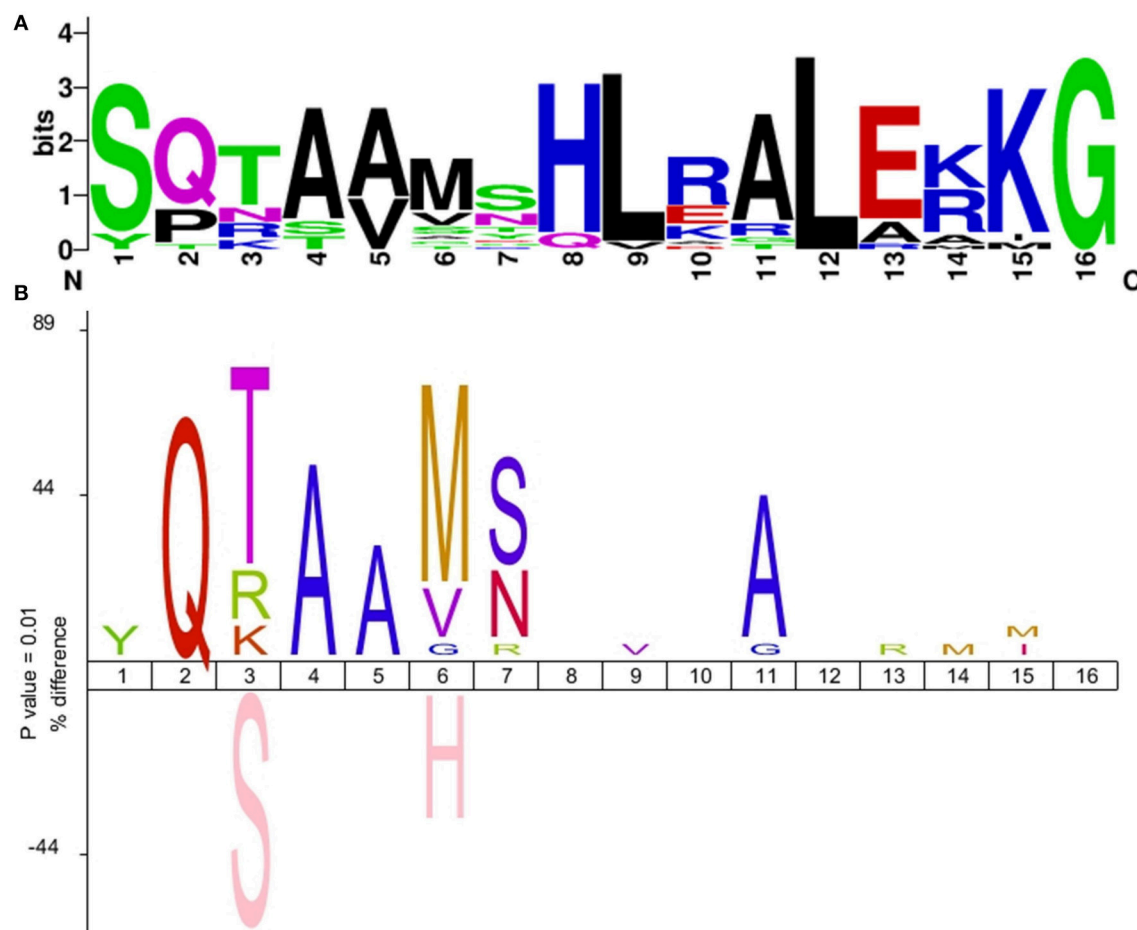


Previous work has established that the  $\alpha 3$  helix of the N-terminal helix-turn-helix motif is responsible for the majority of the specific contacts with monomer binding sites of LexA-binding motifs (Oertel-Buchheit et al., 1990; Otteleben et al., 1991; Thliveris and Mount, 1992; Groban et al., 2005; Zhang et al., 2010). Comparison of the  $\alpha 3$  helix sequence in *Verrucomicrobia* LexA proteins with previously reported LexA  $\alpha 3$  helix motifs (Sanchez-Alberola et al., 2015) shows that the *Verrucomicrobia* LexA  $\alpha 3$  helix is most closely related to those of the *Betaproteobacteria*, *Firmicutes*, and *Actinobacteria*. As shown in **Figure 2**, the majority of the changes observed in the  $\alpha 3$  helix of *Verrucomicrobia* localize to the N-terminal part of the helix, affecting residues that change sequence specificity through direct readout, but that are not essential for DNA bending and structural motif recognition (Oertel-Buchheit et al., 1990; Thliveris et al., 1991; Thliveris and Mount, 1992; Groban et al., 2005; Zhang et al., 2010). Furthermore, the overall distribution of hydrogen donors and hydrophobic residues is preserved across the entire  $\alpha 3$  helix

(Supplementary Material 4). These observations suggest that the structural similarities between *Firmicutes*, *Actinobacteria*, *Betaproteobacteria*, and *Verrucomicrobia* LexA-binding motifs are the result of an evolutionary process in the LexA DNA-binding motif that has modified the specific readout of monomer sites without altering the recognition of the overall motif structure.

### The *Verrucomicrobia* LexA Protein Targets Tandem Binding Sites in LexA Promoters

Close inspection of the *V. spinosum* *lexA* promoter reveals a poorly conserved LexA-binding site immediately downstream (1 bp) of the putative LexA-binding site identified *in silico* (**Figure 3**). To confirm that both these putative motif instances are involved in LexA binding, we performed EMSA with purified *V. spinosum* LexA protein on the *lexA* promoter. The results shown in **Figure 4A** revealed the distinct formation of two retardation bands on the *lexA* promoter at low protein concentrations, corresponding to LexA binding at either one



**FIGURE 2 | (A)** Sequence logo of the *Verrucomicrobia* LexA  $\alpha 3$  helix motif. **(B)** IcelLogo depicting the difference in amino acid frequency at each position of the LexA  $\alpha 3$  helix between the multiple sequence alignment of *Verrucomicrobia* LexA protein sequences and a reference set of *Betaproteobacteria*, *Actinobacteria*, and *Firmicutes* LexA  $\alpha 3$  helix sequences. The upper part of the plot shows residues overrepresented in the *Verrucomicrobia* LexA  $\alpha 3$  helix; the bottom part shows residues overrepresented in the reference set. Only differences with significant z-score under a confidence interval of 0.01 are shown.

V. spinosum DSM 4136	GAC <b>TGCACAACCGAACAG</b> <b>TGTTTCACATCTAACC</b> <u>CATG</u> CCTTACGGA
V. sp. BvORR106	GAC <b>TGCACAATCGAACAG</b> <b>TGTTTCACATCTAACC</b> <u>CATG</u> CCTTACGGA
V. sp. BvORR034	GAC <b>TGCACAATCGAACAG</b> <b>TGTTTCACATCTAACC</b> <u>CATG</u> CCTTACAGA
R. marina DSM 17716	CTA <b>TGTTCTTAAGAACAG</b> <b>TGTTCAAATTATGA</b> GTGATGCATTAA
R. squalenifaciens DSM 18772	TTG <b>TGTTCTTAAGAACAG</b> <b>TGTTCAATTAGGTC</b> <u>ATG</u> GCAAAAGGA
V. SCGC AAA027-I19	TTT <b>TGTTCTGTCTGAACAC</b> <b>TGTTTCATCAGATCA</b> AGCTCAAACCAAC
MEint.metabat.12173	TTT <b>TGTTCTGTCTGAACAC</b> <b>TGTTTCATCAGATCA</b> AGCTCAAACCAAC
MEint.metabat.12612	ATT <b>TGTTCTCAAGAACAC</b> <b>TGTTCAACGAACCA</b> <u>ATG</u> GCACACAACC
TBhypo.metabat.2519	TAG <b>TGTTCAAAAAGAACAC</b> <b>TGTTTCATATGAGTA</b> AGCAATTAACGG
TBepi.metabat.1800	TAG <b>TGTTCAAAAAGAACAC</b> <b>TGTTTCATATGAGTA</b> AGCAATTAACGG
MEint.metabat.3880	CTT <b>TGTTCCCTCGAACAC</b> <b>TGTTCAAAAATCATC</b> CACACTAACCTC
Haloferula BvORR071	GTT <b>TGTTCTTAAGAACAC</b> <b>TGTTTCACTAAGCCA</b> TGCACGAAGGGT
V. EBPR_Bin_208	GAA <b>TGTTCTCTCAAACAG</b> <b>TGTTTCAGAATCAAG</b> TTAACCTCACGC
V. EBPR_Bin_287	TGA <b>TGTTCTGTGGAACAG</b> <b>TGTTTCATAATGAAA</b> TTAACCGCCCGC
V. IMCC26134	ATG <b>TGTTCACTTGAACAC</b> <b>GGTTCACAAGAACA</b> CACCGCC <u>ATG</u> CA
Opitutae-129	AGG <b>TGTTTCATTTGAACC</b> <b>TGTTCAATTGAACC</b> CT <u>ATG</u> TTAACCG
O. terrae PB90-1	GGC <b>TGTTTCGCTTGAACA</b> TG <b>TGTTCAAGAGAACA</b> CATCCGCC <u>ATG</u>

**FIGURE 3 | Multiple sequence alignment of *Verrucomicrobia* *lexA* promoter sequences displaying the tandem LexA-binding site arrangement observed in *V. spinosum*.** The primary LexA-binding site is highlighted in yellow, and the secondary one in green. Bases matching the LexA-binding motif consensus are in bold typeface. Predicted translation start sites are shown in red and underlined.

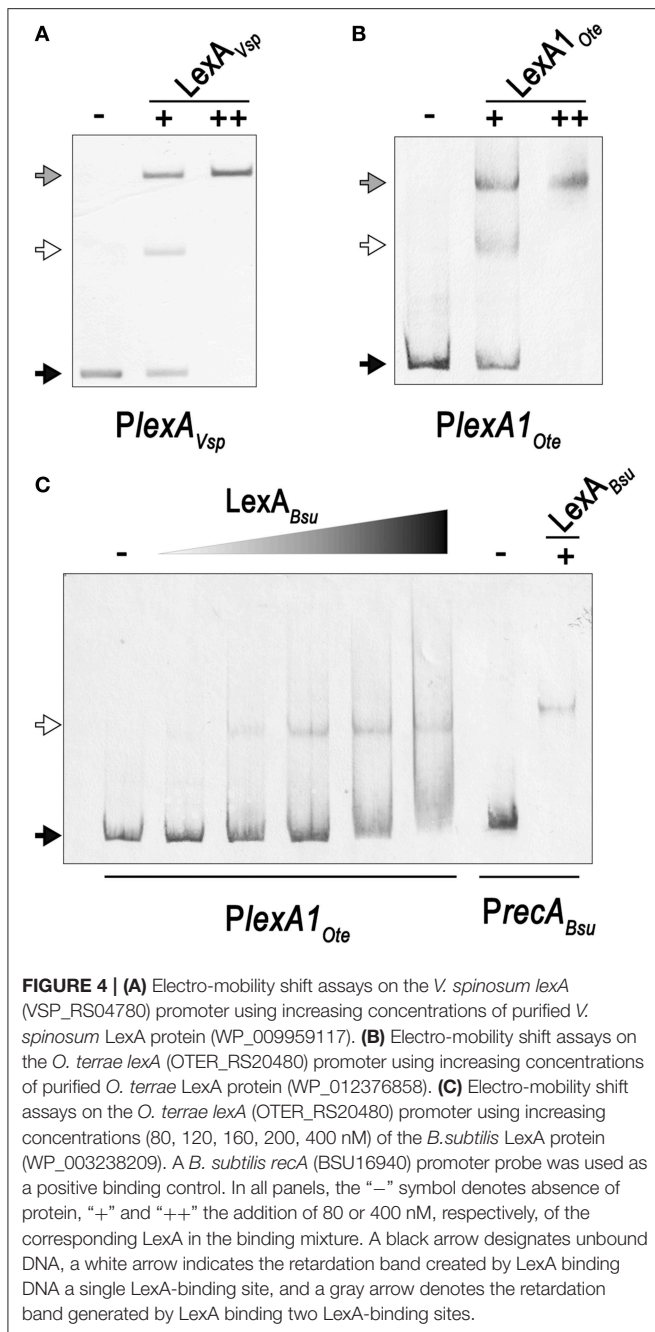
or the two LexA-binding sites identified in the *lexA* promoter. Further, increasing protein concentration resulted in a single retardation band corresponding to LexA recognizing both LexA-binding sites. Taken together, these results indicate that the two identified LexA-binding sites in the promoter region of the *V. spinosum* *lexA* gene are bound cooperatively by LexA. A systematic analysis of the promoter regions of *V. spinosum* *lexA* gene orthologs in the *Verrucomicrobia* revealed that more than half of the *lexA* ortholog promoters with predicted LexA-binding sites display similar tandem site configurations (Supplementary Material 5). Most of these tandem arrangements involve a conserved TGTTTC-N4-GAACA motif instance followed by a degenerate site in which only the first TGTTTC element is conserved, but a tandem site arrangement with both conserved sites can be observed in at least two species (Figure 3).

In *Opitutae terrae* PB90-1, there are two fully conserved *Verrucomicrobia* LexA-binding motifs in the promoter region of a putative *lexA-imuA-imuB-dnaE2* operon [OTER\_RS20480-OTER\_RS20475-OTER\_RS20470-OTER\_RS20465] separated by 2 bp. This arrangement generates an instance of the canonical GAAC-N4-GTTC LexA-binding motif of *Firmicutes* and *Actinobacteria*. Using purified *O. terrae* and *B. subtilis* LexA proteins, we performed EMSA to validate the functionality of this tandem arrangement in *O. terrae* (Figure 4). EMSA with *O. terrae* LexA [WP\_012376858] reveals two retardation bands at low protein concentration, confirming that this protein also binds to both *Verrucomicrobia* LexA target sites in the *lexA-imuA-imuB-dnaE2* promoter (Figure 4B). Mobility assays with incremental concentrations of *B. subtilis* LexA [WP\_003238209] show that *B. subtilis* LexA binds a unique element in the *O.*

*terrae* *lexA-imuA-imuB-dnaE2* promoter, yielding a single retardation band similar to the one observed on the *B. subtilis* *recA* [BSU16940] promoter (Figure 4C). These results suggest that *B. subtilis* LexA binds the *Firmicutes*-like LexA-binding motif instance generated by the tandem arrangement of *Verrucomicrobia* LexA-binding sites.

## The Core *Verrucomicrobia* LexA Regulon Comprises Three Operons Involved in DNA Repair and Mutagenesis

Having established the LexA-binding motif of the *Verrucomicrobia*, we performed a comparative genomics analysis of the LexA regulon in this phylum. We compiled 15 whole-genome shotgun assemblies for members of all the major classes of *Verrucomicrobia* (*Opitutae*, *Spartobacteria*, and *Verrucomicrobiae*) presenting a *V. spinosum* LexA homolog and searched for putative LexA-binding sites in the promoter region [−250, +50] of predicted operons. The results of the comparative genomics analysis (Figure 5; Supplementary Materials 6, 7) reveal a core LexA regulon present in all classes of the *Verrucomicrobia* phylum and composed of three operons: *lexA*, *splB*, and *imuA-imuB-dnaE2*. The *lexA* gene displays high-scoring sites in all the analyzed species, except for *Verrucomicrobium* sp. 3C (TaxID: 1134055), *Verrucomicrobia bacterium* LP2A (TaxID: 478741), and *Pedospaera parvula* Ellin514 (TaxID: 320771). The LexA proteins of these species display significant changes to the α3 helix of the LexA DNA-binding domain, suggesting that they may target a divergent LexA-binding motif. Consistent with this



result, the genomes of these organisms do not reveal any instance of the *Verrucomicrobia* LexA-binding motif in the promoter regions of previously documented SOS genes (Erill et al., 2007). The promoter region of the *splB* gene shows evidence of LexA regulation in several *Verrucomicrobiae*, one *Opitutaceae* (*O. terrae*) and the only available assembly of a *Spartobacteria* species (*Chthoniobacter flavus* Ellin428; TaxID: 497964). The product of the *splB* gene contains a radical SAM domain (PFAM04055) and has homology to COG1533 (ENOG4105DCH), classified as a DNA repair photolyase. Members of this orthologous group have been reported to be regulated by LexA in the *Actinobacteria*,

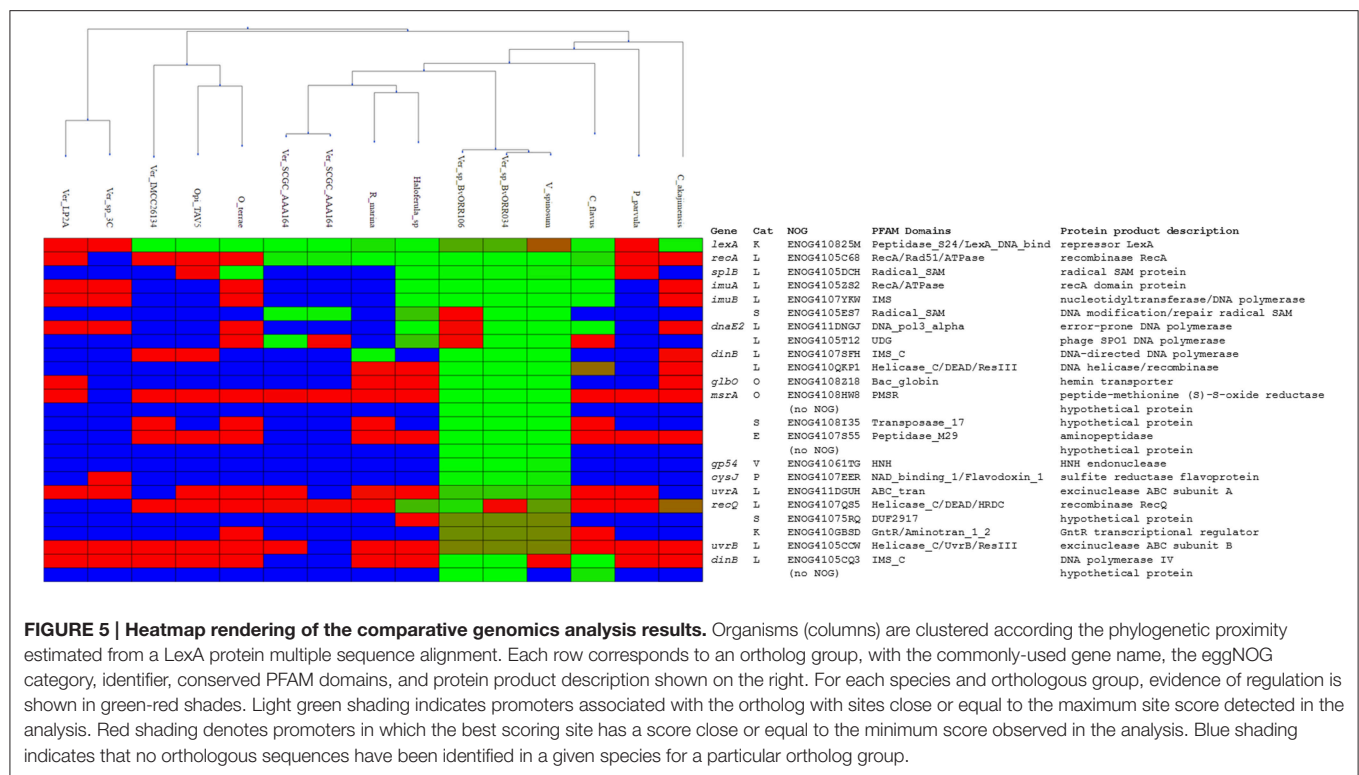
the *Gammaproteobacteria*, the *Betaproteobacteria*, and the *Alphaproteobacteria* (Davis et al., 2002; Cirz et al., 2006; Sanchez-Alberola et al., 2012, 2015; Ulrich et al., 2013), suggesting that it may be a previously unrecognized core component of the SOS response. Lastly, the promoter region of the *imuA-imuB-dnaE2* operon presents *Verrucomicrobia* LexA-binding motif instances in *C. flavus* and the same *Verrucomicrobiae* species as *splB*. As noted above, *O. terrae* presents a putative *lexA-imuA-imuB-dnaE2* with verified *O. terrae* LexA binding in its promoter region (Figure 4). Even though the intergenic distance between *lexA* and *imuA* is larger than the genomic average for this species (264 bp), the prevalence of *lexA-imuA-imuB-dnaE2* arrangements across the *Bacteria* domain suggests that this direction constitutes a functional operon in *O. terrae* (Erill et al., 2006).

### The *Verrucomicrobia* LexA Regulon is Highly Variable and Incorporates Novel Functions

The results of the comparative genomics analysis reveal remarkable variation in the size and composition of the inferred LexA regulon. In the *Verrucomicrobiae*, the predicted regulon ranges from one operon (*Verrucomicrobia bacterium* IMCC26134; TaxID: 1637999) to over 14 (*V. spinosum*), with several species displaying intermediate sizes [2 operons in *Rubritalea marina* DSM 17716 (TaxID: 1123070) or 5 in *Haloferula* sp. BvORR071 (TaxID: 1396141)]. The only available *Spartobacteria* representative (*C. flavus*) shows a moderate regulon size (5 operons). In contrast, the LexA regulon appears to have shrunk noticeably in the *Opitutaceae*, where it encompasses at the most two operons. Two members of this family [*O. terrae* and *Opitutaceae bacterium* TAV5 (TaxID: 794903)] present a duplication of the *lexA* gene. The products of the two *Opitutaceae* TAV5 *lexA* genes (OPIT5\_RS22040 and OPIT5\_RS25725) present 91% identity and their promoter regions contain almost identical LexA-binding sites. The *lexA* genes in *O. terrae* (OTER\_RS20480 and OTER\_RS11645) have diverged substantially (42% protein sequence identity) and only the promoter of the *lexA1* gene (OTER\_RS20480) presents *Verrucomicrobia* LexA-binding motif instances, following the tandem arrangement discussed above (Figure 3).

The overall composition of the inferred *Verrucomicrobia* LexA regulon is in broad agreement with experimental and computational descriptions of the SOS regulatory network in other phyla (Erill et al., 2007; Sanchez-Alberola et al., 2012, 2015). Beyond the core regulon described above (*lexA*, *splB*, *imuA-imuB-dnaE2*), the *Verrucomicrobia* LexA regulon encompasses genes coding for the recombination protein RecA (COG0468; ENOG4105C68), the excinuclease ABC subunits A (COG0178; ENOG411DGUH) and B (COG0556; ENOG4105CCW), two DNA helicase RecQ homologs (COG0514; ENOG4107QS5 and ENOG410QKP1) and two homologs of the error-prone DNA polymerase IV (COG0389; ENOG4105CCW and ENOG4105CQ3). In addition to these previously established SOS genes, the *Verrucomicrobia* LexA regulon shows evidence of regulation for an operon encoding proteins matching the





**FIGURE 5 | Heatmap rendering of the comparative genomics analysis results.** Organisms (columns) are clustered according to the phylogenetic proximity estimated from a LexA protein multiple sequence alignment. Each row corresponds to an ortholog group, with the commonly-used gene name, the eggNOG category, identifier, conserved PFAM domains, and protein product description shown on the right. For each species and orthologous group, evidence of regulation is shown in green-red shades. Light green shading indicates promoters associated with the ortholog with sites close or equal to the maximum site score detected in the analysis. Red shading denotes promoters in which the best scoring site has a score close or equal to the minimum score observed in the analysis. Blue shading indicates that no orthologous sequences have been identified in a given species for a particular ortholog group.

TIGR03916 (ENOG4105E87) and TIGR03915 (ENOG4105T12) models. These models are present in about 20% of sequenced bacterial genomes, arranged always in operon configuration, and are thought to constitute a DNA base excision repair system involving a uracil-DNA glycosylase (UDG) domain that is conserved in all *Verrucomicrobia* TIGR03915-matching homologs.

To validate the predictions of the comparative genomics approach and further establish the LexA regulon of the *Verrucomicrobia*, we performed EMSA with purified *V. spinosum* and *O. terrae* proteins on the promoter region of several genes with predicted LexA-binding sites in these organisms and evidence of regulation in at least three different genomes. The results, shown in **Figure 6**, confirm that LexA binds to the promoter region of the *splB* gene in *O. terrae* (OTER\_RS07185) and *V. spinosum* (VSP\_RS12190). *V. spinosum* LexA also binds the *imuA-imuB-dnaE2* operon promoter, the promoters of genes coding for DNA polymerase IV (VSP\_RS08510) and RecQ (VSP\_RS32195) homologs, and the *recA* (VSP\_RS04780) and *uvrA* (VSP\_RS32650) promoters. Together with the comparative genomics analysis, these results confirm the existence of a conserved core LexA-regulon in the *Verrucomicrobia* and demonstrate that, in some *Verrucomicrobia* species, LexA controls a network of similar size and function to those reported in well-studied bacterial phyla, using a novel LexA-binding motif.

## DISCUSSION

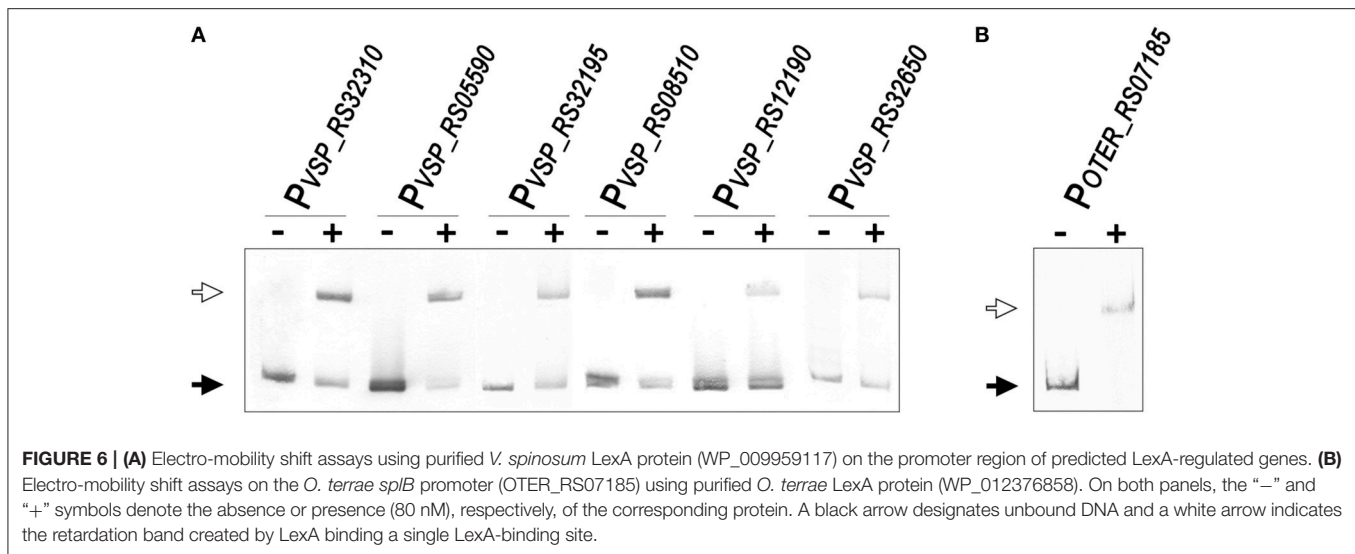
This work reports the combined use of *in silico* and *in vitro* techniques to characterize a novel binding motif for the SOS

transcriptional repressor LexA in the *Verrucomicrobia*, and its use to define the LexA regulon in this bacterial phylum of emerging interest. The results provide further context to illustrate the complex evolutionary history of the SOS response, and put to the fore the plasticity and versatility of this transcriptional system.

## Variability and Core Elements of the SOS Regulatory Network

Phylogenetic and protein signature analyses have firmly established the *Verrucomicrobia* as one of the major phyla in the PVC super-phylum, an ancient bacterial group estimated to have diverged from other bacterial clades almost two billion years ago (Gupta et al., 2012; Kamneva et al., 2012; Lagkouvardos et al., 2014). The analysis of the LexA regulon performed in this work hence provides for the first time insights into the organization of a complex transcriptional system in this large bacterial clade. Our results show evidence of a functional LexA protein targeting the same LexA-binding motif in all three major *Verrucomicrobia* classes (*Opitutae*, *Spartobacteria*, and *Verrucomicrobiae*), suggesting that a functional LexA regulatory network was present in the ancestor of the *Verrucomicrobia* (**Figure 1**). However, the *Verrucomicrobia* also display substantial heterogeneity in the size of their predicted LexA regulons (**Figure 5**). Some families, such as the *Methylococcoides* do not present LexA homologs, while many members of the *Verrucomicrobiales* display small (1–3 operon) regulons, a setup that appears to be the rule in the *Opitutales*. Small SOS regulatory networks have been experimentally reported for several species, but mostly in association with drastic changes in the LexA-binding motif





(Jara et al., 2003; Campoy et al., 2005; Mazon et al., 2006). In these instances, the LexA regulon is typically constrained to the regulation of translesion synthesis polymerases (Erill et al., 2007). Conversely, moderately large (10–40 genes) LexA regulons incorporating several DNA repair pathways have been documented in the *Gammaproteobacteria*, the *Betaproteobacteria*, the *Actinobacteria*, and the *Firmicutes* (Fernandez De Henestrosa et al., 2000; Davis et al., 2002; Au et al., 2005; Ulrich et al., 2013; Sanchez-Alberola et al., 2015). These findings have substantiated the notion that translesion synthesis is the primordial function of the SOS response, and the identification of a translesion synthesis operon (*imuA-imuB-dnaE2*) in the core LexA regulon of the *Verrucomicrobia* confirms the ancestral role of this mechanism in the SOS response. Nonetheless, the presence of a putative photolyase (*splB*) in the core *Verrucomicrobia* LexA regulon, with documented LexA-regulated orthologs in several bacterial clades (Davis et al., 2002; Cirz et al., 2006; Sanchez-Alberola et al., 2012, 2015; Ulrich et al., 2013), suggests that photoreactivation might have played an essential DNA repair role in the primordial SOS response.

Beyond the presence of a putative photolyase, the SOS response of the *Verrucomicrobia* presents several interesting differences with the canonical SOS response of *E. coli* and *B. subtilis*. The *Opitutae*, for instance, show a consistent absence of LexA regulation for the *recA* gene. The lack of *recA* regulation by LexA has been reported in several bacterial groups, such as the *Acidobacteria* and the *Deltaproteobacteria* (Jara et al., 2003; Campoy et al., 2005; Mazon et al., 2006). Loss of *recA* regulation is often associated with small LexA regulons and *lexA* gene duplication, which are both features of the *Opitutae* LexA network inferred in this work. Another distinct feature of the *Verrucomicrobia* LexA regulon is the regulation of multiple RecQ homologs (ENOG4107QS5 and ENOG410QKP1; Figure 5). One of these RecQ homologs (ENOG4107QS5) shares functional domains with *B. subtilis* RecS and RecQ proteins and therefore likely fulfills similar repair functions. The other RecQ homolog

(ENOG410QKP1) lacks DNA-binding HRDC (Helicase and RNase D C-terminal) and RecQ-C-terminal (RQC) domains and presents weaker evidence of homology with *B. subtilis* RecS and RecQ proteins (Fernández et al., 1998). RecQ helicases are involved in the initiation and reversal of recombination and are known to act in concert with the product of SOS genes (*recA*, *ssb*), and to facilitate the onset of the SOS response (Heyer, 2004; Nakayama, 2005). Although SOS regulation of RecQ homologs has not been documented to date, LexA regulation of other DNA helicases (UvrD, PcrA, DinG) is a well-established feature of the SOS response in several organisms (Fernandez De Henestrosa et al., 2000; Au et al., 2005; Abella et al., 2007). These helicases do not appear to be regulated in the *Verrucomicrobia*, suggesting that the putative LexA regulation of RecQ homologs might be fulfilling a complementary role in this phylum. Our analysis also provides evidence of LexA regulation for an operon encoding radical SAM and uracil-DNA glycosylase domain-containing proteins (ENOG4105ES7 and ENOG4105T12; Figure 5), presumed to function as a DNA base excision repair system. SOS-regulated error-prone polymerases have been shown to have poor sugar discrimination, leading to the frequent misincorporation of ribonucleotides (Schroeder et al., 2015). Misincorporated ribonucleotides are usually removed by RNase HII-mediated ribonucleotide excision repair, but SOS-regulated nucleotide excision repair has also been shown to address ribonucleotide incorporation (Vaisman et al., 2013). The presence of putative LexA-regulated translesion synthesis polymerases in the *Verrucomicrobia* (Figures 5, 6) hence suggests that the regulation of this base excision repair operon by LexA may play a role in addressing uracil misincorporation resulting from SOS induction in this phylum.

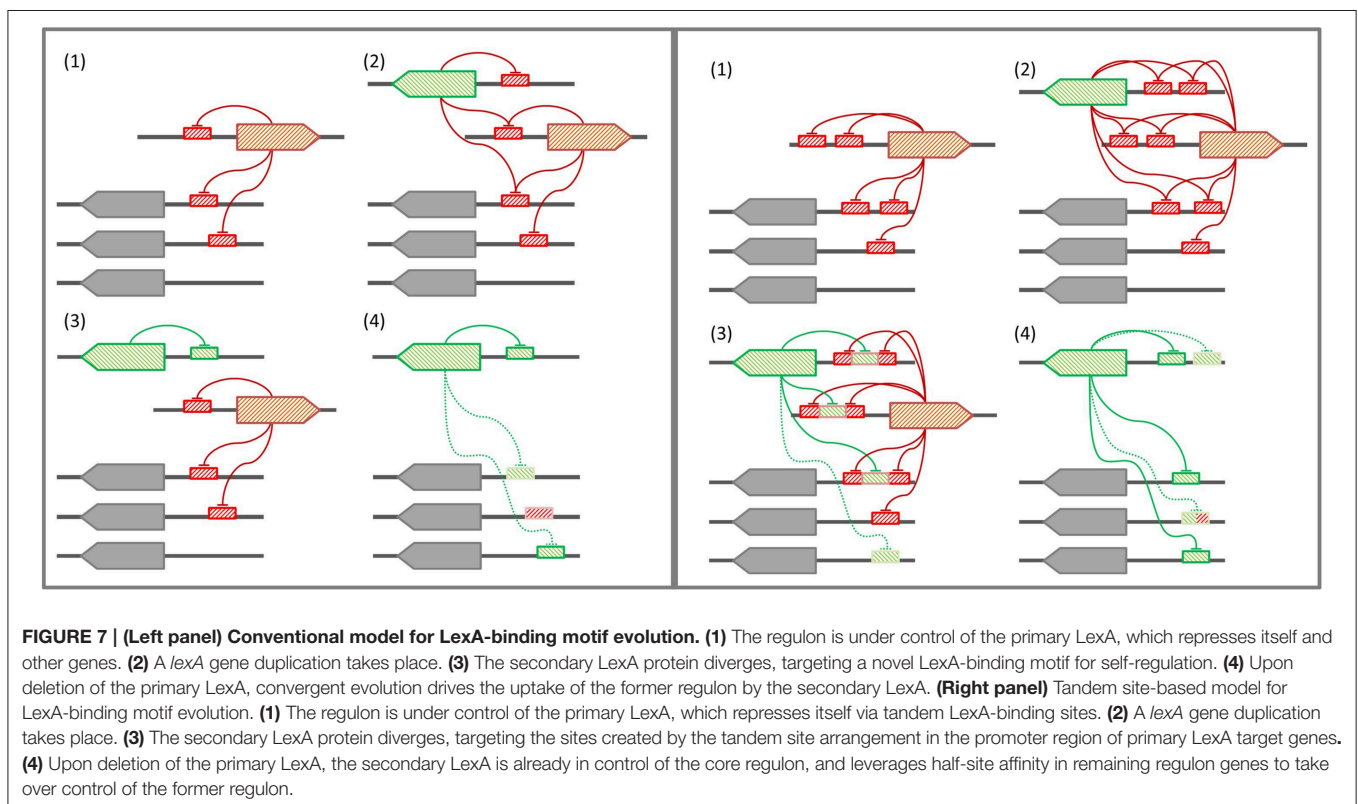
## A Tandem Model for the Evolution of the LexA-Binding Motif

In those bacterial phyla where the SOS response has been experimentally documented, the LexA-binding motif shows

evidence of high conservation, punctuated by periods of rapid divergence and further stabilization (Erill et al., 2007). In the *Firmicutes* and the *Actinobacteria*, LexA targets a conserved GAAC-N4-GTTC LexA-binding motif that is monophyletic for both clades (Cornish et al., 2012), and variations of this motif are also seen in other bacterial groups, such as the *Cyanobacteria* or the *Chloroflexi* (Fernandez de Henestrosa et al., 2002; Mazon et al., 2004b). In the *Proteobacteria*, however, LexA shows an extraordinary diversity of binding motifs (Erill et al., 2007). In *Proteobacteria* classes with abundant sequence information (*Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria*), the LexA-binding motif has been found to be extremely well-conserved, but exceptions to the canonical LexA-binding motif of *Gammaproteobacteria* and *Betaproteobacteria* have been reported in several subgroups (Campoy et al., 2002; Abella et al., 2007; Sanchez-Alberola et al., 2015). These exceptions are associated with duplications of the *lexA* gene, suggesting a model for LexA-binding motif evolution (Figure 7) in which *lexA* duplication leads to progressive divergence in the LexA-binding motif of the duplicated *lexA*, until the primary *lexA* gene is deleted and the divergent LexA takes control of the regulon (Abella et al., 2007; Yang et al., 2008; Sanchez-Alberola et al., 2015). While this model provides a causal mechanism for LexA-motif divergence, it does not address how a divergent LexA can swiftly take control over a regulatory network defined, up to the deletion event, by LexA-binding sites matching the primary LexA-binding motif. Furthermore, the model does not provide a mechanistic explanation for the recurrence of very

similar LexA-binding motifs in distantly related bacterial clades, such as the *Firmicutes* and the *Gallionellales*, recognized through seemingly unrelated LexA  $\alpha 3$  helices (Sanchez-Alberola et al., 2015).

Many bacterial transcription factors bind cooperatively to tandem sites (Barnard et al., 2004). The existence of tandem sites for LexA was first reported in the promoter region of the *E. coli* *lexA* gene (Brent, 1982) and then shown to be a common feature of *lexA* genes in the *Gammaproteobacteria* (Garriga et al., 1992), the *Betaproteobacteria*, and *Alphaproteobacteria* (Sanchez-Alberola et al., 2012), and the *Firmicutes* and *Actinobacteria* (Cornish et al., 2012). These arrangements feature highly conserved and spatially close tandem sites (1–10 bp apart). Tandem LexA-binding sites have also been experimentally reported for other SOS genes, such as the *ydjM* gene of *E. coli* (Fernandez De Henestrosa et al., 2000) or the *umuDC*-like operon (*yqjW-yqzH*) of *B. subtilis* (Au et al., 2005). Furthermore, the use of cooperative LexA-binding to enhance repression has been experimentally demonstrated for several colicin genes, which display a tandem arrangement with a strong and a weak LexA-binding site overlapping at their terminal positions (Gillor et al., 2008). In the *Verrucomicrobia*, there is evidence of a recent *lexA* duplication in the *Opitutaceae* and tandem LexA-binding sites separated by short distances appear to be a conserved feature of the *lexA* promoter (Figure 3). The ability of the *Verrucomicrobia* LexA to cooperatively bind degenerate sites and the fact that at least in one of these species the tandem arrangement generates a functional *B. subtilis* LexA-binding



motif (**Figure 4**) suggest that tandem site arrangements can yield a simple mechanistic process for the evolution of LexA-binding motifs.

In the tandem site model (**Figure 7**), LexA binds consecutive sites in its own promoter and in the promoter of key SOS genes that need to be tightly regulated (Gillor et al., 2008). Upon *lexA* duplication, the site generated by the tandem arrangement provides the secondary LexA with a conserved target for motif divergence. This allows the secondary LexA to maintain cross-regulation with the primary LexA and a subset of its regulon, while incorporating novel elements to its network. After deletion of the primary *lexA* gene, the secondary LexA is hence already in control of a core LexA regulon, and can rapidly evolve sites on other target genes by exploiting its partial overlap, and presumable weak binding affinity, with primary *lexA* sites. The tandem site model therefore provides a conservative mechanism for the evolution of LexA-binding motifs that is capable of addressing outstanding questions regarding the complex evolutionary history of the SOS response. On the one hand, the conservative nature of the model provides a natural explanation for the persistence of conserved SOS response networks under divergent LexA-binding motifs, without the need for a strong selective process driving convergent evolution of similar networks (Sanchez-Alberola et al., 2015). On the other hand, the implicit reuse of LexA monomer-binding sites in the tandem model helps explain the observation of many LexA-binding motifs involving the rearrangement of similar monomer binding sites on different motif structures (Mazon et al., 2004a; Sanchez-Alberola et al., 2015).

Several lines of evidence provide indirect support for a tandem site-based model of LexA-binding motif evolution. As mentioned above, the prevalence of such arrangements in the promoter of *lexA* and other SOS genes has been documented in several bacterial clades. Furthermore, *lexA* duplications targeting identical and divergent motifs have also been experimentally reported, and cross-regulation between duplicated *lexA* genes has been demonstrated in these systems (Jara et al., 2003; Abella et al., 2007; Yang et al., 2008). Lastly, previous work has shown that LexA can bind to degenerate sites that partially match other LexA-binding motifs, indicating that transitional stages of LexA divergence in which the secondary LexA could partially bind the original and tandem-generated motifs are possible (Mazon et al., 2004a). Due to its broad distribution in several phyla, the *Firmicutes* and *Actinobacteria* LexA-binding motif has long been assumed to represent the ancestral motif of LexA. The mirror image relationship between *Firmicutes* and *Verrucomicrobia* LexA-binding motifs, and the generation of functional *B. subtilis* LexA-binding sites from tandem *Verrucomicrobia* LexA-binding sites, hence suggest that the *Verrucomicrobia* LexA-binding motif might have originated after the duplication of a *lexA* gene targeting a tandem arrangement of *Firmicutes*-like LexA-binding sites in a common ancestor of these lineages. The analysis of the  $\alpha 3$  helix of the *Verrucomicrobia* LexA DNA-binding domain (**Figure 2**) supports this hypothesis, revealing overall

conservation of amino acid properties and a substitution pattern consistent with changes in the specific readout of monomer sites, but not in overall motif recognition, as expected in the tandem site evolution model.

## CONCLUSIONS

By combining *in silico* and *in vitro* methods, in this work we have characterized a novel LexA-binding motif in the *Verrucomicrobia*. Using this motif, which presents structural similarities with LexA-binding motifs previously described in other phyla, we performed a comparative genomics analysis of the LexA regulon in this understudied phylum. Our computational analysis, validated through *in vitro* assays, revealed significant variability in the size and composition of the LexA regulatory network of this phylum, and identified novel core and ancillary components of the SOS response. The characterization of the *Verrucomicrobia* LexA-binding motif and regulon also allowed us to postulate for the first time a model for LexA-binding motif evolution that satisfactorily addresses open questions in the evolution of this system via gene duplication events. Future biochemical and genetic experiments, such as determining the conformation of LexA in solution and analyzing expression patterns in mutants for core SOS genes, should provide a more comprehensive characterization of the *Verrucomicrobia* SOS response and its evolution.

## AUTHOR CONTRIBUTIONS

IE and SK performed the *in silico* analyses. SK developed scripts for comparative genomics. SC performed the *in vitro* analyses. IE, SC, and JB discussed the findings and interpreted the results. IE and JB conceived the experiment and coordinated the research. IE drafted the manuscript.

## FUNDING

This work was supported by Spanish Ministry of Science and Innovation (BFU2011-23478) and Generalitat de Catalunya (2014SGR572) awards to JB and by a U.S. National Science Foundation (MCB-1158056) award to IE.

## ACKNOWLEDGMENTS

The authors wish to thank Joan Ruiz and Susana Escribano for their technical support during some of the experimental procedures.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmolb.2016.00033>



## REFERENCES

- Abella, M., Campoy, S., Erill, I., Rojo, F., and Barbe, J. (2007). Cohabitation of two different *lexA* regulons in *Pseudomonas putida*. *J. Bacteriol.* 189, 8855–8862. doi: 10.1128/JB.01213-07
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Au, N., Kuester-Schoeck, E., Mandava, V., Bothwell, L. E., Canny, S. P., Chachu, K., et al. (2005). Genetic composition of the *Bacillus subtilis* SOS system. *J. Bacteriol.* 187, 7655–7666. doi: 10.1128/JB.187.22.7655-7666.2005
- Bailey, T. L., Williams, N., Mistle, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34, W369–W373. doi: 10.1093/nar/gkl198
- Barnard, A., Wolfe, A., and Busby, S. (2004). Regulation at complex bacterial promoters: how *bacteria* use different promoter organizations to produce different regulatory outcomes. *Curr. Opin. Microbiol.* 7, 102–108. doi: 10.1016/j.mib.2004.02.011
- Beaber, J. W., Hochhut, B., and Waldor, M. K. (2004). SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* 427, 72–74. doi: 10.1038/nature02241
- Bergmann, G. T., Bates, S. T., Eilers, K. G., Lauber, C. L., Caporaso, J. G., Walters, W. A., et al. (2011). The under-recognized dominance of *Verrucomicrobia* in soil bacterial communities. *Soil Biol. Biochem.* 43, 1450–1455. doi: 10.1016/j.soilbio.2011.03.012
- Bhargava, N., and Erill, I. (2010). xFITOM: a generic GUI tool to search for transcription factor binding sites. *Bioinformatics* 5, 49–50. doi: 10.6026/97320630005049
- Brent, R. (1982). Regulation and autoregulation by LexA protein. *Biochimie* 64, 565–569. doi: 10.1016/S0300-9084(82)80088-6
- Cambray, G., Sanchez-Alberola, N., Campoy, S., Guerin, E., Da Re, S., Gonzalez-Zorn, B., et al. (2011). Prevalence of SOS-mediated control of integron integrase expression as an adaptive trait of chromosomal and mobile integrons. *Mob. DNA* 2:6. doi: 10.1186/1759-8753-2-6
- Campoy, S., Mazon, G., Fernandez de Henestrosa, A. R., Llagostera, M., Monteiro, P. B., and Barbe, J. (2002). A new regulatory DNA motif of the gamma subclass *Proteobacteria*: identification of the LexA protein binding site of the plant pathogen *Xylella fastidiosa*. *Microbiology* 148, 3583–3597. doi: 10.1099/00221287-148-11-3583
- Campoy, S., Salvador, N., Cortes, P., Erill, I., and Barbe, J. (2005). Expression of canonical SOS genes is not under LexA repression in *Bdellovibrio bacteriovorus*. *J. Bacteriol.* 187, 5367–5375. doi: 10.1128/JB.187.15.5367-5375.2005
- Cirz, R. T., O'Neill, B. M., Hammond, J. A., Head, S. R., and Romesberg, F. E. (2006). Defining the *Pseudomonas aeruginosa* SOS response and its role in the global response to the antibiotic ciprofloxacin. *J. Bacteriol.* 188, 7101–7110. doi: 10.1128/JB.00807-06
- Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009). Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods* 6, 786–787. doi: 10.1038/nmeth1109-786
- Cornish, J. P., Matthews, F., Thomas, J. R., and Erill, I. (2012). Inference of self-regulated transcriptional networks by comparative genomics. *Evol. Bioinform. Online* 8, 449–461. doi: 10.4137/EBO.S9205
- Cornish, J. P., Sanchez-Alberola, N., O'Neill, P. K., O'Keefe, R., Gheba, J., and Erill, I. (2014). Characterization of the SOS meta-regulon in the human gut microbiome. *Bioinformatics* 30, 1193–1197. doi: 10.1093/bioinformatics/btt753
- Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- Davis, E. O., Dullaghan, E. M., and Rand, L. (2002). Definition of the mycobacterial SOS box and use to identify LexA-regulated genes in *Mycobacterium tuberculosis*. *J. Bacteriol.* 184, 3287–3295. doi: 10.1128/JB.184.12.3287-3295.2002
- Dubourg, G., Lagier, J.-C., Armougom, F., Robert, C., Audoly, G., Papazian, L., et al. (2013). High-level colonisation of the human gut by *Verrucomicrobia* following broad-spectrum antibiotic treatment. *Int. J. Antimicrob. Agents* 41, 149–155. doi: 10.1016/j.ijantimicag.2012.10.012
- Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195
- Erill, I., Campoy, S., and Barbe, J. (2007). Aeons of distress: an evolutionary perspective on the bacterial SOS response. *FEMS Microbiol. Rev.* 31, 637–656. doi: 10.1111/j.1574-6976.2007.00082.x
- Erill, I., Campoy, S., Mazon, G., and Barbe, J. (2006). Dispersal and regulation of an adaptive mutagenesis cassette in the bacteria domain. *Nucleic Acids Res.* 34, 66–77. doi: 10.1093/nar/gkj412
- Erill, I., Escibano, M., Campoy, S., and Barbe, J. (2003). *In silico* analysis reveals substantial variability in the gene contents of the gamma *proteobacteria* LexA-regulon. *Bioinformatics* 19, 2225–2236. doi: 10.1093/bioinformatics/btg303
- Erill, I., Jara, M., Salvador, N., Escibano, M., Campoy, S., and Barbe, J. (2004). Differences in LexA regulon structure among *Proteobacteria* through *in vivo* assisted comparative genomics. *Nucleic Acids Res.* 32, 6617–6626. doi: 10.1093/nar/gkh996
- Federhen, S. (2012). The NCBI Taxonomy database. *Nucleic Acids Res.* 40, D136–D143. doi: 10.1093/nar/gkr1178
- Fernandez de Henestrosa, A. R., Cune, J., Erill, I., Magnuson, J. K., and Barbe, J. (2002). A green nonsulfur bacterium, *Dehalococcoides ethenogenes*, with the LexA binding sequence found in gram-positive organisms. *J. Bacteriol.* 184, 6073–6080. doi: 10.1128/JB.184.21.6073-6080.2002
- Fernandez De Henestrosa, A. R., Ogi, T., Aoyagi, S., Chafin, D., Hayes, J. J., Ohmori, H., et al. (2000). Identification of additional genes belonging to the LexA regulon in *Escherichia coli*. *Mol. Microbiol.* 35, 1560–1572. doi: 10.1046/j.1365-2958.2000.01826.x
- Fernandez de Henestrosa, A. R., Rivera, E., Tapias, A., and Barbe, J. (1998). Identification of the *Rhodobacter sphaeroides* SOS box. *Mol. Microbiol.* 28, 991–1003. doi: 10.1046/j.1365-2958.1998.00860.x
- Fernández, S., Sorokin, A., and Alonso, J. C. (1998). Genetic recombination in *Bacillus subtilis* 168: effects of recU and recS mutations on DNA repair and homologous recombination. *J. Bacteriol.* 180, 3405–3409.
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344
- Garriga, X., Calero, S., and Barbe, J. (1992). Nucleotide sequence analysis and comparison of the *lexA* genes from *Salmonella typhimurium*, *Erwinia carotovora*, *Pseudomonas aeruginosa* and *Pseudomonas putida*. *Mol. Gen. Genet.* 236, 125–134.
- Garrity, G. M., and Holt, J. G. (2001). “The road map to the manual,” in *Bergey's Manual® of Systematic Bacteriology*, eds D. R. Boone, R. W. Castenholz, and G. M. Garrity (New York, NY: Springer), 119–166. Available online at: //link.springer.com/chapter/10.1007/978-0-387-21609-6\_15 (Accessed May 12, 2016).
- Gillor, O., Vriezen, J. A. C., and Riley, M. A. (2008). The role of SOS boxes in enteric bacteriocin regulation. *Microbiol. Read. Engl.* 154, 1783–1792. doi: 10.1099/mic.0.2007/016139-0
- Green, M. R., and Sambrook, J. (2012). *Molecular Cloning: A Laboratory Manual (Fourth Edition): Three-volume set, 4th Edn.* New York, NY: Cold Spring Harbor Laboratory Press.
- Groban, E. S., Johnson, M. B., Banky, P., Burnett, P. G., Calderon, G. L., Dwyer, E. C., et al. (2005). Binding of the *Bacillus subtilis* LexA protein to the SOS operator. *Nucleic Acids Res.* 33, 6287–6295. doi: 10.1093/nar/gki939
- Guerin, E., Cambray, G., Sanchez-Alberola, N., Campoy, S., Erill, I., Da Re, S., et al. (2009). The SOS response controls integron recombination. *Science* 324, 1034. doi: 10.1126/science.1172914
- Gupta, R. S., Bhandari, V., and Naushad, H. S. (2012). Molecular signatures for the PVC Clade (*Planctomycetes*, *Verrucomicrobia*, *Chlamydiae*, and *Lentisphaerae*) of *Bacteria* provide insights into their evolutionary relationships. *Front. Microbiol.* 3:327. doi: 10.3389/fmicb.2012.00327
- Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., and Noble, W. S. (2007). Quantifying similarity between motifs. *Genome Biol.* 8:R24. doi: 10.1186/gb-2007-8-2-r24
- Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., and Beck, E. (2013). TIGRFAMs and genome properties in 2013. *Nucleic Acids Res.* 41, D387–D395. doi: 10.1093/nar/gks1234
- Heyer, W. D. (2004). Damage signaling: RecQ sends an SOS to you. *Curr. Biol.* 14, R895–R897. doi: 10.1016/j.cub.2004.09.073
- Jara, M., Nunez, C., Campoy, S., Fernandez de Henestrosa, A. R., Lovley, D. R., and Barbe, J. (2003). *Geobacter sulfurreducens* has two autoregulated *lexA* genes whose products do not bind the *recA* promoter: differing responses of *lexA* and



- recA to DNA damage. *J. Bacteriol.* 185, 2493–2502. doi: 10.1128/JB.185.8.2493-2502.2003
- Kamneva, O. K., Knight, S. J., Liberles, D. A., and Ward, N. L. (2012). Analysis of genome content evolution in PVC bacterial super-phylum: assessment of candidate genes associated with cellular organization and lifestyle. *Genome Biol. Evol.* 4, 1375–1390. doi: 10.1093/gbe/evs113
- Kılıç, S., White, E. R., Sagitova, D. M., Cornish, J. P., and Erill, I. (2014). CollecTF: a database of experimentally validated transcription factor-binding sites in *Bacteria*. *Nucleic Acids Res.* 42, D156–D160. doi: 10.1093/nar/gk1123
- Lagkouvardos, I., Jehl, M. A., Rattei, T., and Horn, M. (2014). Signature protein of the PVC superphylum. *Appl. Environ. Microbiol.* 80, 440–445. doi: 10.1128/AEM.02655-13
- Liou, A. P., Paziuk, M., Luevano, J.-M., Machineni, S., Turnbaugh, P. J., and Kaplan, L. M. (2013). Conserved shifts in the gut microbiota due to gastric bypass reduce host weight and adiposity. *Sci. Transl. Med.* 5, 178ra41. doi: 10.1126/scitranslmed.3005687
- Little, J. W. (1991). Mechanism of specific LexA cleavage: autodigestion and the role of RecA coprotease. *Biochimie* 73, 411–421. doi: 10.1016/0300-9084(91)90108-D
- Little, J. W., and Mount, D. W. (1982). The SOS regulatory system of *Escherichia coli*. *Cell* 29, 11–22. doi: 10.1016/0092-8674(82)90085-X
- Maiques, E., Ubeda, C., Campoy, S., Salvador, N., Lasa, I., Novick, R. P., et al. (2006). beta-lactam antibiotics induce the SOS response and horizontal transfer of virulence factors in *Staphylococcus aureus*. *J. Bacteriol.* 188, 2726–2729. doi: 10.1128/JB.188.7.2726-2729.2006
- Markowitz, V. M., Chen, I.-M. A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M., et al. (2014). IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* 42, D568–D573. doi: 10.1093/nar/gkt919
- Mazon, G., Campoy, S., Erill, I., and Barbe, J. (2006). Identification of the *Acidobacterium capsulatum* LexA box reveals a lateral acquisition of the *Alphaproteobacteria* *lexA* gene. *Microbiology* 152, 1109–1118. doi: 10.1099/mic.0.28376-0
- Mazon, G., Erill, I., Campoy, S., Cortes, P., Forano, E., and Barbe, J. (2004a). Reconstruction of the evolutionary history of the LexA-binding sequence. *Microbiology* 150, 3783–3795. doi: 10.1099/mic.0.27315-0
- Mazon, G., Lucena, J. M., Campoy, S., Fernandez de Henestrosa, A. R., Candau, P., and Barbe, J. (2004b). LexA-binding sequences in Gram-positive and cyanobacteria are closely related. *Mol. Genet. Genomics* 271, 40–49. doi: 10.1007/s00438-003-0952-x
- Nakayama, H. (2005). *Escherichia coli* RecQ helicase: a player in thymineless death. *Mutat. Res. Mol. Mech. Mutagen.* 577, 228–236. doi: 10.1016/j.mrfmmm.2005.02.015
- Oertel-Buchheit, P., Lamerichs, R. M., Schnarr, M., and Granger-Schnarr, M. (1990). Genetic analysis of the LexA repressor: isolation and characterization of LexA(Def) mutant proteins. *Mol. Gen. Genet.* 223, 40–48. doi: 10.1007/BF00315795
- Ottleben, G., Messori, L., Rüterjans, H., Kaptein, R., Granger-Schnarr, M., and Schnarr, M. (1991). 1H-NMR investigation of the interaction of the amino terminal domain of the LexA repressor with a synthetic half-operator. *J. Biomol. Struct. Dyn.* 9, 447–461. doi: 10.1080/07391102.1991.10507928
- Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., et al. (2014). eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.* 42, D231–D239. doi: 10.1093/nar/gkt1253
- Pruitt, K. D., Tatusova, T., Brown, G. R., and Maglott, D. R. (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* 40, D130–D135. doi: 10.1093/nar/gkr1079
- Sait, M., Kamneva, O. K., Fay, D. S., Kirienko, N. V., Polek, J., Shirasu-Hiza, M. M., et al. (2011). Genomic and experimental evidence suggests that *Verrucomicrobium spinosum* Interacts with *Eukaryotes*. *Front. Microbiol.* 2:211. doi: 10.3389/fmicb.2011.00211
- Sanchez-Alberola, N., Campoy, S., Barbe, J., and Erill, I. (2012). Analysis of the SOS response of *Vibrio* and other bacteria with multiple chromosomes. *BMC Genomics* 13:58. doi: 10.1186/1471-2164-13-58
- Sanchez-Alberola, N., Campoy, S., Emerson, D., Barbe, J., and Erill, I. (2015). A SOS regulon under control of a non-canonical LexA-binding motif in the *Betaproteobacteria*. *J. Bacteriol.* 197, 2622–2630. doi: 10.1128/JB.00035-15
- Sassanfar, M., and Roberts, J. W. (1990). Nature of the SOS-inducing signal in *Escherichia coli*. The involvement of DNA replication. *J. Mol. Biol.* 212, 79–96. doi: 10.1016/0022-2836(90)90306-7
- Schlieper, D., Oliva, M. A., Andreu, J. M., and Löwe, J. (2005). Structure of bacterial tubulin BtubA/B: evidence for horizontal gene transfer. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9170–9175. doi: 10.1073/pnas.0502859102
- Schneider, T. D. (1997). Information content of individual genetic sequences. *J. Theor. Biol.* 189, 427–441. doi: 10.1006/jtbi.1997.0540
- Schroeder, J. W., Randall, J. R., Matthews, L. A., and Simmons, L. A. (2015). Ribonucleotides in bacterial DNA. *Crit. Rev. Biochem. Mol. Biol.* 50, 181–193. doi: 10.3109/10409238.2014.981647
- Söding, J., Biegert, A., and Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33, W244–W248. doi: 10.1093/nar/gki408
- Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. doi: 10.1093/nar/28.1.33
- Thliveris, A. T., Little, J. W., and Mount, D. W. (1991). Repression of the *E. coli* *recA* gene requires at least two LexA protein monomers. *Biochimie* 73, 449–456. doi: 10.1016/0300-9084(91)90112-E
- Thliveris, A. T., and Mount, D. W. (1992). Genetic identification of the DNA binding domain of *Escherichia coli* LexA protein. *Proc. Natl. Acad. Sci. U.S.A.* 89, 4500–4504. doi: 10.1073/pnas.89.10.4500
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673
- Ubeda, C., Maiques, E., Knecht, E., Lasa, I., Novick, R. P., and Penades, J. R. (2005). Antibiotic-induced SOS response promotes horizontal dissemination of pathogenicity island-encoded virulence factors in staphylococci. *Mol. Microbiol.* 56, 836–844. doi: 10.1111/j.1365-2958.2005.04584.x
- Ubeda, C., Maiques, E., Tormo, M. A., Campoy, S., Lasa, I., Barbe, J., et al. (2007). SaPI operon I is required for SaPI packaging and is controlled by LexA. *Mol. Microbiol.* 65, 41–50. doi: 10.1111/j.1365-2958.2007.05758.x
- Ulrich, R. L., Deshazer, D., Kenny, T. A., Ulrich, M. P., Moravusova, A., Opperman, T., et al. (2013). Characterization of the *Burkholderia thailandensis* SOS response by using whole-transcriptome shotgun sequencing. *Appl. Environ. Microbiol.* 79, 5830–5843. doi: 10.1128/AEM.00538-13
- Vaisman, A., McDonald, J. P., Huston, D., Kuban, W., Liu, L., Houten, B. V., et al. (2013). Removal of misincorporated ribonucleotides from prokaryotic genomes: an unexpected role for nucleotide excision repair. *PLoS Genet.* 9:e1003878. doi: 10.1371/journal.pgen.1003878
- Walker, G. C., Smith, B. T., and Sutton, M. D. (2000). “The SOS response to DNA damage,” in *Bacterial Stress Responses*, eds G. Storz and R. Hengge-Aronis (Washington, DC: ASM Press), 131–144.
- Yang, M. K., Hsu, C. H., and Sung, V. L. (2008). Analyses of binding sequences of the two LexA proteins of *Xanthomonas axonopodis* pathovar citri. *Mol. Genet. Genomics* 280, 49–58. doi: 10.1007/s00438-008-0344-3
- Zhang, A. P. P., Pigli, Y. Z., and Rice, P. A. (2010). Structure of the LexA-DNA complex and implications for SOS box measurement. *Nature* 466, 883–886. doi: 10.1038/nature09200

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Erill, Campoy, Kılıç and Barbé. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# MgaSpn and H-NS: Two Unrelated Global Regulators with Similar DNA-Binding Properties

Virtu Solano-Collado<sup>1†</sup>, Mário Hüttener<sup>2</sup>, Manuel Espinosa<sup>1</sup>, Antonio Juárez<sup>2,3\*</sup> and Alicia Bravo<sup>1\*</sup>

<sup>1</sup> Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain, <sup>2</sup> Departament de Microbiologia, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain, <sup>3</sup> Institut de Bioenginyeria de Catalunya, Barcelona, Spain

## OPEN ACCESS

### Edited by:

Brian M. Baker,  
University of Notre Dame, USA

### Reviewed by:

Trevor P. Creamer,  
University of Kentucky, USA  
Kurt Henry Piepenbrink,  
University of Maryland School of  
Medicine, USA  
Aaron L. Lucius,  
University of Alabama at Birmingham,  
USA

### \*Correspondence:

Antonio Juárez  
ajuarez@ub.edu  
Alicia Bravo  
abravo@cib.csic.es

### † Present Address:

Virtu Solano-Collado,  
Institute of Medical Sciences, School  
of Medicine, Medical Sciences and  
Nutrition, University of Aberdeen,  
Aberdeen, UK

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

Received: 01 July 2016

Accepted: 15 September 2016

Published: 29 September 2016

### Citation:

Solano-Collado V, Hüttener M,  
Espinosa M, Juárez A and Bravo A  
(2016) MgaSpn and H-NS: Two  
Unrelated Global Regulators with  
Similar DNA-Binding Properties.  
Front. Mol. Biosci. 3:60.  
doi: 10.3389/fmolb.2016.00060

Global regulators play an essential role in the adaptation of bacterial cells to specific niches. Bacterial pathogens thriving in the tissues and organs of their eukaryotic hosts are a well-studied example. Some of the proteins that recognize local DNA structures rather than specific nucleotide sequences act as global modulators in many bacteria, both Gram-negative and -positive. To this class of regulators belong the H-NS-like proteins, mainly identified in  $\gamma$ -Proteobacteria, and the MgaSpn-like proteins identified in Firmicutes. H-NS and MgaSpn from *Escherichia coli* and *Streptococcus pneumoniae*, respectively, neither have sequence similarity nor share structural domains. Nevertheless, they display common features in their interaction with DNA, namely: (i) they bind to DNA in a non-sequence-specific manner, (ii) they have a preference for intrinsically curved DNA regions, and (iii) they are able to form multimeric complexes on linear DNA. Using DNA fragments from the hemolysin operon regulatory region of the *E. coli* plasmid pHly152, we show in this work that MgaSpn is able to recognize particular regions on extended H-NS binding sites. Such regions are either located at or flanked by regions of potential bendability. Moreover, we show that the regulatory region of the pneumococcal *P1623B* promoter, which is recognized by MgaSpn, contains DNA motifs that are recognized by H-NS. These motifs are adjacent to regions of potential bendability. Our results suggest that both regulatory proteins recognize similar structural characteristics of DNA.

**Keywords:** global transcriptional regulators, nucleoid-associated proteins, Mga/AtxA family, protein-DNA interactions, DNA bendability

## INTRODUCTION

Global modulators play key roles in the ability of bacterial cells to rapidly adapt to environmental fluctuations by adjusting their gene expression pattern. As a consequence, they enable the pathogenic bacteria to colonize and survive in different niches of their eukaryotic hosts. Whereas, some global modulators recognize specific DNA sequences, others exhibit a preference for particular DNA structures. Examples of the latter group are the H-NS-like proteins, mainly found in  $\gamma$ -Proteobacteria, and the MgaSpn-like proteins identified in Firmicutes.

In *Escherichia* and *Salmonella*, the DNA-binding properties of the nucleoid-associated protein H-NS (137 amino acids) have been studied in detail (for a review see Winardhi et al., 2015).

Like other nucleoid-associated proteins, H-NS is involved in both organization of the bacterial chromosome and regulation of gene expression. H-NS functions generally as a repressor or gene silencer, but it can also act indirectly as a transcriptional activator (Ko and Park, 2000). Many genes regulated by H-NS encode virulence determinants. H-NS consists of an N-terminal oligomerization domain and a C-terminal DNA-binding domain. In solution, H-NS is able to form higher-order oligomers. This ability correlates with its ability to form nucleoprotein filaments (Lim et al., 2012). Imaging studies have revealed that H-NS-like proteins are able to organize large DNA molecules into various conformations, including extended nucleoprotein filaments, hairpin-like large DNA bridges, and higher-order DNA condensations (Dame et al., 2000; Liu et al., 2010).

*In vitro* DNA binding experiments have shown that H-NS binds to DNA in a non-specific manner, although it has a strong preference for intrinsically curved AT-rich DNA regions. Furthermore, high-affinity DNA-binding sites for H-NS have been identified in AT-rich regions of the chromosomal DNA (Lang et al., 2007). Formation of H-NS nucleoprotein filaments from such high-affinity sites (nucleation sites) may lead to selective gene silencing either by inhibiting the binding of the RNA polymerase to the promoter region or by blocking RNA polymerase translocation (reviewed by Winardhi et al., 2015). Studies on the *LEE5* promoter supported an additional model for H-NS-mediated repression. In this model, H-NS spreading from a site located upstream of the *LEE5* promoter to a site located at the promoter would facilitate specific contacts between H-NS and the RNA polymerase (Shin et al., 2012).

H-NS is known to modulate the expression of the thermoregulated hemolysin (*hly*) operon (genes *hlyC*, *hlyA*, *hlyB*, and *hlyD*), which encodes the toxin  $\alpha$ -hemolysin and additional gene products required for its activation and export. This toxin is produced by several uropathogenic *E. coli* strains. The *hly* operon of the *E. coli* plasmid pHly152 has been studied in detail. First, Vogel et al. (1988) identified an essential regulatory sequence located ~2 kbp upstream of the *hly* operon, the so-called *hlyR* sequence. Between the *hlyR* sequence and the promoter region of the *hly* operon there is an IS2 insertion element. Subsequently, Madrid et al. (2002) identified two extended H-NS binding sites upstream of the *hly* operon (see **Figure 1A**). One of them (site I; nucleotides 190 to 350 of plasmid pHly152) is located within the *hlyR* sequence. The second one (site II; nucleotides 2180 to 2330) overlaps the promoter region of the *hly* operon. A deletion analysis confirmed the relevance of site I for thermoregulation of the *hly* operon (Madrid et al., 2002).

The MgaSpn transcriptional regulator (493 amino acids) contributes to the virulence of *Streptococcus pneumoniae*. Some DNA-binding properties of MgaSpn resemble the ones reported for H-NS. Specifically, *in vitro* DNA binding studies (gel retardation, footprinting and electron microscopy) have shown that MgaSpn generates multimeric complexes on linear

double-stranded DNA. Furthermore, MgaSpn binds to DNA with little or no sequence specificity, and shows a preference for DNA regions that contain a potential intrinsic curvature (Solano-Collado et al., 2013). Nevertheless, despite this fact, MgaSpn and H-NS are unrelated proteins. They neither have sequence similarity nor share structural domains. MgaSpn is a member of a new class of global response regulators known as the Mga/AtxA family, which includes the Mga, AtxA and MafR proteins from *S. pyogenes*, *Bacillus anthracis* and *Enterococcus faecalis*, respectively (Hondorp et al., 2013; Hammerstrom et al., 2015; Ruiz-Cruz et al., 2016). According to the Pfam database (Finn et al., 2016), MgaSpn has two putative N-terminal DNA-binding domains, the so-called HTH\_Mga (residues 6 to 65) and Mga (residues 71 to 158) domains (Solano-Collado et al., 2012). These domains do not exhibit sequence similarity with the C-terminal DNA-binding domain of H-NS.

MgaSpn plays a significant role in both nasopharyngeal colonization and lung infection in mice (Hemsley et al., 2003). *In vivo* experiments showed that MgaSpn activates the pneumococcal *P1623B* promoter and, consequently, the transcription of a four-gene operon (*spr1623-spr1626*) of unknown function. This activation requires a 70-bp region (*PB* activation region) located upstream of the *P1623B* promoter (Solano-Collado et al., 2012) (see **Figure 1B**). Interestingly, MgaSpn recognizes the *PB* activation region as a primary binding site when it is located at internal position on a 222-bp DNA fragment, but not when it is positioned at one end of the DNA fragment (Solano-Collado et al., 2013). According to the bend.it program (Vlahovicek et al., 2003), the *PB* activation region contains a potential intrinsic curvature flanked by regions of bendability (**Figure 1B**).

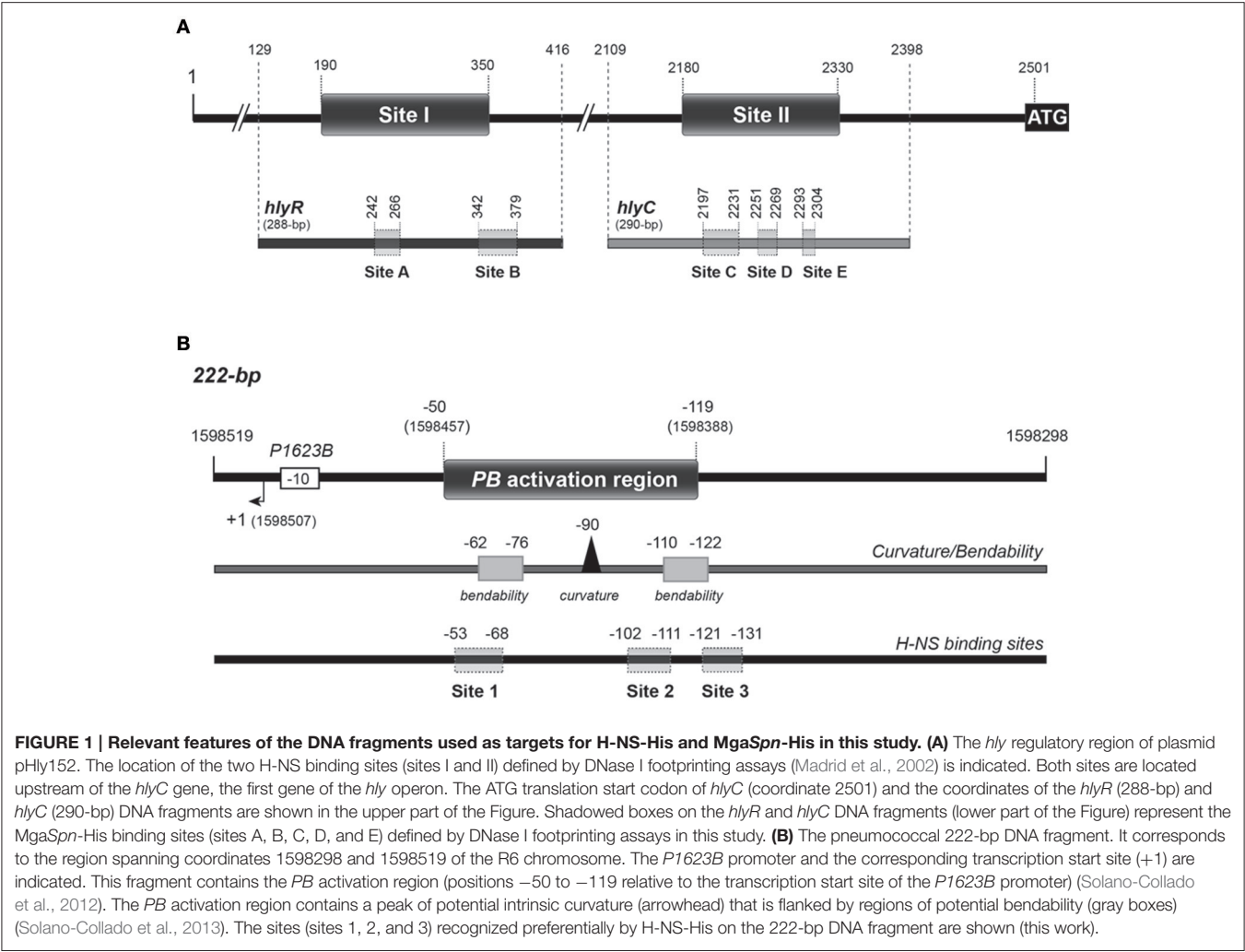
From the information mentioned above, it is apparent that H-NS and MgaSpn share some DNA-binding properties. In light of these observations, we hypothesized that both unrelated proteins might recognize similar characteristics of DNA. In this work we present evidence that supports this hypothesis. By gel retardation and DNase I footprinting assays, we show that MgaSpn is able to recognize particular regions on extended H-NS binding sites and vice versa.

## MATERIALS AND METHODS

### Polymerase Chain Reaction (PCR)

The Phusion High-Fidelity DNA polymerase (Thermo Scientific) was used. Reaction mixtures (50  $\mu$ l) contained 5–20 ng of template DNA, 20 pmol of each primer, 200  $\mu$ M each deoxynucleoside triphosphate (dNTP), and one unit of DNA polymerase. A initial denaturation step was performed at 98°C for 1 min, followed by 30 cycles that included the next steps: (i) denaturation at 98°C for 10 s, (ii) annealing of the primers around 55°C (depending on the primer melting temperature) for 20–30 s, and (iii) extension at 72°C for 20–40 s (depending on the amplicon length). A final extension step was performed at 72°C for 10 min. PCR products were cleaned up with the QIAquick PCR purification kit (Qiagen).

**Abbreviations:** BSA, bovine serum albumin; EMSA, electrophoretic mobility shift assays; *hly*, hemolysin; PCR, polymerase chain reaction; T4 PNK, T4 polynucleotide kinase.



PCR-Amplification of DNA Regions

Oligonucleotides used for PCR amplifications are listed in Table 1. As DNA templates, chromosomal DNA from the pneumococcal R6 strain (Hoskins et al., 2001) and pANN202312R plasmid DNA from *E. coli* (Godessart et al., 1988) were used. Chromosomal DNA from *S. pneumoniae* was prepared as described previously (Ruiz-Cruz et al., 2010). For small-scale preparations of plasmid DNA, the High Pure Plasmid Isolation Kit (Roche Applied Science) was used. The 222-bp DNA region of the R6 chromosome (coordinates 1598298–1598519) was amplified using the 1622H and 1622I primers. The 288-bp *hlyR* DNA fragment (coordinates 129–416 in Madrid et al., 2002) was amplified using the *hlyR*-Fw and *hlyR*-Rev primers. The 290-bp *hlyC* DNA fragment (coordinates 2109–2398 in Madrid et al., 2002) was amplified using the *hlyC*-Fw and *hlyC*-Rev primers.

Radioactive Labeling of DNA Fragments

Oligonucleotides were radioactively labeled at the 5'-end using [ $\gamma$ -<sup>32</sup>P]ATP (PerkinElmer) and T4 polynucleotide kinase (T4 PNK; New England Biolabs). Reactions (25  $\mu$ l) contained 30

TABLE 1 | Oligonucleotides used in this work.

Name	Sequence 5' to 3'
1622H	CGGATTAAACCTCTTGAATTATACC
1622I	CAAATTCTTTAATTGTTGCTATTA
<i>hlyR</i> -Fw	GCCCACTGCATTGAATACTTAC
<i>hlyR</i> -Rev	CATACTTACCTACAGCTATAAG
<i>hlyC</i> -Fw	CAGATAAAGAAGAGTAGTTC
<i>hlyC</i> -Rev	CAAGTTTTTATTGCGCTGACTAAC
Hly Sal/Eco5	CAGACCACACCTGGAAAAAC
Hly Sal/Hind3	GGGCTTCACTGCGAAATTCA

pmol of oligonucleotide, 2.5  $\mu$ l of 10  $\times$  kinase buffer (provided by the supplier), 50 pmol of [ $\gamma$ -<sup>32</sup>P]ATP (3000 Ci/mmol, 10 mCi/ml) and 10 units of T4 PNK. After incubation at 37°C for 30 min, additional T4 PNK (10 units) was added. Reaction mixtures were then incubated at 37°C for 30 min. The enzyme was inactivated by incubation at 65°C for 20 min. Non-incorporated nucleotide was removed using Illustra MicroSpin™ G-25 columns (GE Healthcare). The 5'-labeled oligonucleotides were used for



manual sequencing and for PCR amplification to obtain double-stranded DNA fragments labeled at either the coding or the non-coding strand.

### Purification of MgaSpn-His and H-NS-His

Gene *mgaSpn* was engineered to encode a His-tagged MgaSpn protein (MgaSpn-His). This variant of MgaSpn carries six additional His residues at the C-terminal end. The procedure used to overproduce and purify MgaSpn-His was reported previously (Solano-Collado et al., 2012). Purified H-NS-His protein was obtained as described by Nieto et al. (1991).

### Electrophoretic Mobility Shift Assays (EMSA)

Binding reactions (10  $\mu$ l) contained 40 mM Tris-HCl, pH 7.6, 1 mM DTT, 0.4 mM EDTA, 1–2% glycerol, 50 mM NaCl, 10 mM MgCl<sub>2</sub>, 500  $\mu$ g/ml bovine serum albumin (BSA), 2 nM of 5'-labeled DNA and varying concentrations of MgaSpn-His (20 to 180 nM). Reactions were incubated at ambient temperature for 20 min. Free and bound DNA forms were separated on native polyacrylamide (5%) gels (Mini-PROTEAN system, Bio-Rad) using Tris-borate-EDTA, pH 8.3, buffer (TBE). Gels were run at 100 V and ambient temperature. Labeled DNA was visualized using a Fujifilm Image Analyzer FLA-3000 and quantified using the Quantity One software (Bio-Rad).

For competitive EMSA, a 407-bp DNA fragment from the 5' untranslated region of the *hlyR* gene was generated by PCR using primers Hly Sal/Eco5 and Hly Sal/Hind3 (Table 1). For each reaction, 50 ng of the pneumococcal 222-bp DNA and 150 ng of the 407-bp DNA (competitor DNA) were mixed with increasing concentrations of H-NS-His in binding buffer (250 mM HEPES, pH 7.4, 350 mM KCl, 5 mM EDTA, 5 mM DTT, 500  $\mu$ g/ml BSA, 25% glycerol) and incubated at 37°C for 30 min. Samples (20  $\mu$ l) were loaded onto native polyacrylamide (5%) gels (TBE buffer). Bands were visualized using a Gel-doc system (Bio-Rad).

### DNase I Footprinting Assays

In the case of H-NS-His, binding reactions (10  $\mu$ l) contained 30 mM Tris-HCl, pH 7.6, 1 mM DTT, 1 mM CaCl<sub>2</sub>, 10 mM MgCl<sub>2</sub>, 100 mM NaCl, 1% glycerol, 2 nM <sup>32</sup>P-labeled DNA and different concentrations of H-NS-His (2 nM to 500 nM). For MgaSpn-His, binding reactions (10  $\mu$ l) contained 40 mM Tris-HCl, pH 7.6, 1.2 mM DTT, 0.2 mM EDTA, 1 mM CaCl<sub>2</sub>, 10 mM MgCl<sub>2</sub>, 50 mM NaCl, 1% glycerol, 500  $\mu$ g/ml BSA, 2 nM <sup>32</sup>P-labeled DNA and different concentrations of MgaSpn-His (10 to 200 nM). In all cases, after 20 min at ambient temperature, 0.03 units of DNase I (Roche Applied Science) was added and the reaction proceeded for 5 min at the same temperature. DNase I digestion was stopped by adding 1  $\mu$ l of 250 mM EDTA. Then, 4  $\mu$ l of loading buffer (80% formamide, 1 mM EDTA, 10 mM NaOH, 0.1% bromophenol blue and 0.1% xylene cyanol) was added. After heating at 95°C for 5 min, samples were loaded onto 8 M urea-6% polyacrylamide gels. Dideoxy-mediated chain termination sequencing reactions were run in the same gel. Labeled products were visualized using a Fujifilm Image Analyser FLA-3000. The intensity of the bands was quantified using the Quantity One software (Bio-Rad).

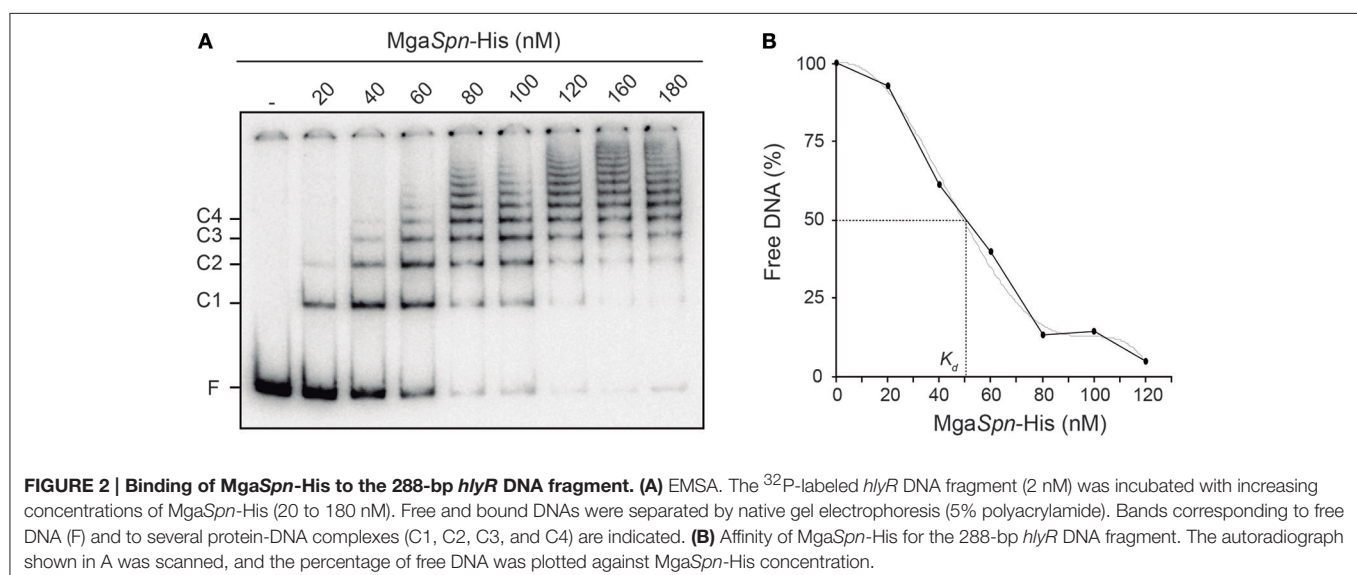
### In silico Prediction of Intrinsic Curvature

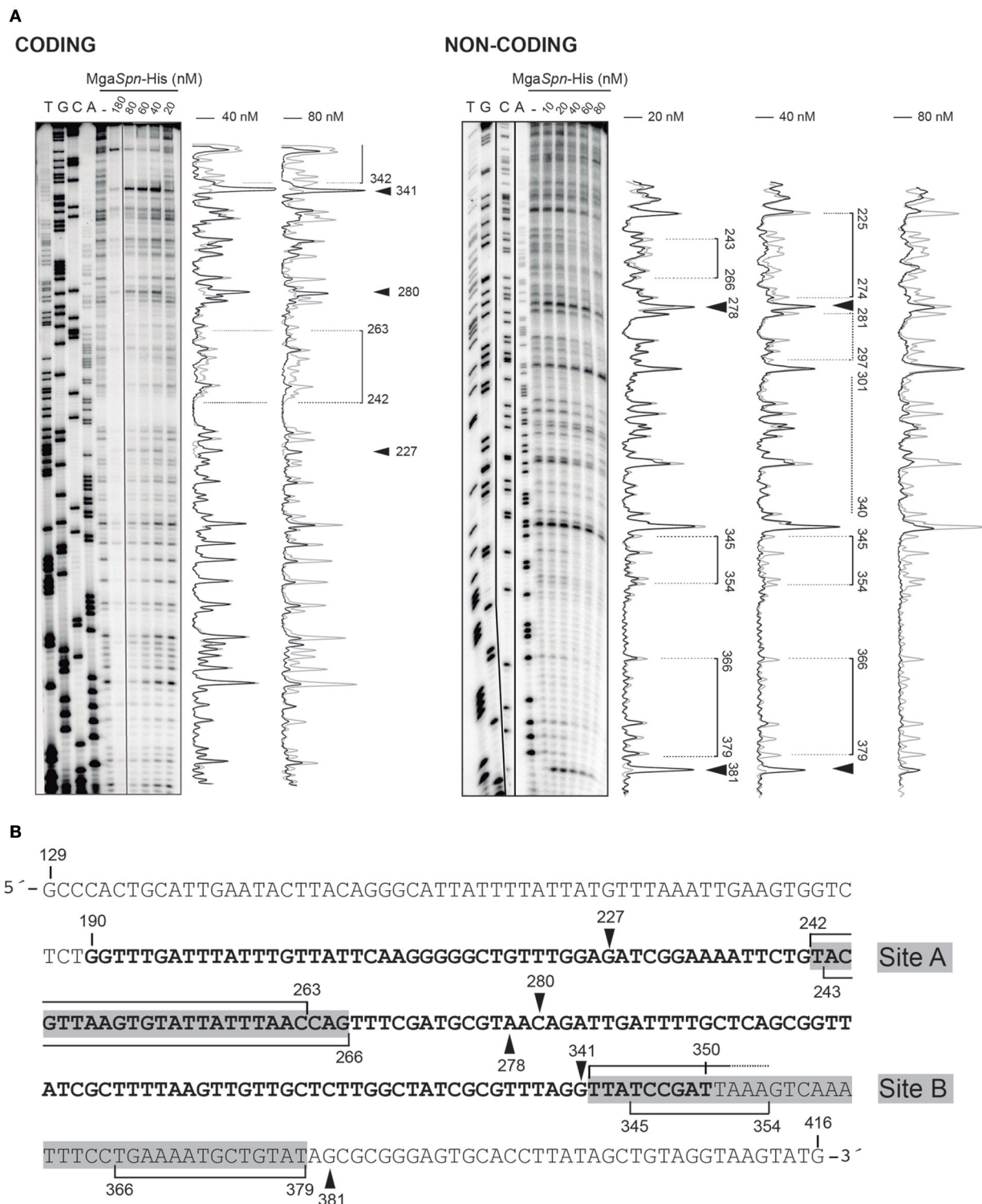
The bendability/curvature propensity plots of the DNA fragments used in this study were calculated with the bend.it server (Vlahovicek et al., 2003; [http://hydra.icgeb.trieste.it/dna/bend\\_it.html](http://hydra.icgeb.trieste.it/dna/bend_it.html)) as described previously (Solano-Collado et al., 2013).

## RESULTS

### Binding of MgaSpn-His to the *hlyR* Region of the *E. coli* *hly* Operon

In the *E. coli* plasmid pHly152, previous DNase I footprinting assays revealed the existence of two extended H-NS binding sites upstream of the *hly* operon (sites I and II in Figure 1A)





**FIGURE 3 | DNase I footprints of complexes formed by MgaSpn-His on the 288-bp *hlyR* DNA fragment. (A)** The *hlyR* DNA fragment (coordinates 129–416) was radioactively labeled at the 5'-end of either the coding or the non-coding strand. Labeled DNA (2 nM) was incubated with the indicated concentrations of MgaSpn-His and then digested with DNase I. Dideoxy-mediated chain termination sequencing reactions were run in the same gel (lanes A, C, G, T). Densitometer scans corresponding to free DNA (gray line) and DNA with protein (black line) are shown. Brackets represent the MgaSpn-His protected regions. Sites more sensitive to DNase I cleavage are indicated with arrowheads. **(B)** Nucleotide sequence of the *hlyR* DNA fragment. The H-NS binding site I (coordinates 190–350) is highlighted in bold. The two sites recognized by MgaSpn-His (sites A and B) are marked with gray boxes. MgaSpn-His protected regions on either the coding or the non-coding strand (brackets) as well as sites more sensitive to DNase I cleavage (arrowheads) are indicated.

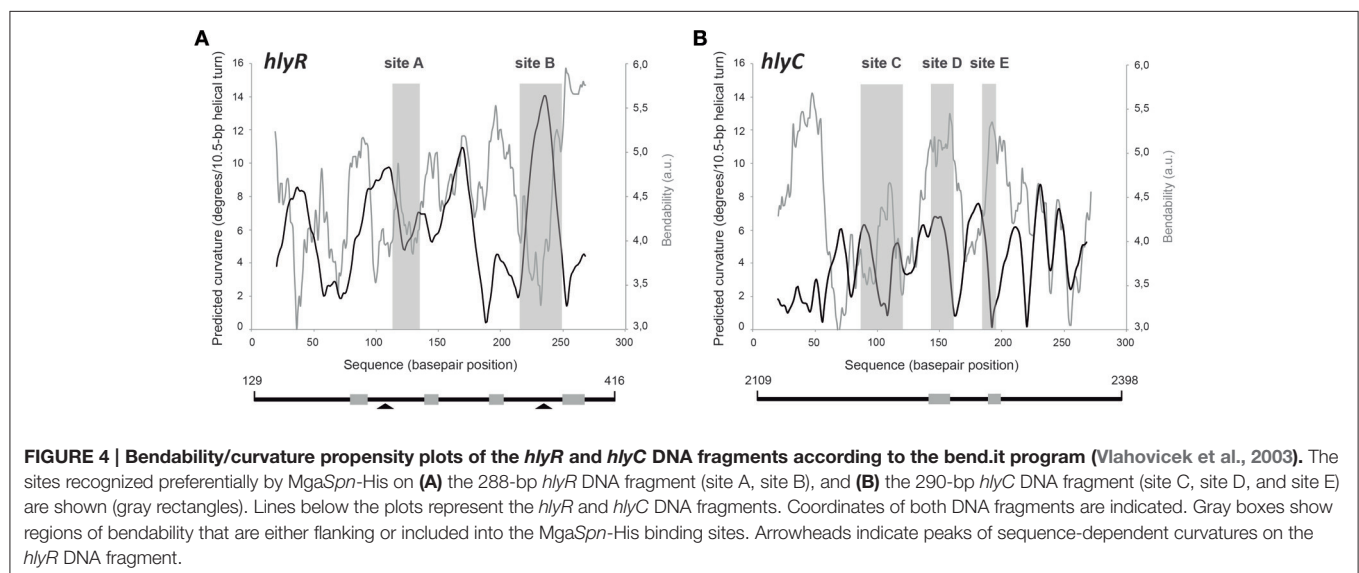
(Madrid et al., 2002). One of them (site I; coordinates 190–350 of pHly152) is included within the so-called *hlyR* regulatory sequence (Vogel et al., 1988), and was shown to play a significant role in the thermoregulation of the *hly* operon (Madrid et al., 2002). To investigate whether the pneumococcal MgaSpn protein is able to recognize particular regions on the H-NS binding site I, we performed EMSA and DNase I footprinting experiments. We used a His-tagged version of MgaSpn (MgaSpn-His) and a 288-bp DNA fragment (here named *hlyR*; coordinates 129–416) that contains the site I (Figure 1A). The presence of a His-tag at the C-terminal end of MgaSpn does not affect its DNA-binding properties (Solano-Collado et al., 2012, 2013). For EMSA, radioactively labeled DNA was incubated with increasing concentrations of MgaSpn-His (Figure 2A). At 20 nM of MgaSpn-His, free DNA and two protein-DNA complexes were detected. However, in agreement with previous results (Solano-Collado et al., 2013), as the concentration of MgaSpn-His was increased, complexes of lower electrophoretic mobility appeared sequentially and complexes moving faster disappeared gradually, indicating that multiple protein units bind orderly on the same DNA molecule (formation of multimeric protein-DNA complexes).

By EMSA, we estimated the affinity of MgaSpn-His for the 288-bp *hlyR* DNA fragment (Figure 2B). Since MgaSpn-His generates multiple protein-DNA complexes, the protein concentration required to bind half the DNA was determined by measuring the decrease in free DNA rather than the increase in complexes, which gives an indication of the approximate magnitude of the dissociation constant,  $K_d$  (Carey, 1988). Such a concentration was about 50 nM. This value is similar to the apparent  $K_d$  of MgaSpn for the pneumococcal 222-bp DNA fragment that harbors the *PB* activation region (MgaSpn binding site) (see Figure 1B) (Solano-Collado et al., 2013).

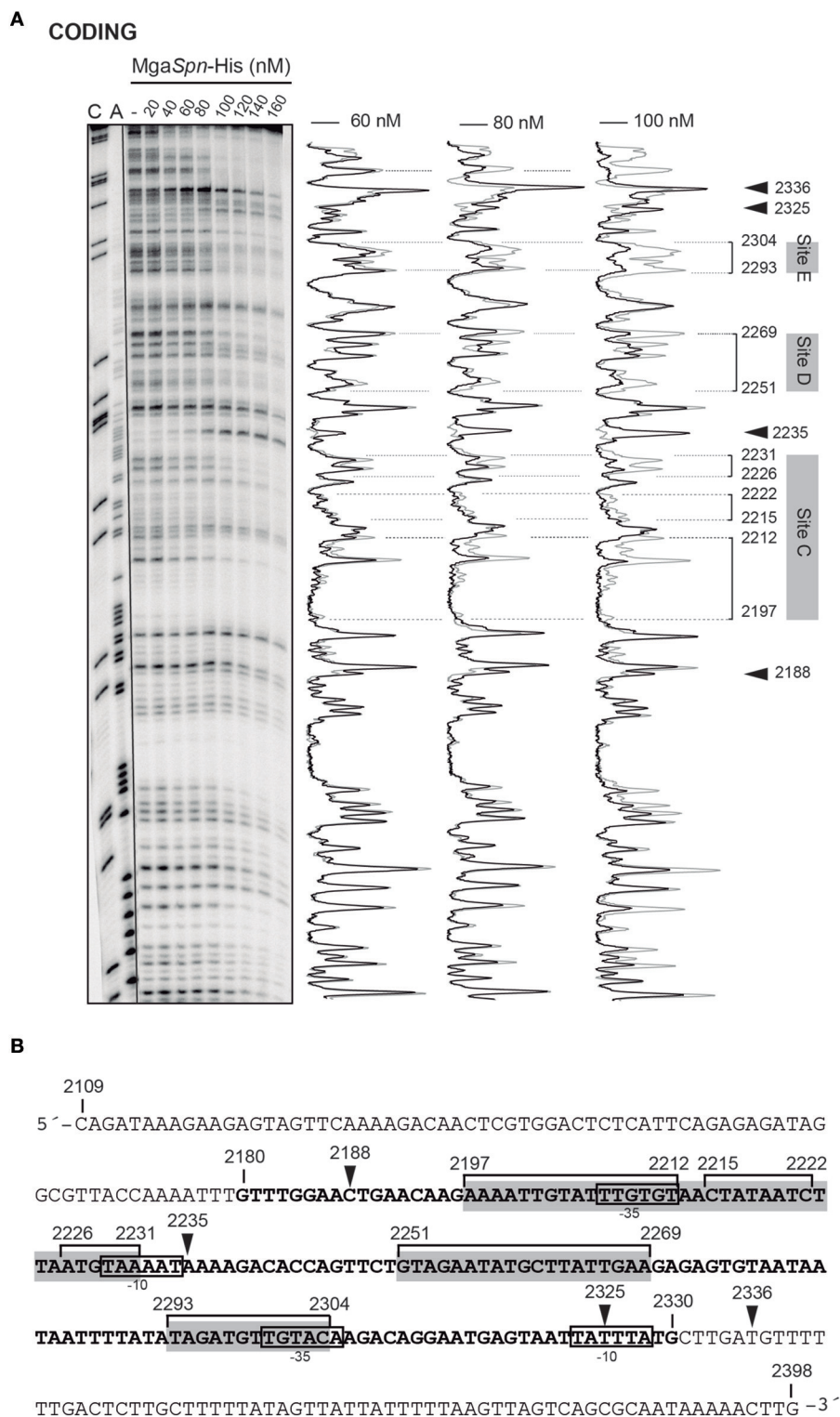
The position of MgaSpn-His on the *hlyR* DNA fragment was further analyzed by DNase I footprinting assays (Figure 3). The DNA fragment was labeled either at the 5'-end of the coding

strand or at the 5'-end of the non-coding strand. On the coding strand and at 40 nM of MgaSpn-His, protections against DNase I digestion were observed at a particular region (from position 242 to 263) (Figure 3A). Diminished cleavages were also observed from position 342 onwards (no resolution in the gel). Moreover, positions 227, 280, and 341 were more sensitive to DNase I cleavage. On the non-coding strand and at 20 nM of MgaSpn-His, diminished cleavages were observed from 243 to 266, from 345 to 354, and from 366 to 379 (Figure 3A). Additionally, the 278 and 381 positions were more sensitive to DNase I cleavage. These results indicated that MgaSpn-His recognizes preferentially two sites on the *hlyR* DNA fragment (Figure 3B). One of them (site A; positions 242 to 266) is located within the H-NS binding site I, whereas the other one (site B, positions 342 to 379) is adjacent to it. On both strands and at 80 nM of MgaSpn-His, regions protected against DNase I digestion were observed along the DNA fragment (Figure 3A), which is consistent with the pattern of protein-DNA complexes observed by EMSA (Figure 2A).

Figure 4A shows the bendability/curvature propensity plot of the 288-bp *hlyR* DNA fragment (coordinates 129–416) according to the bend.it program (Vlahovicek et al., 2003). The profile contains a peak of potential sequence-dependent curvature at position 238, just adjacent to the MgaSpn-His binding site A (positions 242–266). Its magnitude (9.7 degrees per helical turn) is within the values calculated for experimentally tested curved motifs (Gabrielian et al., 1997). Furthermore, the profile reveals that site A is flanked by regions of potential bendability (positions 215–222 and 271–278). The MgaSpn-His binding site B (positions 342–379) contains a peak of predicted curvature at position 364 (magnitude 14) which is also flanked by regions of potential bendability (positions 319–335 and 380–396). Thus, on the *hlyR* DNA fragment (see Figure 1A), MgaSpn-His binds preferentially to two sites (sites A and B) that are flanked by regions of potential bendability. Whereas, site A is located within the extended H-NS binding site I, site B is just adjacent to it.







**FIGURE 5 | DNase I footprints of complexes formed by MgaSpn-His on the 290-bp *hlyC* DNA fragment. (A)** The *hlyC* DNA fragment (coordinates 2109–2398) was radioactively labeled at the 5'-end of the coding strand. Labeled DNA (2 nM) was incubated with the indicated concentrations of MgaSpn-His and then digested with DNase I. Dideoxy-mediated chain-termination sequencing reactions were run in the same gel (lanes A, C). Densitometer scans corresponding to free DNA (gray line) and DNA with protein (black line; 60, 80, and 100 nM) are shown. The three primary sites (site C, site D and site E) recognized by MgaSpn-His on

(Continued)



**FIGURE 5 | Continued**

the *hlyC* DNA fragment are indicated. Arrowheads indicate sites more sensitive to DNase I digestion. **(B)** Nucleotide sequence of the *hlyC* DNA fragment. The  $-10$  and  $-35$  elements of two of the three promoters described for the *hly* operon (Koronakis and Hughes, 1988) are indicated. The H-NS binding site II (coordinates 2180–2330) identified by Madrid et al. (2002) is shown (bold letters). The three primary sites (site C, site D, and site E) recognized by MgaSpn-His are indicated with gray boxes. MgaSpn-His protected regions (brackets) as well as sites more sensitive to DNase I cleavage (arrowheads) are shown.

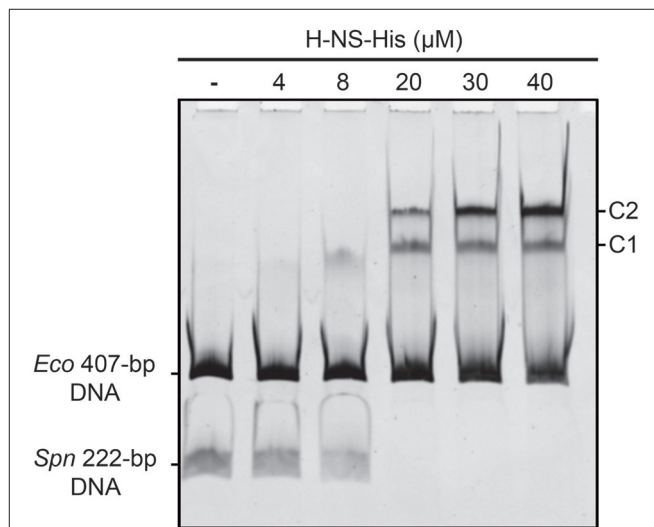
## Binding of MgaSpn-His to the Promoter Region of the *E. coli hly* Operon

H-NS interacts not only with the site I of plasmid pHly152 but also with the site II (coordinates 2180–2330), which is located upstream of *hlyC*, the first gene of the *hly* operon (Madrid et al., 2002) (see **Figure 1A**). Site II includes two of the three promoters described for the *hly* operon (Koronakis and Hughes, 1988). To analyze whether MgaSpn-His recognizes particular regions on the H-NS binding site II, we used a 290-bp DNA fragment (here named *hlyC*; coordinates 2109–2398) that contains the site II at internal position (**Figure 1A**). Radioactively labeled DNA was incubated with increasing concentrations of MgaSpn-His. By EMSA, we found that MgaSpn-His also generates multimeric complexes on the *hlyC* DNA fragment (**Supplementary Figure 1**). The apparent  $K_d$  of MgaSpn-His for the *hlyC* DNA fragment was about 75 nM, slightly higher than for the *hlyR* DNA fragment (**Figure 2B**).

The position of MgaSpn-His on the *hlyC* DNA fragment was further analyzed by DNase I footprinting assays (**Figure 5A**). To this end, the 290-bp DNA fragment was labeled at the 5'-end of the coding strand. At 60 and 80 nM of MgaSpn-His, diminished DNase I cleavages were mainly observed from coordinate 2197 to 2231 (site C), from 2251 to 2269 (site D), and from 2293 to 2304 (site E). Moreover, positions 2188, 2235, 2325, and 2336 were more sensitive to DNase I digestion. At higher protein/DNA ratios, MgaSpn-mediated protections were observed along the *hlyC* DNA fragment (**Figure 5A**). Therefore, MgaSpn-His recognizes preferentially three regions within the extended H-NS binding site II (**Figure 5B**). Compared to the 288-bp *hlyR* DNA fragment (**Figure 4A**), the magnitude of the curvatures predicted in the 290-bp *hlyC* DNA fragment is slightly lower ( $<9$  degrees per helical turn) (**Figure 4B**). Nevertheless, the MgaSpn-His binding sites D (2251–2269) and E (2293–2304) are located at regions of conspicuous bendability (positions 2248–2267 and 2296–2309, respectively).

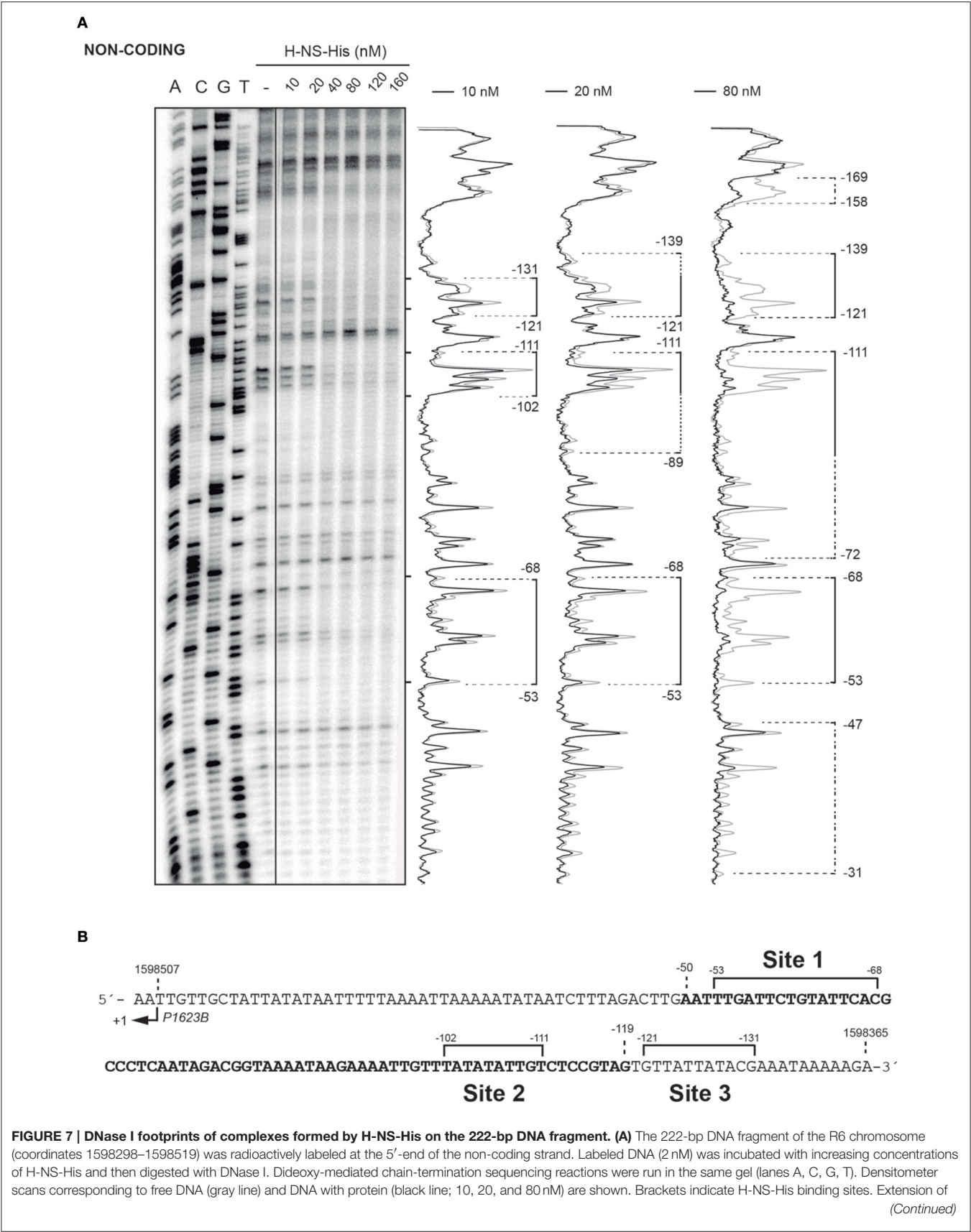
## Binding of H-NS-His to the Pneumococcal *PB* Activation Region

*In vivo* activation of the *P1623B* promoter requires a 70-bp region (*PB* activation region) located upstream of the promoter (from position  $-50$  to  $-119$ ) (**Figure 1B**). By DNase I footprinting experiments, we demonstrated previously that MgaSpn-His interacts with the *PB* activation region (positions  $-52$  to  $-102$ ) when it is located at internal position on a 222-bp DNA fragment (coordinates 1598298 to 1598519 of the pneumococcal R6 genome; see **Figure 1B**) (Solano-Collado et al., 2012). Similar results were obtained using an untagged form of the MgaSpn protein (Solano-Collado et al., 2013). In the present study, we analyzed whether protein H-NS-His is able to



**FIGURE 6 | Binding of H-NS-His to the pneumococcal 222-bp DNA fragment (competitive EMSA).** The pneumococcal 222-bp DNA fragment (50 ng) was incubated with increasing concentrations of H-NS-His in the presence of competitor DNA (407-bp DNA fragment from the *E. coli* plasmid pHly152; 150 ng). Free and bound DNAs were separated by native gel electrophoresis (5% polyacrylamide). Bands corresponding to free DNAs and protein-DNA complexes (C1 and C2) are indicated.

bind to the pneumococcal 222-bp DNA fragment that contains the *PB* activation region. First, we performed a competitive gel retardation assay (**Figure 6**). The pneumococcal 222-bp DNA fragment was mixed with a 407-bp DNA fragment (competitor DNA) from the *E. coli* plasmid pHly152, and both DNAs were incubated with increasing concentrations of H-NS-His. The 407-bp DNA fragment was reported to lack preferential binding sites for H-NS (Madrid et al., 2002). As shown in **Figure 6**, protein H-NS-His showed a higher affinity for the pneumococcal 222-bp DNA fragment. Next, we used DNase I footprinting to identify the sites recognized by H-NS-His. The pneumococcal 222-bp DNA fragment was radioactively labeled at the 5'-end of the non-coding strand, and then it was incubated with increasing concentrations of H-NS-His (**Figure 7A**). At 10 nM of H-NS-His, changes in DNase I sensitivity (diminished cleavages) were observed from positions  $-53$  to  $-68$  (site 1),  $-102$  to  $-111$  (site 2), and  $-121$  to  $-131$  (site 3). Whereas, site 1 and site 2 are located within the *PB* activation region, site 3 is adjacent to it (**Figure 7B**). At 80 nM of protein, H-NS-His mediated protections were observed along the entire DNA fragment (**Figure 7A**). According to predictions of intrinsic DNA curvature (Solano-Collado et al., 2013), the pneumococcal 222-bp DNA fragment contains one peak of potential curvature (magnitude 9.5, position  $-90$ )



**FIGURE 7 | Continued**

the H-NS-His mediated protections as the protein concentration increases is indicated with dotted lines. **(B)** Nucleotide sequence of the region spanning coordinates 1598365 and 1598509 of the 222-bp DNA fragment. The transcription start site from the *P1623B* promoter (+1) is shown. The *PB* activation region (positions –50 to –119), which includes the *MgaSpn* binding site, is highlighted in bold. The primary sites (sites 1, 2, and 3) recognized by H-NS-His on the 222-bp DNA fragment are indicated (brackets).

<b>A</b>		<b>A+T (%)</b>
<b><i>PB</i> region</b>	GTATTTCACGCCCTCAATAGACGGTAAATAAGAAATTGT	65
<b><i>Pmga</i> region</b>	ATTCTTTGTGGTATAATTGCAAGAGGTTTAAATCCGATAATTTAT	72.7
<b>B</b>		<b>A+T (%)</b>
<b>Site A</b>	TACGTTAAGTCTATTATTAAACCAG	72
<b>Site B</b>	TTATCCGATTAAACTCAAATTTCCGAAAACTGCTGTAT	71.1
<b>Site C</b>	AAAATTGTATTTGTCTAACTATAATCTTAATCTAA	82.9
<b>Site D</b>	GTACAATATGCTTATTGAA	73.7
<b>Site E</b>	TAGATGTTGTAC	66.7

**FIGURE 8 | Sites recognized by MgaSpn. (A)** Nucleotide sequence alignment of the *PB* activation region and the *Pmga* promoter region of the pneumococcal R6 genome. Both *MgaSpn* binding sites were defined by hydroxyl radical footprinting assays (Solano-Collado et al., 2013). Identical nucleotides are highlighted in gray boxes. **(B)** Nucleotide sequence alignment of the regions recognized by MgaSpn-His on the *hly* regulatory region of the *E. coli* plasmid pHly152 (this work). These *MgaSpn*-His binding sites were defined by DNase I footprinting assays. Nucleotides of such sites that are present in the *PB* activation region and/or the *Pmga* promoter region are highlighted in black boxes. The A+T content of the different sites is indicated on the right. The global A+T content of the pneumococcal R6 genome is 60.3%. By EMSA, the minimum DNA size required for *MgaSpn* binding was shown to be between 20 and 26 bp (Solano-Collado et al., 2013).

within the *PB* activation region (**Figure 1B**). Furthermore, there are two regions of potential bendability (from –62 to –76 and from –110 to –122) flanking such a curvature. Hence, the three sites recognized by H-NS-His on the pneumococcal 222-bp DNA fragment are adjacent to regions of potential bendability.

## DISCUSSION

Proteins *MgaSpn* and H-NS are very different in size. They neither exhibit sequence similarity nor share a common domain structure. *MgaSpn* is a member of an emerging class of global response regulators (the Mga/AtxA family) that contain phosphoenolpyruvate phosphotransferase system (PTS) regulation domains (PRDs) (Hammerstrom et al., 2015). *MgaSpn* is predicted to have two N-terminal helix-turn-helix DNA-binding motifs, a central PRD and a C-terminal region with amino acid similarity to the PTS protein EIIB. On the other hand, proteins in the H-NS family consist of a coiled-coil N-terminal oligomerization domain and a C-terminal DNA-binding domain. Both domains are joined via an unstructured flexible linker (reviewed by Winardhi et al., 2015). Despite these differences between *MgaSpn* and H-NS, previous studies on the *MgaSpn* transcriptional regulator suggested that it shares certain DNA-binding properties with H-NS (Solano-Collado et al., 2013).

Based on *in vitro* DNA binding studies, we have proposed that *MgaSpn* regulates the expression of numerous genes by a mechanism that involves recognition of particular DNA conformations (Solano-Collado et al., 2013). By hydroxyl radical footprinting experiments, *MgaSpn* was

shown to bind to two regions of the *S. pneumoniae* R6 chromosome: the *PB* activation region (positions –60 to –99 of the *P1623B* promoter) and the *Pmga* promoter region (positions –23 to +21 of the *Pmga* promoter) (Solano-Collado et al., 2013). Both *MgaSpn* binding regions share the GGT(A/T)(A/T)AAT(A/C)(A/C)GA(A/T)AATT sequence element (**Figure 8A**). Moreover, they contain a potential intrinsic curvature flanked by regions of bendability (Solano-Collado et al., 2013). Results presented in this work support that *MgaSpn* recognizes structural features in its DNA targets. Specifically, DNase I footprinting experiments allowed us to identify five primary sites for *MgaSpn* on the *hly* operon regulatory region of the *E. coli* plasmid pHly152. All of them are located within or adjacent to extended H-NS binding sites (**Figure 1A**). Moreover, four out of the five *MgaSpn* binding sites (i) have a size between 19 and 38 bp, (ii) display a high A+T content (71.1–82.9%), (iii) share short DNA sequence motifs with the two pneumococcal *MgaSpn* binding sites (**Figure 8**), and (iv) are either located at or flanked by regions of potential bendability (**Figure 4**). Considering the global A+T content (60.3%) of the pneumococcal R6 genome, these results reinforce the conclusion that *MgaSpn*, like H-NS, has a preference for AT-rich DNA sites. Most likely, a preference for AT-rich DNA regions rather than for specific DNA sequences is a general feature of the global regulators that constitute the Mga/AtxA family. Sequence alignments of all established Mga binding regions revealed that they exhibit only 13.4% identity. Furthermore, a mutational analysis in some target promoters indicated that Mga binds to DNA in a promoter-specific manner (Hause and McIver, 2012). In the case of AtxA, sequence similarities in its target



promoters are not apparent, and it has been shown that the promoter regions of several target genes are intrinsically curved (Hadjifrangiskou and Koehler, 2008).

Proteins of the H-NS family have a high degree of sequence similarity in the DNA-binding domain, as it is the case of H-NS and Ler (Shindo et al., 1995; Cordeiro et al., 2011). The three-dimensional structure of a complex between the DNA-binding domain of Ler and a 15-mer DNA duplex has been solved (Cordeiro et al., 2011). This structure revealed that the DNA-binding domain of Ler does not participate in base-specific contacts but recognizes specific structural features in the DNA minor groove. Thus, Ler, and likely other members of the H-NS family, recognizes specific DNA shapes. By DNase I footprinting experiments, we have found that H-NS recognizes three particular sites on the regulatory region of the pneumococcal *P1623B* promoter (Figure 1B). The three sites, ranging in size from 10 to 16 bp, are adjacent to regions of potential bendability, which agrees with the preference of H-NS for AT-rich DNA regions. Moreover, two out of the three H-NS binding sites occur in the *PB* activation region, which includes an *MgaSpn* binding site. Hence, the regulatory region of the pneumococcal *P1623B* promoter contains structural motifs that are recognized by H-NS.

In conclusion, two categories of protein-DNA interactions, namely those in which the protein recognizes unique chemical signatures of the DNA bases (base readout), and those in which the protein recognizes a sequence-dependent DNA shape (shape readout) were defined by Rohs et al. (2010). Our present work suggests that two unrelated DNA-binding proteins from phylogenetically distant bacteria are able to recognize similar structural characteristics in their DNA targets. It is intriguing

that unrelated bacterial species have evolved to encode proteins that seem to use a similar strategy to regulate the expression of a number of genes (silencing or activations). We take this as an indication of a successful strategy for proteins recognizing DNA regions that show intrinsic bendability/flexibility.

## AUTHOR CONTRIBUTIONS

VS and MH performed laboratory work. VS, MH, ME, AJ, and AB designed the study, performed data analysis and wrote the manuscript. All authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

This work was supported by grants CSD2008-00013-INTERMODS, BIO2013-49148-C2-1-R, BIO2013-49148-C2-2-R and BIO2015-69085-REDC from the Spanish Ministry of Economy and Competitiveness.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmolb.2016.00060>

**Supplementary Figure 1 | Binding of MgaSpn-His to the 290-bp *hlyC* DNA fragment. (A)** EMSA. The  $^{32}$ P-labeled *hlyC* DNA fragment (2 nM) was incubated with increasing concentrations of MgaSpn-His (20 to 180 nM). Free and bound DNAs were separated by native gel electrophoresis (5% polyacrylamide). Bands corresponding to free DNA (F) and to several protein-DNA complexes (C1, C2, C3, and C4) are indicated. **(B)** Affinity of MgaSpn-His for the 290-bp *hlyC* DNA fragment. The autoradiograph shown in A was scanned, and the percentage of free DNA was plotted against MgaSpn-His concentration.

## REFERENCES

- Carey, J. (1988). Gel retardation at low pH resolves trp repressor-DNA complexes for quantitative study. *Proc. Natl. Acad. Sci. U.S.A.* 85, 975–979. doi: 10.1073/pnas.85.4.975
- Cordeiro, T. N., Schmidt, H., Madrid, C., Juárez, A., Bernadó, P., Griesinger, C., et al. (2011). Indirect DNA readout by an H-NS related protein: structure of the DNA complex of the C-terminal domain of Ler. *PLoS Pathog.* 7:e1002380. doi: 10.1371/journal.ppat.1002380
- Dame, R. T., Wyman, C., and Goosen, N. (2000). H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Res.* 28, 3504–3510. doi: 10.1093/nar/28.18.3504
- Finn, R. D., Cogill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344
- Gabrielian, A., Vlahovicek, K., and Pongor, S. (1997). Distribution of sequence-dependent curvature in genomic DNA sequences. *FEBS Lett.* 406, 69–74. doi: 10.1016/S0014-5793(97)00236-6
- Godessart, N., Muñoz, F. J., Regué, M., and Juárez, A. (1988). Chromosomal mutations that increase the production of a plasmid-encoded haemolysin in *Escherichia coli*. *J. Gen. Microbiol.* 134, 2779–2787.
- Hadjifrangiskou, M., and Koehler, T. M. (2008). Intrinsic curvature associated with the coordinately regulated anthrax toxin gene promoters. *Microbiology* 154, 2501–2512. doi: 10.1099/mic.0.2007/016162-0
- Hammerstrom, T. G., Horton, L. B., Swick, M. C., Joachimiak, A., Osipiuk, J., and Koehler, T. M. (2015). Crystal structure of *Bacillus anthracis* virulence regulator AtxA and effects of phosphorylated histidines on multimerization and activity. *Mol. Microbiol.* 95, 426–441. doi: 10.1111/mmi.12867
- Hause, L. L., and McIver, K. S. (2012). Nucleotides critical for the interaction of the *Streptococcus pyogenes* Mga virulence regulator with Mga-regulated promoter sequences. *J. Bacteriol.* 194, 4904–4919. doi: 10.1128/JB.00809-12
- Hemsley, C., Joyce, E., Hava, D. L., Kawale, A., and Camilli, A. (2003). MgrA, an orthologue of Mga, acts as a transcriptional repressor of the genes within the *rlrA* pathogenicity islet in *Streptococcus pneumoniae*. *J. Bacteriol.* 185, 6640–6647. doi: 10.1128/JB.185.22.6640-6647.2003
- Hondorp, E. R., Hou, S. C., Hause, L. L., Gera, K., Lee, C.-E., and McIver, K. S. (2013). PTS phosphorylation of Mga modulates regulon expression and virulence in the group A streptococcus. *Mol. Microbiol.* 88, 1176–1193. doi: 10.1111/mmi.12250
- Hoskins, J., Alborn, W. E. Jr., Arnold, J., Blaszcak, L. C., Burgett, S., Dehoff, B. S., et al. (2001). Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J. Bacteriol.* 183, 5709–5717. doi: 10.1128/JB.183.19.5709-5717.2001
- Ko, M., and Park, C. (2000). H-NS-dependent regulation of flagellar synthesis is mediated by a LysR family protein. *J. Bacteriol.* 182, 4670–4672. doi: 10.1128/JB.182.16.4670-4672.2000
- Koronakis, V., and Hughes, C. (1988). Identification of the promoters directing *in vivo* expression of hemolysin genes in *Proteus vulgaris* and *Escherichia coli*. *Mol. Gen. Genet.* 213, 99–104. doi: 10.1007/BF00333404
- Lang, B., Blot, N., Bouffartigues, E., Buckle, M., Geertz, M., Gualerzi, C. O., et al. (2007). High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. *Nucleic Acids Res.* 35, 6330–6337. doi: 10.1093/nar/gkm712
- Lim, C. J., Lee, S. Y., Kenney, L. J., and Yan, J. (2012). Nucleoprotein filament formation is the structural basis for bacterial protein H-NS gene silencing. *Sci. Rep.* 2:509. doi: 10.1038/srep00509



- Liu, Y., Chen, H., Kenney, L. J., and Yan, J. (2010). A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes Dev.* 24, 339–344. doi: 10.1101/gad.1883510
- Madrid, C., Nieto, J. M., Paytubi, S., Falconi, M., Gualerzi, C. O., and Juárez, A. (2002). Temperature- and H-NS-dependent regulation of a plasmid-encoded virulence operon expressing *Escherichia coli* hemolysin. *J. Bacteriol.* 184, 5058–5066. doi: 10.1128/JB.184.18.5058-5066.2002
- Nieto, J. M., Carmona, M., Bolland, S., Jubete, Y., De La Cruz, F., and Juárez, A. (1991). The *hha* gene modulates haemolysin expression in *Escherichia coli*. *Mol. Microbiol.* 5, 1285–1293. doi: 10.1111/j.1365-2958.1991.tb01902.x
- Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., and Mann, R. S. (2010). Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* 79, 233–269. doi: 10.1146/annurev-biochem-060408-091030
- Ruiz-Cruz, S., Espinosa, M., Goldmann, O., and Bravo, A. (2016). Global regulation of gene expression by the MafR protein of *Enterococcus faecalis*. *Front. Microbiol.* 6:1521. doi: 10.3389/fmicb.2015.01521
- Ruiz-Cruz, S., Solano-Collado, V., Espinosa, M., and Bravo, A. (2010). Novel plasmid-based genetic tools for the study of promoters and terminators in *Streptococcus pneumoniae* and *Enterococcus faecalis*. *J. Microbiol. Methods* 83, 156–163. doi: 10.1016/j.mimet.2010.08.004
- Shin, M., Lagda, A. C., Lee, J. W., Bhat, A., Rhee, J. H., Kim, J.-S., et al. (2012). Gene silencing by H-NS from distal DNA site. *Mol. Microbiol.* 86, 707–719. doi: 10.1111/mmi.12012
- Shindo, H., Iwaki, T., Ieda, R., Kurumizaka, H., Ueguchi, C., Mizuno, T., et al. (1995). Solution structure of the DNA binding domain of a nucleoid-associated protein, H-NS, from *Escherichia coli*. *FEBS Lett.* 360, 125–131. doi: 10.1016/0014-5793(95)00079-O
- Solano-Collado, V., Espinosa, M., and Bravo, A. (2012). Activator role of the pneumococcal Mga-like virulence transcriptional regulator. *J. Bacteriol.* 194, 4197–4207. doi: 10.1128/JB.00536-12
- Solano-Collado, V., Lurz, R., Espinosa, M., and Bravo, A. (2013). The pneumococcal MgaS<sub>pn</sub> virulence transcriptional regulator generates multimeric complexes on linear double-stranded DNA. *Nucleic Acids Res.* 41, 6975–6991. doi: 10.1093/nar/gkt445
- Vlahovicek, K., Kaján, L., and Pongor, S. (2003). DNA analysis servers: plot.it, bend.it, model.it and IS. *Nucleic Acids Res.* 31, 3686–3687. doi: 10.1093/nar/gkg559
- Vogel, M., Hess, J., Then, I., Juárez, A., and Goebel, W. (1988). Characterization of a sequence (*hlyR*) which enhances synthesis and secretion of hemolysin in *Escherichia coli*. *Mol. Gen. Genet.* 212, 76–84. doi: 10.1007/BF00322447
- Winardhi, R. S., Yan, J., and Kenney, L. J. (2015). H-NS regulates gene expression and compacts the nucleoid: insights from single-molecule experiments. *Biophys. J.* 109, 1321–1329. doi: 10.1016/j.bpj.2015.08.016

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Solano-Collado, Hüttener, Espinosa, Juárez and Bravo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# An Interplay among FIS, H-NS, and Guanosine Tetraphosphate Modulates Transcription of the *Escherichia coli* *cspA* Gene under Physiological Growth Conditions

Anna Brandi, Mara Giangrossi, Anna M. Giuliadori and Maurizio Falconi\*

Laboratory of Genetics, School of Bioscience and Veterinary Medicine, University of Camerino, Camerino, Italy

## OPEN ACCESS

### Edited by:

Antonio Juárez,  
Universitat de Barcelona, Spain

### Reviewed by:

Eduard Torrents,  
Institute for Bioengineering of  
Catalonia, Spain  
Alicia Bravo,  
Centro de Investigaciones Biológicas,  
Consejo Superior de Investigaciones  
Científicas, Spain

### \*Correspondence:

Maurizio Falconi  
maurizio.falconi@unicam.it

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 29 January 2016

**Accepted:** 01 May 2016

**Published:** 24 May 2016

### Citation:

Brandi A, Giangrossi M, Giuliadori AM  
and Falconi M (2016) An Interplay  
among FIS, H-NS, and Guanosine  
Tetraphosphate Modulates  
Transcription of the *Escherichia coli*  
*cspA* Gene under Physiological  
Growth Conditions.  
Front. Mol. Biosci. 3:19.  
doi: 10.3389/fmolb.2016.00019

CspA, the most characterized member of the *csp* gene family of *Escherichia coli*, is highly expressed not only in response to cold stress, but also during the early phase of growth at 37°C. Here, we investigate at molecular level the antagonistic role played by the nucleoid proteins FIS and H-NS in the regulation of *cspA* expression under non-stress conditions. By means of both probing experiments and immunological detection, we demonstrate *in vitro* the existence of binding sites for these proteins on the *cspA* regulatory region, in which FIS and H-NS bind simultaneously to form composite DNA-protein complexes. While the *in vitro* promoter activity of *cspA* is stimulated by FIS and repressed by H-NS, a compensatory effect is observed when both proteins are added in the transcription assay. Consistently with these findings, inactivation of *fis* and *hns* genes reversely affect the *in vivo* amount of *cspA* mRNA. In addition, by means of strains expressing a high level of the alarmone guanosine tetraphosphate ((p)ppGpp) and *in vitro* transcription assays, we show that the *cspA* promoter is sensitive to (p)ppGpp inhibition. The (p)ppGpp-mediated expression of *fis* and *hns* genes is also analyzed, thus clarifying some aspects of the regulatory loop governing *cspA* transcription.

**Keywords:** *cspA* gene, FIS, H-NS, guanosine tetraphosphate, DNA-protein interaction, gene regulation in Bacteria, transcription

## INTRODUCTION

The *cspA* gene of *Escherichia coli* encodes a nucleic acid-binding protein of 70 amino acid residues (CspA) that interacts preferentially with single stranded RNA and DNA (Jiang et al., 1997; Bae et al., 2000). CspA belongs to the *csp* gene family, a group which includes in *E. coli* a total of nine paralogs, called from *cspA* to *cspI*. CspA is known as the “major cold-shock protein” (Goldstein et al., 1990; Jones and Inouye, 1994) by virtue of the original observation that cold-shock induces its synthesis *ex novo*. However, CspA is also synthesized at 37°C, particularly during the early phase of growth (Brandi et al., 1999; Brandi and Pon, 2012). In addition to *cspA*, also *cspB*, *cspE*, *cspG*, *cspI* are cold shock inducible, unlike *cspD*, *cspC*, *cspF* and *cspH* which are expressed only at 37°C. In particular, *cspD* is expressed exclusively during stationary phase or nutritional stress, *cspC* is constitutively synthesized at 37°C, and *cspF* and *cspH*, the most distant related genes, remain to be characterized (Jones and Inouye, 1994; Yamanaka et al., 1998; Ermolenko and Makhataadze, 2002).

During cold-shock, CspA was found to affect both transcription (La Teana et al., 1991; Jones et al., 1992b) and translation (Brandi et al., 1996; Giuliadori et al., 2004) of other genes and it was also suggested to function as an RNA chaperone (Jiang et al., 1997; Bae et al., 2000). On the other hand, little is known about the role played by CspA at 37°C. The expression of *cspA* displays a multilevel regulation (transcription, mRNA stability and translation) modulated by multiple factors that may differently contribute to ensure a rapid and precise response to nutritional or environmental changes (Gualerzi et al., 2011). Interestingly, the regulation of *cspA* expression is different under stress and non-stress conditions. In fact, while the elevated production of CspA following a temperature down-shift is due mainly to post-transcriptional events (i.e., increased stability and preferential translation of *cspA* mRNA) rather than to the transcriptional stimulation of *cspA* promoter (Brandi et al., 1996; Goldenberg et al., 1996; Giuliadori et al., 2010), the synthesis of CspA at 37°C is prevalently regulated at transcriptional level (Brandi et al., 1999; Brandi and Pon, 2012).

Bacteria contain a heterogeneous group of polypeptides, collectively known as nucleoid-associated proteins (NAPs), that are cataloged as a family on the basis of functional similarities. These proteins bind to nucleic acids, are basic and have low molecular mass (Azam and Ishihama, 1999). In addition to their common architectural role in the organization of bacterial chromosome, they are able to modulate transcription initiation and to control DNA replication, segregation and repair (Browning et al., 2010; Dillon and Dorman, 2010; Rimsky and Travers, 2011; Wang et al., 2011). H-NS, one of the most abundant NAPs, preferentially binds to tracts of intrinsically curved DNA (A/T-rich sequences) and/or actively induces bending (Yamada et al., 1990; Spurio et al., 1997; Gordon et al., 2011). Thus, by interacting with curved DNA, often found in upstream promoter regions, this nucleoid protein, referred to as a “universal repressor,” silences transcription of its target genes (Atlung and Ingmer, 1997; Hommais et al., 2001; Dorman, 2004, 2007; Bouffartigues et al., 2007; Lang et al., 2007). FIS (Factor for Inversion Stimulation) is another NAP that binds to DNA and modulates the topology of DNA in a growth-phase dependent manner (Schneider et al., 1997; Muskhelishvili and Travers, 2003). Unlike H-NS, FIS is a positive regulator activating transcription of genes and operons associated with primary metabolism as stable RNAs (Ross et al., 1990; Gonzalez-Gil et al., 1996). Thus, H-NS and FIS, through direct and indirect effects, control the expression of a large number of genes and are viewed as global regulators of transcription in response to growth phase and environmental changes (reviewed in Kahramanoglou et al., 2011).

Since *cspA* belongs to the set of genes controlled by FIS and H-NS under non-stress conditions (Brandi et al., 1999), the prime aim of this study was to elucidate the molecular aspects of *cspA* regulation by these two NAPs expanding and deepening our previous knowledge. Our results demonstrate the existence of a functional interplay between FIS and H-NS, which are able to bind separately or simultaneously the *cspA* promoter region. The composition of the DNA-protein complexes thus formed has a different impact on the transcription of *cspA*, that is inhibited by

H-NS, stimulated by FIS, and unaltered when both factors are present, in fully agreement with the *in vivo* data (Brandi et al., 1999). Finally, seeking further factors which could participate to this regulatory circuit, we found that *cspA* promoter is sensitive to (p)ppGpp, the mediator molecule of stringent response.

## MATERIALS AND METHODS

### Strains

*E. coli* strains used in this study were: MRE600 (*F*<sup>-</sup>, *rna*) (Cammack and Wade, 1965); DH5α (Sambrook and Russell, 2001); WM2482 (correspond to MG1655 reference strain K-12, *F*<sup>-</sup>  $\lambda$ <sup>-</sup> *ilvG*<sup>-</sup> *rfb-50* *rph-1*), WM2648 (MG1655 *hns::hyg*), WM2649 (MG1655 *fis::kan*) and WM 2650 (*hns/fis* double mutant of MG1655) a kind gift of Walter Messer's laboratory (Berlin) (Afflerbach et al., 1998); *E. coli* KT793 carrying IPTG-inducible *relA* in pKT31 (Tedin et al., 1995). Cells were grown at 37°C in Luria-Bertani broth or in M9 minimal medium supplemented with 0.4% glucose (Sambrook and Russell, 2001) or in “Phosphates-free” medium (100 mM Tris-HCl, pH 7.7, 0.5% glucose, 0.5% peptone, 10 mM NH<sub>4</sub>Cl, 0.7 mM NaNO<sub>3</sub>, 1 mM Na<sub>2</sub>SO<sub>4</sub>, 0.5 mM MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.05 mM MnCl<sub>2</sub>·4H<sub>2</sub>O, 0.02 mM FeSO<sub>4</sub>·7H<sub>2</sub>O) where indicated.

### DNA Manipulations and General Procedures

The plasmid pTZ310 was constructed by cloning the HpaII-HpaII fragment, containing the *cspA* promoter region (from pos. -145 to pos. +165), into the AccI site of the pTZ19R polylinker. H-NS was purified as described in Falconi et al. (1988); FIS was a kind gift from the laboratory of Regine Kahmann. DNA isolation, agarose gel electrophoresis, polymerase chain reaction and other DNA manipulations were performed according to standard procedures (Sambrook and Russell, 2001). Radioactivity associated to DNA or RNA was detected and quantified by Molecular Imager (Bio-Rad, model FX).

### Northern Blot Analysis

Total RNA was purified by hot phenol extraction from cells harvested at the indicated times and levels of individual mRNAs were detected by Northern blots probed with specific 5'-end-labeled oligonucleotides (Brandi et al., 1999). The hybridization was performed in the range of temperature 37–48°C, depending on the oligonucleotide used. The oligonucleotides used as specific probes are: 5'-CTTTCGATGGTGAAGGACACT-3' for *cspA*; 5'-GCGCACGAAGAGTACGG-3' for *hns* and 5'-CAGGGGTTTTTGGGTACCT-3' for *fis*.

### Electrophoretic Mobility Shift Assay (EMSA)

The 310 bp DNA fragment, excised from pTZ310 by BamHI/HindIII digestion, was end-labeled with  $\alpha$ -[<sup>32</sup>P]-dATP by fill-in reaction using the DNA polymerase Klenow fragment. About 5–10 ng of the radioactive DNA fragment were incubated with the indicated amounts of purified FIS and H-NS at 25°C in a reaction mixture (15  $\mu$ l) containing 10 mM Tris HCl (pH 8), 10 mM MgCl<sub>2</sub>, 100 mM NaCl, 10 mM KCl, 1 mM

spermidine, 0.5 mM dithiothreitol, 5% glycerol, 0.08 mg ml<sup>-1</sup> BSA, and 50 ng Poly dI-dC as competitor DNA. After 15 min of incubation, samples were subjected to electrophoresis on 6% polyacrylamide gel in TBE buffer (Sambrook and Russell, 2001).

The combined EMSA-Western blot analysis was carried out essentially as described above except that each reaction mix contained 120 ng of a cold DNA fragment corresponding to *cspA* promoter. This fragment (340 bp) was obtained by PCR amplification using the primer pairs 5'-CAACCCGGCATTAAGTAAGC-3' and 5'-CCATTTTACGATACCAGTCA-3' on a 1200 bp DNA fragment cloned in pTZ19R (Brandi et al., 1996). Samples were loaded in duplicate on 6% polyacrylamide gel which was subsequently electro-transferred (35 min at 2.5 mA/cm<sup>2</sup>) into a nitrocellulose membrane. The filter was divided into the two duplicates: one half was incubated with polyclonal antibodies anti-FIS and the other half with antibodies anti-H-NS. Finally, proteins detection was carried out using Alkaline phosphatase Conjugated anti-rabbit IgG and NBT/BCIP as substrates.

## DNase I Footprinting

The DNA fragment used in footprints was excised from pTZ310 with BamHI/PstI or HindIII/SmaI and end-labeled by fill-in reaction with  $\alpha$ -[<sup>32</sup>P]-dATP using Klenow fragment of DNA polymerase. The radioactive DNA was pre-incubated with the indicated amounts of FIS or/and H-NS for 20 min at 25°C in 30  $\mu$ l of *in vitro* transcription buffer (see below). After addition of 15 ng of DNase I, the reaction was prolonged for further 45 s and then stopped on ice with 1.5  $\mu$ l of 0.5 M EDTA (pH 8) and 10  $\mu$ l of 10 M NH<sub>4</sub> acetate (pH 7.3). Partially digested DNA was ethanol precipitated in presence of 1  $\mu$ g of tRNA as carrier and then loaded on a 7% polyacrylamide-urea gel (Sambrook and Russell, 2001).

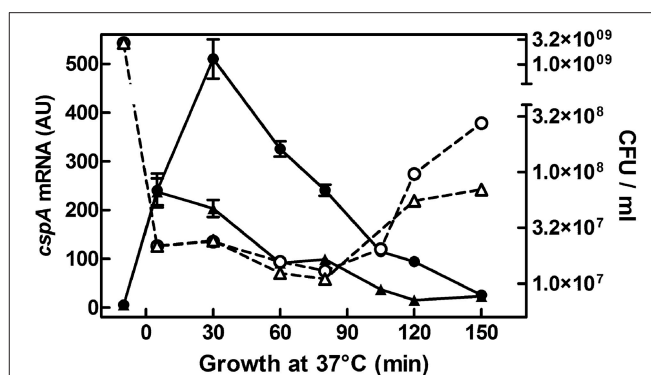
## In vitro Transcription

*In vitro* transcription assays were programmed with pKK*cspA310::cat*, a pKK232-8 derivative, carrying a 310 bp DNA fragment of *cspA* promoter region (from pos. -145 to pos. +165; Goldenberg et al., 1997). The reactions, carried out at 37°C in 25  $\mu$ l of transcription buffer (10 mM Tris HCl, pH 8, 10 mM MgCl<sub>2</sub>, 100 mM NaCl, 2 mM spermidine, 2 mM dithiothreitol, 0.1 mg ml<sup>-1</sup> BSA), contain 0.15 U of *E. coli* RNA Polymerase (USB), 0.5 mM of each NTP, 5 U of human placental ribonuclease inhibitor and 100 ng of DNA template. At the indicated times, the reaction was stopped with 1.5  $\mu$ l of 0.5 M EDTA (pH 8) and 10  $\mu$ l of 10 M NH<sub>4</sub> acetate (pH 7.5) and mRNA ethanol precipitated. The amount of the *cat* reporter gene transcribed *in vitro* was determined by Northern blotting, probed with a <sup>32</sup>P-labeled *cat* fragment and quantified by Molecular Imager (Bio-Rad, model FX).

## RESULTS

### The Cold-Shock *cspA* Gene Is Expressed at 37°C

In previous studies, we have shown that *cspA* is highly expressed not only during cold-shock but also under non-stress conditions



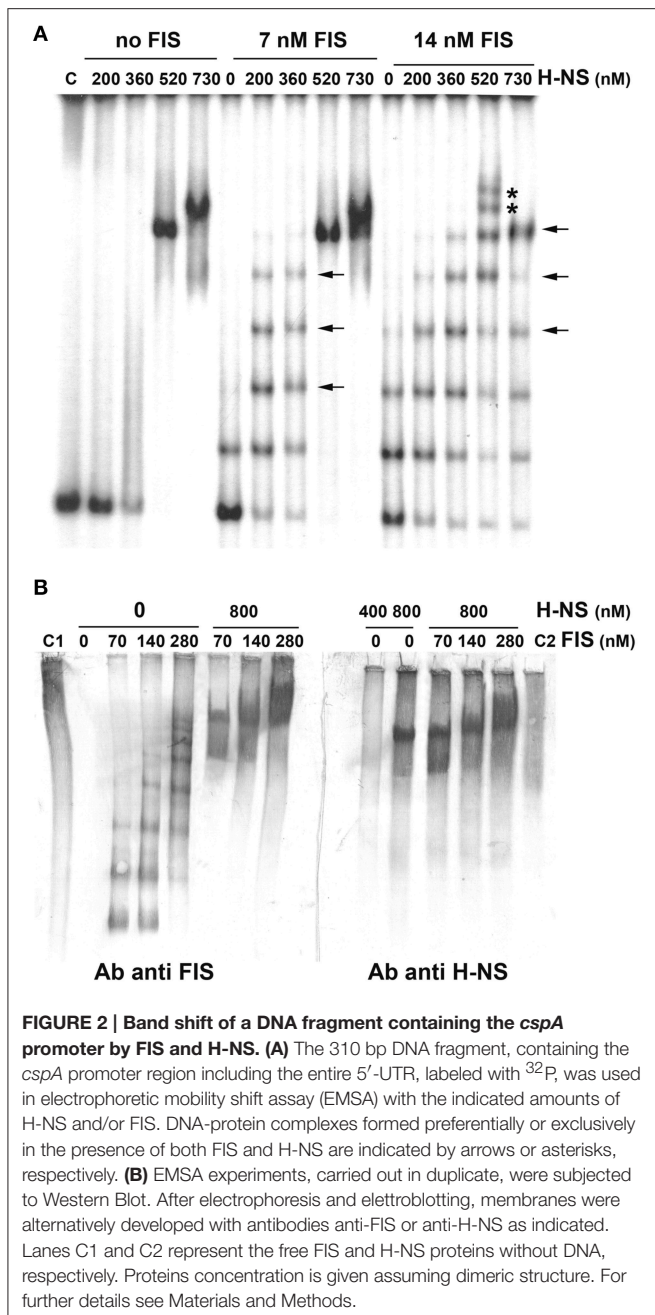
**FIGURE 1 | *cspA* mRNA steady state level following growth resumption at 37°C.** An overnight culture of *E. coli* MRE600 grown at 37°C in M9 medium was diluted (1:100) in parallel with LB (●) or with M9 (▲) fresh medium and cell aliquots were harvested for RNA extraction at the indicated time after resumption from stationary phase. About 6  $\mu$ g of total RNA were subjected to Northern analysis. The cellular level of *cspA* transcripts were evaluated by imager quantification of the radioactivity associated with the mRNA and expressed as Arbitrary Units (AU). Data represent the average of at least two independent experiments and standard deviation is indicated. The profile of Colony Forming Units (CFU) during growth in rich (LB, ○) or minimal (M9, △) medium as a function of time after nutritional up-shift is also reported. Further details are provided in Materials and Methods.

when cells grow at 37°C (Brandi et al., 1999; Brandi and Pon, 2012). Here, to understand how the physiological state of the cell could affect the expression of *cspA*, we monitored the *cspA* mRNA levels in cells escaping from stationary phase as a function of the availability of nutrients. To accomplish this goal, an overnight culture of *E. coli* grown at 37°C was diluted in rich (LB) or minimal (M9) fresh medium and bulk RNA, extracted from cells at increasing times after the nutritional up-shift, was used for Northern analysis (Figure 1). Transcription of *cspA* is promptly induced upon cell dilution, in both media, albeit to different extents (~40 fold in M9 and ~100 fold in LB). Furthermore, the level of transcript augments within the initial 80 min, a time preceding the first bacterial division as denoted by the constant number of cells (CFU), and drops off to the initial level after the cells start dividing. These observations are confirmed by detecting the *cspA* mRNA by RT-qPCR in cells growing in LB (Figure S6C). In light of this finding, we focused our study on those factors, likely affecting *cspA* regulation at 37°C, that are known to couple gene expression to growth conditions.

### Interaction of H-NS and FIS with the *cspA* Promoter Region

Although, FIS and H-NS have been shown to influence *cspA* expression (Brandi et al., 1999), so far the evidence of a direct binding of these proteins to the promoter region of the target gene was lacking. Thus, we investigated the interaction of these two NAPs with the *cspA* DNA by electrophoretic mobility shift assay (EMSA). For this purpose, a fragment of 310 bp spanning from position -145 to position +165, was incubated with H-NS, FIS or with a mixture containing both proteins at different molar ratios. As seen in Figure 2A, when tested individually, H-NS produces an EMSA pattern typical of an all-or-none response,





suggesting that this protein binds in a co-operative manner to the 310 bp DNA fragment containing the *cspA* promoter as described for other genetic systems (Falconi et al., 1993, 1998; Madrid et al., 2002; Giangrossi et al., 2005; Ulissi et al., 2014). In the absence of FIS, H-NS has little or no effect below the critical concentration of 360 nM, whereas it forms a stable nucleoprotein complex at 520 nM (Figure 2A). The addition of more H-NS (730 nM) causes the appearance of a new band with reduced mobility, suggesting that, at the maximum concentration tested, H-NS oligomerizes occupying all high and low affinity sites presents on this DNA fragment. Unlike H-NS, a discrete retardation band is detected even at very low FIS concentrations (7 nM) and

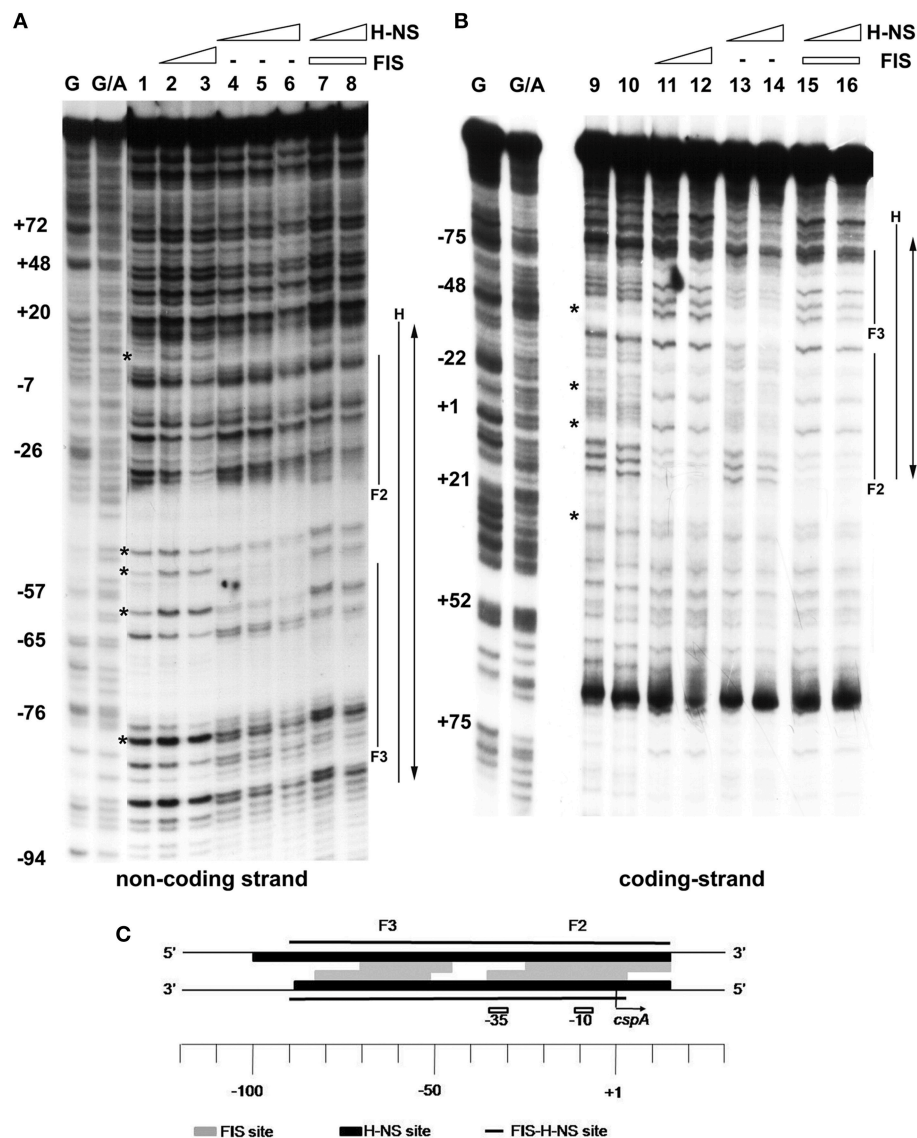
additional bands with progressively lowered mobility appear as a function of increasing FIS concentrations (14 nM in Figure 2A, 28 and 56 nM in Figure S1). This pattern suggests that the 310 bp DNA fragment contains multiple sites for which FIS displays different affinities and that are saturated in a concentration-dependent manner by this protein. Furthermore, compared to H-NS, a relative little amount of FIS is sufficient to occupy, at least partially, all sites.

Interestingly, the addition of 200–300 nM of H-NS to low concentrations of FIS produces retarded bands (indicated by horizontal arrows in Figure 2A, Figure S1) with a mobility similar to those bands present when FIS alone is added at higher concentrations. Nevertheless, when both H-NS and FIS are added at a certain concentration ratio, additional bands (indicated by asterisks in Figure 2A, Figure S1), not found in the individual H-NS and FIS patterns, appear. On the other hand, when FIS and H-NS are added at low ( $\leq 14$  nM) and high ( $> 500$  nM) concentrations, respectively, the most retarded complex seems to prevalently contain H-NS, since it displays a mobility similar to that obtained with H-NS alone at 520 nM (Figure 2A). All together, these observations indicate that FIS and H-NS might simultaneously bind to the same DNA molecule (the 310 bp DNA fragment) when are present in a given range of concentration ratios.

To verify this hypothesis, we carried out a band shift assay in which the DNA fragment was not radioactive and therefore nucleoprotein complexes were immunodetected using antibodies against FIS or H-NS (Figure 2B). Under the native conditions used for electrophoresis, unbound FIS and H-NS appear diffused throughout the lane (control samples C1 and C2) whereas discrete retarded bands are visible only in the presence of DNA. As expected, when FIS and H-NS are individually tested, DNA-protein complexes visualized by antibodies are superimposable to those obtained with labeled DNA (compare panels B and A of Figure 2). Nevertheless, when FIS (70–280 nM) and H-NS (800 nM) are combined, the same DNA-protein aggregates are revealed using either anti-FIS or anti-H-NS antibodies (Figure 2B). The control experiment shown in Figure S2 rules out the possibility that this result could be due to a cross-reaction between anti-FIS and H-NS or anti-H-NS and FIS. Therefore, all together, these data strongly suggest that both proteins can simultaneously interact with the 310 bp DNA fragment containing the *cspA* promoter to originate miscellaneous complexes.

## Identification of FIS and H-NS Binding Sites on *cspA* Promoter Region

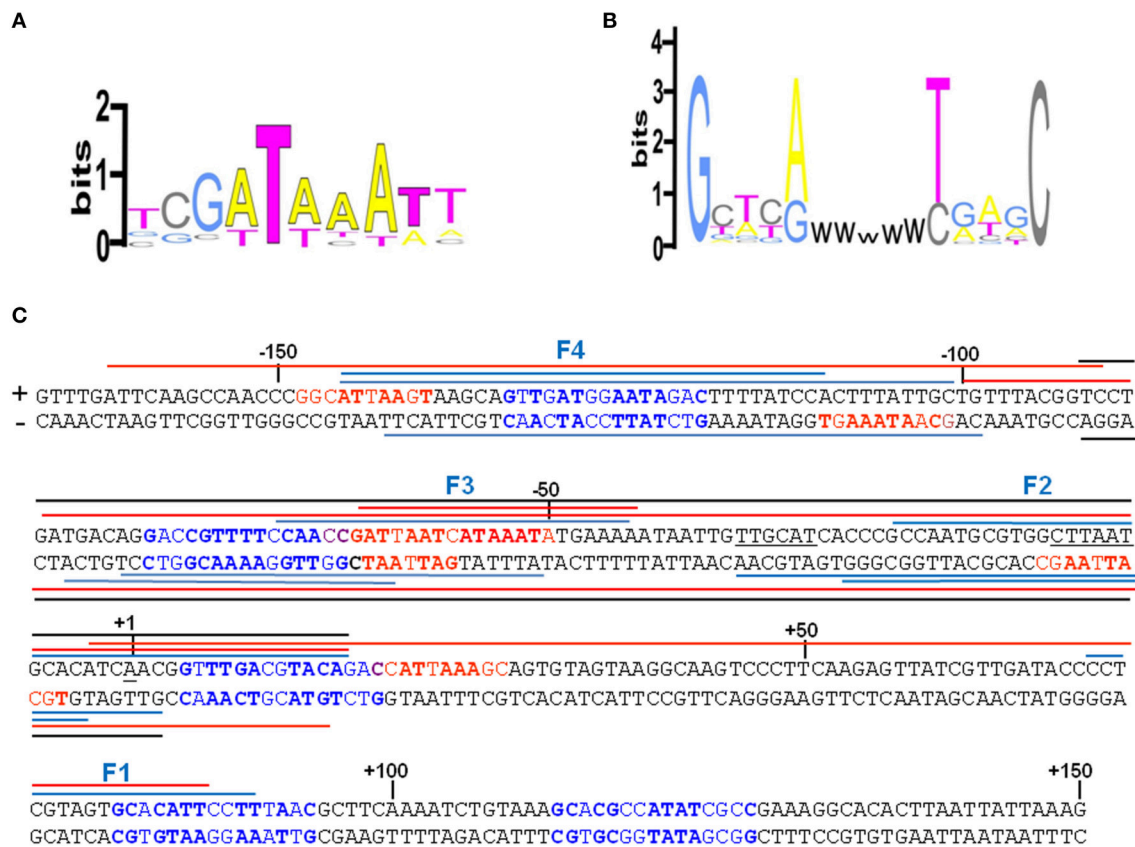
EMSA experiments prompted us to localize FIS and H-NS binding sites on *cspA* promoter, in an attempt to characterize also the nature of the complexes containing both proteins. To this end, we carried out DNase I footprints and compared the digestion patterns obtained with single proteins to that observed with a mixture of FIS and H-NS (Figures 3A,B). When individually tested, FIS interacts with two sites of  $\sim 35$  bp in length centered at positions  $-10$  (site 2) and  $-60$  (site 3) producing typical hypersensitive points, while H-NS covers a



**FIGURE 3 | Mapping of FIS and H-NS binding sites on *cspA* promoter region by DNase I footprinting.** Footprinting analysis was carried out on both non-coding strand (A) and coding strand (B) of a 310 bp DNA fragment corresponding to *cspA* promoter in the presence of the following FIS and HNS dimeric concentrations: no proteins, lanes 1, 9, and 10; 278 nM FIS, lanes 2 and 11; 555 nM FIS, lanes 3 and 12; 93 nM H-NS, lanes 4 and 13; 185 nM H-NS, lanes 5 and 14; 277 nM H-NS, lane 6; 278 nM FIS and 93 nM H-NS, lanes 7 and 15; 278 nM FIS and 185 nM H-NS, lanes 8 and 16. FIS and H-NS sites are indicated by solid lines while the double-headed arrows show protections resulting from the concomitant bond of FIS and H-NS. Sites hypersensitive to DNase I due to FIS-DNA interaction are indicated by asterisk. G and G+A represent the Maxam and Gilbert sequencing reactions. Localization of FIS (gray boxes), H-NS (black boxes), and FIS-H-NS (solid line) binding sites are schematically indicated on *cspA* promoter region (C).

fairly wide region (~100 nucleotides) centered at position -40. When mixed together FIS and H-NS cover all the available sites, giving rise to an extended protection spanning from position -90 to position +20 on *cspA* promoter. Remarkably, the DNase I digestion pattern observed at this cumulative site is essentially a merge of protected and hypersensitive positions characteristic of FIS and H-NS individual sites. Therefore, although FIS and H-NS protections are almost completely overlapping on both DNA strands (see scheme in Figure 3C), the two proteins apparently do not compete for binding to the same target DNA.

An extensive scanning of the regions adjacent to the *cspA* minimum promoter (Figures S3, S4) reveals the existence of other positions recognized by these nucleoid proteins. FIS covers two other distinct sites numbered F1 and F4 and centered at positions +32 and -120, respectively, while H-NS extends its protection to basically the entire *cspA* promoter due to its extensive oligomerization (Spurio et al., 1997; Badaut et al., 2002; Stella et al., 2005; Giangrossi et al., 2014). According to this premise, we found several *in silico* predicted binding sites for H-NS (Figure 4A) that are all imperfect fits to its consensus sequences (Lang et al., 2007) and can be considered as



**FIGURE 4 | Localization of FIS and H-NS binding sites found within the *cspA* promoter region. (A)** Logo representation of the H-NS binding motif (Lang et al., 2007). **(B)** Simplified version of the Logo for the FIS binding motif proposed by Shao et al. (2008). W = A or T. **(C)** The sequence of the promoter region of *cspA* (–167 to +150) is shown, with the transcriptional start point (+1), the –10 and –35 consensus promoter elements underlined in black. The bases in red represent the *in silico* predicted H-NS binding sites matching at least 4/6 bases of the core sequence of the H-NS binding motif and displaying less frequent bases in the other positions, while the bases in blue represent the *in silico* predicted FIS binding sites according to the FIS logo shown in (B). The bases in bold match those found more frequently in the H-NS and FIS binding sites, respectively. The colored lines placed above or below the + and – strand, respectively, indicate the protections from DNase I digestion found with H-NS alone (red), FIS alone (blue), or with a mixture of FIS and H-NS (black). Overlapping lines indicate DNase I footprintings found in independent experiments (Figure 3, Figures S3, S4 in Supplementary Material and other not shown). The protections were identified using either a 310 bp *cspA* fragment (for FIS, H-NS, and FIS+H-NS) or a 1200 bp *cspA* fragment (for FIS and H-NS alone).

nucleation sites where the protein initially binds before spreading to adjacent DNA tracts on *cspA* sequence. The same *in silico* approach allowed us to identify also five potential FIS binding sites matching the FIS Logo (Figure 4B) proposed by Shao et al. (2008), four of which fall in the regions shielded or exposed to DNase I cleavage by FIS. The overall H-NS and FIS protections on the sequence of *cspA* regulatory region along with their *in silico* predicted binding sites are summarized in Figure 4C.

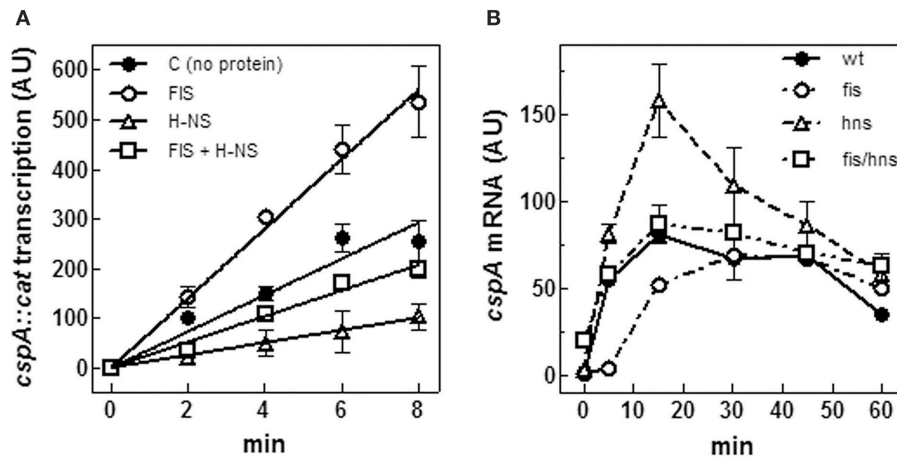
## Transcription of *cspA* Is Modulated by FIS, H-NS, and (p)ppGpp

In a previous paper, we provided evidence of a functional antagonism between FIS and H-NS on *cspA* expression (Brandi et al., 1999). This observation is consistent with the location, reported here, of H-NS and FIS binding sites, extending over the whole promoter region of *cspA*. Concerning the role of these two NAPs, while the inhibitory action of H-NS is commonly accepted, the function of FIS is still a point of debate since contradictory results have been reported about this protein that was also found

to negatively regulate *cspA* (Yamanaka and Inouye, 2001). Thus, to address this issue, we assayed H-NS and FIS for their capability to affect *cspA* transcription *in vitro*. As seen in Figure 5A, the activity of *cspA* promoter is stimulated by FIS and repressed by H-NS, totally confirming our previous data. Under the experimental conditions used, the extents of FIS stimulation and H-NS inhibition are similar (~3-fold) as compared to transcription carried out in the absence of proteins. Consistently, when FIS and H-NS are added together, their opposed effects on transcription neutralize each other, thus restoring the basal activity of *cspA* promoter (compare C and FIS+H-NS curves in Figure 5A).

The effect of H-NS and FIS on *cspA* expression was also studied *in vivo*. To this end, the steady state level of *cspA* mRNA was estimated at 37°C upon resumption from stationary phase of growth in wt, *hns*- and *fis*- strains and in *hns/fis* double deletion mutant. In agreement with Brandi et al. (1999) and Brandi and Pon (2012) in wt cells *cspA* expression is very high in early exponential growth and then progressively declines (Figure 5B). Interestingly, the lack of FIS causes a





**FIGURE 5 | FIS and H-NS modulate *cspA* transcription. (A)** *In vitro* transcription assay programmed with pKKcspA310::cat. DNA template was pre-incubated at 37°C in the absence of proteins (●), with 50 nM FIS (○), 375 nM H-NS (Δ) and with both FIS (50 nM) and H-NS (375 nM) (□) as dimers. The reaction was started by adding 0.15 units of RNA polymerase and the incubation was prolonged for the indicated times at 37°C, as described in Materials and Methods. **(B)** Steady-state levels of *CspA* mRNA were determined in wt *E. coli* cells (●), a *fis* null allele (○), an *hns* null allele (Δ), and a double mutant *fis-hns* (□). After a 10-fold dilution with fresh medium of saturated cultures grown in LB at 37°C, total RNA was extracted at the indicated times and subjected (10 μg) to Northern analysis. The radioactivity associated with *cspA* mRNA, normalized for the corresponding amounts of 16S rRNA, was quantified by Molecular Imager (Bio-Rad) and expressed as Arbitrary Units (AU). Data represent the average of at least two independent experiments and standard deviation is indicated.

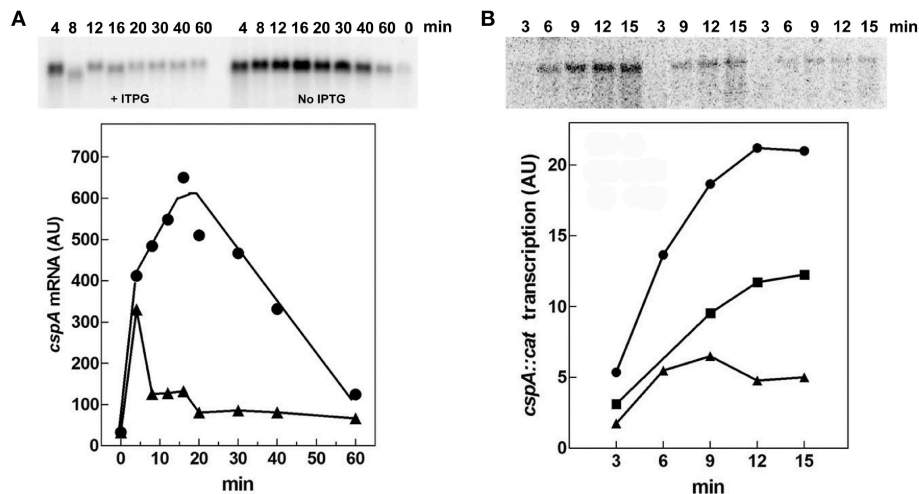
reduction of *cspA* transcript as compared to the wt within the initial 15 min, a time window that usually precedes the first cell division. However, at later stages, wt and *fis*- cells show comparable amounts of *cspA* mRNA (Figure 5B). On the contrary, inactivation of *hns* gene induces an increase of the *cspA* mRNA level that almost doubles in the first 20 min of growth with respect to the wt condition, and then declines. Finally, in agreement with the *in vitro* transcription assay (Figure 5A), the concomitant absence of FIS and H-NS results in a compensatory effect and ultimately has no significant consequences on the level of *cspA* mRNA (Figure 5B). Taken together, *in vitro* and *in vivo* data are fully consistent with each other and strongly indicate that H-NS and FIS, acting as negative and positive regulators, respectively, play an opposed role in modulating *cspA* transcription.

The alarmone guanosine tetraphosphate ((p)ppGpp) is a global regulator which is produced in most circumstances and modulates bacterial physiology (Hauryliuk et al., 2015). This small effector molecule is known to influence the expression of many genes thereby coupling the overall level of transcription to growth-rate (Potrykus and Cashel, 2008). Furthermore, overproduction of this unusual nucleotide prior to cold-shock was reported to lower the induction of most cold-shock genes, including *cspA* (Jones et al., 1992a).

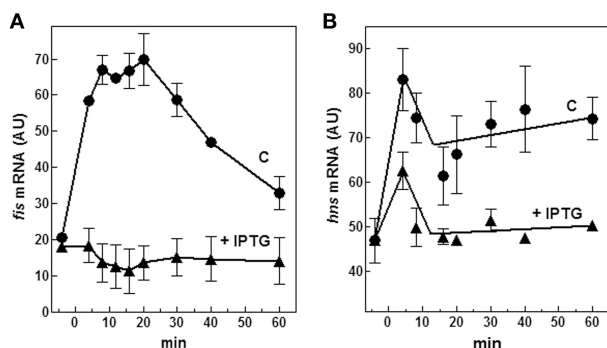
Therefore, we evaluated both *in vivo* and *in vitro* whether *cspA* promoter could respond to (p)ppGpp regulation also at 37°C (Figure 6). First of all, the intracellular (p)ppGpp level was artificially increased by IPTG induction of extra-chromosomal copies of *relA* gene placed under the control of the *lacUV5* promoter in plasmid pTK31 (see Materials and Methods). The induction of (p)ppGpp synthesis from pTK31 was verified (not shown) by thin layer chromatography as previously described (Sarubbi et al., 1989). As expected, when

cells in stationary phase were subjected to a nutritional up-shift, a sudden burst of *cspA* transcription was observed. By contrast, when the fresh medium was supplemented with IPTG, the high levels of the unusual nucleotide in induced cells significantly counteracted the characteristic promoter activation resulting in a substantial reduction of *cspA* messenger (Figure 6A). Thus, hypothesizing a direct action of (p)ppGpp, we investigated the effect of this molecule on *cspA* promoter activity in an *in vitro* purified system, following *cspA* transcription as a function of increasing reaction times. This experiment demonstrates that the level of *cspA* mRNA is decreased 2- and 4-fold in the presence of 200 and 400 μM of (p)ppGpp, respectively, as compared to the control curve obtained in the absence of the regulatory nucleotide (Figure 6B). According to the finding that (p)ppGpp-dependent inhibition of transcription of sensitive promoters results from the competition between the mediator molecule and NTPs substrates for access to the active center of RNA polymerase (Jöres and Wagner, 2003), the use of higher concentrations of NTPs (0.5 mM) alleviates the negative action of (p)ppGpp on *cspA* transcription (Figure S5). Altogether these results indicate that the stimulation of *cspA* expression, observed in early exponential growth at 37°C, is almost completely prevented by high levels of (p)ppGpp and that this effect reasonably resides on the ability of this molecule to directly repress the synthesis of mRNA from *cspA* promoter. Finally, to better understand the three components (FIS, H-NS, and (p)ppGpp) regulatory loop governing *cspA* expression, we analyzed, under our experimental conditions (cells carrying the pTK31 vector) *fis* and *hns* transcription as a function of increased intracellular concentrations of (p)ppGpp. As seen in Figure 7, activation of *fis* and *hns* promoters consequent to cell resumption from stationary phase is completely (panel A) and partially (panel B) abolished,





**FIGURE 6 | (p)ppGpp negatively affects *cspA* transcription. (A)** The steady state levels of *cspA* transcript were determined during growth in Phosphates-free medium at 37°C in cells overproducing or not (p)ppGpp. Total RNA was extracted from cells in stationary phase (time zero) and at the indicated times after dilution with fresh medium alone (●) or supplemented with 400 μM IPTG (▲). The levels of *cspA* mRNA were estimated by Northern blotting analysis. **(B)** Effect of (p)ppGpp on the *in vitro* activity of *cspA* promoter. The transcriptional activity was estimated in the absence (●) or in presence of 200 (■) and 400 μM (▲) of (p)ppGpp. Reactions and processing of samples were performed as described in Materials and Methods. The supercoiled plasmid pKK310::cat used as DNA template was pre-incubated with (p)ppGpp and RNA polymerase for 5 min at 37°C. Then reactions were started by adding NTPs at a final concentration of 100 μM each. At the indicated incubation times, aliquots were withdrawn and transcription stopped with EDTA (f.c. 30 mM). Quantization of the Northern blots (upper panels), expressed as Arbitrary Units (AU), is plotted as a function of time (lower panels).



**FIGURE 7 | Effect of (p)ppGpp on *fis* and *hns* expression. Levels of *fis* (A) and *hns* (B) mRNAs were estimated by Northern blotting analysis of total RNA extracted from *E. coli* cells growing at 37°C and carrying the plasmid pTK31. Cells were collected in stationary phase (time zero) and at the indicated times after dilution with fresh medium. As described in Figure 6A, the control culture was grown in the absence of IPTG (▲) whereas the induced culture contained 400 μM IPTG (●) to activate the *lacUV5::relA* gene. Data represent the average of at least two independent experiments and standard deviation is indicated.**

respectively, by an elevated level of (p)ppGpp suggesting that transcription of both genes is negatively regulated by this alarmone.

## DISCUSSION

There is evidence that regulation of many bacterial genes is based on structural/functional interplays between two or more

nucleoid-associated proteins which may play synergetic and/or antagonistic roles. In particular, FIS and H-NS cooperate to regulate several unrelated genetic systems in different bacteria. Some examples are *rrnB* (Afflerbach et al., 1999), *oriC* (Roth et al., 1994), *nir* (Browning et al., 2000), *dps* (Grainger et al., 2008), *pel* (Ouafa et al., 2012) as well as those systems that we have contributed to characterize, such as the FIS-H-NS-mediated regulation of the *E. coli hns* gene itself (Falconi et al., 1996), or the virulence gene *virF* of *Shigella flexneri* (Falconi et al., 2001).

In this study, in addition to confirming our earlier observations on the involvement of FIS and H-NS in controlling *cspA* (Brandi et al., 1999), we dissected their interplay at *cspA* promoter. Furthermore, we identified another factor, the (p)ppGpp, participating in this regulation, thus expanding our knowledge of the regulatory circuit governing the expression of *cspA* at 37°C.

Firstly, through both EMSA and footprinting assays, we have demonstrated the direct interaction of FIS and H-NS with the promoter region of *cspA* and identified their target sites as summarized in Figure 4. The appearance of discrete bands with shifted mobility indicates that FIS binds at least five sites on the *cspA* promoter region spanning from position -167 to position +150. This interaction is also confirmed by the FIS-dependent protections from DNase I cleavage of four discrete stretches of nucleotides (~30–40 bp). Notably, all these protected targets overlap a degenerated FIS consensus sequence predicted *in silico* using a software that allows identification of sequence motifs (Hu et al., 2003) based on the FIS binding site logo (Shao et al., 2008). With regard to H-NS, we found six sequences on *cspA* DNA partially matching the H-NS binding

motif as identified by Lang et al. (2007). These sequences may represent high-affinity nucleation sites where H-NS initially binds before oligomerizing along DNA and interacting with adjacent lower affinity sites. Such binding property accounts for the extended regions protected by H-NS in footprinting experiments (Figure 3, Figure S4) and ultimately determines the complete coverage of *cspA* promoter region including the four mapped FIS sites (F1–F4). Both the standard EMSA and a modified EMSA, in which the DNA-protein complexes were localized by immunodetection (Figure 2), reveal that FIS and H-NS, instead of structurally competing for the binding to the DNA, at certain concentration ratios can contact simultaneously the region containing the *cspA* promoter each one recognizing their own targets. The formation of high-order aggregates containing both proteins is also evidenced by the occurrence of a single merged protection observed in footprints carried out with FIS and H-NS together. The coexistence of these two proteins is likely due to their different binding properties: while FIS interacts with the major groove (Osuna et al., 1991), H-NS contacts the minor groove of the DNA, as demonstrated by the finding that two DNA minor groove-binding molecules, distamycin and netropsin, effectively compete with H-NS for the binding to an AT-rich sequence (Yamada et al., 1990; Gordon et al., 2011). Furthermore, our EMSA experiments strongly suggest that the binding affinity of FIS and H-NS for *cspA* DNA increases when the proteins are combined (Figure 2, Figure S1). Two not mutually exclusive circumstances can explain this observation: FIS can bend DNA upon binding (Pan et al., 1996), thus facilitating the interaction of H-NS that is known to preferentially recognize curved DNA (Yamada et al., 1990) or H-NS can actively curve not curved DNA (Spurio et al., 1997), thus easing the positioning of FIS at its binding sites. Since the regions protected by FIS (site F2) and H-NS overlap the –35 and –10 elements of the promoter (Figure 4), their occupancy may allow the formation of contacts between the two nucleoid proteins and RNA polymerase thereby accounting for the stimulation/repression of *cspA* transcription by FIS and H-NS (Figure 5). This scenario shares many similarities with the models proposed for the regulation of *hns* and *virF* genes by H-NS and FIS in which the protein molar ratio reflects the nature of the hetero-complex formed and ultimately controls the switch between a transcriptionally active or repressed state (Falconi et al., 1996, 2001).

The composition of the population of NAPs in the cell is not fixed and fluctuations of these proteins is thought to mediate global changes in nucleoid structure and transcriptional activity (Azam and Ishihama, 1999). In fact, growth phase and variations of other environmental parameters (i.e., temperature, pH, availability of nutrients and oxygen) produce a characteristic profile of nucleoid-associated proteins. While the intracellular level of H-NS is generally high and quite constant, the expression of FIS is strongly dependent on growth conditions, being elevated in early exponential phase and upon a nutrient up-shift (Ball et al., 1992; Dillon and Dorman, 2010).

Interestingly, the expression pattern of *cspA* at 37°C seems to mirror that of FIS, since *cspA* transcription is maximal before growth resumption from stationary phase (in the period

of time immediately preceding the first cell division), and it is progressively reduced at later stages of growth. Since FIS is able to directly stimulate the transcription of *cspA* and to contrast the inhibitory effect of H-NS, as demonstrated by *in vitro* transcription assays (Figure 5A), it is likely that changes of FIS intracellular levels help to link the physiological state of cells and the environmental conditions to *cspA* expression. This model of regulation is confirmed also by *in vivo* experiments performed in different genetic backgrounds. In fact, with respect to wt strain, the raise in *cspA* mRNA level during the initial minutes after escape from stationary phase is reduced and delayed in *fis*- background and increased in *hns*- strain (Figure 5B). Furthermore, in line with our model, at increased culture age the *cspA* mRNA level in both wt and *hns*- strains becomes comparable to that of *fis*- strain. Therefore, FIS seems able to bind *cspA* promoter and sponsor the activation of *cspA* transcription in a dose-dependent manner so that when its level drops below a certain concentration, it can no longer contrast the inhibition by H-NS.

In addition to NAPs, *cspA* promoter is able to respond to changes of (p)ppGpp. Although, the intracellular concentration of this mediator molecule is quite stable under physiological growth conditions, its synthesis is affected by several types of nutritional limitations and other environmental stimuli, like an abrupt change of temperature (Magnusson et al., 2005). Here, we show that a high level of (p)ppGpp abolishes the *in vivo* induction of *cspA* after a nutritional up-shift and that this effect is due to a direct inhibition of *cspA* promoter activity (Figure 6). Numerous hypotheses have been made to explain the negative regulation of transcription by (p)ppGpp. These mechanisms, not mutually exclusive but possibly working in concert, rely on three main conditions: (i) the presence of particular features of promoters sensitive to (p)ppGpp; (ii) the direct interaction between (p)ppGpp and RNAP; (iii) the regulatory role of DksA protein (reviewed in Magnusson et al., 2005; Potrykus and Cashel, 2008; Hauryliuk et al., 2015). Interestingly, the *cspA* promoter contains, in the region between the TATA box (–10) and the transcriptional start site (+1), a GC-rich sequence termed discriminator (Figure 4), that is a key element of those promoters repressed by (p)ppGpp. Furthermore, in our genetic system, a direct interaction of (p)ppGpp with RNAP is supported by the fact that the *in vitro* inhibition of *cspA* transcription by this effector is dependent on the amount of NTPs added. In fact, an excess of NTPs (from 0.1 to 0.5 mM) causes a general decreased capability of the alarmone to negatively affect the promoter activity of *cspA* and makes the transcription repression independent from the concentration of (p)ppGpp (compare mRNA levels at 200 and 400  $\mu$ M of (p)ppGpp in Figure 6B, Figure S5). Similarly, a sensitivity to the nature and concentration of initiating nucleotide was observed in the (p)ppGpp-mediated regulation of rRNA transcription (Jöres and Wagner, 2003; Kolmsee et al., 2011) where the major step affected by (p)ppGpp is the formation of the ternary transcription initiation complex. This finding strongly indicates that the unusual nucleotide and NTPs compete for access to the active center of RNAP and is consistent also with the (p)ppGpp-RNAP cocrystals that positioned (p)ppGpp in the secondary channel of the enzyme

near the catalytic center (Artsimovitch et al., 2004). Notably, the amount of (p)ppGpp added in our *in vitro* transcription assays is very close to that used in similar studies (Heinemann and Wagner, 1997) and compatible with that estimated *in vivo* (~900  $\mu$ M) in response to amino acid starvation (Traxler et al., 2008).

The *cspA* and *fis* genes are apparently regulated in parallel and this may be attributed to the role of the common regulatory molecule, (p)ppGpp. As *cspA*, also *fis* expression is controlled by this unusual nucleotide as demonstrated by the fact that the stringent response produced by the artificial induction of (p)ppGpp in early log cells causes a dramatic reduction of *fis* mRNA while transcription of *hns*, under the same experimental conditions, is less affected (Figure 7). Interestingly, also *fis* promoter contains a GC-rich discriminator sequence downstream the -10 position and its transcription peaks during the early logarithmic growth phase, a condition characterized by a low concentration of (p)ppGpp, to decrease soon thereafter, as the level of (p)ppGpp begins to rise in cells approaching the stationary phase (Mallik et al., 2004). Therefore, (p)ppGpp is a fundamental element of this regulatory circuit and is able to repress *cspA* transcription by a dual mechanism: one direct, exerting a negative action on the functionality of *cspA* promoter and one indirect, depriving the target gene of its natural transcriptional activator FIS. On the other hand, a low level of FIS favors the silencing of *cspA* by H-NS.

A recent expression analysis of *csp* genes based on quantitative RT-PCR (Czapski and Trun, 2014) has demonstrated that the levels of *csp* mRNAs change with growth phase and type of medium. In particular, in rich defined medium the transcripts of *cspA*, *B*, *G*, and *I* were found to accumulate preferentially in early log phase, those of *cspC* and *cspD* in mid-log phase

and stationary phase, respectively, and *cspE* mRNA was found to be constitutively present. Our comparison of *cspB*, *G*, and *I* expression pattern using both Northern blotting and quantitative RT-qPCR (Figure S6) essentially confirms that these genes display a growth-dependent expression similar to that of *cspA*. Therefore, it is tempting to speculate that at 37°C also the fluctuations of other *csp* genes could be regulated by a network similar to that found for *cspA* and based on the antagonistic role of FIS/H-NS and the transcription inhibition by (p)ppGpp.

## AUTHOR CONTRIBUTIONS

AB designed and performed most of the experiments giving an important contribution also to the analysis and interpretation of data. Additionally AB in collaboration with MG has been dealing with the preparation of figures, material and methods and references. MG and MF planned and carried out some experiments. MF and AG have mainly done the analysis and interpretation of results, drafting the work (including figures) and revising it critically.

## FUNDING

Fondi di Ricerca di Ateneo (FAR), Università di Camerino, to MF.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmolb.2016.00019>

## REFERENCES

- Afflerbach, H., Schroder, O., and Wagner, R. (1999). Conformational changes of the upstream DNA mediated by H-NS and Fis regulate *E. coli* rrnB P1 promoter activity. *J. Mol. Biol.* 286, 339–353. doi: 10.1006/jmbi.1998.2494
- Afflerbach, H., Schröder, O., and Wagner, R. (1998). Effects of the *Escherichia coli* DNA-binding protein H-NS on rRNA synthesis *in vivo*. *Mol. Microbiol.* 28, 641–653. doi: 10.1046/j.1365-2958.1998.00829.x
- Artsimovitch, I., Patlan, V., Sekine, S., Vassilyeva, M. N., Hosaka, T., Ochi, K., et al. (2004). Structural basis for transcriptional regulation by alarmone ppGpp. *Cell* 117, 299–310. doi: 10.1016/S0092-8674(04)00401-5
- Atlung, T., and Ingmer, H. (1997). H-NS: a modulator of environmentally regulated gene expression. *Mol. Microbiol.* 24, 7–17. doi: 10.1046/j.1365-2958.1997.3151679.x
- Azam, T. A., and Ishihama, A. (1999). Twelve species of nucleoid-associate protein from *Escherichia coli*: sequence recognition specificity and binding affinity. *J. Biol. Chem.* 274, 33105–33113. doi: 10.1074/jbc.274.46.33105
- Badaut, C., Williams, R., Arluison, V., Bouffartigues, E., Robert, B., Buc, H., et al. (2002). The degree of oligomerization of the H-NS nucleoid structuring protein is related to specific binding to DNA. *J. Biol. Chem.* 277, 41657–41666. doi: 10.1074/jbc.M206037200
- Bae, W., Xia, B., Inouye, M., and Severinov, K. (2000). *Escherichia coli* CspA-family RNA chaperone are transcription antiterminators. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7784–7789. doi: 10.1073/pnas.97.14.7784
- Ball, C. A., Osuna, R., Ferguson, K. C., and Johnson, R. C. (1992). Dramatic changes in Fis levels upon nutrient upshift in *Escherichia coli*. *J. Bacteriol.* 174, 8043–8056.
- Bouffartigues, E., Buckle, M., Badaut, C., Travers, A., and Rimsky, S. (2007). H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing. *Nat. Struct. Mol. Biol.* 14, 441–448. doi: 10.1038/nsmb1233
- Brandi, A., Pietroni, P., Gualerzi, C. O., and Pon, C. L. (1996). Post-transcriptional regulation of CspA expression in *Escherichia coli*. *Mol. Microbiol.* 19, 231–240. doi: 10.1046/j.1365-2958.1996.362897.x
- Brandi, A., and Pon, C. L. (2012). Expression of *Escherichia coli* cspA during early exponential growth at 37°C. *Gene* 492, 382–388. doi: 10.1016/j.gene.2011.10.047
- Brandi, A., Spurio, R., Gualerzi, C. O., and Pon, C. L. (1999). Massive presence of the *Escherichia coli* “major cold-shock protein” CspA under non stress conditions. *EMBO J.* 18, 1653–1659. doi: 10.1093/emboj/18.6.1653
- Browning, D. F., Cole, J. A., and Busby, S. (2000). Suppression of FNR-dependent transcription activation at the *Escherichia coli* nir promoter by FIS, IHF and H-NS: modulation of transcription initiation by a complex nucleoid-protein assembly. *Mol. Microbiol.* 37, 1258–1269. doi: 10.1046/j.1365-2958.2000.02087.x
- Browning, D. F., Grainger, D. C., and Busby, S. J. (2010). Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr. Opin. Microbiol.* 13, 773–780. doi: 10.1016/j.mib.2010.09.013
- Cammack, K. A., and Wade, H. E. (1965). The sedimentation behaviour of ribonuclease-active and -inactive ribosomes from bacteria. *Biochem. J.* 96, 671–680. doi: 10.1042/bj0960671
- Czapski, T. R., and Trun, N. (2014). Expression of *csp* genes in *E. coli* K-12 in defined rich and defined minimal media during normal growth, and after cold-shock. *Gene* 547, 91–97. doi: 10.1016/j.gene.2014.06.033



- Dillon, S. C., and Dorman, C. J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* 8, 185–195. doi: 10.1038/nrmicro2261
- Dorman, C. J. (2004). H-NS: a universal regulator for a dynamic genome. *Nat. Rev. Microbiol.* 2, 391–400. doi: 10.1038/nrmicro883
- Dorman, C. J. (2007). H-NS, the genome sentinel. *Nat. Rev. Microbiol.* 5, 157–161. doi: 10.1038/nrmicro1598
- Ermolenko, D. N., and Makhatadze, G. I. (2002). Bacterial cold-shock proteins. *Cell. Mol. Life Sci.* 59, 1902–1913. doi: 10.1007/PL00012513
- Falconi, M., Brandi, A., La Teana, A., Gualerzi, C. O., and Pon, C. L. (1996). Antagonistic involvement of FIS and H-NS proteins in the transcriptional control of *hns* expression. *Mol. Microbiol.* 19, 965–975. doi: 10.1046/j.1365-2958.1996.436961.x
- Falconi, M., Colonna, B., Prosseda, G., Micheli, G., and Gualerzi, C. O. (1998). Thermoregulation of *Shigella* and *Escherichia coli* EIEC pathogenicity. A temperature-dependent structural transition of DNA modulates accessibility of *virF* promoter to transcriptional repressor H-NS. *EMBO J.* 17, 7033–7043. doi: 10.1093/emboj/17.23.7033
- Falconi, M., Gualtieri, M. T., La Teana, A., Losso, M. A., and Pon, C. L. (1988). Proteins from prokaryotic nucleoid: primary and quaternary structure of the 15-KD *Escherichia coli* DNA binding protein H-NS. *Mol. Microbiology* 2, 323–329. doi: 10.1111/j.1365-2958.1988.tb00035.x
- Falconi, M., Higgins, N. P., Spurio, R., Pon, C. L., and Gualerzi, C. O. (1993). Expression of the gene encoding the major bacterial nucleoid protein H-NS is subject to transcriptional auto-repression. *Mol. Microbiol.* 10, 273–282. doi: 10.1111/j.1365-2958.1993.tb01953.x
- Falconi, M., Prosseda, G., Giangrossi, M., Beghetto, E., and Colonna, B. (2001). Involvement of FIS in the H-NS-mediated regulation of *virF* gene of *Shigella* and enteroinvasive *Escherichia coli*. *Mol. Microbiol.* 42, 439–452. doi: 10.1046/j.1365-2958.2001.02646.x
- Giangrossi, M., Wintraecken, K., Spurio, R., and de Vries, R. (2014). Probing the relation between protein-protein interactions and DNA binding for a linker mutant of the bacterial nucleoid protein H-NS. *Biochim. Biophys. Acta.* 1844, 339–345. doi: 10.1016/j.bbapap.2013.11.010
- Giangrossi, M., Zattoni, S., Tramonti, A., De Biase, D., and Falconi, M. (2005). Antagonistic role of H-NS and GadX in the regulation of the glutamate decarboxylase-dependent acid resistance system in *Escherichia coli*. *J. Biol. Chem.* 280, 21498–21505. doi: 10.1074/jbc.M413255200
- Giuliodori, A. M., Brandi, A., Gualerzi, C. O., and Pon, C. L. (2004). Preferential translation of cold-shock mRNAs during cold adaptation. *RNA* 10, 265–276. doi: 10.1261/rna.5164904
- Giuliodori, A. M., Di Pietro, F., Marzi, S., Masquida, B., Wagner, R., Romby, P., et al. (2010). The *cspA* mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. *Mol. Cell* 37, 21–33. doi: 10.1016/j.molcel.2009.11.033
- Goldenberg, D., Azar, I., and Oppenheim, A. B. (1996). Differential mRNA stability of the *cspA* gene in the cold-shock response of *Escherichia coli*. *Mol. Microbiol.* 19, 241–248. doi: 10.1046/j.1365-2958.1996.363898.x
- Goldenberg, D., Azar, I., Oppenheim, A. B., Brandi, A., Pon, C. L., and Gualerzi, C. O. (1997). Role of *Escherichia coli* *cspA* promoter sequences and adaptation of translational apparatus to the cold shock response. *Mol. Gen. Genet.* 256, 282–290. doi: 10.1007/s004380050571
- Goldstein, J., Pollitt, N. S., and Inouye, M. (1990). Major cold-shock protein of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 87, 283–287. doi: 10.1073/pnas.87.1.283
- Gonzalez-Gil, G., Bringmann, P., and Kahmann, R. (1996). FIS a regulator of metabolism in *Escherichia coli*. *Mol. Microbiol.* 22, 21–29. doi: 10.1111/j.1365-2958.1996.tb02652.x
- Gordon, B. R., Li, Y., Cote, A., Weirauch, M. T., Ding, P., Hughes, T. R., et al. (2011). Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10690–10695. doi: 10.1073/pnas.1102544108
- Grainger, D. C., Goldberg, M. D., Lee, D. J., and Busby, S. J. (2008). Selective repression by Fis and H-NS at the *Escherichia coli* *dps* promoter. *Mol. Microbiol.* 68, 1366–1377. doi: 10.1111/j.1365-2958.2008.06253.x
- Gualerzi, C. O., Giuliodori, A. M., Brandi, A., Di Pietro, F., Piersimoni, L., Fabbretti, A., et al. (2011). “Translation Initiation at the root of the cold-shock translation bias,” in *RIBOSOMES Structure, Function and Dynamics*, eds M. V. Rodnina, W. Wintermeyer and R. Green (New York, NY: Springer Wien), 143–154.
- Hauryliuk, V., Atkinson, G. C., Murakami, K. S., Tenson, T., and Gerdes, K. (2015). Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nat. Rev. Microbiol.* 13, 298–309. doi: 10.1038/nrmicro3448
- Heinemann, M., and Wagner, R. (1997). Guanosine 3',5'-bis(diphosphate) (ppGpp)-dependent inhibition of transcription from stringently controlled *Escherichia coli* promoters can be explained by an altered initiation pathway that traps RNA polymerase. *FEBS* 247, 990–999. doi: 10.1111/j.1432-1033.1997.00990.x
- Hommais, F., Krin, E., Laurent-Winter, C., Soutourina, O., Malpertuy, A., Le Caer, J. P., et al. (2001). Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. *Mol. Microbiol.* 40, 20–36. doi: 10.1046/j.1365-2958.2001.02358.x
- Hu, Z., Frith, M., Niu, T., and Weng, Z. (2003). SeqVISTA: a graphical tool for sequence feature visualization and comparison. *BMC Bioinform.* 4:1. doi: 10.1186/1471-2105-4-1
- Jiang, W., Hou, Y., and Inouye, M. (1997). CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. *J. Biol. Chem.* 272, 196–202. doi: 10.1074/jbc.272.1.196
- Jones, P. G., Cashel, M., Glaser, G., and Neidhardt, F. C. (1992a). Function of a relaxed-like state following temperature downshifts in *Escherichia coli*. *J. Bacteriol.* 174, 3903–3914.
- Jones, P. G., Krah, R., Tafuri, S. R., and Wolffe, A. P. (1992b). DNA gyrase, CS7.4, and the cold shock response in *Escherichia coli*. *J. Bacteriol.* 174, 5798–5802.
- Jones, P. J., and Inouye, M. (1994). The cold-shock response - a hot topic. *Mol. Microbiol.* 11, 811–818. doi: 10.1111/j.1365-2958.1994.tb00359.x
- Jöres, L., and Wagner, R. (2003). Essential steps in the ppGpp-dependent regulation of bacterial ribosomal RNA promoters can be explained by substrate competition. *J. Biol. Chem.* 278, 16834–16843. doi: 10.1074/jbc.M300196200
- Kahramanoglou, C., Seshasayee, A. S., Prieto, A. I., Ibberson, D., Schmidt, S., Zimmermann, J., et al. (2011). Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Res.* 39, 2073–2091. doi: 10.1093/nar/gkq934
- Kolmsee, T., Delic, D., Agyenim, T., Calles, C., and Wagner, R. (2011). Differential stringent control of *Escherichia coli* rRNA promoters: effects of ppGpp, DksA and the initiating nucleotides. *Microbiology* 157, 2871–2879. doi: 10.1099/mic.0.052357-0
- Lang, B., Blot, N., Bouffartigues, E., Buckle, M., Geertz, M., Gualerzi, C. O., et al. (2007). High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. *Nucleic Acids Res.* 35, 6330–6337. doi: 10.1093/nar/gkm712
- La Teana, A., Brandi, A., Falconi, M., Spurio, R., Pon, C. L., and Gualerzi, C. O. (1991). Identification of a cold shock transcriptional enhancer of the *Escherichia coli* gene encoding nucleoid protein H-NS. *Proc. Natl. Acad. Sci. U.S.A.* 88, 10907–10911. doi: 10.1073/pnas.88.23.10907
- Madrid, C., Nieto, J. M., Paytubi, S., Falconi, M., Gualerzi, C. O., and Juárez, A. (2002). Temperature- and H-NS-dependent regulation of a plasmid-encoded virulence operon expressing *Escherichia coli* hemolysin. *J. Bacteriol.* 184, 5058–5066. doi: 10.1128/JB.184.18.5058-5066.2002
- Magnusson, L. U., Farewell, A., and Nyström, T. (2005). ppGpp: a global regulator in *Escherichia coli*. *Trends Microbiol.* 13, 236–242. doi: 10.1016/j.tim.2005.03.008
- Mallik, P., Pratt, T. S., Beach, M. B., Bradley, M. D., Undamatla, J., and Osuna, R. (2004). Growth phase-dependent regulation and stringent control of *fis* are conserved processes in enteric bacteria and involve a single promoter (*fisP*) in *Escherichia coli*. *J. Bacteriol.* 186, 122–135. doi: 10.1128/JB.186.1.122-135.2004
- Muskhelishvili, G., and Travers, A. (2003). Transcription factor as a topological homeostat. *Front. Biosci.* 8, d279–d285. doi: 10.2741/969
- Osuna, R., Finkel, S. E., and Johnson, R. C. (1991). Identification of two functional regions in Fis: the N-terminus is required to promote Hin-mediated DNA inversion but not  $\lambda$  excision. *EMBO J.* 10, 1593–1603.
- Ouafa, Z. A., Reverchon, S., Lautier, T., Muskhelishvili, G., and Nasser, W. (2012). The nucleoid-associated proteins H-NS and FIS modulate the DNA supercoiling response of the *pel* genes, the major virulence factors in the plant pathogen bacterium *Dickeya dadantii*. *Nucleic Acids Res.* 40, 4306–4319. doi: 10.1093/nar/gks014



- Pan, C. Q., Finkel, S. E., Cramton, S. E., Feng, J. A., Sigman, D. S., and Johnson, R. C. (1996). Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. *J. Mol. Biol.* 264, 675–695. doi: 10.1006/jmbi.1996.0669
- Potrykus, K., and Cashel, M. (2008). (p)ppGpp: still magical? *Annu. Rev. Microbiol.* 62, 35–51. doi: 10.1146/annurev.micro.62.081307.162903
- Rimsky, S., and Travers, A. (2011). Pervasive regulation of nucleoid structure and function by nucleoid-associated proteins. *Curr. Opin. Microbiol.* 14, 136–141. doi: 10.1016/j.mib.2011.01.003
- Ross, W., Thompson, J. F., Newlands, J. T., and Gourse, R. L. (1990). *E. coli* Fis protein activates ribosomal RNA transcription *in vitro* and *in vivo*. *EMBO J.* 9, 3733–3742.
- Roth, A., Urmoneit, B., and Messer, W. (1994). Functions of histone-like proteins in the initiation of DNA replication at *oriC* of *Escherichia coli*. *Biochimie* 76, 917–923. doi: 10.1016/0300-9084(94)90016-7
- Sambrook, J., and Russell, D. W. (2001). *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor, NY: CSHL Press.
- Sarubbi, E., Rudd, K. E., Xiao, H., Ikehara, K., Kalman, M., and Cashel, M. (1989). Characterization of the *spoT* gene of *Escherichia coli*. *J. Biol. Chem.* 264, 15074–15082.
- Schneider, R., Travers, A., and Muskhelishvili, G. (1997). FIS modulates growth phase-dependent topological transitions of DNA in *Escherichia coli*. *Mol. Microbiol.* 26, 519–530. doi: 10.1046/j.1365-2958.1997.5951971.x
- Shao, Y., Feldman-Cohen, L. S., and Osuna, R. (2008). Functional characterization of the *Escherichia coli* Fis-DNA binding sequence. *J. Mol. Biol.* 376, 771–785. doi: 10.1016/j.jmb.2007.11.101
- Spurio, R., Falconi, M., Brandi, A., Pon, C. L., and Gualerzi, C. O. (1997). The oligomeric structure of nucleoid protein H-NS is necessary for recognition of intrinsically curved DNA and for DNA bending. *EMBO J.* 16, 1795–1805. doi: 10.1093/emboj/16.7.1795
- Stella, S., Spurio, R., Falconi, M., Pon, C. L., and Gualerzi, C. O. (2005). Nature and mechanism of the *in vivo* oligomerization of nucleoid protein H-NS. *EMBO J.* 24, 2896–2905. doi: 10.1038/sj.emboj.7600754
- Tedin, K., Witte, A., Reisinger, G., Lubitz, W., and Bläsi, U. (1995). Evaluation of the *E. coli* *rrnB* P1 promoter and phage-derived lysis genes for the use in a biological containment system: a concept study. *J. Biotechnol.* 39, 137–148. doi: 10.1016/0168-1656(95)00003-9
- Traxler, M. F., Summers, S. M., Nguyen, H. T., Zacharia, V. M., and Hightower, G. A. (2008). The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Mol. Microbiol.* 68, 1128–1148. doi: 10.1111/j.1365-2958.2008.06229.x
- Ulissi, U., Fabbretti, A., Sette, M., Giuliodori, A. M., and Spurio, R. (2014). Time-resolved assembly of a nucleoprotein complex between *Shigella flexneri* *virF* promoter and its transcriptional repressor H-NS. *Nucleic Acids Res.* 42, 13039–13050. doi: 10.1093/nar/gku1052
- Wang, W., Li, G. W., Chen, C., Xie, X. S., and Zhuang, X. (2011). Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* 9, 1445–1449. doi: 10.1126/science.1204697
- Yamada, H., Muramatsu, S., and Mizuno, T. (1990). An *Escherichia coli* protein that preferentially binds to sharply curved DNA. *J. Biochem.* 108, 420–425.
- Yamanaka, K., Fang, L., and Inouye, M. (1998). The CspA family in *Escherichia coli*: multiple gene duplication for stress adaptation. *Mol. Microbiol.* 27, 247–255. doi: 10.1046/j.1365-2958.1998.00683.x
- Yamanaka, K., and Inouye, M. (2001). Induction of CspA, an *E. coli* major cold-shock protein, upon nutritional shift at 37°C. *Genes Cells* 6, 279–290. doi: 10.1046/j.1365-2443.2001.00424.x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Brandi, Giangrossi, Giuliodori and Falconi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# DNA-Binding Proteins Regulating pIP501 Transfer and Replication

Elisabeth Grohmann<sup>1,2\*</sup>, Nikolaus Goessweiner-Mohr<sup>3,4,5,6\*</sup> and Sabine Brantl<sup>7\*</sup>

<sup>1</sup> Division of Infectious Diseases, University Medical Center Freiburg, Freiburg im Breisgau, Germany, <sup>2</sup> Life Sciences and Technology, Beuth University of Applied Sciences Berlin, Berlin, Germany, <sup>3</sup> Center for Structural System Biology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, <sup>4</sup> Deutsches Elektronen-Synchrotron, Hamburg, Germany, <sup>5</sup> Institute of Molecular Biotechnology, Austrian Academy of Sciences, Vienna, Austria, <sup>6</sup> Research Institute of Molecular Pathology, Vienna, Austria, <sup>7</sup> Lehrstuhl für Genetik, Biologisch-Pharmazeutische Fakultät, AG Bakteriengenetik, Friedrich-Schiller-Universität Jena, Jena, Germany

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Spanish National Research Council,  
Spain

### Reviewed by:

Itziar Alkorta,  
University of the Basque Country,  
Spain  
Gloria Del Solar,  
Spanish National Research Council,  
Spain

### \*Correspondence:

Elisabeth Grohmann  
elisabeth.grohmann  
@beuth-hochschule.de  
Nikolaus Goessweiner-Mohr  
nikolaus.goessweiner  
@imba.oeaw.ac.at  
Sabine Brantl  
sabine.brantl@uni-jena.de

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 24 June 2016

**Accepted:** 29 July 2016

**Published:** 11 August 2016

### Citation:

Grohmann E, Goessweiner-Mohr N  
and Brantl S (2016) DNA-Binding  
Proteins Regulating pIP501 Transfer  
and Replication.  
Front. Mol. Biosci. 3:42.  
doi: 10.3389/fmolb.2016.00042

pIP501 is a Gram-positive broad-host-range model plasmid intensively used for studying plasmid replication and conjugative transfer. It is a multiple antibiotic resistance plasmid frequently detected in clinical *Enterococcus faecalis* and *Enterococcus faecium* strains. Replication of pIP501 proceeds unidirectionally by a theta mechanism. The minimal replicon of pIP501 is composed of the *repR* gene encoding the essential rate-limiting replication initiator protein RepR and the origin of replication, *oriR*, located downstream of *repR*. RepR is similar to RepE of related streptococcal plasmid pAMβ1, which has been shown to possess RNase activity cleaving free RNA molecules in close proximity of the initiation site of DNA synthesis. Replication of pIP501 is controlled by the concerted action of a small protein, CopR, and an antisense RNA, RNAIII. CopR has a dual function: It acts as transcriptional repressor at the *repR* promoter and, in addition, prevents convergent transcription of RNAIII and *repR* mRNA (RNAII), which indirectly increases RNAIII synthesis. CopR binds asymmetrically as a dimer at two consecutive binding sites upstream of and overlapping with the *repR* promoter. RNAIII induces transcriptional attenuation within the leader region of the *repR* mRNA (RNAII). Deletion of either control component causes a 10- to 20-fold increase of plasmid copy number, while simultaneous deletions have no additional effect. Conjugative transfer of pIP501 depends on a type IV secretion system (T4SS) encoded in a single operon. Its transfer host-range is considerably broad, as it has been transferred to virtually all Gram-positive bacteria including *Streptomyces* and even the Gram-negative *Escherichia coli*. Expression of the 15 genes encoding the T4SS is tightly controlled by binding of the relaxase TraA, the transfer initiator protein, to the operon promoter overlapping with the origin of transfer (*oriT*). The T4SS operon encodes the DNA-binding proteins TraJ (VirD4-like coupling protein) and the VirB4-like ATPase, TraE. Both proteins are actively involved in conjugative DNA transport. Moreover, the operon encodes TraN, a small cytoplasmic protein, whose specific binding to a sequence upstream of the *oriT* *nic*-site was demonstrated. TraN seems to be an effective repressor of pIP501 transfer, as conjugative transfer rates were significantly increased in an *E. faecalis* pIP501  $\Delta$ *traN* mutant.

**Keywords:** conjugative plasmid, replication, copy number control, type IV secretion, broad-host-range, transfer control

## INTRODUCTION

Plasmids are extrachromosomal elements, by definition not encoding any essential functions for the bacterial host but rather contributing additional traits, which can be advantageous or even essential for survival under particular conditions, e.g., in the presence of antibiotic pressure. pIP501 is a considerably small, 30.6-kb broad-host-range self-transmissible plasmid, which was isolated from a clinical *Streptococcus agalactiae* strain (Evans and Macrina, 1983). It belongs to incompatibility group *Inc18* and encodes resistance to antibiotics of the macrolide/lincosamide/streptogramin (MLS) group and to chloramphenicol.

*Inc18* plasmids encode an efficient plasmid stabilization system, the  $\epsilon$ - $\theta$   $\zeta$  locus functioning as a toxin-antitoxin system (Ceglowski et al., 1993). The  $\omega$ - $\epsilon$ - $\zeta$  operon of pSM19035 and of other *Inc18* plasmids is a novel proteic plasmid addiction system in which the  $\epsilon$  and  $\zeta$  genes code for an antitoxin and a toxin, respectively, while  $\omega$ - plays an autoregulatory role. Broad-host-range efficiency of the  $\omega$ - $\epsilon$ - $\zeta$  cassette has been demonstrated in eight different Gram-positive bacteria, including among others the human pathogens, *E. faecalis*, *S. agalactiae*, *Streptococcus pyogenes*, and *Staphylococcus aureus* (Brzozowska et al., 2012). Expression of toxin Zeta was shown to be bactericidal for Gram-positive bacteria and bacteriostatic for the Gram-negative *Escherichia coli*, thus stabilizing plasmids in *E. coli* less efficiently than in Gram-positive bacteria (Zielenkiewicz and Ceglowski, 2005).

pIP501 replicons stabilized by this toxin-antitoxin system have been frequently encountered in *E. faecium* isolates from geographically diverse clinical, human community and poultry fecal origin (Rosvoll et al., 2010). In addition, pIP501-like replicons are often linked with the vancomycin resistance phenotype encoded by *vanA* (Rosvoll et al., 2010).

pIP501 is characterized by a replicon, which is tightly controlled on several levels by protein and RNA key players, and a conjugative transfer (*tra*) region comprising almost half of the plasmid genome encoding 15 putative Tra factors making up a Gram-positive T4SS. The review summarizes (i) key findings on replication and copy number control processes that involve DNA-binding proteins and (ii) current knowledge on key factors of the pIP501 T4SS whose activity involves interaction with DNA. The review ends with a Conclusion and Perspective section on urgent future research needs in the field of plasmid biology.

## PIP501 REPLICATION AND COPY NUMBER CONTROL

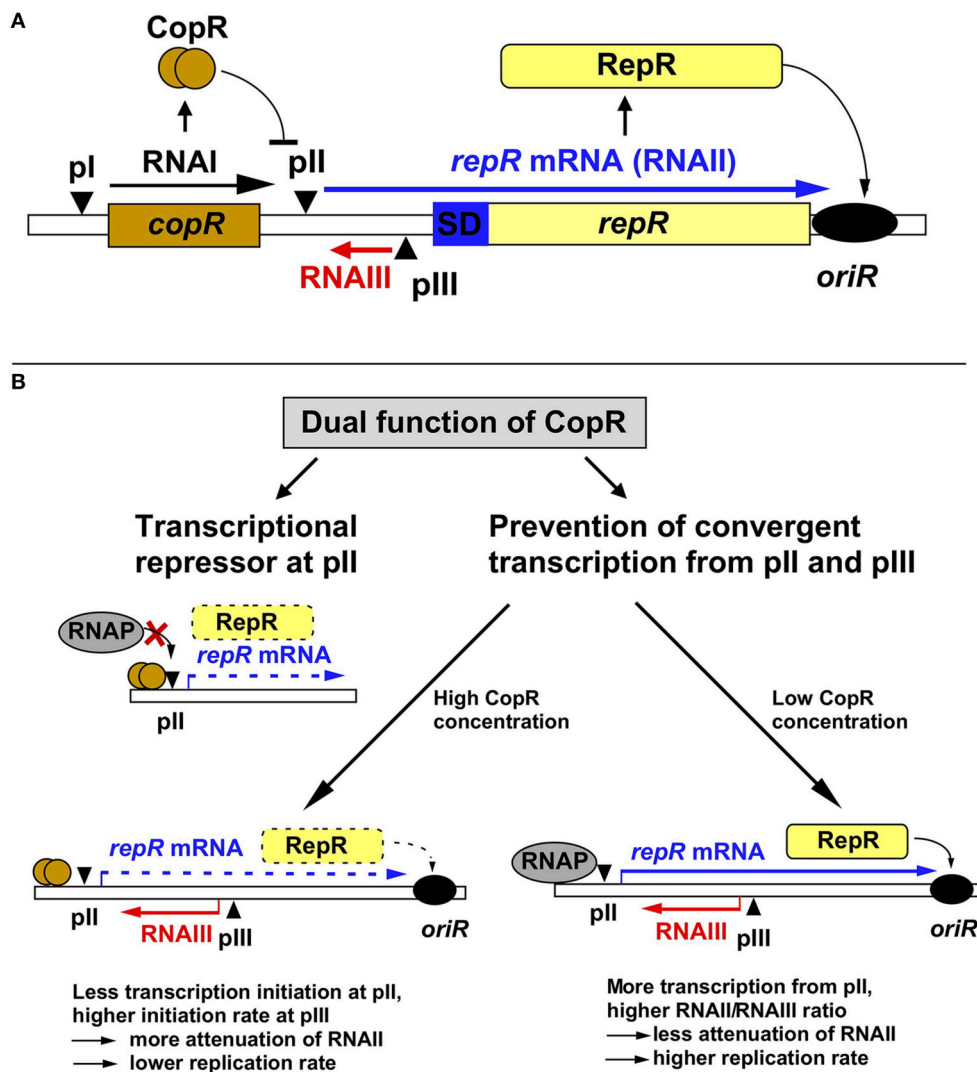
Plasmid pIP501 from *S. agalactiae* belongs, together with pAM $\beta$ 1 from *E. faecalis* and pSM19035 from *S. pyogenes* to the *Inc18* family of plasmids that replicate unidirectionally by the theta mechanism (Brantl et al., 1990; Bruand et al., 1993) in a multitude of Gram-positive bacteria, including *Bacillus subtilis*. All three plasmids show a high degree of sequence identity in their replication regions (Brantl et al., 1989, 1990; Swinfield et al., 1990).

## The RepR Protein

The minimal pIP501 replicon comprises the *repR* gene encoding the essential replication initiator protein RepR (57.4 kDa) and the replication origin, *oriR* (Brantl et al., 1990) located downstream of the *repR* gene (see **Figure 1**). The minimal origin *oriR* has been narrowed down to 52 bp and includes an inverted repeat, both branches of which are essential (Brantl and Behnke, 1992a). The RepR protein is rate-limiting for pIP501 replication (Brantl and Behnke, 1992c) and can both act *in cis* and *in trans* at *oriR*. The *repR* promoter pII is located 300 bp upstream of the Shine-Dalgarno (SD) sequence of the *repR* gene, and this leader region proved to be essential for replication control (see below). RepR of pIP501 has not been analyzed in detail. However, the highly similar RepE protein from pAM $\beta$ 1 has been biochemically characterized (Le Chatelier et al., 2001): RepE is a monomer and binds specifically, rapidly and durably to the origin *oriE*<sub>pAM $\beta$ 1</sub> at a unique binding site immediately upstream of the replication initiation site. RepE binding induces only a weak bend. In addition, it also binds non-specifically to single stranded (ss) DNA with a 2- to 4-fold greater affinity than for double stranded (ds) *oriE*. RepE binding to *oriE*<sub>pAM $\beta$ 1</sub> causes denaturation of the AT-rich sequence downstream of its binding site yielding an open complex that is atypical: Its formation does not require multiple RepE binding sites or a strong *oriE* bending or any co-factors, and its melted region acts as substrate for RepE binding. These properties and the requirement of transcription through the origin for DNA polymerase I to initiate replication as well as a primosome to load the replisome indicate that RepE might assist primer generation at the origin. It has been hypothesized that it might cleave its own *repE* mRNA downstream of the ORF to generate the replication primer. As RepR<sub>pIP501</sub> and RepE<sub>pAM $\beta$ 1</sub> display 97% sequence identity, it can be assumed that these characteristics also apply for RepR<sub>pIP501</sub>.

## Regulation of pIP501 Replication by Two Components

Replication of pIP501 is regulated by the products of two non-essential genes, *copR* and *rnaIII* (Brantl, 2014, 2015), which control the synthesis of the rate-limiting replication initiator protein RepR (Brantl and Behnke, 1992c). RNAIII is a 136-nt long antisense RNA and CopR is a small protein composed of 92 amino acids (aa, see below). RNAIII induces premature termination (attenuation) of transcription of the essential *repR* mRNA (Brantl et al., 1993; Brantl and Wagner, 1994, 1996; Heidrich and Brantl, 2003, 2007). CopR acts as transcriptional repressor at the essential *repR* promoter pII (Brantl, 1994). Point mutations and deletions in either *rnaIII* or *copR* result in the same 10- to 20-fold increase in the copy number of pIP501 derivatives (Brantl and Behnke, 1992b). However, simultaneous deletions do not display additive effects suggesting the involvement of a limiting host factor. Surprisingly, the half-life of RNAIII is with 30 min unusually long (Brantl and Wagner, 1996). Such a long-lived antisense RNA should presumably be a poor regulator since fortuitous decreases in plasmid copy number in individual cells could only be slowly corrected, resulting in unstable maintenance. However, unstable



**FIGURE 1 | Regulation of pIP501 replication. (A)** Working model on regulation of pIP501 replication. The minimal replicon with the *copR* and *repR* genes is shown, separated by the 329-bp long leader region. **(B)** Dual function of CopR. On the one hand, CopR represses transcription from the *repR* promoter *pII* by binding to its operator region upstream of the  $-35$  box, thereby inhibiting RNAP binding. On the other hand, CopR prevents convergent transcription from *pII* and *pIII* that, in its absence reduces initiation at the supercoiling-sensitive antisense promoter *pIII* which results in lower RNAIII compared to RNAII levels, yielding a lower attenuation rate, and, hence, higher replication rate. Additionally, the higher amount of RNAII (*repR* mRNA) titrates the remaining long-lived RNAIII, which further reduces the amount of the inhibitor. In the presence of CopR, RNAP transcribes less frequently through *pIII*, which allows higher initiation rates at *pIII* resulting in increased premature termination of RNAII transcription and, consequently, lower replication rates. Left, the plasmid-copy number is 10-fold lower than right, but relatively more RNAIII is present, the amount of RNAIII is (as determined in Northern blots) approximately the same in both cases, reflected by the same thickness of the red arrows symbolizing RNAIII.

maintenance of pIP501 derivatives was never observed, likely because the second regulator CopR in fact has a dual function and thus provides pIP501 with a strategy to cope with the risk of unstable inheritance (Figure 1). CopR exerts its effect, by the same molecular event, on two levels: transcriptional repression of *repR* mRNA synthesis (see below), and accumulation of RNAIII by prevention of convergent transcription, thereby indirectly increasing transcription initiation from the antisense promoter *pIII* (Brantl and Wagner, 1997). The discovery of a second function for CopR was initiated by the surprising

finding that high copy number pIP501 derivatives lacking *copR* and low copy number derivatives containing *copR* produce the same intracellular amounts of RNAIII. Transcriptional *pI-lacZ* fusions revealed that CopR does not activate its own promoter *pI* (Brantl, 1994) and half-life measurements indicated that CopR does not affect the half-life of RNAIII. Instead, in the presence of both sense promoter *pII* and antisense promoter *pIII* *in cis*, CopR provided *in cis* or *in trans* causes an increase in the intracellular concentration of RNAIII. This effect can be attributed to the CopR protein and not the *copR* mRNA



(Brantl and Wagner, 1997). Apparently, in the absence of CopR, the increased (de-repressed) RNAII transcription interferes, *in cis*, with initiation of RNAIII transcription (“convergent transcription”), yielding a lower RNAIII/plasmid ratio. The crucial factor in convergent transcription is the movement of the RNA polymerase toward or through the pIII promoter region, whereas it does not proceed through pII. Promoter pII as well as promoter pIII are supercoiling sensitive indicated by the effect of the gyrase inhibitor novobiocin on the accumulation of both RNAII and RNAIII (Brantl and Wagner, 1997). Therefore, in the absence of CopR, transcription from pII reduces initiation at pIII by inducing positive supercoils. By contrast, in the presence of CopR, promoter pII is 10-fold repressed, so that convergent transcription is mostly abolished. Consequently, more transcription from promoter pIII can be initiated resulting in increased RNAIII/plasmid ratios. Therefore, we propose the following model for pIP501 copy number control: RNAIII alone is able to adjust increases in copy number. At higher plasmid concentrations, more RNAIII is synthesized which in turn increases transcriptional attenuation of RNAII thus decreasing the replication frequency. In contrast, fortuitous copy number decreases cannot rapidly be corrected by RNAIII, since its long half-life (Brantl and Wagner, 1996) will result in high concentrations of the inhibitor, which threatens to yield a replication frequency inappropriately low for the current copy number. CopR, due to its dual function, can correct downward fluctuations of the plasmid copy number: Decreased synthesis of CopR de-represses promoter pII. This has two consequences (Brantl, 1994): (1) enhanced transcription of RNAII and (2) convergent transcription, which reduces RNAIII transcription. Both effects enhance RepR synthesis resulting in a higher replication frequency. The molecular event of pII de-repression works as an amplifier. In summary, the concerted action of two regulatory components, RNAIII and CopR, efficiently regulates pIP501 replication and ensures stable plasmid maintenance.

## Biochemical Characterization of the CopR Protein

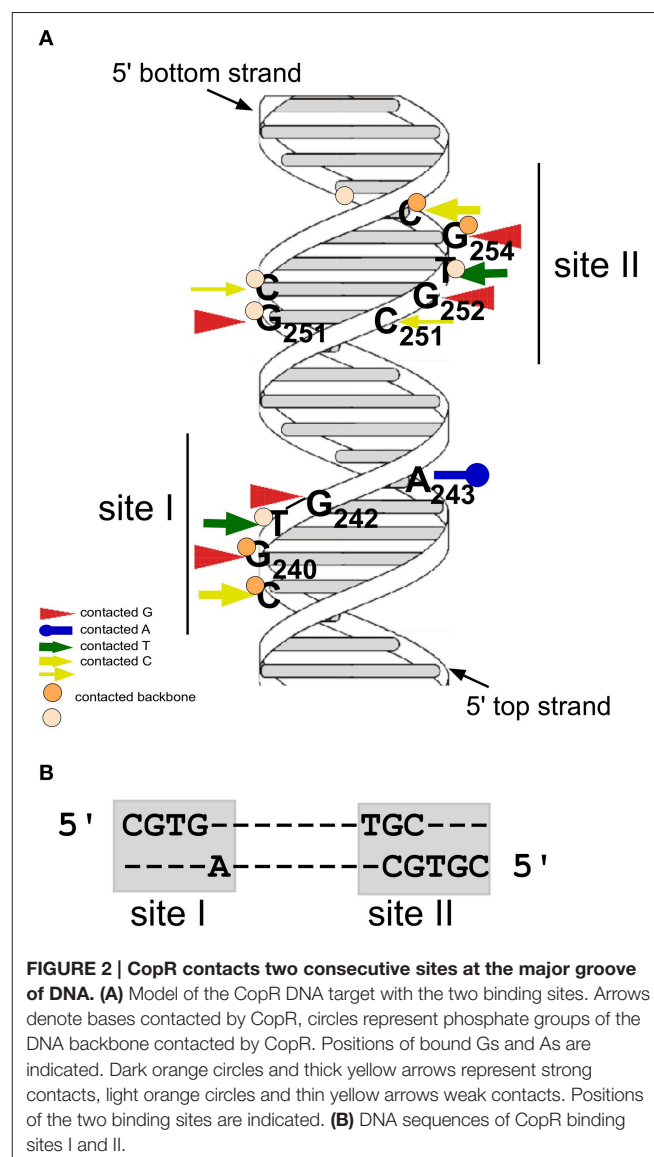
### Three *Inc18* Family Plasmids Encode Almost Identical Cop Proteins

Two almost identical Cop proteins with the same functions are encoded by the related streptococcal plasmids pAMβ1 (CopF) and pSM19035 (CopS) that share a high degree of sequence similarity with CopR at the aa level (Swinfield et al., 1990; Ceglowski et al., 1993): only two conservative aa exchanges at positions 51 and 80 are present (Brantl et al., 1994) and 2 additional aa (CopF) or two lacking aa (CopS) are found at the C-terminal end. To date, CopR is the best characterized Cop protein of this family. The repressor activity of CopF has been demonstrated and its operator sequence was narrowed down to a region of 31 bp (Le Chatelier et al., 1994). CopS has not been characterized in detail.

### Identification of Bases and Phosphate Residues Contacted by CopR

The gene product of the *copR* gene is a small protein composed of 92 aa (predicted MW 10.4 kDa) that acts as transcriptional

repressor at the essential *repR* promoter pII. CopR binds to a 44-bp region containing inverted repeat IR1 upstream of pII (Brantl, 1994). It does not autoregulate its own promoter pI nor does it activate the antisense RNA promoter pIII (Brantl, 1994). To identify bases and phosphates at the backbone directly contacted by CopR, chemical footprinting studies were performed (Steinmetzer and Brantl, 1997; **Figure 2A**). Methylation interference identified three guanines (G240, G242, and G251) and one cytosine (C239) in the top strand and two guanines (G252 and G254) and one cytosine (C255) in the bottom strand that are contacted by CopR in the major groove of DNA (Brantl et al., 1994). Furthermore, missing base interference uncovered the contribution of the bases adjacent to these guanines to the specific DNA-protein contacts. To determine phosphate residues in the DNA backbone essential for CopR binding ethylation interference experiments were employed. In the top strand, ethylation of C239, G240, and

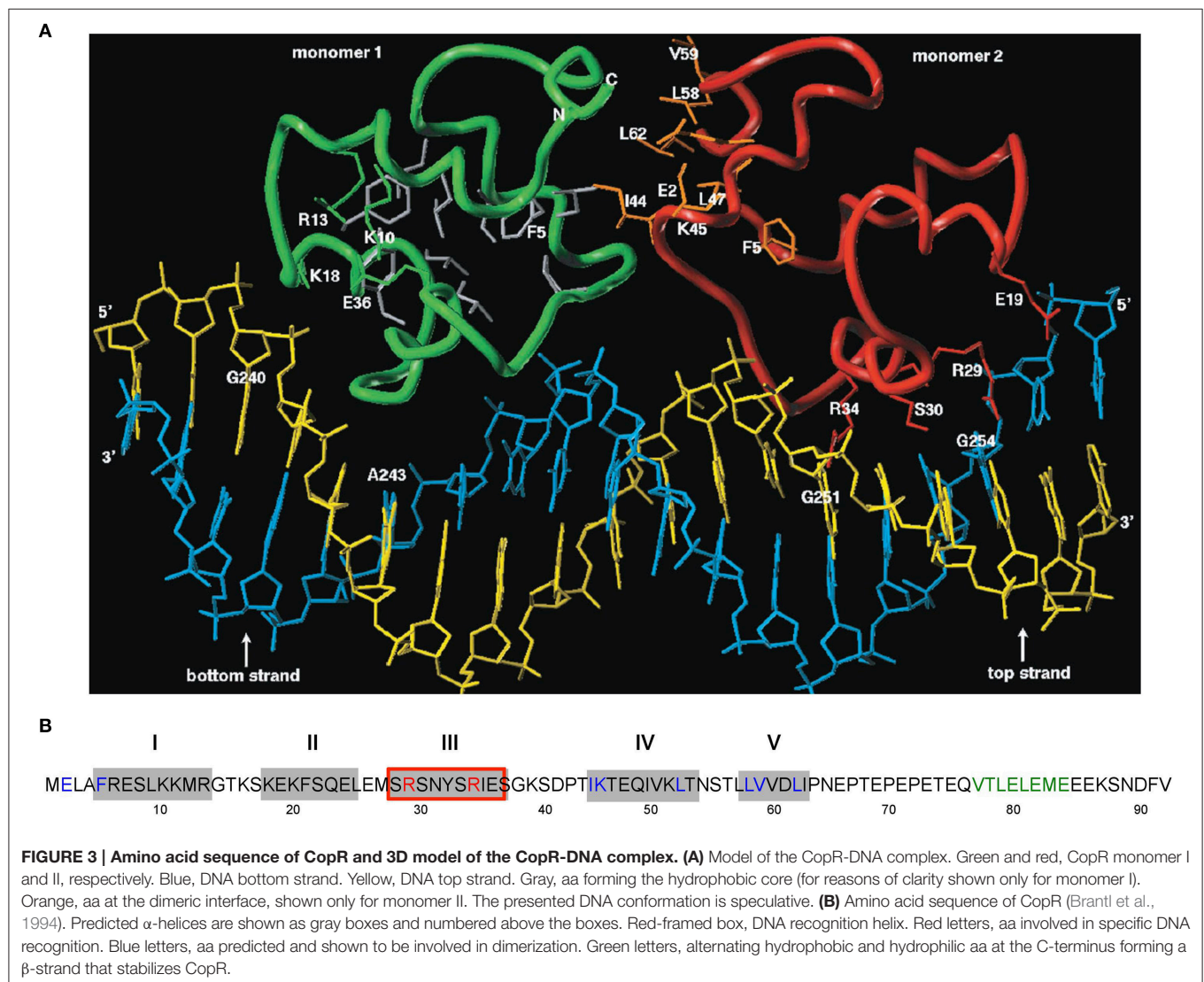


T241 interfered strongly with CopR binding while in the bottom strand, ethylation of T253, G254, and C255 affected binding. The recognition sequence of CopR is situated at the center of inverted repeat IR1. The protein contacts two consecutive major grooves (site I and II) on the same face of the DNA. Both binding sites share the common sequence motif 5'CGTG3', and the outermost G is most important for CopR binding. A243 and G/C251 located within the loop region of the inverted repeat IR1 evoke an imperfect symmetry within the binding sequence (Figure 2). The sequence of the CopR operator was narrowed down to 17 bp. Gel filtration and native gel electrophoresis revealed that CopR is mainly dimeric under the conditions assayed (Steinmetzer and Brantl, 1997). An initially obtained sigmoidal binding curve proved to be the result of two coupled equilibria, on the one hand dimerization of CopR monomers and on the other hand CopR dimer-DNA binding. Using analytical ultracentrifugation, a  $K_{\text{Dimer}}$ -value of  $1.44 \pm 0.49 \times 10^{-6}$  M for CopR dimers was determined (Steinmetzer et al., 1998) indicating relatively weak

interactions between the two monomers. Using the  $K_{\text{Dimer}}$ -value and the binding curve, the equilibrium dissociation constant  $K_2$  for the CopR-DNA complex was calculated to be  $4 \pm 1.3 \times 10^{-10}$  M, i.e.,  $\approx 0.4$  nM. In this concentration range, CopR is mostly monomeric. By quantitative Western blotting, the intracellular concentration of CopR in *B. subtilis* carrying low copy number (*copR<sup>+</sup> rnaIII<sup>+</sup>*) pIP501 derivatives was determined to be 20–30  $\mu$ M. As this value is 10- to 20-fold higher than the  $K_{\text{Dimer}}$ , CopR is preferentially present as a dimer in the cell. Using gel-shift assays with wild-type and a C-terminally truncated CopR species (Cop $\Delta$ 20), it was demonstrated that CopR also binds to the DNA as a preformed dimer (Steinmetzer et al., 1998).

### 3D Model of CopR and Identification of Residues Involved in DNA Recognition and Dimerization

A structural model encompassing the N-terminal 63 aa of CopR was constructed (Figure 3). This model was based on the rather low (14%) sequence similarity to the P22 c2 repressor



(Steinmetzer et al., 2000b). The model lacks the C-terminal 29 aa that had been found previously to be not important for DNA binding. In analogy to the P22 c2 repressor this model proposes that CopR is a HTH (helix-turn-helix) protein and describes the property of the protein to bind to DNA as a dimer at two consecutive major grooves (Steinmetzer et al., 1998). The protein backbone is built up by five  $\alpha$ -helices, two of which are involved in DNA binding. Helix I is situated between aa 5 and 13. Helix II containing aa 18–25 is proposed to be the stabilization helix, while helix III comprising aa 29–37 is suggested to be the recognition helix. Moreover, aa 44–54 and 58–62 are predicted to form helices IV and V. The model proposes that aa R29, S30, S33, and R34 in the recognition-helix contact defined bases in the DNA sequence-specifically. In addition, residues K10, K18, K20, S28, N31, and S40 are supposed to contact the DNA-phosphate backbone sequence-unspecifically. E36 is near K10, R13, and K18 that, in the model, are close together in space and in contact with the phosphate backbone of the DNA. Residues F5, L9, F21, L25, Y32, I35, P42, L47, I50, and L53 build the hydrophobic core of CopR. Residues E2, F5, I44, K45, L47, L58, V59, and L62 are located on the protein surface and are suggested to be part of the dimeric interface. The real conformation of the fifth helix involving residues L58–I63 may—due to the uncertainties in the sequence alignment—differ from the model.

Based on experimental footprinting data (Steinmetzer and Brantl, 1997), the CopR homology model and the crystal structure of the 434 c1 repressor-DNA complex, a model for the complex of CopR with the DNA target was generated (Figure 3A). To test the function of aa involved in sequence and non-sequence specific DNA recognition as well as aa important for correct protein folding, site-directed mutagenesis was employed. CD measurements of CopR variants were carried out to detect structural changes resulting from the mutations. In addition, dimerization was monitored by glutaraldehyde cross-linking and analytical ultracentrifugation. This approach allowed to localize the predicted HTH motif between aa 18 and 37 and to determine two aa within the recognition helix that make specific contacts with the DNA, R29 and R34 (Steinmetzer et al., 2000b; Figure 3A). Variants R29Q and R34Q showed only non-specific DNA binding at very high (micromolar) concentrations, while the protein structure was not affected. Furthermore, mutations of aa predicted to be involved in non-specific binding of the DNA backbone (S28T and K10Q) led to decreased binding affinity while maintaining selectivity. Additionally, substitution of aa necessary for proper folding—E36 and F5—caused significant structural changes. Taken together, these data support the model of CopR as a HTH protein that belongs to the  $\lambda$  repressor superfamily and uses  $\alpha$ -helix III as recognition helix.

To verify the model predictions on aa involved in dimerization of CopR monomers, a combination of site-directed mutagenesis, EMSA, dimerization studies using sedimentation equilibrium centrifugation and CD measurements was used (Steinmetzer et al., 2000a). This allowed locating the dimeric interface between aa I44 and L62 (Figure 3A). As aa F5 situated at the N-terminus is needed for proper folding, it could not be unambiguously assigned to the dimeric interface. CD measurements at protein concentrations below the  $K_{\text{Dimer}}$

value demonstrated that the monomer of CopR is folded. As in the first analysis, double and triple variants were constructed that all exhibited drastically increased dimerization constants, the analysis was complemented with single variants in the dimeric interface (Steinmetzer et al., 2002b). DNA binding and dimerization constants were calculated, urea-induced denaturation experiments were applied to evaluate the *in vitro* stability, and CD spectra of all mutated CopR proteins were measured. Variants I44D, L58D, V59S, and L62D had 4- to 50-fold increased  $K_{\text{Dimer}}$ -values and bound the CopR operator only non-specifically. Thereby, the substitution of aa L58 or L62 that were predicted to form several close interface contacts severely diminished dimerization, while mutation of the weakly interacting aa V59 did not significantly affect dimerization. Whereas the CD spectra did not show drastic structural alterations, the denaturation data revealed that the four variants unfold differently compared with the wild-type. Our results reveal that the four analyzed aa are engaged in dimerization as well as in folding of the monomer, i.e., they stabilize the monomer and, in addition, the dimeric interface (Steinmetzer et al., 2002b). Possibly, for economic reasons, some aa have dual functions in a small protein like CopR. Our data obtained with the four variants carrying single amino acid exchanges indicate that conformational changes are indeed necessary for dimerization. Furthermore, we observed that a single aa can on the one hand contribute to intra-monomeric contacts, when the protein is present as a monomer, and on the other hand contribute to inter-monomeric contacts when the protein dimerizes.

### Structure of the DNA and Shape of the Protein in the CopR-DNA Complex

To determine the DNA conformation in the CopR-DNA complex, a combination of hydroxyl radical footprinting and fluorescence resonance energy transfer (FRET) measurements (Steinmetzer et al., 2002a) was employed. The footprints of CopR covered in total 29 bp and showed three defined areas of protection for each strand. This is comparable with the results obtained for the  $\lambda$ -repressor, 434-repressor, and for phage  $\Phi$ 105 repressor from *B. subtilis* (Tullius and Dombroski, 1986; Van Kaer et al., 1989; Ramesh and Nagaraja, 1996). The area of protection was significantly larger than that calculated earlier by chemical interference experiments, where the distance between the outermost contacts made up 17 bp (Steinmetzer and Brantl, 1997). Protected sites I and II were consistent with the previously identified contacted sites I and II. By contrast, the outer site III had not been identified before. For site III of the bottom strand, protection was weaker. This confirms our former observation that the interaction between CopR and the DNA is slightly asymmetric and also reflects the imperfect symmetry of the operator sequence (Steinmetzer and Brantl, 1997). FRET measurements revealed a bending angle of 20–25° for the DNA around the CopR protein, which is similar to that observed in the 434 c1 repressor-DNA complex and the  $\lambda$  c1 repressor-DNA complex. Furthermore, sedimentation velocity experiments demonstrated an extended shape of CopR dimers which accounts for the relatively large protection area detected with hydroxyl



radical footprinting. To determine the global shape of the DNA in complex with CopR, FRET experiments with two DNA fragments were performed: A 19-bp DNA-fragment comprises only the minimal operator sequence (+2 bp for stabilization) and a 34-bp-fragment includes also the outer contact sites. For both fragments, bending angles of 20–25° were measured. This demonstrates that the center of the DNA bending is within the 17-bp sequence constituting the minimal operator and that the additional outer base contacts did not increase the DNA bending beyond 40–50°. Both outer binding sites do not add more than 10–15° to the overall bent. A slight bent is also in agreement with the fact that no hypersensitive sites were observed in hydroxyl radical footprinting indicating that CopR binding does not cause a drastic distortion of the DNA backbone. Analytical ultracentrifugation revealed that CopR dimers have an extended shape with a size of 8.4 nm for the fully hydrated protein. Due to this extended shape, only a gentle bending of the DNA is needed to enable CopR to make additional contacts outside of its 17-bp operator that reinforce the protein-DNA interaction. The CopR operator contains—similar to the operator of the 434 repressor—two TG-steps 11 bp apart that may constitute bending points by providing the flexibility required for the conformational changes of the DNA (Tzou and Hwang, 1999). As the CopR model does not include the 29 C-terminal aa, it can be assumed that these residues contact the outer binding sites III. Interestingly, variant CopR $\Delta$ 27 lacking the 27 C-terminal aa has a 10-fold increased  $K_D$  value (3.8 nM instead of 0.4 nM) for the CopR-DNA complex (Kuhn et al., 2000). This corroborates the formation of additional contacts between aa of the full-length CopR-C-terminus and the DNA backbone.

### The C-terminus of CopR is Structured and Important for Protein Stability

Previous results showed that the C-terminal 27 aa of CopR were neither necessary for DNA binding nor for dimerization (Steinmetzer et al., 1998, 2000b). However, CopR $\Delta$ 27 was 5-fold impaired in copy number control *in vivo* compared to both the wild-type and CopR $\Delta$ 20. Interestingly, the C-terminus of CopR is very acidic comprising 10 Glu and one Asp residues. Therefore, a series of CopR variants truncated at the C-terminus were investigated for their half-life *in vivo* as well as for dimerization, DNA binding, structure and stability *in vitro* (Kuhn et al., 2000). The last 28 aa were apparently not required for DNA binding and dimerization, although the  $K_D$  was 10-fold increased for CopR $\Delta$ 27. Progressive deletions from the C-terminus significantly shortened the half-life of CopR: The half-life decreased from 42 min (wild-type CopR) over 24 min (CopR $\Delta$ 7),  $\approx$ 4.75 min (CopR $\Delta$ 20), to  $\approx$ 0.3 min (CopR $\Delta$ 27). Guanidine-HCl denaturation assays corroborated that variants with shortened half-lives were also less stable *in vitro*. These results indicate that the C-terminus of CopR is required for protein stability. Amino acid substitutions within the C-terminus indicated that neither length nor charge is important for stabilization. CD measurements revealed that the C-terminus of CopR that contains alternating hydrophilic and hydrophobic aa residues is structured and forms a  $\beta$ -strand (Kuhn et al., 2000). Further analysis of the stabilizing motifs within the

C-terminus (Kuhn et al., 2001) showed that both the wild-type (QVTLELEME, **Figure 3A**) and an artificial (QVTVTVTVT)  $\beta$ -strand structure (variant CopRVT) between aa 76 and 84 stabilized the corresponding protein derivatives. By contrast, replacement of the  $\beta$ -strand by an  $\alpha$ -helix or an unstructured sequence significantly or moderately destabilized the protein. A second stabilization motif was identified in the 7 C-terminal aa, as their deletion from CopR or CopRVT reduced the half-life of the corresponding pIP501 derivatives to  $\approx$ 50% (Kuhn et al., 2001). Our hypothesis is that the structured C-terminus of CopR interacts with other aa sequences in the core protein, thereby preventing its proteolytic degradation.

Surprisingly, variant CopR $\Delta$ 20 with a 10-fold reduced half-life was fully functional *in vivo* in copy number control. The intracellular concentration of this variant was with 1  $\mu$ M 15-fold lower than that of wild-type CopR (Kuhn et al., 2000). Why does wild-type CopR have such a long half-life, if a half-life of 4.75 min is completely sufficient for proper control? de la Hoz and colleagues investigating CopS from related plasmid pSM19035 found that the *copS* promoter is 8-fold down-regulated by the plasmid encoded  $\omega$ - protein (de la Hoz et al., 2000). They suggested that  $\omega$ - might represent a global regulator linking copy-number control with better than random segregation of pSM19035. As pIP501 derivatives lacking  $\omega$  did not display defects in replication control (Brantl and Behnke, 1992b), an  $\omega$ -homolog is apparently not required for replication control of pIP501. An 8-fold down-regulation of *copR* would still result in an intracellular CopR concentration of  $>2 \mu$ M, i.e., twice the amount determined for CopR $\Delta$ 20. In case  $\omega$  were included in pIP501 replication control and repressed *copR* 8-fold, the long CopR half-life would still ensure that sufficient CopR molecules are present to warrant proper control.

### Evolution of CopR Resulted in Maximal DNA Binding Affinity

When pIP501 evolved in its original host, *S. agalactiae*, selection was, apparently, for a low, but not the lowest possible, copy number, that was optimal under the conditions experienced by this bacterial host. This assumption is based on the independent *in vivo* selection of three almost identical (in their core sequences) operators of the related streptococcal plasmids pIP501, pSM19035 and pAM $\beta$ 1. CopR, CopS, and CopF have similar Cop operators with identical binding sites I and II. Only the spacer regions of the *copR* and *copS* operator differ (G244A and T247A), and the flanking sequences of the *copR* and *copF* binding sites display two nt exchanges (T236G and A260G).

One instrument to adjust the copy number of pIP501 is the  $K_D$  value of the CopR-operator DNA complex. Based on the data summarized above we wondered if the *copR* operator found in nature (in pIP501) was optimized for strong DNA binding or if it would be possible to select an operator sequence that is bound more efficiently by CopR and, if yes, how such an operator would behave *in vivo*. To this end, we employed a SELEX experiment with *copR* operator sequences of different lengths combined with subsequent EMSAs with mutated operator fragments, copy-number determinations, and *in vitro* transcription (Freede and



Brantl, 2004). Four experiments were performed: SELEX 1 with a randomized 7-bp spacer region, SELEX 2 with a randomized 17-bp fragment spanning the minimal operator, SELEX 3 with a longer operator (30 bp), and SELEX 4 with randomized 5-bp operator flanking regions. Our results demonstrate that the optimal spacer sequence between the two CopR binding sites comprises 7 bp, is AT rich and requires an A/T and T at the 3' positions. By contrast, broad variations in the sequences flanking the minimal 17-bp operator did not affect CopR binding. These results show that the sequence differences between the *copR*, *copS*, and *copF* operator can be neglected. SELEX 2 for the minimal 17-bp *copR* operator yielded the same sequences as *in vivo* selection except that the completely symmetrical operator was found, too. Three simultaneous nucleotide exchanges outside the bases directly contacted by CopR selected in SELEX 3 did only slightly affect CopR binding *in vitro* or copy numbers *in vivo*. Therefore, we can conclude that *in vivo* evolution of the *copR* operator sequence was for maximal binding affinity.

### Transcriptional Repressor CopR Acts by Inhibiting RNA Polymerase Binding

To investigate the complexes formed by the *B. subtilis* RNA polymerase (RNAP) at the *repR* promoter pII and to elucidate the mechanism exerted by CopR to repress transcription, a combination of DNase I footprinting, EMSA and KMnO<sub>4</sub> footprinting was used (Licht et al., 2011). As shown by DNase I footprinting, the binding sites for CopR and RNAP overlap. EMSA confirmed that CopR and *B. subtilis* RNAP can not bind simultaneously. Instead, they compete for binding at promoter pII. Apparently, CopR prevents the access of RNAP to the promoter region by steric exclusion. We assume that CopR competes with the  $\alpha$ CTD of the RNAP. Additionally, KMnO<sub>4</sub> footprinting experiments revealed that prevention of open complex formation at pII does not further increase the repression effect of CopR. Furthermore, CopR-operator complexes were 18-fold less stable than RNAP-pII complexes in competition assays. However, due to its higher intracellular concentration CopR can effectively compete with RNAP for binding to the same region, where promoter and operator overlap. What are the consequences for copy number control? The half-lives of both CopR-pII and RNAP-pII complexes provide the time window for regulation. As CopR is produced constitutively and has a much higher intracellular concentration than the RNAP, repression can occur quickly inspite of the long half-life of the RNAP-pII

complex. However, upon cell division the CopR concentration decreases, the repressor can be displaced by the RNAP—due to the much shorter half-life of the CopR-DNA complexes—and transcription of *repR* mRNA will be resumed immediately.

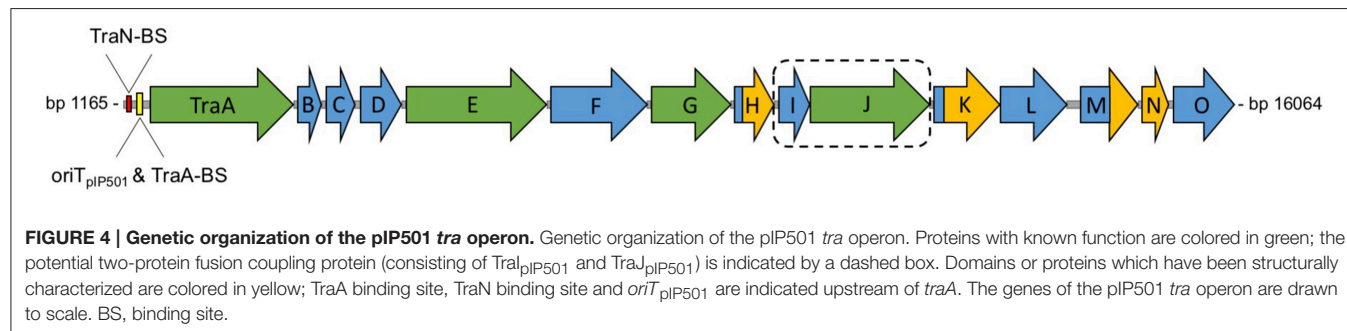
## PIP501 CONJUGATIVE TRANSFER

pIP501 encodes a Gram-positive T4SS, whose key characteristics include the lack of a putative inner membrane transport channel owed to the different membrane composition of Gram-positive organisms and the lack of a third putative conjugative ATPase, a VirB11-like protein (Bhatty et al., 2013). The whole T4SS is encoded by the *tra* operon coding for 15 putative Tra proteins, seven of these show sequence or structural homology with Vir proteins of the Gram-negative prototype T4SS from *Agrobacterium tumefaciens* (Figure 4). Expression of the *tra* operon is controlled by the transfer initiator protein, TraA.

### Biochemical Characterization of the TraA Relaxase

The TraA protein belongs to the family of IncQ-type relaxases, which includes the relaxases of the Gram-positive plasmids pGO1, pSK41, and pMRC01 as well as those of plasmids RSF1010, pSC101, and pTF1 of Gram-negative bacterial origin. The prototype of this relaxase family is the MobA protein encoded by the mobilizable plasmid RSF1010. MobA is a multifunctional protein consisting of an N-terminal relaxase domain and a C-terminal DNA primase domain (Scherzinger et al., 1991; Henderson and Meyer, 1996).

To confirm the postulated relaxase activity of TraA, supercoiled plasmid pVA2241 which contains a 309-bp fragment encompassing *oriT*<sub>pIP501</sub> (Wang and Macrina, 1995) was used as a substrate in an *in vitro* cleavage assay with purified TraA protein. TraA sequence-specifically cleaved the *oriT*<sub>pIP501</sub> containing supercoiled DNA (Kurenbach et al., 2002). TraA relaxase activity was optimal between 42 and 45°C, with the reactions being less efficient at temperatures below 37°C. TraA-mediated cleavage of supercoiled DNA was strictly dependent on Mg<sup>2+</sup> or Mn<sup>2+</sup>. Mg<sup>2+</sup> optimum was 5 mM, and optimal Mn<sup>2+</sup> concentration was 10 mM. As was the case with MobM of pMV158 (Guzman and Espinosa, 1997), the TraI-TraJ *oriT* complexes from plasmid RP4 (Pansegrau et al., 1990), TrwC from R388 (Llosa et al., 1995), and TraI of F (Matson and Morton, 1991), the maximum amount of form FII (relaxed plasmid form)



produced by TraA was about 55%. Total DNA relaxation was never obtained.

Interestingly, the N-terminal part of TraA comprising the first 293 aa also cleaved supercoiled *oriT*<sub>pIP501</sub> containing DNA, albeit less efficiently (Approximately 25% conversion) than the full-length protein. These data coincide with those of MobA from plasmid RSF1010. Experiments with a C-terminally truncated MobA protein demonstrated that MobA-dependent *oriT* nicking activity resides within the first 34% (243 aa) of the 78-kDa MobA protein (Scherzinger et al., 1992).

## pIP501 *tra* Operon Expression is Not Growth-Phase Dependent

The compact organization of the pIP501 *oriT* region is similar to that of rolling-circle-replicating plasmid pMV158, which was shown to be efficiently mobilized by pIP501 (van der Lelie et al., 1990; Kurenbach et al., 2003). The two regions are similar, meaning that the *oriT* *nic*-region, where the relaxase binds to its cognate DNA (Grohmann et al., 1999), lies within the respective promoter region (Fariás et al., 1999). This configuration suggests autoregulation of the putative pIP501 *tra* operon consisting of the genes *traA* to *traO* (Figure 4) by the DNA relaxase TraA.

To study co-transcription of *traA* to *traO*, we conducted Reverse Transcription PCR (RT-PCR) with RNA isolated from *E. faecalis* (pIP501) cells harvested during mid exponential growth phase. Primer pairs were selected to amplify two successive genes of the *tra* region. RT-PCR resulted in products of the expected size (Kurenbach et al., 2002, 2006). We also tested for the existence of transcription products beyond *traO* using primers which would generate a *traO/copR* product of 480 bp. Using RNA as template, the respective product was never observed. Transcription of the pIP501 *tra* operon appears to be terminated by a strong rho-independent transcriptional terminator (Kurenbach et al., 2006).

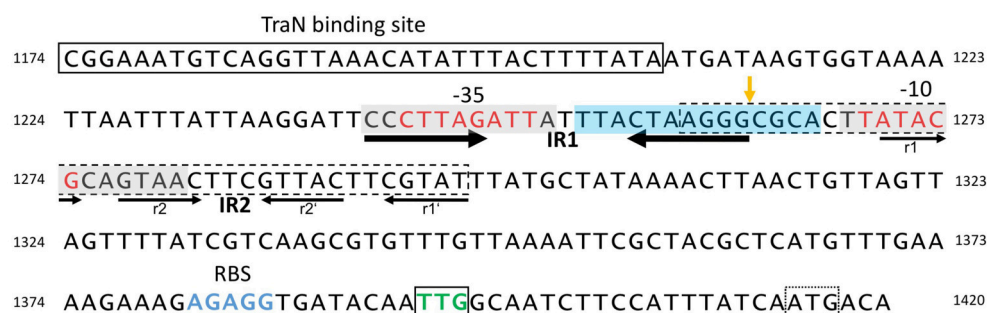
To test the potential impact of the growth phase on the transcription of the *tra* genes, total RNA from *E. faecalis*

(pIP501) was isolated at three different time-points: in the early exponential, the mid exponential, and the stationary growth phase ( $OD_{600} = 1.0$ ). First, we looked if the *tra* genes are transcribed in all three growth phases. The selected RT-PCR amplicons from *traC* to *traD*, *traF* to *traG*, and *traM* to *traN* were generated with RNA from all three time-points (Kurenbach et al., 2006). Semi-quantitative RT-PCRs were carried out for *traA* to *traB*, and for *traM* to *traN* to study the transcription levels of different *tra* genes under differing physiological conditions. As a control, the constitutively expressed GAP-DH gene was amplified by RT-PCR, with RNA from *E. faecalis* cells harvested at the respective time-points as template. Densitometric analysis of the amplification products did not show any significant differences with respect to the growth phase, the same picture was obtained, as expected, for the constitutively expressed GAP-DH (Kurenbach et al., 2006).

However, we cannot exclude that *tra* gene transcription declines at a later stage in stationary phase, as we have seen slightly lower transfer frequencies (2- to 3-fold decrease) for donors and/or recipients at high cell densities ( $OD_{600} > 1$ ) (Kurenbach et al., 2006). However, a phenomenon like “F2 phenocopies,” meaning that  $F^+$  cells get transfer-deficient in stationary phase (Hayes, 1964), was not observed. Transcription of several F-encoded *tra* genes decreases in mid-exponential or stationary phase, which is in agreement with a rapid decrease in transfer frequency in mid-exponential phase (Frost and Manchak, 1998). We conclude that the pIP501 *tra* genes are transcribed during the whole growth cycle of *E. faecalis* and that their level of expression does not depend on the growth phase.

## TraA Relaxase Binds to the *P<sub>tra</sub>* Promoter

The compact structure of the pIP501 *oriT* region (Figure 5), in the sense that the *P<sub>tra</sub>* -10 and -35 boxes overlap with the left half repeat of inverted repeat structures (IR-1 and IR-2), likely representing the TraA recognition and binding site (Kopeck et al., 2005), suggests autoregulation of the *tra* operon by TraA relaxase. To study TraA binding to the *P<sub>tra</sub>* promoter, three



**FIGURE 5 | Structure of the pIP501 *oriT* region on DNA level.** *oriT* region of pIP501 (nucleotides 1174–1420, GenBank L39769.1). Putative -10 and -35 promoter elements are colored in red; the TraN binding site is boxed. The perfect IR (IR1) is indicated with thick arrows; the second, imperfect IR (IR2) is marked with thin arrows—its repeating segments are individually named. Both repeats have the potential to generate a secondary structure. The areas that showed to be either protected from (gray box, mainly on the cleaved strand) or hypersensitive to (blue box, mainly on the non-cleaved strand) DNaseI cleavage in TraA footprinting assays (Kurenbach et al., 2006) are marked. *oriT*<sub>pIP501</sub> core sequence (GenBank L39769.1, bp 1259–1296) is indicated with a dashed box; *oriT* *nic*-site is shown by an orange vertical arrow. *tra* operon ribosomal binding site (RBS) is colored in blue; the start codon of the first *tra* gene (*traA*, TTG) is boxed and colored in green; a second, potential start codon (ATG) is marked with a dotted box.

DNA fragments were selected, the first comprising the –35 and –10 region, the second only the –35 region, and the third the –10 region alone. The shortest N-terminal TraA portion exhibiting relaxase activity, TraAN<sub>246</sub> (Kopec et al., 2005), was used in band shift assays with ds oligonucleotides comprising the different parts of the *P<sub>tra</sub>* promoter. Applying increasing TraAN<sub>246</sub> concentrations to the –10 fragment, we detected one retarded DNA–protein complex. Binding affinity for the –35 region and for the whole promoter region was weaker than for the –10 region fragment. This could be due to presence of the complete left half repeat of IR-2 in the –10 region fragment. This complete left half repeat was present in all ss oligonucleotides that bound TraAN<sub>246</sub> and TraA. An oligonucleotide similar to the –10 region fragment, but additionally comprising the right half repeat, resulted in similar binding affinity (Kopec et al., 2005). For all tested promoter fragments, TraA exhibited similar binding affinities as its N-terminal domain TraAN<sub>246</sub>. We conclude that TraA relaxase binds to the *P<sub>tra</sub>* promoter region and that only the N-terminal TraA relaxase portion, TraAN<sub>246</sub>, is required for efficient binding (Kopec et al., 2005).

DNase I footprinting with a 250-bp DNA fragment encompassing *P<sub>tra</sub>* and the complete IR-1 and IR-2 sequences demonstrated protection of the –35 and –10 region, with hypersensitive sites on the non-cleaved strand in vicinity of the *nic* site, at the *nic* site and two nucleotides 3' of the –10 region (Kurenbach et al., 2006). DNase I protection on the cleaved strand extended eight nucleotides to the *nic* site, the *nic* site itself appeared as hypersensitive site. The DNase I hypersensitive sites are likely due to a conformational change of the *oriT* region induced by TraA binding, making the DNA better accessible for DNase I attack.

We have demonstrated that the left half repeats of IR-1 and IR-2 are the preferential binding sites for TraA. We postulate that binding of TraA to its target DNA is required for recognition and cleavage of DNA at the 5'-GpC-3' dinucleotide in the *nic* site, which would remain accessible to the enzymatic activity of TraA.

## Expression of the *tra* Genes is Controlled by TraA Relaxase

To confirm that TraA binding to the *P<sub>tra</sub>* promoter region affects promoter activity, we cloned the promoterless *lacZ* gene in plasmid pQF120 under control of *P<sub>tra</sub>*. *E. coli* cells with the construct, pQF120-*P<sub>tra</sub>::lacZ*, gave blue colonies on LB X-Gal plates and generated  $\beta$ -galactosidase activity of 401 Miller units (Kurenbach et al., 2006). The impact of *traA* expression *in trans* on *P<sub>tra</sub>* activity was studied by co-transformation of *E. coli* with pQF120-*P<sub>tra</sub>::lacZ* and pACYC184-*P<sub>tac</sub>::GST-traA* which expresses *traA* under control of the *tac* promoter. Upon induction of *traA* expression  $\beta$ -galactosidase activity dropped to 6 Miller units. As a control, the effect of co-resident pACYC184-*P<sub>tac</sub>::GST* on *P<sub>tra</sub>* activity of pQF120-*P<sub>tra</sub>::lacZ* was analyzed. No significant change in  $\beta$ -galactosidase activity (407 Miller units) was observed. The data clearly demonstrated that the *tra* operon is regulated at the transcriptional level by TraA relaxase (Kurenbach et al., 2006).

For Mob, the mobilization protein encoded by the mobilizable broad-host-range plasmid pBBR1, whose *nic* site is identical with that of pMV158, autoregulation by binding of Mob to its promoter region overlapping with *oriT* has also been demonstrated (Szpirer et al., 2001).

Autoregulation of *tra* gene expression mediated by the transfer initiator protein, the DNA relaxase, seems to be an efficient mechanism to shut down plasmid transfer at a very early stage of conjugation, and is likely destined to obtain an optimum balance between the maximum transfer potential and the lowest burden for the host.

## Two Conjugative ATPases Show Non-specific DNA Binding Activity

The pIP501 *tra* region comprises a Gram-positive conjugative T4SS, encoding, like most of the related systems in Gram-positive bacteria, two ATPases, VirB4-like ATPase, TraE, and VirD4-like coupling protein, TraJ (for a recent review on pIP501 T4SS see Goessweiner-Mohr et al., 2013, 2014a). However, in contrast to coupling proteins from other T4SSs from Gram-negative and Gram-positive bacteria alike, pIP501 appears to encode the first coupling protein consisting of two proteins, the TraJ protein and the TraI protein. TraJ has ATP-binding and low ATPase activity *in vitro*, and the membrane-associated TraI protein, encoded immediately upstream of *traJ* in the *tra* operon (Figure 4), presumably recruits TraJ via protein-protein interaction to its putative site of action, the cytoplasmic membrane.

Already in 2002, Llosa and coworkers postulated that during conjugative plasmid transfer, ss-plasmid DNA is “actively pushed into the recipient cell by action of the coupling protein” (Llosa et al., 2002). In the case of pIP501, the energy for this process could be generated by ATP hydrolysis, mediated by TraJ. EMSAs with purified TraJ on ssDNA and dsDNA containing the minimal *oriT*<sub>pIP501</sub> region (GenBank L39769.1, bp 1259–1296) or random DNA of the same size with no similarity to *oriT*<sub>pIP501</sub> were performed. The DNA substrates were 42 bases or 42 bp long; the random 42-mer lacked the ability to form a hairpin-like secondary structure (Kopec et al., 2005), one of the prototypical characteristics of *oriT* regions. TraJ bound non-sequence specifically to both ssDNA substrates, whereas binding to dsDNA substrates was not observed, not even at very high TraJ concentrations (Arends, 2010). These observations are in agreement with the postulated function of TraJ as a conjugative coupling protein, connecting the relaxosome consisting of the TraA relaxase covalently bound to the 5'-end of the processed ss pIP501 DNA with the mating pair formation complex.

The VirB4-like ATPase, TraE, which showed higher *in vitro* ATPase activity than the coupling protein (Çelik, 2011) also bound non-sequence specifically to ss *oriT* 42-mer DNA and random 42-mer DNA in EMSAs performed similarly to those described above for TraJ. As demonstrated for TraJ, dsDNA was no substrate for TraE (Çelik, 2011). TraE and TraJ could be both actively involved in generating energy for the T4SS process, presumably each ATPase producing energy for different step(s)



in the T4S process. Details on these processes have not been unraveled so far.

## TraN is a Putative Transfer Repressor

### TraN Binds Sequence-Specifically to *oriT*<sub>pIP501</sub> DNA

TraN is a small (14.4 kDa) soluble cytoplasmic protein encoded by *traN*, the penultimate gene of the pIP501 *tra* operon. The structure of TraN was solved to 1.35 Å resolution. It contains an internal dimer fold with antiparallel β-sheets in the center and a HTH motif at both ends (Goessweiner-Mohr et al., 2012, 2014b).

Because TraN co-purified with DNA, we investigated if it can interact with radiolabelled ssDNA and dsDNA oligonucleotides. By applying the identical oligonucleotides as described for the EMSAs with TraE and TraJ, TraN showed only a slight shift for the ssDNA oligonucleotides, whereas the dsDNA fragments were significantly shifted. The random and the *oriT*<sub>pIP501</sub> containing oligonucleotide showed the same binding affinity (Goessweiner-Mohr et al., 2014b).

To search for a potential sequence-specific TraN binding site, we conducted EMSAs with dsDNA fragments encompassing the *oriT*<sub>pIP501</sub> and sequences upstream and downstream of this region. At high TraN concentrations, all DNA fragments were cooperatively shifted. A small but significant stepwise shift using an equimolar protein:DNA ratio was visible only for fragments comprising a common 149-bp sequence 5' of the *oriT* sequence, for which we postulate a preferred TraN binding site (Goessweiner-Mohr et al., 2014b).

To delimit the specific TraN binding site within the 149-bp sequence, we designed a new footprinting assay which is based upon 5'-to-3' exonuclease digestion. The TraN binding site was localized to a 34-bp sequence located 55 bp 5' of *oriT*<sub>pIP501</sub> *nic*-site. Interestingly, the TraN binding site has no direct or inverted repeats but is A/T rich (Goessweiner-Mohr et al., 2014b).

Thermal stability of TraN was studied in presence and absence of DNA with a ThermoFluor-based assay. The melting temperature (*T*<sub>m</sub>) of TraN alone amounted to 54.3°C; the binding of a non-specific (random) 34-mer dsDNA oligonucleotide raised the *T*<sub>m</sub> to 65.2°C, whereas addition of DNA containing the specific binding site increased *T*<sub>m</sub> to 70.4°C. The stabilizing effect indicates an enhanced binding affinity for the specific site compared with the random DNA (Goessweiner-Mohr et al., 2014b).

To determine whether there is a difference in the molar ratio of the TraN–DNA interaction between the random and the specific oligonucleotides, as well as to obtain information on the respective binding constants, isothermal titration calorimetry analyses with the oligonucleotide encompassing the binding site and the non-specific (random) oligonucleotides used in the ThermoFluor experiments were carried out. When titrating with non-specific DNA, two TraN molecules bound to one dsDNA fragment (in a 2:1 ratio), whereas, as expected, equimolar stoichiometry (1:1 ratio) was observed for the specific interaction. TraN was found to bind to the specific binding site exothermically with a binding constant of 10<sup>7</sup> M<sup>-1</sup> in comparison to endothermic binding to the non-specific sequence with a binding constant of 10<sup>5</sup> M<sup>-2</sup> (2:1 binding ratio; Goessweiner-Mohr et al., 2014b).

## The Crystal Structure of the TraN–DNA Complex has been Solved to High Resolution

Recently, we solved the 1.9 Å co-crystal structure of TraN bound to its specific 34-bp binding site upstream of the *oriT*<sub>pIP501</sub> *nic*-site, described above (Goessweiner-Mohr et al., in preparation). The binding mode postulated in Goessweiner-Mohr et al. (2014b) could be confirmed: “The recognition helices of the two mirrored HTH motifs enter two adjacent major grooves of the dsDNA binding site.” Furthermore, the tip of the loops between strands 2 and 3 as well as strands 5 and 6, which are close to the internal 2-fold axis, are interacting with the minor groove. While tied to its binding site, TraN slightly bends the dsDNA oligonucleotide used in the crystallization setup (Goessweiner-Mohr et al., in preparation).

## TraN is Not an Essential T4SS Protein

Very recently we generated a markerless *E. faecalis* JH2-2 (pIP501Δ*traN*) mutant by applying a two-step recombination technique developed for construction of mutants in *E. faecalis* (Kristich et al., 2007). Surprisingly, in standard *in vitro* mating tests we could demonstrate that TraN is not an essential T4SS protein but contrary to expectations, *traN* deletion resulted in an increase of pIP501 transfer efficiency.

## TraN Shows Structural Homology with Transcriptional Regulators: Potential Role of TraN in the T4S Process

In searches for proteins structurally similar to TraN we only found hits that resemble one half of the protein. Amongst others, the TraN fold resembles that of the N-terminal domain of transcriptional regulators of the MerR family (Goessweiner-Mohr et al., 2014b), for example a transcriptional activator from *Bacillus thuringiensis* (PDB entry 3gpv; New York SGX Research Center for Structural Genomics). Transcriptional activators of the MerR family comprise an N-terminal winged-helix DNA-binding domain and recognize the specific DNA site as a dimer where the recognition helices of the HTH motifs are inserted into two adjacent major grooves.

The dimerization motif of MerR proteins is completely distinct from the internal dimer configuration of TraN, which requires hydrophobic interactions within a barrel-like motif in its center. Contrarily to MerR family proteins, which contain a C-terminal effector-binding region (Brown et al., 2003), neither in TraN nor TraN-like proteins of related T4SSs such a C-terminal extension was found. All TraN-like proteins found are of enterococcal origin (from conjugative *E. faecalis* plasmids, pRE25 and pAMβ1, *E. faecium* plasmid pVEF3 and two genomically located TraN-like proteins from an *E. faecalis* and *Enterococcus italicus* strain), and their sequence is highly similar to that of TraN. All other proteins found (transposon or bacteriophage-encoded excisionases and MerR family proteins) have only a single TraN-like domain (Goessweiner-Mohr et al., 2014b).

Due to the structural similarity of TraN with MerR-like transcriptional regulators and the fact that *traN* deletion resulted in a 2 log increase of pIP501 transfer efficiency, we postulate that TraN could repress pIP501 transfer by regulating either expression of the pIP501 *tra* operon or TraA activity.



Although MerR-like proteins show only similarity to the fold of a single TraN domain, binding to DNA requires the formation of a homodimer (PDB entry 3gpv) which binds to two adjacent major grooves of dsDNA, as postulated for TraN. Expression of the pIP501 *tra* genes is already autoregulated by TraA relaxase (Kurenbach et al., 2006), TraA binds to the two left half repeats of IR1-1 and IR1-2 (Kurenbach et al., 2006) which overlap with the  $-10$  and  $-35$  box of the  $P_{tra}$  promoter respectively. Specific DNA recognition and binding is required for TraA-mediated site-specific nicking at the 5'-GpC-3' dinucleotide in the *nic*-site (nucleotide position 1262/1263, Acc. Nr. L39769) which thus will be accessible to the enzymatic activity of TraA (Kurenbach et al., 2006; see also **Figure 5**). TraN could act as an additional repressor of the *tra* operon by specifically binding to a 34-bp sequence located 55 bp upstream of the *nic*-site thereby inhibiting RNA polymerase from efficient transcription of the *tra* operon. We hypothesize that this negative regulation could be relieved by binding of putative interaction partners, e.g., TraE or TraJ (Abajy et al., 2007) to TraN, possibly as response to (i) the presence of potential recipient cells/mating partners or (ii) an assembled putative pIP501 T4SS core complex. Experimental studies on the mechanism of TraN regulation are in progress.

## SPECULATIONS ON CONTROL OF PIP501 TRANSFER GENE EXPRESSION

Tight control of *tra* gene expression is a general feature of mobile genetic elements from Gram-negative and Gram-positive bacteria alike, presumably to ensure that costly—referring to bacterial fitness—expression of multiple Tra proteins only takes place when the effort is worthwhile because potential recipients are present or more generally speaking the environmental conditions allow efficient plasmid transfer. Different modes of controlling conjugative transfer are known: The well-characterized Gram-negative conjugative broad-host-range *IncP* plasmids, F plasmid and F-related plasmids, such as R1 and R100, have a very complex regulation system controlling expression of Tra factors at transcriptional and translational level involving not only plasmid-encoded factors but also host-factors (Zatyka et al., 1994, 1997; Taki et al., 1998; Adamczyk and Jagura-Burdzy, 2003; Will and Frost, 2006; Wong et al., 2012). Additionally, in case of TraJ from plasmid R1 and F-related plasmids, regulation of the transfer operon via a sense/antisense RNA system has been shown (Koraimann et al., 1991, 1996; Mark Glover et al., 2015). For Gram-positive bacteria, the sex-pheromone-responsive enterococcal plasmids, particularly pCF10, are those with the best studied regulatory processes controlling conjugative transfer (Tanimoto et al., 1996; Muscholl-Silberhorn, 2000; Dunny, 2007, 2013; Folli et al., 2008).

None of the known conjugation control systems fits to what we have observed for broad-host-range plasmid pIP501. pIP501 *tra* gene expression seems to be always on, independent of the growth phase of the host (Kurenbach et al., 2006) and presence of potential recipients, presumably at a low basic

level. *tra* gene expression was shown to be controlled by the transfer initiator protein, TraA, which regulates its own synthesis and that of the other Tra factors by binding to the  $P_{tra}$  promoter overlapping with *oriT* (Kurenbach et al., 2006; see also **Figure 5**).

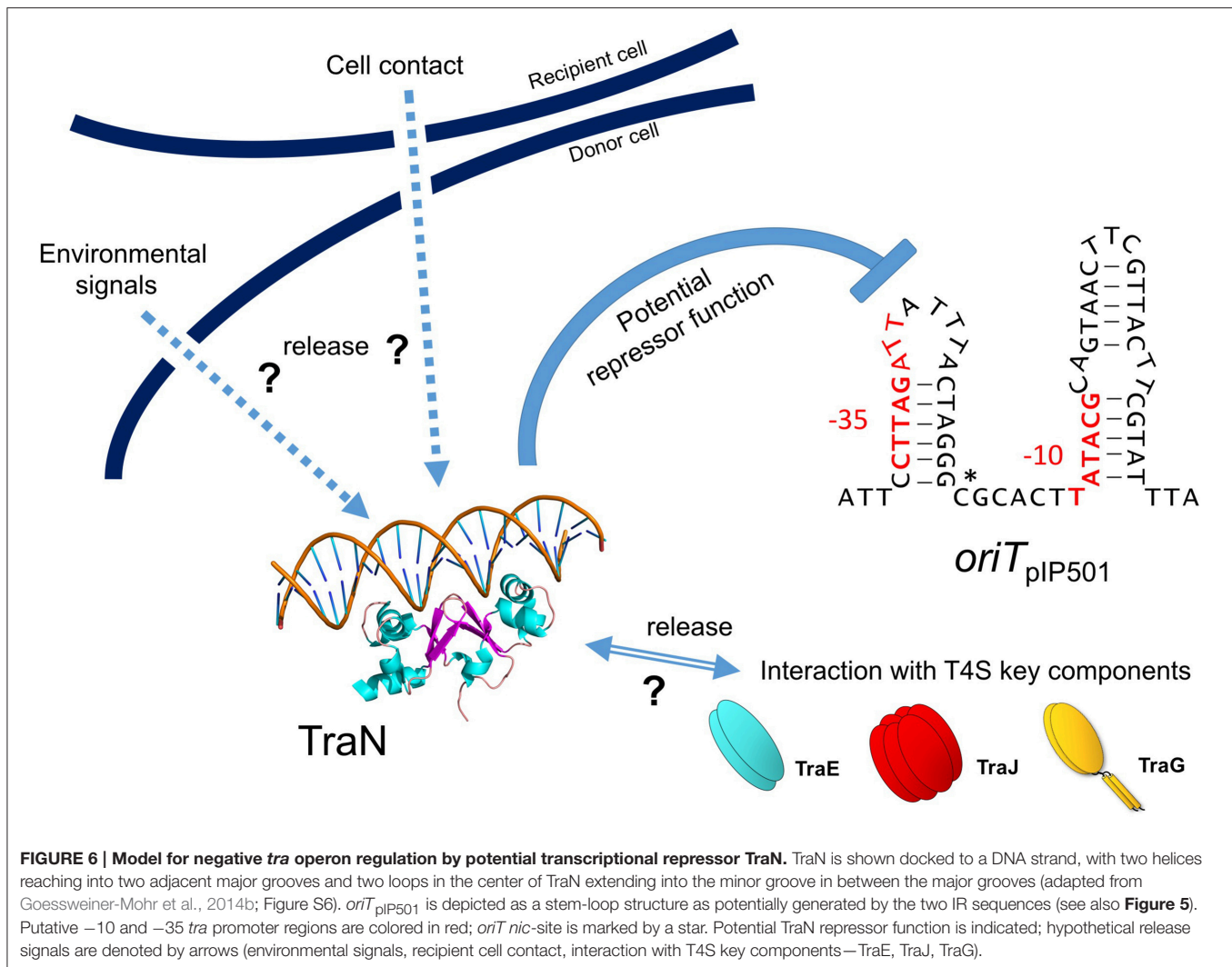
Recently, we detected binding of another Tra factor, TraN, to a region 55 bp upstream of the *oriT nic*-site (Goessweiner-Mohr et al., 2014b; **Figure 5** in this article). This TraN-binding site is located only 35 bp upstream of the  $-35$  region of the  $P_{tra}$  promoter. Thus, we postulate that the *tra* operon might be negatively controlled by two proteins, the TraA relaxase binding to the  $-10$  and  $-35$  region of the promoter thereby leaving the *oriT nic*-site accessible for specific TraA cleavage and by the winged-helix-turn-helix DNA-binding protein TraN, binding to a unique operator site (present only once on the pIP501 genome) upstream of  $P_{tra}$  promoter. We postulate that TraA activity is blocked by binding of TraN upstream of the *oriT nic*-site. Either by (i) receiving environmental signals which could include the presence or already the contact of the donor cell with a potential recipient cell and/or (ii) by interaction of TraN with T4S key components, such as TraE, TraG or TraJ (binding to these proteins has been observed in the yeast two-hybrid system Abajy et al., 2007), TraN would be released from the DNA, likely resulting in a conformational change of the DNA in the vicinity of the TraA binding site triggering *nic*-cleavage by TraA. Our working model of pIP501 *tra* operon regulation is depicted in **Figure 6**.

A putative winged-helix-turn-helix DNA-binding protein, RctA from the symbiotic rhizobial megaplasmids, has been demonstrated to repress transcription of conjugative transfer genes of pRetCFN42d, the symbiotic plasmid (pSym) of *Rhizobium etli* (Pérez-Mendoza et al., 2005; Sepúlveda et al., 2008; Nogales et al., 2013).

The negative regulation of pIP501 *tra* gene expression exerted by two (putative) transcriptional regulators would be in agreement with the generally low transfer frequencies of pIP501, in the range of  $(2 - 8) \times 10^{-5}$  transconjugants per recipient for intraspecies *E. faecalis* matings (Arends et al., 2013; Fercher et al., 2016).

## CONCLUSIONS AND PERSPECTIVES

Conjugative transfer of diverse genetic traits, such as antibiotic or heavy metal resistance genes, virulence genes or genes conferring specific metabolic capabilities such as degradation of anthropogenic compounds is a natural process going on everywhere in nature at diverse transfer rates, as these naturally depend on the plasmids involved and on the habitat. Availability of nutrients and water, or in other words, good physiological conditions of donor and recipients, are generally accepted as conditions favoring horizontal plasmid transfer. Availability of colonizable surfaces for the microorganisms is another very important feature, as the close proximity of microorganisms in surface-associated communities, the so-called biofilms, increases the chances of horizontal gene exchange (Hausner and Wuertz, 1999; Sørensen et al., 2005; Madsen et al., 2012).



The observation that *tra* gene expression seems to be a tightly controlled process not only holds true for plasmids of the *Inc18* group, but seems to be a general feature of self-transmissible plasmids of diverse origin. In particular, the expression of relaxosome components seems to be tightly regulated, in many plasmids it is under autoregulatory control by the relaxase or relaxase accessory proteins. One of the most complex regulatory circuits controlling the production of relaxosome proteins has been deciphered in the prototype Gram-negative broad-host-range plasmid RK2. Zatyka and coworkers argued that the complex regulatory circuits involved in regulation of *IncPα* plasmid RK2 provide an autoregulatory way of ensuring production of enough relaxosome proteins without overburdening the host (Zatyka et al., 1994). Expression of the *tra* genes of F-plasmid is also tightly controlled by a number of factors, including among others, a plasmid-encoded activator and two autoregulators. One of them, TraM, is a component of the F relaxosome (Will and Frost, 2006). In all these plasmids the level and stringency of the

regulatory processes appear to be in good balance with the transfer potential of the host in order to reduce its fitness costs.

Detailed knowledge of these regulatory processes in Gram-positive bacteria is still scarce, thus challenging tasks of the coming years will be to unravel the internal as well as external environmental signals triggering plasmid transfer on the molecular level to develop more efficient interference strategies to efficiently reduce conjugative spread of antibiotic resistance genes.

With regard to replication of pIP501, the identification of a third (upper) regulatory level would be required to unravel why deletion of both regulatory components, CopR and RNAIII, does not show an additive effect. In the related plasmid pSM19035,  $\omega$  protein was found to be this central regulator (de la Hoz et al., 2000). However, in pIP501, no promoter has been detected so far preceding *orf*  $\omega$ .

Thus, for both, pIP501 replication and transfer, the most urgent questions to answer concern global regulatory processes

governing the success of pIP501-like multiple antibiotic resistance replicons in terms of maintenance and wide spread, particularly, in hospital environments.

## AUTHOR CONTRIBUTIONS

EG, NG, and SB contributed to writing of the manuscript. NG, SB designed the figures, all authors approved the final version of the manuscript.

## REFERENCES

- Abajy, M. Y., Kopec, J., Schiwon, K., Burzynski, M., Döring, M., Bohn, C., et al. (2007). A type IV-secretion-like system is required for conjugative DNA transport of broad-host-range plasmid pIP501 in gram-positive bacteria. *J. Bacteriol.* 189, 2487–2496. doi: 10.1128/JB.01491-06
- Adamczyk, M., and Jagura-Burdzy, G. (2003). Spread and survival of promiscuous IncP-1 plasmids. *Acta Biochim. Pol.* 50, 425–453. Available online at: [http://www.actabp.pl/pdf/2\\_2003/425.pdf](http://www.actabp.pl/pdf/2_2003/425.pdf)
- Arends, K. (2010). *Entwicklung von gfp-Basierten Monitoring Tools zur Verfolgung von Horizontalem Gentransfer und Studien zum T4SLS des Konjugativen Plasmids pIP501 aus Enterococcus faecalis*. Ph.D. thesis, Technical University, Berlin.
- Arends, K., Celik, E.-K., Probst, I., Goessweiner-Mohr, N., Fercher, C., Grumet, L., et al. (2013). TraG encoded by the pIP501 type IV secretion system is a two-domain peptidoglycan-degrading enzyme essential for conjugative transfer. *J. Bacteriol.* 195, 4436–4444. doi: 10.1128/JB.02263-12
- Bhatty, M., Laverde Gomez, J. A., and Christie, P. J. (2013). The expanding bacterial type IV secretion lexicon. *Res. Microbiol.* 164, 620–639. doi: 10.1016/j.resmic.2013.03.012
- Brantl, S. (1994). The *copR* gene product of plasmid pIP501 acts as a transcriptional repressor at the essential *repR* promoter. *Mol. Microbiol.* 14, 473–483. doi: 10.1111/j.1365-2958.1994.tb02182.x
- Brantl, S. (2014). Plasmid replication control by Antisense RNAs. *Microbiol. Spectr.* 2:PLAS-0001–2013. doi: 10.1128/microbiolspec.PLAS-0001-2013
- Brantl, S. (2015). Antisense-RNA mediated control of plasmid replication - pIP501 revisited. *Plasmid* 78, 4–16. doi: 10.1016/j.plasmid.2014.07.004
- Brantl, S., and Behnke, D. (1992a). Characterization of the minimal origin required for replication of the streptococcal plasmid pIP501 in *Bacillus subtilis*. *Mol. Microbiol.* 6, 3501–3510. doi: 10.1111/j.1365-2958.1992.tb01785.x
- Brantl, S., and Behnke, D. (1992b). Copy number control of the streptococcal plasmid pIP501 occurs at three levels. *Nucleic Acids Res.* 20, 395–400. doi: 10.1093/nar/20.3.395
- Brantl, S., and Behnke, D. (1992c). The amount of RepR protein determines the copy number of plasmid pIP501 in *Bacillus subtilis*. *J. Bacteriol.* 174, 5475–5478.
- Brantl, S., Behnke, D., and Alonso, J. C. (1990). Molecular analysis of the replication region of the conjugative *Streptococcus agalactiae* plasmid pIP501 in *Bacillus subtilis*. Comparison with plasmids pAM $\beta$ 1 and pSM 19035. *Nucleic Acids Res.* 18, 4783–4790. doi: 10.1093/nar/18.16.4783
- Brantl, S., Birch-Hirschfeld, E., and Behnke, D. (1993). RepR protein expression on plasmid pIP501 is controlled by an antisense RNA-mediated transcription attenuation mechanism. *J. Bacteriol.* 175, 4052–4061.
- Brantl, S., Kummer, C., and Behnke, D. (1994). Complete nucleotide sequence of plasmid pGB3631, a derivative of the *Streptococcus agalactiae* plasmid pIP501. *Gene* 142, 155–156. doi: 10.1016/0378-1119(94)90372-7
- Brantl, S., Nowak, A., Behnke, D., and Alonso, J. C. (1989). Revision of the nucleotide sequence of the *Streptococcus pyogenes* plasmid pSM19035 *repS* gene. *Nucleic Acids Res.* 17:10110. doi: 10.1093/nar/17.23.10110
- Brantl, S., and Wagner, E. G. (1994). Antisense RNA-mediated transcriptional attenuation occurs faster than stable antisense/target RNA pairing: an *in vitro* study of plasmid pIP501. *EMBO J.* 13, 3599–3607.

## ACKNOWLEDGMENTS

We apologize for not having mentioned all valuable contributions from colleagues in the field due to space restrictions. We thank K. Arends and C. Fercher for critical reading of the manuscript. Work in the Grohmann lab was supported by DLR grants 50WB1166 and 50WB1466 from Deutsches Zentrum für Luft und Raumfahrt, work in the Brantl lab was supported by grants BR1552/4-1 to 4-3, 6-1 to 6-3, and BR1552/8-1 from Deutsche Forschungsgemeinschaft.

- Brantl, S., and Wagner, E. G. (1996). An unusually long-lived antisense RNA in plasmid copy number control: *in vivo* RNAs encoded by the streptococcal plasmid pIP501. *J. Mol. Biol.* 255, 275–288. doi: 10.1006/jmbi.1996.0023
- Brantl, S., and Wagner, E. G. (1997). Dual function of the *copR* gene product of plasmid pIP501. *J. Bacteriol.* 179, 7016–7024.
- Brown, N. L., Stoyanov, J. V., Kidd, S. P., and Hobman, J. L. (2003). The MerR family of transcriptional regulators. *FEMS Microbiol. Rev.* 27, 145–163. doi: 10.1016/S0168-6445(03)00051-2
- Bruand, C., Le Chatelier, E., Ehrlich, S. D., and Jannière, L. (1993). A fourth class of theta-replicating plasmids: the pAM beta 1 family from gram-positive bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 90, 11668–11672. doi: 10.1073/pnas.90.24.11668
- Brzozowska, I., Brzozowska, K., and Zielenkiewicz, U. (2012). Functioning of the TA cassette of streptococcal plasmid pSM19035 in various gram-positive bacteria. *Plasmid* 68, 51–60. doi: 10.1016/j.plasmid.2012.01.010
- Ceglowski, P., Lurz, R., and Alonso, J. C. (1993). Functional analysis of pSM19035-derived replicons in *Bacillus subtilis*. *Gene* 136, 1–12. doi: 10.1016/0378-1119(93)90441-5
- Çelik, E. K. (2011). *Studien zur Aufklärung der Rolle von Proteinkomponenten für die Assemblierung eines Konjugativen Typ IV Sekretionssystems in Gram-Positiven Bakterien sowie Nachweis Konjugativer Plasmide und Antibiotika-Resistenzgene in Bodenproben*. Ph.D. thesis, Karl-Franzens-University Graz, Graz.
- de la Hoz, A. B., Ayora, S., Sitkiewicz, I., Fernández, S., Pankiewicz, R., Alonso, J. C., et al. (2000). Plasmid copy-number control and better-than-random segregation genes of pSM19035 share a common regulator. *Proc. Natl. Acad. Sci. U.S.A.* 97, 728–733. doi: 10.1073/pnas.97.2.728
- Dunny, G. M. (2007). The peptide pheromone-inducible conjugation system of *Enterococcus faecalis* plasmid pCF10: cell-cell signalling, gene transfer, complexity and evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 1185–1193. doi: 10.1098/rstb.2007.2043
- Dunny, G. M. (2013). Enterococcal sex pheromones: signaling, social behavior, and evolution. *Annu. Rev. Genet.* 47, 457–482. doi: 10.1146/annurev-genet-111212-133449
- Evans, R. P. J. R., and Macrina, F. L. (1983). Streptococcal R plasmid pIP501: endonuclease site map, resistance determinant location, and construction of novel derivatives. *J. Bacteriol.* 154, 1347–1355.
- Fariás, M. E., Grohmann, E., and Espinosa, M. (1999). Expression of the *mobM* gene of the streptococcal plasmid pMV158 in *Lactococcus lactis* subsp. *lactis*. *FEMS Microbiol. Lett.* 176, 403–410. doi: 10.1016/S0378-1097(99)00265-7
- Fercher, C., Probst, I., Kohler, V., Goessweiner-Mohr, N., Arends, K., Grohmann, E., et al. (2016). VirB8-like protein TraH is crucial for DNA transfer in *Enterococcus faecalis*. *Sci. Rep.* 6:24643. doi: 10.1038/srep24643
- Folli, C., Mangiarotti, L., Folloni, S., Alfieri, B., Gobbo, M., Berni, R., et al. (2008). Specificity of the TraA-DNA interaction in the regulation of the pPD1-encoded sex pheromone response in *Enterococcus faecalis*. *J. Mol. Biol.* 380, 932–945. doi: 10.1016/j.jmb.2008.05.058
- Frede, P., and Brantl, S. (2004). Transcriptional Repressor CopR: use of SELEX to study the *copR* operator indicates that evolution was directed at maximal binding affinity. *J. Bacteriol.* 186, 6254–6264. doi: 10.1128/JB.186.18.6254-6264.2004
- Frost, L. S., and Manchak, J. (1998). F-phenocopies: characterization of expression of the F transfer region in stationary phase. *Microbiology* 144(Pt 9), 2579–2587. doi: 10.1099/00221287-144-9-2579



- Goessweiner-Mohr, N., Arends, K., Keller, W., and Grohmann, E. (2013). Conjugative type IV secretion systems in gram-positive bacteria. *Plasmid* 70, 289–302. doi: 10.1016/j.plasmid.2013.09.005
- Goessweiner-Mohr, N., Arends, K., Keller, W., and Grohmann, E. (2014a). Conjugation in gram-positive bacteria. *Microbiol. Spectr.* 2:PLAS-0004-2013. doi: 10.1128/microbiolspec.PLAS-0004-2013
- Goessweiner-Mohr, N., Eder, M., Hofer, G., Fercher, C., Arends, K., Birner-Gruenberger, R., et al. (2014b). Structure of the double-stranded DNA-binding type IV secretion protein TraN from *Enterococcus*. *Acta Crystallogr. D Biol. Crystallogr.* 70, 2376–2389. doi: 10.1107/S1399004714014187
- Goessweiner-Mohr, N., Fercher, C., Abajy, M. Y., Grohmann, E., and Keller, W. (2012). Crystallization and first data collection of the putative transfer protein TraN from the Gram-positive conjugative plasmid pIP501. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 68, 1402–1405. doi: 10.1107/S174430911204184X
- Grohmann, E., Guzmán, L. M., and Espinosa, M. (1999). Mobilisation of the streptococcal plasmid pMV158: interactions of MobM protein with its cognate *oriT* DNA region. *Mol. Gen. Genet.* 261, 707–715. doi: 10.1007/s004380050014
- Guzman, L. M., and Espinosa, M. (1997). The mobilization protein, MobM, of the streptococcal plasmid pMV158 specifically cleaves supercoiled DNA at the plasmid *oriT*. *J. Mol. Biol.* 266, 688–702. doi: 10.1006/jmbi.1996.0824
- Hausner, M., and Wuerzt, S. (1999). High rates of conjugation in bacterial biofilms as determined by quantitative *in situ* analysis. *Appl. Environ. Microbiol.* 65, 3710–3713.
- Hayes, W. (1964). *The Genetics of Bacteria and their Viruses. Studies in Basic Genetics and Molecular Biology*. Oxford: Blackwell Scientific Publications.
- Heidrich, N., and Brantl, S. (2003). Antisense-RNA mediated transcriptional attenuation: importance of a U-turn loop structure in the target RNA of plasmid pIP501 for efficient inhibition by the antisense RNA. *J. Mol. Biol.* 333, 917–929. doi: 10.1016/j.jmb.2003.09.020
- Heidrich, N., and Brantl, S. (2007). Antisense RNA-mediated transcriptional attenuation in plasmid pIP501: the simultaneous interaction between two complementary loop pairs is required for efficient inhibition by the antisense RNA. *Microbiology* 153, 420–427. doi: 10.1099/mic.0.2006/002329-0
- Henderson, D., and Meyer, R. J. (1996). The primase of broad-host-range plasmid R1162 is active in conjugal transfer. *J. Bacteriol.* 178, 6888–6894.
- Kopeck, J., Bergmann, A., Fritz, G., Grohmann, E., and Keller, W. (2005). TraA and its N-terminal relaxase domain of the Gram-positive plasmid pIP501 show specific *oriT* binding and behave as dimers in solution. *Biochem. J.* 387, 401–409. doi: 10.1042/BJ20041178
- Koraimann, G., Koraimann, C., Koronakis, V., Schlager, S., and Högenauer, G. (1991). Repression and derepression of conjugation of plasmid R1 by wild-type and mutated *finP* antisense RNA. *Mol. Microbiol.* 5, 77–87. doi: 10.1111/j.1365-2958.1991.tb01828.x
- Koraimann, G., Teferle, K., Markolin, G., Woger, W., and Högenauer, G. (1996). The FinOP repressor system of plasmid R1: analysis of the antisense RNA control of *traJ* expression and conjugative DNA transfer. *Mol. Microbiol.* 21, 811–821. doi: 10.1046/j.1365-2958.1996.361401.x
- Kristich, C. J., Chandler, J. R., and Dunny, G. M. (2007). Development of a host-genotype-independent counterselectable marker and a high-frequency conjugative delivery system and their use in genetic analysis of *Enterococcus faecalis*. *Plasmid* 57, 131–144. doi: 10.1016/j.plasmid.2006.08.003
- Kuhn, K., Steinmetzer, K., and Brantl, S. (2000). Transcriptional repressor CopR: the structured acidic C terminus is important for protein stability. *J. Mol. Biol.* 300, 1021–1031. doi: 10.1006/jmbi.2000.3929
- Kuhn, K., Steinmetzer, K., and Brantl, S. (2001). Transcriptional repressor CopR: dissection of stabilizing motifs within the C terminus. *Microbiology* 147, 3387–3392. doi: 10.1099/00221287-147-12-3387
- Kurenbach, B., Bohn, C., Prabhu, J., Abudukerim, M., Szewzyk, U., and Grohmann, E. (2003). Intergenic transfer of the *Enterococcus faecalis* plasmid pIP501 to *Escherichia coli* and *Streptomyces lividans* and sequence analysis of its *tra* region. *Plasmid* 50, 86–93. doi: 10.1016/S0147-619X(03)00044-1
- Kurenbach, B., Grothe, D., Farias, M. E., Szewzyk, U., and Grohmann, E. (2002). The *tra* region of the conjugative plasmid pIP501 is organized in an operon with the first gene encoding the relaxase. *J. Bacteriol.* 184, 1801–1805. doi: 10.1128/JB.184.6.1801-1805.2002
- Kurenbach, B., Kopeck, J., Mägdefrau, M., Andreas, K., Keller, W., Bohn, C., et al. (2006). The TraA relaxase autoregulates the putative type IV secretion-like system encoded by the broad-host-range *Streptococcus agalactiae* plasmid pIP501. *Microbiology* 152, 637–645. doi: 10.1099/mic.0.28468-0
- Le Chatelier, E., Ehrlich, S. D., and Jannière, L. (1994). The pAM beta 1 CopF repressor regulates plasmid copy number by controlling transcription of the *repE* gene. *Mol. Microbiol.* 14, 463–471. doi: 10.1111/j.1365-2958.1994.tb02181.x
- Le Chatelier, E., Jannière, L., Ehrlich, S. D., and Canceill, D. (2001). The RepE initiator is a double-stranded and single-stranded DNA-binding protein that forms an atypical open complex at the onset of replication of plasmid pAMBeta 1 from Gram-positive bacteria. *J. Biol. Chem.* 276, 10234–10246. doi: 10.1074/jbc.M010118200
- Licht, A., Freede, P., and Brantl, S. (2011). Transcriptional repressor CopR acts by inhibiting RNA polymerase binding. *Microbiology* 157, 1000–1008. doi: 10.1099/mic.0.047209-0
- Llosa, M., Gomis-Ruth, F. X., Coll, M., and de la Cruz, F. (2002). Bacterial conjugation: a two-step mechanism for DNA transport. *Mol. Microbiol.* 45, 1–8. doi: 10.1046/j.1365-2958.2002.03014.x
- Llosa, M., Grandoso, G., and de la Cruz, F. (1995). Nicking activity of TrwC directed against the origin of transfer of the IncW plasmid R388. *J. Mol. Biol.* 246, 54–62. doi: 10.1006/jmbi.1994.0065
- Madsen, J. S., Burmølle, M., Hansen, L. H., and Sørensen, S. J. (2012). The interconnection between biofilm formation and horizontal gene transfer. *FEMS Immunol. Med. Microbiol.* 65, 183–195. doi: 10.1111/j.1574-695X.2012.00960.x
- Mark Glover, J. N., Chaulk, S. G., Edwards, R. A., Arthur, D., Lu, J., and Frost, L. S. (2015). The FinO family of bacterial RNA chaperones. *Plasmid* 78, 79–87. doi: 10.1016/j.plasmid.2014.07.003
- Matson, S. W., and Morton, B. S. (1991). *Escherichia coli* DNA helicase I catalyzes a site- and strand-specific nicking reaction at the F plasmid *oriT*. *J. Biol. Chem.* 266, 16232–16237.
- Muscholl-Silberhorn, A. B. (2000). Pheromone-regulated expression of sex pheromone plasmid pAD1-encoded aggregation substance depends on at least six upstream genes and a cis-acting, orientation-dependent factor. *J. Bacteriol.* 182, 3816–3825. doi: 10.1128/JB.182.13.3816-3825.2000
- Nogales, J., Blanca-Ordóñez, H., Olivares, J., and Sanjuán, J. (2013). Conjugal transfer of the *Sinorhizobium meliloti* 1021 symbiotic plasmid is governed through the concerted action of one- and two-component signal transduction regulators. *Environ. Microbiol.* 15, 811–821. doi: 10.1111/1462-2920.12073
- Pansegau, W., Balzer, D., Kruff, V., Lurz, R., and Lanka, E. (1990). *In vitro* assembly of relaxosomes at the transfer origin of plasmid RP4. *Proc. Natl. Acad. Sci. U.S.A.* 87, 6555–6559. doi: 10.1073/pnas.87.17.6555
- Pérez-Mendoza, D., Sepúlveda, E., Pando, V., Muñoz, S., Nogales, J., Olivares, J., et al. (2005). Identification of the *rctA* gene, which is required for repression of conjugative transfer of rhizobial symbiotic megaplasmids. *J. Bacteriol.* 187, 7341–7350. doi: 10.1128/JB.187.21.7341-7350.2005
- Ramesh, V., and Nagaraja, V. (1996). Sequence-specific DNA binding of the phage Mu C protein: footprinting analysis reveals altered DNA conformation upon protein binding. *J. Mol. Biol.* 260, 22–33. doi: 10.1006/jmbi.1996.0379
- Rosvoll, T. C. S., Pedersen, T., Sletvold, H., Johnsen, P. J., Sollid, J. E., Simonsen, G. S., et al. (2010). PCR-based plasmid typing in *Enterococcus faecium* strains reveals widely distributed pRE25-, pRUM-, pIP501- and pHTbeta-related replicons associated with glycopeptide resistance and stabilizing toxin-antitoxin systems. *FEMS Immunol. Med. Microbiol.* 58, 254–268. doi: 10.1111/j.1574-695X.2009.00633.x
- Scherzinger, E., Haring, V., Lurz, R., and Otto, S. (1991). Plasmid RSF1010 DNA replication *in vitro* promoted by purified RSF1010 RepA, RepB and RepC proteins. *Nucleic Acids Res.* 19, 1203–1211. doi: 10.1093/nar/19.6.1203
- Scherzinger, E., Lurz, R., Otto, S., and Dobrinski, B. (1992). *In vitro* cleavage of double- and single-stranded DNA by plasmid RSF1010-encoded mobilization proteins. *Nucleic Acids Res.* 20, 41–48. doi: 10.1093/nar/20.1.41
- Sepúlveda, E., Pérez-Mendoza, D., Ramírez-Romero, M. A., Soto, M. J., López-Lara, I. M., Geiger, O., et al. (2008). Transcriptional interference and repression modulate the conjugative ability of the symbiotic plasmid of *Rhizobium etli*. *J. Bacteriol.* 190, 4189–4197. doi: 10.1128/JB.00041-08
- Sørensen, S. J., Bailey, M., Hansen, L. H., Kroer, N., and Wuerzt, S. (2005). Studying plasmid horizontal transfer *in situ*: a critical review. *Nat. Rev. Microbiol.* 3, 700–710. doi: 10.1038/nrmicro1232



- Steinmetzer, K., Behlke, J., and Brantl, S. (1998). Plasmid pIP501 encoded transcriptional repressor CopR binds to its target DNA as a dimer. *J. Mol. Biol.* 283, 595–603. doi: 10.1006/jmbi.1998.2122
- Steinmetzer, K., Behlke, J., Brantl, S., and Lorenz, M. (2002a). CopR binds and bends its target DNA: a footprinting and fluorescence resonance energy transfer study. *Nucleic Acids Res.* 30, 2052–2060. doi: 10.1093/nar/30.9.2052
- Steinmetzer, K., and Brantl, S. (1997). Plasmid pIP501 encoded transcriptional repressor CopR binds asymmetrically at two consecutive major grooves of the DNA. *J. Mol. Biol.* 269, 684–693. doi: 10.1006/jmbi.1997.1083
- Steinmetzer, K., Hillisch, A., Behlke, J., and Brantl, S. (2000a). Transcriptional repressor CopR: amino acids involved in forming the dimeric interface. *Proteins* 39, 408–416. doi: 10.1002/(SICI)1097-0134(20000601)39:4<408::AID-PROT130>3.0.CO;2-0
- Steinmetzer, K., Hillisch, A., Behlke, J., and Brantl, S. (2000b). Transcriptional repressor CopR: structure model-based localization of the deoxyribonucleic acid binding motif. *Proteins* 38, 393–406. doi: 10.1002/(SICI)1097-0134(20000301)38:4<393::AID-PROT5>3.0.CO;2-H
- Steinmetzer, K., Kuhn, K., Behlke, J., Golbik, R., and Brantl, S. (2002b). Plasmid pIP501 encoded transcriptional repressor CopR: single amino acids involved in dimerization are also important for folding of the monomer. *Plasmid* 47, 201–209. doi: 10.1016/S0147-619X(02)00002-1
- Swinfield, T. J., Oultram, J. D., Thompson, D. E., Brehm, J. K., and Minton, N. P. (1990). Physical characterisation of the replication region of the *Streptococcus faecalis* plasmid pAM beta 1. *Gene* 87, 79–90.
- Szpirer, C. Y., Faelen, M., and Couturier, M. (2001). Mobilization function of the pBHR1 plasmid, a derivative of the broad-host-range plasmid pBBR1. *J. Bacteriol.* 183, 2101–2110. doi: 10.1128/JB.183.6.2101-2110.2001
- Taki, K., Abo, T., and Ohtsubo, E. (1998). Regulatory mechanisms in expression of the *traY-I* operon of sex factor plasmid R100: involvement of *traJ* and *traY* gene products. *Genes Cells* 3, 331–345. doi: 10.1046/j.1365-2443.1998.00194.x
- Tanimoto, K., Tomita, H., and Ike, Y. (1996). The *traA* gene of the *Enterococcus faecalis* conjugative plasmid pPD1 encodes a negative regulator for the pheromone response. *Plasmid* 36, 55–61. doi: 10.1006/plas.1996.0032
- Tullius, T. D., and Dombroski, B. A. (1986). Hydroxyl radical “footprinting”: high-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein. *Proc. Natl. Acad. Sci. U.S.A.* 83, 5469–5473. doi: 10.1073/pnas.83.15.5469
- Tzou, W. S., and Hwang, M. J. (1999). Modeling helix-turn-helix protein-induced DNA bending with knowledge-based distance restraints. *Biophys. J.* 77, 1191–1205. doi: 10.1016/S0006-3495(99)76971-7
- van der Lelie, D., Wösten, H. A., Bron, S., Oskam, L., and Venema, G. (1990). Conjugal mobilization of streptococcal plasmid pMV158 between strains of *Lactococcus lactis* subsp. *lactis*. *J. Bacteriol.* 172, 47–52.
- Van Kaer, L., Van Montagu, M., and Dhaese, P. (1989). Purification and *in vitro* DNA-binding specificity of the *Bacillus subtilis* phage phi 105 repressor. *J. Biol. Chem.* 264, 14784–14791.
- Wang, A., and Macrina, F. L. (1995). Streptococcal plasmid pIP501 has a functional *oriT* site. *J. Bacteriol.* 177, 4199–4206.
- Will, W. R., and Frost, L. S. (2006). Hfq is a regulator of F-plasmid TraJ and TraM synthesis in *Escherichia coli*. *J. Bacteriol.* 188, 124–131. doi: 10.1128/JB.188.1.124-131.2006
- Wong, J. J. W., Lu, J., and Glover, J. N. M. (2012). Relaxosome function and conjugation regulation in F-like plasmids - a structural biology perspective. *Mol. Microbiol.* 85, 602–617. doi: 10.1111/j.1365-2958.2012.08131.x
- Zatyka, M., Jagura-Burdzy, G., and Thomas, C. M. (1994). Regulation of transfer genes of promiscuous IncP alpha plasmid RK2: repression of TraI region transcription both by relaxosome proteins and by the Tra2 regulator TrbA. *Microbiology* 140(Pt 1), 2981–2990. doi: 10.1099/13500872-140-11-2981
- Zatyka, M., Jagura-Burdzy, G., and Thomas, C. M. (1997). Transcriptional and translational control of the genes for the mating pair formation apparatus of promiscuous IncP plasmids. *J. Bacteriol.* 179, 7201–7209.
- Zielenkiewicz, U., and Ceglowski, P. (2005). The toxin-antitoxin system of the streptococcal plasmid pSM19035. *J. Bacteriol.* 187, 6094–6105. doi: 10.1128/JB.187.17.6094-6105.2005

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer GS and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2016 Grohmann, Goessweiner-Mohr and Brantl. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Conjugative DNA Transfer Is Enhanced by Plasmid R1 Partitioning Proteins

Christian J. Gruber<sup>1</sup>, Silvia Lang<sup>1</sup>, Vinod K. H. Rajendra<sup>1</sup>, Monika Nuk<sup>1</sup>, Sandra Raffl<sup>1</sup>, Joel F. Schildbach<sup>2</sup> and Ellen L. Zechner<sup>1\*</sup>

<sup>1</sup> Institute of Molecular Biosciences, University of Graz, BioTechMed-Graz, Graz, Austria, <sup>2</sup> Department of Biology, Johns Hopkins University, Baltimore, MD, USA

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Consejo Superior de Investigaciones  
Científicas, Spain

### Reviewed by:

Guenther Muth,  
Universitaet Tuebingen, Germany  
Fabián Lorenzo,  
Universidad de La Laguna, Spain

### \*Correspondence:

Ellen L. Zechner  
ellen.zechner@uni-graz.at

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 01 June 2016

**Accepted:** 01 July 2016

**Published:** 19 July 2016

### Citation:

Gruber CJ, Lang S, Rajendra VKH,  
Nuk M, Raffl S, Schildbach JF and  
Zechner EL (2016) Conjugative DNA  
Transfer Is Enhanced by Plasmid R1  
Partitioning Proteins.  
Front. Mol. Biosci. 3:32.  
doi: 10.3389/fmolb.2016.00032

Bacterial conjugation is a form of type IV secretion used to transport protein and DNA directly to recipient bacteria. The process is cell contact-dependent, yet the mechanisms enabling extracellular events to trigger plasmid transfer to begin inside the cell remain obscure. In this study of plasmid R1 we investigated the role of plasmid proteins in the initiation of gene transfer. We find that Tral, the central regulator of conjugative DNA processing, interacts physically, and functionally with the plasmid partitioning proteins ParM and ParR. These interactions stimulate Tral catalyzed relaxation of plasmid DNA *in vivo* and *in vitro* and increase ParM ATPase activity. ParM also binds the coupling protein TraD and VirB4-like channel ATPase TraC. Together, these protein-protein interactions probably act to co-localize the transfer components intracellularly and promote assembly of the conjugation machinery. Importantly these data also indicate that the continued association of ParM and ParR at the conjugative pore is necessary for plasmid transfer to start efficiently. Moreover, the conjugative pilus and underlying secretion machinery assembled in the absence of Par proteins mediate poor biofilm formation and are completely dysfunctional for pilus specific R17 bacteriophage uptake. Thus, functional integration of Par components at the interface of relaxosome, coupling protein, and channel ATPases appears important for an optimal conformation and effective activation of the transfer machinery. We conclude that low copy plasmid R1 has evolved an active segregation system that optimizes both its vertical and lateral modes of dissemination.

**Keywords:** type IV secretion system, conjugative transfer, plasmid segregation, relaxase, pilus

## INTRODUCTION

Extrachromosomal DNA elements such as plasmids are responsible for their own propagation in dividing host cells. Low copy number plasmids rely on active segregation mechanisms for stable inheritance. In addition, many acquire the capacity for horizontal dissemination via bacterial conjugation. Because bacterial resistance to antibiotics is an immense problem in human health, research has focused on gaining detailed knowledge of the initiation stage of conjugation and its control. The process has been best studied in Gram-negative organisms where multiple mating pore formation (Mpf) proteins assemble a cell envelope spanning transport channel and cell surface pili or adhesins mediate contact between cells. A receptor, called the type IV coupling protein (T4CP), is positioned at the cytoplasmic entrance of the secretion channel to recognize specific plasmid-bound protein complexes and deliver them to the channel. Following an initiation signal that has

never been defined, the nucleoprotein cargo is then pumped through the transport apparatus in a reaction requiring ATP.

Regulation of conjugation involves donor cell perception of environmental signals. Knowledge of the control circuits coupling extracellular quorum signals and other stimuli to transcription of conjugation genes is increasing (Frost and Koraimann, 2010; Christie and Gordon, 2014; Clewell et al., 2014; Gibert et al., 2014). Yet, it remains challenging to discover how a potential recipient cell stimulates donor conjugative DNA transfer upon cell contact. We have postulated that bacteriophage might mimic potential recipient cells and initiate a signaling pathway that activates mechanisms typically involved in gene transfer. Thus, studies of bacteriophage that exploit conjugative pili as receptors for penetration of host cells are a promising approach to discover how cell contact-activated regulation of a type IV apparatus might operate.

Our work with the group 1 RNA phage R17 and the IncFII paradigm conjugation system R1 (Lang et al., 2011; Lang and Zechner, 2012) established that infection of the host required not only pilus biogenesis factors including TraA pilin, the Mpf proteins, the lytic transglycosylase P19 and the T4CP ATPase, but additionally the relaxosome, a complex of proteins generally required for binding and preparing the plasmid DNA origin of transfer (*oriT*) for export to recipient cells. The relaxosomes of F-like systems are well characterized (de la Cruz et al., 2010; Zechner et al., 2012). TraI is a bifunctional relaxase that cleaves one plasmid strand at *oriT* forming a covalent linkage to the nicked strand in the process (Matson et al., 1993). Recognition motifs enable TraI to bind the T4CP receptor and secretion of the TraI-DNA adduct delivers the plasmid to the recipient (Lang et al., 2010). A distinct functional region of TraI provides the essential helicase activity to generate single-stranded DNA (ssDNA) for export (Matson et al., 2001). In contrast to conjugative DNA transfer, R17 uptake via the R1-16 type IV apparatus does not require the entire TraI protein. This finding allowed us to define a novel domain of TraI necessary for activation of the nucleoprotein transfer via phage-generated signals (Lang et al., 2011). This work and previous biochemical studies support a model where the T4CP has a key role in coupling perceived signals of extracellular origin with intracellular cues provided by the relaxosome to activate the type IV channel (Berry and Christie, 2011; Lang et al., 2011). A following study showed that the activation domain of TraI is not only crucial to priming the T4CP for phage and conjugative transfer but also in signaling activation of the transporter for mobilization of competing plasmids such as ColE1 under conditions where the conjugative R1-16 plasmid is transfer deficient (Lang et al., 2014).

Another general function of conjugative pili is to form contacts with other cells and abiotic surfaces to promote biofilm development (Ghigo, 2001). Studies investigating the underlying mechanisms using F-like plasmids have highlighted the importance of pilus structure (Ghigo, 2001; Reisner et al., 2003). The *E. coli* biofilm phenotype and pilus-specific phage sensitivity can therefore be combined with general mutagenesis to identify proteins of host or plasmid origin that alter the conformation or function of the envelope spanning apparatus.

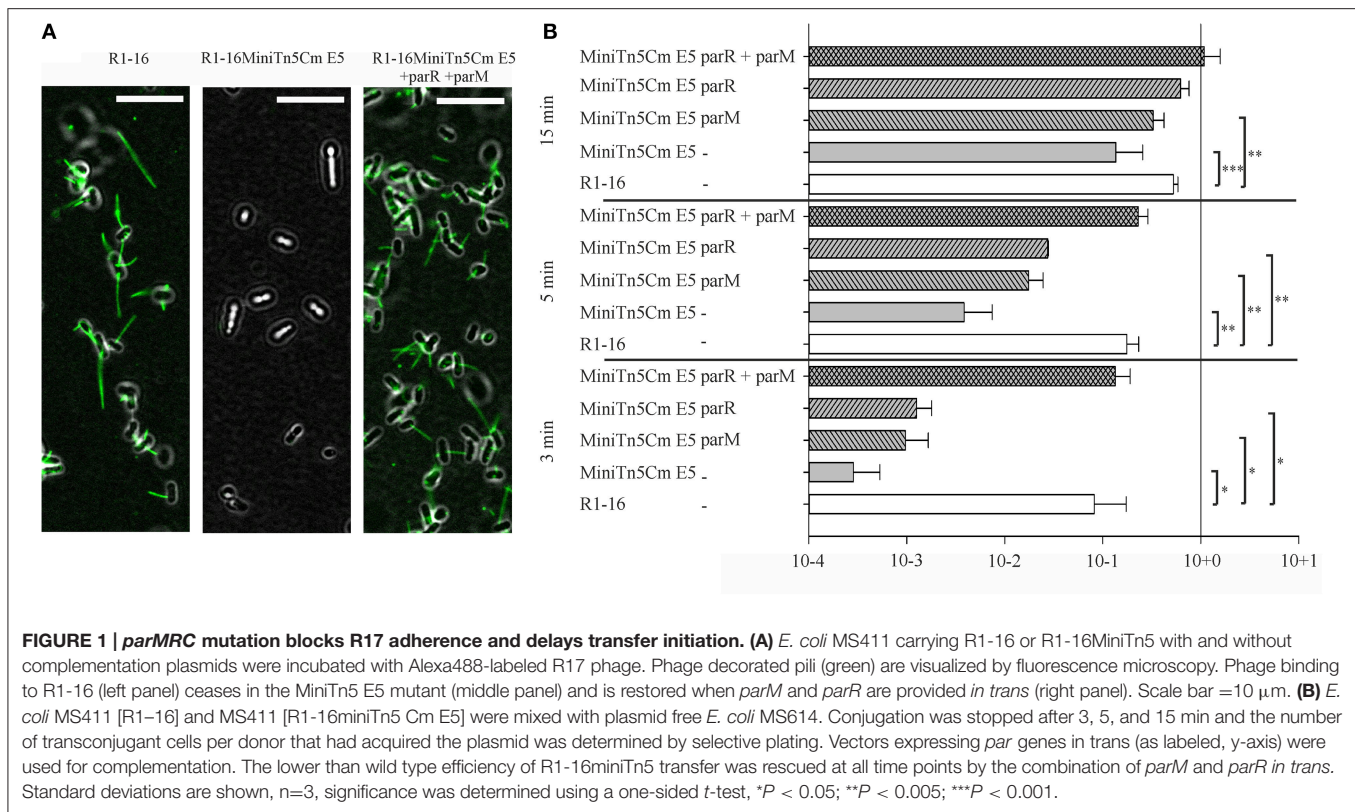
Using a screen of this type we identified a miniTn5 mutant derivative of plasmid R1-16, which assembled conjugation machinery able to transfer DNA with wild type efficiency yet the pili promoted poor biofilm formation and were completely deficient for R17 phage infection even with overnight incubation (Nuk et al., 2011). Surprisingly, the site of transposon insertion was the R1-16 *parMRC* operon, which is involved in active segregation (partitioning) of the low copy plasmid. The system involves a centromere-like sequence *parC* bound by the adapter protein ParR and the actin-like ATPase ParM to form bipolar spindles, which push sister plasmids to the cell poles during cell division (Moller-Jensen et al., 2003; Salje and Lowe, 2008; Bharat et al., 2015). Segregation systems like *parMRC* are key to faithful plasmid inheritance. Moreover, type I ParA-like proteins of plasmid and chromosomal origin are also involved in intracellular partitioning of cellular organelles and proteins (Lutkenhaus, 2012; Roberts et al., 2012; Jones and Armitage, 2015). A connection between plasmid partitioning factors and DNA transfer machinery was established for the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens*. In that system the ParA-like protein VirC1 spatially coordinates early DNA transfer events by mediating interactions between the T4CP VirD4 and the relaxase VirD2-DNA transfer intermediate (Atmakuri et al., 2007).

In this study we investigate the contribution of the R1 partitioning proteins ParM and ParR to the nucleoprotein transfer activities of the plasmid type IV secretion system (T4SS).

## RESULTS

### *parMRC* Mutation Blocks R17 Adherence and Delays Transfer Initiation

Mutagenesis of plasmid R1-16 used the transposon delivery system pUT-miniTn5Cm (Nuk et al., 2011). A selection step requiring conjugative transfer of the R1-16 mutant derivatives was included to eliminate those with transposon insertions in the plasmid *tra* genes. One biofilm deficient mutant, R1-16miniTn5Cm E5, carried the transposon inserted at position 488 of *parM* (Accession Number X04268), effectively blocking transcription of *parM* and *parR*. Disruption of this locus did not lead to an immediate loss of plasmid from the population and donor cultures conjugated normally in a standard 30 min mating experiment (Nuk et al., 2011). Nonetheless, the poor biofilm formation and the complete R17 resistance of hosts carrying this mutant suggested that the *parMRC* locus could be involved in the assembly and function of the conjugation machinery. We first asked whether conjugative pili assembled in the absence of *parMRC* were defective for bacteriophage attachment. *E. coli* MS411 carrying R1-16 or R1-16miniTn5Cm E5 were combined with fluorescently labeled R17 and visualized microscopically (Figure 1A). Attachment of the labeled R17 to wild type pili was apparent but strikingly absent from hosts carrying the mutant. The attachment defect was complemented by expression of *parM* and *parR* *in trans*. These data suggest that pili assembled in the absence of *parMRC* have an abnormal conformation.



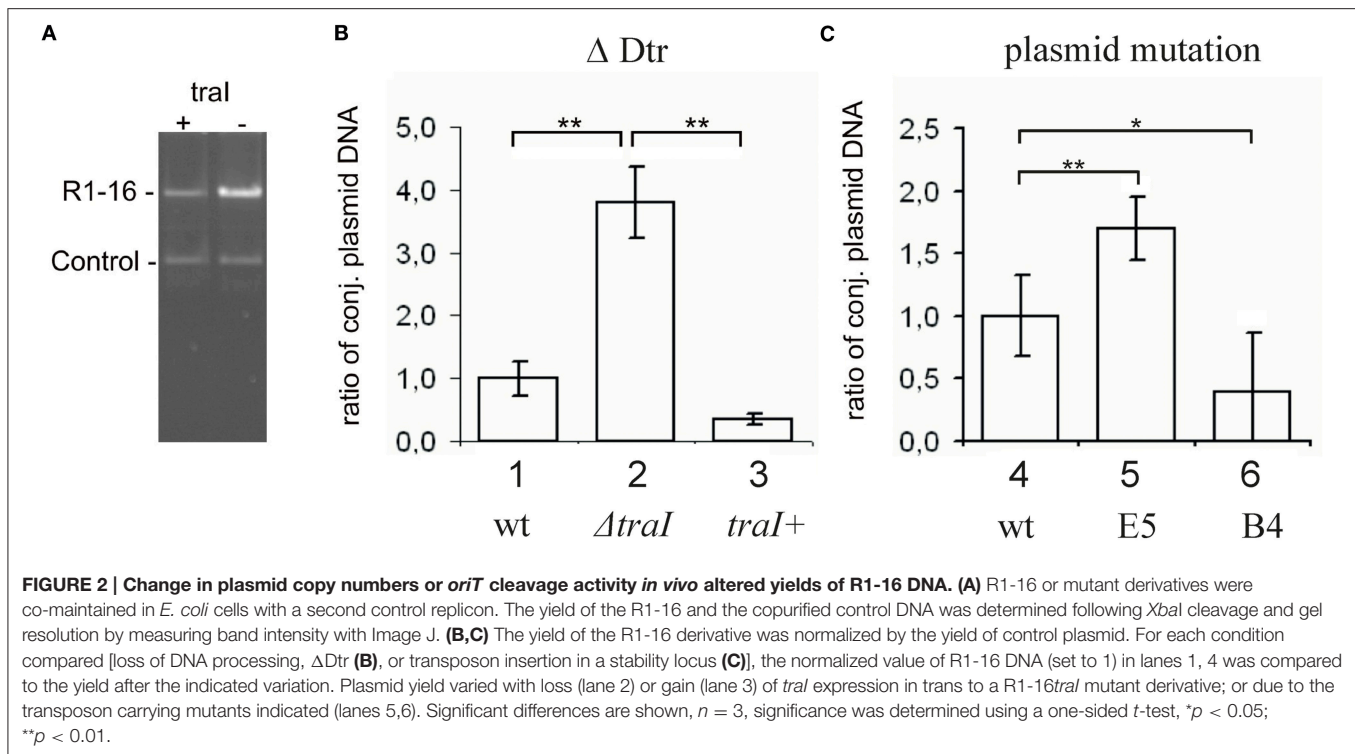
We then asked whether a defect in early stages of plasmid transfer could be detected. *E. coli* MS411 [R1-16miniTn5Cm E5] donors were combined with *E. coli* MS614 recipient cells in broth culture. Conjugation was interrupted after 3–15 min and transconjugants were selected on agar plates. After 3 min of coincubation, transfer of the R1-16miniTn5Cm E5 was detected, but at frequencies 2–3 log units lower than transfer of wild type R1-16 (**Figure 1B**). Complete complementation of transconjugant formation at this time point was observed by providing *parM* and *parR* in combination on an expression plasmid *in trans*. Presence of either *parM* or *parR* was not sufficient. Significantly lower transfer frequencies were also observed for R1-16miniTn5Cm E5 compared to wild type after 5 and 15 min of conjugation. The magnitude of this difference decreased with increasing time, however. In every case, addition of *parM* and *parR* *in trans* fully complemented mating efficiency to wild type levels. We conclude that *E. coli* carrying the *parMRC* transposon insertion in R1-16 exhibits a delayed initiation of transfer phenotype that is fully overcome when cultures are allowed to conjugate for longer than 15 min.

### ***parMRC* Disruption Reduced *oriT* Cleavage *In vivo***

Transfer initiation requires the activity of the DNA processing relaxosome complex. We asked next whether the *parMRC* disruption influences nicking of R1-16 *oriT* by TraI *in vivo*. R1-16 plasmids express conjugative genes constitutively and therefore support continuous relaxosome assembly. Within the relaxosome TraI maintains an equilibrium of cleaving and resealing of the

nick site, meaning that a fraction of R1-16 will be in a nicked state at any moment (Zechner et al., 1997). When host cells are lysed for plasmid DNA isolation, the population of nicked molecules covalently attached to TraI should be lost during phenol extraction, lowering the yield. In contrast, any condition that disrupts *oriT* cleavage would allow R1-16 to remain supercoiled, increasing DNA recovery. To validate this assay of relative plasmid yield we combined R1-16 or mutant derivatives in *E. coli* M1174 cells with a second independent replicon. The two plasmids were copurified by phenol extraction, linearized once with *XbaI* and applied to agarose gels to detect quantitative variation in the apparent copy number of the conjugative plasmid relative to the second replicon. **Figure 2** illustrates changes in plasmid ratios obtained by controlled manipulation of *oriT* DNA processing ( $\Delta$  Dtr) through deletion of *traI*. The band intensities of the recovered R1-16 derivatives were normalized to the internal control plasmid and compared (**Figures 2A,B**). Values obtained for R1-16 wild type plasmid were set to 1. Disruption of the *traI* gene resulted in a nearly four-fold higher relative yield compared to wild type R1-16 DNA. Induction of *traI* expression *in trans* restored nicking activity, resulting in a plasmid yield significantly lower than in the absence of *traI*. We have used a similar assay previously (Nuk et al., 2011) to test whether random insertions of transposon miniTn5 in plasmid R1-16 result in plasmid instability. Validation of the assay in that case relied on controlled manipulation of the plasmid R1 copy number control system. In the current study, the plasmid recovery assay was applied to *E. coli* cells carrying R1-16 wild type, R1-16miniTn5Cm E5 and, as positive control, an additional





transposon insertion in the *yjcA* gene close to the *kis/kid* stability locus of plasmid R1. Due to the insertion in the *parMRC* locus, we predicted relatively low yields of R1-16miniTn5Cm E5 compared to the reference replicon (Jensen and Gerdes, 1999), yet surprisingly the *parMRC* mutant derivative was obtained in higher relative yields than wild type R1-16 (**Figure 2C**). In comparison the control mutant B4, carrying the miniTn5 at the *kis/kid* locus, was poorly recovered, as expected for a destabilized plasmid. A possible explanation for the unexpectedly high yield of the *parMRC* mutant is that the assay outcome reflects a stronger defect in relaxosome activity than in plasmid stability. In that case, the partitioning components of the ParMRC system appear to enhance *oriT* DNA processing *in vivo*.

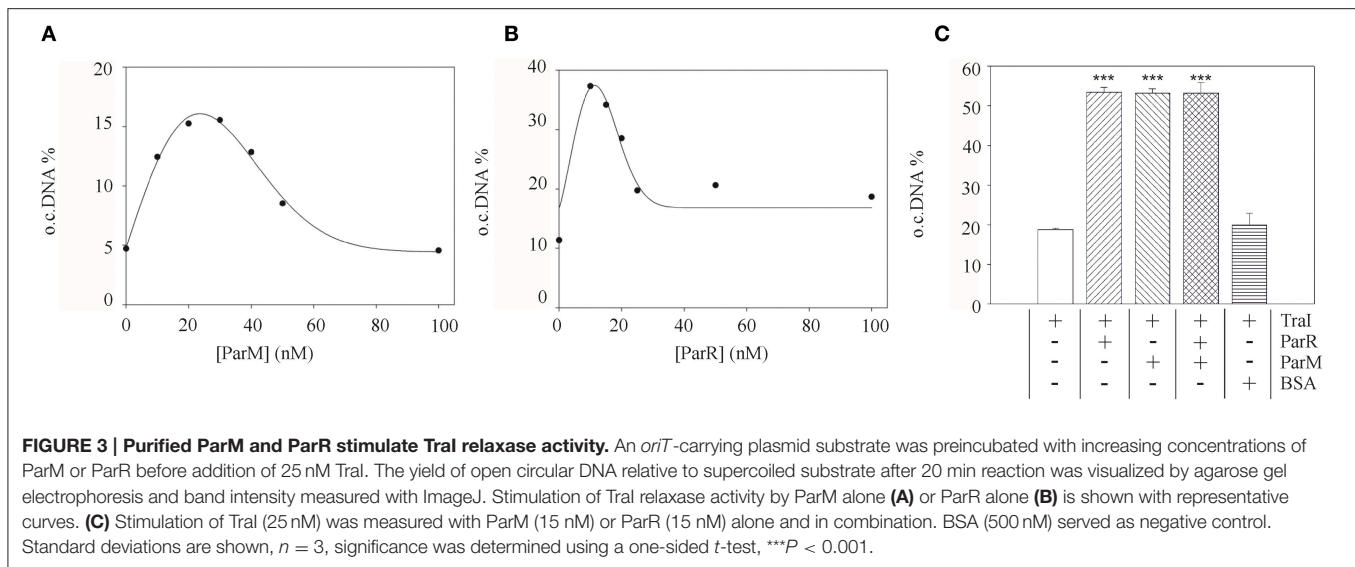
## The Relaxase of TraI is Stimulated by ParM and ParR *In vitro*

To test whether this effect of the partitioning components directly involves TraI we purified ParR and ParM proteins and measured the impact of these effectors on known enzymatic activities of the TraI enzyme *in vitro*. A standard *oriT* DNA-cleavage assay used to monitor relaxase enzyme activity measures the conversion of supercoiled plasmid substrate to the open circular form using agarose gel electrophoresis (Lanka and Wilkins, 1995; Csitkovits et al., 2004). A supercoiled substrate plasmid (4 nM) carrying 420 bp of R1 *oriT* (pDE100) was preincubated with putative effector proteins ParM or ParR and the reaction started by the addition of 25 nM TraI (**Figure 3**). The percentage of *oriT* DNA captured in open circular form was significantly enhanced by the additional presence of ParM or ParR in a concentration dependent manner.

Maximum stimulation (~three-fold, 5–16%) was observed when ParM was present in equimolar amounts (20–30 nM) relative to TraI. ParR alone (10 nM) stimulated TraI relaxase activity nearly four-fold (11–38%). At higher concentrations ParM and ParR failed to stimulate. Moreover, no superstimulation was observed when both factors were present. We then asked whether the Par proteins also stimulate truncated versions of TraI in this assay (not shown). N-terminal fragment TraI<sub>1–308</sub> (TraIN308) forms the minimal relaxase domain, and residues 1–992 (TraIN992) comprise the relaxase and the complete activation domain absolutely required for all T4SS activities we have tested thus far. Titration of either Par effector protein to the reactions containing truncated forms of TraI did not result in stimulation of *oriT* cleavage. We conclude therefore, that both ParM and ParR stimulate the *oriT* cleaving and joining activity of TraI independently. Stimulation was observed exclusively with the full length TraI protein.

## ParM and ParR Mediated Stimulation of TraI is Specific for the Relaxase Reaction

In addition to DNA transesterase activity TraI also acts as a single-stranded (ss)DNA dependent ATPase and helicase that unwinds the plasmid DNA duplex in preparation for transfer to recipient cells. We next asked whether the Par proteins affect these enzyme activities of TraI. We measured ATP hydrolysis by purified TraI on single-stranded circular M13 DNA in the absence or presence of increasing concentrations of ParM and ParR. The specific activity of TraI was 226 kmol ATP/h/mol protein. No stimulation of this activity was observed with additional proteins present.



We next tested whether the Par proteins affect the helicase activity of TraI. The enzyme initiates unwinding on any DNA substrate if it is able to first bind to a stretch of ssDNA 5' to the duplex junction (Kuhn et al., 1979; Csitkovits and Zechner, 2003). We generated two dsDNA substrates with a 60 bp central region of unwound DNA to support helicase loading (Sut et al., 2009). The substrates contained R1 *oriT* DNA for specific TraI binding (Williams and Schildbach, 2006) or non-specific sequences. The extent of DNA unwinding on these substrates agreed well with our previous results (Sut et al., 2009). However, unlike the stimulatory effect we obtain with the auxiliary effectors TraM, IHF, or TraD, no enhancement of helicase activity was observed in the additional presence of ParM or ParR alone, or in combination, under any conditions we tested (not shown). These findings indicate that the effects of Par proteins on TraI are specific for the enzyme's relaxase activity.

### TraI Stimulates ParM ATPase

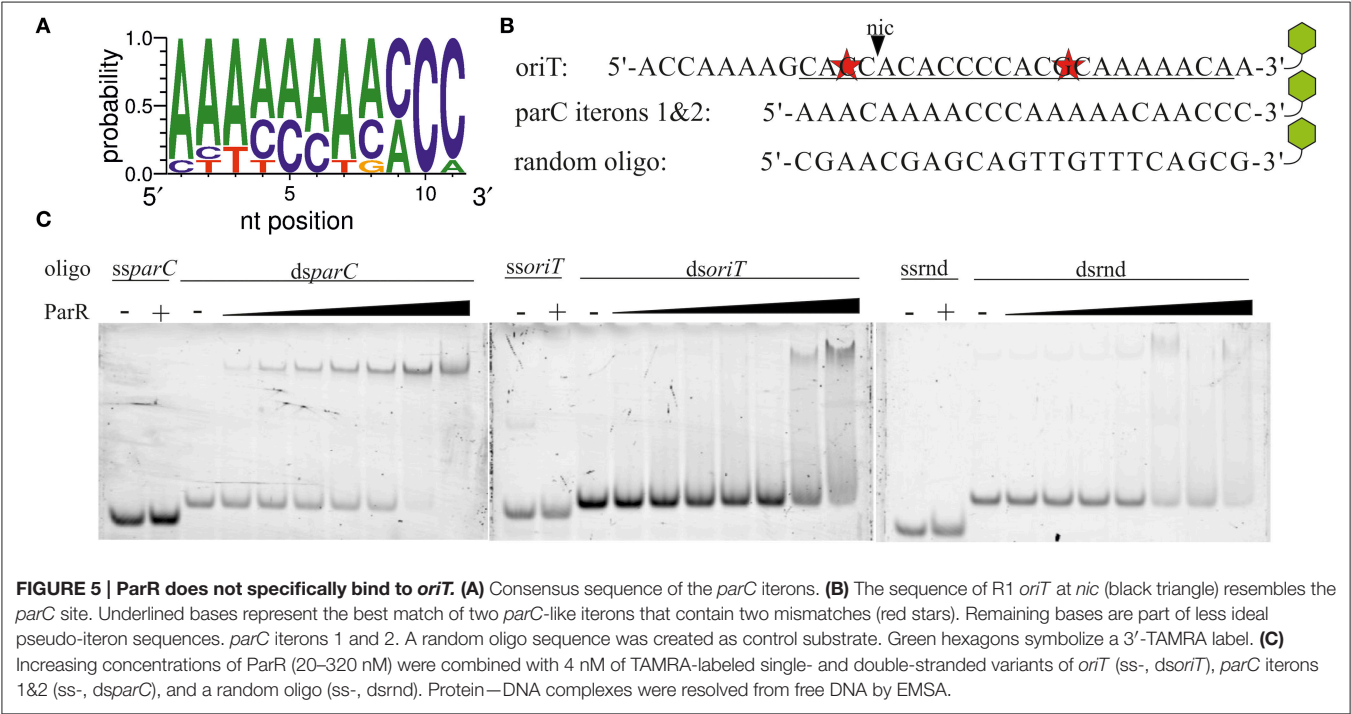
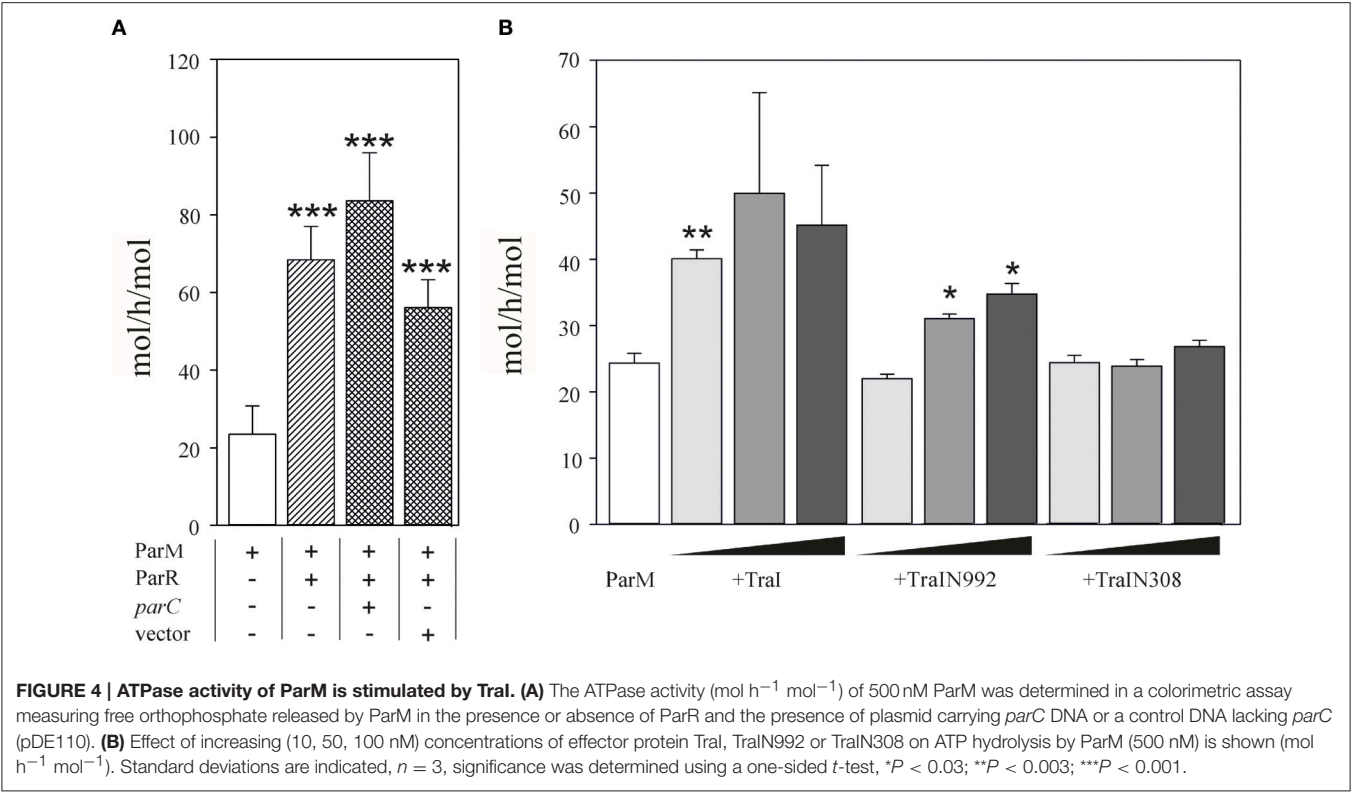
We then performed the reciprocal test and asked whether TraI could stimulate ParM ATPase (Figure 4). Using conditions standardized by the K. Gerdes laboratory we measured  $23.5 \pm 7.3$  mol ATP hydrolysed per hour per mol ParM (mol/h/mol). In good agreement with Jensen et al. (Jensen and Gerdes, 1997), this activity was increased three-fold in the additional presence of excess ParR (9 mM). ParR mediated stimulation was increased to ~four-fold in the additional presence of 17 nM double-stranded (ds)DNA containing *parC*. No additional ParR enhancement of ParM ATPase was observed when the dsDNA lacked *parC*. The effect of TraI protein on ParM ATPase was then tested using increasing concentrations of the full length TraI, N-terminal fragments TraIN992 and TraIN308 (Figure 4B). Significant stimulation of ParM ATPase was observed with the full length TraI and the TraIN992 fragment but not by the smallest TraIN308 variant. No ssDNA effector was present in the reaction mixtures effectively silencing ATP hydrolysis by

TraI itself. Moreover, TraIN992 lacks ATPase activity (Matson and Ragonese, 2005). We conclude that TraI increased ATP hydrolysis by ParM.

Taken together these results imply that the mutual stimulation of ParM ATPase and TraI DNA transesterase activities is due to protein-protein interactions supported most efficiently with the full length TraI protein.

### Par Protein-TraI Interactions Do Not Alter DNA Binding Activities

All of the proteins known to stimulate either the relaxase or helicase activities of TraI (Mihajlovic et al., 2009; Sut et al., 2009) bind to *oriT* DNA, specifically, in the case of TraM, TraY, and IHF or, in the case of the coupling protein TraD, in a sequence independent manner (Tsai et al., 1990; Nelson et al., 1995; Verdino et al., 1999; Schröder et al., 2002; Wong et al., 2011). ParR binds specifically to two sets of five direct repeats at the *parC* site (Moller-Jensen et al., 2003, 2007). These authors also showed that a minimum of two iterons is sufficient to support ParR binding (Moller-Jensen et al., 2003). We noted that *oriT* of plasmid R1 contains an A/C rich sequence that may constitute two *parC*-like iterons with a single mismatch to the consensus in each (Figure 5). The *parC*-like sequence overlaps part of the inverted repeat and neighboring bases specifically recognized by TraI (Stern and Schildbach, 2001) raising the possibility that ParR stimulates TraI activity through *oriT* binding. We compared ParR binding to different DNA fragments using an electrophoretic mobility shift assay (EMSA) (Figure 5). No binding of ParR to any ssDNA was observed. As a positive dsDNA control, a 22 bp fragment containing iterons 1 and 2 of *parC* was used. Mobility shift of the *parC* fragment was observed beginning at protein/DNA molar ratios of five to one, comparable to published values (Moller-Jensen et al., 2003). By contrast a mobility shift of the *oriT* sequence by ParR required a 40-fold molar excess of protein to DNA, equivalent to amounts necessary to shift a random-sequence substrate that served as



negative control. We conclude that ParR does not specifically bind *oriT*.

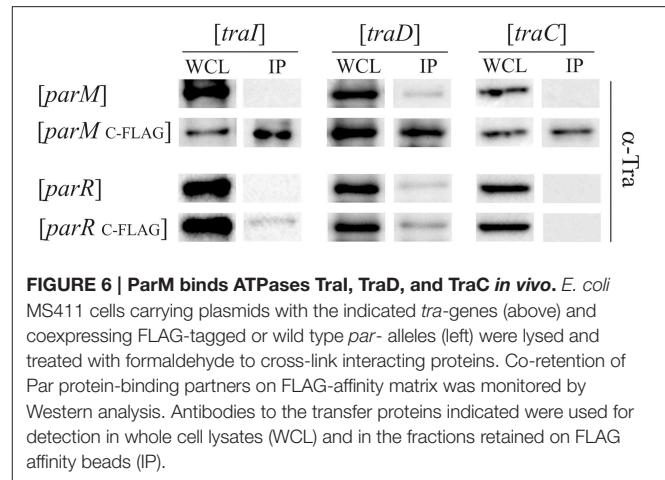
Due to this finding, we next asked whether the Par proteins alter TraI-DNA interactions. We tested two hypotheses:

Par-mediated stimulation of relaxase activity is due to (i) a higher rate of TraI association with the substrate or (ii) a stabilization of the product in a cleaved state. Fluorescence intensity and anisotropy measurements of TraI association with a 3'-TAMRA

labeled 17mer *nic*-substrate have been described in detail (Stern and Schildbach, 2001; Harley et al., 2002; Williams and Schildbach, 2006; Hekman et al., 2008; Dostal and Schildbach, 2010). In a set of experiments using this approach (Figure S1) we investigated whether ParR or ParM induce variation in TraI DNA binding.  $K_D$ s for TraI, TraIN308 (relaxase domain), and Tra $\Delta$ N308 (helicase domain) were determined. The presence of ParM or ParR (both 10 nM) did not change affinity of TraI for *oriT*-DNA. Thus, we conclude that the partitioning factors do not stimulate the relaxase reaction by altering the enzyme's DNA binding properties (see Supplementary Material for full description of results).

### ParM Binds Conjugation Proteins *In vivo*

Given that ParM and ParR increase TraI relaxase activity but do not bind to *oriT* DNA we next tested for direct protein-protein interactions. Par protein fusions were engineered with terminal FLAG-epitopes. Expression plasmids for the tagged fusion proteins were maintained in *E. coli* MS411 cells carrying either R1-16, or a second vector expressing a candidate *tra* gene. Following induction of fusion protein expression cells were lysed, protein complexes briefly cross-linked with formaldehyde and the Par proteins captured on FLAG-affinity beads. Bound proteins and their specific interaction partners were washed, eluted, and visualized by western immunoblotting. Based on the phenotypic and biochemical results, candidates for specific Par protein binding included TraI and the second key factor involved in pilus specific R17 phage infection, T4CP TraD, as well as the VirB4-like ATPase TraC, which is essential for assembly and function of R1 pili. Tra proteins retained by Par proteins on the FLAG affinity matrix were detected with polyclonal antibodies generated to specific transfer proteins. The amounts of Tra proteins detected in the whole cell lysates reflect native levels produced from R1-16. The Tra specific antibodies revealed that TraI, TraD and TraC were co-retained by affinity beads together with ParM<sub>C-FLAG</sub> (Figure S2). The specificity of these interactions was confirmed with a *par* allele lacking FLAG. In contrast ParR retained very small amounts of TraI and TraC but only with the C-terminal FLAG epitope (Figure S2). To assess whether the observed Tra-Par binding interactions can occur in the absence of the other segregation and transfer components, we co-produced the proteins in a pairwise manner. As shown in **Figure 6**, TraI, TraD, and TraC were again co-purified with ParM<sub>C-FLAG</sub>. Retention of low amounts of TraI by ParR was confirmed. No specific interaction with TraD or TraC was detected. Relaxase accessory factor TraM binds *oriT*, TraD, and the membrane (Schwab et al., 1991; Disque-Kochem and Dreiseikermann, 1997; Lu and Frost, 2005), but antibodies to TraM did not detect co-retention of this protein with either ParM or ParR (not shown). The abundance of Par protein present in each cell extract and in the pull down fractions was quantitated using antibodies to the FLAG epitope (Figure S3). We conclude that ParR binds TraI less strongly than ParM. Moreover ParM protein in the cell binds not only TraI, but additionally the T4CP TraD and TraC. In every case, these interactions take place in the absence of an intact transfer machinery and filament formation. There is a possibility that ParM is bound



**FIGURE 6 | ParM binds ATPases TraI, TraD, and TraC *in vivo*.** *E. coli* MS411 cells carrying plasmids with the indicated *tra*-genes (above) and coexpressing FLAG-tagged or wild type *par*- alleles (left) were lysed and treated with formaldehyde to cross-link interacting proteins. Co-retention of Par protein-binding partners on FLAG-affinity matrix was monitored by Western analysis. Antibodies to the transfer proteins indicated were used for detection in whole cell lysates (WCL) and in the fractions retained on FLAG affinity beads (IP).

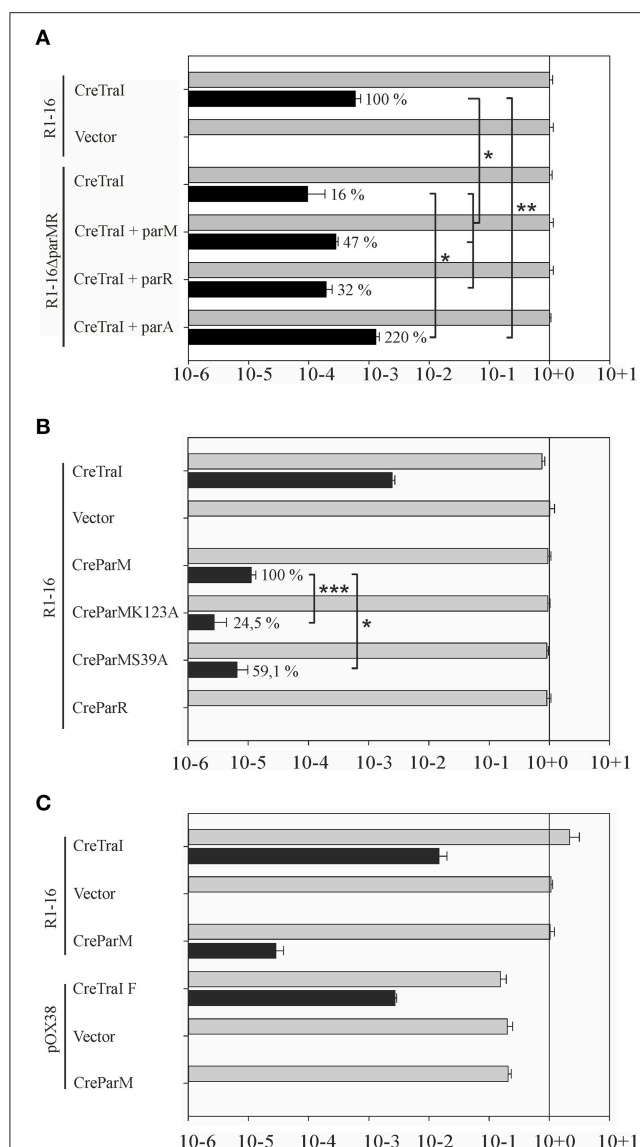
by additional Tra proteins of plasmid R1 but currently we do not have antibodies available for the entire suite of purified components.

To test whether the additional binding partners, TraD and TraC, alter ParM ATPase activity, the purified proteins were assayed in combination *in vitro*. Our published (Mihajlovic et al., 2009) and unpublished (V.K.H. Rajendra and E.L. Zechner) observations show that the soluble form of TraD (lacking the N-terminal transmembrane domain; Tra $\Delta$ N130) and full length TraC interact with several protein and DNA ligands *in vitro*, yet neither Tra protein increased ATP hydrolysis compared to reactions containing ParM alone (not shown).

### Lack of ParM and ParR Lowers Protein Translocation by the T4SS

Some conjugative T4SS have been shown to translocate specific proteins in addition to protein-DNA adducts to recipient cells in a manner that strictly requires the T4CP. TraI is thus far the only protein known to be secreted by F-like transporters in both its nucleoprotein and DNA free forms. In both activities TraD is expected to be directly involved in recognition and binding of the (nucleo)protein substrate. We tested whether the absence of Par proteins affect protein translocation using the Cre recombinase assay for translocation (CRAfT). This technique fuses the reporter enzyme to a protein specifically secreted by the T4SS (Vergunst et al., 2000). Transfer to recipients is then quantitated by a switch in antibiotic resistance catalyzed by Cre recombination. In our assay (Lang et al., 2010) donor cells carry R1-16 to provide all the essential components for substrate recognition, conjugative DNA processing, and transport including wild type TraI protein. Here we compared the frequency of Cre-TraI transmission supported by R1-16 carrying cells to hosts carrying the double deletion R1-16 $\Delta$ *parMR*. The frequency of plasmid DNA transfer was measured to provide an internal standard for the performance of the T4SS in every experiment. As shown in **Figure 7A** *E. coli* R1-16 transferred Cre-TraI efficiently, in good agreement with our prior results. Significantly less (six-fold)





**FIGURE 7 | Secretion of Cre-TraI and Cre-ParM to recipient cells. (A)**

Protein translocation was detected by recombination events per donor. The frequencies of protein transfer (black bars) were normalized to conjugation efficiency for each culture (gray bars). Relative differences in Cre-TraI transfer by R1-16 *par* mutant derivatives compared to wild type (100%) are indicated. Empty complementation vector pMS119EH was used as a control. **(B)** Frequencies of translocation of the indicated Cre-fusion proteins (left) by wild type R1-16 are shown with black bars. Percent of Cre-ParM transfer observed with mutant ParM variants compared to wild type ParM (100%) is indicated. **(C)** F transfer proteins expressed by pOX38 mediate DNA transfer (gray bars) and translocation of the F TraI protein fused to Cre, but not Cre-ParM<sub>R1</sub>. Standard deviations are shown,  $n = 3$ , significance was determined using a one-sided *t*-test, \* $P < 0.05$ ; \*\* $P < 0.05$ ; \*\*\* $P < 0.005$ .

Cre-TraI was transferred in the absence of both *par* genes (R1-16Δ*parMR*). Expression of both *parM* or *parR* *in trans* to R1-16Δ*parMR* complemented the protein transfer defect to higher than wild type efficiency. Neither factor alone was sufficient. These data imply a role for the Par proteins in efficient TraI transfer.

## Cre-ParM Fusion Proteins are Specifically Transported to Recipient Cells

We next asked whether the reciprocal activity, namely the direct transfer of either Par protein by the T4SS to recipient cells, could be detected. The *cre* gene was fused 5' to each *par* gene and translocation of ParM and ParR was analyzed by CRAFT. No Cre-ParR transfer was detected (**Figure 7B**). Remarkably, however, Cre-ParM translocation was measured. The observed frequency was low compared to Cre-TraI transfer. To address whether Cre-ParM transfer is the result of specific recognition, we tested ParM variants with amino acid exchanges in residues exposed on outside loops and along the surface of ParM filaments (Salje and Lowe, 2008). These residues are important to ParR binding, and crucial to stable filament formation. Both mutant variants were transferred, but CreParMK123A and CreParMS39A secretion was significantly less efficient (24.5 and 59% of wild type levels, respectively). The impact of single residue exchanges on transfer efficiency strengthens the evidence for specific ParM recognition by the T4CP. We then asked whether the highly related conjugation system of plasmid F would also mediate transfer of ParM, despite the absence of *parMRC* on that plasmid. For this test CRAFT assays were performed with pOX38 (**Figure 7C**). Although Cre fused to F TraI was efficiently secreted, we measured no Cre-ParM transfer. We conclude that translocation of ParM protein is unique for plasmid R1-16.

## DISCUSSION

Partitioning systems are classified by their motor proteins as type I (ParA-like), type II (ParM-like), and type III (TubZ-like; Salje et al., 2010). These dynamic systems assemble into higher order structures that organize and move subcellular components. They segregate not only plasmids and bacterial chromosomes, but also partition cell organelles and proteins intracellularly (Lutkenhaus, 2012; Roberts et al., 2012; Ptacin et al., 2014; Jones and Armitage, 2015). Type I loci encode ATPases with a deviant Walker A nucleotide binding motif (Szardenings et al., 2011). The type I enzyme VirC1 of *A. tumefaciens* is required for efficient T-DNA transfer. Ground breaking work in the Christie laboratory revealed that the VirC1 motor protein promotes conjugative DNA transfer by coordinating two early steps of that process. First, VirC1 acts with auxiliary factor VirC2 to promote assembly of the relaxosome at *oriT*-like T-DNA border sequences. VirC1 then acts to spatially position the transfer intermediate at the cell pole and stimulate docking of this substrate to the T4SS channel (Atmakuri et al., 2007).

Functional links between segregation and conjugation machineries have been observed in other systems as well. The stability locus *stbABC* characterized in plasmid R388 is conserved in several Mob families (Guynet et al., 2011). Loss of R388 stability through *stbA* inactivation is caused by the plasmid's mislocalization to nucleoid-free regions of the cell. Given that the R388 T4 secretion machinery preferentially assembles at the cell poles, the accumulation of plasmid copies at the poles in the absence of StbA favors higher than normal frequencies of conjugative transfer. This functional organization

thus coordinates vertical and lateral modes of plasmid R388 dissemination, i.e., conditions that jeopardize faithful plasmid inheritance are compensated by enhanced horizontal transfer. The logic of this elegant regulatory circuit is apparent but the mechanistic basis remains unknown. In particular the function of the ParA-like ATPase StbB in plasmid positioning and any possible direct contribution to conjugative transfer remain unresolved. Finally we note with interest that *parA* and *parB* of the gonococcal genetic island of *Neisseria gonorrhoeae* are involved in DNA secretion by the T4SS (Hamilton et al., 2005) but the functions performed by these proteins also await further study (Pachulec et al., 2014).

Here we report experimental evidence suggesting that the type II partitioning locus *parMRC* of plasmid R1 has been appropriated by the conjugation machinery to facilitate early steps in the assembly and function of the T4SS. Mechanistic similarities with VirC1 include the capacity of ParM and ParR to stimulate the *oriT* cleavage reaction of TraI *in vitro* and of the relaxosome *in vivo*. Although VirC1 binds to a sequence called *overdrive* adjacent to the T-DNA right border and the *oriT*-like sequence cleaved by VirD2 relaxase, ParR does not bind *oriT* of plasmid R1 and no evidence was found for an effect of the Par proteins on TraI DNA binding properties. It follows that enzyme stimulation probably occurs via direct interactions of the proteins. Indeed TraI stimulates ParM ATPase activity in the absence of DNA. Moreover the mutual stimulation of both ParM and TraI proteins was decreased when truncated versions of TraI were assayed. Finally, binding of the partner proteins *in vivo* was confirmed with pull down assays.

VirC1 is able to act as a central organizer of the relaxosome because of its DNA binding activity and because it binds pairwise with the accessory factors VirC2, VirD1, and relaxase VirD2. These interactions were detected in the absence of the Ti plasmid, therefore VirB channel components are not involved. VirC1 also associates with the polar membrane and binds the T4CP VirD4. Together these properties enable VirC1 to actively recruit the relaxosome to the cell poles and to the colocalized assembly of T4CP and the VirB T4 secretion channel.

In the R1 system, active relaxosomes form *in vivo* in the complete absence of *parMRC* (Karl et al., 2001). Nonetheless here we see that the absence of Par proteins decreases *in vivo* cleavage activity in the absence of conjugation and delays DNA transfer during conjugation. In the simplest model, recruitment of the R1 relaxosome to the conjugative pore would simply involve recognition of relaxase translocation signals by the T4CP receptor and docking interactions between the coupling protein and factors of the relaxosome (TraM, TraI, DNA). Alternatively however, recruitment may be enhanced by spatial determinants provided by the *parMRC* segregation system (Figure 8). We show that partitioning proteins ParM and ParR interact physically and functionally with several proteins of the T4 secretion machinery. Transfer proteins may have acquired an affinity for ParM to exploit the protein's cytomotive force for intracellular localization. Plasmid R1 is expected to produce very few intracellular ParM filaments that are situated close to the edge of the nucleoid (Salje et al., 2009; Bharat et al., 2015). These authors propose that the capture of plasmid DNA may take place within the nucleoid periphery. ParM filaments

form continuously but are subject to dynamic instability (Garner et al., 2004, 2007). Filaments that fail to capture a ParR bound plasmid centromere rapidly disassemble. By contrast the growth of plasmid-bound filaments is stabilized long enough to push plasmids to the cell's polar extremes. Affinity of the conjugation proteins and the relaxosome for ParM should thus concentrate these along the filaments aligned on the longitudinal axis of the cell (Moller-Jensen et al., 2003; Campbell and Mullins, 2007). Once plasmids are positioned at the poles, ParM depolymerizes, allowing the plasmids to drift randomly until recaptured by a ParM filament (Campbell and Mullins, 2007). The dynamic nature of this "positioning network" should facilitate not only rapid nucleation of transporter assembly but also colocalization of T4CP and relaxosome at those sites (Figure 8). The central role of ParM in multivalent binding interactions could additionally promote efficient joining of the different subassemblies including channel ATPases and other transporter components, the T4CP and finally, docking of the relaxosome.

Importantly, our data additionally show that not only early stages of protein colocalization and *oriT*-DNA processing are enhanced, but also that the continued association of Par proteins at the interface of relaxosome, T4CP and the conjugative ATPases is important to optimal T4SS function (Figure 8). This functional interdependence governs several T4SS-mediated activities: (i) R17 phage entry via a pathway otherwise dependent on pilus conformation, and productively docked, enzymatically active complex of T4CP and relaxosome, (ii) host biofilm formation; a process relying on pilus mediated contracts with surfaces and other cells, and (iii) rapid completion of plasmid DNA transfer. Moreover, ParM and ParR interact with TraI and may directly enhance TraI protein secretion. Finally, ParM positioning at the conjugative pore and specific binding by the T4CP receptor is demonstrated by secretion of this protein to recipient cells. We have not demonstrated a specific function for ParM protein in recipient cells, but it is tempting to speculate that co-transfer of ParM may support TraI-catalyzed recircularization of T-DNA helping to stabilize this strand in the new host.

We propose a working model where positioning of the Par proteins at the T4SS channel opening induces a conformational change in the envelope-spanning complex and/or the conjugative pilus, which is important to productive biofilm formation and essential for bacteriophage R17 to penetrate the host. It follows that the altered structure generated by integration of the Par proteins seems best able to perceive or process signals conveyed by cell to surface and cell-cell contacts during biofilm formation and pilus mediated uptake of phage RNA. In the absence of this conformational shift, plasmid transfer does occur, but is indeed delayed. In summary we conclude that the simple, self-organizing *parMRC* system actively promotes not only faithful vertical transmission of the low copy number plasmid R1, but also streamlines lateral spread via conjugation.

## MATERIALS AND METHODS

**Strains:** All strains used are listed in Table 1. Cultures were grown in LB-media. Antibiotic concentrations were as follows: kanamycin (Km) 50 µg/ml, chloramphenicol (Cm) 10 µg/ml,

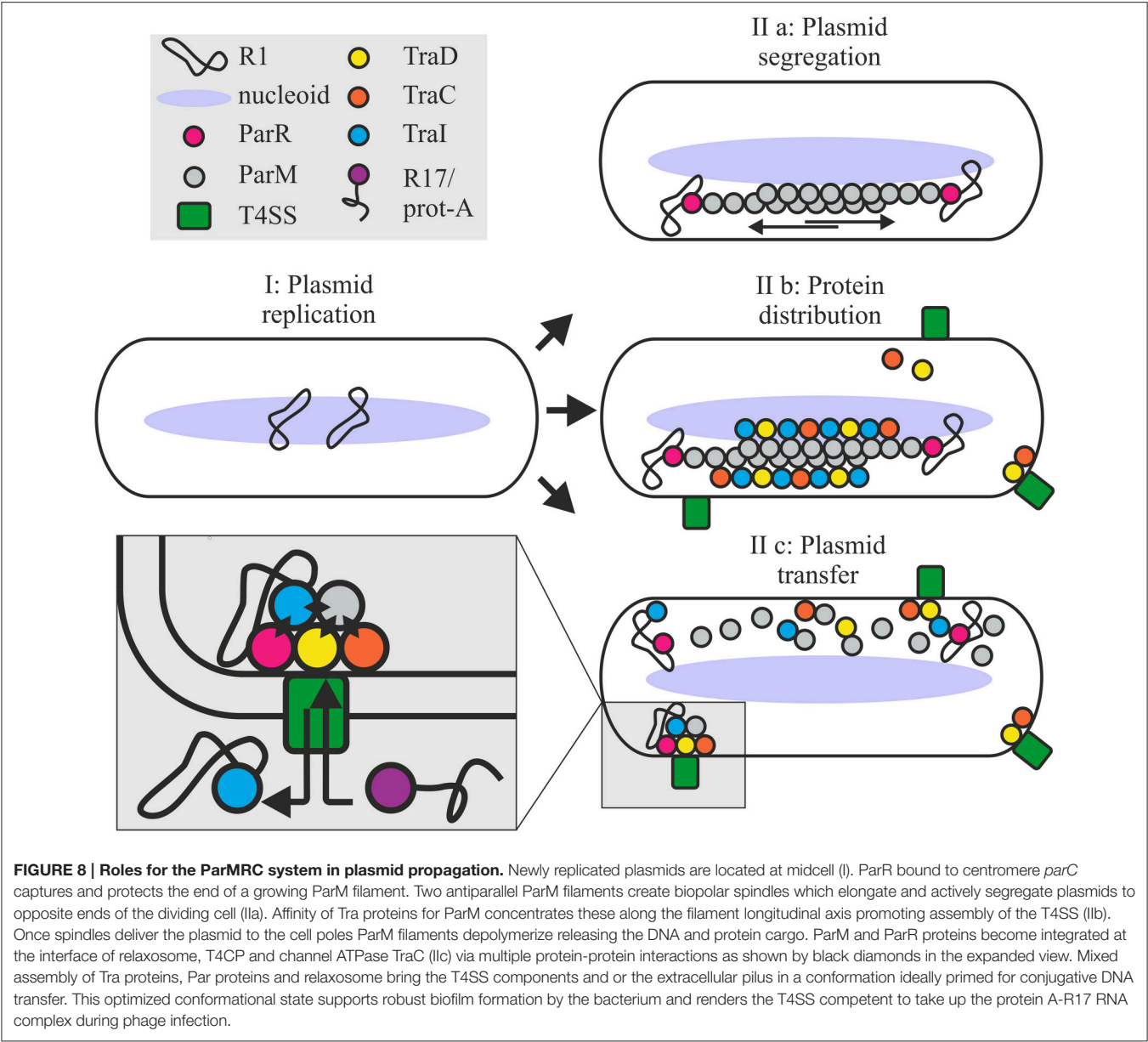


TABLE 1 | Strains used in this study.

Strain	Description <sup>a</sup>	References
<i>E. coli</i> DH5α	F <sup>−</sup> Φ80lacZΔM15 Δ(lacZYA-argF) U169 <i>recA1 endA1 hsdR17</i> (rK <sup>−</sup> , mK <sup>+</sup> ) <i>phoA supE44 λ</i> – <i>thi-1 gyrA96 relA1</i>	Woodcock et al., 1989
<i>E. coli</i> C41(DE3)	F- <i>ompT hsdS</i> (rB-mB-) <i>dcm</i> + Tetr gal I (DE3) <i>endA Hte</i> [ <i>argU ileY leuW Cm</i> <sup>R</sup> ]*]	Miroux and Walker, 1996
<i>E. coli</i> BL21(DE3)	<i>E. coli</i> B F <sup>−</sup> <i>dcm ompT hsdS</i> (rB– mB–) gal λ(DE3)	Studier and Moffatt, 1986
<i>E. coli</i> MS411	<i>ilvG rfb-50 thi</i>	M. Schembri, DTU, Denmark
<i>E. coli</i> CSH26Cm::LTL	Tc <sup>R</sup> , CSH26 <i>galK::cat::loxP-Tet-loxP</i>	Lang et al., 2010
<i>E. coli</i> DY330	W3110 Δ <i>lacU169 gal 490 ts λ</i> .cl857 Δ ( <i>cro-bioA</i> )	Yu et al., 2000
<i>E. coli</i> MS614	Sm <sup>R</sup> , <i>ilvG, rfb-50, thi, rpsL</i>	Beutin and Achtman, 1979
<i>E. coli</i> M1174	<i>strA</i> +, P–, B1–, <i>trp</i> –, <i>lys</i> –, <i>gal</i> –, <i>malA</i> –, <i>strR</i> , Su–, Δ( <i>lac</i> )X74	Silverman et al., 1991

<sup>a</sup>resistances; Tc<sup>R</sup>, tetracycline; Sm<sup>R</sup>, streptomycin; Cm<sup>R</sup>, chloramphenicol.

ampicillin (Amp) 100 µg/ml, tetracycline (Tc) 10 µg/ml, and streptomycin (Sm) 25 µg/ml.

Plasmids are listed in **Table 2**.

## DNA Preparation and Modification

All enzymes were used according to Manufacturer's recommendations.

Oligonucleotides are shown in **Table 3**.

## Construction of Complementation-, Cre-, and Expression-Plasmids

Inserts for pMS\_*parM* and derivatives were amplified with parMBamHI<sub>fw</sub>/parMEcoRI<sub>rev</sub>, for pMS\_*parR* with parRBamHI<sub>fw</sub>/parREcoRI<sub>rev</sub> using pMS\_*parA*, pJSC1-S39A, or pJSC1-K123A as templates. Amplicons were cut with *Bam*HI/*Eco*RI and ligated with pMS119HE. N-terminal Cre fusions were constructed by inserting amplicons from

**TABLE 2 | Plasmids used in this study.**

Plasmids	Description <sup>a</sup>	Reference
<b>R1 DERIVATIVES</b>		
R1-16miniTn5 E5	Km <sup>R</sup> , Cm <sup>R</sup> , R1-16 with miniTn5Cm inserted in the <i>parA</i> locus	Nuk et al., 2011
R1-16miniTn5 B4	Km <sup>R</sup> , Cm <sup>R</sup> , R1-16 with miniTn5Cm inserted in the <i>yjcA</i> locus	Nuk et al., 2011
R1-16Δ <i>parM</i>	Km <sup>R</sup> , R1-16 with gene specific <i>parM</i> null mutation	This study
R1-16Δ <i>parR</i>	Km <sup>R</sup> , R1-16 with gene specific <i>parR</i> null mutation	This study
R1-16Δ <i>parMR</i>	Km <sup>R</sup> , R1-16 with <i>parMR</i> double gene null mutation	This study
R1-16	Km <sup>R</sup> , IncFII, <i>fin</i> -	Goebel et al., 1977
R1-16Δ <i>tral</i>	Km <sup>R</sup> Tc <sup>R</sup> <i>tral::tetRA</i>	Lang et al., 2010
pOX38	Km <sup>R</sup> , derivative of F	Chandler and Galas, 1983
<b>COMPLEMENTATION/CLONING</b>		
pMS_ <i>parMR</i>	Amp <sup>R</sup> , <i>Eco</i> RI/ <i>Bam</i> HI fragment of pKG321 (Gerdes et al., 1985) in pMS119HE	This study
pMS_ <i>parM</i>	Amp <sup>R</sup> , pMS119HE with wild type R1 <i>parM</i> encoding residues 1–320	This study
pMS_ <i>parR</i>	Amp <sup>R</sup> , pMS119HE with wild type <i>parR</i> encoding residues 1–117	This study
pMS_ <i>parMS39A</i>	Amp <sup>R</sup> , pMS_ <i>parM</i> with a S39A sense mutation	This study
pMS_ <i>parMK123A</i>	Amp <sup>R</sup> , pMS_ <i>parM</i> with a K123A sense mutation	This study
pMS119HE	Amp <sup>R</sup> , P <sub>tac</sub> expression vector	Strack et al., 1992
pGZ119EH	Cm <sup>R</sup> , P <sub>tac</sub> expression vector	Lessl et al., 1992
<b>PROTEIN TRANSLOCATION</b>		
CFP B	Amp <sup>R</sup> , Cre-fusion vector	Parker and Meyer, 2007
CFP B Sm	Sm <sup>R</sup> , Cre-fusion vector	Lang et al., 2010
CreParM Sm	Sm <sup>R</sup> , CFP B Sm with R1 <i>parM</i> encoding residues 2–320	This study
CreParMS39A Sm	Sm <sup>R</sup> , CreParM Sm with a S39A mutation	This study
CreParMK123A Sm	Sm <sup>R</sup> , CreParM Sm with a K123A mutation	This study
CreParR Sm	Sm <sup>R</sup> , CFP B Sm with R1 <i>parR</i> encoding residues 2–117	This study
CreTral(3-1756) Sm	Sm <sup>R</sup> , CFP B Sm with R1 <i>tral</i> encoding residues 3–1756	Lang et al., 2010
CreTralF	Amp <sup>R</sup> , <i>Kpn</i> I- <i>Sal</i> I fragment from p99I+ (Haft et al., 2006) cloned into CFP B	Lang et al., 2010
<b>PROTEIN EXPRESSION</b>		
pJSC21	Amp <sup>R</sup> , P <sub>tac</sub> ParR expression construction	Salje and Lowe, 2008
pET3A-ParM	Amp <sup>R</sup> , P <sub>tac</sub> ParM expression construction	This study
pHP2	Cm <sup>R</sup> , P <sub>tac</sub> Tral expression construct, pGZ119HE with 6,1 kb <i>Asn</i> I-fragment from R1-16 in <i>Sma</i> I-site	Zechner et al., 1997
pGZTraD	Cm <sup>R</sup> , P <sub>tac</sub> TraD expression vector	This study
pGZTraC	Cm <sup>R</sup> , P <sub>tac</sub> TraC expression vector	This study
pMS-CFLAG <i>parM</i>	Amp <sup>R</sup> , P <sub>tac</sub> , C-terminal fusion of FLAG tag with <i>parM</i>	This study
pMS-NFLAG <i>parR</i>	Amp <sup>R</sup> , P <sub>tac</sub> , N-terminal fusion of FLAG tag with <i>parR</i>	This study
pMS-CFLAG <i>parR</i>	Amp <sup>R</sup> , P <sub>tac</sub> , C-terminal fusion of FLAG tag with <i>parR</i>	This study
pASKIBA7PLUSTraC	Amp <sup>R</sup> , P <sub>tet</sub> , N-terminal fusion of Strep tag with <i>traC</i>	This study
<b>RELAXASE/ATPASE ASSAY</b>		
pDE100	Amp <sup>R</sup> , <i>Bgl</i> II/ <i>Pst</i> I fragment of R1 <i>oriT</i> in <i>Eco</i> RI/ <i>Pst</i> I pBluescript II KS-	Csitkovits et al., 2004
pDE110	Amp <sup>R</sup> 388-bp insert in pBluescript II KS, including a 363-bp <i>Cla</i> I/ <i>Pst</i> I fragment of the R1 <i>traD</i> gene (positions 1766–2129, GenBank <sup>TM</sup> accession number AY684127)	Csitkovits et al., 2004
pBlue <i>parC</i>	Amp <sup>R</sup> , <i>Eco</i> RI- <i>parC-Eco</i> RI PCR fragment cloned via <i>Eco</i> RI restriction site into pBluescript II KS(-)	This study

<sup>a</sup>resistances; Amp<sup>R</sup>, ampicillin; Km<sup>R</sup>, kanamycin; Sm<sup>R</sup>, streptomycin; Cm<sup>R</sup>, chloramphenicol.



**TABLE 3 | Oligos used in this study.**

Name	5' -3' Sequence
<b>CLONING</b>	
parMRCFW	GACGAAT7CCACTTTTGTACCGCC
parCrev	GACGAAT7CTTAATTATAAACTCCTTATGG
parM NdeI FW	TTTTCATATGTTGGTATTCATTGATGACG
parM BamHI REV	TTTGGATCCCTTAATTACCTATGAGATACATACC
ParM_SFW	ATAGTAGGTACCTTGGTATTCATTGATGACG
ParM_SRev	GCAATCGTCGACTTAATTACCTATGAGATACATACCGT
ParR_FW	ATAGTAGGTACC-ATGGACAAGCGCAGAACC
ParR_Rev	GCAATCGTCGACTTATTAATTTATTAGCTTCATCGC
parR_NheIFW	GTAGCTAGCATGATGGACAAGCGCAGAACC
parR_NheIRev	GTAGCTAGCATTATTAGCTTCATCGC
parMBamHIfw	TTTGGATCCATGTTGGTATTCATTGATGACG
parMEcoRIrev	TTTGAAT7CTTAATTACCTATGAGATACATACC
parRBamHIfw	TTTGGATCCATGATGGACAAGCGCAGAACC
parREcoRIrev	TTTGAAT7CTTAATTATTAGCTTCATCG
ParMloxFW	CCCCAAACATACCCAAACACACACCAAAAAACACCATAAGGAGTTTTATAatataactcgtataG
ParMloxRev	GTTTGATTACATCTGGATTAGTTTGAAGGCAATGGTTCTGCCTTGTCATCAGGataactcgtataA
FW_parR_TetRA	AACCAATAACTCTCAATATGATTAGTTAACGGTATGTATCTCATAGGTAATTACAAGAATTGCCGGCGGAT
Rev_parR_TetRA	GTTCCCTTTATCCAGCCTGATAGTGGATAAAGGGAACCAATAAATTGAAGGTATTTACACCGCATAGC
SS01	GCCGAAT7CATGAGTTTTAACGCAAAG
SS02	CGTGAAGCTTTTCAGAAATCATCTCCCG
parM_CFLAG_EcoRI_rev	TTATAGAAT7CTTACTTGTGCATCGTCATCCTTGTAAATCATTACCTATGAGATACATACC
parR_NFLAG_BamHI_fw	TATAGGATCCATGATTACAAGGATGACGATGACAAGATGGACAAGCGCAGAACC
parR_CFLAG_EcoRI_rev	TTATAGAAT7CTTACTTGTGCATCGTCATCCTTGTAAATCATTATTAGCTTCATCGC
StrepTraCEcorIFw	AAGAAT7CAATAACCCACTTGAGGCCG
StrepTraCHindIIIR	TTTAAGCT7TCATGCAACACTCCTGTATT
<b>EMSA AND FLUORESCENCE INTENSITY</b>	
parC*	AAACAAAACCCAAAAACAACCC
parCcomp	GGGTTGTTTTGGGTTTTGTTT
oriT*	ACCAAAAGCACCCACCCACGCAAAAACAAG
oriTcomp	CTTGTTTTGCGTGGGGTGTGGTGCTTTTGGT
randomseq*	CGAACGAGCAGTTGTTTCAGCG
randomseqcomp	CGCTGAAACAACTGCTCGTTCG
oriT17*	TTTGCGTGGGGTGTGGT
2 x G144C	TTTTCGCTGGGCTGTGGTCTTTGCGTGGGCTGTGGTCTTT

\* = 3'-carboxytetramethyl-rhodamine (TAMRA) labeled, Restriction sites are marked in cursive, FLAG-sequences are underscored, lox-sites are in lower-case letters.

templates pMS\_*parM*, pMS\_*parR*, or derivatives in CFB B Sm plasmid via *KpnI/SalI*. Primers for the *parM* inserts: ParM\_SFW/ParM\_SRev; for *parR* insert: ParR\_FW/ParR\_Rev. To generate the C-terminal CreParR fusion plasmid *parR* was amplified with parR\_NheIFW/parR\_NheIRev from template pMS\_*parA*, and ligated to *NheI* cut CFP B. The *parM* PCR fragment from parMNdeI\_fw/parMBamHI\_rev and R1-16 template were cut with *NdeI* and *BamHI* and ligated to yield pet3A-*parM*. pASKIBA7PLUSTraC was generated by amplification of *traC* from R1-16 with primers StrepTraCEcorIFw and StrepTraCHindIIIR, cutting with *EcoRI* and *HindIII* and ligation in pASKIBA7PLUS.

pGZ*traD* was created by amplification of *traD* from R1-16 with primers SS01 and SS02, *EcoRI* and *HindIII* treated and inserted in pGZ119EH. To generate pGZ*traC*, *traC* was cut from R1-16 with *EcoRI* and *SmaI* and ligated into pGZ119EH. pMS-CFLAGparM was constructed by amplification of *parM* with parM\_CFLAG\_EcoRI\_rev/parMBamHI\_fw, cutting with *EcoRI* and *BamHI* and ligation into pMS119EH. pMS-NFLAGparR and pMS-CLFAGparR were constructed by amplification of *parR* with primers parR\_NFLAG\_BamHI and parREcoRIrev or parRBamHIfw and parR\_CFLAG\_EcoRI\_rev, respectively, restriction with *BamHI/EcoRI* and ligation with pMS119EH. pBlueparC was constructed by amplification of *parC* from R1-16

with primers parMRCFW and parCrev, restriction with *EcoRI* and ligation with pBluescript II KS(–).

## Construction of *parM*, *parR*, and *parMR* Null Derivatives

To generate R1-16Δ*parM*, primers ParMloxFW/ParMloxRev were used to amplify a *loxP*-TetRA-*loxP* cassette from *E. coli* CSH26Cm::LTL. For R1-16Δ*parR*, FW\_parR\_TetRA/Rev\_parR\_TetRA were used to amplify a tetracycline resistance cassette from pAR183 (Reisner et al., 2006). For R1-16Δ*parM R*, FW\_R1parM/Rev\_R1parR were used to amplify a streptomycin resistance cassette from pAH144 (Haldimann and Wanner, 2001). The amplified fragments were introduced into *E. coli* DY330 [R1-16] and integrated via homologous recombination (Reisner et al., 2002). Introduction of the CFP B plasmid into strains carrying R1-16 mutants catalyzed a Cre/*loxP* mediated recombination reaction excising the *tetRA* cassette.

## Fluorescence Microscopy

Alexa488 labeled R17 phage was prepared as described (Lang et al., 2014). *E. coli* MS411 carrying R1-derivatives with and without complementation plasmids were grown to an  $A_{600}$  of 0.6–0.8, diluted in PBS to an  $A_{600}$  of 0.5 and incubated with 0.01 vol. R17 phage for 10 min at RT. Five microliters were mounted on a glass slide, pictures were taken with an Eclipse Ti fluorescence microscope (Nikon).

## Copy Number Determination

For quantification of apparent copy number, plasmid yields of R1 derivatives (R1-16, R1-16::B4, R1-16miniTn5 E5, or R1-16miniTn5 B4) were determined and compared to the yields of an independent replicon (pGZ119EH) as described in Nuk et al. (2011). *E. coli* M1174 carrying the desired plasmid combinations were grown in 5 ml LB medium with antibiotics. One millimolar IPTG induced *traI* (pHP2) expression.

## Mating Experiments and Cre Recombinase Assay for Translocation (CRAFT)

*E. coli* MS411 carrying the plasmids were grown overnight in LB media with antibiotics at 37°C. Hundred microliters of donor cells were centrifuged for 3 min at 3000 × g, resuspended in 1 ml 0.9% NaCl, subcultured in drug free LB for 1 h at 37°C and adjusted to  $A_{600}$  0.02. A 10-fold excess of recipient MS614 was added and the mixture was incubated for 3, 5, 15, or 30 min at 37°C without shaking. DNA transfer was stopped by vortexing for 1 min and rapid cooling on ice. Donors were selected on LB-plates with antibiotics (see Table 2) and transconjugants were selected on kanamycin (40 µg/ml) and streptomycin (25 µg/ml). Conjugation frequencies were calculated as transconjugants per donor.

CRAFT was performed as described previously (Lang et al., 2010). *E. coli* MS411 carrying the plasmids of interest and recipient CSH26Cm::LTL were used. Donors were selected on plates with antibiotics (Table 2). Transconjugants and recombinants were identified by plating serial dilutions on LB-Kan/Tc or chloramphenicol, respectively.

Conjugation and protein translocation frequencies are calculated as transconjugants or recombinants per donor, respectively.

## Protein Purification

TraI, TraIN309, TraIN992, TraDΔN130 were purified as described (Csitkovits et al., 2004; Mihajlovic et al., 2009; Sut et al., 2009; Lang et al., 2011).

ParR purification: 2 l *E. coli* C41(DE3) [pJSC21] were grown in LB with 100 µg/ml ampicillin (LB-Amp) at 37°C with shaking to an OD<sub>600</sub> of 0.6. One millimolar IPTG was used for induction. After 6 h, cells were harvested (10 min, 4000 × g, 4°C), resuspended in 20 ml buffer I (50 mM Tris-HCl pH 7, 100 mM KCl, 1 mM EDTA, 1 mM DTT, 0.001% PMSF) with proteinase inhibitor cocktail (cOmplete EDTA-free, Roche) and frozen at –80°C. Cells were lysed, incubated with DNase I (AppliChem) for 20 min on ice and centrifuged (140,000 × g, 1 h, 4°C). The cytoplasmic fraction was filtered through a 0.45 µm PVDF syringe filter and loaded (1 ml/min) on HiTrap Heparin columns equilibrated with buffer I. After washing (2 column volumes buffer I), bound protein was eluted with buffer I and a linear gradient of 0–1 M NaCl. Partially pure ParR eluted at ~300–450 mM NaCl. These fractions were pooled and dialyzed overnight against 100 × vol. buffer I. Dialyzed fractions were loaded (1 ml/min) on 2 serially connected 5 ml HiTrap SP HP columns equilibrated with buffer I. After washing (2 ml/min) with 10 ml of buffer I, protein was eluted with a linear gradient of buffer I + 0–1 M NaCl. ParR eluted at ~450–500 mM NaCl. Purity was confirmed by SDS-PAGE and fractions were pooled and dialyzed overnight [100 vol. buffer II (20 mM Tris-HCl pH 9, 50 mM KCl, 1 mM EDTA, 1 mM Na<sub>3</sub>N)]. ParR was concentrated, adjusted to 20% glycerol and frozen at –80°C.

## ParM Purification

*E. coli* BL21 (DE3) carrying pet3A-*parM* were grown in 1 l LB-Amp at 37°C. At OD<sub>600</sub> 0.6 expression was induced with 1 mM IPTG. After 3 h, cells were harvested (4000 × g, 10 min, 4°C), resuspended in 20 ml buffer I (30 mM Tris-HCl pH 7.5, 25 mM KCl, 1 mM MgCl<sub>2</sub>, 1 mM DTT, 0.001% PMSF) with protease inhibitor (cOmplete EDTA-free, Roche) and frozen at –80°C. Lysis was performed as described above. The lysate was cleared by centrifugation (21,000 × g, 1 h, 4°C) and the supernatant was precipitated by addition of solid (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> to 30% w/v with stirring on ice (1 h). After centrifugation (77,000 × g, 1 h, 4°C) the pellet was dissolved in 20 ml buffer I and dialyzed overnight at 4°C (100 vol buffer I). Dialyzed solution was filtered (0.45 µm) loaded on 2 × 5 ml HiTrap Heparin columns equilibrated with buffer I. Flowthrough was collected and loaded onto 2 × 5 ml HitrapQ columns equilibrated with buffer I and washed with 10 ml of buffer I. Protein was eluted with a 4 step gradient with buffer I + 1 M KCl (12, 15, 20, 30%, 5 ml buffer flow between steps, followed by 10 ml isocratic elution for each step). ParM eluted at ~200 mM KCl and purity of fractions was confirmed with SDS-PAGE. Pure fractions were pooled and dialyzed overnight at 4°C (100 vol buffer I), concentrated, adjusted to 20% glycerol and frozen at –80°C.

## TraC Purification

*E. coli* BL21 C41 (DE3) carrying pASKIBA7PLUSTraC were grown in 1 l LB-Amp at 37°C to an OD<sub>600</sub> of 0.5. Overexpression was induced by addition of anhydrotetracyclin (AHT, 0.2 mg/l). Cells were harvested after 4 h shaking at 37°C, pellets were frozen at −80°C. Pellet was resuspended in 20 ml of buffer I (50 mM Tris-Cl pH 7.7, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 1 tablet protease inhibitor (cOmplete, Roche) and lysed. The lysate was centrifuged for 45 min at 50,000 × g at 4°C. The supernatant was filtered (0.45 μm PVDF syringe filters) and loaded onto a Hitrap strep HP column pre-equilibrated with buffer I. After washing (5–10 column volumes) with buffer I, TraC was eluted with 30 ml buffer II (buffer I containing 2.5 mM desthiobiotin) in one step. Fractions containing TraC were pooled and concentrated (Amicon centrifugal filter, Millipore) and loaded onto a Hiload 16/60 Superdex 200 column. The protein was eluted with buffer III (25 mM Tris pH 7.6, 100 mM NaCl, 1 mM DTT, 1 mM MgCl<sub>2</sub>, 1 mM PMSF). Pure TraC fractions were pooled, concentrated and frozen at −80°C (with 20% glycerol). Identity of TraC was confirmed by mass spectrometry and apparent molecular mass 99 kDa was determined (SDS-PAGE and Coomassie staining).

## Relaxase Assay

*oriT* specific cleavage activity was determined as described in Csitkovits et al. (2004). Indicated ParM and ParR concentrations were combined with 4 nM of pDE100 or pDE110 (negative control) independently or in combination, then the cleavage reaction was started by TraI (25 nM), TraIN308 (300 nM), or TraIN992 (100 nM). Statistical significance was determined using maximum stimulatory concentrations of ParM (15 nM) and ParR (15 nM) and 500 nM BSA as a negative control.

## T-Strand Cleavage and Unwinding

Construction of heteroduplex-substrates and the assay conditions were as described in Csitkovits et al. (2004) and Sut et al. (2009). Each heteroduplex substrate G2028 or IR (1 nM) was combined with effector proteins ParM and ParR independently or in combination in concentrations that were most stimulatory for TraI in the relaxase assay. Twenty-five nanomolar TraI was added to start the reaction. Resolution and quantitation of unwound DNA product were as described (Sut et al., 2009).

## ATPase Assay

Enzyme activities were measured with the Malachite Green Assay Kit (Bioassay Systems). Briefly, different ParM concentrations (0–1 mM) with and without ParR (9 mM) and *parC* (pBlueparC, 17 nM) or DNA not containing *parC* (pDE110, 17 nM) were incubated (30 mM Tris-HCl pH 7.5, 50 mM KCl, 0.2 mM MgCl<sub>2</sub>, 1 mM DTT, 0.1 mg/ml bovine serum albumin, 0.1 mM ATP; Jensen and Gerdes, 1997) at 30°C in a total volume of 25 μl. After 10 min, the reaction was stopped and color development after 30 min at RT was recorded at 595 nm. Stimulation of ParM ATPase activity by TraI, TraD, or TraC: 0.5 or 1 mM ParM were titrated with TraI, TraIN308, or TraIN992 (10–100 nM), TraD (20–500 nM), or TraC (0.5–8 mM), respectively, in the reaction buffer described above.

Impact of Par-Proteins on TraI ATP-hydrolysis: 1 fmol M13mp18 single-stranded phage DNA (New England BioLabs) was preincubated with ParR (0.5 mM) and ParM (9 mM) in buffer containing ATP (25 mM Tris HCl pH 7.5, 20 mM NaCl, 3 mM MgCl<sub>2</sub>, 5 mM DTT, 2 mM ATP). TraI addition (0–10 nM) started the reaction. Basal TraI ATPase activity was 225,892 ± 83,485 mol/h/mol.

## Electrophoretic Mobility Shift (EMSA)

Oligos for fluorescence studies were reconstituted in 1 × STE (10 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA). To create dsDNA substrates 3'-TAMRA labeled oligos were mixed in a 1:1 ratio with the unlabelled complementary strand, heated for 10 min, 96°C and re-hybridized at RT for 1 h. DNA binding by ParR was tested with ssDNA and dsDNA substrates. The consensus sequence of *parC* (Moller-Jensen et al., 2003) was created with Weblogo 3.4 (Crooks et al., 2004), a random DNA sequence was created with SMS (Stothard, 2000). Briefly, ds- or ssDNA *parC*\*, *oriT*\*, and *randomseq*\* (all 20 nM) were titrated with ParR in a total volume of 15 μl in 10 mM Tris HCl pH7.5, 50 mM NaCl, 0.005% NONIDET P-40, 1 mM DTT, 1 mM EDTA. The reaction was incubated for 20 min at 25°C, then 2 μl 87% glycerol were added. The products were resolved on 8% PAGE-gel without SDS at 15 V/cm for 30 min. Gels were scanned with a Typhoon 9400.

## Affinity for ssDNA

TraI affinity for 3'-TAMRA *oriT* DNA was measured on an ATF-105 fluorometer (Aviv Biomedical, Inc., Lakewood, NJ) as described in Dostal and Schildbach (2010). Briefly, 4 nM substrates (*oriT*17\*) with and without unlabelled competitor (50 nM 2 × G144C, Dostal et al., 2011) were preincubated with protein (ParR 10 nM, ParM 10 nM, 20 nM) at RT for 10 min in binding buffer (20 mM TrisHCl pH 7.5, 100 mM NaCl, 1 mM EDTA), the reaction was started with 0–100 nM TraI with a Microlab 500 titrator (Hamilton). A constant temperature of 25°C was maintained. Excitation wavelength was 520 nm, change in fluorescence intensity was followed at 580 nm. Equilibration time between each titration step was 3 min with mixing. Datapoints were averaged over 10 s. Volume correction and data fitting was done as described (Dostal and Schildbach, 2010).

## Co-Immunoprecipitation of Par- with Tra-Proteins

The protocol was adapted from the TrIP assay (Cascales, Christie 2004). Hundred milliliters LB was inoculated with ONCs of *E. coli* MS411 strains carrying plasmid combinations. Thirty OD of exponentially growing (LB with antibiotics, 1 mM IPTG) cells were pelleted and resuspended in 45 ml 20 mM sodium phosphate buffer (pH = 6.8) with 1% formaldehyde and incubated for 15 min at RT. Five milliliters glycine (1.2 M in 20 mM sodium phosphate buffer pH = 6.8) were added, cells were pelleted and washed with 50 ml/30 OD of buffer A (50 mM Tris-HCl pH 6.8, 100 mM NaCl). For lysis, pellet was resuspended in 1 ml buffer B (10 mM Tris-HCl pH 8.0, 10 mM MgCl<sub>2</sub>, 1 mg/ml lysozyme) and transferred to a 2 ml Eppendorf

tube and frozen for 2 min with liquid N<sub>2</sub>, thawed on ice, frozen and thawed again. Each sample was sonicated for 10 s on ice. One, two milliliters of buffer C was added and adjusted with Triton X-100 to a final concentration of 4% and the lysate was incubated for 15 min with rotation at RT. Two hundred and thirty microliters cOmplete EDTA-free, Roche (1 tablet in 1 ml 25 mM MgCl<sub>2</sub>) was added and the mixture was incubated for 30 min at 37°C with shaking. 6.4 ml of buffer C (150 mM Tris-HCl pH 8, 0.5 M sucrose, 10 mM EDTA) were added and insoluble material was removed by centrifugation (14,000 × g, 15 min). At this point, 200 µl aliquots of the supernatant were saved and stored at -80°C (whole cell lysate fraction, WCL). Remaining supernatant was transferred to tubes with Anti-FLAG affinity gel (90 µl/sample, A2220, Sigma) and incubated over night at 4°C. Beads were pelleted at 5000 × g, supernatant was removed. Beads were washed once with 950 µl buffer C supplemented with 1% Triton X-100 and twice with 950 µl buffer C supplemented with 0.1% Triton X-100. Immunoprecipitates (IP) were eluted by incubation of the beads for 30 min with FLAG-peptide (F3290, Sigma) at RT [80 µl FLAG-peptide (1 mg/ml in buffer C)/40 µl beads]. Beads were pelleted and the supernatant was collected (IP fraction).

## Western Analysis

A<sub>600</sub> 0.015–0.03 equivalents of the WCL and IP fractions were mixed with sample buffer containing DTT and SDS and resolved on SDS-PAGE (10%). Gels were blotted for 1.5 h on PVDF membranes. Blocking was done for 2 h (3% milk in TST). Detection of TraI and TraD -proteins was done with rabbit-antisera and α-rabbit HRP-conjugated antibody (7074S, Cellsignalling). Affinity purified TraC antibodies raised against TraC were produced by immunoGlobe. FLAG-tagged proteins

were detected by HRP-conjugated α-FLAG antibody (A8592, Sigma). After washing (3 × 5 min) with 1 × TST, secondary antibody was incubated for 1 h. Blot development was done with ECL (Clarity Western, Bio-Rad).

## Computer Programs

ImageJ (Schneider et al., 2012) was used to quantify bands in gel-electrophoresis assays and SigmaPlot 12.2.0.45 and Qtiplot 0.9.8.3.3 were used to plot data.

## AUTHOR CONTRIBUTIONS

CG, SL, JS, and EZ designed the research. CG, SL, VR, SR, and MN did the experiments. CG, SL, JS, and EZ analyzed data. CG and EZ wrote the paper.

## ACKNOWLEDGMENTS

This research was supported by Austrian Science Fund (FWF) grants P-24016 and W901 DK Molecular Enzymology, the European Fund for Regional Development and the province of Styria (grant A3-11.B-38/2011-6) and BioTechMed-Graz. Work in the Schildbach laboratory was supported by National Institutes of Health grant GM61017. We are grateful to Chris Larkin, Sonja Kirchberger, and Bettina Konrad for their contributions to this study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmolb.2016.00032>

## REFERENCES

- Atmakuri, K., Cascales, E., Burton, O. T., Banta, L. M., and Christie, P. J. (2007). *Agrobacterium* ParA/MinD-like VirC1 spatially coordinates early conjugative DNA transfer reactions. *EMBO J.* 26, 2540–2551. doi: 10.1038/sj.emboj.7601696
- Berry, T. M., and Christie, P. J. (2011). Caught in the act: the dialogue between bacteriophage R17 and the type IV secretion machine of plasmid R1. *Mol. Microbiol.* 82, 1039–1043. doi: 10.1111/j.1365-2958.2011.07870.x
- Beutin, L., and Achtman, M. (1979). Two *Escherichia coli* chromosomal cistrons, *sfrA* and *sfrB*, which are needed for expression of F factor *tra* functions. *J. Bacteriol.* 139, 730–737.
- Bharat, T. A. M., Murshudov, G. N., Sachse, C., and Löwe, J. (2015). Structures of actin-like ParM filaments show architecture of plasmid-segregating spindles. *Nature* 523, 106–110. doi: 10.1038/nature14356
- Campbell, C. S., and Mullins, R. D. (2007). *In vivo* visualization of type II plasmid segregation: bacterial actin filaments pushing plasmids. *J. Cell Biol.* 179, 1059–1066. doi: 10.1083/jcb.200708206
- Chandler, M., and Galas, D. J. (1983). Cointegrate formation mediated by Tn9. II. Activity of IS1 is modulated by external DNA sequences. *J. Mol. Biol.* 170, 61–91. doi: 10.1016/S0022-2836(83)80227-7
- Christie, P. J., and Gordon, J. E. (2014). The *Agrobacterium* ti plasmids. *Microbiol. Spectr.* 2. doi: 10.1128/microbiolspec.PLAS-0010-2013
- Clewell, D. B., Weaver, K. E., Dunny, G. M., Coque, T. M., Francia, M. V., and Hayes, F. (2014). “Extrachromosomal and mobile elements in enterococci: transmission, maintenance, and epidemiology: enterococci: from commensals to leading causes of drug resistant infection,” in *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection*, eds M. S. Gilmore, D. B. Clewell, Y. Ike, and N. Shankar (Boston, MA: Massachusetts Eye and Ear Infirmary), 1–84.
- Crooks, G., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- Csitkovits, V. C., Dermic, D., and Zechner, E. L. (2004). Concomitant reconstitution of TraI-catalyzed DNA transesterase and DNA helicase activity *in vitro*. *J. Biol. Chem.* 279, 45477–45484. doi: 10.1074/jbc.M407970200
- Csitkovits, V. C., and Zechner, E. L. (2003). Extent of single-stranded DNA required for efficient TraI helicase activity *in vitro*. *J. Biol. Chem.* 278, 48696–48703. doi: 10.1074/jbc.M310025200
- de la Cruz, F., Frost, L. S., Meyer, R. J., and Zechner, E. L. (2010). Conjugative DNA metabolism in Gram-negative bacteria. *FEMS Microbiol. Rev.* 34, 18–40. doi: 10.1111/j.1574-6976.2009.00195.x
- Disque-Kochem, C., and Dreiseikelmann, B. (1997). The cytoplasmic DNA-binding protein TraM binds to the inner membrane protein TraD *in vitro*. *J. Bacteriol.* 179, 6133–6137.
- Dostal, L., and Schildbach, J. F. (2010). Single-stranded DNA binding by F TraI relaxase and helicase domains is coordinately regulated. *J. Bacteriol.* 192, 3620–3628. doi: 10.1128/JB.00154-10
- Dostal, L., Shao, S., and Schildbach, J. F. (2011). Tracking F plasmid TraI relaxase processing reactions provides insight into F plasmid transfer. *Nucleic Acids Res.* 39, 2658–2670. doi: 10.1093/nar/gkq1137



- Frost, L. S., and Koraimann, G. (2010). Regulation of bacterial conjugation: balancing opportunity with adversity. *Future Microbiol.* 5, 1057–1071. doi: 10.2217/fmb.10.70
- Garner, E. C., Campbell, C. S., and Mullins, R. D. (2004). Dynamic instability in a DNA-segregating prokaryotic actin homolog. *Science* 306, 1021–1025. doi: 10.1126/science.1101313
- Garner, E. C., Campbell, C. S., Weibel, D. B., and Mullins, R. D. (2007). Reconstitution of DNA segregation driven by assembly of a prokaryotic actin homolog. *Science* 315, 1270–1274. doi: 10.1126/science.1138527
- Gerdes, K., Larsen, J. E., and Molin, S. (1985). Stable inheritance of plasmid R1 requires two different loci. *J. Bacteriol.* 161, 292–298.
- Ghigo, J. M. (2001). Natural conjugative plasmids induce bacterial biofilm development. *Nature* 412, 442–445. doi: 10.1038/35086581
- Gibert, M., Juarez, A., Zechner, E. L., Madrid, C., and Balsalobre, C. (2014). TrhR, TrhY and HtdA, a novel regulatory circuit that modulates conjugation of the IncHI plasmids. *Mol. Microbiol.* 94, 1146–1161. doi: 10.1111/mmi.12823
- Goebel, W., Lindenmaier, W., Schrempp, H., Kollek, R., and Blohm, D. (1977). “Dissociation and recombination of fragments with defined functions of the antibiotic resistance factor R1,” in *Topics in Infectious Diseases*, eds J. Drews and G. Högenauer (New York, NY: Springer Verlag), 261–275.
- Guynet, C., Cuevas, A., Moncalian, G., and de la Cruz, F. (2011). The *stb* operon balances the requirements for vegetative stability and conjugative transfer of plasmid R388. *PLoS Genet.* 7:e1002073. doi: 10.1371/journal.pgen.1002073
- Haft, R. J., Palacios, G., Nguyen, T., Mally, M., Gachelet, E. G., Zechner, E. L., et al. (2006). General mutagenesis of F plasmid TraI reveals its role in conjugative regulation. *J. Bacteriol.* 188, 6346–6353. doi: 10.1128/JB.00462-06
- Haldimann, A., and Wanner, B. (2001). Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J. Bacteriol.* 183, 6384–6393. doi: 10.1128/JB.183.21.6384-6393.2001
- Hamilton, H. L., Dominguez, N. M., Schwartz, K. J., Hackett, K. T., and Dillard, J. P. (2005). *Neisseria gonorrhoeae* secretes chromosomal DNA via a novel type IV secretion system. *Mol. Microbiol.* 55, 1704–1721. doi: 10.1111/j.1365-2958.2005.04521.x
- Harley, M. J., Toptygin, D., Troxler, T., and Schildbach, J. F. (2002). R150A mutant of F TraI relaxase domain: reduced affinity and specificity for single-stranded DNA and altered fluorescence anisotropy of a bound labeled oligonucleotide. *Biochemistry* 41, 6460–6468. doi: 10.1021/bi011969i
- Hekman, K., Guja, K., Larkin, C., and Schildbach, J. F. (2008). An intrastrand three-DNA-base interaction is a key specificity determinant of F transfer initiation and of F TraI relaxase DNA recognition and cleavage. *Nucleic Acids Res.* 36, 4565–4572. doi: 10.1093/nar/gkn422
- Jensen, R. B., and Gerdes, K. (1997). Partitioning of plasmid R1. The ParM protein exhibits ATPase activity and interacts with the centromere-like ParR-parC complex. *J. Mol. Biol.* 269, 505–513. doi: 10.1006/jmbi.1997.1061
- Jensen, R. B., and Gerdes, K. (1999). Mechanism of DNA segregation in prokaryotes: parM partitioning protein of plasmid R1 co-localizes with its replicon during the cell cycle. *EMBO J.* 18, 4076–4084. doi: 10.1093/emboj/18.14.4076
- Jones, C. W., and Armitage, J. P. (2015). Positioning of bacterial chemoreceptors. *Trends Microbiol.* 23, 247–256. doi: 10.1016/j.tim.2015.03.004
- Karl, W., Bamberger, M., and Zechner, E. L. (2001). Transfer protein TraY of plasmid R1 stimulates TraI-catalyzed *oriT* cleavage *in vivo*. *J. Bacteriol.* 183, 909–914. doi: 10.1128/JB.183.3.909-914.2001
- Kuhn, B., Abdel-Monem, M., Krell, H., and Hoffmann-Berling, H. (1979). Evidence for two mechanisms for DNA unwinding catalyzed by DNA helicases. *J. Biol. Chem.* 254, 11343–11350.
- Lang, S., Gruber, C. J., Raffl, S., Reisner, A., and Zechner, E. L. (2014). Common requirement for the relaxosome of plasmid R1 in multiple activities of the conjugative type IV secretion system. *J. Bacteriol.* 196, 2108–2121. doi: 10.1128/JB.00045-13
- Lang, S., Gruber, K., Mihajlovic, S., Arnold, R., Gruber, C. J., Steinlechner, S., et al. (2010). Molecular recognition determinants for type IV secretion of diverse families of conjugative relaxases. *Mol. Microbiol.* 78, 1539–1555. doi: 10.1111/j.1365-2958.2010.07423.x
- Lang, S., Kirchberger, P. C., Gruber, C. J., Redzej, A., Raffl, S., Zellnig, G., et al. (2011). An activation domain of plasmid R1 TraI protein delineates stages of gene transfer initiation. *Mol. Microbiol.* 82, 1071–1085. doi: 10.1111/j.1365-2958.2011.07872.x
- Lang, S., and Zechner, E. L. (2012). General requirements for protein secretion by the F-like conjugation system R1. *Plasmid* 67, 128–138. doi: 10.1016/j.plasmid.2011.12.014
- Lanka, E., and Wilkins, B. M. (1995). DNA processing reactions in bacterial conjugation. *Annu. Rev. Biochem.* 64, 141–169. doi: 10.1146/annurev.bi.64.070195.001041
- Lessl, M., Balzer, D., Lurz, R., Waters, V. L., Guiney, D. G., and Lanka, E. (1992). Dissection of IncP conjugative plasmid transfer: definition of the transfer region Tra2 by mobilization of the TraI region in trans. *J. Bacteriol.* 174, 2493–2500.
- Lu, J., and Frost, L. S. (2005). Mutations in the C-terminal region of TraM provide evidence for *in vivo* TraM-TraD interactions during F-plasmid conjugation. *J. Bacteriol.* 187, 4767–4773. doi: 10.1128/JB.187.14.4767-4773.2005
- Lutkenhaus, J. (2012). The ParA/MinD family puts things in their place. *Trends Microbiol.* 20, 411–418. doi: 10.1016/j.tim.2012.05.002
- Matson, S. W., Nelson, W. C., and Morton, B. S. (1993). Characterization of the reaction product of the *oriT* nicking reaction catalyzed by *Escherichia coli* DNA helicase I. *J. Bacteriol.* 175, 2599–2606.
- Matson, S. W., and Ragonese, H. (2005). The F-plasmid TraI protein contains three functional domains required for conjugative DNA strand transfer. *J. Bacteriol.* 187, 697–706. doi: 10.1128/JB.187.2.697-706.2005
- Matson, S. W., Sampson, J. K., and Byrd, D. R. (2001). F plasmid conjugative DNA transfer: the TraI helicase activity is essential for DNA strand transfer. *J. Biol. Chem.* 276, 2372–2379. doi: 10.1074/jbc.M008728200
- Mihajlovic, S., Lang, S., Sut, M. V., Strohmaier, H., Gruber, C. J., Koraimann, G., et al. (2009). Plasmid R1 conjugative DNA processing is regulated at the coupling protein interface. *J. Bacteriol.* 191, 6877–6887. doi: 10.1128/JB.00918-09
- Miroux, B., and Walker, J. E. (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* 260, 289–298. doi: 10.1006/jmbi.1996.0399
- Moller-Jensen, J., Borch, J., Dam, M., Jensen, R. B., Roepstorff, P., and Gerdes, K. (2003). Bacterial mitosis: parM of plasmid R1 moves plasmid DNA by an actin-like insertional polymerization mechanism. *Mol. Cell* 12, 1477–1487. doi: 10.1016/S1097-2765(03)00451-9
- Moller-Jensen, J., Ringgaard, S., Mercogliano, C. P., Gerdes, K., and Lowe, J. (2007). Structural analysis of the ParR/parC plasmid partition complex. *EMBO J.* 26, 4413–4422. doi: 10.1038/sj.emboj.7601864
- Nelson, W. C., Howard, M. T., Sherman, J. A., and Matson, S. W. (1995). The *traY* gene product and integration host factor stimulate *Escherichia coli* DNA helicase I-catalyzed nicking at the F plasmid *oriT*. *J. Biol. Chem.* 270, 28374–28380. doi: 10.1074/jbc.270.47.28374
- Nuk, M. R., Reisner, A., Neuwirth, M., Schilcher, K., Arnold, R., Jehl, A., et al. (2011). Functional analysis of the *finO* distal region of plasmid R1. *Plasmid* 65, 159–168. doi: 10.1016/j.plasmid.2010.12.002
- Pachulec, E., Siewering, K., Bender, T., Heller, E.-M., Salgado-Pabon, W., Schmolter, S. K., et al. (2014). Functional analysis of the gonococcal genetic island of *Neisseria gonorrhoeae*. *PLoS ONE* 9:e109613. doi: 10.1371/journal.pone.0109613
- Parker, C., and Meyer, R. J. (2007). The R1162 relaxase/primase contains two, type IV transport signals that require the small plasmid protein MobB. *Mol. Microbiol.* 66, 252–261. doi: 10.1111/j.1365-2958.2007.05925.x
- Ptacin, J. L., Gahlmann, A., Bowman, G. R., Perez, A. M., von Diezmann, A. R. S., Eckart, M. R., et al. (2014). Bacterial scaffold directs pole-specific centromere segregation. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2046–E2055. doi: 10.1073/pnas.1405188111
- Reisner, A., Haagensen, J. A. J., Schembri, M. A., Zechner, E. L., and Molin, S. (2003). Development and maturation of *Escherichia coli* K-12 biofilms. *Mol. Microbiol.* 48, 933–946. doi: 10.1046/j.1365-2958.2003.03490.x
- Reisner, A., Holler, B. M., Molin, S., and Zechner, E. L. (2006). Synergistic effects in mixed *Escherichia coli* biofilms: conjugative plasmid transfer drives biofilm expansion. *J. Bacteriol.* 188, 3582–3588. doi: 10.1128/JB.188.10.3582-3588.2006
- Reisner, A., Molin, S., and Zechner, E. L. (2002). Recombinogenic engineering of conjugative plasmids with fluorescent marker cassettes. *FEMS Microbiol. Ecol.* 42, 251–259. doi: 10.1111/j.1574-6941.2002.tb01015.x

- Roberts, M. A. J., Wadhams, G. H., Hadfield, K. A., Tickner, S., and Armitage, J. P. (2012). ParA-like protein uses nonspecific chromosomal DNA binding to partition protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6698–6703. doi: 10.1073/pnas.1114000109
- Salje, J., Gayathri, P., and Lowe, J. (2010). The ParMRC system: molecular mechanisms of plasmid segregation by actin-like filaments. *Nat. Rev. Microbiol.* 8, 683–692. doi: 10.1038/nrmicro2425
- Salje, J., and Lowe, J. (2008). Bacterial actin: architecture of the ParMRC plasmid DNA partitioning complex. *EMBO J.* 27, 2230–2238. doi: 10.1038/emboj.2008.152
- Salje, J., Zuber, B., and Lowe, J. (2009). Electron cryomicroscopy of *E. coli* reveals filament bundles involved in plasmid DNA segregation. *Science* 323, 509–512. doi: 10.1126/science.1164346
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi: 10.1038/nmeth.2089
- Schröder, G., Krause, S., Zechner, E. L., Traxler, B., Yeo, H. J., Lurz, R., et al. (2002). TraG-like proteins of DNA transfer systems and of the *Helicobacter pylori* type IV secretion system: inner membrane gate for exported substrates? *J. Bacteriol.* 184, 2767–2779. doi: 10.1128/JB.184.10.2767-2779.2002
- Schwab, M., Gruber, H., and Högenauer, G. (1991). The TraM protein of plasmid R1 is a DNA-binding protein. *Mol. Microbiol.* 5, 439–446. doi: 10.1111/j.1365-2958.1991.tb02127.x
- Silverman, P. M., Wickersham, E., and Harris, R. (1991). Regulation of the F plasmid *traY* promoter in *Escherichia coli* by host and plasmid factors. *J. Mol. Biol.* 218, 119–128. doi: 10.1016/0022-2836(91)90878-A
- Stern, J. C., and Schildbach, J. F. (2001). DNA recognition by F factor Tra136: highly sequence-specific binding of single-stranded DNA. *Biochemistry* 40, 11586–11595. doi: 10.1021/bi010877q
- Stothard, P. (2000). The sequence manipulation suite: javascript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28, 1102–1104.
- Strack, B., Lessl, M., Calendar, R., and Lanka, E. (1992). A common sequence motif, -E-G-Y-A-T-A-, identified within the primase domains of plasmid-encoded I- and P-type DNA primases and the alpha protein of the *Escherichia coli* satellite phage P4. *J. Biol. Chem.* 267, 13062–13072.
- Studier, F. W., and Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189, 113–130. doi: 10.1016/0022-2836(86)90385-2
- Sut, M. V., Mihajlovic, S., Lang, S., Gruber, C. J., and Zechner, E. L. (2009). Protein and DNA effectors control the TraI conjugative helicase of plasmid R1. *J. Bacteriol.* 191, 6888–6899. doi: 10.1128/JB.00920-09
- Szardenings, F., Guymer, D., and Gerdes, K. (2011). ParA ATPases can move and position DNA and subcellular structures. *Curr. Opin. Microbiol.* 14, 712–718. doi: 10.1016/j.mib.2011.09.008
- Tsai, M. M., Fu, Y. H., and Deonier, R. C. (1990). Intrinsic bends and integration host factor binding at F plasmid oriT. *J. Bacteriol.* 172, 4603–4609.
- Verdino, P., Keller, W., Strohmaier, H., Bischof, K., Lindner, H., and Koraimann, G. (1999). The essential transfer protein TraM binds to DNA as a tetramer. *J. Biol. Chem.* 274, 37421–37428. doi: 10.1074/jbc.274.52.37421
- Vergunst, A. C., Schrammeijer, B., den Dulk-Ras, A., Vlaam, C. M., de Regensburg-Tuink, T. J., and Hooykaas, P. J. (2000). VirB/D4-dependent protein translocation from *Agrobacterium* into plant cells. *Science* 290, 979–982. doi: 10.1126/science.290.5493.979
- Williams, S. L., and Schildbach, J. F. (2006). Examination of an inverted repeat within the F factor origin of transfer: context dependence of F TraI relaxase DNA specificity. *Nucleic Acids Res.* 34, 426–435. doi: 10.1093/nar/gkj444
- Wong, J. J., Lu, J., Edwards, R. A., Frost, L. S., and Glover, J. N. (2011). Structural basis of cooperative DNA recognition by the plasmid conjugation factor, TraM. *Nucleic Acids Res.* 39, 6775–6788. doi: 10.1093/nar/gkr296
- Woodcock, D. M., Crowther, P. J., Doherty, J., Jefferson, S., DeCruz, E., Noyer-Weidner, M., et al. (1989). Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic Acids Res.* 17, 3469–3478. doi: 10.1093/nar/17.9.3469
- Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., and Court, D. L. (2000). An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 5978–5983. doi: 10.1073/pnas.100127597
- Zechner, E. L., Lang, S., and Schildbach, J. F. (2012). Assembly and mechanisms of bacterial type IV secretion machines. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1073–1087. doi: 10.1098/rstb.2011.0207
- Zechner, E. L., Pruger, H., Grohmann, E., Espinosa, M., and Högenauer, G. (1997). Specific cleavage of chromosomal and plasmid DNA strands in gram-positive and gram-negative bacteria can be detected with nucleotide resolution. *Proc. Natl. Acad. Sci. U.S.A.* 94, 7435–7440. doi: 10.1073/pnas.94.14.7435

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Gruber, Lang, Rajendra, Nuk, Raffl, Schildbach and Zechner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# A Functional *oriT* in the Ptw Plasmid of *Burkholderia cenocepacia* Can Be Recognized by the R388 Relaxase TrwC

Esther Fernández-González<sup>1</sup>, Sawsane Bakioui<sup>2,3</sup>, Margarida C. Gomes<sup>2,3</sup>, David O'Callaghan<sup>2,3</sup>, Annette C. Vergunst<sup>2,3\*</sup>, Félix J. Sangari<sup>1</sup> and Matxalen Llosa<sup>1\*</sup>

<sup>1</sup> Departamento de Biología Molecular, Instituto de Biomedicina y Biotecnología de Cantabria, Universidad de Cantabria, UC-SODERCAN-Consejo Superior de Investigaciones Científicas, Santander, Spain, <sup>2</sup> Institut National de la Santé et de la Recherche Médicale, U1047, Nîmes, France, <sup>3</sup> UFR de Médecine Site de Nîmes, U1047, Université de Montpellier, France

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Consejo Superior de Investigaciones  
Científicas, Spain

### Reviewed by:

Fabián Lorenzo,  
Universidad de La Laguna, Spain  
Radosław Pluta,  
International Institute of Molecular and  
Cell Biology in Warsaw, Poland

### \*Correspondence:

Matxalen Llosa  
llosam@unican.es;  
Annette C. Vergunst  
annette.vergunst@umontpellier.fr

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 11 March 2016

**Accepted:** 14 April 2016

**Published:** 03 May 2016

### Citation:

Fernández-González E, Bakioui S,  
Gomes MC, O'Callaghan D,  
Vergunst AC, Sangari FJ and Llosa M  
(2016) A Functional *oriT* in the Ptw  
Plasmid of *Burkholderia cenocepacia*  
Can Be Recognized by the R388  
Relaxase TrwC.  
Front. Mol. Biosci. 3:16.  
doi: 10.3389/fmolb.2016.00016

*Burkholderia cenocepacia* is both a plant pathogen and the cause of serious opportunistic infections, particularly in cystic fibrosis patients. *B. cenocepacia* K56-2 harbors a native plasmid named Ptw for its involvement in the Plant Tissue Watersoaking phenotype. Ptw has also been reported to be important for survival in human cells. Interestingly, the presence of PtwC, a homolog of the conjugative relaxase TrwC of plasmid R388, suggests a possible function for Ptw in conjugative DNA transfer. The *ptw* region includes Type IV Secretion System genes related to those of the F plasmid. However, genes in the adjacent region shared stronger homology with the R388 genes involved in conjugative DNA metabolism. This region included the putative relaxase *ptwC*, a putative coupling protein and accessory nicking protein, and a DNA segment with high number of inverted repeats and elevated AT content, suggesting a possible *oriT*. Although we were unable to detect conjugative transfer of the Ptw resident plasmid, we detected conjugal mobilization of a co-resident plasmid containing the *ptw* region homologous to R388, demonstrating the cloned *ptw* region contains an *oriT*. A similar plasmid lacking *ptwC* could not be mobilized, suggesting that the putative relaxase PtwC must act *in cis* on its *oriT*. Remarkably, we also detected mobilization of a plasmid containing the Ptw *oriT* by the R388 relaxase TrwC, yet we could not detect PtwC-mediated mobilization of an R388 *oriT*-containing plasmid. Our data unambiguously show that the Ptw plasmid harbors DNA transfer functions, and suggests the Ptw plasmid may play a dual role in horizontal DNA transfer and eukaryotic infection.

**Keywords:** bacterial conjugation, type IV secretion, *Burkholderia cenocepacia*, plasmid R388, plasmid Ptw, conjugative relaxases, origin of transfer

## INTRODUCTION

The *Burkholderia* genus contains more than 60 species of clinical, environmental and agro-biotechnological value (Estrada-De Los Santos et al., 2013); most of them occupy a high diversity of ecological niches, in which they act as plant pathogens and catabolically active soil saprophytes. Only two of them, *B. mallei* and *B. pseudomallei* are considered primary pathogens of

**Abbreviations:** Ptw, plant tissue watersoaking; Dtr, DNA transfer and replication; *oriT*, origin of transfer; T4SS, Type IV Secretion System.

animals and humans. The *B. cepacia* complex (Bcc) is a collection of currently 18 related species, sharing 97.5% rDNA sequence similarity, although only 30–60% genome-wide variability was shown by DNA-DNA hybridization (Vandamme and Dawyndt, 2011). It was initially described as one of the agents that promotes soft rot disease in onions. Importantly, Bcc bacteria pose serious health problems in vulnerable patients, particularly in individuals with cystic fibrosis (CF) or chronic granulomatous disease (CGD), and they are emerging pathogens in nosocomial infections (Coenye and Vandamme, 2003). Members of the Bcc can cause a life-threatening respiratory infection in patients with CF (Gonzalez et al., 1997; Berriatua et al., 2001; LiPuma, 2003). *B. cenocepacia* and *B. multivorans* are the most prevalent species (85–90%) of the Bcc isolated from patients with CF (LiPuma, 1998a,b; Speert, 2002). *B. cenocepacia* is correlated with increased morbidity and mortality, and it has caused several major epidemics (Goven et al., 1993).

The genome of *B. cenocepacia* strain J2315, isolated from an infected CF patient, has been sequenced (Holden et al., 2009), showing the presence of three circular chromosomes and a plasmid of 92 kb. Most of the coding DNA sequences (CDSs) present in chromosome I are involved in housekeeping functions, while chromosomes II and III [the latter recently renamed as a mega virulence plasmid (Agnoli et al., 2012)] contain CDSs with accessory functions like protective responses and horizontal gene transfer, among others (Holden et al., 2009). Many studies conducted on *B. cenocepacia* utilize another CF clinical isolate, strain K56-2 (Darling et al., 1998), which is clonally related to J2315 (Mahenthiralingam et al., 2000) but easier to grow and less resistant to antibiotics, making it more amenable to genetic manipulation. K56-2 has a shorter chromosome I due to the absence of a large duplication in J2315, and its 92 kb plasmid has one less copy of an insertion element (Varga et al., 2013).

Bacterial pathogens use secretion mechanisms for the delivery of virulence determinants (Christie and Vogel, 2000), ranging from one component systems to complex multi-component machineries. Type IV Secretion Systems (T4SS) constitute a family of molecular transporters able to deliver DNA, proteins, or nucleoprotein complexes to the extracellular milieu or other cells (prokaryotic or eukaryotic). Many T4SS mediate secretion of proteins into eukaryotic cells, being implicated in infection processes, and adaptability inside the host (Alvarez-Martinez and Christie, 2009). T4SS involved in conjugative DNA transfer (Zechner et al., 2000) are associated with a DNA transfer region (Dtr), composed by an origin of transfer (*oriT*), a coupling protein, and a conjugative relaxase; accessory nicking proteins and regulators are also often present. The relaxase catalyzes the initial and final stages of conjugative DNA transfer, cleaving the *oriT* in the donor to produce the DNA strand to be transferred, and resealing the transferred DNA in the recipient. Relaxases are defined by a series of conserved motifs, and have been classified in different families (Garcillán-Barcia et al., 2009). The *oriT* is the only DNA requirement *in cis* for mobilization of a DNA molecule. *oriTs* are usually located in intergenic regions showing a higher AT content than the rest of the molecule, to allow strand separation during initiation of single-strand transfer. *oriT* regions usually contain abundance of direct and inverted repeats

(DRs and IRs), many of which have been shown to be the relaxase or accessory nicking protein binding sites (Moncalián et al., 1997; Becker and Meyer, 2000; Williams and Schildbach, 2006, 2007; Lucas et al., 2010; Wong et al., 2011). The Dtr region of plasmid R388 includes a 330 bp *oriT* and three genes transcribed from a single operon, *trwABC* (Llosa et al., 1994). The accessory nicking protein TrwA belongs to the Ribbon-helix-helix (RHH) family proteins (Moncalián and De La Cruz, 2004). TrwB is the coupling protein, interacting both with the transferred substrate and with the Type IV secretion machine (Llosa et al., 2003). The conjugative relaxase TrwC, as other members of the MOB<sub>F</sub> family of relaxases, is composed by two functional domains, the N-terminus harboring the relaxase activity, and the C-terminus showing DNA helicase activity which is also required for conjugation (Llosa et al., 1996).

Two T4SS have been described in *B. cenocepacia*, the Ptw (Plant tissue water-soaking) T4SS, and the VirB/D4 T4SS (Engledow et al., 2004). The *B. cenocepacia* VirB/D4 T4SS is located on chromosome II and bears high homology with the prototypical VirB/D4 T4SS of *A. tumefaciens*. Although its function is still unknown, a possible role in DNA transfer has been reported (Zhang et al., 2009). The Ptw T4SS is encoded on a 45 kb region (Holden et al., 2009) of a native plasmid of 92 kb. Based on amino acid sequence similarity, it was proposed that the Ptw T4SS is a chimera of various translocation and/or conjugation related proteins similar to VirB/D4 and F-specific subunits (Engledow et al., 2004). One of them, named as PtwC, presents 33% amino acid identity with TrwC, the relaxase of plasmid R388 (Engledow et al., 2004). The presence of a relaxase homolog, a protein specifically associated with conjugative DNA transfer, argued for a conjugative role of the Ptw T4SS. However, to date Ptw functions have been associated with PtwC-independent secretion of a plant cytotoxic factor (Engledow et al., 2004), and intracellular survival in both professional and non-professional phagocytes (Sajjan et al., 2008).

In this study we address the characterization of the conjugative functions encoded by the 92 kb plasmid of *B. cenocepacia* K56-2 (named pK56-2 by Engledow et al., 2004), hereon referred to as the Ptw plasmid, and the possible relationship with the R388 transfer machinery. We have mapped a functional *oriT* in Ptw, suggesting that Ptw is a conjugative plasmid; therefore, Ptw may play a dual role in *B. cenocepacia*, promoting horizontal DNA transfer among bacteria and assisting infection of eukaryotic hosts.

## MATERIALS AND METHODS

### Bioinformatic Analysis

The *B. cenocepacia* strain J2315 plasmid pBCJ2315 sequence is available in GenBank under the accession number NC\_011003. The RAST (Rapid Annotation Subsystems Technology) program was used to re-annotate the ORFs present in the plasmid sequence (Overbeek et al., 2014). PHYRE (Protein Homology/analogy Recognition Engine) was used for the prediction of secondary and 3-dimensional structure of a protein amino acid sequence (Kelley et al., 2015). The phylogenetic tree was constructed using MEGA (Molecular Evolutionary Genetics



Analysis) software, which employs the Maximum Likelihood method based on the JTT matrix-based model (Koichiro et al., 2011). GenScript bioinformatics software was selected to analyze AT content (<http://www.genscript.com>). The program Scan for Matches was used to localize IRs having at least 5 bp in length and less than two mismatches ([http://www.theseed.org/servers/downloads/scan\\_for\\_matches.tgz](http://www.theseed.org/servers/downloads/scan_for_matches.tgz)). Other intergenic regions of the same size selected randomly from the Ptw plasmid were also analyzed for comparison.

Bacterial Strains and Plasmids

Bacterial strains used in this study are listed in Table S1. All bacterial strains were grown in Lysogeny Broth (LB), supplemented with 1.5% agar for culture on solid medium, at 37°C. Selective media included the following antibiotics at the indicated concentrations: ampicillin (Ap), 100 µg/ml; chloramphenicol (Cm), 25 µg/ml (*E. coli*) or 100 µg/ml (*Burkholderia*); kanamycin (Km), 50 µg/ml; nalidixic acid (Nx), 20 µg/ml; streptomycin (Sm), 300 µg/ml; gentamicin (Gm), 10 µg/ml; and trimethoprim (Tp), 20 µg/ml (*E. coli*), or 250 µg/ml (*Burkholderia*). Plasmids are listed in Table 1. Plasmids were maintained in the *E. coli lacI<sup>q</sup>* strain D1210 (Sadler et al., 1980).

Plasmid Constructions

Plasmids were constructed using standard methodological techniques (Sambrook and Russell, 2001). *E. coli* DH5α (Grant et al., 1990) was used in all cloning procedures. Table S2 shows details of the constructions for each plasmid. All cloned DNA inserts were obtained by PCR on genomic DNA isolated from *B. cenocepacia* K56-2. Genomic DNA was extracted using Instagene Matrix (BioRad). Restriction enzymes, shrimp alkaline phosphatase, and T4 DNA ligase were purchased from Fermentas. High-fidelity Kapa Taq polymerase was purchased from KapaBiosystems. DNA sequences of all cloned PCR segments were determined (MACROGEN Inc. DNA Sequencing Service; Amsterdam, The Netherlands). Plasmid pEF031 was cloned in two sequential steps. In a first step, the 949 bp region where the possible *oriT* is located was cloned together with the hypothetical accessory nicking protein gene *ptwA* and 1269 bp of the coupling protein gene *ptwB*, adding a HindIII restriction site

in *ptwB* which did not alter the predicted ORE, obtaining plasmid pEF022. In a second step, the rest of *ptwB* and *ptwC* were added to pEF022, re-creating the *ptwABC* operon.

Colony Analysis by PCR

All oligonucleotides pairs used for plasmid confirmation by PCR are summarized in Table S3. Colonies obtained from the three plasmid curing methods were checked for the presence of the Ptw plasmid by PCR, using two oligonucleotide pairs (P1-P2) and (P3-P4) amplifying *ptwC* and *pwaC* internal fragments. All the colonies tested, at least 50 for each curing method, were positive for the presence of both genes. Four oligonucleotide pairs (P5-P6; P7-P8; P9-P10; and P11-P12) were selected to detect the presence of different regions of the 45 kb *ptw* cluster in the two *B. cepacia* strains (Table S1) by PCR. K56-2 total DNA was always used as a positive control. Transconjugants obtained in the mobilization assay were checked by PCR using oligonucleotide pairs (P13-P14) to amplify the R388 *oriT* sequence of plasmid pSU1445, and (P15-P6) pair to confirm the presence of the *cat* gene in the pBBR1 plasmid.

Plasposon Insertion

For preparation of electrocompetent cells, bacteria were grown to OD<sub>600</sub> = 0.5–0.6, and pelleted by centrifugation at 4°C. Two series of washes and centrifugations (6000 rpm) of 1 vol milliQ water and a final wash in 1/50 vol 10% glycerol at 4°C were applied. Cells were resuspended in 1/500 vol 10% glycerol and aliquotted in 50 µl samples. K56-2 electrocompetent cells were transformed with <10 ng of the pTnMob-OCm plasposon DNA (Dennis and Zylstra, 1998) in a 0.2 cm Gene Pulser® cuvette (BioRad) and subjected to an electric pulse (2.5 kV, 25 Mf, and 200 Ω) in a MicroPulser TM (BioRad). Electroporated cells were added to 1 ml LB and incubated with shaking at 37°C to allow expression of antibiotic resistance genes. After incubation cells were plated on antibiotic containing media. The pool of bacterial colonies growing on Cm plates was used as donor cells to test Ptw conjugation. Considering the size of the plasmid and the total length of the genomic DNA, it is estimated that on average one out of 100 colonies would have the plasposon inserted in the Ptw plasmid, and from those, about half would not affect

TABLE 1 | Plasmids used in this work.

Plasmid	Antibiotic Resistance	Description	References
pBBR1-MCS	Cm <sup>R</sup>	Broad host range vector from <i>Bordetella bronchiseptica</i>	Kovach et al., 1994
pEF022	Cm <sup>R</sup>	pBBR1::Ptw-oriT+ptwA+2/3ptwB	This work
pEF031	Cm <sup>R</sup>	pBBR1::Ptw-oriT+ptwA+ptwB+ptwC	This work
pEF033	Cm <sup>R</sup>	pBBR1::Ptw-oriT (700 bp)	This work
pEF034	Cm <sup>R</sup>	pBBR1::Ptw-oriT (700 bp)+ptwA	This work
pKM101	Ap <sup>R</sup>	R46 deletion derivative	Langer and Walker, 1981
pOX38	Km <sup>R</sup>	F plasmid derivative	Chandler and Galas, 1983
pSU1443	Km <sup>R</sup> Tp <sup>R</sup>	R388::Tn5tac in <i>tnwB</i>	Llosa et al., 1994
pSU1445	Km <sup>R</sup> Tp <sup>R</sup>	R388::Tn5tac in <i>tnwC</i>	Llosa et al., 1994
pSU2007	Km <sup>R</sup>	R388 with a Km <sup>R</sup> cassette	Martinez and de la Cruz, 1988
pTnMob-OCm	Cm <sup>R</sup>	Self-cloning minitransposon	Dennis and Zylstra, 1998

any transfer-related functions. Therefore, in a mating experiment involving  $10^6$  donor cells, we would expect to have at least  $5 \times 10^3$  cells carrying a Cm-resistant, transfer-proficient Ptw plasmid.

## Mating Experiments

Standard *E. coli* quantitative mating assays were performed as described previously (Grandoso et al., 2000): equal amounts of donor and recipient strains from overnight cultures were mixed and placed on Millipore filters on a prewarmed LB agar plate for 1 h at 37°C. Strains D1210 and DH5 $\alpha$  were used as donors and recipients, as indicated. Results are shown as the frequency of transconjugants per donor and are the mean of 3–5 independent experiments. For mating assays using *B. cenocepacia* as a donor and *E. coli*  $\beta$ 2163 (Demarre et al., 2005) or *B. cepacia* strains as recipients, bacterial cultures were adjusted to an OD<sub>600</sub> = 0.5, mixed at an equal ratio, washed twice and transferred to a Millipore filter on a prewarmed LB agar plate for 18 h at 37°C.

## Plant Watersoaking Assay

The plant watersoaking assay was used to assess the functionality of the Ptw T4SS in *B. cenocepacia*, as described (Engledow et al., 2004). Several onion types, like echalote, red onion and white onion were used throughout the study, and no significant differences were found among them, so all experiments were performed with white onions. Bacterial suspensions of *B. cenocepacia* strains were adjusted to OD<sub>600</sub> = 0.5 and individual onion scales were wounded on the abaxial (inner) surface with a sterile blade. 10  $\mu$ l of bacterial suspension ( $10^6$  c.f.u per scale) were inoculated into the wound. Sterile double-distilled water was used as a negative control. Onion scales were placed on a sheet of sterile aluminum foil in containers containing Whatman paper towels moistened with sterile distilled water, sealed, and incubated at 37°C. Ptw activity was estimated by the appearance of water drops on the onion tissue at 24 hpi.

## Plasmid Curing

In order to cure the Ptw plasmid from *B. cenocepacia*, bacteria were incubated under different stress conditions, where mutations in essential plasmid genes can be induced, promoting the loss of the plasmid. Two methods using high temperatures and a third one growing bacteria in the presence of the mutagenic agent ethidium bromide were used: (i) *Growth on TN Medium* (Gonzalez et al., 1997). Bacteria containing the plasmid to be cured were grown in TN broth [5 g Tryptone, 1 g Dextrose (D-Glucose), 2.5 g Yeast extract, and 8.5 g NaCl per liter] for 18 h at 37°C with shaking (200 r.p.m). Bacteria were subcultured into pre-warmed TN broth (42–44°C) to a final concentration of  $10^4$  c.f.u ml<sup>-1</sup> and grown with shaking in a water bath (at 42–44°C) for 18 h. Temperature-treated cultures were diluted and plate on TN agar. Individual colonies were checked by PCR. (ii) *Growth at High Temperature* (Asheshov, 1966). Bacteria were grown overnight at 37°C. Overnight cultures were diluted 1/4 in fresh LB medium for 1 1/2 h. Samples of this culture were added to tubes containing 5 ml of fresh LB medium previously pre-warmed at 37°C and at 43–44°C respectively. Tubes were then incubated at the appropriate temperature for

5 1/2 h. Cultures were then diluted and spread on LB agar plates and incubated at 37°C overnight; the resulting colonies were then tested for the loss of the native plasmid by PCR. (iii) *Growth with Ethidium Bromide* (Cramer et al., 1986). Ethidium bromide (25  $\mu$ g/ml) was added to a *B. cenocepacia* liquid culture and incubated with shaking at 37°C overnight protected from the light. A 1/10 dilution of the overnight culture was prepared with fresh LB medium supplemented with Ethidium bromide. This procedure was repeated over 15 days; every 2 days, dilutions of the culture were plated, and colonies were checked for the presence of the plasmid by PCR.

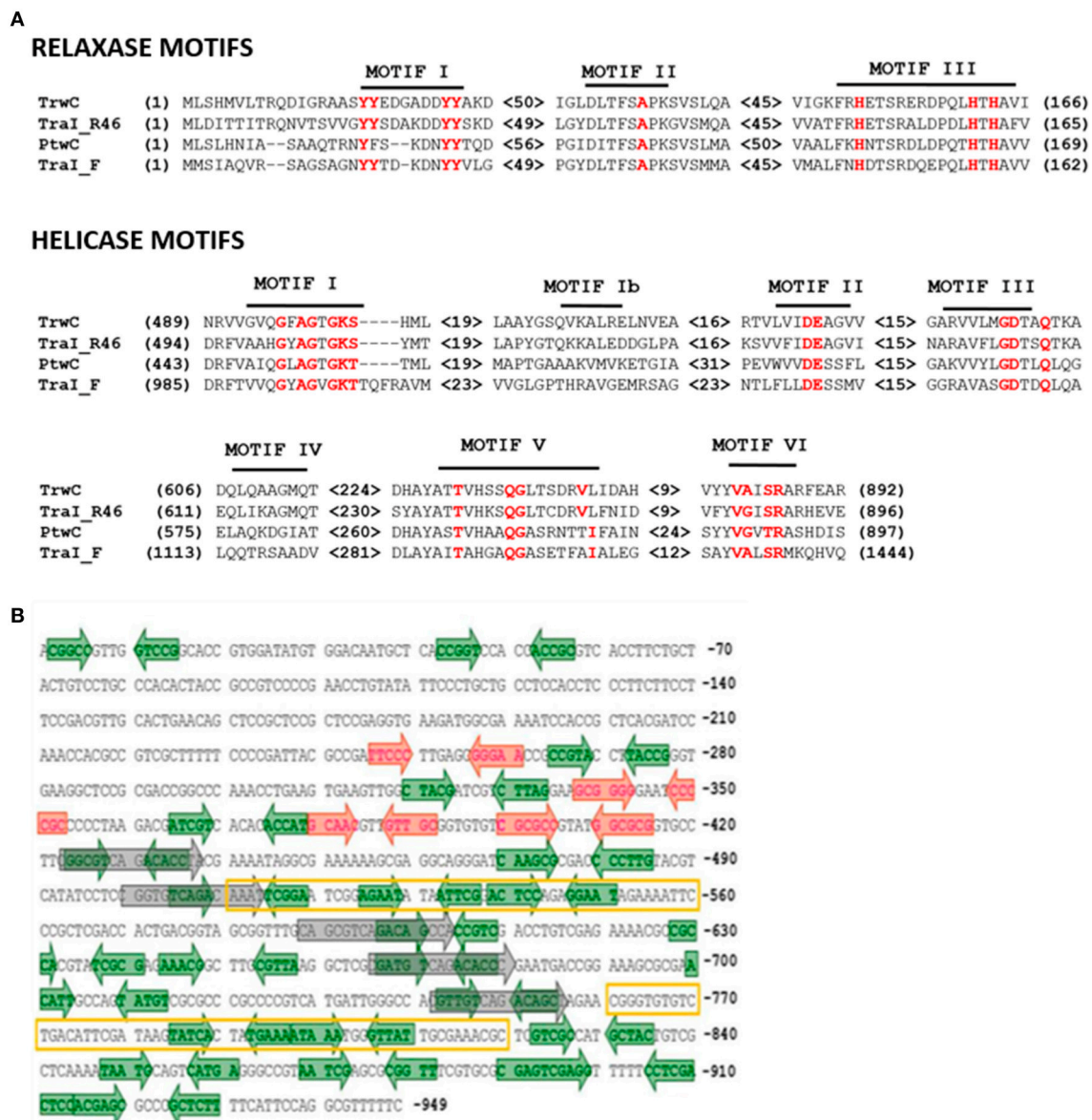
## RESULTS

### Bioinformatic Analysis of the Ptw Plasmid

A preliminary annotation of the Ptw plasmid sequence for *B. cenocepacia* J2315 was available (Holden et al., 2009). We have used the RAST Bioinformatic Program to obtain a detailed annotation of the ORFs present in the plasmid, based on a comparative study against the NCBI database. As previously reported (Engledow et al., 2004; Holden et al., 2009), a 45 kb region was found to encode genes with similarity to the F plasmid T4SS genes (Lawley et al., 2003). However, the adjacent DNA region to T4SS genes, which included the putative relaxase and coupling protein genes previously described, *ptwC* and *ptwD4*, showed higher similarity to the Dtr region of conjugative plasmid R388: PtwC shows 33% of amino acid identity with R388 relaxase TrwC, and 27% with relaxase TraI of the F plasmid. PtwD4 presents 20% identity with VirD4 from *A. tumefaciens*, 28% with TraD of the F plasmid, and 30% with R388 TrwB. Due to the stronger similarity to TrwB, PtwD4 will be named PtwB from here on.

Both R388-TrwC and F-TraI belong to the MOB<sub>F</sub> family of relaxases, defined by three common motifs based on the known atomic structure of the TrwC relaxase and similar proteins (Francia et al., 2004). CLUSTALW alignment of PtwC and its homologs showed that the three relaxase motifs were conserved in PtwC (Figure 1A). The homology between TrwC and PtwC in this amino terminal domain goes up to 42%, while the C-terminal domains share 29% amino acid identity. In addition, PtwC shares the seven conserved amino acid motifs in the C-terminal domain which define the DNA helicase superfamily I (Matson, 1990; Figure 1A). The MEGA5 program (Koichiro et al., 2011) was used to determine its phylogenetic position in MOB<sub>F</sub> family (Figure S1), confirming that PtwC belongs to the MOB<sub>F</sub> family, and R388-TrwC is its closest relative out of the *Burkholderia* spp homologs.

In most conjugative systems, the *oriT* is located close to the relaxase; accessory nicking proteins are also often encoded in this DNA region. Analysis of the DNA region upstream of *ptwB* led to the identification of a small ORF of 429 bp (pBCA060) coding for a hypothetical protein with a predicted RHH structure related to other members of the RHH superfamily (Heidelberg et al., 2000), and thus a likely accessory nicking protein of PtwC; we named it PtwA after TrwA, the accessory nicking protein of TrwC, although both proteins do not share significant amino acid



**FIGURE 1 | Conserved motifs in PtwC and related proteins and features of the hypothetical *oriT* region. (A)** Conserved motifs in the relaxase and DNA helicase domains, with the most conserved residues highlighted in red. **(B)** Features in hypothetical Ptw *oriT* region (coordinates 66,438–65,490 in Accession number NC\_011003.1). IRs with one mismatch are represented in green. A set of DRs of 15 bp with three mismatches are represented in gray. High AT regions are marked with a yellow box.

identity. The most likely region to contain the Ptw *oriT* sequence would be in a 949 bp intergenic region located upstream of *ptwA*. It contains DNA segments with high AT content (squared regions in Figure 1B), candidates to contain the *nic* site, and two times more IRs than in a randomly selected intergenic region of the same size (not shown), especially in the region including the two AT-rich DNA segments (Figure 1B). All these features lead us to propose this as a candidate region to contain the Ptw *oriT*.

Compiling the results of the bioinformatics analysis, we propose that the Ptw plasmid codes for a conjugative transfer region with a T4SS related to that of the F plasmid, while the

Dtr region is closely related to that of the R388 transfer system (Figure 2).

## Involvement of the Ptw Plasmid in Conjugative DNA Transfer

Our next goal was to experimentally test conjugative transfer of the Ptw plasmid. Entry exclusion is a common phenomenon in conjugative DNA transfer, which prevents conjugation into a recipient cell already containing the conjugative plasmid present in the donor (Garcillán-Barcia and de la Cruz, 2008). To prevent this possible effect, a Ptw-free recipient strain was required. We





the Ptw plasmid which provides the complete putative transfer machinery. However, a pBBR-plasmid containing the complete Ptw Dtr region was mobilized to both *B. cepacia* strains; no mobilization to *B. cenocepacia* K56-2 or to *E. coli* was detected. Transconjugants were verified by PCR amplification to confirm the presence of both plasmid pSU1445 and the mobilized plasmid. This result proves the presence of a functional *oriT* in the Ptw plasmid.

## Recognition of the Ptw *oriT* by the Conjugative Relaxase TrwC

It has previously been reported that conjugative systems can use heterologous T4SS with different affinities: the TrwC-DNA complex could be secreted through the T4SS of different conjugative plasmids, and moreover, it could also use the T4SS of *Bartonella henselae* to be transferred to human cells (Llosa et al., 2003; Fernández-González et al., 2011). Mobilization of a plasmid coding for the Ptw Dtr components by other conjugative T4SS was therefore tested. Plasmids containing different parts of the Ptw Dtr were tested for mobilization between *E. coli* strains. Three different conjugative plasmids were used as helper plasmids in the donor strain: pSU2007 (R388 derivative) (Martinez and de la Cruz, 1988), pOX38-Km (F derivative) (Chandler and Galas, 1983), and pKM101 (R46 derivative; Langer and Walker, 1981; Table 1), all coding for relaxases sharing homology with PtwC (Figure 1). Results are

shown in Table 3. While no transconjugants were observed with the F or R46 transfer systems, a low level of mobilization of the Ptw-*oriT* containing plasmids was observed when the R388 transfer system was present in the donor cell. Transconjugants were analyzed by restriction analysis (data not shown), showing both the mobilizable and helper plasmids as separate entities in the transconjugants (clearly distinguishable by their different copy number), thus ruling out plasmid conduction by cointegrate formation with the conjugative plasmid. Restriction analysis with HindIII and SacI rendered the expected fragments for the helper pSU2007 (32.5 kb) and mobilizable plasmids pEF022 or pEF031: 2640 bp fragment in all cases (*ptw oriT* + *ptwA* + 2/3 *ptwB*) plus a fragment of 4700 or 8800 bp, respectively.

Surprisingly, similar results were obtained when the mobilizable plasmid carried only the Ptw *oriT*, *ptwA*, and 2/3 of *ptwB*, instead of the complete Ptw Dtr region (Table 3, compare mobilization by pSU2007 of pEF022 and pEF031). Thus, mobilization through the T4SS of R388 is not due to PtwC, but presumably to recognition of the Ptw *oriT* by TrwC *in trans*. To confirm if mobilization is TrwC dependent, a TrwC-deficient R388 plasmid (pSU1445) was used as helper; in this case, no plasmid mobilization was detected (Table 3). These assays also show that there is no functional complementation of TrwB or TrwC by PtwB and PtwC, at least *in trans*, since R388 *trwB*- (pSU1443) and R388 *trwC*- (pSU1445) deficient plasmids cannot be mobilized in the presence of a plasmid harboring the Ptw

TABLE 3 | Mobilization of plasmids carrying the Ptw-*oriT* by different conjugative plasmids between *E. coli* strains.

Plasmids in donor				Transfer frequency	
Helper ( <i>TRA Rel CP</i> ) <sup>a</sup>		Mobilizable ( <i>Ptw Dtr</i> )		Helper	Mobilizable
pKM101	IncN Tral TraD			0.21 ± 0.11	
pKM101	IncN Tral TraD	pBBR1		0.19 ± 0.12	<3.9 × 10 <sup>-8</sup>
pKM101	IncN Tral TraD	pEF022	<i>Ptw-oriT</i> + <i>ptwA</i> <sup>b</sup>	0.19 ± 0.11	<2.3 × 10 <sup>-8</sup>
pKM101	IncN Tral TraD	pEF031	<i>Ptw-oriT</i> + <i>ptwABC</i>	0.25 ± 0.08	<2.5 × 10 <sup>-8</sup>
pOX38	IncF Tral TraD			0.51 ± 0.03	
pOX38	IncF Tral TraD	pBBR1		0.34 ± 0.07	<5.6 × 10 <sup>-8</sup>
pOX38	IncF Tral TraD	pEF022	<i>Ptw-oriT</i> + <i>ptwA</i> <sup>b</sup>	0.36 ± 0.37	<2.2 × 10 <sup>-8</sup>
pOX38	IncF Tral TraD	pEF031	<i>Ptw-oriT</i> + <i>ptwABC</i>	0.19 ± 0.02	<1.9 × 10 <sup>-8</sup>
pSU2007	IncW TrwC TrwB			0.11 ± 0.08	
pSU2007	IncW TrwC TrwB	pBBR1		0.20 ± 0.16	<1.3 × 10 <sup>-8</sup>
pSU2007	IncW TrwC TrwB	pEF022	<i>Ptw-oriT</i> + <i>ptwA</i> <sup>b</sup>	0.12 ± 0.11	<b>1.5 × 10<sup>-7</sup> ± 1.1 × 10<sup>-7</sup></b>
pSU2007	IncW TrwC TrwB	pEF031	<i>Ptw-oriT</i> + <i>ptwABC</i>	0.19 ± 0.11	<b>8.2 × 10<sup>-8</sup> ± 5.0 × 10<sup>-9</sup></b>
pSU2007	IncW TrwC TrwB	pEF033	<i>Ptw-oriT</i> (700)	0.21 ± 0.06	<b>7.9 × 10<sup>-8</sup> ± 1.3 × 10<sup>-8</sup></b>
pSU2007	IncW TrwC TrwB	pEF034	<i>Ptw-oriT</i> (700) + <i>ptwA</i>	0.16 ± 0.15	<b>1.3 × 10<sup>-7</sup> ± 1.0 × 10<sup>-7</sup></b>
pSU1443	IncW TrwC			<1.5 × 10 <sup>-7</sup>	
pSU1443	IncW TrwC	pBBR1		<2.2 × 10 <sup>-7</sup>	<3.1 × 10 <sup>-8</sup>
pSU1443	IncW TrwC	pEF031	<i>Ptw-oriT</i> + <i>ptwABC</i>	<2.9 × 10 <sup>-7</sup>	<4.5 × 10 <sup>-8</sup>
pSU1445	IncW TrwB			<2.8 × 10 <sup>-7</sup>	
pSU1445	IncW TrwB	pBBR1		<2.5 × 10 <sup>-7</sup>	<2.8 × 10 <sup>-8</sup>
pSU1445	IncW TrwB	pEF022	<i>Ptw-oriT</i> + <i>ptwA</i> <sup>b</sup>	<2.9 × 10 <sup>-7</sup>	<3.6 × 10 <sup>-8</sup>
pSU1445	IncW TrwB	pEF031	<i>Ptw-oriT</i> + <i>ptwABC</i>	<3.1 × 10 <sup>-7</sup>	<1.9 × 10 <sup>-8</sup>

DH5α was used as a donor and D1210 as a recipient. Transfer frequencies (transconjugants/donor) represent the mean ± SD of three independent experiments.

<sup>a</sup>Description of the Helper plasmid: TRA transfer system (defined by plasmid Inc group), Rel relaxase, CP coupling protein.

<sup>b</sup>A truncated *ptwB* copy was also present, which was omitted for clarity. Positive mobilization frequencies are highlighted in bold.

Dtr region (Table 3, negative transfer frequency of pSU1443 and pSU1445 in the presence of pEF031).

Mobilization of Ptw-*oriT* containing plasmids by R388 allowed us to refine mapping of the Ptw *oriT*. Plasmids were constructed carrying only the region where the *oriT*-like features were found, with and without *ptwA*. A plasmid carrying region 65.490–66.190 of the Ptw-*oriT* (nucleotides 249–949 in Figure 1B) was mobilized by TrwC as efficiently as the plasmid containing the complete Ptw Dtr region (Table 3, compare mobilization frequency of pEF031 and pEF033 by pSU2007).

## DISCUSSION

Previous studies proposed that the T4SS encoded on the Ptw plasmid of *B. cenocepacia* played a role in interactions with the eukaryotic host. The T4SS was associated with the plant tissue watersoaking phenotype, produced by unknown translocated effector proteins (Engledow et al., 2004). In addition, the Ptw T4SS was shown to have a role in intracellular survival and replication of *B. cenocepacia* inside macrophages (Sajjan et al., 2008), although Valvano and colleagues were unable to show defects in intracellular survival in *B. cenocepacia* mutants lacking all possible secretion systems (Valvano, 2015). However, the presence of the *ptwC* gene coding for a putative conjugative relaxase, not involved in the Ptw phenotype (Engledow et al., 2004), led us to hypothesize that Ptw is a conjugative plasmid. Relaxases are the signature proteins of conjugative systems; all known close homologs are involved in horizontal DNA transfer (Garcillán-Barcia et al., 2009). PtwC belongs to the MOB<sub>F</sub> family (Garcillán-Barcia et al., 2009) of relaxases-helicases, according to phylogenetic analysis (Figure S1). Conservation in PtwC of the signature motifs associated to relaxase and DNA helicase activity (Figure 1) reinforce the idea that PtwC is an active conjugative relaxase.

A straightforward way to determine the role of the Ptw plasmid would be to cure it from *B. cepacia* strains; however our repeated attempts were unsuccessful. Low-copy-number plasmids have developed a number of mechanisms to promote their stable maintenance, such as Toxin-Antitoxin (TA) systems (Yamaguchi et al., 2011). The Ptw plasmid indeed encodes genes showing similarity to the toxin and antitoxin of the VapBC TA system of *Mycobacterium tuberculosis* (Ramage et al., 2009), which could explain the failure to cure the Ptw plasmid. Similarly, recent studies have shown that the pC3 mega-plasmid (previously chromosome 3) in *B. cenocepacia* K56-2 showed unexpected stability due to the presence of a toxin-antitoxin system, which made it difficult to cure this non-essential plasmid from K56-2, in contrast to other Bcc strains (Agnoli et al., 2012).

The Ptw T4SS was reported to be a chimera between VirB and F-like specific subunits (Engledow et al., 2004). Our bioinformatic analyses show that the plasmid possesses all the components of a conjugative system, composed of a T4SS gene set related to that of the F plasmid, and a Dtr region whose closest homolog is that of plasmid R388 (Figure 2). In the Dtr region, upstream the putative coupling protein and relaxase previously reported (Engledow et al., 2004), we described a protein, PtwA, with similar size to accessory nicking proteins, predicted to have the characteristic Ribbon-Helix-Helix secondary structure. More

importantly, we delimited a region upstream of this putative accessory nicking protein candidate to be the Ptw *oriT*, due to its high AT content and the elevated number of IRs and DRs (Figure 1).

Despite the presence of a DNA transfer region, Ptw plasmid conjugation could not be detected under different test conditions. However, we detected mobilization of a Ptw-Dtr containing plasmid introduced in the host of the Ptw plasmid. It is reasonable to assume that mobilization will occur through the Ptw T4SS, although we cannot discard that the plasmid may be using the VirB T4SS, previously shown to be involved in DNA transfer (Zhang et al., 2009). Whatever the T4SS used for secretion, the absence of mobilization of the Ptw plasmid, which includes the Dtr region, suggests that self-transfer is repressed *in cis*. This is also the case for one of the symbiotic plasmids (pSym) present in *Rhizobium etli*, which possesses a full set of genes involved in conjugation. Plasmid transfer has never been detected under laboratory conditions, although there are several lines of evidence for the movement of pSym plasmids among naturally occurring rhizobial populations (Wernegreen and Riley, 1999). The absence of conjugation in this symbiotic plasmid has been explained by the presence of a conjugative repressor, *rctA*, which inhibits the conjugation of the symbiotic plasmid by decreasing *virB* transcription (Sepúlveda et al., 2008). No similar genes to *rctA* have been described until now in the Ptw plasmid, however the presence of such a repressor cannot be discarded.

Mobilization of the plasmid containing the Ptw Dtr occurred to two different *B. cepacia* strains, but not to *B. cenocepacia* or to *E. coli* (Table 2). The absence of transfer to *B. cenocepacia* could reflect an entry-exclusion phenomenon, frequent in bacterial conjugative plasmids (Garcillán-Barcia and de la Cruz, 2008), due to the presence of the Ptw plasmid in the recipient. The absence of transfer to *E. coli* could reflect a narrow host range of this conjugative system, similar to that of the F plasmid to which the T4SS is most closely related (Zhong et al., 2005).

Plasmids containing the putative *oriT* but no *ptwC* were not mobilized (Table 2), supporting the sequence homology and phylogenetic analysis which indicate that PtwC is a conjugative relaxase. Since, *ptwC* is present in the co-resident Ptw plasmid and no Ptw transfer is detected, PtwC may act only *in cis*, as previously reported for other conjugative relaxases (Pérez-Mendoza et al., 2006; Cho and Winans, 2007). Alternatively, this could be due to *cis*-repression of the *ptwC* copy present in the Ptw plasmid, as discussed above. The *ptwC* copy present in the multicopy plasmid would escape repression *in trans*, or if the presence of multiple *ptwC* copies outnumbers the repressor, leading to partial expression (in accordance with the low transfer frequencies obtained).

Interestingly, we showed that the Ptw *oriT* can be also recognized by the conjugative relaxase of plasmid R388, TrwC (Table 3), while the reverse is not true: PtwC does not act on R388 *oriT*, at least *in trans*. The fact that TrwC can mobilize Ptw-*oriT* containing plasmids *in trans* while PtwC cannot, can be explained by the different nature of both relaxases. *cis*- or *trans*-acting is an attribute of the relaxase, not the *oriT*. PtwC may act *in cis* as has been reported for a few other relaxases, but most of them act *in trans*, including TrwC (Llosa et al., 1996). If alternatively the lack of trans-mobilization by PtwC were due to *cis*-repression of the

*ptwC* gene, as suggested above, this repression would obviously not affect the *trwC* copy present in a different plasmid.

Traditionally considered as highly sequence specific, conjugative relaxases have been shown to act on related sequences with relaxed specificity (Jandle and Meyer, 2006; Fernández-López et al., 2013; O'Brien et al., 2015; Pollet et al., 2016). However, it is surprising that TrwC can act on another *oriT* with no apparent homology to its natural target; we have not found any sequence within the 700 bp defined as the Ptw *oriT* which resembles the R388 *nic* site and the surrounding 17 bp which are required for TrwC binding and nicking (Lucas et al., 2010). TrwC has been reported to recognize sequences which differ from its natural target (Agúndez et al., 2012), and it is also efficiently recruited by heterologous T4SS (Llosa et al., 2003; Fernández-González et al., 2011), so this result adds to previous ones to highlight the promiscuity of this enzyme and its conjugative transfer system. In addition, this result opens up the possibility that the Ptw plasmid can be transferred in nature by the conjugative system of a horizontally transmitted promiscuous plasmid such as is R388.

Earlier reports have suggested a role for the Ptw T4SS in the Plant Tissue Water soaking phenotype and survival in mammalian host cells through the secretion of putative effector proteins (Sajjan et al., 2008). Our results unambiguously show that the Ptw plasmid encodes an active *oriT* adjacent to the *ptw* T4SS gene cluster. It would be interesting to find out whether conjugative transfer during infection of mammalian hosts would contribute to disease. A role for the Ptw plasmid in transfer of DNA and effector proteins may give an advantage to *B. cenocepacia* in the environment, while at the same time contributing to their opportunistic character.

## REFERENCES

- Agnoli, K., Schwager, S., Uehlinger, S., Vergunst, A., Viteri, D. F., Nguyen, D. T., et al. (2012). Exposing the third chromosome of *Burkholderia cepacia* complex strains as a virulence plasmid. *Mol. Microbiol.* 83, 362–378. doi: 10.1111/j.1365-2958.2011.07937.x
- Agúndez, L., González-Prieto, C., Machón, C., and Llosa, M. (2012). Site-specific integration of foreign dna into minimal bacterial and human target sequences mediated by a conjugative relaxase. *PLoS ONE* 7:e31047. doi: 10.1371/journal.pone.0031047
- Alvarez-Martinez, C. E., and Christie, P. J. (2009). Biological diversity of prokaryotic type IV secretion systems. *Microbiol. Mol. Biol. Rev.* 73, 775–808. doi: 10.1128/MMBR.00023-09
- Asheshov, E. H. (1966). Loss of antibiotic resistance in *Staphylococcus aureus* resulting from growth at high temperature. *J. Gen. Microbiol.* 42, 403–410. doi: 10.1099/00221287-42-3-403
- Becker, E. C., and Meyer, R. J. (2000). Recognition of *oriT* for DNA processing at termination of a round of conjugal transfer. *J. Mol. Biol.* 300, 1067–1077. doi: 10.1006/jmbi.2000.3902
- Berriatua, E., Ziluaga, I., Miguel-Virto, C., Uribarren, P., Juste, R., Laevens, S., et al. (2001). Outbreak of subclinical mastitis in a flock of dairy sheep associated with *Burkholderia cepacia* complex infection. *J. Clin. Microbiol.* 39, 990–994. doi: 10.1128/JCM.39.3.990-994.2001
- Chandler, M., and Galas, D. J. (1983). Cointegrate formation mediated by Tn9. II. Activity of IS1 is modulated by external DNA sequences. *J. Mol. Biol.* 170, 61–91. doi: 10.1016/S0022-2836(83)80227-7

## AUTHOR CONTRIBUTIONS

EF contributed to the experimental set-up, analysis, interpretation of data and drafting the work; SB and MG contributed to the experimental results; DO and FS contributed to the conception of the work and critical revision of the manuscript; FS, AV, and ML contributed to the design of the work, interpretation of data, drafting, and revising the manuscript. All the authors approved the final version of the manuscript.

## FUNDING

This work was supported by grants BIO2013-46414-P from the Spanish Ministry of Economy and Competitiveness to ML and BFU2011-25658 from the Spanish Ministry of Science and Innovation to FS. U1047 is supported by INSERM and the Université de Montpellier. AV was recipient of a “Chercheur d’avenir” award from La Region Languedoc-Roussillon, and SB was supported by a doctoral fellowship co-funded by the Region Languedoc-Roussillon and INSERM.

## ACKNOWLEDGMENTS

We are grateful to Juan M. García-Lobo for assistance with the bioinformatics analysis.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmolb.2016.00016>

- Cho, H., and Winans, S. C. (2007). TraA, TraC and TraD autorepress two divergent quorum-regulated promoters near the transfer origin of the Ti plasmid of *Agrobacterium tumefaciens*. *Mol. Microbiol.* 63, 1769–1782. doi: 10.1111/j.1365-2958.2007.05624.x
- Christie, P. J., and Vogel, J. P. (2000). Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol.* 8, 354–360. doi: 10.1016/S0966-842X(00)01792-3
- Coenye, T., and Vandamme, P. (2003). Diversity and significance of *Burkholderia* species occupying diverse ecological niches. *Environ. Microbiol.* 5, 719–729. doi: 10.1046/j.1462-2920.2003.00471.x
- Cramer, R., Davies, J. E., and Hütter, R. (1986). Plasmid curing and generation of mutations induced with ethidium bromide in streptomycetes. *J. Gen. Microbiol.* 132, 819–824. doi: 10.1099/00221287-132-3-819
- Darling, P., Chan, M., Cox, A. D., and Sokol, P. A. (1998). Siderophore production by cystic fibrosis isolates of *Burkholderia cepacia*. *Infect. Immun.* 66, 874–877.
- Demarre, G., Guérout, A. M., Matsumoto-Mashimo, C., Rowe-Magnus, D. A., Marlière, P., and Mazel, D. (2005). A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPα) conjugative machineries and their cognate *Escherichia coli* host strains. *Res. Microbiol.* 156, 245–255. doi: 10.1016/j.resmic.2004.09.007
- Dennis, J. J., and Zylstra, G. J. (1998). Plasmids: modular self-cloning minitransposon derivatives for rapid genetic analysis of gram-negative bacterial genomes. *Appl. Environ. Microbiol.* 64, 2710–2715.
- Engledow, A. S., Medrano, E. G., Mahenthiralingam, E., LiPuma, J. J., and Gonzalez, C. F. (2004). Involvement of a plasmid-encoded type IV secretion



- system in the plant tissue watersoaking phenotype of *Burkholderia cenocepacia*. *J. Bacteriol.* 186, 6015–6024. doi: 10.1128/JB.186.18.6015-6024.2004
- Estrada-De Los Santos, P., Vinuesa, P., Martínez-Aguilar, L., Hirsch, A. M., and Caballero-Mellado, J. (2013). Phylogenetic analysis of burkholderia species by multilocus sequence analysis. *Curr. Microbiol.* 67, 51–60. doi: 10.1007/s00284-013-0330-9
- Fernández-González, E., de Paz, H. D., Alperi, A., Agúndez, L., Faustmann, M., Sangari, F. J., et al. (2011). Transfer of R388 derivatives by a pathogenesis-associated type IV secretion system into both bacteria and human cells. *J. Bacteriol.* 193, 6257–6265. doi: 10.1128/JB.05905-11
- Fernández-López, C., Lorenzo-Díaz, F., Pérez-Luque, R., Rodríguez-González, L., Boer, R., Lurz, R., et al. (2013). Nicking activity of the pMV158 MobM relaxase on cognate and heterologous origins of transfer. *Plasmid* 70, 120–130. doi: 10.1016/j.plasmid.2013.03.004
- Francia, M. V., Varsaki, A., Garcillán-Barcia, M. P., Latorre, A., Drinas, C., and De La Cruz, F. (2004). A classification scheme for mobilization regions of bacterial plasmids. *FEMS Microbiol. Rev.* 28, 79–100. doi: 10.1016/j.femsre.2003.09.001
- Garcillán-Barcia, M. P., and de la Cruz, F. (2008). Why is entry exclusion an essential feature of conjugative plasmids? *Plasmid* 60, 1–18. doi: 10.1016/j.plasmid.2008.03.002
- Garcillán-Barcia, M. P., Francia, M. V., and De La Cruz, F. (2009). The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol. Rev.* 33, 657–687. doi: 10.1111/j.1574-6976.2009.00168.x
- Gonzalez, C. F., Pettit, E. A., Valadez, V. A., and Provin, E. M. (1997). Mobilization, cloning, and sequence determination of a plasmid-encoded polygalacturonase from a phytopathogenic *Burkholderia* (*Pseudomonas*) *cepacia*. *Mol. Plant. Microbe Interact.* 10, 840–851. doi: 10.1094/MPMI.1997.10.7.840
- Goven, J. R. W., Brown, P. H., Maddison, J., Doherty, C. J., Nelson, J. W., Dodd, M., et al. (1993). Evidence for transmission of *Pseudomonas cepacia* by social contact in cystic fibrosis. *Lancet* 342, 15–19. doi: 10.1016/0140-6736(93)91881-L
- Grandoso, G., Avila, P., Cayón, A., Hernando, M. A., Llosa, M., and de la Cruz, F. (2000). Two active-site tyrosyl residues of protein TrwC act sequentially at the origin of transfer during plasmid R388 conjugation. *J. Mol. Biol.* 295, 1163–1172. doi: 10.1006/jmbi.1999.3425
- Grant, S. G., Jessee, J., Bloom, F. R., and Hanahan, D. (1990). Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc. Natl. Acad. Sci. U.S.A.* 87, 4645–4649. doi: 10.1073/pnas.87.12.4645
- Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A., Gwinn, M. L., Dodson, R. J., et al. (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406, 477–483. doi: 10.1038/35020000
- Holden, M. T. G., Seth-Smith, H. M. B., Crossman, L. C., Sebahia, M., Bentley, S. D., Cerdeño-Tárraga, A. M., et al. (2009). The genome of *Burkholderia cenocepacia* J2315, an epidemic pathogen of cystic fibrosis patients. *J. Bacteriol.* 191, 261–277. doi: 10.1128/JB.01230-08
- Jandle, S., and Meyer, R. (2006). Stringent and relaxed recognition of oriT by related systems for plasmid mobilization: implications for horizontal gene transfer. *J. Bacteriol.* 188, 499–506. doi: 10.1128/JB.188.2.499-506.2006
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053
- Koichiro, T., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Kovach, M. E., Phillips, R. W., Elzer, P. H., Roop, R. M., and Peterson, K. M. (1994). pBBR1MCS: a broad-host-range cloning vector. *Biotechniques* 16, 800–802.
- Langer, P. J., and Walker, G. C. (1981). Restriction endonuclease cleavage map of pKM101: relationship to parental plasmid R46. *Mol. Gen. Genet.* 182, 268–272. doi: 10.1007/BF00269669
- Lawley, T. D., Klimke, W. A., Gubbins, M. J., and Frost, L. S. (2003). F factor conjugation is a true type IV secretion system. *FEMS Microbiol. Lett.* 224, 1–15. doi: 10.1016/S0378-1097(03)00430-0
- LiPuma, J. J. (1998a). Burkholderia cepacia epidemiology and pathogenesis: implications for infection control. *Curr. Opin. Pulm. Med.* 4, 337–341.
- LiPuma, J. J. (1998b). Burkholderia cepacia. Management issues and new insights. *Clin. Chest Med.* 19, 473–486, vi.
- LiPuma, J. J. (2003). Burkholderia and emerging pathogens in Cystic Fibrosis. *Semin. Respir. Crit. Care Med.* 24, 681–692. doi: 10.1055/s-2004-815664
- Llosa, M., Bolland, S., and de la Cruz, F. (1994). Genetic organization of the conjugal DNA processing region of the IncW plasmid R388. *J. Mol. Biol.* 235, 448–464. doi: 10.1006/jmbi.1994.1005
- Llosa, M., Grandoso, G., Hernando, M. A., and de la Cruz, F. (1996). Functional domains in protein TrwC of plasmid R388: dissected DNA strand transferase and DNA helicase activities reconstitute protein function. *J. Mol. Biol.* 264, 56–67. doi: 10.1006/jmbi.1996.0623
- Llosa, M., Zunzunegui, S., and de la Cruz, F. (2003). Conjugative coupling proteins interact with cognate and heterologous VirB10-like proteins while exhibiting specificity for cognate relaxosomes. *Proc. Natl. Acad. Sci. U.S.A.* 100, 10465–10470. doi: 10.1073/pnas.1830264100
- Lucas, M., González-Pérez, B., Cabezas, M., Moncalián, G., Rivas, G., and De La Cruz, F. (2010). Relaxase DNA binding and cleavage are two distinguishable steps in conjugative DNA processing that involve different sequence elements of the nic site. *J. Biol. Chem.* 285, 8918–8926. doi: 10.1074/jbc.M109.057539
- Mahenthalingam, E., Coenye, T., Chung, J. W., Speert, D. P., Govan, J. R. W., Taylor, P., et al. (2000). Diagnostically and experimentally useful panel of strains from the Burkholderia cepacia complex. *J. Clin. Microbiol.* 38, 910–913.
- Martinez, E., and de la Cruz, F. (1988). Transposon Tn21 encodes a RecA-independent site-specific integration system. *Mol. Gen. Genet.* 211, 320–325. doi: 10.1007/BF00330610
- Matson, S. (1990). DNA Helicases. *Annu. Rev. Biochem.* 59, 289–329. doi: 10.1146/annurev.bi.59.070190.001445
- Moncalián, G., and De La Cruz, F. (2004). DNA binding properties of protein TrwA, a possible structural variant of the Arc repressor superfamily. *Biochim. Biophys. Acta* 1701, 15–23. doi: 10.1016/j.bbapap.2004.05.009
- Moncalián, G., Grandoso, G., Llosa, M., and de la Cruz, F. (1997). oriT-processing and regulatory roles of TrwA protein in plasmid R388 conjugation. *J. Mol. Biol.* 270, 188–200. doi: 10.1006/jmbi.1997.1082
- O'Brien, F. G., Eto, K. Y., Murphy, R. J. T., Fairhurst, H. M., Coombs, G. W., Grubb, W. B., et al. (2015). Origin-of-transfer sequences facilitate mobilisation of non-conjugative antimicrobial-resistance plasmids in *Staphylococcus aureus*. *Nucleic Acids Res.* 43, 7971–7983. doi: 10.1093/nar/gkv755
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi: 10.1093/nar/gkt1226
- Pérez-Mendoza, D., Lucas, M., Muñoz, S., Herrera-Cervera, J. A., Olivares, J., De La Cruz, F., et al. (2006). The relaxase of the *Rhizobium etli* symbiotic plasmid shows nic site cis-acting preference. *J. Bacteriol.* 188, 7488–7499. doi: 10.1128/JB.00701-06
- Pollet, R. M., Ingle, J. D., Hymes, J. P., Eakes, T. C., Eto, K. Y., Kwong, S. M., et al. (2016). Processing of non-conjugative resistance plasmids by conjugation nicking enzyme of Staphylococci. *J. Bacteriol.* 198, 888–897. doi: 10.1128/JB.00832-15
- Ramage, H. R., Connolly, L. E., and Cox, J. S. (2009). Comprehensive functional analysis of *Mycobacterium tuberculosis* toxin-antitoxin systems: Implications for pathogenesis, stress responses, and evolution. *PLoS Genet.* 5:e1000767. doi: 10.1371/journal.pgen.1000767
- Sadler, J. R., Tecklenburg, M., and Betz, J. L. (1980). Plasmids containing many tandem copies of a synthetic lactose operator. *Gene* 8, 279–300. doi: 10.1016/0378-1119(80)90005-0
- Sajjan, S. U., Carmody, L. A., Gonzalez, C. F., and Lipuma, J. J. (2008). A type IV secretion system contributes to intracellular survival and replication of *Burkholderia cenocepacia*. *Infect. Immun.* 76, 5447–5455. doi: 10.1128/IAI.00451-08
- Sambrook, J. W., and Russell, D. (2001). *Molecular cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Sepúlveda, E., Pérez-Mendoza, D., Ramírez-Romero, M. A., Soto, M. J., López-Lara, I. M., Geiger, O., et al. (2008). Transcriptional interference and repression modulate the conjugative ability of the symbiotic plasmid of *Rhizobium etli*. *J. Bacteriol.* 190, 4189–4197. doi: 10.1128/JB.00041-08
- Speert, D. P. (2002). Advances in *Burkholderia cepacia* complex. *Paediatr. Respir. Rev.* 3, 230–235. doi: 10.1016/S1526-0542(02)00185-9



- Valvano, M. A. (2015). Intracellular survival of *Burkholderia cepacia* complex in phagocytic cells. *Can. J. Microbiol.* 615, 607–615. doi: 10.1139/cjm-2015-0316
- Vandamme, P., and Dawyndt, P. (2011). Classification and identification of the *Burkholderia cepacia* complex: past, present and future. *Syst. Appl. Microbiol.* 34, 87–95. doi: 10.1016/j.syapm.2010.10.002
- Varga, J. J., Losada, L., Zelazny, A. M., Kim, M., McCorrison, J., Brinkac, L., et al. (2013). Draft Genome Sequences of *Burkholderia cenocepacia* ET12 Lineage Strains K56-2 and BC7. *Genome Announc.* 1, e00841–e00813. doi: 10.1128/genomeA.00841-13
- Wernegreen, J. J., and Riley, M. A. (1999). Comparison of the evolutionary dynamics of symbiotic and housekeeping loci: a case for the genetic coherence of rhizobial lineages. *Mol. Biol. Evol.* 16, 98–113. doi: 10.1093/oxfordjournals.molbev.a026041
- Williams, S. L., and Schildbach, J. F. (2006). Examination of an inverted repeat within the F factor origin of transfer: context dependence of F TraI relaxase DNA specificity. *Nucleic Acids Res.* 34, 426–435. doi: 10.1093/nar/gkj444
- Williams, S. L., and Schildbach, J. F. (2007). TraY and integration host factor oriT binding sites and F conjugal transfer: sequence variations, but not altered spacing, are tolerated. *J. Bacteriol.* 189, 3813–3823. doi: 10.1128/JB.01783-06
- Wong, J. J. W., Lu, J., Edwards, R. A., Frost, L. S., and Glover, J. N. M. (2011). Structural basis of cooperative DNA recognition by the plasmid conjugation factor, TraM. *Nucleic Acids Res.* 39, 6775–6788. doi: 10.1093/nar/gkr296
- Yamaguchi, Y., Park, J.-H., and Inouye, M. (2011). Toxin-Antitoxin systems in bacteria and Archaea. *Annu. Rev. Genet.* 45, 61–79. doi: 10.1146/annurev-genet-110410-132412
- Zechner, E. L., La Cruz, F., Eisenbrandt, R., Grahm, A. M., Koraimann, G., Lanka, E., et al. (2000). “Conjugative-DNA transfer processes,” in *The Horizontal Gene Pool*, ed C. Thomas (Reading: Harwood Academic Publisher), 87–174.
- Zhang, R., LiPuma, J. J., and Gonzalez, C. F. (2009). Two type IV secretion systems with different functions in *Burkholderia cenocepacia* K56-2. *Microbiology* 155, 4005–4013. doi: 10.1099/mic.0.033043-0
- Zhong, Z., Helinski, D., and Toukdarian, A. (2005). Plasmid host-range: restrictions to F replication in *Pseudomonas*. *Plasmid* 54, 48–56. doi: 10.1016/j.plasmid.2004.11.001

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Fernández-González, Bakioui, Gomes, O’Callaghan, Vergunst, Sangari and Llosa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Comparative Genomics of the Conjugation Region of F-like Plasmids: Five Shades of F

Raul Fernandez-Lopez<sup>1\*</sup>, Maria de Toro<sup>2</sup>, Gabriel Moncalian<sup>1</sup>, M. Pilar Garcillan-Barcia<sup>1</sup> and Fernando de la Cruz<sup>1\*</sup>

<sup>1</sup> Instituto de Biomedicina y Biotecnología de Cantabria, Santander, Spain, <sup>2</sup> Centro de Investigación Biomedica de la Rioja, Logroño, Spain

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas Medical  
Branch, USA

### Reviewed by:

Teresa M. Coque,  
Instituto Ramón y Cajal de  
Investigación Sanitaria, Spain  
Günther Koraimann,  
University of Graz, Austria

### \*Correspondence:

Fernando de la Cruz  
delacruz@unican.es  
Raul Fernandez Lopez  
fernandr@unican.es

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 29 August 2016

**Accepted:** 18 October 2016

**Published:** 10 November 2016

### Citation:

Fernandez-Lopez R, de Toro M,  
Moncalian G, Garcillan-Barcia MP and  
de la Cruz F (2016) Comparative  
Genomics of the Conjugation Region  
of F-like Plasmids: Five Shades of F.  
Front. Mol. Biosci. 3:71.  
doi: 10.3389/fmolb.2016.00071

The F plasmid is the foremost representative of a large group of conjugative plasmids, prevalent in *Escherichia coli*, and widely distributed among the Enterobacteriaceae. These plasmids are of clinical relevance, given their frequent association with virulence determinants, colicins, and antibiotic resistance genes. Originally defined by their sensitivity to certain male-specific phages, IncF plasmids share a conserved conjugative system and regulatory circuits. In order to determine whether the genetic architecture and regulation circuits are preserved among these plasmids, we analyzed the natural diversity of F-like plasmids. Using the relaxase as a phylogenetic marker, we identified 256 plasmids belonging to the IncF/ MOB<sub>F12</sub> group, present as complete DNA sequences in the NCBI database. By comparative genomics, we identified five major groups of F-like plasmids. Each shows a particular operon structure and alternate regulatory systems. Results show that the IncF/MOB<sub>F12</sub> conjugation gene cluster conforms a diverse and ancient group, which evolved alternative regulatory schemes in its adaptation to different environments and bacterial hosts.

**Keywords:** plasmids, plasmid conjugation, IncF incompatibility group, plasmid evolution, antibiotic resistance

## INTRODUCTION

The IncF incompatibility group comprises a diverse set of conjugative plasmids frequently found in enterobacterial species like *E. coli* and *Salmonella*. This group was named after F: the factor found by Joshua Lederberg to be responsible for bacterial conjugation in *E. coli* K-12 (Lederberg and Tatum, 1946). The F factor was originally thought to be involved in some sort of para-sexual reproduction in *E. coli* (Makela et al., 1962), thus it was originally named the fertility factor, or F. Bacterial strains able to transmit genetic traits by conjugation were deemed fertile, or F<sup>+</sup>. It was also believed that fertility was different from R factors: self-transmissible episomes conferring antibiotic resistance to their hosts (Watanabe, 1967; Meynell et al., 1968a). Soon it was found that many R factors were sensitive to male-specific phages that infected F-bearing cells (Brinton et al., 1964; Caro and Schnös, 1966; Dennison, 1972). Serological testing revealed that many of these plasmids produced immunological cross-reactions (Orskov and Orskov, 1960; Ishibashi, 1967). Besides, it was observed that they were often unable to co-reside within the same recipient cell (Meynell et al., 1968b). Thus, it was concluded that F and some R plasmids constituted a distinct group, probably sharing a similar genetic structure, and a common ancestor (Meynell et al., 1968a). With the advent of DNA sequencing techniques, this idea was partially confirmed: IncF plasmids share a common set of genes involved in the genesis of the conjugative pilus. This is the reason behind their

common phage sensitivity profile and serological cross-reactivity. Besides their common mating apparatus, F-like plasmids appear to be functionally diverse. For instance, they may encode different replication and partition systems (Ogura and Hiraga, 1983; Gerdes and Molin, 1986), and a wide diversity of cargo genes (Lanza et al., 2014; Johnson et al., 2016).

The F pilus is thus the common denominator of the IncF/MOB<sub>F12</sub> group. F pili are distinct from other sex-related pili, such as the P, N, W, or X pili (4). The conjugation regions of these plasmids show similarity at the protein level to the VirB system of *Agrobacterium*, constituting prototypic Type IV secretion systems (T4SS) (Krause et al., 2000; Smillie et al., 2010; Chandran Darbari and Waksman, 2015). The F-pilus, however, is a more distant relative from VirB systems, albeit a true T4SS (Lawley et al., 2003). Unlike the short, rigid VirB-like pili, F pili are long and flexible, and able to retract upon contacting a recipient cell (Clarke et al., 2008). The genetic region involved in F conjugation is significantly longer and contains more genes than those of VirB-like pili forming plasmids (roughly 34 kb vs. 15 kb) (Kennedy et al., 1977; Frost et al., 1994; Lawley et al., 2003). One of its most conspicuous features is that all *tra* genes are transcribed from a single promoter (Helmuth and Achtman, 1975). The *tra* operon spans nearly 40 kb, making it, to the best of our knowledge, the longest transcript ever found in *E. coli*. Despite this simple operon arrangement, regulation of the transfer functions in IncF/MOB<sub>F12</sub> plasmids is complicated. Expression of F conjugative functions is controlled by three transcriptional regulators: TraM, TraJ, and TraY (Frost and Koraimann, 2010; Arutyunov and Frost, 2013). From these three proteins, TraM and TraY play an additional role in relaxosome assembly (Wong et al., 2012; Lang et al., 2014). TraJ is the key activator of the Py promoter, responsible for the transcription of *tra* genes (Finnegan and Willetts, 1973; Frost and Koraimann, 2010). TraJ is regulated at the translational level by a small antisense RNA, FinP. FinP binds *traJ* mRNA, blocking its translation (Timmis et al., 1978; Arthur et al., 2003; Mark Glover et al., 2015). This process is assisted by the action of a key RNA chaperone, FinO (Ghetu et al., 2000). The *finOP* regulatory system constitutes the major controller of *tra* expression, and thus was named fertility inhibition system (Mark Glover et al., 2015). Besides this plasmid-encoded system, a relatively large number of host-encoded factors also modulate the expression of F transfer functions. In classical F-like plasmids, the key host factors regulating transfer expression are the transcriptional regulators ArcA, which co-activates the Py along with TraJ, and HNS, which acts as a silencing factor of the PY promoter. Besides these, other host factors like Lrp (leucine responsive regulatory protein), ArcB (anaerobic repressor of the arc modulon) and RNase E have been shown to modulate the expression of *tra* functions (Frost and Koraimann, 2010). The action of these host-encoded factors is often plasmid-specific, thus not all IncF/MOB<sub>F12</sub> plasmids are linked to the host regulatory network in the same fashion. Paradoxically, the oddest case among naturally-isolated IncF plasmids is factor F itself. The F plasmid is a *finO*- mutant, produced by insertion of a IS3 insertion sequence (Yoshioka et al., 1987). Thus, it contains a non-functional fertility inhibition system, and exhibits

conjugation frequencies two or three orders of magnitude above other naturally occurring IncF/MOB<sub>F12</sub> plasmids like R1, R100, or pSLT (Frost and Koraimann, 2010).

These three plasmids (R1, R100, and pSLT) are considered “classical” IncF plasmids because they were extensively studied in the pre-genomic era. Plasmid R1 was transferred from its original host *Salmonella enterica* (serovar Paratyphi) to *E. coli*, conferring resistance to ampicillin, kanamycin, chloramphenicol and sulfonamides (Meynell and Datta, 1966). Plasmid R100 (also named NR1) was isolated from *Shigella flexneri* 2b, and encoded resistances to chloramphenicol, tetracycline and streptomycin (Nakaya et al., 1960). It was later found that R100 also provided the host cell with resistance to organomercury compounds (Womble and Rownd, 1988). pSLT was intensively studied because its role in the virulence of *Salmonella enterica* (serovar Typhimurium). Although the repertoire of classical F-like plasmids was reduced, it was observed that these plasmids presented significant differences in the regulation of the conjugative functions. With the advent of next-generation sequencing techniques, the genomes of hundreds of plasmids similar to classical IncF prototypes became available. Indeed, systematic studies of *E. coli* epidemic clones, like the widely distributed ST131, revealed an extraordinary prevalence of IncF plasmids in natural isolates (Lanza et al., 2014). This opens a number of questions regarding the conservation of IncF *tra* functions, and specially the regulatory circuits governing them. It is not known, for example, whether *finO*- plasmids like F itself are frequent among natural populations, or whether alternate regulatory schemes of the F conjugation machinery exist in nature. Using the relaxase as a phylogeny marker, we identified 256 IncF/MOB<sub>F12</sub> plasmids in the NCBI plasmid database. By comparing the genomic structure of their conjugation regions, we identified five major groups displaying idiosyncratic genetic structures and alternative regulatory schemes. These five groups correspond to well-supported branches in the relaxase phylogenetic tree, indicating that these five groups represent radiations of an ancestral MOB<sub>F</sub> conjugation system.

## MATERIALS AND METHODS

### Selection of the Plasmid Dataset

In order to identify MOB<sub>F12</sub> plasmids, an initial search using a set of 26 known MOB<sub>F12</sub> relaxases (**Table S1**) taken from previous studies (Garcillán-Barcia et al., 2009; Alvarado et al., 2012) was carried out. MOB<sub>F12</sub> relaxases were defined as those having (D/E)NYY and D(L/F)TF amino acid motifs in the N-terminal relaxase domain of the protein (Garcillán-Barcia et al., 2009). These 26 relaxases were used as baits in protein BLAST searches of the NCBI plasmid database (6079 plasmids, 20th October 2015) using a threshold *e*-value of 1E-25. In this way, we retrieved a total of 256 plasmids containing relaxases that showed at least 40% sequence identity at the protein level to its closest database reference. This threshold was selected upon realizing that relaxases with low identity (down to 26%) were retrieved in an initial search. However, when we compared two of the most distant MOB<sub>F12</sub> plasmids, for instance, F and pAsa5, they showed an ID of 49%. We therefore established as selection criteria that

IncF/MOB<sub>F12</sub> plasmids are those whose relaxase showed at least 40% ID with respect our homemade MOB<sub>F12</sub> relaxase DB.

## Construction of Protein Profiles

To construct the presence/absence profile of the *tra* proteins of MOB<sub>F12</sub> plasmids, Psi-blast searches (Altschul et al., 1997) were performed against a protein database constructed from the annotations of the 256 MOB<sub>F12</sub> plasmids (Table S1) using the following conjugative proteins as queries: F plasmid proteins TraJ, TraA, TraL, TraE, TraK, TraB, TraP, TrbD, TrbG, TraV, TraR, TraC, TrbI, TraW, TraU, TrbC, TraN, TrbE, TraF, TrbA, ArtA, TraQ, TrbB, TrbJ, TrbF, TraH, TraG, TraS, TraT, TraD, TrbH, TraI (N-terminal 300 amino acids), and TraX as well as TraM, TraY, and FinO from plasmid R100. By default, hits below an *e*-value of 1E-3 were considered as positive hits. Selected protein hits were aligned using MUSCLE (Edgar, 2004). The resulting global alignments were used to reconstruct maximum-likelihood (ML) phylogenies using RAXML version 7.2.7 (Stamatakis, 2006). Twenty ML trees were executed using the JTTGAMMA model, and 100 bootstrap trees were inferred to obtain the confidence values for each node of the best ML tree.

## Co-occurrence Matrix

To compute the co-occurrence matrix between *tra* genes, we used the presence/absence profile of all *tra* genes in our 256 plasmid dataset. Thus, for each *tra* gene, we defined a vector of 256 elements, with values 1 or 0. We then calculated the Hamming distance between each pair of vectors. Co-occurrence between two *tra* genes was expressed as the maximum possible distance (256) minus the Hamming distance between the pair of genes.

## Structural Modeling

The 3D structures of TraJ<sub>V</sub>, EntFR, and SphTR were predicted by homology modeling using the Phyre2 server (Kelley and Sternberg, 2009). Images of the resulting 3D models were generated using Pymol (DeLano Scientific, Palo Alto, CA, USA).

## RESULTS

### The MOB<sub>F12</sub> Phylogenetic Tree Includes Plasmids from $\alpha$ and $\gamma$ -Proteobacteria

In order to identify IncF/MOB<sub>F12</sub> plasmids present in the databases, we used the conjugative relaxase gene as the lowest common denominator. This property of the relaxase to serve as classification guide was shown in previous works, and is widely used for plasmid typing (Garcillán-Barcia et al., 2009; Alvarado et al., 2012). In order to identify MOB<sub>F12</sub> relaxase-containing plasmids present in the NCBI plasmid database (6079 plasmids; October 20, 2015), we employed a set of 26 known MOB<sub>F12</sub> relaxases as baits (Table S1), as indicated in Materials and Methods. We retrieved a total of 256 plasmids containing relaxases that showed at least 40% sequence identity at the amino acid level to their closest database reference (see M&M). The resulting plasmid list is shown in Table S1. To reconstruct the phylogeny of MOB<sub>F12</sub> plasmids, we aligned the N-300 residues of the relaxase proteins as described in Materials and Methods and constructed a ML phylogenetic tree. We rooted

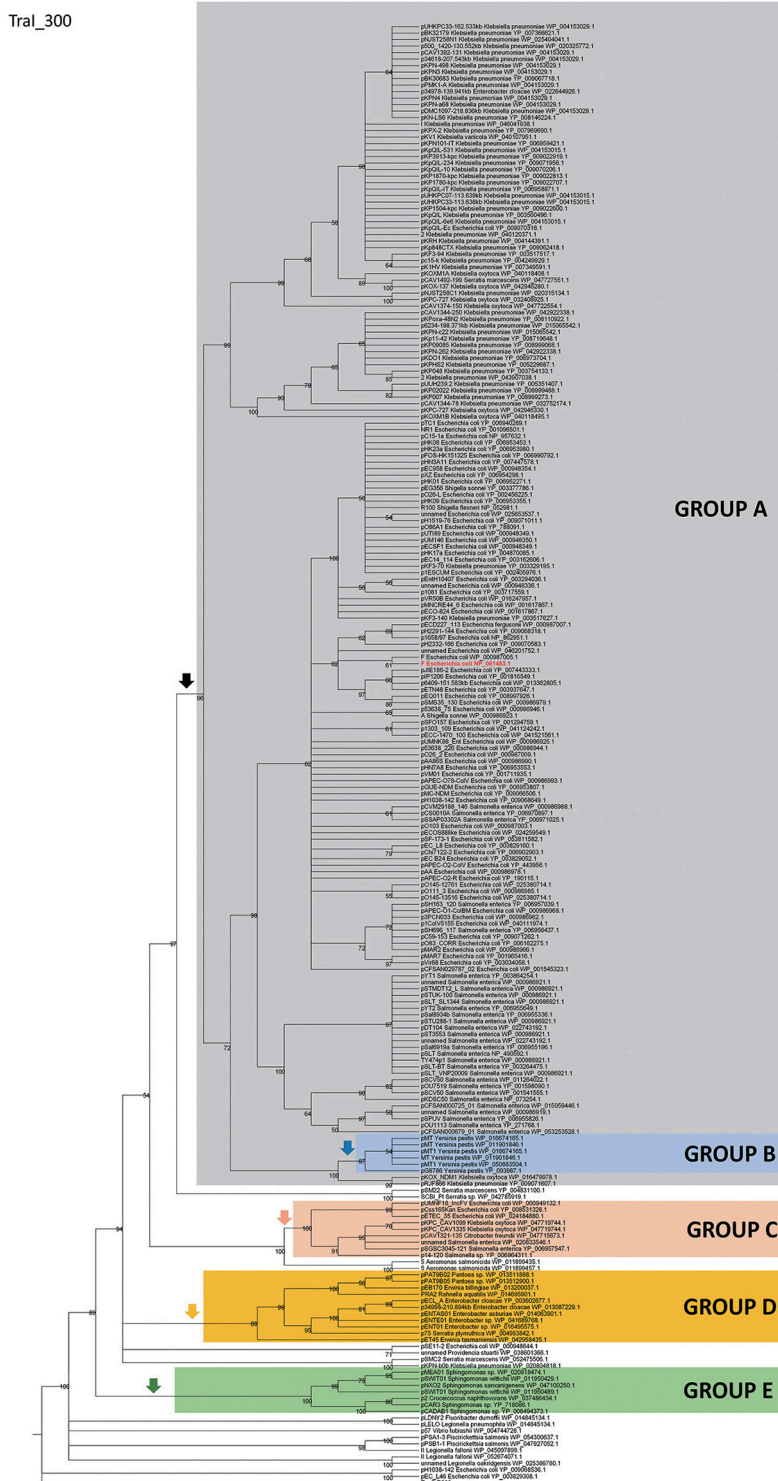
the tree using the MOB<sub>F11</sub> relaxase TrwC\_R388 as outgroup. The resulting tree is shown in Figure 1. The MOB<sub>F12</sub> phylogenetic tree includes plasmids isolated from  $\alpha$  and  $\gamma$ -Proteobacteria, with 91% coming from species within *Enterobacteria*. The tree showed that 88% of the plasmids coming from enterobacterial species clustered in a monophyletic branch, well-supported by the bootstrap value (Figure 1, black vertical arrow). This branch included the “classical” IncF plasmids F, R1, R100, and pSLT. Enterobacterial plasmids not belonging to this branch included several from *Enterobacter*, which instead clustered in a second monophyletic branch (Figure 1, orange arrow). A third set of plasmids from *Escherichia*, *Salmonella*, and *Klebsiella* appeared in a third monophyletic branch (Figure 1, red arrow). As we will show later, these clusters contain plasmids harboring a typical MOB<sub>F12</sub> conjugation region, but showing different regulatory systems. Plasmids from  $\alpha$ -Proteobacteria appeared in an ancestral, monophyletic group (Figure 1, green arrow).

### Relaxases of the MOB<sub>F12</sub> Branch Are Associated to F-like Conjugative Systems

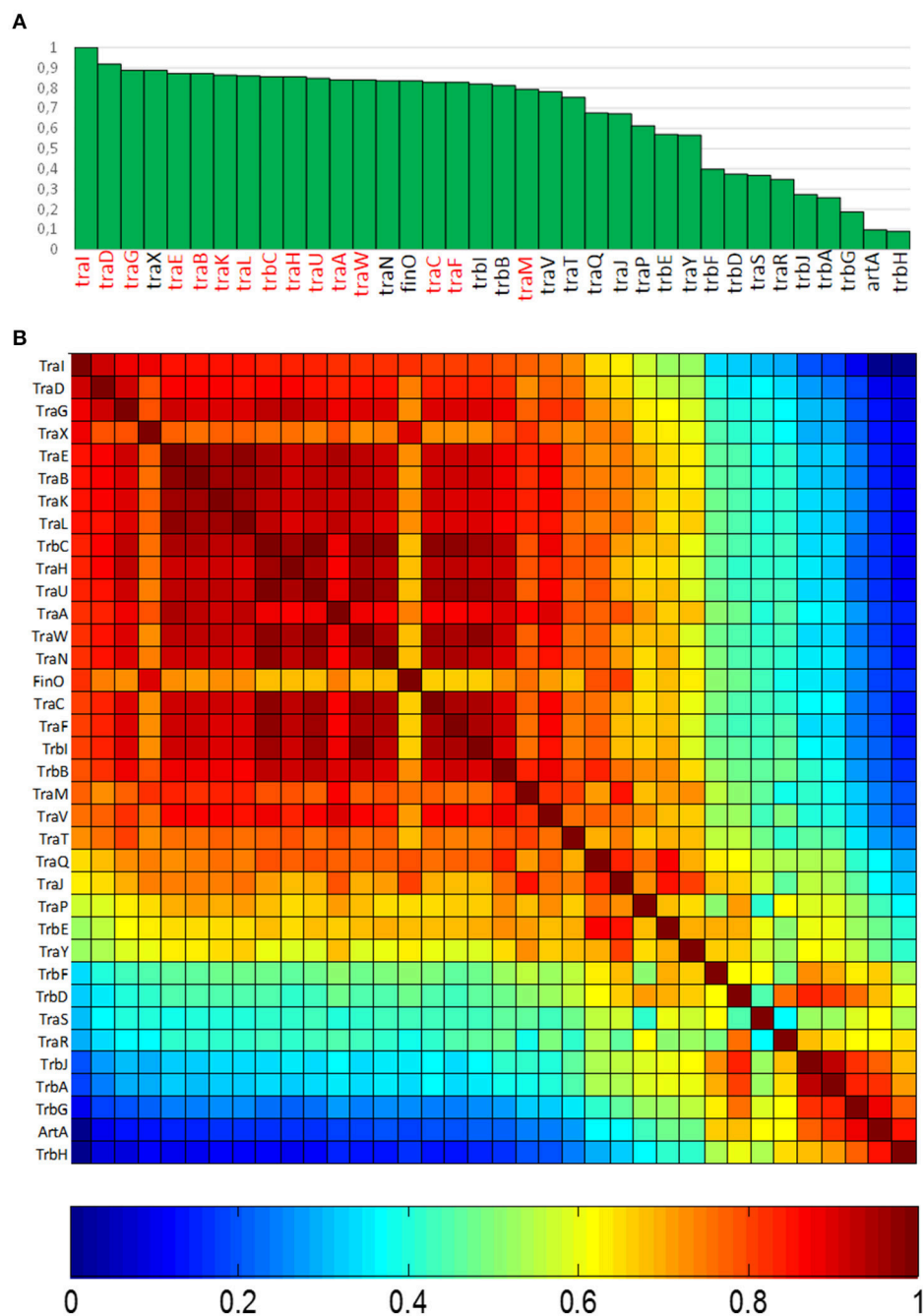
In order to determine the degree of association of MOB<sub>F12</sub> relaxases to the canonical F pilus, we determined the presence/absence of homologs of the F conjugation genes in our 256 plasmid dataset. For this purpose, we selected a total of 36 genes present in the conjugation region of plasmids F and R100, two well-studied IncF prototypes. As described in M&M, the presence of homologs of these 36 reference genes was determined by PSI-BLAST. Table S1 shows the accession numbers of each homolog identified for the 256 plasmids analyzed. Table S1 consists of a matrix in which rows (i) correspond to each MOB<sub>F12</sub> plasmid, while columns (j) indicate each *tra* gene. Thus, reference numbers in the i,jth position of the matrix correspond to the homolog to protein j present in plasmid i. We transformed this table into a binary matrix, such that each i,jth position was 1 if there was a homolog detectable by PSI-Blast, and 0 otherwise. This allowed us to determine the level of overall conservation of F conjugation genes among the 256 plasmids. Results showed that a TraD-like protein (the coupling protein (T4CP) of F-like plasmids) could be detected in 91% of the plasmids (Figure 2A). This intimate phylogenetic association between the relaxase and the coupling protein was shown previously to be a hallmark of *mob* genes (Fernández-López, et al., 2006; Garcillán-Barcia et al., 2011). The analysis of other plasmid groups showed, however, that the association between MOB and MPF genes is less stringent. For example, MOB<sub>F11</sub> relaxases are associated to N or W pili (variants of MPF<sub>T</sub>) (Fernández-López, et al., 2006; Garcillán-Barcia et al., 2011). As shown in Figure 2A, the presence of MPF<sub>F</sub> genes could be detected in more than 80% of the MOB<sub>F12</sub> plasmids in our dataset. Results also indicated that the conservation of MPF<sub>F</sub> genes was not uniform. The most conserved MPF gene was *traG*, responsible for mating pair stabilization, which appeared in 88% of the plasmids. The least conserved gene was *trbH*, which could only be detected in 8% of the plasmids. Interestingly, genes that have been described as essential for F transfer appeared in more than 80% of the plasmids (Figure 2A,



TraL\_300



**FIGURE 1 | Phylogenetic tree of MOB<sub>F12</sub> relaxases.** The maximum-likelihood (ML) phylogenetic tree was built with the relaxase domain (N-terminal 300 residues) of 260 TraL\_F homologs encoded by 255 plasmids present in our dataset (plasmid pCFSAN029787\_01 relaxase was left out because it lacks the N-terminal relaxase domain). Bootstrap values are indicated at the corresponding nodes of the ML tree. The cut-off value for the condensed tree was chosen at bootstrap value = 50%. For each taxon, the plasmid name, the bacterial host, the GenBank protein accession number (excluding 13 non-annotated relaxases), and the GenBank plasmid accession numbers are indicated. The MOB<sub>F12</sub> prototype (TraL\_F) is highlighted in bold red letters. The MOB<sub>F11</sub> relaxase TrwC of plasmid R388 (the first N-terminal 300 residues of GenBank Acc. No. FA00039.1) was used as outgroup. Branches containing MOB<sub>F11</sub> relaxases are drawn in gray. The MOB<sub>F12</sub> cluster is indicated in the corresponding ancestral node. MOB<sub>F12</sub> groups according to the organization and regulation of the conjugation system (A-E) are shadowed in different colors.



**FIGURE 2 | Conservation of MPF<sub>F</sub> conjugation genes in MOB<sub>F12</sub> plasmids. (A)** The graph shows the percentage of plasmids (y axis) showing a gene homologous to each of the MPF<sub>F</sub> genes indicated in the x axis, for the 256 plasmids that showed a MOB<sub>F12</sub> relaxase. Genes were ordered according to their overall conservation. **(B)** Co-occurrence matrix of MPF<sub>F</sub> genes in the 256 MOB<sub>F12</sub> plasmids. The color matrix indicates the probability of co-occurrence (1, always appearing together in the same plasmid, 0 never appearing together in the same plasmid) for all MPF<sub>F</sub> gene pairs.

highlighted in red), while non-essential genes tend to appear at lower frequencies (**Figure 2A**, in black). An exception to this rule was *traX*, a gene that encodes an acetylase of pilin subunits. This gene was deemed non-essential for F plasmid conjugation (Maneewannakul et al., 1995), yet it was detected in 88% of the plasmids analyzed.

Since not all MPF<sub>F</sub> genes showed the same degree of conservation, we wondered whether there were genes that showed preferential co-occurrence. To determine this, we built a co-occurrence matrix for the 36 MPF<sub>F</sub> genes (Materials and Methods). Results (**Figure 2B**) showed that the highest co-occurrence values corresponded to the gene clusters *traEBKL*,

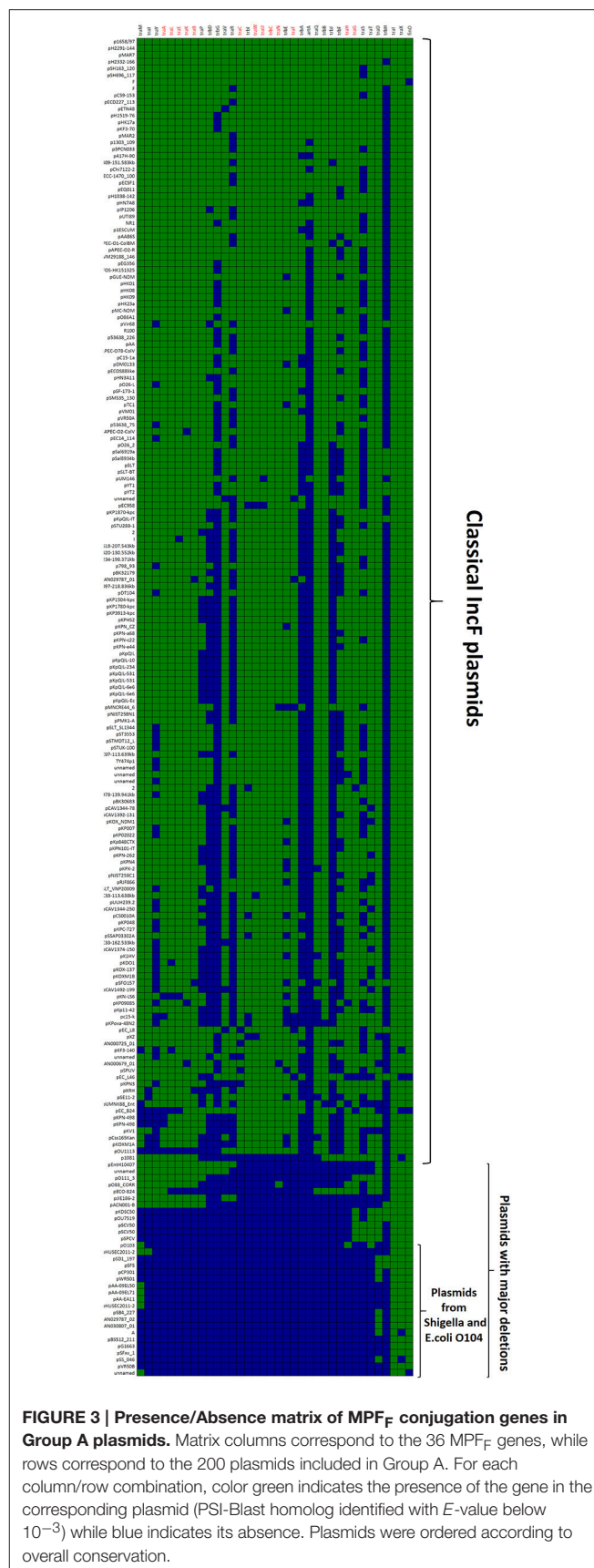
*trbCI*, and *traHUNCF*. These genes are essential components for the synthesis and function of F-pili, and thus their co-occurrence suggests the presence of functional F transfer systems. Interestingly, genes *trbA*, *trbG*, and *artA* also showed a high level of co-occurrence, despite their overall conservation is among the lowest overall (<25% identity). This indicates that this gene cluster is specific of a certain set of F-like plasmids. It was also noteworthy that regulatory genes (*traM*, *traY*, *traJ*, and *finO*) showed a lower degree of co-occurrence than structural genes, suggesting that alternative regulatory mechanisms could exist for MPF<sub>F</sub> conjugation systems.

## Clustering of F-like (MOB<sub>F12</sub>) Plasmids Based on Key Regulatory Genes of the Transfer Region

Since regulatory genes showed lower conservation than structural genes, we looked specifically at three key regulatory genes, namely *traM*, *traJ*, and *finO*. We did not include *traY* because, given its small size (around 225 bp), it is often not properly annotated. In order to distinguish between alternative regulatory schemes and major deletions that might have eliminated a substantial fraction of the transfer region, we also included a marker gene for the presence of the MPF apparatus. For this purpose, we used the essential ATPase *traC*. Using these genes as guidelines, we identified five major groups of F-like plasmids, which corresponded to major branches in the MOB<sub>F12</sub> phylogenetic tree.

### GROUP A: Classical F-like Plasmids. Prototype: Plasmid R1

The first and major MOB<sub>F12</sub> group includes a total of 200 plasmids (78% of the total), residing in the genera *Escherichia*, *Salmonella*, and *Klebsiella*. In the MOB<sub>F12</sub> tree of **Figure 1**, these relaxase genes are monophyletic. If we extrapolate from the relaxase to the whole TRA system, the tree structure implies that group A TRA<sub>F</sub> system arose from a common ancestor that spread among these three bacterial genera. Plasmids of this group share a “classical” F plasmid conformation. Transfer genes are organized in a long, polycistronic operon, where gene synteny is preserved. **Figure 3** shows the presence/absence table of *tra* genes in group A. It shows that we could detect the entire set of proteins deemed essential for F conjugation in a total of 150 plasmids. The essential genes, which were determined by transposon insertion analysis (Ippen-Ihler et al., 1972; Wu et al., 1987, 1988; Moore et al., 1990; Kathir and Ippen-Ihler, 1991; Maneewannakul et al., 1991, 1992; Maneewannakul and Ippen-Ihler, 1993), are shown in red in **Figure 3**. Remarkably, the regulatory components are also preserved among members of this group. TraM and TraJ appear in 100% of these plasmids. TraY homologs could be identified in 80% of the plasmids, but the real figure is probably higher, given the small size of the protein and lack of proper annotation. All MOB<sub>F12</sub> group A plasmids were *finO*+, the only exception being plasmid F itself. Since the *finO* phenotype in the F plasmid is due to an IS3 insertion, it is thus highly likely that the “original” F plasmid was also repressed.



**FIGURE 3 | Presence/Absence matrix of MPF<sub>F</sub> conjugation genes in Group A plasmids.** Matrix columns correspond to the 36 MPF<sub>F</sub> genes, while rows correspond to the 200 plasmids included in Group A. For each column/row combination, color green indicates the presence of the gene in the corresponding plasmid (PSI-Blast homolog identified with *E*-value below  $10^{-3}$ ) while blue indicates its absence. Plasmids were ordered according to overall conservation.



Within group A, a total of 50 plasmids lacked some essential component of the transfer machinery. Within this group, we identified a set of small plasmids (about 15 kb long) present in *Shigella* sp. and *E. coli* O104:H4 Shiga-toxin containing species (Figure 3, bottom). These plasmids seem to have suffered massive deletion of the TRA region, with only *traI*, *traX*, and *finO* genes remaining. This should result in a non-transmissible plasmid, since these plasmids neither contain the essential genes for pilus formation, nor the coupling protein TraD. In some cases, the relaxase itself appears truncated. The presence of plasmids with this particular structure among *Shigella* and Shiga-toxin containing *E. coli* is puzzling. The presumptive inability of these plasmids for horizontal mobilization would point out to the vertical propagation of a single deletion event in the ancestral line shared by *Shigella* sp. and *E. coli* O104:H4. However, members of this group do not form a monophyletic branch in the relaxase tree (Figure 1), which would suggest repeated but independent deletion events. Further research is needed to clarify the evolutionary history of these plasmids, the functional advantage of these deletions, if any, and their relationship to the pathogenesis of Shiga-toxin containing enterobacteria.

## GROUP B: MOB<sub>F12</sub> Plasmids from *Yersinia*. Prototype: pMT1

A second group of F-like plasmids comprises a set of plasmids from *Yersinia pestis* (Figure 4A). Their relaxases appear as a monophyletic branch in the MOB<sub>F12</sub> tree (Figure 1), showing an ancestral relationship to plasmids from group A (Figure 1). Structurally, they are characterized by a bipartite operon structure (Figure 4B), with genes involved in relaxosome formation (*traD*, *traI*) transcribed divergently from genes involved in conjugative pilus formation. Group B preserves all F essential genes and also *traP*, *traR*, *traR*, *trbI*, *traQ*, *trbB*, and *traX*. They all contain *finO*, yet none contain homologs of *traM*, *traJ*, or *traY*. Moreover, we could not find any putative transcriptional regulator in the vicinity of their conjugation regions, opening the questions of (a) how this mating system is regulated and (b) whether these plasmids are self-transmissible, given the lack of relaxase-accessory proteins or a recognizable origin of transfer. Outside the conjugation region, group B plasmids show extensive homology to plasmid pMT. pMT plasmids are a fundamental component of *Yersinia* pathogenesis, carrying essential virulence determinants for flea colonization (Hu et al., 1998). Besides, it is known that pMT plasmids from all three *Yersinia pestis* biovars (Antiqua, Medievalis, and Orientalis) are not self-transmissible and contain no transfer genes. Indeed, all *Yersinia* plasmids with a MOB<sub>F12</sub> conjugation systems belong to isolates of *Yersinia pestis pestoides*, an atypical *Y. pestis* group, probably the closest to the ancestral lineage that gave rise to the pandemic biovars (Garcia et al., 2007). Incorporation of pMT to *Y. pestis* has been traditionally linked to horizontal gene transfer from other enterobacterial species (Hu et al., 1998; Lindler et al., 1998). According to the phylogeny shown in Figure 1, MOB<sub>F12</sub> group B pMT plasmids stemmed from group A plasmids. Specifically, group B plasmids are monophyletic with two group A plasmids from *Klebsiella* sp. (pKOX\_NDM1 and pRJF866) that contain the

entire repertoire of essential F genes. However, given the lack of relaxase-accessory proteins (*traM* and *traY*), it is unclear whether group B plasmids are self-transmissible.

## GROUP C: MOB<sub>F12</sub> Plasmids Related to IncFV. Prototype: pUMNF18

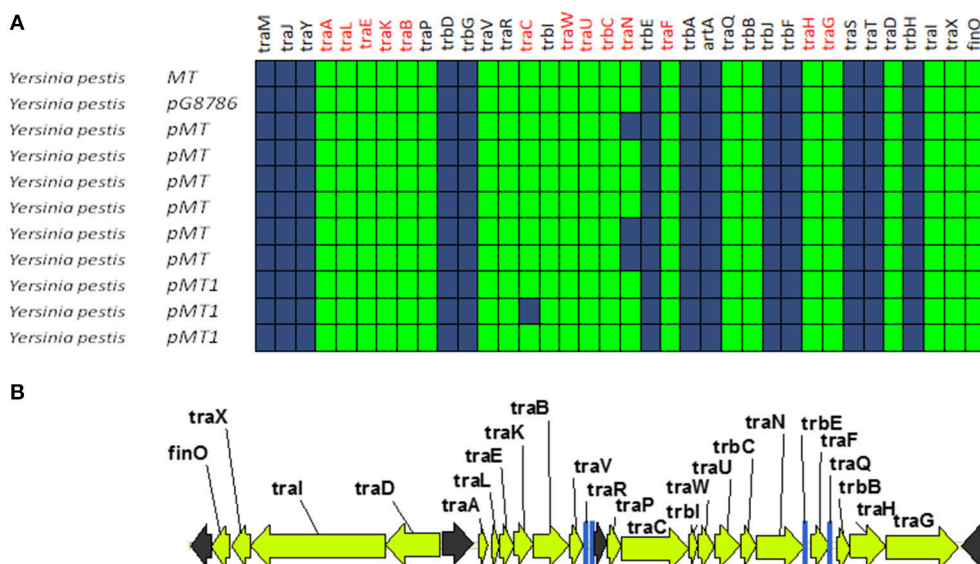
Although all plasmids from the MOB<sub>F12</sub> group share a common mating apparatus (MPF<sub>F</sub>), only some IncF plasmids contain the same replication and partition machineries. Thus, some IncF plasmids are able to co-reside together in the same cell, while others are not. Classical incompatibility testing identified several IncF subgroups, which were numbered from I to VII (de la Cruz et al., 1979). One of these IncF subgroups, IncFV, stood out because of its particular regulatory scheme. In our analysis, we identified a set of plasmids that display a typical IncFV arrangement (Figure 5). Plasmids from group C form a monophyletic branch in the MOB<sub>F12</sub> relaxase tree (Figure 1). They are characterized by a single promoter architecture. Conserved homologs include the same genes as in group A plasmids (Lu et al., 2002). The most conspicuous difference is that, although they encode a protein that is called TraJ, this protein actually shows no detectable homology to TraJ from Group A plasmids. Moreover, group C plasmids lack any recognizable homolog for FinO, yet experimental analysis showed that these plasmids are not de-repressed (Lu et al., 2002). To avoid confusion with the classical TraJ protein, hereafter we shall name this protein TraJ<sub>V</sub>.

According to structural analysis of TraJ<sub>V</sub> with Phyre2, this protein is predicted to contain a DNA/RNA-binding domain (DBD) formed by a 3-helical bundle fold in the 70 N-terminal amino acids. This type of DBD is found in transcriptional regulators such as LuxR/UhpA family. Most LuxR-type regulators act as transcriptional activators. They contain an HTH domain in the C-terminal part of the protein and an effector binding domain in the N-terminal domain. However, in TraJ<sub>V</sub> the HTH domain is located in the N-terminal half of the protein. A comparison between the predicted 3D structure for TraJ<sub>V</sub> N-terminal domain and the solved structure for F plasmid TraJ N-terminal domain (pdb 4KQD) (Lu et al., 2014) showed no structural homology (Figure 5C). In fact, TraJ is predicted to be similar to the canonical LuxR protein, with an N-terminal effector binding domain and a C-terminal DBD. Thus, TraJ<sub>V</sub> retains the DBD of TraJ, but the different position of this DBD and the differences in the rest of the protein suggest that TraJ<sub>V</sub> could play a different role than TraJ in the control of MOB<sub>F12</sub> group C plasmid conjugation.

## GROUP D: MOB<sub>F12</sub> Plasmids from *Enterobacter*. Prototype pENT01

A fourth group of F-like plasmids comprises members from another monophyletic branch in the MOB<sub>F12</sub> tree. This branch includes plasmids mainly coming from *Enterobacter* and the closely-related genus *Pantoea*, but also contains plasmids from other enterobacteria like *Erwinia*, *Rahnella*, and *Kluyvera* (Figure 6A). These plasmids maintain the classical group A genetic organization: a single operon including all *tra* genes,





**FIGURE 4 | Genetic structure and gene conservation in plasmids from Group B. (A)** Presence/Absence matrix of MPF<sub>F</sub> genes for plasmids included in group B. Color green indicates the presence of the gene in the plasmid (PSI-Blast homolog identified with *E*-value below  $10^{-3}$ ), while blue indicates its absence. **(B)** Genetic structure of plasmid pMT1, group B prototype. Yellow arrows and blue bars indicate ORFs corresponding to MPF<sub>F</sub> genes conserved in other groups. Black arrows indicate ORFs for genes without detectable homology to other IncF-like plasmids.

except *traM* and *traJ* regulators, (Figure 6B). Genes that were deemed essential for F transfer are also conserved among plasmids of group D, along with *traV*, *trbI*, *traN*, and *trbB*. Remarkably, group D plasmids are FinO and TraJ negative. However, the locus occupied by *traJ* in Group A contains a DNA-binding protein that is conserved within the group. We named this putative regulator EntFR.

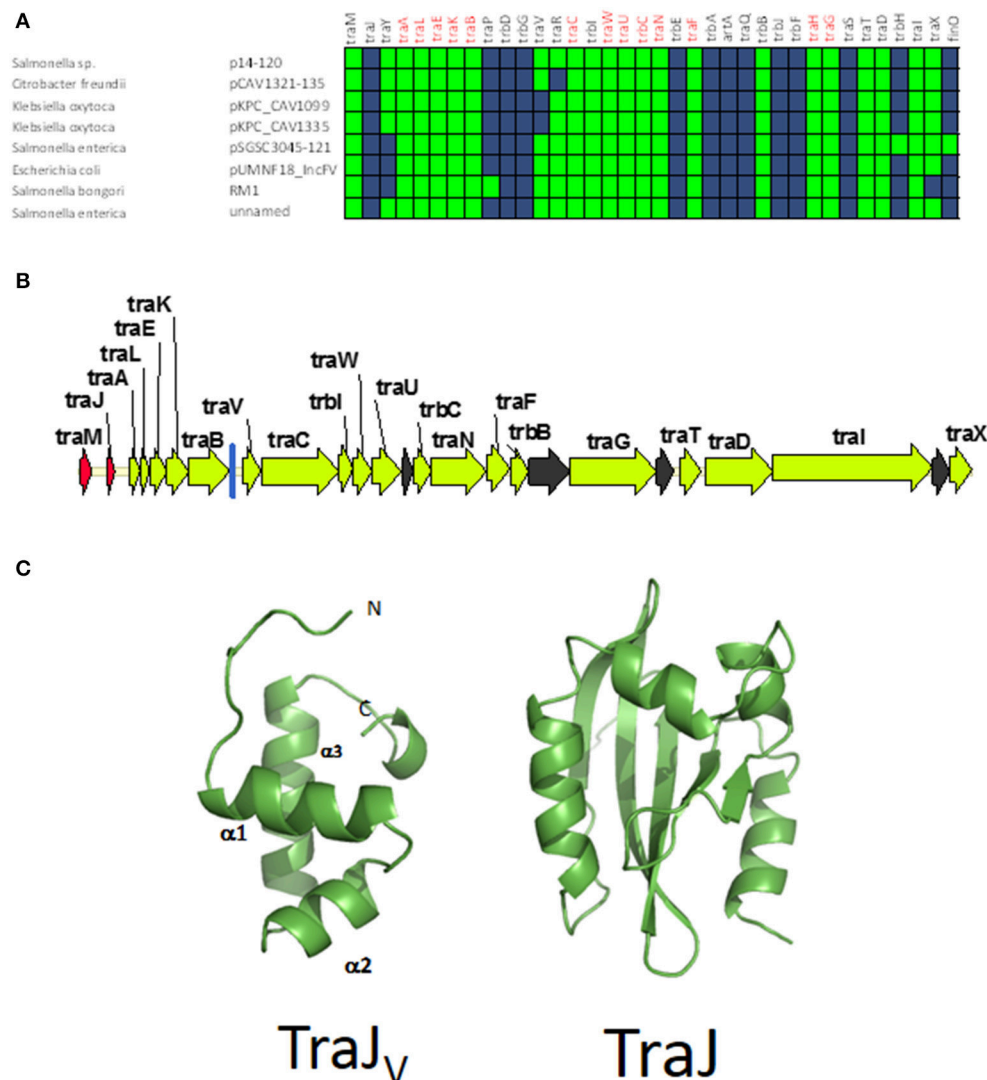
According to the Phyre2 prediction, EntFR is a ribbon-helix-helix (RHH) DBD (Figure 6C). Although this DBD is not as common as the helix-turn-helix (HTH) domain, it is frequently found in accessory proteins that bind to the origin of conjugative transfer (*TrwA* in plasmid R388, *TraJ* in plasmid RP4). In F plasmids *TraM* and *TraY* proteins present this fold. Structurally, EntFR is thus more similar to these proteins than to *TraJ*. EntFR contains two RHH domains, a feature that is also shared by *TraY*. Based on this, it is possible that EntFR acts as the functional homolog of *TraY* in group D plasmids. However, given that there are known RHH containing proteins able to act as transcriptional activators (Schreiter and Drennan, 2007), the possibility of EntFR being the functional homolog of *TraJ* cannot be ruled out.

## GROUP E: MOB<sub>F12</sub> Plasmids from *Sphingomonas*. Prototype pCAR3

A fifth group of F-like plasmids includes plasmids from the genus *Sphingomonas*, the relaxases of which form a monophyletic branch in the MOB<sub>F12</sub> tree (Figure 1). These plasmids exhibit a conserved architecture that is different from the arrangement in the other MOB<sub>F12</sub> groups (Figure 7). Instead of a long, single operon, *tra* genes from group E plasmids are split in two convergent operons. The first contains the genes responsible

for the formation of the conjugative pilus (MPF<sub>F</sub> genes), while the second includes the genes involved in relaxosome formation (Figure 7B). This architecture is reminiscent of that of conjugative plasmids with VirB-like mating apparatus, such as the MOB<sub>F11</sub> (IncW and IncN) groups. However, although the architecture of group E is VirB-like, all the constituent genes are related to those of plasmid F, i.e., they belong to the MPF<sub>F</sub> family. Genes deemed essential for F conjugation are preserved in group E plasmids, although *traA*, the gene coding for the conjugative pilin, is significantly different. Apart from these essential genes, group E plasmids also maintain clear homologs of *trbI* and *traN*. Regarding the regulatory components of the transfer machinery, group E plasmids lack homologs of *TraM*, *TraJ*, *TraY*, or *FinO*. The only putative regulator that can be identified by BLAST analysis is a small, conserved protein present immediately upstream *traD*. In plasmids with MPF<sub>T</sub> (VirB-like) conjugation systems, this position is usually occupied by a protein coding for a ribbon helix-helix relaxase-accessory protein (e.g., *TrwA* in plasmid R388) (Moncalián and de la Cruz, 2004; Varsaki et al., 2009). Relaxase-accessory proteins participate in relaxosome assembly as well as in regulation of the expression of relaxosome components. In group E plasmids, this small protein can be thus considered as the hallmark of the group; it will be named SphTR (for *Sphingomonas* transfer regulator). Structural modeling of SphTR using Phyre2 showed that this protein belongs to the ribbon helix helix (RHH) superfamily of DNA binding proteins. Thus, it is likely that SphTR fulfills the mobilization-accessory role in this group of plasmids.

Plasmids containing MOB<sub>F11</sub> relaxases, like IncN and IncW plasmids, show an operon structure similar to group E plasmids. However, while in IncN and IncW plasmids it is possible to



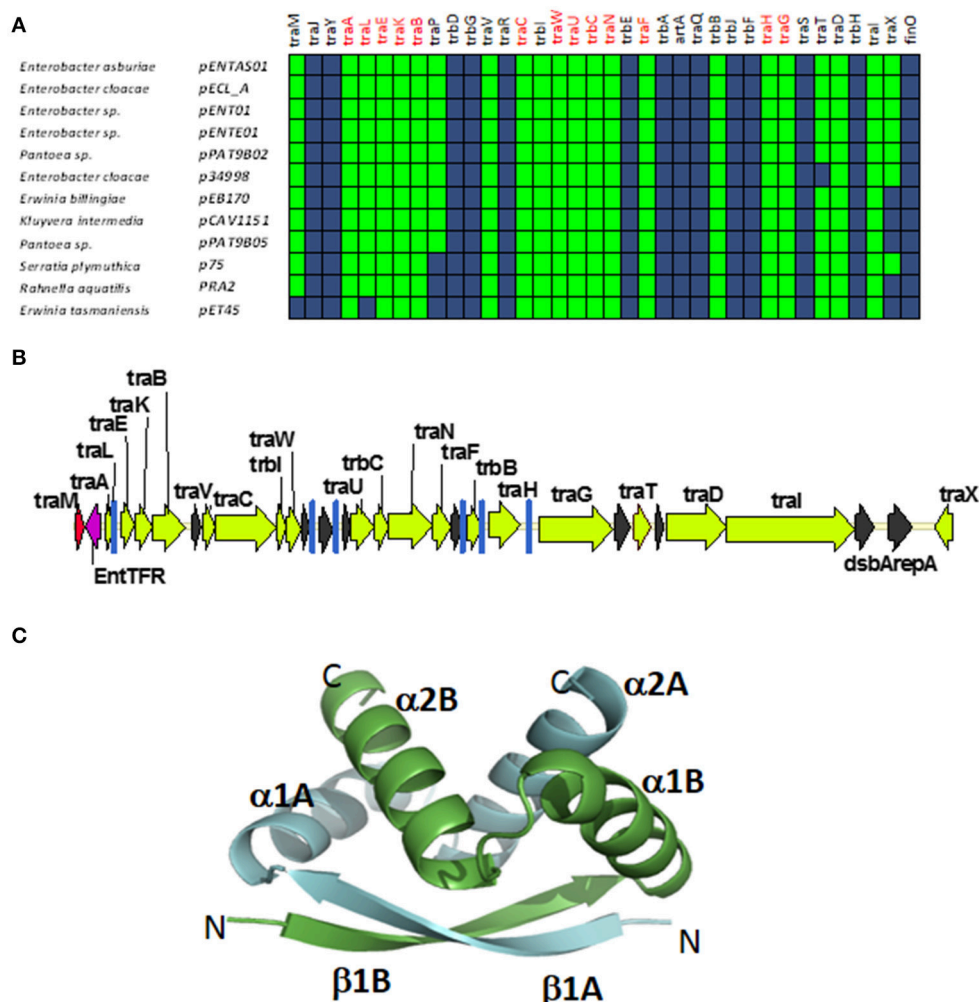
**FIGURE 5 | Genetic structure and gene conservation in plasmids from Group C. (A)** Presence/Absence matrix of MPF<sub>F</sub> genes for plasmids included in group C. Color green indicates the presence of the gene in the plasmid (PSI-Blast homolog identified with  $E$ -value below  $10^{-3}$ ), while blue indicates its absence. **(B)** Genetic structure of plasmid pUMNF18, group C prototype. Yellow arrows and blue bars indicate ORFs corresponding to MPF<sub>F</sub> genes conserved in other groups. Black arrows indicate ORFs for genes without detectable homology to other IncF-like plasmids. Red arrows indicate putative transcriptional regulators. **(C)** Structural prediction of the N-terminal domain of TraJ<sub>V</sub> using the Phyre2 server (left) compared to the 3D structure of TraJ (right).

identify the regulators responsible for the independent control of the two operons (Fernandez-Lopez et al., 2014), we were unable to find any other putative DNA binding proteins in the vicinity of Group E transfer genes. Thus, it is unclear whether these genes are controlled by other plasmid/host regulators, by SphTR, or are not transcriptionally regulated. In any case, Group E plasmids constitute a valuable divergent evolutionary line of MOB<sub>F12</sub> plasmids that, at least in genome organization and regulatory components, represents a bridge between MOB<sub>F11</sub> and MOB<sub>F12</sub>.

### Other MOB<sub>F12</sub> Plasmids outside *Enterobacteriaceae*

A set of 14 plasmids remain unassigned in our group classification. They correspond to MOB<sub>F12</sub> plasmids that

were found outside the *Enterobacteriaceae*. Members of this group include plasmids from *Vibrio*, *Aeromonas*, *Legionella*, *Fluoribacter*, and *Piscirickettsia*. Although many of them contain an entire set of *tra* genes (like for example pLELO-like plasmids from *Legionella*), their genetic organization and putative regulators (indicated in Table S2) do not seem to be shared by the different plasmids (nor with Group E plasmids from *Sphingomonas*). It is entirely possible that this is due to under-representation of these genera in the nucleotide databases, compared to clinically-relevant enterobacteria. Anyhow, these plasmids serve to demonstrate that the IncF conjugation system is not restricted to the enterobacteria and that the MOB<sub>F12</sub> plasmid clade can assimilate a number alternative of regulatory proteins.



**FIGURE 6 | Genetic structure and gene conservation in plasmids from Group D. (A)** Presence/Absence matrix of MPF<sub>F</sub> genes for plasmids included in group D. Color green indicates the presence of the gene in the plasmid (PSI-Blast homolog identified with  $E$ -value below  $10^{-3}$ ), while blue indicates its absence. **(B)** Genetic structure of plasmid pENT01, prototype of group D plasmids. Yellow arrows and blue bars indicate ORFs corresponding to MPF<sub>F</sub> genes conserved in other groups. Black arrows indicate ORFs for genes without detectable homology among other IncF-like plasmids. Red arrows indicate putative transcriptional regulators. **(C)** Structural prediction of the putative regulator EntTFR using Phyre2. The double RHH configuration is highlighted, with one RHH indicated in green (comprising the beta strand  $\beta 1A$  and alpha-helices  $\alpha 1A$  and  $\alpha 2A$ ), while the second RHH motif is indicated in blue (comprising the beta strand  $\beta 1B$  and alpha-helices  $\alpha 1B$  and  $\alpha 2B$ ).

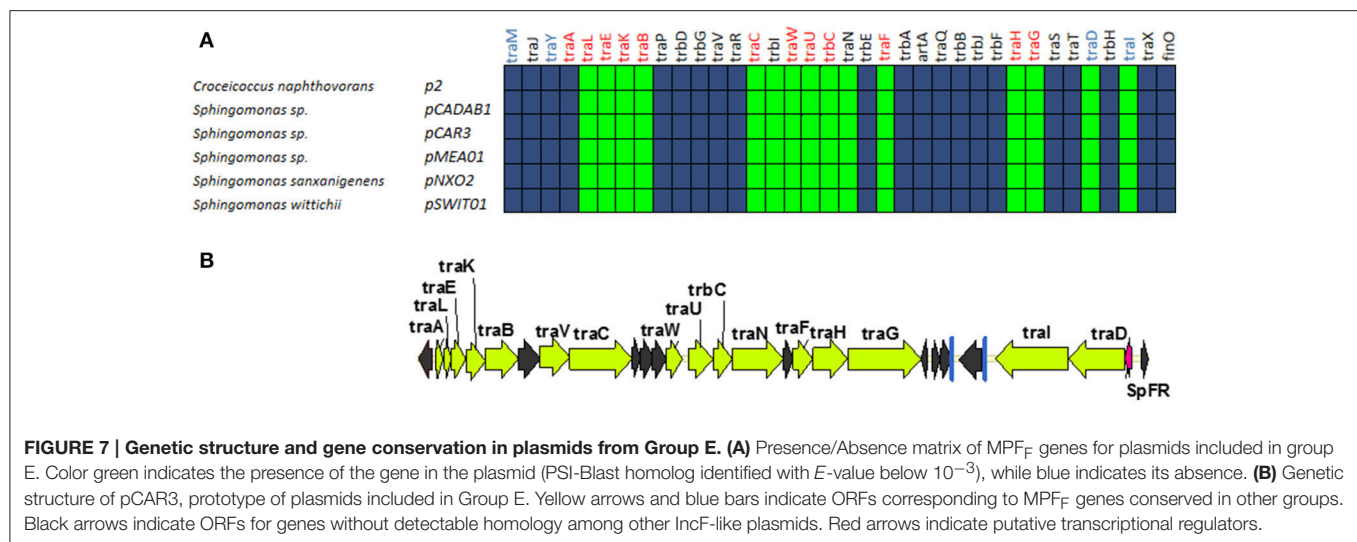
## Comparison with Other IncF Typing Systems: within Group Diversity

IncF plasmids are frequent carriers of antibiotic resistance genes and virulence factors, and a common finding in clinically relevant enterobacteria. Clinical microbiologists differentiate IncF plasmids using a sequence-typing system that takes advantage of the allelic diversity that IncF plasmids present in their replication regions (Villa et al., 2010). Analyzing replicon variants, Villa et al. were able to differentiate several IncF groups, according to their replicon sequence type (RST) (Villa et al., 2010). A comparison between replicon typing and the analysis of conjugation regions, revealed that plasmids with an assignable IncF RST belonged to Group A and Group B IncF plasmids, two groups that are monophyletic in the relaxase phylogenetic tree (Figure 1, black arrow). Group A plasmids, which include all

classical IncF plasmids, present different RST profiles (Villa et al., 2010), indicating that within this broad group there is substantial sequence variation. Plasmids from groups C, D and E, however, cannot be assigned a typical RST profile, indicating that IncF plasmids in these groups are likely to exhibit different replication mechanisms.

## DISCUSSION

The F plasmid was the first example of a conjugative plasmid found in bacteria (Lederberg and Tatum, 1946). IncF plasmids were also among the first plasmids known to provide antibiotic resistances (Meynell and Datta, 1966; Meynell et al., 1968b), colicins (Ozeki et al., 1962), and virulence determinants (Rotger and Casadesús, 1999). Because this historical relevance and



their frequent association to clinically-relevant enterobacteria, F-like plasmids occupy a prominent place among bacterial plasmids. In order to analyze their conservation and diversity, we studied 256 plasmids that contained a MOB<sub>F12</sub> relaxase. Analysis of these plasmids revealed that MOB<sub>F12</sub> relaxases are associated exclusively with MPF<sub>F</sub> conjugation systems. However, the MOB<sub>F12</sub> conjugation systems analyzed presented a wider diversity than anticipated. Using the regulatory proteins as the most conspicuous indicators of this diversity, we could identify five major groups of IncF-like plasmids. As shown in **Figure 1**, these groups correspond to different branches of the relaxase phylogenetic tree. This indicates that these groups represent different radiations in the common branch of F-like plasmids. Interestingly, we also found a strong correlation between the MOB<sub>F12</sub> phylogenetic groups and their bacterial hosts, suggesting that these groups might represent adaptations to different host genetic backgrounds. Group A was the most populated group, and included “classical” F-like plasmids like F, R1, pSLT and R100. Plasmids from this group are restricted to enterobacteria, with *E. coli*, *Klebsiella*, and *Salmonella* as the most frequent hosts. The overpopulation of this group compared to others, however, should not be taken as an indicator of particular evolutionary success. *E. coli*, *Klebsiella*, and *Salmonella* are clinically-relevant pathogens, much more represented in the genome databases than other species. Thus, the abundance of plasmids from group A could be just an indicator of sequencing bias. Plasmids from this group have been studied for decades, yet it yielded some surprising facts. First of all, it indicated how rare the F plasmid is. One of the motivations of this study was to determine whether de-repressed F-like plasmids were often found in clinical and environmental samples. Our analysis showed that de-repression by FinO inactivation is a property exclusive to the F-plasmid itself. Other genome alterations, particularly deletions, however, were far more common. At least 25% of the plasmids from group A lacked some of the genes deemed essential for F conjugation. This indicates that MOB<sub>F12</sub> plasmids suffer frequent insertions and deletions, and that the presence of certain genes (such as the

MOB<sub>F12</sub> relaxase) cannot be taken as a guarantee that the plasmid is going to be self-transmissible.

Our results also revealed that some species are more prone to delete genes from the IncF conjugation region. The phylogenetic tree of **Figure 1** indicates that these deletions can come from a single event, such as the monophyletic plasmids from Group B in *Yersinia pestis*. In *Shigella* sp. and *E. coli* O104 even more radical deletions have occurred multiple times along the course of evolution. Importantly, all plasmids from these species showed plasmids with major deletions, indicating a strong selective pressure against MOB<sub>F12</sub>-conjugation genes.

Group A plasmids were the most common and the only ones to show a clear-cut fertility inhibition system (as judged from the presence of *traJ/finO* genes). Since group A is monophyletic, fertility inhibition was an innovation incorporated in some enterobacterial plasmid that then invaded *Escherichia*, *Salmonella*, and *Klebsiella* species. Although we cannot compare Group A abundance to groups outside the enterobacteria due to probable sequencing bias, it is interesting to note that there is another plasmid group which is exclusively found in enterobacteria, but much less populated. Group C plasmids, which include plasmids similar to classical IncFV plasmids, is restricted to the same species as Group A. This means that the sequencing bias between these two groups is less pronounced, yet Group A plasmids are much more abundant than Group C plasmids. Both plasmid groups share a common genetic structure, and their main difference is the presence of the fertility inhibition system. Thus, it is possible that the incorporation of the fertility inhibition system enhanced the ability of Group A plasmids to spread among enterobacterial species.

Groups D and E represent adaptations of the F conjugation machinery to other bacterial clades. Interestingly, the greater the phylogenetic distance between the hosts, the higher the level of divergence between IncF/MOB<sub>F12</sub> plasmids. Thus, plasmids belonging to Group D are typically found in *Enterobacter* sp., a member of the *Enterobacteriaceae*, and their main difference



with Group A plasmids is the presence of a different regulatory scheme, with EntFR likely playing the role of TraY. Meanwhile, Group E plasmids are present in *Sphingomonas*, an alpha-proteobacteria, and plasmids from this group present not only a different regulatory scheme, but also a different operon structure. Because their bipartite operon structure and the presence of a RHH protein in the same operon as the relaxase and the coupling protein, plasmids from this group resemble plasmids with MOB<sub>F11</sub> relaxases, such as IncN and IncW plasmid groups (Fernández-López, et al., 2006). Judging from the relaxase phylogenetic tree, members of this group are also the closest phylogenetically to MOB<sub>F11</sub> relaxases. This indicates that Group E plasmids represent an interesting intermediate link between VirB-like pilus containing plasmids and F-like pilus containing plasmids.

In summary, Groups A to E represent five alternate configurations for F-like plasmids. All these configurations present a shared protein core, which includes the 13 pilus genes deemed essential for plasmid F conjugation by the seminal work of Karin Ippen-Ihler (Ippen-Ihler et al., 1972; Maneewannakul et al., 1987, 1991, 1992, 1995; Wu et al., 1987, 1988; Moore et al., 1990; Kathir and Ippen-Ihler, 1991; Maneewannakul and Ippen-Ihler, 1993; Frost et al., 1994). Thus, while the mechanism of transfer is probably conserved, different regulatory mechanisms and operon structures exist in nature. It is also likely that other proteins not included in the *tra* operon, but known to play a role in conjugation like VirB1-like lytic transglycosylases show similar variation (Zahrl et al., 2005). This diversity has been found focusing exclusively in the conjugation region. However, it is known that plasmids belonging to group A present different replication and partition systems (Ogura and Hiraga, 1983; Gerdes and Molin, 1986; Villa et al., 2010). Thus, the exploration of the entire diversity of IncF plasmids requires analysis of other plasmid regions apart from the conjugation machinery. In particular, the diversity of replication strategies found among

these plasmids is worthy of further analysis (Osborn et al., 2000; Villa et al., 2010). IncF plasmids are the most abundant plasmid type found in enterobacteria (de Toro et al., 2014). They are key to the generation and spread of clonal groups like *E. coli* ST-131 (Lanza et al., 2014; Johnson et al., 2016), and they are fundamental vehicles for the spread of antibiotic resistances (de Been et al., 2014). We hope that analysis using comparative genomics, like the one presented here, will help unraveling the causes behind the prevalence and evolutionary success of IncF plasmids.

## AUTHOR CONTRIBUTIONS

RF-L, MdT, GM, MPG-B, and FdC retrieved, analyzed the data, and wrote the paper.

## FUNDING

The work performed by the FdC group was supported by grants BFU2014-55534-C2-1-P and BFU2014-62190-EXP from the Spanish Ministry of Economy and Competitiveness and 612146/FP7-ICT-2013-10 from the European Seventh Framework Programme.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmolb.2016.00071/full#supplementary-material>

**Table S1 | (A)** List of MOB<sub>F12</sub> relaxases used as bait in BLAST searches. **(B)** List of plasmids retrieved from the NCBI plasmid database. Columns D to AM indicate the NCBI reference number of the protein homolog for each of the F conjugation genes indicated in column headers, as explained in Materials and Methods.

**Table S2 | Putative alternative regulators of MOB<sub>F12</sub> transfer genes in groups C, D, E and plasmids not assigned to any group.**

## REFERENCES

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Alvarado, A., Garcillán-Barcia, M. P., and de la Cruz, F. (2012). A degenerate primer MOB typing (DPMT) method to classify gamma-proteobacterial plasmids in clinical and environmental settings. *PLoS ONE* 7:e40438. doi: 10.1371/journal.pone.0040438
- Arthur, D. C., Ghetu, A. F., Gubbins, M. J., Edwards, R. A., Frost, L. S., and Glover, J. N. (2003). FinO is an RNA chaperone that facilitates sense-antisense RNA interactions. *EMBO J.* 22, 6346–6355. doi: 10.1093/emboj/cdg607
- Arutyunov, D., and Frost, L. S. (2013). F conjugation: back to the beginning. *Plasmid* 70, 18–32. doi: 10.1016/j.plasmid.2013.03.010
- Brinton, C. C. Jr., Genski, P. Jr., and Carnahan, J. (1964). A new type of bacterial pilus genetically controlled by the fertility factor of *E. coli* K 12 and its role in chromosome transfer. *Proc. Natl. Acad. Sci. U.S.A.* 52, 776–783. doi: 10.1073/pnas.52.3.776
- Caro, L. G., and Schnös, M. (1966). The attachment of the male-specific bacteriophage F1 to sensitive strains of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 56, 126–132. doi: 10.1073/pnas.56.1.126
- Chandran Darbari, V., and Waksman, G. (2015). structural biology of bacterial type IV secretion systems. *Annu. Rev. Biochem.* 84, 603–629. doi: 10.1146/annurev-biochem-062911-102821
- Clarke, M., Maddera, L., Harris, R. L., and Silverman, P. M. (2008). F-pili dynamics by live-cell imaging. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17978–17981. doi: 10.1073/pnas.0806786105
- de Been, M., Lanza, V. F., de Toro, M., Scharringa, J., Dohmen, W., Du, Y., et al. (2014). Dissemination of cephalosporin resistance genes between *Escherichia coli* strains from farm animals and humans by specific plasmid lineages. *PLoS Genet.* 10:e1004776. doi: 10.1371/journal.pgen.1004776
- de la Cruz, F., Zabala, J. C., and Ortiz, J. M. (1979). Incompatibility among alpha-hemolytic plasmids studied after inactivation of the alpha-hemolysin gene by transposition of Tn802. *Plasmid* 2, 507–519. doi: 10.1016/0147-619X(79)90050-7
- Dennison, S. (1972). Naturally occurring R factor, derepressed for pilus synthesis, belonging to the same compatibility group as the sex factor F of *Escherichia coli* K-12. *J. Bacteriol.* 109, 416–422.
- de Toro, M., Garcillán-Barcia, M. P., and De La Cruz, F. (2014). plasmid diversity and adaptation analyzed by massive sequencing of *Escherichia coli* plasmids. *Microbiology Spectr.* 2:PLAS-0031-2014. doi: 10.1128/microbiolspec.plas-0031-2014

- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Fernandez-Lopez, R., Del Campo, I., Revilla, C., Cuevas, A., and de la Cruz, F. (2014). Negative feedback and transcriptional overshooting in a regulatory network for horizontal gene transfer. *PLoS Genet.* 10:e1004171. doi: 10.1371/journal.pgen.1004171
- Fernández-López, R., Garcillán-Barcia, M. P., Revilla, C., Lázaro, M., Vielva, L., and de la Cruz, F. (2006). Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution. *FEMS Microbiol. Rev.* 30, 942–966. doi: 10.1111/j.1574-6976.2006.00042.x
- Finnegan, D., and Willetts, N. (1973). The site of action of the F transfer inhibitor. *Mol. Gen. Genet.* 127, 307–316. doi: 10.1007/BF00267101
- Frost, L. S., Ippen-Ihler, K., and Skurray, R. A. (1994). Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev.* 58, 162–210.
- Frost, L. S., and Koraimann, G. (2010). Regulation of bacterial conjugation: balancing opportunity with adversity. *Future Microbiol.* 5, 1057–1071. doi: 10.2217/fmb.10.70
- Garcia, E., Worsham, P., Bearden, S., Malfatti, S., Lang, D., Larimer, F., et al. (2007). Pestoides F, an atypical *Yersinia pestis* strain from the former Soviet Union. *Adv. Exp. Med. Biol.* 603, 17–22. doi: 10.1007/978-0-387-72124-8\_2
- Garcillán-Barcia, M. P., Alvarado, A., and de la Cruz, F. (2011). Identification of bacterial plasmids based on mobility and plasmid population biology. *FEMS Microbiol. Rev.* 35, 936–956. doi: 10.1111/j.1574-6976.2011.00291.x
- Garcillán-Barcia, M. P., Francia, M. V., and de la Cruz, F. (2009). The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol. Rev.* 33, 657–687. doi: 10.1111/j.1574-6976.2009.00168.x
- Gerdes, K., and Molin, S. (1986). Partitioning of plasmid R1. Structural and functional analysis of the parA locus. *J. Mol. Biol.* 190, 269–279. doi: 10.1016/0022-2836(86)90001-X
- Ghetu, A. F., Gubbins, M. J., Frost, L. S., and Glover, J. N. (2000). Crystal structure of the bacterial conjugation repressor finO. *Nat. Struct. Biol.* 7, 565–569. doi: 10.1038/76790
- Helmuth, R., and Achtman, M. (1975). Operon structure of DNA transfer cistrons on the F sex factor. *Nature* 257, 652–656. doi: 10.1038/257652a0
- Hu, P., Elliott, J., McCready, P., Skowronski, E., Garnes, J., Kobayashi, A., et al. (1998). Structural organization of virulence-associated plasmids of *Yersinia pestis*. *J. Bacteriol.* 180, 5192–5202.
- Ippen-Ihler, K., Achtman, M., and Willetts, N. (1972). Deletion map of the *Escherichia coli* K-12 sex factor F: the order of eleven transfer cistrons. *J. Bacteriol.* 110, 857–863.
- Ishibashi, M. (1967). F pilus as f+ antigen. *J. Bacteriol.* 93, 379–389.
- Johnson, T. J., Danzeisen, J. L., Youmans, B., Case, K., Llop, K., Munoz-Aguayo, J., et al. (2016). Separate F-type plasmids have shaped the evolution of the H30 subclone of *Escherichia coli* sequence type 131. *mSphere* 1:e00121-16. doi: 10.1128/mSphere.00121-16
- Kathir, P., and Ippen-Ihler, K. (1991). Construction and characterization of derivatives carrying insertion mutations in F plasmid transfer region genes, trbA, artA, traQ, and trbB. *Plasmid* 26, 40–54. doi: 10.1016/0147-619X(91)90035-U
- Kelley, L. A., and Sternberg, M. J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* 4, 363–371. doi: 10.1038/nprot.2009.2
- Kennedy, N., Beutin, L., Achtman, M., Skurray, R., Rahmsdorf, U., and Herrlich, P. (1977). Conjugation proteins encoded by the F sex factor. *Nature* 270, 580–585. doi: 10.1038/270580a0
- Krause, S., Pansegrau, W., Lurz, R., de la Cruz, F., and Lanka, E. (2000). Enzymology of type IV macromolecule secretion systems: the conjugative transfer regions of plasmids RP4 and R388 and the cag pathogenicity island of *Helicobacter pylori* encode structurally and functionally related nucleoside triphosphate hydrolases. *J. Bacteriol.* 182, 2761–2770. doi: 10.1128/JB.182.10.2761-2770.2000
- Lang, S., Gruber, C. J., Raffl, S., Reisner, A., and Zechner, E. L. (2014). Common requirement for the relaxosome of plasmid R1 in multiple activities of the conjugative type IV secretion system. *J. Bacteriol.* 196, 2108–2121. doi: 10.1128/JB.00045-13
- Lanza, V. F., de Toro, M., Garcillán-Barcia, M. P., Mora, A., Blanco, J., Coque, T. M., et al. (2014). Plasmid flux in *Escherichia coli* ST131 sublineages, analyzed by plasmid constellation network (PLACNET), a new method for plasmid reconstruction from whole genome sequences. *PLoS Genet.* 10:e1004766. doi: 10.1371/journal.pgen.1004766
- Lawley, T. D., Klimke, W. A., Gubbins, M. J., and Frost, L. S. (2003). F factor conjugation is a true type IV secretion system. *FEMS Microbiol. Lett.* 224, 1–15. doi: 10.1016/S0378-1097(03)00430-0
- Lederberg, J., and Tatum, E. L. (1946). Gene recombination in *Escherichia coli*. *Nature* 158, 558. doi: 10.1038/158558a0
- Lindler, L. E., Plano, G. V., Burland, V., Mayhew, G. F., and Blattner, F. R. (1998). Complete DNA sequence and detailed analysis of the *Yersinia pestis* KIM5 plasmid encoding murine toxin and capsular antigen. *Infect. Immun.* 66, 5731–5742.
- Lu, J., Manchak, J., Klimke, W., Davidson, C., Firth, N., Skurray, R. A., et al. (2002). Analysis and characterization of the IncFV plasmid pED208 transfer region. *Plasmid* 48, 24–37. doi: 10.1016/S0147-619X(02)00007-0
- Lu, J., Wu, R., Adkins, J. N., Joachimiak, A., and Glover, J. N. (2014). Crystal structures of the F and pSLT plasmid TraJ N-terminal regions reveal similar homodimeric PAS folds with functional interchangeability. *Biochemistry* 53, 5810–5819. doi: 10.1021/bi500244m
- Makela, P. H., Lederberg, J., and Lederberg, E. M. (1962). Patterns of sexual recombination in enteric bacteria. *Genetics* 47, 1427–1439.
- Maneewannakul, K., and Ippen-Ihler, K. (1993). Construction and analysis of F plasmid traR, trbJ, and trbH mutants. *J. Bacteriol.* 175, 1528–1531.
- Maneewannakul, K., Maneewannakul, S., and Ippen-Ihler, K. (1995). Characterization of traX, the F plasmid locus required for acetylation of F-pilin subunits. *J. Bacteriol.* 177, 2957–2964.
- Maneewannakul, S., Kathir, P., and Ippen-Ihler, K. (1992). Characterization of the F plasmid mating aggregation gene traN and of a new F transfer region locus trbE. *J. Mol. Biol.* 225, 299–311. doi: 10.1016/0022-2836(92)90923-8
- Maneewannakul, S., Kathir, P., Moore, D., Le, L. A., Wu, J. H., and Ippen-Ihler, K. (1987). Location of F plasmid transfer operon genes traC and traW and identification of the traW product. *J. Bacteriol.* 169, 5119–5124.
- Maneewannakul, S., Maneewannakul, K., and Ippen-Ihler, K. (1991). Characterization of trbC, a new F plasmid tra operon gene that is essential to conjugative transfer. *J. Bacteriol.* 173, 3872–3878.
- Mark Glover, J. N., Chaulk, S. G., Edwards, R. A., Arthur, D., Lu, J., and Frost, L. S. (2015). The FinO family of bacterial RNA chaperones. *Plasmid* 78, 79–87. doi: 10.1016/j.plasmid.2014.07.003
- Meynell, E., and Datta, N. (1966). The relation of resistance transfer factors to the F-factor (sex-factor) of *Escherichia coli* K12. *Genet. Res.* 7, 134–140. doi: 10.1017/S0016672300009538
- Meynell, E., Datta, N., Lawn, A. M., and Meynell, G. G. (1968a). Ancestral relationships of transmissible bacterial plasmids. *J. Gen. Microbiol.* 50(Suppl. v).
- Meynell, E., Meynell, G. G., and Datta, N. (1968b). Phylogenetic relationships of drug-resistance factors and other transmissible bacterial plasmids. *Bacteriol. Rev.* 32, 55–83.
- Moncalián, G., and de la Cruz, F. (2004). DNA binding properties of protein TrwA, a possible structural variant of the Arc repressor superfamily. *Biochim. Biophys. Acta* 1701, 15–23. doi: 10.1016/j.bbapap.2004.05.009
- Moore, D., Maneewannakul, K., Maneewannakul, S., Wu, J. H., Ippen-Ihler, K., and Bradley, D. E. (1990). Characterization of the F-plasmid conjugative transfer gene traU. *J. Bacteriol.* 172, 4263–4270.
- Nakaya, R., Nakamura, A., and Murata, Y. (1960). Resistance transfer agents in *Shigella*. *Biochem. Biophys. Res. Commun.* 3, 654–659. doi: 10.1016/0006-291X(60)90081-4
- Ogura, T., and Hiraga, S. (1983). Partition mechanism of F plasmid: two plasmid gene-encoded products and a cis-acting region are involved in partition. *Cell* 32, 351–360. doi: 10.1016/0092-8674(83)90454-3
- Orskov, I., and Orskov, F. (1960). An antigen termed f-plus occurring in F-plus *E. coli* strains. *Acta Pathol. Microbiol. Scand.* 48, 37–46. doi: 10.1111/j.1699-0463.1960.tb04734.x
- Osborn, A. M., da Silva Tatley, F. M., Steyn, L. M., Pickup, R. W., and Saunders, J. R. (2000). Mosaic plasmids and mosaic replicons: evolutionary lessons from the analysis of genetic diversity in IncFII-related replicons. *Microbiology* 146 (Pt 9), 2267–2275. doi: 10.1099/00221287-146-9-2267
- Ozeki, H., Stocker, B. A., and Smith, S. M. (1962). Transmission of colicinogeny between strains of *Salmonella typhimurium* grown together. *J. Gen. Microbiol.* 28, 671–687. doi: 10.1099/00221287-28-4-671
- Rotger, R., and Casadesús, J. (1999). The virulence plasmids of *Salmonella*. *Int. Microbiol.* 2, 177–184.

- Schreiter, E. R., and Drennan, C. L. (2007). Ribbon-helix-helix transcription factors: variations on a theme. *Nat. Rev. Microbiol.* 5, 710–720. doi: 10.1038/nrmicro1717
- Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P., and de la Cruz, F. (2010). Mobility of plasmids. *Microbiol. Mol. Biol. Rev.* 74, 434–452. doi: 10.1128/MMBR.00020-10
- Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi: 10.1093/bioinformatics/btl446
- Timmis, K. N., Andrés, I., and Achtman, M. (1978). Fertility repression of F-like conjugative plasmids: physical mapping of the R6–5 finO and finP cistrons and identification of the finO protein. *Proc. Natl. Acad. Sci. U.S.A.* 75, 5836–5840. doi: 10.1073/pnas.75.12.5836
- Varsaki, A., Moncalian, G., Garcillan-Barcia Mdel, P., Drainas, C., and de la Cruz, F. (2009). Analysis of ColE1 MbeC unveils an extended ribbon-helix-helix family of nicking accessory proteins. *J. Bacteriol.* 191, 1446–1455. doi: 10.1128/JB.01342-08
- Villa, L., García-Fernández, A., Fortini, D., and Carattoli, A. (2010). Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. *J. Antimicrob. Chemother.* 65, 2518–2529. doi: 10.1093/jac/dkq347
- Watanabe, T. (1967). Evolutionary relationships of R factors with other episomes and plasmids. *Fed. Proc.* 26, 23–28.
- Womble, D. D., and Rownd, R. H. (1988). Genetic and physical map of plasmid NR1: comparison with other IncFII antibiotic resistance plasmids. *Microbiol. Rev.* 52, 433–451.
- Wong, J. J., Lu, J., and Glover, J. N. (2012). Relaxosome function and conjugation regulation in F-like plasmids - a structural biology perspective. *Mol. Microbiol.* 85, 602–617. doi: 10.1111/j.1365-2958.2012.08131.x
- Wu, J. H., Kathir, P., and Ippen-Ihler, K. (1988). The product of the F plasmid transfer operon gene, traF, is a periplasmic protein. *J. Bacteriol.* 170, 3633–3639.
- Wu, J. H., Moore, D., Lee, T., and Ippen-Ihler, K. (1987). Analysis of *Escherichia coli* K12 F factor transfer genes: traQ, trbA, and trbB. *Plasmid* 18, 54–69. doi: 10.1016/0147-619X(87)90078-3
- Yoshioka, Y., Ohtsubo, H., and Ohtsubo, E. (1987). Repressor gene finO in plasmids R100 and F: constitutive transfer of plasmid F is caused by insertion of IS3 into F finO. *J. Bacteriol.* 169, 619–623.
- Zahl, D., Wagner, M., Bischof, K., Bayer, M., Zavec, B., Beranek, A., et al. (2005). Peptidoglycan degradation by specialized lytic transglycosylases associated with type III and type IV secretion systems. *Microbiology* 151, 3455–3467. doi: 10.1099/mic.0.28141-0

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Fernandez-Lopez, de Toro, Moncalian, Garcillan-Barcia and de la Cruz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Stabilization of the Virulence Plasmid pSLT of *Salmonella* Typhimurium by Three Maintenance Systems and Its Evaluation by Using a New Stability Test

Damián Lobato-Márquez<sup>1\*</sup>, Laura Molina-García<sup>2†</sup>, Inma Moreno-Córdoba<sup>3†</sup>, Francisco García-del Portillo<sup>4</sup> and Ramón Díaz-Orejas<sup>3</sup>

<sup>1</sup> Section of Microbiology, Department of Medicine, Centre for Molecular Bacteriology and Infection, Imperial College London, London, UK, <sup>2</sup> Department of Cell and Developmental Biology, University College London, London, UK, <sup>3</sup> Departamento de Microbiología Molecular y Biología de las Infecciones, Centro de Investigaciones Biológicas-Spanish National Research Council, Madrid, Spain, <sup>4</sup> Departamento de Biotecnología Microbiana, Centro Nacional de Biotecnología-Spanish National Research Council, Madrid, Spain

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas Medical Branch,  
USA

### Reviewed by:

Gloria Del Solar,  
Spanish National Research Council,  
Spain  
Virtu Solano-Collado,  
University of Aberdeen, UK

### \*Correspondence:

Damián Lobato-Márquez  
d.marquez@imperial.ac.uk

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 31 July 2016

**Accepted:** 27 September 2016

**Published:** 17 October 2016

### Citation:

Lobato-Márquez D, Molina-García L,  
Moreno-Córdoba I, García-del  
Portillo F and Díaz-Orejas R (2016)  
Stabilization of the Virulence Plasmid  
pSLT of *Salmonella* Typhimurium by  
Three Maintenance Systems and Its  
Evaluation by Using a New Stability  
Test. *Front. Mol. Biosci.* 3:66.  
doi: 10.3389/fmolb.2016.00066

Certain *Salmonella enterica* serovars belonging to subspecies I carry low-copy-number virulence plasmids of variable size (50–90 kb). All of these plasmids share the *spv* operon, which is important for systemic infection. Virulence plasmids are present at low copy numbers. Few copies reduce metabolic burden but suppose a risk of plasmid loss during bacterial division. This drawback is counterbalanced by maintenance modules that ensure plasmid stability, including partition systems and toxin-antitoxin (TA) loci. The low-copy number virulence pSLT plasmid of *Salmonella enterica* serovar Typhimurium encodes three auxiliary maintenance systems: one partition system (*parAB*) and two TA systems (*ccdAB*<sub>ST</sub> and *vapBC2*<sub>ST</sub>). The TA module *ccdAB*<sub>ST</sub> has previously been shown to contribute to pSLT plasmid stability and *vapBC2*<sub>ST</sub> to bacterial virulence. Here we describe a novel assay to measure plasmid stability based on the selection of plasmid-free cells following elimination of plasmid-containing cells by ParE toxin, a DNA gyrase inhibitor. Using this new maintenance assay we confirmed a crucial role of *parAB* in pSLT maintenance. We also showed that *vapBC2*<sub>ST</sub>, in addition to contribute to bacterial virulence, is important for plasmid stability. We have previously shown that *ccdAB*<sub>ST</sub> encodes an inactive CcdB<sub>ST</sub> toxin. Using our new stability assay we monitored the contribution to plasmid stability of a *ccdAB*<sub>ST</sub> variant containing a single mutation (R99W) that restores the toxicity of CcdB<sub>ST</sub>. The “activation” of CcdB<sub>ST</sub> (R99W) did not increase pSLT stability by *ccdAB*<sub>ST</sub>. In contrast, *ccdAB*<sub>ST</sub> behaves as a canonical type II TA system in terms of transcriptional regulation. Of interest, *ccdAB*<sub>ST</sub> was shown to control the expression of a polycistronic operon in the pSLT plasmid. Collectively, these results show that the contribution of the CcdB<sub>ST</sub> toxin to pSLT plasmid stability may depend on its role as a co-repressor in coordination with CcdA<sub>ST</sub> antitoxin more than on its toxic activity.

**Keywords:** virulence plasmid, toxin-antitoxin, plasmid stability, transcriptional regulation, *Salmonella* Typhimurium



## INTRODUCTION

During evolution bacterial pathogens acquire new genes dedicated to manipulate host processes. Many of these pathogen functions are encoded by chromosomal genes. Others, however, can be encoded by genes present in mobile genetic elements such as virulence plasmids. Horizontal transfer of these mobile genetic components has shaped the host adaptation strategies in several bacterial pathogens (Jackson et al., 2011). The presence of a virulence gene in a mobile element also facilitates its rapid acquisition or loss under distinct selective pressures. Enteric bacteria such as *Escherichia coli*, *Shigella* spp. and *Salmonella enterica*, frequently carry virulence genes in large transmissible low-copy-number plasmids (Sasakawa et al., 1986; Makino et al., 1988; Gulig et al., 1993). The *S. enterica* species are facultative intracellular bacteria that cause disease ranging from self-limiting gastroenteritis to more severe systemic infections (Rivera-Chávez and Bäumler, 2015). *S. enterica* subdivides into seven subspecies (I, II, IIIa, IIIb, IV, VI, and VII) (Tindall et al., 2005; Grimont and Weill, 2007) and subspecies I includes more than 2500 serovars (Grimont and Weill, 2007). Most of these serovars have adapted to infect warm-blooded hosts. One of the most extensively studied serovars of subspecies I is Typhimurium, which infects both humans and livestock. Serovar Typhimurium, together with a few other serovars of subspecies I, possesses a virulence plasmid (Jones et al., 1982). These plasmids have a variable size of 50–90 kb and share common features such as low copy number (1–2 plasmids per chromosome), a similar *repC* replicon (similar to the *repFIB* family) and a conserved set of virulence genes encoding toxins and fimbrial proteins (including *spv* and *pef* operons) (Bäumler et al., 1998; Rotger and Casadesús, 1999). The low copy number of the *S. Typhimurium* virulence plasmid (also called pSLT) could theoretically compromise its heritability to daughter cells during cell division. Despite this, pSLT is extremely stable in the host with  $\sim 10^{-7}$  segregants per cell generation, in a similar rate to that observed for other low-copy-number plasmids such as F and P1 (Austin and Abeles, 1983; Kline, 1985; Tinge and Curtiss, 1990a). Low-copy-number plasmids carry maintenance modules such as partition systems and toxin-antitoxin (TA) systems that ensure their proper segregation to nascent cells (Million-Weaver and Camps, 2014). Partition systems significantly increase the stability of plasmids by ensuring segregation of one copy of the plasmid to each sibling cell (Ebersbach and Gerdes, 2005). On the other hand, TA modules are typically bicistronic operons that encode an unstable antitoxin and a stable toxin (Chan et al., 2016; Lobato-Márquez et al., 2016). As a consequence of their different stabilities, antitoxin must be continuously produced to efficiently neutralize its cognate toxin (Gerdes et al., 1986). However, if the TA-encoding plasmid is lost, the antitoxin cannot be replenished and the free toxin eliminates or reduces the growth of daughter cells thus diluting plasmid-free cells in the population (Yamaguchi and Inouye, 2011). This phenomenon is called post-segregational killing (Gerdes et al., 1986).

Classically, plasmid stability has been measured using antibiotic-resistance plasmid derivatives. Cells harboring the studied plasmid are positively selected in the presence of the

selection antibiotic and those that have lost the plasmid are killed (Gerdes et al., 1985; del Solar et al., 1987). The main drawback of this technique is its sensitivity. Highly stable plasmids such as *S. Typhimurium* pSLT are below the sensitivity range of these assays. To solve this problem other methods relying in the direct selection of plasmid-free cells have been developed; for instance, the one based on the *tetAR*-chlortetracycline system (Bochner et al., 1980; Maloy and Nunn, 1981). The *tetA* gene encodes a protein which resides in the cytoplasmic membrane and prevents cellular accumulation of tetracycline, thereby conferring resistance (Reyrat et al., 1998). However, TetA location in the bacterial membrane also causes the cell to become hypersensitive to lipophilic chelators such as fusaric or quinalic acids (Bochner et al., 1980). Therefore, it is possible to select those cells that have lost the *tetAR* cassette. Inserted in a plasmid, the *tetAR* cassette can be used to select plasmid-free cells in special agar plates (Bochner-Maloy) containing fusaric acid (García-Quintanilla et al., 2006). Limitations of this method include poor reproducibility and the frequent occurrence of false positives (Li et al., 2013). Here, we have developed a novel, highly sensitive stability assay based on the negative selection of plasmid-containing cells. This assay is based on a cassette containing the ParE toxin-encoding gene of the *parDE* TA system and a kanamycin resistance gene (*aph*). ParE toxin targets DNA gyrase, blocks DNA replication and induces DNA breaks leading to cell death (Jiang et al., 2002). In our system ParE synthesis is controlled by a rhamnose-inducible promoter (*P<sub>parE</sub>*) (Maisonneuve et al., 2011). Once the *aph-parE* cassette has been inserted in the plasmid of interest and upon induction of *P<sub>parE</sub>*, only plasmid-free cells survive. Using this new tool we studied the contribution of the three main maintenance modules of the pSLT virulence plasmid of *S. Typhimurium*: the *parAB* partition system (Tinge and Curtiss, 1990b) and the *ccdAB<sub>ST</sub>* and *vapBC2<sub>ST</sub>* TA loci (Lobato-Márquez et al., 2015). We show that *vapBC2<sub>ST</sub>* TA module, which we recently demonstrated to be important to *S. Typhimurium* survival during non-phagocytic cells infection (Lobato-Márquez et al., 2015, 2016), also stabilizes pSLT plasmid. We show that the *ccdAB<sub>ST</sub>* TA system, known to impact pSLT heritability and encoding an inactive toxin (García-Quintanilla et al., 2006; Lobato-Márquez et al., 2015), conserves its TA transcriptional regulatory activity. Of interest, the *ccdAB<sub>ST</sub>* operon extends beyond the toxin gene including four additional open reading frames. Moreover, CcdAB<sub>ST</sub> TA complexes influence expression of downstream genes. We also demonstrate that stability of pSLT plasmid is not affected by a mutation (R99W) that restores CcdB<sub>ST</sub> toxicity. We propose that the contribution of *ccdAB<sub>ST</sub>* to pSLT stability could be related to the regulatory activity of CcdA<sub>ST</sub>-CcdB<sub>ST</sub> complexes rather than to a post-segregational killing effect mediated only by CcdB<sub>ST</sub> toxicity.

## RESULTS

### Development of a New Assay to Measure Plasmid Stability

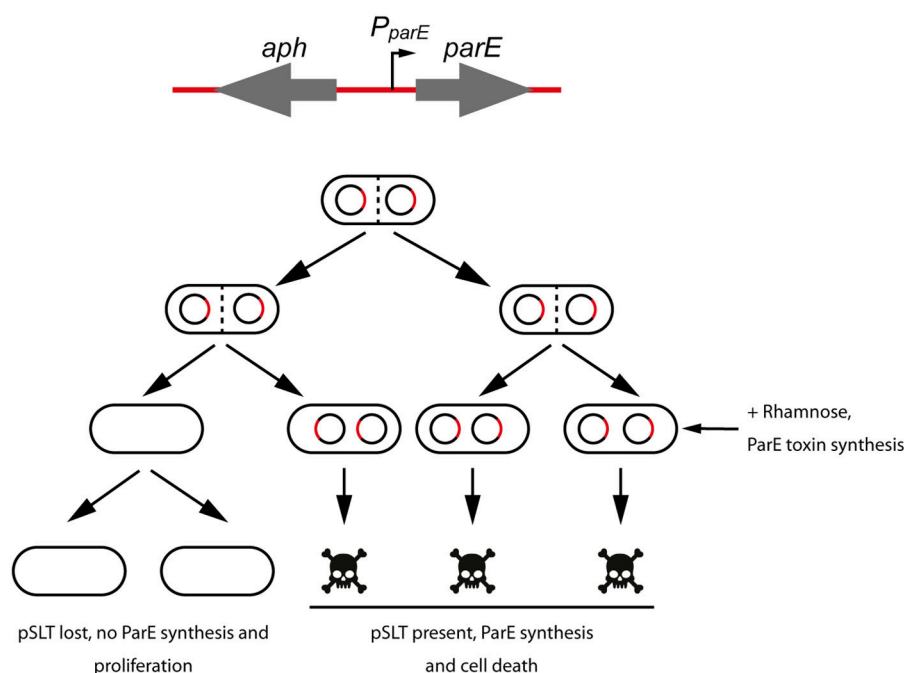
Due to the recognized problems of the *tetAR*-chlortetracycline method to measure plasmid stability, we decided to develop a

novel negative selection method to measure the contribution to stability of the different maintenance modules encoded in pSLT plasmid (**Figure 1**). We took advantage of an *aph-parE* cassette of the pKD267 plasmid (Maisonneuve et al., 2011). This cassette carries a kanamycin resistance gene (*aph*) and the *parE* gene, which encodes the toxin of the *parDE* TA system. ParE toxin interacts with and blocks the DNA gyrase, causing inhibition of DNA synthesis, induction of breaks and nicks in the DNA and finally cell death (Jiang et al., 2002). In the *aph-parE* cassette, previously used for chromosomal scarless deletions (Maisonneuve et al., 2011; Lobato-Márquez et al., 2015), the toxin-encoding *parE* gene is controlled by a rhamnose-inducible promoter. Thus, when rhamnose is present as the only carbon source in the medium, ParE is synthesized and the cell is killed (**Figure 1**). Using *aph-parE* cassette to disrupt the maintenance modules of pSLT plasmid we could select plasmid-free bacteria. To distinguish plasmid curing from other events causing rhamnose resistance (e.g., mutations in *P<sub>parE</sub>* promoter or *parE* gene), we took advantage of the kanamycin resistance gene also present in the *aph-parE* cassette. The resulting pSLT plasmid derivatives were thus tagged with two different markers.

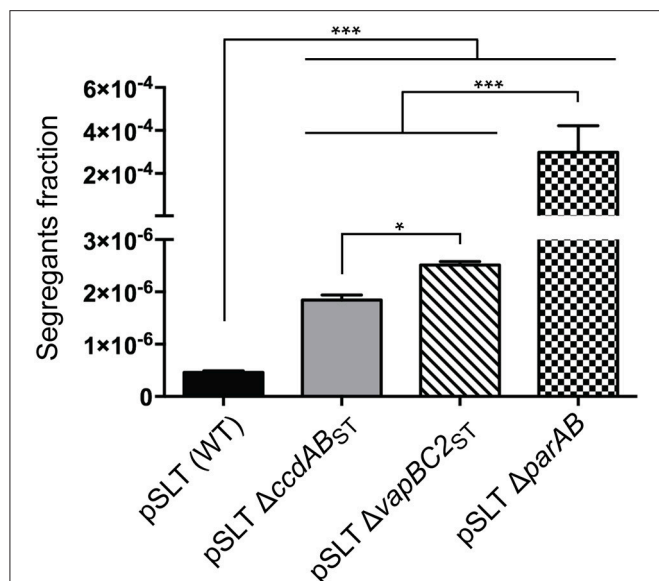
### ***parAB* Partition System and *vapBC2<sub>ST</sub>* Promote Stability of *S. Typhimurium* pSLT Plasmid**

Previous studies proposed two important regions involved in *S. Typhimurium* pSLT plasmid stability: the *parAB* partition

system and the TA module *ccdAB<sub>ST</sub>* (Tinge and Curtiss, 1990b; García-Quintanilla et al., 2006). Additionally, we identified another TA system, called *vapBC2<sub>ST</sub>*, encoded within the *trbH* gene (Lobato-Márquez et al., 2015) and homologous to the *mvpAT* locus encoded in the virulence plasmid of *Shigella flexneri* (Sayeed et al., 2000). We reported that *vapBC2<sub>ST</sub>* promotes *Salmonella* survival inside infected host cells. We now evaluated if similarly to *ccdAB<sub>ST</sub>*, and to other plasmidic TA loci, *vapBC2<sub>ST</sub>* may play a role in pSLT stability. Additionally, to test the sensitivity and the reproducibility of our method, we reevaluated the contribution of *parAB* and *ccdAB<sub>ST</sub>* using the new stability assay. We compared pSLT plasmids derivatives lacking *parAB*, *ccdAB<sub>ST</sub>* or *vapBC2<sub>ST</sub>* with an isogenic strain in which *aph-parE* cassette was inserted in the gene *spvA*, which was previously shown to be innocuous for the stability of pSLT (García-Quintanilla et al., 2006). Stability assays demonstrated that disruption of *vapBC2<sub>ST</sub>* TA system resulted in a  $5.5 \pm 0.1$  fold increase in the fraction of segregants after  $\sim 10$  generations of growth without selection pressure (**Figure 2**). This increase was more important than in the case of the pSLT derivative lacking *ccdAB<sub>ST</sub>* ( $4 \pm 0.2$ ) under the same growth conditions (**Figure 2**). In accordance with previous studies, disruption of *parAB* or *ccdAB<sub>ST</sub>* decreased pSLT stability (Tinge and Curtiss, 1990a; García-Quintanilla et al., 2006). The *parAB* partition system stabilizes pSLT plasmid  $119 \pm 3$  and  $163 \pm 9$  fold more efficiently than the *vapBC2<sub>ST</sub>* or *ccdAB<sub>ST</sub>* TA systems, respectively (**Figure 2**). Moreover, the pSLT wild type plasmid was  $650.1 \pm 190.2$  fold more stable than pSLT lacking *parAB* (**Figure 2**). These



**FIGURE 1 | Scheme of the method used to measure plasmid stability of *S. Typhimurium* pSLT derivatives.** A cassette containing a kanamycin resistance gene (*aph*) and the DNA gyrase inhibitor ParE-encoding gene was used to disrupt genes of interest (**top**). Grown cultures were plated in rhamnose-containing agar plates and synthesis of ParE toxin was induced. Cells that kept pSLT plasmid and therefore *aph-parE* cassette, were selectively killed (**bottom**).



**FIGURE 2 | Plasmid stability assays of *S. Typhimurium* pSLT derivatives.**

Segregants fraction measurement of pSLT plasmid comparing bacteria bearing the following plasmids: (i) wild type pSLT, (ii) pSLT lacking *ccdAB*<sub>ST</sub> TA module, (iii) pSLT lacking *vapBC2*<sub>ST</sub> TA system, and (iv) pSLT lacking *parAB* partition system. Disruption of all maintenance modules significantly decreased pSLT stability. The fraction of segregants was determined dividing the number of colony forming units grown in M9-rhamnose-agar plates (plasmid-free bacteria) by total bacteria grown in LB-agar plates. Data represent the means and standard deviations from five independent experiments. \**P* < 0.05; \*\*\**P* < 0.001 by one-way ANOVA and Tukey's multiple comparison post-test.

data strongly suggested that *parAB* is the main contributor to pSLT heritability. However, *ccdAB*<sub>ST</sub> and *vapBC*<sub>ST</sub> TA systems showed a moderate contribution to pSLT stability. Together, these results demonstrated the potential of this new stability assay to determine accurately plasmid lost rates, being able to detect ~1 segregant in 2·10<sup>6</sup> bacteria. Moreover, we demonstrated that *vapBC2*<sub>ST</sub>, apart from its contribution to *S. Typhimurium* virulence, also mediates pSLT heritability.

### CcdB<sub>ST</sub> Toxicity Is Not Required for *ccdAB*<sub>ST</sub>-Mediated Stability of pSLT

Our stability assays agreed with a previous study reporting contribution of *ccdAB*<sub>ST</sub> to pSLT plasmid stability (García-Quintanilla et al., 2006). We have recently demonstrated that CcdB<sub>ST</sub> toxin of *S. Typhimurium* is not functional due to an amino acid substitution in the position 99 (W99R) (Lobato-Márquez et al., 2015). This residue is essential for the binding of CcdB to the subunit A of DNA Gyrase (GyrA) (Bahassi et al., 1995; Dao-Thi et al., 2005). The lack of toxic activity in CcdB<sub>ST</sub> was further confirmed in liquid cultures of *S. Typhimurium* expressing either wild type (inactive) or active (R99W) versions of CcdB<sub>ST</sub> (Figure 3A).

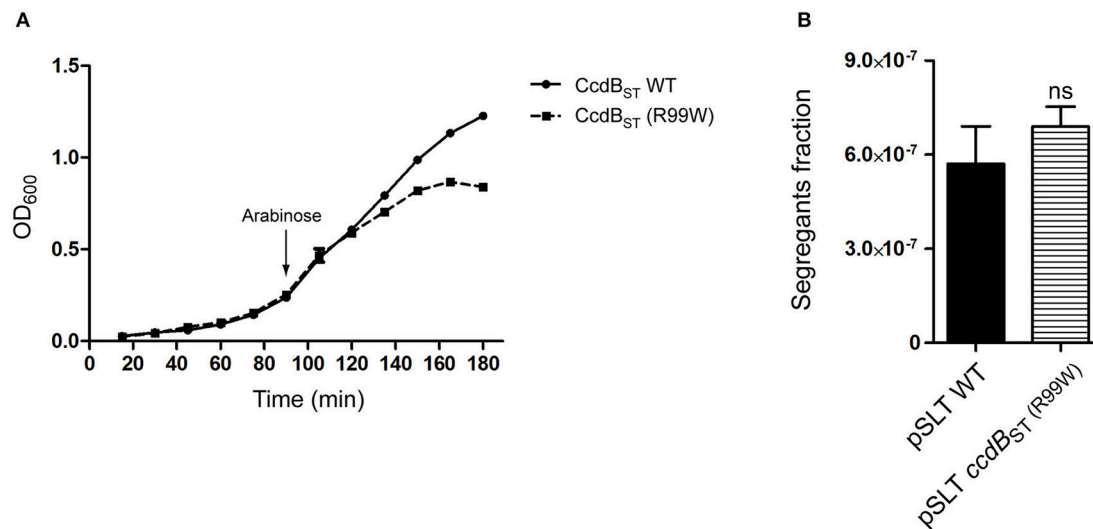
In the F plasmid, the ortholog TA module *ccdAB* contributes to plasmid stability through a mechanism called “post-segregational killing” (Gerdes et al., 1986). Cells that do not inherit a copy of the TA-encoding plasmid cannot synthesize

new antitoxin, leading the toxin free to kill or reduce the growth of plasmid-free cells (Van Melder et al., 1994). We asked whether in *S. Typhimurium* the toxicity of CcdB<sub>ST</sub> could be important for pSLT stability. To test this hypothesis, we carried out stability assays using a pSLT plasmid in which the non-functional *ccdB*<sub>ST</sub> was substituted by an activated *ccdB*<sub>ST</sub> (R99W) variant. Stability assays showed no differences between the pSLT plasmid derivatives containing wild type *ccdB*<sub>ST</sub> or the toxic version *ccdB*<sub>ST</sub> (R99W), suggesting that CcdB<sub>ST</sub> toxicity is dispensable for *ccdAB*<sub>ST</sub>-dependent stability (Figure 3B). Due to the ability of CcdAB<sub>ST</sub> TA system to stabilize pSLT plasmid independently of CcdB<sub>ST</sub> toxicity, we characterized the *ccdAB*<sub>ST</sub> operon in more detail.

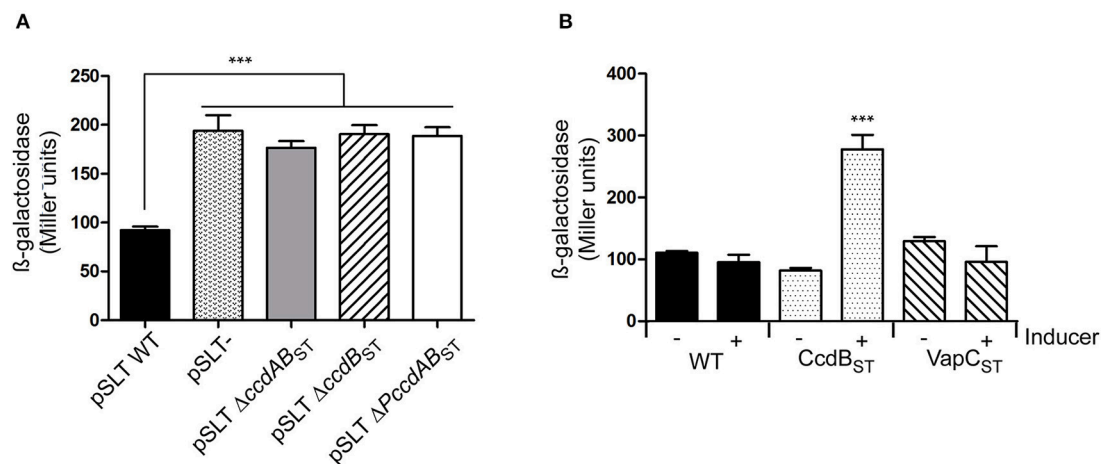
### The Non-functional TA System *ccdAB*<sub>ST</sub> of *S. Typhimurium* Conserves Transcriptional Regulatory Activity

We tested if the type II TA module *ccdAB*<sub>ST</sub> of *S. Typhimurium* behaves as a *bona fide* TA system in terms of transcriptional regulation. In the F plasmid the antitoxin CcdA of the *ccdAB* ortholog acts as a transcriptional repressor and the toxin enhances the repressor activity when TA complexes are formed in a proper stoichiometry (Tam and Kline, 1989; Salmon et al., 1994). Mutations in the last three amino acids of the CcdB toxin in the F plasmid eliminate its toxicity while maintain its regulatory activity (Bahassi et al., 1995). To test the transcriptional activity of *S. Typhimurium* *ccdAB*<sub>ST</sub>, we fused the promoter of the TA system (*PccdAB*<sub>ST</sub>) to a promoterless *lacZ* reporter gene. We measured β-galactosidase activity in the following genetic backgrounds: (i) pSLT wild type, (ii) pSLT plasmid cured, (iii) pSLT deficient for *ccdB*<sub>ST</sub> gene, (iv) pSLT deficient for *ccdAB*<sub>ST</sub> operon, and (v) pSLT only lacking promoter *PccdAB*<sub>ST</sub>. β-galactosidase assays demonstrated that *ccdAB*<sub>ST</sub> TA module behaves as a classical type II TA system. When the whole system is present (wild type background), transcription of the operon is repressed. However, this repression is lost in the absence of CcdAB<sub>ST</sub> repressor complexes due to the loss of either *ccdB*<sub>ST</sub> or *ccdAB*<sub>ST</sub> (Figure 4A). Interestingly, we did not observe differences in β-galactosidase activity when the system lacked only the toxin *ccdB*<sub>ST</sub> or the whole operon arguing for an important role of CcdB<sub>ST</sub> in transcriptional regulation.

In many type II TA modules, transcriptional regulation relies on the toxin:antitoxin ratio. Thus, an excess of antitoxin results in TA complexes that are efficient repressors; however, when the number of toxin molecules increases, the stoichiometry of the complex changes and repression is relieved. This regulation feature is termed “conditional cooperativity” (Overgaard et al., 2008). Taking advantage of the inactive CcdB<sub>ST</sub> toxin, we analyzed the conditional cooperativity phenomenon in the *ccdAB*<sub>ST</sub> TA module of pSLT plasmid by supplying *in trans* an extra dose of the inactive CcdB<sub>ST</sub> toxin. We employed a plasmid that contains inactive *ccdB*<sub>ST</sub> gene controlled by an arabinose-inducible promoter. To discard unspecific effects derived from protein over-production, the same experiment was carried out with the unrelated non-toxic VapC<sub>ST</sub> toxin encoded in the



**FIGURE 3 | The toxic activity of CcdB<sub>ST</sub> is dispensable for the CcdAB<sub>ST</sub>-mediated pSLT stability.** (A) Growth curves of *S. Typhimurium* strains expressing wild type non-active or CcdB<sub>ST</sub> (R99W) toxic proteins. Bacteria were grown at 37°C with shaking in LB medium. The expression of the *ccdB<sub>ST</sub>* (R99W) toxin-encoding gene ceased bacterial growth. The arrow indicates the time point (90 min) at which CcdB<sub>ST</sub> synthesis was induced by arabinose addition. (B) Segregants fraction measurement of pSLT plasmid comparing pSLT wild type and a pSLT variant harboring the toxic version *ccdB<sub>ST</sub>* (R99W). Data represent the means and standard deviations from five independent experiments. Data were compared using Student's *T*-test. ns, non significant.



**FIGURE 4 | Transcriptional regulation of the *ccdAB<sub>ST</sub>* TA system.** The 300 bp region upstream to *ccdA<sub>ST</sub>* was cloned as a transcriptional fusion with a promoter-less *lacZ* reporter gene in plasmid pMP220. (A) β-galactosidase was measured in different genetic backgrounds of *S. Typhimurium* SV5015 transformed with pPccAB-lacZ: (i) pSLT wild type, (ii) a strain cured of pSLT, (iii) a pSLT plasmid deficient for *ccdAB<sub>ST</sub>* TA system (including the promoter of the system), (iv) a pSLT lacking the toxin *ccdB<sub>ST</sub>* gene, and (v) a pSLT in which the *PccdAB<sub>ST</sub>* promoter was eliminated. Absence of TA complexes releases the CcdAB<sub>ST</sub> transcriptional repression, resulting in increasing β-galactosidase activity (B) Wild type *S. Typhimurium* SV5015 was transformed with pPccAB-lacZ and either a plasmid expressing the non-functional copies of CcdB<sub>ST</sub> or VapC<sub>ST</sub> to further measure β-galactosidase activity. Cultures were grown to OD<sub>600</sub> of 0.3 and then expression of *ccdB<sub>ST</sub>* or *vapC<sub>ST</sub>* genes was induced by adding 0.3% arabinose during 1 h. Excess of CcdB<sub>ST</sub> specifically shows conditional cooperativity effect as its overexpression derepresses transcription at *PccdAB<sub>ST</sub>* promoter of pSLT plasmid. Data represent the means and standard deviations from four independent experiments. \*\*\**P* < 0.001 by one-way ANOVA.

*S. Typhimurium* chromosome (Lobato-Márquez et al., 2015). Upon arabinose addition, we specifically observed an increased transcriptional activity of the *PccdAB<sub>ST</sub>* promoter following CcdB<sub>ST</sub> but not VapC<sub>ST</sub> production (Figure 4B). These data demonstrate that the *ccdAB<sub>ST</sub>* TA system responds to conditional cooperativity.

### ***ccdAB<sub>ST</sub>* of *S. Typhimurium* pSLT Plasmid Conforms a Six-Gene Polycistronic Operon**

In the *E. coli* F plasmid, *ccdAB* maps upstream of the resolvase-encoding gene *resD*. However, analysis of the regions flanking *ccdAB<sub>ST</sub>* TA system of pSLT showed that this locus could be genetically linked to four other downstream genes (Figure 5A).

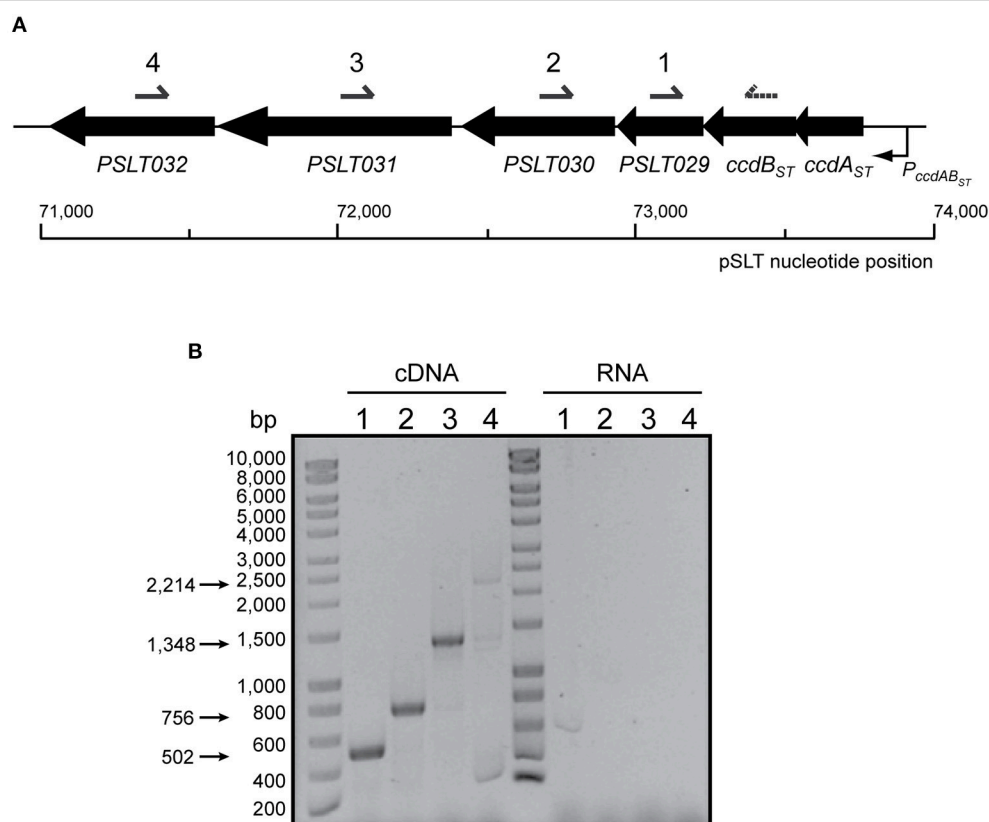


*ccdB<sub>ST</sub>* gene is separated by only one single nucleotide from the downstream gene *SL1344\_P1\_0078* (*PSLT029*), which itself overlaps 4 bp with *SL1344\_P1\_0077* (*PSLT030*). The next downstream gene is *SL1344\_P1\_0076* (*PSLT031* or *rsdB*). *PSLT031* maps 33 bp downstream from the 3'-end of *SL1344\_P1\_0077* (*PSLT030*) and 8 bp upstream from the 5'-end of *SL1344\_P1\_0075* (*PSLT032*) (**Figure 5A**). These short intergenic regions led us hypothesize that the TA system *ccdAB<sub>ST</sub>* of pSLT plasmid could be encoded within a six-gene polycistronic operon. RT-PCR assays confirmed a polycistronic operon encompassing from *ccdAB<sub>ST</sub>* to *PSLT032* (**Figure 5B**).

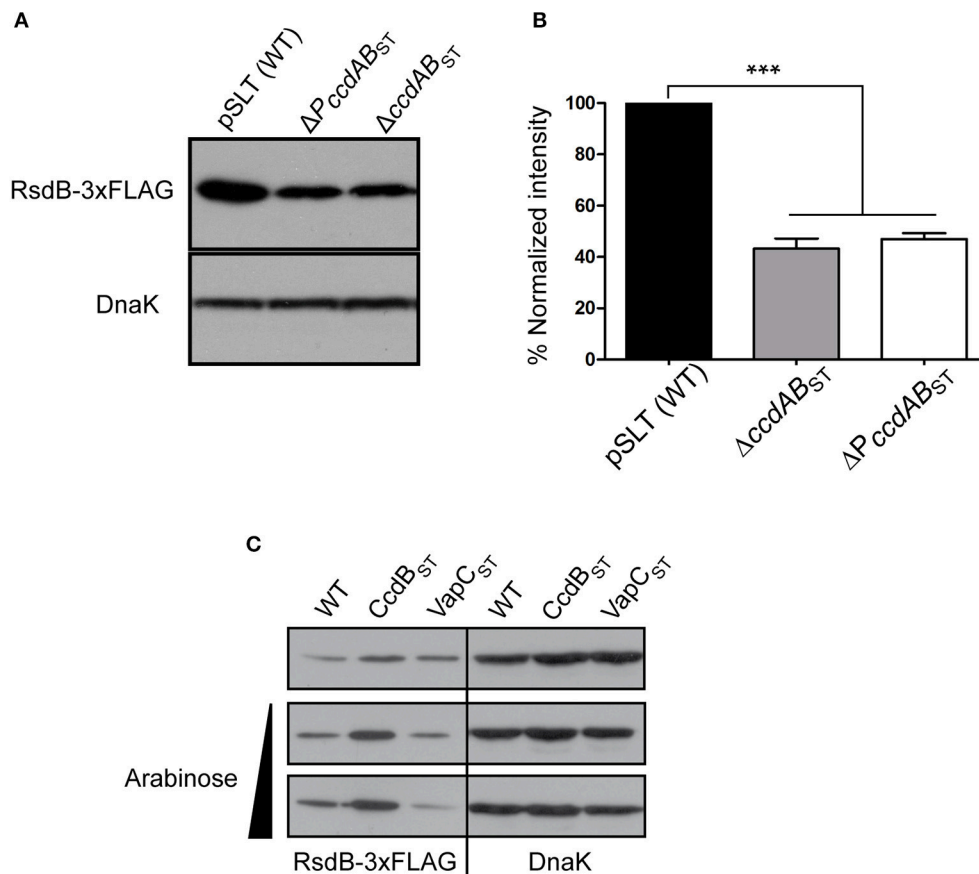
## *ccdAB<sub>ST</sub>* Transcriptional Regulation Is Important to Control the Polycistronic Operon

To further analyze the role of *ccdAB<sub>ST</sub>* in the polycistronic operon transcriptional control we asked if placed at the beginning of the operon, CcdA<sub>ST</sub>-CcdB<sub>ST</sub> TA complexes could modulate transcriptional expression of the operon in a TA system “classic” manner. *PSLT031* or *rsdB*, placed at the penultimate position of

the polycistronic operon, is annotated as a putative resolvase that could be important in multimer resolution during pSLT plasmid replication (Krause and Guiney, 1991). Thus, we used *rsdB* as a reporter to monitor the operon transcriptional regulation exerted by *ccdAB<sub>ST</sub>*. We tagged the *rsdB* gene with a 3xFLAG epitope at the 3'-end, and measured its protein levels in strains carrying: (i) wild type pSLT, (ii) engineered pSLT lacking the whole *ccdAB<sub>ST</sub>*, and (iii) pSLT lacking the 300 bp containing the *ccdAB<sub>ST</sub>* promoter. RsdB levels significantly decreased when the *ccdAB<sub>ST</sub>* TA system was altered, thus indicating that *ccdAB<sub>ST</sub>* acts as transcriptional repressor for the polycistronic operon (**Figures 6A,B**). As described above for the *ccdAB<sub>ST</sub>* TA system, we tested if the polycistronic operon could also respond to conditional cooperativity. We expressed the non-toxic CcdB<sub>ST</sub> variant and measured RsdB levels. Complementary, we used as a negative control the production of the unrelated toxin VapC<sub>ST</sub>. When CcdB<sub>ST</sub> was provided *in trans* (**Supplementary Figure 1**), RsdB levels increased accordingly to conditional cooperativity (**Figure 6C**). Altogether, these data demonstrate that CcdA<sub>ST</sub> TA complexes influence the transcription of the polycistronic operon.



**FIGURE 5 | Genetic organization and transcriptional analysis of the *ccdAB<sub>ST</sub>* locus present in *S. Typhimurium* pSLT plasmid. (A) Genetic organization of *ccdAB<sub>ST</sub>* in *S. Typhimurium* pSLT. Black arrows represent the location and orientation of the genes (scaled). Under the scheme it is marked pSLT DNA sequence coordinates. Primers used for RT-PCR (number 4) and cDNA amplification (numbers 1–4) are denoted above. (B) Co-transcription analysis of *ccdAB<sub>ST</sub>* and downstream genes. cDNA was synthesized using a primer annealing with *PSLT032* (named as 4). Primers named as 1, 2, 3, and 4 were used in combination to a primer annealing in 3'-end of *ccdB<sub>ST</sub>* (dashed arrow) to PCR-amplify each region, and PCR products were resolved in a 0.8% agarose gel. DNA marker is shown on the left. The expected sizes of PCR-amplified DNA fragments are indicated with arrows on the left. All genes are co-transcribed with those encoding the *ccdAB<sub>ST</sub>* TA system.**



**FIGURE 6 | *ccdAB<sub>ST</sub>* TA system influences the transcription of the polycistronic operon.** *rsdB* of pSLT plasmid was tagged with a 3xFLAG epitope in its 3'-end. **(A,B)** RsdB-3xFLAG levels were analyzed in pSLT plasmid lacking the *PccdAB<sub>ST</sub>* promoter or the entire *ccdAB<sub>ST</sub>* locus (including toxin, antitoxin and the promoter of the system). Bands intensity in **(A)** were quantified taking into account the number of cells loaded in the acrylamide gels and the band intensity of the loading control DnaK **(B)** Deletion of either the *ccdAB<sub>ST</sub>* operon or the promoter of the system, *PccdAB<sub>ST</sub>*, reduced RsdB 2.3 and 2.1 fold, respectively. **(C)** Conditional cooperativity phenomenon was analyzed in a wild type pSLT carrying the tagged 3xFLAG *rsdB*. *CcdB<sub>ST</sub>* or *VapC<sub>ST</sub>* (negative control) proteins were supplement *in trans* from pCcdB and pVapC plasmids. Due to the conditional cooperativity phenomenon, overexpression of *CcdB<sub>ST</sub>* results in derepression of *PccdAB<sub>ST</sub>* and increased amount of RsdB. Quantification was carried out as described for A and B. Protein expression of *CcdB<sub>ST</sub>* and *VapC<sub>ST</sub>* was confirmed in SDS-acrylamide gels stained with Coomassie blue (**Supplementary Figure 1**). Histograms represent the quantification of at least four independent experiments. \*\*\**p* < 0.001 by one-way ANOVA.

## DISCUSSION

In this report we describe a novel method to measure plasmid stability in bacteria. This procedure is based on the use of an *aph-parE* cassette in which a rhamnose-inducible promoter controls synthesis of ParE toxin. When the *aph-parE* cassette is inserted in the plasmid of interest and rhamnose is present in the medium as the only carbon source, ParE is synthesized and plasmid-containing cells are selectively eliminated. This methodology allows direct selection of plasmid-free segregants in a reproducible and highly sensitive manner. As it has been described previously for many low-copy-number plasmids, the pSLT virulence plasmid of *S. Typhimurium* possesses at least three main mechanisms to ensure its stable maintenance in the cell: (i) a copy number control of replication mediated by *repB* and *repC* replicons; (ii) the *parAB* partition system; and (iii) the TA systems *ccdAB<sub>ST</sub>* and *vapBC2<sub>ST</sub>*. In our study, we did not

considered the influence of the conjugation machinery because although *S. Typhimurium* SV5015 pSLT is mobilizable, it is not self-transmissible (Ahmer et al., 1999). Using our novel stability assay, we reevaluated the contribution of ParAB and *CcdAB<sub>ST</sub>* to pSLT plasmid stability as a proof of concept for the reliability of our methodology. In accordance with the literature, we show that the ParAB partition system stabilizes the pSLT plasmid very efficiently. Moreover, as described for other plasmids, the partition system appeared more important for pSLT stability than the *vapBC2<sub>ST</sub>* or *ccdAB<sub>ST</sub>* TA systems (Sia et al., 1995; Sengupta and Austin, 2011; Hernández-Arriaga et al., 2014). Several studies have demonstrated a moderately stabilizing effect of TA systems. Two examples are the *ccdAB* TA module of the fertility factor F (Ogura and Hiraga, 1983) and the *kis-kid* (also called *parD*) TA locus of the R1 plasmid (Bravo et al., 1987). These systems increase the stability of their host plasmids around 10-fold compared to mini-derivate plasmids (Hernández-Arriaga

et al., 2014). However, there are exceptions to this rule. For instance, the *parDE* module of RK2 has a more important role in the stabilization of this plasmid than other TA systems (Roberts et al., 1994; Easter et al., 1997). Interestingly, the *mvpTA* TA system of the virulence plasmid pWR100 in *S. flexneri* is the principal contributor to plasmid stability, more than the partition system (Sayeed et al., 2005). This differs from the stability contribution of its ortholog in *S. Typhimurium*, *vapBC<sub>ST</sub>*. Of note, MvpAT and VapBC<sub>ST</sub> show more than 96% amino acid sequence identity. However, it has also been described that diverse experimental variables, including temperature, growth media or the strain analyzed in the assay can alter plasmid stability (Easter et al., 1997; Sayeed et al., 2005). The toxin MvpT is a specific endonuclease that cleaves the initiator tRNA (Winther and Gerdes, 2011), and the *mvpTA* TA system has been shown to stabilize the virulence plasmid of *S. flexneri* by post-segregational killing (Sayeed et al., 2000). On the other hand, the plasmidic toxin VapC<sub>ST</sub> and its chromosomal paralog VapC<sub>ST</sub> of *S. Typhimurium* conserve 82% amino acid sequence identity (Lobato-Márquez et al., 2015). Moreover, similar to MvpT toxin, the chromosomal VapC<sub>ST</sub> toxin possesses tRNA endonuclease activity (Winther and Gerdes, 2011). These evidences imply that VapBC<sub>ST</sub> may mediate pSLT plasmid stability by post-segregational killing.

The other TA system of pSLT is *ccdAB<sub>ST</sub>*. In this work we demonstrate that this TA module shows classic characteristics of type II TA loci, such as autorepression and conditional cooperativity. Moreover, *ccdAB<sub>ST</sub>* is highly conserved to its ortholog present in the F plasmid: 90 and 83% amino acid identity to CcdA and CcdB, respectively. One important amino acid substitution is the tryptophan 99 to arginine in CcdB<sub>ST</sub> of pSLT, an indispensable residue for the toxic activity of CcdB (Bahassi et al., 1995). Using a pSLT plasmid derivative encoding a CcdB<sub>ST</sub> (R99W) variant we demonstrate that CcdB<sub>ST</sub> toxicity is not necessary for the contribution of this TA module to plasmid stability. Intriguingly, *ccdAB<sub>ST</sub>* forms part of a polycistronic operon with four other downstream genes. Moreover, CcdA<sub>ST</sub>-CcdB<sub>ST</sub> TA complexes contribute to the regulation of the expression of this operon. This result is surprising given that few exceptions escape the general rule of TA operons organization. These exceptions include TA modules with a third gene acting as the transcriptional repressor of the system (Zielenkiewicz and Ceglowski, 2005; Hallez et al., 2010) and a single case in which a chaperone, co-transcribed with a TA operon, facilitates the folding of the antitoxin and, therefore, its activity (Bordes et al., 2011). Although, RsdB levels decreased upon deletion of either the promoter of *ccdAB<sub>ST</sub>* TA module or the whole TA locus, we still detected RsdB by western blot. These results indicate that *PccdAB<sub>ST</sub>* does control the transcription of the operon but it may exist at least another additional promoter regulating the operon.

Future work should address how the unprecedented TA genomic organization of this novel polycistronic operon including *ccdAB<sub>ST</sub>* and its transcriptional regulation influence pSLT stability. pSLT is evolutionary related to F plasmid, yet in F *ccdAB* does not constitute such a polycistronic operon. The study of this particular TA system could shed light on the evolution and adaptation of TA modules to its bacterial host.

## MATERIALS AND METHODS

### Bacterial Strains, Plasmids and Growth Conditions

*S. enterica* serovar Typhimurium SV5015 (a SL1344 His<sup>+</sup> derivative strain Mariscotti and García-del Portillo, 2009) was used as parental strain (*S. Typhimurium* SL1344 accession number: NC\_016810.1). All strains and plasmids used in this study are listed in **Supplementary Table 1**. Bacteria were grown at 37°C with shaking at 150 rpm in Luria-Bertani (LB) medium. When necessary antibiotics were added at the following concentrations: kanamycin, 50 µg/ml; ampicillin, 50 µg/ml; chloramphenicol, 20 µg/ml.

A transcriptional fusion *PccdAB<sub>ST</sub>-lacZ* was designed to measure the transcriptional activity of *PccdAB<sub>ST</sub>* promoter. A 300 bp DNA sequence upstream of *ccdA<sub>ST</sub>* containing the promoter of *ccdAB<sub>ST</sub>* (Tam and Kline, 1989; Madl et al., 2006) was PCR-amplified, digested with EcoRI-KpnI and ligated with the large EcoRI-KpnI fragment of plasmid pMP220 (Spaink et al., 1987). The resulting plasmid was confirmed by DNA sequencing.

### Construction of *S. Typhimurium* Mutants

Oligonucleotide primers used in these procedures are listed in **Supplementary Table 2**. For disruption of pSLT plasmid maintenance modules, the deletion method described by Maisonneuve et al., was used (Maisonneuve et al., 2011). The strain used as control on stability assays was design inserting an *aph-parE* cassette in the *spvA* gene of pSLT. Disruption of this gene does not alter pSLT stability (Ahmer et al., 1999; García-Quintanilla et al., 2006).

A similar protocol to that involving generation of deletion mutants was used to introduce the amino acid substitution R99W in CcdB<sub>ST</sub>. Briefly, the *aph-parE* module was first introduced in *ccdB<sub>ST</sub>* gene. Then the cassette was cleaned up with a PCR-amplified DNA fragment bearing the nucleotide change C-T in the position 73,232 of pSLT plasmid corresponding to the first nucleotide of the arginine 99 (R99) codon.

Construction of *S. Typhimurium* recombinant strain expressing tagged RsdB-3xFLAG was carried out as previously described (Uzzau et al., 2001). 3x-FLAG tagging was performed at the 3'-end of the *PSLT031* gene.

All mutants were verified and confirmed by PCR.

### Plasmid Stability Assays

Before starting stability assays, bacteria were grown in LB containing 50 µg/ml kanamycin. For plasmid stability assays all bacterial strains were grown in 10 ml LB medium (10:1 flask:medium volume ratio) without selection pressure for 16 h (~10 generations) at 37°C and 150 rpm. We did not observe alterations in the growth rate of the pSLT plasmid derivatives lacking *parAB*, *ccdAB<sub>ST</sub>* or *vapBC<sub>2ST</sub>* compared to pSLT wild type plasmid. Aliquots of 1 ml of the culture were collected into 1.5 ml eppendorf tubes and bacteria were pelleted in a MiniSpin® Eppendorf centrifuge 1 min at 12,000 rpm at room temperature. Supernatants were discarded and bacterial pellets were washed twice with phosphate buffered saline (PBS) pH 7.4. This ensures proper elimination of LB medium traces that

otherwise could interfere with the growth in M9-rhamnose plates. Serial dilutions were done in PBS pH 7.4 and 100  $\mu$ l of the appropriate aliquots plated onto LB- or M9-rhamnose-agar plates. Typically a  $1:10^7$  dilution was used to quantify total bacterial population in LB-agar plates, and dilutions in the range  $1:1-10^3$  were used to determine the number of segregants in M9-rhamnose-agar plates. Plates were incubated for 24 h (LB-agar) or 48–72 h (M9-rhamnose-agar) at 37°C before counting of the colony forming units. Colony forming units grown in M9-rhamnose-agar were tested for their kanamycin resistance on antibiotic-containing LB plates. This is a sensitive assay that effectively eliminates plasmid-containing cells, thus allowing a direct selection of plasmid-free segregants.

## **$\beta$ -Galactosidase Activity Measurements**

Bacteria containing the plasmid with transcriptional fusion *PccdAB-lacZ* were grown to an optical density (OD)<sub>600</sub> of 0.6 at 37°C and 150 rpm in LB. Then,  $\beta$ -galactosidase activity was measured as previously described (Miller, 1972).

For the conditional cooperativity experiments, bacteria containing pCcdB or pVapC plasmids (Lobato-Márquez et al., 2015) were grown in LB to an OD<sub>600</sub> of 0.3 at 37°C and 150 rpm in the presence of 50  $\mu$ g/ml kanamycin. Inactive CcdB<sub>ST</sub> or VapC<sub>ST</sub> toxins were synthesized upon induction with 0.3 % (w/v) L-arabinose.  $\beta$ -galactosidase activity was assessed as in the rest of strains after 1 h of induction. The chromosomally-encoded *S. Typhimurium* VapC<sub>ST</sub> was used as a control to discard unspecific effects of protein expression in  $\beta$ -galactosidase measurements.

## **Reverse Transcriptase PCR (RT-PCR)**

To determine the presence of a polycistronic operon controlled by *ccdAB*<sub>ST</sub> total RNA was extracted from wild type *S. Typhimurium* SV5015 (Mariscotti and García-del Portillo, 2009) grown in LB at 37°C until OD<sub>600</sub>  $\sim$  0.3. Volume corresponding to 1 absorbance unit at OD<sub>600</sub> was lysed in 100  $\mu$ l lysis buffer (lysozyme 50 mg/ml, 0.3% SDS). Cells extracts were processed using *RNeasy minit kit* (#74104, Qiagen). cDNA was constructed employing ThermoScript RT-PCR system (#11146-016, Invitrogen), using 600 ng of total RNA as template, a *t<sub>m</sub>* of 60°C and 0.6  $\mu$ M of an oligonucleotide annealing with the 3'-end of *PSLT032* (Supplementary Table 2). cDNA was amplified by PCR (Pfu DNA polymerase, #M774B, Promega) using 0.5  $\mu$ M of primers annealing with *ccdB*<sub>ST</sub>, *SL1344\_P1\_0078* (*PSLT029*), *SL1344\_P1\_0077* (*PSLT030*), *rsdB*, and *SL1344\_P1\_0075* (*PSLT032*) (Supplementary Table 2). PCR amplification was carried out in duplicate using cDNA and RNA as a negative control. PCR products were visualized in 0.8% (w/v) agarose gels stained with ethidium bromide.

## **Detection of RsdB Levels by Western Blotting and Protein Levels Quantification**

Bacterial cultures were grown 16 h at 37°C and 150 rpm. Same amount of bacterial cells were collected (volumes were adjusted based on OD<sub>600</sub>), centrifuged (1 min at 12,000 rpm) and resuspend in Laemmli buffer (Laemmli, 1970). Bacterial protein extracts were resolved in SDS-PAGE using 15% polyacrylamide

gels and processed for Western blot assays. Levels of the *S. Typhimurium* DnaK protein were used as loading control. RsdB or DnaK detection were performed using anti-FLAG antibody (#F3165, Sigma-Aldrich) 1:2000 (2 h) or anti-DNAK 1:10,000 (1 h), respectively, dissolved in TBS-Tween buffer (137 mM NaCl, 0.1% m/v Tween 20 and 20 mM Tris-HCl pH 7.5) containing 3% non-fat milk. RsdB expression levels were calculated by western blotting experiments using extracts prepared from at least four independent experiments and pSLT plasmid variants expressing 3xFLAG-tagged RsdB. Mean data were taken as the relative expression levels of the proteins. Band densitometry was determined using Quantity One v.4.6.3 software (Bio-Rad, Berkeley, CA) as previously described (Molina-García and Giraldo, 2014; López-Villarejo et al., 2015).

## **Statistical Analyses**

Statistical significance was analyzed with GraphPad Prism v7 software (GraphPad Inc., La Jolla, CA) using one-way analysis of variance (ANOVA) with Dunnett's multiple comparison post-test for Figures 2, 4, 6B. In the comparison test used for Figure 3B a Student's *T*-test analysis was used. A  $P \leq 0.05$  was considered significant. Data are presented as mean  $\pm$  standard deviation of the mean (SEM).

## **AUTHOR CONTRIBUTIONS**

DL and RD: Conceived and designed the experiments; DL, LM, and IM: Performed the experiments; DL, LM, IM, FG, and RD: Analyzed the data; DL: Wrote the paper.

## **ACKNOWLEDGMENTS**

We are grateful to Josep Casadesus Pursals for his critical comments about the manuscript and for providing the strain SV3081 of *S. Typhimurium*. We thank RD and FG lab members for their comments and help, and Dr. Serge Mostowy and Alexandra Willis for their critical review of the manuscript. The work in RD and FG's laboratories is supported by grants BFU2011-25939 (RD), CSD2008-00013 (RD and FG), and BIO2013-46281-P/BIO2015-69085-REDC (FG) from the Spanish Ministry of Economy and Competitiveness.

## **SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmolb.2016.00066>

**Supplementary Figure 1 | Control assays showing proper protein synthesis of CcdB<sub>ST</sub> and VapC<sub>ST</sub> in the experiments involving conditional cooperativity regulation of *rsdB* in Figure 6C (main text).** Equal amounts of total protein extracts were loaded in each lane. Bacteria were grown in LB medium to OD<sub>600</sub> of 0.3, time at which CcdB<sub>ST</sub> or VapC<sub>ST</sub> expression was induced with 0.3% arabinose.

**Supplementary Table 1 | Bacterial strains and plasmids used in this study.**

**Supplementary Table 2 | Oligonucleotides used in this study.**



## REFERENCES

- Ahmer, B. M., Tran, M., and Heffron, F. (1999). The virulence plasmid of *Salmonella* Typhimurium is self-transmissible. *J. Bacteriol.* 181, 1364–1368.
- Austin, S., and Abeles, A. (1983). Partition of unit-copy miniplasmids to daughter cells. I. P1 and F miniplasmids contain discrete, interchangeable sequences sufficient to promote equipartition. *J. Mol. Biol.* 169, 353–372. doi: 10.1016/S0022-2836(83)80055-2
- Bahassi, E. M., Salmon, M. A., Van Melder, L., Bernard, P., and Couturier, M. (1995). F plasmid CcdB killer protein: *ccdB* gene mutants coding for non-cytotoxic proteins which retain their regulatory functions. *Mol. Microbiol.* 15, 1031–1037. doi: 10.1111/j.1365-2958.1995.tb02278.x
- Bäumler, A. J., Tsolis, R. M., Ficht, T. A., and Adams, L. G. (1998). Evolution of host adaptation in *Salmonella enterica*. *Infect. Immun.* 66, 4579–4587.
- Bochner, B. R., Huang, H. C., Schieven, G. L., and Ames, B. N. (1980). Positive selection for loss of tetracycline resistance. *J. Bacteriol.* 143, 926–933.
- Bordes, P., Cirinesi, A. M., Ummels, R., Sala, A., Sakr, S., Bitter, W., et al. (2011). SecB-like chaperone controls a toxin-antitoxin stress-responsive system in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8438–8443. doi: 10.1073/pnas.1101189108
- Bravo, A., de Torrontegui, G., and Díaz, R. (1987). Identification of components of a new stability system of plasmid R1, ParD, that is close to the origin of replication of this plasmid. *Mol. Gen. Genet.* 210, 101–110. doi: 10.1007/BF00337764
- Chan, W. T., Espinosa, M., and Yeo, C. C. (2016). Keeping the wolves at Bay: antitoxins of prokaryotic type II toxin-antitoxin systems. *Front. Mol. Biosci.* 3:9. doi: 10.3389/fmolb.2016.00009
- Dao-Thai, M. H., Van Melder, L., De Genst, E., Afif, H., Buts, L., Wyns, L., et al. (2015). Molecular basis of gyrase poisoning by the addiction toxin CcdB. *J. Mol. Biol.* 348, 1091–1102. doi: 10.1016/j.jmb.2005.03.049
- del Solar, G. H., Puyet, A., and Espinosa, M. (1987). Initiation signals for the conversion of single stranded to double stranded DNA forms in the streptococcal plasmid pLS1. *Nucleic Acids Res.* 15, 5561–5580. doi: 10.1093/nar/15.14.5561
- Easter, C. L., Sobecky, P. A., and Helinski, D. R. (1997). Contribution of different segments of the *par* region to stable maintenance of the broad-host-range plasmid RK2. *J. Bacteriol.* 179, 6472–6479.
- Ebersbach, G., and Gerdes, K. (2005). Plasmid segregation mechanisms. *Annu. Rev. Genet.* 39, 453–479. doi: 10.1146/annurev.genet.38.072902.091252
- García-Quintanilla, M., Prieto, A. I., Barnes, L., Ramos-Morales, F., and Casadesús, J. (2006). Bile-induced curing of the virulence plasmid in *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.* 188, 7963–7965. doi: 10.1128/JB.00995-06
- Gerdes, K., Larsen, J. E., and Molin, S. (1985). Stable inheritance of plasmid R1 requires two different loci. *J. Bacteriol.* 161, 292–298.
- Gerdes, K., Rasmussen, P. B., and Molin, S. (1986). Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proc. Natl. Acad. Sci. U.S.A.* 83, 3116–3120. doi: 10.1073/pnas.83.10.3116
- Grimont, P., and Weill, F. (2007). *Antigenic Formulae of the Salmonella Serovars, 9th Edn*. Paris: World Health Organization Collaborating Center for Reference and Research on *Salmonella*. Pasteur Institute.
- Gulig, P. A., Danbara, H., Guiney, D. G., Lax, A. J., Norel, F., and Rhen, M. (1993). Molecular analysis of *spv* virulence genes of the *Salmonella* virulence plasmids. *Mol. Microbiol.* 7, 825–830. doi: 10.1111/j.1365-2958.1993.tb01172.x
- Hallez, R., Geeraerts, D., Sterckx, Y., Mine, N., Loris, R., and Van Melder, L. (2010). New toxins homologous to ParE belonging to three-component toxin-antitoxin systems in *Escherichia coli* O157:H7. *Mol. Microbiol.* 76, 719–732. doi: 10.1111/j.1365-2958.2010.07129.x
- Hernández-Arriaga, A. M., Chan, W. T., Espinosa, M., and Díaz-Orejas, R. (2014). Conditional activation of toxin-antitoxin systems: postsegregational killing and beyond. *Microbiol. Spectr.* 2:PLAS-0009-2013. doi: 10.1128/microbiolspec.PLAS-0009-2013
- Jackson, R. W., Vinatzer, B., Arnold, D. L., Dorus, S., and Murillo, J. (2011). The influence of the accessory genome on bacterial pathogen evolution. *Mob. Genet. Elem.* 1, 55–65. doi: 10.4161/mge.1.1.16432
- Jiang, Y., Pogliano, J., Helinski, D. R., and Konieczny, I. (2002). ParE toxin encoded by the broad-host-range plasmid RK2 is an inhibitor of *Escherichia coli* gyrase. *Mol. Microbiol.* 44, 971–979. doi: 10.1046/j.1365-2958.2002.02921.x
- Jones, G. W., Rabert, D. K., Svinarich, D. M., and Whitfield, H. J. (1982). Association of adhesive, invasive, and virulent phenotypes of *Salmonella* Typhimurium with autonomous 60-megadalton plasmids. *Infect. Immun.* 38, 476–486.
- Kline, B. C. (1985). A review of mini-F plasmid maintenance. *Plasmid* 14, 1–16. doi: 10.1016/0147-619X(85)90027-7
- Krause, M., and Guiney, D. G. (1991). Identification of a multimer resolution system involved in stabilization of the *Salmonella* Dublin virulence plasmid pSDL2. *J. Bacteriol.* 173, 5754–5762.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685. doi: 10.1038/227680a0
- Li, X. T., Thomason, L. C., Sawitzke, J. A., Costantino, N., and Court, D. L. (2013). Positive and negative selection using the *tetA-sacB* cassette: recombineering and P1 transduction in *Escherichia coli*. *Nucleic Acids Res.* 41, e204. doi: 10.1093/nar/gkt1075
- Lobato-Márquez, D., Díaz-Orejas, R., and García-del Portillo, F. (2016). Toxin-antitoxins and bacterial virulence. *FEMS Microbiol. Rev.* 40, 592–609. doi: 10.1093/femsre/fuw022
- Lobato-Márquez, D., Moreno-Córdoba, I., Figueroa, V., Díaz-Orejas, R., and García-del Portillo, F. (2015). Distinct type I and type II toxin-antitoxin modules control *Salmonella* lifestyle inside eukaryotic cells. *Sci. Rep.* 5:9374. doi: 10.1038/srep09374
- López-Villarejo, J., Lobato-Márquez, D., and Díaz-Orejas, R. (2015). Coupling between the basic replicon and the Kis-Kid maintenance system of plasmid R1: modulation by Kis antitoxin levels and involvement in control of plasmid replication. *Toxins (Basel)* 7, 478–492. doi: 10.3390/toxins7020478
- Madl, T., Van Melder, L., Mine, N., Respondek, M., Oberer, M., Keller, W., et al. (2006). Structural basis for nucleic acid and toxin recognition of the bacterial antitoxin CcdA. *J. Mol. Biol.* 364, 170–185. doi: 10.1016/j.jmb.2006.08.082
- Maisonneuve, E., Shakespeare, L. J., Jørgensen, M. G., and Gerdes, K. (2011). Bacterial persistence by RNA endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13206–13211. doi: 10.1073/pnas.1100186108
- Makino, S., Sasakawa, C., and Yoshikawa, M. (1988). Genetic relatedness of the basic replicon of the virulence plasmid in *shigellae* and enteroinvasive *Escherichia coli*. *Microb. Pathog.* 5, 267–274. doi: 10.1016/0882-4010(88)90099-X
- Maloy, S. R., and Nunn, W. D. (1981). Selection for loss of tetracycline resistance by *Escherichia coli*. *J. Bacteriol.* 145, 1110–1111.
- Mariscotti, J. F., and García-del Portillo, F. (2009). Genome expression analyses revealing the modulation of the *Salmonella* Rcs regulon by the attenuator IgaA. *J. Bacteriol.* 191, 1855–1867. doi: 10.1128/JB.01604-08
- Miller, J. H. (1972). *Experiments in Molecular Genetics*. New York, NY: Cold Spring Harbor Laboratory.
- Million-Weaver, S., and Camps, M. (2014). Mechanisms of plasmid segregation: have multicopy plasmids been overlooked? *Plasmid* 75, 27–36. doi: 10.1016/j.plasmid.2014.07.002
- Molina-García, L., and Giraldo, R. (2014). Aggregation interplay between variants of the RepA-WH1 prionoid in *Escherichia coli*. *J. Bacteriol.* 196, 2536–2542. doi: 10.1128/JB.01527-14
- Ogura, T., and Hiraga, S. (1983). Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proc. Natl. Acad. Sci. U.S.A.* 80, 4784–4788. doi: 10.1073/pnas.80.15.4784
- Overgaard, M., Borch, J., Jørgensen, M. G., and Gerdes, K. (2008). Messenger RNA interferase RelE controls *relBE* transcription by conditional cooperativity. *Mol. Microbiol.* 69, 841–857. doi: 10.1111/j.1365-2958.2008.06313.x
- Reyrt, J. M., Pelicic, V., Gicquel, B., and Rappuoli, R. (1998). Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. *Infect. Immun.* 66, 4011–4017.
- Rivera-Chávez, F., and Bäumler, A. J. (2015). The pyromaniac inside you: *Salmonella* metabolism in the host gut. *Annu. Rev. Microbiol.* 69, 31–48. doi: 10.1146/annurev-micro-091014-104108
- Roberts, R. C., Ström, A. R., and Helinski, D. R. (1994). The *parDE* operon of the broad-host-range plasmid RK2 specifies growth inhibition associated with plasmid loss. *J. Mol. Biol.* 237, 35–51. doi: 10.1006/jmbi.1994.1207
- Rotger, R., and Casadesús, J. (1999). The virulence plasmids of *Salmonella*. *Int. Microbiol.* 2, 177–184.
- Salmon, M. A., Van Melder, L., Bernard, P., and Couturier, M. (1994). The antidote and autoregulatory functions of the F plasmid CcdA protein: a genetic and biochemical survey. *Mol. Gen. Genet.* 244, 530–538. doi: 10.1007/BF00583904
- Sasakawa, C., Kamata, K., Sakai, T., Murayama, S. Y., Makino, S., and Yoshikawa, M. (1986). Molecular alteration of the 140-megadalton plasmid associated with

- loss of virulence and Congo red binding activity in *Shigella flexneri*. *Infect. Immun.* 51, 470–475.
- Sayed, S., Brendler, T., Davis, M., Reaves, L., and Austin, S. (2005). Surprising dependence on postsegregational killing of host cells for maintenance of the large virulence plasmid of *Shigella flexneri*. *J. Bacteriol.* 187, 2768–2773. doi: 10.1128/JB.187.8.2768-2773.2005
- Sayed, S., Reaves, L., Radnedge, L., and Austin, S. (2000). The stability region of the large virulence plasmid of *Shigella flexneri* encodes an efficient postsegregational killing system. *J. Bacteriol.* 182, 2416–2421. doi: 10.1128/JB.182.9.2416-2421.2000
- Sengupta, M., and Austin, S. (2011). Prevalence and significance of plasmid maintenance functions in the virulence plasmids of pathogenic bacteria. *Infect. Immun.* 79, 2502–2509. doi: 10.1128/IAI.00127-11
- Sia, E. A., Roberts, R. C., Easter, C., Helinski, D. R., and Figurski, D. H. (1995). Different relative importances of the *par* operons and the effect of conjugal transfer on the maintenance of intact promiscuous plasmid RK2. *J. Bacteriol.* 177, 2789–2797.
- Spaink, H. P., Okker, R. J., Wijffelman, C. A., Pees, E., and Lugtenberg, B. J. (1987). Promoters in the nodulation region of the *Rhizobium leguminosarum* Sym plasmid pRL1J1. *Plant Mol. Biol.* 9, 27–39. doi: 10.1007/BF00017984
- Tam, J. E., and Kline, B. C. (1989). Control of the *ccd* operon in plasmid F. *J. Bacteriol.* 171, 2353–2360.
- Tindall, B. J., Grimont, P. A., Garrity, G. M., and Euzeby, J. P. (2005). Nomenclature and taxonomy of the genus *Salmonella*. *Int. J. Syst. Evol. Microbiol.* 55(Pt 1), 521–524. doi: 10.1099/ijs.0.63580-0
- Tinge, S. A., and Curtiss, R. III. (1990a). Conservation of *Salmonella* Typhimurium virulence plasmid maintenance regions among *Salmonella* serovars as a basis for plasmid curing. *Infect. Immun.* 58, 3084–3092.
- Tinge, S. A., and Curtiss, R. III. (1990b). Isolation of the replication and partitioning regions of the *Salmonella* Typhimurium virulence plasmid and stabilization of heterologous replicons. *J. Bacteriol.* 172, 5266–5277.
- Uzzau, S., Figueroa-Bossi, N., Rubino, S., and Bossi, L. (2001). Epitope tagging of chromosomal genes in *Salmonella*. *Proc. Natl. Acad. Sci. U.S.A.* 98, 15264–15269. doi: 10.1073/pnas.261348198
- Van Melder, L., Bernard, P., and Couturier, M. (1994). Lon-dependent proteolysis of CcdA is the key control for activation of CcdB in plasmid-free segregant bacteria. *Mol. Microbiol.* 11, 1151–1157. doi: 10.1111/j.1365-2958.1994.tb00391.x
- Winther, K. S., and Gerdes, K. (2011). Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7403–7407. doi: 10.1073/pnas.1019587108
- Yamaguchi, Y., and Inouye, M. (2011). Regulation of growth and death in *Escherichia coli* by toxin-antitoxin systems. *Nat. Rev. Microbiol.* 9, 779–790. doi: 10.1038/nrmicro2651
- Zielenkiewicz, U., and Ceglowski, P. (2005). The toxin-antitoxin system of the streptococcal plasmid pSM19035. *J. Bacteriol.* 187, 6094–6105. doi: 10.1128/JB.187.17.6094-6105.2005

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer GDS declared a shared affiliation, though no other collaboration, with several of the authors IM, RD to the handling Editor, who ensured that the process nevertheless met the standards of a fair and objective review.

Copyright © 2016 Lobato-Márquez, Molina-García, Moreno-Córdoba, García-del Portillo and Díaz-Orejás. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Multifaceted Activity of the VirF Regulatory Protein in the *Shigella* Lifestyle

Maria Letizia Di Martino<sup>1</sup>, Maurizio Falconi<sup>2</sup>, Gioacchino Micheli<sup>3</sup>, Bianca Colonna<sup>1\*</sup> and Gianni Prosseda<sup>1\*</sup>

<sup>1</sup> Dipartimento di Biologia e Biotecnologie C. Darwin, Istituto Pasteur Italia-Fondazione Cenci Bolognietti, Sapienza Università di Roma, Roma, Italy, <sup>2</sup> Laboratorio di Genetica Molecolare e dei Microrganismi, Scuola di Bioscienze e Medicina Veterinaria, Università di Camerino, Camerino, Italy, <sup>3</sup> Istituto di Biologia e Patologia Molecolari, Consiglio Nazionale Delle Ricerche, Roma, Italy

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas Medical  
Branch-Galveston, USA

### Reviewed by:

Alessandra Polissi,  
University of Milan, Italy  
Paolo Landini,  
University of Milan, Italy

### \*Correspondence:

Bianca Colonna  
bianca.colonna@uniroma1.it  
Gianni Prosseda  
gianni.prosseda@uniroma1.it

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 28 July 2016

**Accepted:** 15 September 2016

**Published:** 29 September 2016

### Citation:

Di Martino ML, Falconi M, Micheli G,  
Colonna B and Prosseda G (2016)  
The Multifaceted Activity of the VirF  
Regulatory Protein in the *Shigella*  
Lifestyle. *Front. Mol. Biosci.* 3:61.  
doi: 10.3389/fmolb.2016.00061

*Shigella* is a highly adapted human pathogen, mainly found in the developing world and causing a severe enteric syndrome. The highly sophisticated infectious strategy of *Shigella* banks on the capacity to invade the intestinal epithelial barrier and cause its inflammatory destruction. The cellular pathogenesis and clinical presentation of shigellosis are the sum of the complex action of a large number of bacterial virulence factors mainly located on a large virulence plasmid (pINV). The expression of pINV genes is controlled by multiple environmental stimuli through a regulatory cascade involving proteins and sRNAs encoded by both the pINV and the chromosome. The primary regulator of the virulence phenotype is VirF, a DNA-binding protein belonging to the AraC family of transcriptional regulators. The *virF* gene, located on the pINV, is expressed only within the host, mainly in response to the temperature transition occurring when the bacterium transits from the outer environment to the intestinal milieu. VirF then acts as anti-H-NS protein and directly activates the *icsA* and *virB* genes, triggering the full expression of the invasion program of *Shigella*. In this review we will focus on the structure of VirF, on its sophisticated regulation, and on its role as major player in the path leading from the non-invasive to the invasive phenotype of *Shigella*. We will address also the involvement of VirF in mechanisms aimed at withstanding adverse conditions inside the host, indicating that this protein is emerging as a global regulator whose action is not limited to virulence systems. Finally, we will discuss recent observations conferring VirF the potential of a novel antibacterial target for shigellosis.

**Keywords:** VirF, *Shigella*, pathogenic *E. coli*, AraC proteins, transcriptional regulators, DNA binding proteins, bacterial virulence, antivirulence therapy

## INTRODUCTION

Bacterial pathogens must often survive within fundamentally diverse habitats. Dynamic adaptation to the surroundings depends on the ability to sense environmental variations and to respond in an appropriate manner. This involves drastic changes in the transcriptional program of the cell. The capability to swiftly modulate gene expression requires investment by the bacterium in numerous gene functions that not only allow adaptation to different milieus but also enable the cell to co-ordinately rework its response to mutable conditions (Cases and de Lorenzo, 2005). The complexity of gene expression in pathogenic bacteria can be viewed as an evolutionary

response to the challenge of surviving in changing environments (McAdams et al., 2004). Many pathogenic *E. coli*, including *Shigella*, are able to live in complex environments and have evolved intricate control systems. The expression of their pathogenic phenotype is the result of a multifactorial process which requires the synthesis of a large set of virulence determinants that may not be simultaneously needed during all stages of the infection process. These determinants are controlled by global regulatory networks, integrating specific regulators with conserved housekeeping processes.

*Shigella* is a facultative intracellular pathogen causing human bacillary dysentery, also known as shigellosis, a highly infectious disease widespread in developing countries. Although usually self-limiting, shigellosis can be fatal, especially in children (Anderson et al., 2016; Njamkepo et al., 2016; The et al., 2016). The pathogenicity of *Shigella* relies on its capacity to kill macrophages and invade colonic epithelial cells. Bacteria multiply intracellularly and spread to adjacent cells, with consequent cell death and inflammatory destruction of the mucosa (Schroeder and Hilbi, 2008; Ashida et al., 2015). The invasive process requires the coordinated expression of several genes located on the chromosome as well as on a virulence plasmid (pINV) (Sansone et al., 1982; Parsot, 2005). The acquisition of the pINV, a large F-type plasmid, by horizontal gene transfer (HTG) constitutes one of the most critical events in the evolution of the pathogenic lifestyle of *Shigella* since it encodes crucial elements of the molecular machinery required for invasion and intracellular survival (Pupo et al., 2000; The et al., 2016). pINV plasmids isolated from different *Shigella* spp. share significant homology and carry about 100 genes and an equivalent number of IS sequences (Buchrieser et al., 2000; Yang et al., 2005). Most genes required for host cell invasion and macrophage killing are contained in a conserved 31 kb region which is arranged as a PAI-like structure (Maurelli et al., 1985; Sasakawa et al., 1988). This so-called “entry region” consists of 34 genes organized into two large, divergently transcribed clusters. It contains the genes encoding the Ipa proteins, their chaperons and a specific T3SS system (Parsot, 2005). Besides these structural genes, the entry region hosts two regulatory genes coding for VirB and MxiE, the transcriptional activators required for the sequential expression of most pINV virulence genes (Beloin et al., 2002; Mavris et al., 2002). Scattered along the pINV, outside the entry region, are other genes encoding proteins crucial for the invasive process: the IcsA protein, responsible for the recruitment and polymerization of the host actin at one pole of the bacterial cell (Bernardini et al., 1989); the PhoN2 protein, required for proper IcsA localization (Scribano et al., 2014); the OspG protein, involved in the modulation of the host innate immune response (Kim et al., 2005); the IpaH proteins,

which interfere with the host ubiquitination-dependent protein degradation (Ashida and Sasakawa, 2015); and the VirF protein, the primary virulence regulator (Sakai et al., 1988; Prosseda et al., 1998).

In this review we will focus on VirF summarizing its structure, its sophisticated regulation, its role as major player in the cascade of events leading to the activation of the virulence program of *Shigella*, and its involvement in mechanisms aimed at withstanding adverse conditions inside the host. We will also address how these features confer VirF the potential of a novel antibacterial target for shigellosis.

## VirF, AN AraC-LIKE REGULATOR

VirF is a member of the AraC family of transcriptional regulators. This large family contains more than 800 different proteins, found mainly among gram negative bacteria and involved in the regulation of carbon metabolism, stress response, and virulence (Gallegos et al., 1997; Egan, 2002). The AraC proteins are characterized by two structural domains: a conserved C-terminal DNA binding domain and a more variable N-terminal signaling domain. The two domains are connected by an unstructured linker. DNA binding involves two helix-turn-helix (HTH) motifs and has been studied by analyzing the structure of the MarA (Rhee et al., 1998) and Rob (Kwon et al., 2000) proteins in complex with their target DNA, leading to the proposal that two different modes of DNA binding might exist involving either one (Rob) or both (MarA) HTH motifs. The N-terminal region is responsible for multimerization and/or binding of cofactors (Egan, 2002). The AraC family has been traditionally divided into three classes (Gallegos et al., 1997). The first one consists of proteins which, like AraC, act as regulators in response to a chemical signal (Schleif, 2010). Proteins involved in stress response, such as MarA and Rob, constitute the second class. VirF belongs to the third class, whose members control transcription in response to a physical signal (like temperature) and mostly serve as virulence gene regulators. While the proteins of the first and third class act as homodimers, proteins of the second class operate mainly as monomers. Members of the AraC family are frequently insoluble proteins. Precise molecular characterization of VirF and detailed information on its mechanism of action are still scarce as purification of VirF has been obtained only in a few cases (Porter et al., 2004; Tran et al., 2011). Most groups (Tobe et al., 1993; Porter and Dorman, 2002; Koppolu et al., 2013; Emanuele et al., 2014; Emanuele and Garcia, 2015) have reported difficulties in isolating VirF in quantities suitable to *in vitro* analysis and have therefore used a MalE—VirF fusion protein which still retains VirF functionality despite the lack of the first 10 N-terminal aminoacids (Tobe et al., 1993; Koppolu et al., 2013; Emanuele and Garcia, 2015).

VirF carries the two canonical AraC DNA-binding HTH motifs within its C-terminus. To acquire structural and functional information about the VirF protein, the entire *virF* gene has been subjected to both, random and site-directed mutagenesis and mutated proteins have been assayed for their ability to activate the plasmid-encoded invasive genes by

**Abbreviations:** FIS, factor for inversion stimulation; HGT, horizontal gene transfer; H-NS, heat stable nucleoid-structuring protein; HTH, helix-turn-helix; IC<sub>50</sub>, half maximal inhibitory concentration; IHE, integration host factor; Ipa, invasion plasmid antigen; IS, insertion sequence; MiaA, tRNA-N<sup>6</sup>-isopentyladenosine synthetase; PAI, pathogenicity island; PCr, Congo red phenotype; sRNA, small RNA; SAT, spermidine acetyl transferase; T3SS, type III secretion system; TCS, two component system; TGT, tRNA-guanine transglycosylase.



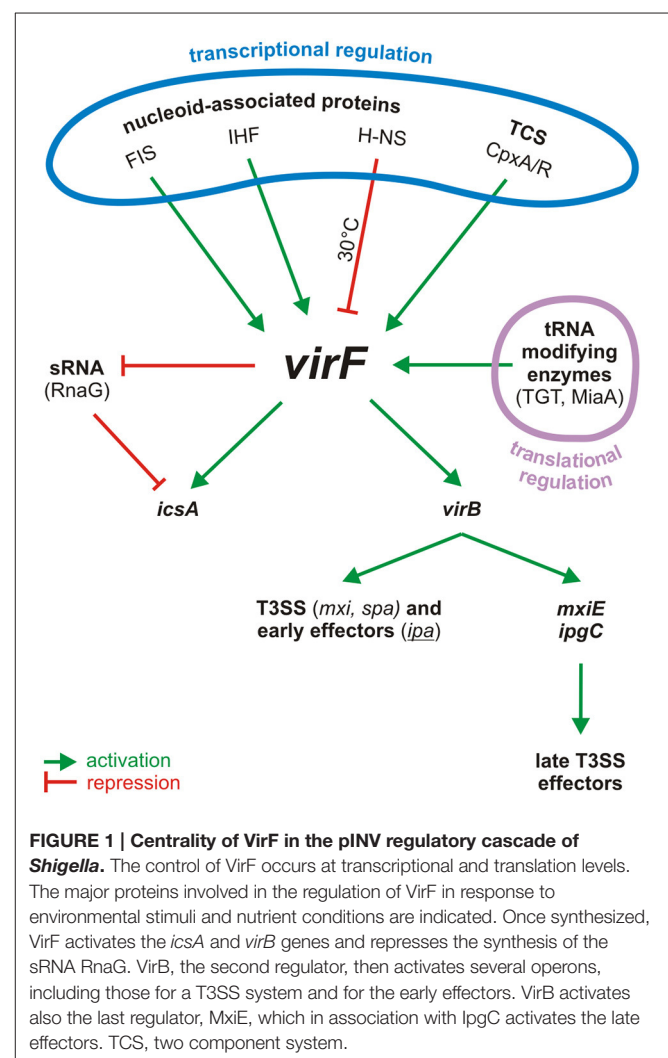
analyzing the expression of a *mxiC-lacZ* fusion (Porter and Dorman, 2002). Mutating the key residues of the first HTH motif, in particular in the positioning helix (I180) or in the recognition helix (K193), inactivates VirF. The second HTH motif is essential as well, since mutating the key residues (Y239 and I241), which according to the MarA-DNA co-crystal (Rhee et al., 1998) form specific contacts with DNA, leads again to VirF inactivation. The functionality of VirF is hampered also by modifications in the hydrophobic core of HTH2 or by deletion of the C-terminal HTH2 region. This strongly suggests that VirF interacts with its target sequences via both HTH motifs. Additional evidence on the relevance of the two motifs to DNA binding stems from studies on another AraC regulator, PerA, which shares significant homology with VirF. This protein is required for the activation of the bundle-forming pili in ETEC and is able to fully substitute VirF in the activation of the *Shigella virB* promoter (Porter et al., 2004). Mutations affecting the DNA-interacting nucleotides from either of the two C-terminal HTH motifs of PerA have been shown to inactivate the protein, indicating that also PerA requires both HTH motifs to interact with its DNA targets (Porter et al., 2004). However, while PerA and Rns, another AraC-like regulator from ETEC (Porter et al., 1998), are able to complement *Shigella virF* mutants, VirF is unable to restore the expression of the PerA or Rns regulated genes, suggesting that a particular DNA structure which forms only at *Shigella* VirF-regulated promoters is required as signal for the activation of the VirF protein (Porter and Dorman, 2002; Porter et al., 2004).

AraC and several members of the AraC family are known to form homodimers (Gallegos et al., 1997; Egan, 2002; Schleif, 2010) and the residues required for self-association are contained in the N-terminal domain. As yet the only evidence suggesting that VirF may act as dimer/oligomer stems from the observation that two mutants unable to bind DNA are dominant negative when co-expressed with the wild type VirF. However, the fact that other HTH1 or HTH2 mutations have no trans-dominant effect rises the possibility (Porter and Dorman, 2002) that VirF binds first as monomer and then, following DNA bending and DNA-DNA interactions, forms a multisubunit nucleoprotein complex, as previously observed in the case of the melibiose-dependent activator MelR (Bourgerie et al., 1997). This is consistent with the observation that a MalE—VirF fusion protein recognizes a single, large (about 100 bp) region on the *virB* promoter (Tobe et al., 1993). However, recent footprinting analyses using purified VirF and revealing the presence of four distinct VirF binding sites (each spanning 40–60 bp) within the *icsA-RnaG* regulatory region (Tran et al., 2011) indicate that, rather than interacting at a single large spot, VirF recognizes its target sites with different affinities and may give rise to a large nucleoprotein complex only at higher protein concentration.

## VirF IS AT THE TOP OF THE VIRULENCE REGULATORY CASCADE

The transcriptional activation of the invasive operons relies on the response to environmental stimuli like temperature, pH, and osmolarity, commonly encountered in the human intestine

(Schroeder and Hilbi, 2008). This process requires the VirF protein (Sakai et al., 1988; Adler et al., 1989; Falconi et al., 1998; Durand et al., 2000) which, through a sophisticated regulatory cascade (Dorman and Porter, 1998; Prosseda et al., 2002; Parsot, 2005), leads to the full expression of the virulence phenotype (Figure 1). Temperature is a crucial factor since transcription of the pINV invasion genes is strongly repressed at 30°C by the chromosome-encoded protein H-NS (Maurelli and Sansonetti, 1988; Falconi et al., 1998). The primary event following the upshift of *Shigella* to the host temperature (37°C) is the synthesis of VirF which acts as an antisilencer, relieving H-NS mediated repression at the *virB* (Tobe et al., 1993) and *icsA* (Tran et al., 2011) loci. This is not uncommon: in many bacterial pathogens H-NS, one of the major components of the bacterial nucleoid, acts as transcriptional silencer of virulence genes located on mobile genetic elements (Dorman, 2004, 2007). As for the *icsA* gene, it encodes an outer membrane protein which promotes the intra- and intercellular spreading of *Shigella* among the epithelial cells of the colon (Bernardini et al., 1989; Lett et al., 1989). The other VirF-activated gene encodes a secondary transcriptional



activator, VirB, which antagonizes the repressive activity of H-NS at the promoters of several pINV virulence genes by remodeling the DNA within H-NS-DNA nucleoprotein complexes (Beloin et al., 2002; McKenna et al., 2003). In particular, VirB activates, among others, the genes coding for the *Shigella* T3SS (Mxi and Spa proteins), for its early effectors and their chaperones (Ipa and Ipg proteins), and for MxiE, the last regulator in the cascade (Parsot, 2005; Schroeder and Hilbi, 2008). MxiE is another AraC-like protein (Kane et al., 2002; Mavris et al., 2002) which positively controls the expression of the late effectors but it becomes available as activator only when the IpgC chaperone facilitates its release from an anti-activator complex (OspD/Spa15).

The role of VirF as positive activator was first evidenced by observing that in *S. flexneri* (Sakai et al., 1988) and in *S. sonnei* (Kato et al., 1989) the expression of the four Ipa antigen proteins and of IcsA (known also as VirG) is silenced in the absence of VirF and increased when VirF is over expressed. It was then demonstrated (Adler et al., 1989; Tobe et al., 1991) that the regulatory activity of VirF is exerted directly when it interacts with the *icsA* gene, while it occurs indirectly, i.e., through the mediation of VirB, on the *ipa* operons. This gave the first basis for the classic cascade model. Originally, the regulatory role of VirF remained unclear and the protein was simply regarded as the element responsible for the capacity of *Shigella* to bind Congo red, a phenotype (Pcr<sup>+</sup>) associated with the expression of virulence (Sakai et al., 1986a). Initial molecular analyses of the pINV had led to the identification of an about 1 kb region responsible for the Pcr<sup>+</sup> phenotype in *E. coli*. This region turned out to be located within the F fragment of a *Sall* pINV restriction digests and contained a gene, hence named *virF*, which was shown to be essential but not sufficient for the full expression of virulence in *S. flexneri* (Sakai et al., 1986a,b). Additional support to the active role played by VirF as regulator banks on the observations that its intracellular concentration is strictly related to the expression of virulence genes, that a VirF threshold level is required for the activation of the second regulator, *virB*, and that overexpression of *virF* fully restores the invasive phenotype at non-permissive temperature or in pINV-integrated strains (Adler et al., 1989; Dagberg and Uhlin, 1992; Colonna et al., 1995; Prosseda et al., 1998; Durand et al., 2000).

Cloning and sequencing analyses of the pINV plasmid have revealed that, in contrast to *virB* which is localized within the 31 kb PAI, the *virF* gene constitutes a “desert island” located 60 kb outside the PAI and surrounded by a mosaic of IS sequences (Buchrieser et al., 2000; Venkatesan et al., 2001; Prosseda et al., 2006), suggesting that it might have been acquired by the plasmid genome independently from the “entry region” (Prosseda et al., 2002). Besides its position, one of the striking features of the *virF* sequence is its low GC content—about 30%; interestingly this is true also for other virulence-related genes in the 31 kb PAI—as compared to that of the whole pINV plasmid and of the *Shigella* chromosome, which are about 48 and 50%, respectively (Buchrieser et al., 2000; Venkatesan et al., 2001; Yang et al., 2005). The *virF* gene exhibits canonical −35 and −10 regulatory boxes, however no clear ribosome binding site is found. While attempting to purify the protein it has been observed that VirF

exists in three different forms: 30, 27, and 21 kDa (Sakai et al., 1986b; Kato et al., 1989). The larger form has been considered the active one while as yet no functional activity has been attributed to the two minor forms.

Strong evidence that VirF acts as a DNA binding protein capable to activate *virB* transcription by interacting with the upstream region of the *virB* promoter comes from the studies of Tobe et al. (1993). By deletion analysis of the *virB* promoter the authors have shown that the activation of *virB* requires a region of 110 bp upstream the transcription start site. The relevance of this region has been confirmed in the same study by footprinting with a MalE—VirF fusion protein. This allowed to identify a VirF binding site spanning from −117 to −17 which, by *in vitro* transcription, proved to be essential for the activation of *virB*. Considering the position of the VirF binding site and its role as activator, it has been proposed that VirF could act either by recruiting RNA polymerase at the *virB* promoter or by improving the ability of RNA polymerase to form an open complex (Tobe et al., 1993). The positive effect of VirF on *virB* transcription is counteracted by H-NS which has been shown to occupy a *virB* promoter region (spanning from −20 to +20) encompassing the RNA polymerase binding site (Tobe et al., 1993). Since the H-NS and VirF binding sites at the *virB* promoter are contiguous, it has been speculated that the binding of VirF may disrupt the H-NS-DNA repression complex, possibly being favored by thermally-induced changes in local supercoiling (Tobe et al., 1993). The relevance of superhelicity *per se* is supported by the finding that *virB* activation by VirF occurs much more efficiently on supercoiled templates than on relaxed ones (Tobe et al., 1993).

As mentioned before, also the *icsA* gene is controlled by VirF (Sakai et al., 1988). This gene encodes a protein responsible for the motility of *Shigella* and is located on the pINV plasmid outside the “entry region” (Bernardini et al., 1989; Lett et al., 1989). In contrast to all other structural genes involved in the invasive process, *icsA* does not require VirB (Figure 1). As is the case of *virB*, also *icsA* is repressed by the binding of H-NS to its regulatory region (Prosseda et al., 1998). This occurs at three sites (H-NS I, H-NS II, and H-NS III) and severely reduces *icsA* transcription at 30°C (Tran et al., 2011). Recent evidence shows that the regulation of *icsA* also depends on RnaG, a small antisense RNA transcribed on the complementary strand of *icsA* (Giangrossi et al., 2010). RnaG downregulates *icsA* transcription by means of two independent mechanisms, transcriptional interference, and transcriptional attenuation. In the first case the expression of RnaG decreases the activity of the *icsA* promoter. In the second case RnaG causes a premature termination of the *icsA* transcript (Giangrossi et al., 2010).

Experiments performed with purified VirF indicate that this protein is able to directly stimulate the expression of *icsA* and that it binds to four 40–60 bp sites, two of which overlap the *icsA* and RnaG promoters (Tran et al., 2011). This region hosts H-NS sites I and II. The relative position of binding sites for H-NS and VirF provides a physical basis for a possible functional competition between the two proteins. By monitoring the *icsA* promoter activity in the presence of both H-NS and VirF it has been shown that VirF is able to significantly counteract the H-NS-dependent inhibition at the *icsA* promoter, thus acting as

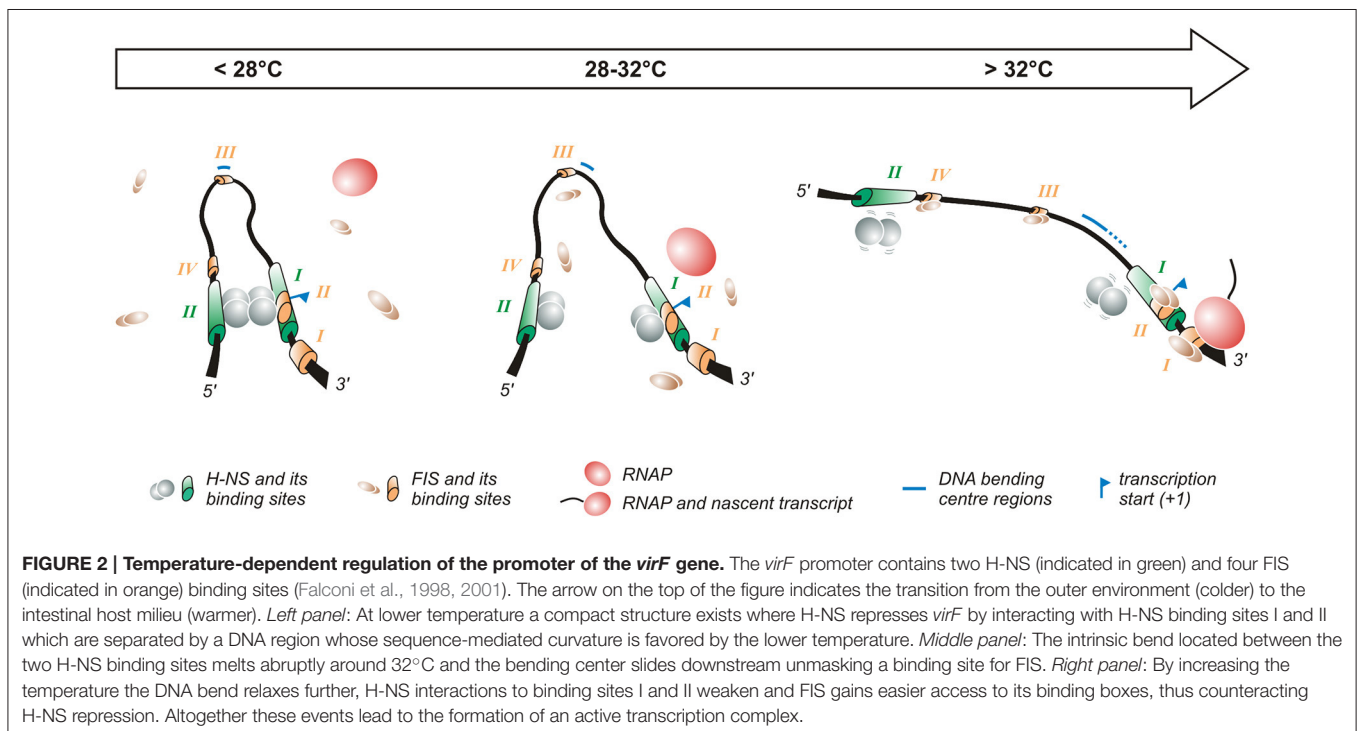
an anti-H-NS protein. VirF expression increases rapidly when temperature is raised above 32°C (Falconi et al., 1998; Prosseda et al., 1998; Durand et al., 2000; **Figure 2**). It is possible to speculate that at the host temperature (37°C) the increased amount of VirF may facilitate the interaction with its sites and consequently disrupt the H-NS-DNA complexes by forming a putative H-NS-*icsA*-VirF intermediate able to promote the switch from a repressed to an active state (Tran et al., 2011). Besides inducing *icsA* activation, VirF is also able to repress the expression of RnaG (**Figure 1**) by binding to a specific site overlapping the RnaG promoter. Thus, VirF promotes *icsA* expression in two ways: by acting on *icsA* both directly and indirectly, i.e., via repression of RnaG transcription (Tran et al., 2011). Altogether the experimental evidence existing on the role of VirF as major regulator in the cascade of events regulating the virulence of *Shigella* highlights the antagonism between H-NS and AraC proteins, suggesting that this may be a common evolutionary strategy that pathogens adopt to control virulence genes (Egan, 2002; Dorman, 2004).

## THE REGULATION OF THE *virF* GENE

The switch from the non-invasive to the invasive phenotype in *Shigella* is an elaborate process and the complexity of the regulatory mechanisms the *virF* gene undergoes is therefore not surprising. The expression of *virF* is affected by different environmental signals that act through highly diversified mechanisms. Nucleoid-associated proteins are known for their contribution to the transcriptional control of several genes including virulence genes (Rimsky and Travers, 2011). This is

well exemplified in the case of *virF* (Prosseda et al., 2002) where H-NS (Falconi et al., 1998), FIS (Falconi et al., 2001), and IHF (Porter and Dorman, 1997b) are deeply involved as regulative elements.

The first studies aimed at understanding the temperature-regulated expression of the invasion phenotype of *Shigella* demonstrate that H-NS, originally designated *virR* (Maurelli and Sansonetti, 1988; Hromockyj et al., 1992), is responsible for silencing invasive genes at 30°C. Indeed, *hns*-defective mutants were shown to express the virulence determinants also at non-permissive temperature. Successive reports have revealed that the thermo-dependent expression of *virF* is lost in a *Shigella* *hns*-defective background, providing clear evidence that H-NS is able to affect the regulation of the *virF* gene (Dagberg and Uhlin, 1992; Colonna et al., 1995) and that this capacity is based on the ability of H-NS to bind the *virF* promoter with strong specificity (Prosseda et al., 1998). A detailed analysis (Falconi et al., 1998) of how H-NS interacts with the *virF* gene indicates that, both *in vivo* and *in vitro*, this protein is able to recognize and repress *virF* only below a critical temperature (32°C). This temperature-dependence relies on the interaction of H-NS with two binding sites within the *virF* promoter, spaced by a DNA linker region. The accessibility of the target DNA to H-NS varies significantly with temperature and H-NS is able to recognize its binding boxes only at lower temperature: with increasing temperature H-NS binding decreases, and at a temperature higher than 30°C the *virF* promoter becomes insensitive to H-NS repression. The position of the two H-NS binding sites on the *virF* promoter has been investigated in detail (Falconi et al., 1998). They have been mapped through *in vivo* and *in vitro* footprinting experiments and are localized around positions





–250 (binding site II) and –1 (binding site I) with respect to the *virF* transcription start site. Site I overlaps the canonical –35 and –10 regulatory elements, suggesting that this region is involved in the transcriptional repression of the *virF* gene by H-NS: the interaction of H-NS with DNA may prevent the –35 and –10 boxes to be accessed by RNA polymerase (Figure 2). Binding site II maps far upstream the –35 and –10 sequences. In the absence of this site, H-NS forms a very unstable nucleoprotein complex, which cannot compete with RNA polymerase effectively. The DNA linker region separating the two H-NS binding sites is endowed with sequence-mediated curvature, a feature whose amplitude is inversely dependent on temperature (Prosseda et al., 2004, 2010). This DNA bend progressively relaxes with increasing temperature and is rapidly abolished above 32°C (Figure 2).

The fact that H-NS binding to the *virF* promoter, bending of the promoter region and repression of *virF* transcription are all sharply decreased by raising the temperature above a critical threshold suggests that the physical basis for the thermoregulated expression of *virF* resides in a temperature-dependent structural transition of the *virF* promoter, with the curved DNA tract within the promoter operating essentially as a thermosensor (Falconi et al., 1998). According to this model, with increasing temperature the transcriptionally inactive DNA architecture prevailing at low temperature would be replaced by a more relaxed geometry which no longer hinders the formation of a functional transcription complex (Falconi et al., 1998). This view has received further support by successive experiments (Prosseda et al., 2004) where the decreased DNA curvature of the *virF* promoter induced by raising the temperature is mimicked, at constant temperature, by templates (obtained by a targeted *in vitro* mutagenesis procedure) endowed with decreasing intrinsic DNA curvature. The results of footprinting and *in vivo* transcription assays using these templates highlight a strict correlation between intrinsic DNA curvature and the capability of H-NS to bind to its target sites and repress *virF* transcription. Moreover, the reciprocal rotational position of the two H-NS sites appears crucial for the thermoregulation of *virF*: moving the sites by about a half helix turn further apart produces templates which, despite a comparable overall bending level, are almost insensitive to H-NS repression, likely because the protein can no longer form a tight, transcription-blocking nucleoprotein complex. Temperature does not only affect the amplitude of the DNA bend within the *virF* promoter but also the position of the bending center, identified by means of circular permutation assays (Prosseda et al., 2004): at 4°C this spot maps at position –137 but with increasing temperature it slides downstream (up to position –55 at 60°C) within the region bounded by the two H-NS sites (Figure 2).

The complexity of the interaction between H-NS and bent DNA is increased further by the participation of FIS (Falconi et al., 2001). In agreement with the well-known growth phase dependent expression of FIS (Rimsky and Travers, 2011), the effect of this protein is higher in early exponential phase (Falconi et al., 2001), suggesting that it contributes to rapidly increase *virF* expression once bacteria have entered the host environment. FIS

exerts, both *in vivo* and *in vitro*, a direct positive control on the transcription of *virF* and is able to bind to the *virF* promoter. Four FIS binding sites (I to IV), centered around positions +55 (site I), –1 (site II), –130 (site III, almost coinciding with the –137 position of the DNA bending center), and –200 (site IV), have been identified in the *virF* promoter region (Falconi et al., 2001). The interaction of FIS with site II likely hampers H-NS binding to its boxes. The downstream sliding of the DNA bending center and the concomitant curvature decrease occurring with increasing temperature favor the binding of FIS to the other sites inducing an adjustment to the geometry of the DNA that hinders long range H-NS/H-NS interaction, thus promoting the transcription of *virF*. Several reports have stressed the validity of this model—envisaging an environmentally induced structural collapse of a promoter's intrinsic bend as a regulative key—also among other pathogenic bacteria, like *Y. enterocolitica* *virF* (Rohde et al., 1999), and *E. coli* *hly* (Madrid et al., 2002).

Besides FIS also IHF is involved in the positive control of the pINV regulon by stimulating the transcription of the *virF* and *virB* genes (Porter and Dorman, 1997b). IHF is composed by two subunits encoded by the *ihfA* and *ihfB* genes (Rimsky and Travers, 2011). IHF binds to the *virF*, *virB*, and *icsA* promoters. In the case of *virF* it recognizes a 127 bp fragment spanning from the promoter-proximal sequence to the start of the ORE, with a putative binding site located between +45 and +57. In *ihfA*-defective strains entering stationary phase the expression of the Ipa proteins and of the T3SS components is decreased as a consequence of the reduced level of VirF and VirB (Porter and Dorman, 1997b). This growth-phase-dependent effect is consistent with the increased level of IHF during late growth phases (Rimsky and Travers, 2011). However, the observation that in a *ihfA* mutant virulence gene expression is reduced only two- to three-fold suggests that IHF, rather than constituting an absolute requirement, plays a modulatory role.

The expression of the *Shigella* invasive genes is repressed at low pH and low osmolarity (Porter and Dorman, 1997a). Also in these cases regulation occurs mainly at the level of *virF* expression. While it has been demonstrated that H-NS is involved, at least in part, in the repression of *virF* under low osmotic conditions (Mitobe et al., 2009), the pH-dependent regulation of *virF* requires the two component regulatory system CpxA/CpxR (Nakayama and Watanabe, 1995, 1998; Figure 1). The Cpx system responds to a broad range of stimuli including, besides pH, also salt, metals, lipids, and misfolded proteins that cause perturbations in the bacterial envelope (Raivio, 2014). In particular, it has been shown that the response regulator CpxR binds directly to a fragment containing the upstream promoter region of *virF* between positions –103 and –37. Phosphorylation of CpxR enhances its binding capacity and directly activates the transcription of *virF* (Nakayama and Watanabe, 1998). Due to the relevance of the CpxA/CpxR system in response to several stimuli connected to envelope stress, the integration of *virF* within the CpxR-controlled network represents a further regulatory layer acting on the fine control of VirF.



Besides being subject to a complex transcriptional control the *virF* gene is also regulated at the translational level. It has been shown that post-transcriptional modifications of tRNA affect the translation of VirF (Figure 1). The full expression of *S. flexneri* virulence genes depends on the presence of two modified tRNA nucleosides (queuosine, at position 34, and 2-methylthio-N<sup>6</sup>-isopentenyladenosine, adjacent to 3'-end of the anticodon). The synthesis of these nucleosides depends on the products of the *tgt* (encoding TGT, the tRNA-guanine transglycosylase) and *miaA* (coding for MiaA, a tRNA-N<sup>6</sup>-isopentenyladenosine synthetase) genes, respectively. The intracellular concentration of VirF decreases in *tgt* mutants and in *miaA* mutants, inducing a low virulence phenotype and an avirulent phenotype, respectively (Durand et al., 1994, 1997). Overall, tRNA modifications are required for VirF to reach the threshold level necessary to activate the virulence cascade (Durand et al., 2000). Besides being involved in tRNA modification TGT also recognizes (Hurt et al., 2007) the *virF* mRNA *in vitro* and this recognition results in a site-specific modification at position 421 responsible for the change from guanine to adenine in the mRNA. As yet the physiological significance of this modification is unclear. A contribution of tRNA modification enzymes in virulence has been reported also in other systems: the GidA and MnmE enzymes, which act together as a complex, are both required for the expression of the full virulence phenotype in *Salmonella enterica*, *Aeromonas hydrophila*, and in the plant pathogen *Pseudomonas syringae* (Shippy and Fadle, 2014). Moreover, the GidA/MnmE complex has been shown to be relevant for the stress response in *Salmonella* and MiaA is essential in *E. coli* for growth at higher temperature (Tsui et al., 1996).

Finally, a connection between bacterial metabolism and virulence gene expression has been found in *S. flexneri*, observing that the addition of ornithine to minimal medium reduces *virF* expression while the addition of putrescine, lysine and few other aminated metabolites counteract this inhibition (Durand and Björk, 2003, 2009). Moreover, it has been observed that proteins involved in the glycolytic pathway, most notably the carbon regulator CsrA, are required for the full expression of the *Shigella* virulence phenotype: indeed, the inactivation of *csrA* causes a significant reduction of *virF* and *virB* expression (Gore and Payne, 2010). As yet the molecular mechanism adopted by CsrA to control these genes is unknown.

## VirF AS GLOBAL REGULATOR

The evolution of bacterial pathogens from harmless ancestors mainly depends on the acquisition of virulence gene clusters on plasmids, phages, and pathogenicity islands by HGT. This process is complemented by the progressive adaptation to a specific niche determined by so-called pathoadaptive events such as mutations, rearrangements or deletions of genes unnecessary, or even deleterious, for optimal fitness to the new environment (Ochman and Moran, 2001). These events usually involve the concomitant acquisition or loss of regulatory factors which modify the transcriptional profile of

the host to a significant extent. The evolution of *Shigella* from *E. coli* is a most studied exemplification of these events (Prosseda et al., 2012).

Global transcriptional analyses of *E. coli* cells expressing or lacking the *virF* gene have contributed to understand to what extent the arrival of VirF by the acquisition of the pINV plasmid in *Shigella* has altered the transcriptional program of the ancestor. These studies have revealed that the activity of VirF is not restricted to the regulation of plasmid-encoded virulence genes as *icsA* and *virB* but extends also to chromosomally-located genes (Barbagallo et al., 2011). Genes activated by VirF can be grouped into two categories: those which are functional both in *Shigella* and *E. coli*, and those that have been inactivated in *Shigella*. Among the more VirF-sensitive genes of the first group there are those encoding the heat shock proteins IbpA, GroESL, HtpG, DnaK, and Lon. Given the protective role exerted by these stress proteins, it can be hypothesized that the functional significance of VirF resides in activating genes which contribute to better withstand adverse conditions inside the host.

The existence of VirF-regulated genes silent in *Shigella* (i.e., the second group) suggests that some of them may encode factors perturbing the invasive process which likely have been silenced during evolution in order to optimize bacterial fitness within the host. Most of these genes encode yet unknown products. In this respect the *speG* gene, encoding the enzyme spermidine acetyl transferase (SAT), is an exception. In *E. coli*, SAT prevents the accumulation of spermidine (and the consequent toxic effects) by catalyzing the conversion of this polyamine into its physiologically inert form, acetylspermidine (Barbagallo et al., 2011). The *speG* gene belongs to the *ynfB-speG* operon (no function is as yet known for the *ynfB* product), whose induction by VirF has been observed only at 37°C (Barbagallo et al., 2011) in agreement with the thermodependency of *virF* expression (Falconi et al., 1998). The loss of *speG* constitutes a pathoadaptive mutation (Prosseda et al., 2012; Campilongo et al., 2014) since spermidine accumulation induced by the lack of SpeG activity increases *Shigella* survival inside macrophages and its resistance to oxidative stress conditions (Barbagallo et al., 2011). These data have led to hypothesize that, during the evolutionary transition from *E. coli* to *Shigella*, the acquisition of *virF* on the pINV plasmid might have caused an increased expression of *speG*, thus lowering the intracellular spermidine content (Di Martino et al., 2013). Evidence supporting this view stems from the observation that in *Shigella* *speG* expression is decreased in a VirF-defective background and that restoration of *speG* expression lowers the bacterial fitness within macrophages (Barbagallo et al., 2011).

Recently VirF has been shown to be involved in the regulation of *mdtJI*, another operon related to polyamine function in bacteria. In *E. coli* the *mdtJI* operon, which encodes an efflux pump belonging to the small multidrug resistance family of transporters, is almost silent because of a strong repression by H-NS (Leuzzi et al., 2015). Despite the high homology between *Shigella* and *E. coli*, in *Shigella* the *mdtJI* pump is expressed because the H-NS repressive effect is counteracted by VirF and favored by high levels of spermidine (Leuzzi et al., 2015). Intracellular spermidine does not affect the synthesis of VirF (Leuzzi et al., 2015) and the molecular mechanism by which

this polyamine contributes towards the activation of *mdtJI* is still unclear. Genetic studies on the expression of *mdtJI* in *Shigella* indicate the presence of a VirF binding site around the promoter consensus boxes which partially overlaps an H-NS site, supporting the occurrence of a competition between the two proteins (Leuzzi et al., 2015), in close analogy with the observations on the *virB* and *icsA* promoters (Tobe et al., 1993; Tran et al., 2011). In *Shigella* MdtJI promotes the secretion of putrescine, the precursor of spermidine, and it has been proposed that VirF- and spermidine-mediated activation of the *mdtJI* operon represents a safety mechanism allowing spermidine to accumulate within the bacterial cell to a level optimally suited for survival within infected macrophages and, at the same time, avoid toxic side effects on bacterial viability due to spermidine excess (Leuzzi et al., 2015). Altogether, the observations on the ability of *Shigella* to activate chromosomal genes evidence the extent to which the acquisition of a new regulator by HGT represents a crucial event for reshaping the transcriptional profile of the core genome, facilitating bacterial adaptation to specific niches within infected hosts.

## VirF AS NOVEL TARGET FOR ANTI-VIRULENCE THERAPIES

Each year *Shigella* is responsible for 125 million cases of illness, mainly in low income countries (The et al., 2016). Despite the enormous clinical relevance of these infections and the emergence of multiresistance strains, no vaccine has been as yet released for public use (Anderson et al., 2016). Several recent studies have focused on the development of novel treatment strategies targeting virulence instead of bacterial viability, since this is regarded as a highly effective approach to combat bacterial infections while minimizing the emergence of antibiotic resistances (Rasko and Sperandio, 2010). The expression of virulence factors is not required for cell viability and therefore bacterial pathogens should be subject to less selective pressure to develop resistance to inhibitors of virulence determinants. AraC proteins are considered very interesting candidate targets in anti-virulence strategies due to their critical role in controlling virulence in pathogenic bacteria (Rasko and Sperandio, 2010). Specific inhibitors can affect AraC-mediated processes at different stages, such as self-association, DNA binding, and recruitment of RNA polymerase.

As for shigellosis it is therefore not surprising that VirF has been considered a very good antivirulence target, since its silencing prevents host cell invasion and intercellular spreading without affecting bacterial viability (Sakai et al., 1988; Falconi et al., 1998). So far, two approaches have been adopted to identify potential VirF inhibitors: a high-throughput screening for small molecules and a targeted search among already characterized AraC inhibitors. In the first case (Hurt et al., 2010), the expression of *virB* in response to VirF has been analyzed using a pINV-cured *S. flexneri* strain harboring a recombinant plasmid containing the *virF* gene and a *virB-lacZ* fusion. The screening, performed initially on 42,000 compounds

from several small-molecule libraries and then extended to an additional set of 100,000 compounds, has led to the identification of about 600 molecules meeting the selection criteria for VirF inhibition (Hurt et al., 2010; Emanuele et al., 2014). After selecting for candidates with favorable medicinal chemistry, low toxicity and a dose-dependent activity, five compounds able to inhibit VirF-driven transcriptional activation with very low IC<sub>50</sub> values were considered for further studies. These compounds were characterized by the presence of aromatic or heterocyclic rings that could interact with DNA or with aromatic acids that typically associate with DNA (Hurt et al., 2010; Emanuele et al., 2014). *In vivo* assays on cell cultures have shown that only one polycyclic compound, named 19615 (methyl-[2-(2-phenyl-4a,9b-dihydro-benzo[4,5]furo[3,2-d]pyrimidin-4-yl)oxy-ethyl]-amine), was able to severely affect intercellular spreading and induce a significant inhibition of cell invasion (Emanuele and Garcia, 2015). In particular this molecule has been demonstrated to be able to directly interact with a MalE—VirF fusion protein and inhibit its binding to the *virB* promoter. Further studies to structurally optimize the selected compound are required to fully clarify its effectiveness in an antivirulence therapy.

The second approach (Koppolu et al., 2013) adopted so far to identify possible VirF inhibitors has been centered on a small molecule, SE-1, a quinoline derivative [1-butyl-4-nitromethyl-3-(quinolin-yl)-4Hquinoline] previously identified as an inhibitor of two AraC activators, RhaS, and RhaR (Skredenske et al., 2013). Since SE-1 interacts with the conserved DNA-binding domains of the AraC proteins (Skredenske et al., 2013), it has been considered able to potentially inhibit also other AraC activators. This turned out to be true: the data revealed that SE-1 induced a significant reduction in the expression of all VirF-controlled genes, consequently inhibiting the invasion of epithelial cells. SE-1 binds directly to a MalE—VirF fusion protein, a feature likely responsible for inhibiting VirF to interact with DNA and activate transcription. On account that SE-1 does not affect the growth of *Shigella* and does not have detectable toxicity in human cell cultures, it has been considered as another good candidate as novel antibacterial agent.

## CONCLUSIONS AND PERSPECTIVES

Altogether the available data stress how *Shigella*, in its long route to becoming a successful pathogen, has evolved an elaborate regulatory system to ensure the coordinated activation of virulence determinants or prevent their wasteful expression, depending on environmental signals. The complexity of the circuitry regulating virulence in *Shigella* highlights the relevance of the firing of a major regulator, the *virF* gene, to the successful development of the invasive program. Through studies analyzing the interplay, either synergistic or antagonistic, among nucleoid-associated proteins, sRNA, global, and specific regulators, and intrinsic features of the DNA, the complex nature of the regulation of VirF and of the genes under its control is emerging with intriguing detail. However, while the master regulator has

been identified almost three decades ago, several open questions still exist, such as the capacity of VirF to form dimers or more complex aggregates and the molecular mechanisms underlying its DNA binding specificity and its interactions with RNA polymerase. Understanding how the gene regulatory circuitry has evolved in bacterial pathogens represents a challenge. From the evolutionary standpoint this relates to the need to understand how genes acquired by HGT have integrated into existing regulatory networks and how newly acquired regulators have shaped the genome of the new bacterial host. In the long run a better understanding of the structure of the VirF protein can be expected to positively impact on the development of new therapeutic approaches based on the use of specific inhibitory compounds.

## REFERENCES

- Adler, B., Sasakawa, C., Tobe, T., Makino, S., Komatsu, K., and Yoshikawa, M. (1989). A dual transcriptional activation system for the 230 kb plasmid genes coding for virulence-associated antigens of *Shigella flexneri*. *Mol. Microbiol.* 3, 627–635. doi: 10.1111/j.1365-2958.1989.tb00210.x
- Anderson, M., Sansonetti, P. J., and Marteyn, B. S. (2016). *Shigella* diversity and changing landscape: insights for the twenty-first century. *Front. Cell. Infect. Microbiol.* 6:45. doi: 10.3389/fcimb.2016.00045
- Ashida, H., Mimuro, H., and Sasakawa, C. (2015). *Shigella* manipulates host immune responses by delivering effector proteins with specific roles. *Front. Immunol.* 6:219. doi: 10.3389/fimmu.2015.00219
- Ashida, H., and Sasakawa, C. (2015). *Shigella* IpaH family effectors as a versatile model for studying pathogenic bacteria. *Front. Cell. Infect. Microbiol.* 5:100. doi: 10.3389/fcimb.2015.00100
- Barbagallo, M., Di Martino, M. L., Marcocci, L., Pietrangeli, P., De Carolis, E., Casalino, M., et al. (2011). A new piece of the *Shigella* Pathogenicity puzzle: spermidine accumulation by silencing of the *speG* gene. *PLoS ONE* 6:e27226. doi: 10.1371/journal.pone.0027226
- Beloin, C., McKenna, S., and Dorman, C. J. (2002). Molecular dissection of VirB, a key regulator of the virulence cascade of *Shigella flexneri*. *J. Biol. Chem.* 277, 15333–15344. doi: 10.1074/jbc.M111429200
- Bernardini, M. L., Mounier, J., D'Hauteville, H., Coquis-Rondon, M., and Sansonetti, P. J. (1989). Identification of *icsA*, a plasmid locus of *Shigella flexneri* that governs bacterial intra- and intercellular spread through interaction with F-actin. *Proc. Natl. Acad. Sci. U.S.A.* 86, 3867–3871. doi: 10.1073/pnas.86.10.3867
- Bourgerie, S. J., Michán, C. M., Thomas, M. S., Busby, S. J., and Hyde, E. I. (1997). DNA binding and DNA bending by the MelR transcription activator protein from *Escherichia coli*. *Nucleic Acids Res.* 25, 1685–1693. doi: 10.1093/nar/25.9.1685
- Buchrieser, C., Glaser, P., Rusniok, C., Nedjari, H., D'Hauteville, H., Kunst, F., et al. (2000). The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. *Mol. Microbiol.* 38, 760–771. doi: 10.1046/j.1365-2958.2000.02179.x
- Campilongo, R., Di Martino, M. L., Marcocci, L., Pietrangeli, P., Leuzzi, A., Grossi, M., et al. (2014). Molecular and functional profiling of the polyamine content in enteroinvasive *E. coli*: looking into the gap between commensal *E. coli* and harmful *Shigella*. *PLoS ONE* 9:e106589. doi: 10.1371/journal.pone.0106589
- Cases, I., and de Lorenzo, V. (2005). Promoters in the environment: transcriptional regulation in its natural context. *Nat. Rev. Microbiol.* 3, 105–118. doi: 10.1038/nrmicro1084
- Colonna, B., Casalino, M., Fradiani, P. A., Zagaglia, C., Naitza, S., Leoni, L., et al. (1995). H-NS regulation of virulence gene expression in enteroinvasive *Escherichia coli* harboring the virulence plasmid integrated into the host chromosome. *J. Bacteriol.* 177, 4703–4712.
- Dagberg, B., and Uhlin, B. E. (1992). Regulation of virulence-associated plasmid genes in enteroinvasive *Escherichia coli*. *J. Bacteriol.* 174, 7606–7612.
- Di Martino, M. L., Campilongo, R., Casalino, M., Micheli, G., Colonna, B., and Prosseda, G. (2013). Polyamines: emerging players in bacteria-host interactions. *Int. J. Med. Microbiol.* 303, 484–491. doi: 10.1016/j.ijmm.2013.06.008
- Dorman, C. J. (2004). H-NS: a universal regulator for a dynamic genome. *Nat. Rev. Microbiol.* 2, 391–400. doi: 10.1038/nrmicro883
- Dorman, C. J. (2007). H-NS, the genome sentinel. *Nat. Rev. Microbiol.* 5, 157–161. doi: 10.1038/nrmicro1598
- Dorman, C. J., and Porter, M. E. (1998). The *Shigella* virulence gene regulatory cascade: a paradigm of bacterial gene control mechanisms. *Mol. Microbiol.* 29, 677–684.
- Durand, J. M. B., and Björk, G. R. (2003). Putrescine or a combination of methionine and arginine restores virulence gene expression in a tRNA modification-deficient mutant of *Shigella flexneri*: a possible role in adaptation of virulence. *Mol. Microbiol.* 47, 519–527. doi: 10.1046/j.1365-2958.2003.03314.x
- Durand, J. M. B., and Björk, G. R. (2009). Metabolic control through ornithine and uracil of epithelial cell invasion by *Shigella flexneri*. *Microbiology* 155(Pt 8), 2498–2508. doi: 10.1099/mic.0.028191-0
- Durand, J. M., Björk, G. R., Kuwae, A., Yoshikawa, M., and Sasakawa, C. (1997). The modified nucleoside 2-methylthio-N6-isopentenyladenosine in tRNA of *Shigella flexneri* is required for expression of virulence genes. *J. Bacteriol.* 179, 5777–5782.
- Durand, J. M., Dagberg, B., Uhlin, B. E., and Björk, G. R. (2000). Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of *Shigella flexneri*: the expression of the *virF* gene. *Mol. Microbiol.* 35, 924–935. doi: 10.1046/j.1365-2958.2000.01767.x
- Durand, J. M., Okada, N., Tobe, T., Watarai, M., Fukuda, I., Suzuki, T., et al. (1994). *vacC*, a virulence-associated chromosomal locus of *Shigella flexneri*, is homologous to *tgt*, a gene encoding tRNA-guanine transglycosylase (Tgt) of *Escherichia coli* K-12. *J. Bacteriol.* 176, 4627–4634.
- Egan, S. M. (2002). Growing repertoire of AraC/XylS activators. *J. Bacteriol.* 184, 5529–5532. doi: 10.1128/JB.184.20.5529-5532.2002
- Emanuele, A. A., Adams, N. E., Chen, Y.-C., Maurelli, A. T., and Garcia, G. A. (2014). Potential novel antibiotics from HTS targeting the virulence-regulating transcription factor, VirF, from *Shigella flexneri*. *J. Antibiot.* 67, 379–386. doi: 10.1038/ja.2014.10.
- Emanuele, A. A., and Garcia, G. A. (2015). Mechanism of action and initial, *in vitro* sar of an inhibitor of the *Shigella flexneri* virulence regulator VirF. *PLoS ONE* 10:e0137410. doi: 10.1371/journal.pone.0137410
- Falconi, M., Colonna, B., Prosseda, G., Micheli, G., and Gualerzi, C. O. (1998). Thermoregulation of *Shigella* and *Escherichia coli* EIEC pathogenicity. A temperature-dependent structural transition of DNA modulates accessibility of *virF* promoter to transcriptional repressor H-NS. *EMBO J.* 17, 7033–7043. doi: 10.1093/emboj/17.23.7033
- Falconi, M., Prosseda, G., Giangrossi, M., Beghetto, E., and Colonna, B. (2001). Involvement of FIS in the H-NS-mediated regulation of *virF* gene of *Shigella* and enteroinvasive *Escherichia coli*. *Mol. Microbiol.* 42, 439–452. doi: 10.1046/j.1365-2958.2001.02646.x

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication. In particular MLDM, BC and GP conceived the review and wrote the manuscript. MLDM and GM conceived the figures. MF and GM critically edited the final version of the review.

## FUNDING

This research was supported by grants from Ministero della Ricerca e dell'Istruzione (PRIN 2012/WWJSX8K), Sapienza Università di Roma, Consiglio Nazionale delle Ricerche, and Institut Pasteur (PTR-24-16).



- Gallegos, M. T., Schleif, R., Bairoch, A., Hofmann, K., and Ramos, J. L. (1997). AraC/XylS family of transcriptional regulators. *Microbiol. Mol. Biol. Rev.* 61, 393–410.
- Giangrossi, M., Prosseda, G., Tran, C. N., Brandi, A., Colonna, B., and Falconi, M. (2010). A novel antisense RNA regulates at transcriptional level the virulence gene *icsA* of *Shigella flexneri*. *Nucleic Acids Res.* 38, 3362–3375. doi: 10.1093/nar/gkq025
- Gore, A. L., and Payne, S. M. (2010). CsrA and Cra influence *Shigella flexneri* pathogenesis. *Infect. Immun.* 78, 4674–4682. doi: 10.1128/IAI.00589-10
- Hromockyj, A. E., Tucker, S. C., and Maurelli, A. T. (1992). Temperature regulation of *Shigella* virulence: identification of the repressor gene *virR*, an analogue of *hns*, and partial complementation by tyrosyl transfer RNA (tRNA<sup>Tyr</sup>). *Mol. Microbiol.* 6, 2113–2124. doi: 10.1111/j.1365-2958.1992.tb01385.x
- Hurt, J. K., McQuade, T. J., Emanuele, A., Larsen, M. J., and Garcia, G. A. (2010). High-throughput screening of the virulence regulator VirF: a novel antibacterial target for shigellosis. *J. Biomol. Screen.* 15, 379–387. doi: 10.1177/1087057110362101
- Hurt, J. K., Olgen, S., and Garcia, G. A. (2007). Site-specific modification of *Shigella flexneri* virF mRNA by tRNA-guanine transglycosylase *in vitro*. *Nucleic Acids Res.* 35, 4905–4913. doi: 10.1093/nar/gkm473
- Kane, C. D., Schuch, R., Day, W. A., and Maurelli, A. T. (2002). MxiE regulates intracellular expression of factors secreted by the *Shigella flexneri* 2a type III secretion system. *J. Bacteriol.* 184, 4409–4419. doi: 10.1128/JB.184.16.4409-4419.2002
- Kato, J., Ito, K., Nakamura, A., and Watanabe, H. (1989). Cloning of regions required for contact hemolysis and entry into LLC-MK2 cells from *Shigella sonnei* form I plasmid: *virF* is a positive regulator gene for these phenotypes. *Infect. Immun.* 57, 1391–1398.
- Kim, D. W., Lenzen, G., Page, A.-L., Legrain, P., Sansonetti, P. J., and Parsot, C. (2005). The *Shigella flexneri* effector OspG interferes with innate immune responses by targeting ubiquitin-conjugating enzymes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14046–14051. doi: 10.1073/pnas.0504466102
- Koppolu, V., Osaka, I., Skredenske, J. M., Kettle, B., Hefty, P. S., Li, J., et al. (2013). Small-molecule inhibitor of the *Shigella flexneri* master virulence regulator VirF. *Infect. Immun.* 81, 4220–4231. doi: 10.1128/IAI.00919-13
- Kwon, H. J., Bennik, M. H., Demple, B., and Ellenberger, T. (2000). Crystal structure of the *Escherichia coli* Rob transcription factor in complex with DNA. *Nat. Struct. Biol.* 7, 424–430. doi: 10.1038/75213
- Lett, M. C., Sasakawa, C., Okada, N., Sakai, T., Makino, S., Yamada, M., et al. (1989). *virG*, a plasmid-coded virulence gene of *Shigella flexneri*: identification of the VirG protein and determination of the complete coding sequence. *J. Bacteriol.* 171, 353–359.
- Leuzzi, A., Di Martino, M. L., Campilongo, R., Falconi, M., Barbagallo, M., Marcocci, L., et al. (2015). Multifactor regulation of the MdtJI polyamine transporter in *Shigella*. *PLoS ONE* 10:e0136744. doi: 10.1371/journal.pone.0136744
- Madrid, C., Nieto, J. M., Paytubi, S., Falconi, M., Gualerzi, C. O., and Juárez, A. (2002). Temperature- and H-NS-dependent regulation of a plasmid-encoded virulence operon expressing *Escherichia coli* hemolysin. *J. Bacteriol.* 184, 5058–5066. doi: 10.1128/JB.184.18.5058-5066.2002
- Maurelli, A. T., Baudry, B., d'Hauteville, H., Hale, T. L., and Sansonetti, P. J. (1985). Cloning of plasmid DNA sequences involved in invasion of HeLa cells by *Shigella flexneri*. *Infect. Immun.* 49, 164–171.
- Maurelli, A. T., and Sansonetti, P. J. (1988). Identification of a chromosomal gene controlling temperature-regulated expression of *Shigella* virulence. *Proc. Natl. Acad. Sci. U.S.A.* 85, 2820–2824. doi: 10.1073/pnas.85.8.2820
- Mavris, M., Page, A.-L., Tournebise, R., Demers, B., Sansonetti, P., and Parsot, C. (2002). Regulation of transcription by the activity of the *Shigella flexneri* type III secretion apparatus. *Mol. Microbiol.* 43, 1543–1553. doi: 10.1046/j.1365-2958.2002.02836.x
- McAdams, H. H., Srinivasan, B., and Arkin, A. P. (2004). The evolution of genetic regulatory systems in bacteria. *Nat. Rev. Genet.* 5, 169–178. doi: 10.1038/nrg1292
- McKenna, S., Beloin, C., and Dorman, C. J. (2003). In vitro DNA-binding properties of VirB, the *Shigella flexneri* virulence regulatory protein. *FEBS Lett.* 545, 183–187. doi: 10.1016/S0014-5793(03)00524-6
- Mitobe, J., Morita-Ishihara, T., Ishihama, A., and Watanabe, H. (2009). Involvement of RNA-binding protein Hfq in the osmotic-response regulation of *invE* gene expression in *Shigella sonnei*. *BMC Microbiol.* 9:110. doi: 10.1186/1471-2180-9-110
- Nakayama, S., and Watanabe, H. (1995). Involvement of *cpxA*, a sensor of a two-component regulatory system, in the pH-dependent regulation of expression of *Shigella sonnei* *virF* gene. *J. Bacteriol.* 177, 5062–5069.
- Nakayama, S., and Watanabe, H. (1998). Identification of *cpxR* as a positive regulator essential for expression of the *Shigella sonnei* *virF* gene. *J. Bacteriol.* 180, 3522–3528.
- Njamkepo, E., Fawal, N., Tran-Dien, A., Hawkey, J., Strockbine, N., Jenkins, C., et al. (2016). Global phylogeography and evolutionary history of *Shigella dysenteriae* type 1. *Nat. Microbiol.* 1, 16027. doi: 10.1038/nmicrobiol.2016.27
- Ochman, H., and Moran, N. A. (2001). Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* 292, 1096–1099. doi: 10.1126/science.1058543
- Parsot, C. (2005). *Shigella* spp. and enteroinvasive *Escherichia coli* pathogenicity factors. *FEMS Microbiol. Lett.* 252, 11–18. doi: 10.1016/j.femsle.2005.08.046
- Porter, M. E., and Dorman, C. J. (1997a). Differential regulation of the plasmid-encoded genes in the *Shigella flexneri* virulence regulon. *Mol. Gen. Genet.* 256, 93–103.
- Porter, M. E., and Dorman, C. J. (1997b). Positive regulation of *Shigella flexneri* virulence genes by integration host factor. *J. Bacteriol.* 179, 6537–6550.
- Porter, M. E., and Dorman, C. J. (2002). In vivo DNA-binding and oligomerization properties of the *Shigella flexneri* AraC-like transcriptional regulator VirF as identified by random and site-specific mutagenesis. *J. Bacteriol.* 184, 531–539. doi: 10.1128/JB.184.2.531-539.2002
- Porter, M. E., Mitchell, P., Roe, A. J., Free, A., Smith, D. G. E., and Gally, D. L. (2004). Direct and indirect transcriptional activation of virulence genes by an AraC-like protein, PerA from enteropathogenic *Escherichia coli*. *Mol. Microbiol.* 54, 1117–1133. doi: 10.1111/j.1365-2958.2004.04333.x
- Porter, M. E., Smith, S. G., and Dorman, C. J. (1998). Two highly related regulatory proteins, *Shigella flexneri* VirF and enterotoxigenic *Escherichia coli* Rns, have common and distinct regulatory properties. *FEMS Microbiol. Lett.* 162, 303–309.
- Prosseda, G., Di Martino, M. L., Campilongo, R., Fioravanti, R., Micheli, G., Casalino, M., et al. (2012). Shedding of genes that interfere with the pathogenic lifestyle: the *Shigella* model. *Res. Microbiol.* 163, 399–406. doi: 10.1016/j.resmic.2012.07.004
- Prosseda, G., Falconi, M., Gualerzi, C. O., Micheli, G., and Colonna, B. (2004). The *virF* promoter in *Shigella*: more than just a curved DNA stretch. *Mol. Microbiol.* 51, 523–537. doi: 10.1046/j.1365-2958.2003.03848.x
- Prosseda, G., Falconi, M., Nicoletti, M., Casalino, M., Micheli, G., and Colonna, B. (2002). Histone-like proteins and the *Shigella* invasivity regulon. *Res. Microbiol.* 153, 461–468. doi: 10.1016/S0923-2508(02)01346-3
- Prosseda, G., Fradiani, P. A., Di Lorenzo, M., Falconi, M., Micheli, G., Casalino, M., et al. (1998). A role for H-NS in the regulation of the *virF* gene of *Shigella* and enteroinvasive *Escherichia coli*. *Res. Microbiol.* 149, 15–25. doi: 10.1016/S0923-2508(97)83619-4
- Prosseda, G., Latella, M. C., Casalino, M., Nicoletti, M., Michienzi, S., and Colonna, B. (2006). Plasticity of the P<sub>junc</sub> promoter of ISEc11, a new insertion sequence of the IS1111 family. *J. Bacteriol.* 188, 4681–4689. doi: 10.1128/JB.00332-06
- Prosseda, G., Mazzola, A., Di Martino, M. L., Tielker, D., Micheli, G., and Colonna, B. (2010). A temperature-induced narrow DNA curvature range sustains the maximum activity of a bacterial promoter *in vitro*. *Biochemistry* 49, 2778–2785. doi: 10.1021/bi902003g
- Pupo, G. M., Lan, R., and Reeves, P. R. (2000). Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc. Natl. Acad. Sci. U.S.A.* 97, 10567–10572. doi: 10.1073/pnas.180094797
- Raivio, T. L. (2014). Everything old is new again: an update on current research on the Cpx envelope stress response. *Biochim. Biophys. Acta* 1843, 1529–1541. doi: 10.1016/j.bbamcr.2013.10.018
- Rasko, D. A., and Sperandio, V. (2010). Anti-virulence strategies to combat bacteria-mediated disease. *Nat. Rev. Drug Discov.* 9, 117–128. doi: 10.1038/nrd3013
- Rhee, S., Martin, R. G., Rosner, J. L., and Davies, D. R. (1998). A novel DNA-binding motif in MarA: the first structure for an AraC



- family transcriptional activator. *Proc. Natl. Acad. Sci. U.S.A.* 95, 10413–10418.
- Rimsky, S., and Travers, A. (2011). Pervasive regulation of nucleoid structure and function by nucleoid-associated proteins. *Curr. Opin. Microbiol.* 14, 136–141. doi: 10.1016/j.mib.2011.01.003
- Rohde, J. R., Luan, X. S., Rohde, H., Fox, J. M., and Minnich, S. A. (1999). The *Yersinia enterocolitica* pYV virulence plasmid contains multiple intrinsic DNA bends which melt at 37 degrees C. *J. Bacteriol.* 181, 4198–4204.
- Sakai, T., Sasakawa, C., Makino, S., Kamata, K., and Yoshikawa, M. (1986a). Molecular cloning of a genetic determinant for Congo red binding ability which is essential for the virulence of *Shigella flexneri*. *Infect. Immun.* 51, 476–482.
- Sakai, T., Sasakawa, C., Makino, S., and Yoshikawa, M. (1986b). DNA sequence and product analysis of the *virF* locus responsible for Congo red binding and cell invasion in *Shigella flexneri* 2a. *Infect. Immun.* 54, 395–402.
- Sakai, T., Sasakawa, C., and Yoshikawa, M. (1988). Expression of four virulence antigens of *Shigella flexneri* is positively regulated at the transcriptional level by the 30 kiloDalton *virF* protein. *Mol. Microbiol.* 2, 589–597.
- Sansonetti, P. J., Kopecko, D. J., and Formal, S. B. (1982). Involvement of a plasmid in the invasive ability of *Shigella flexneri*. *Infect. Immun.* 35, 852–860.
- Sasakawa, C., Kamata, K., Sakai, T., Makino, S., Yamada, M., Okada, N., et al. (1988). Virulence-associated genetic regions comprising 31 kilobases of the 230-kilobase plasmid in *Shigella flexneri* 2a. *J. Bacteriol.* 170, 2480–2484.
- Schleif, R. (2010). AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiol. Rev.* 34, 779–796. doi: 10.1111/j.1574-6976.2010.00226.x
- Schroeder, G. N., and Hilbi, H. (2008). Molecular pathogenesis of *Shigella* spp.: controlling host cell signaling, invasion, and death by type III secretion. *Clin. Microbiol. Rev.* 21, 134–156. doi: 10.1128/CMR.00032-07
- Scribano, D., Petrucca, A., Pompili, M., Ambrosi, C., Bruni, E., Zagaglia, C., et al. (2014). Polar localization of PhoN2, a periplasmic virulence-associated factor of *Shigella flexneri*, is required for proper IcsA exposition at the old bacterial pole. *PLoS ONE* 9:e90230. doi: 10.1371/journal.pone.0090230
- Shippy, D. C., and Fadle, A. A. (2014). tRNA modification enzymes GidA and MnmE: potential role in virulence of bacterial pathogens. *Int. J. Mol. Sci.* 15, 18267–18280. doi: 10.3390/ijms151018267
- Skredenske, J. M., Koppolu, V., Kolin, A., Deng, J., Kettle, B., Taylor, B., et al. (2013). Identification of a small-molecule inhibitor of bacterial AraC family activators. *J. Biomol. Screen.* 18, 588–598. doi: 10.1177/1087057112474690
- The, H. C., Thanh, D. P., Holt, K. E., Thomson, N. R., and Baker, S. (2016). The genomic signatures of *Shigella* evolution, adaptation and geographical spread. *Nat. Rev. Microbiol.* 14, 235–250. doi: 10.1038/nrmicro.2016.10
- Tobe, T., Nagai, S., Okada, N., Adler, B., Yoshikawa, M., and Sasakawa, C. (1991). Temperature-regulated expression of invasion genes in *Shigella flexneri* is controlled through the transcriptional activation of the *virB* gene on the large plasmid. *Mol. Microbiol.* 5, 887–893.
- Tobe, T., Yoshikawa, M., Mizuno, T., and Sasakawa, C. (1993). Transcriptional control of the invasion regulatory gene *virB* of *Shigella flexneri*: activation by *virF* and repression by H-NS. *J. Bacteriol.* 175, 6142–6149.
- Tran, C. N., Giangrossi, M., Prosseda, G., Brandi, A., Di Martino, M. L., Colonna, B., et al. (2011). A multifactor regulatory circuit involving H-NS, VirF and an antisense RNA modulates transcription of the virulence gene *icsA* of *Shigella flexneri*. *Nucleic Acids Res.* 39, 8122–8134. doi: 10.1093/nar/gkr521
- Tsui, H. C., Feng, G., and Winkler, M. E. (1996). Transcription of the *mutL* repair, *miaA* tRNA modification, *hfq* pleiotropic regulator, and *hflA* region protease genes of *Escherichia coli* K-12 from clustered Esigma32-specific promoters during heat shock. *J. Bacteriol.* 178, 5719–5731.
- Venkatesan, M. M., Goldberg, M. B., Rose, D. J., Grotbeck, E. J., Burland, V., and Blattner, F. R. (2001). Complete DNA sequence and analysis of the large virulence plasmid of *Shigella flexneri*. *Infect. Immun.* 69, 3271–3285. doi: 10.1128/IAI.69.5.3271-3285.2001
- Yang, F., Yang, J., Zhang, X., Chen, L., Jiang, Y., Yan, Y., et al. (2005). Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery. *Nucleic Acids Res.* 33, 6445–6458. doi: 10.1093/nar/gki954

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Di Martino, Falconi, Micheli, Colonna and Prosseda. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# FapR: From Control of Membrane Lipid Homeostasis to a Biotechnological Tool

Daniela Albanesi\* and Diego de Mendoza

Laboratorio de Fisiología Microbiana, Instituto de Biología Molecular y Celular de Rosario, Consejo Nacional de Investigaciones Científicas y Técnicas, Universidad Nacional de Rosario, Rosario, Argentina

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas Medical Branch,  
USA

### Reviewed by:

Christian Sohlenkamp,  
National Autonomous University of  
Mexico, Mexico  
Fabián Lorenzo,  
University of La Laguna, Spain

### \*Correspondence:

Daniela Albanesi  
albanesi@ibr-conicet.gov.ar

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 14 July 2016

**Accepted:** 21 September 2016

**Published:** 06 October 2016

### Citation:

Albanesi D and de Mendoza D (2016)  
FapR: From Control of Membrane  
Lipid Homeostasis to a  
Biotechnological Tool.  
Front. Mol. Biosci. 3:64.  
doi: 10.3389/fmolb.2016.00064

Phospholipids and fatty acids are not only one of the major components of cell membranes but also important metabolic intermediates in bacteria. Since the fatty acid biosynthetic pathway is essential and energetically expensive, organisms have developed a diversity of homeostatic mechanisms to fine-tune the concentration of lipids at particular levels. FapR is the first global regulator of lipid synthesis discovered in bacteria and is largely conserved in Gram-positive organisms including important human pathogens, such as *Staphylococcus aureus*, *Bacillus anthracis*, and *Listeria monocytogenes*. FapR is a transcription factor that negatively controls the expression of several genes of the fatty acid and phospholipid biosynthesis and was first identified in *Bacillus subtilis*. This review focuses on the genetic, biochemical and structural advances that led to a detailed understanding of lipid homeostasis control by FapR providing unique opportunities to learn how Gram-positive bacteria monitor the status of fatty acid biosynthesis and adjust the lipid synthesis accordingly. Furthermore, we also cover the potential of the FapR system as a target for new drugs against Gram-positive bacteria as well as its recent biotechnological applications in diverse organisms.

**Keywords:** lipid synthesis, FapR, transcriptional regulation, Gram-positive bacteria, *in vivo* malonyl-CoA sensor, synthetic biology

## INTRODUCTION

The cell membrane, consisting mainly of a fluid phospholipid bilayer in which a variety of proteins are embedded, is an essential structure to bacteria making membrane lipid homeostasis a crucial aspect of bacterial cell physiology. The production of phospholipids requires of the biosynthesis of fatty acids and their subsequent delivery to the membrane-bound glycerol-phosphate acyltransferases. In all organisms fatty acids are synthesized via a repeated cycle of reactions involving the condensation, reduction, hydration, and reduction of carbon-carbon bonds (Rock and Cronan, 1996; Campbell and Cronan, 2001). In mammals and other higher eukaryotes, these reactions are all catalyzed by a large multifunctional protein, known as type I synthase (FAS I), in which the growing fatty acid chain is covalently attached to the protein (Rock and Cronan, 1996; Campbell and Cronan, 2001). In contrast, bacteria, plant chloroplasts, and *Plasmodium falciparum* contain a type II system (FAS II) in which each reaction is catalyzed by a discrete protein. A characteristic of FASII is that all fatty acyl intermediates are covalently connected to a small acidic protein named acyl carrier protein (ACP), and sequentially shuttled from one enzyme to another. A key molecule for fatty acid elongation is malonyl-coenzyme A (CoA) which is formed by

carboxylation of acetyl-CoA by the enzyme acetyl-CoA carboxylase (ACC) (**Figure 1**). This biosynthetic scheme is conserved in all fatty acid producing bacteria, but the substrate specificity of some of the enzymes involved in the pathway leads to the variety of fatty acids found in different bacterial genera (Campbell and Cronan, 2001; Lu et al., 2004). When the acyl-ACPs reach the proper length they become substrates for the acyltransferases that transfer successively the fatty acyl chains into glycerol phosphate to synthesize phosphatidic acid (PtdOH), the universal intermediate in the biosynthesis of membrane glycerophospholipids (**Figure 1**; Campbell and Cronan, 2001; Rock and Jackowski, 2002). There are two enzyme systems that carry out the first transacylation reaction in bacteria. In the first one, present exclusively in Gram-negative bacteria (primarily gamma-proteobacteria), either acyl-ACP or acyl-CoA thioesters are utilized by the membrane-bound PlsB acyltransferase to acylate position 1 of glycerol-P giving 1-acylglycerol phosphate (Parsons and Rock, 2013). The second enzyme system, widely distributed and predominating in Gram-positive bacteria, consist of the PlsX/Y pathway for 1-acyl-glycerol phosphate formation (Lu et al., 2006; Schujman and de Mendoza, 2006; Paoletti et al., 2007). PlsX is a membrane associated protein (Sastre et al., 2016) that catalyzes the formation of a novel acyl donor, acyl phosphate (acyl-P), from acyl-ACP. This activated fatty acid is then used by the membrane-bound PlsY acyl transferase to acylate the position 1 of glycerol phosphate. The PlsX/PlsY system is also present in *E. coli* although its precise role is still an enigma as *plsB* is an essential gene in this bacterium (Parsons and Rock, 2013). Independently of the first enzyme system used, the second acyl transferase in PtdOH formation is PlsC, which is universally expressed in bacteria. This enzyme completes the synthesis of PtdOH by transferring an acyl chain to the position 2 of 1-acyl-glycerol phosphate. In the case of Gram-positive bacteria, PlsC isoforms exclusively utilize acyl-ACP (Lu et al., 2006; Paoletti et al., 2007), while *E. coli* PlsC can use both, acyl-ACP or acyl-CoA, as substrates (Coleman, 1992).

The fluidity of the lipid bilayer is essential for the normal function of the cellular membrane and bacteria normally control its physical state by modifying the incorporation of a mixture of fatty acids with different melting temperatures into phospholipids. In this sense, many bacteria respond to a decrease in temperature, which increases membrane rigidity, by increasing the proportion of unsaturated fatty acids (UFAs) into the phospholipids and viceversa (Zhang and Rock, 2008). Unsaturated double bonds in lipids generate kinks into the otherwise straightened acyl hydrocarbon chain and thereby increase membrane fluidity. Hence, the production of UFAs and its regulation are important processes in membrane homeostasis in bacteria and the underlying diverse mechanisms have been recently revised elsewhere (Mansilla et al., 2008; Parsons and Rock, 2013).

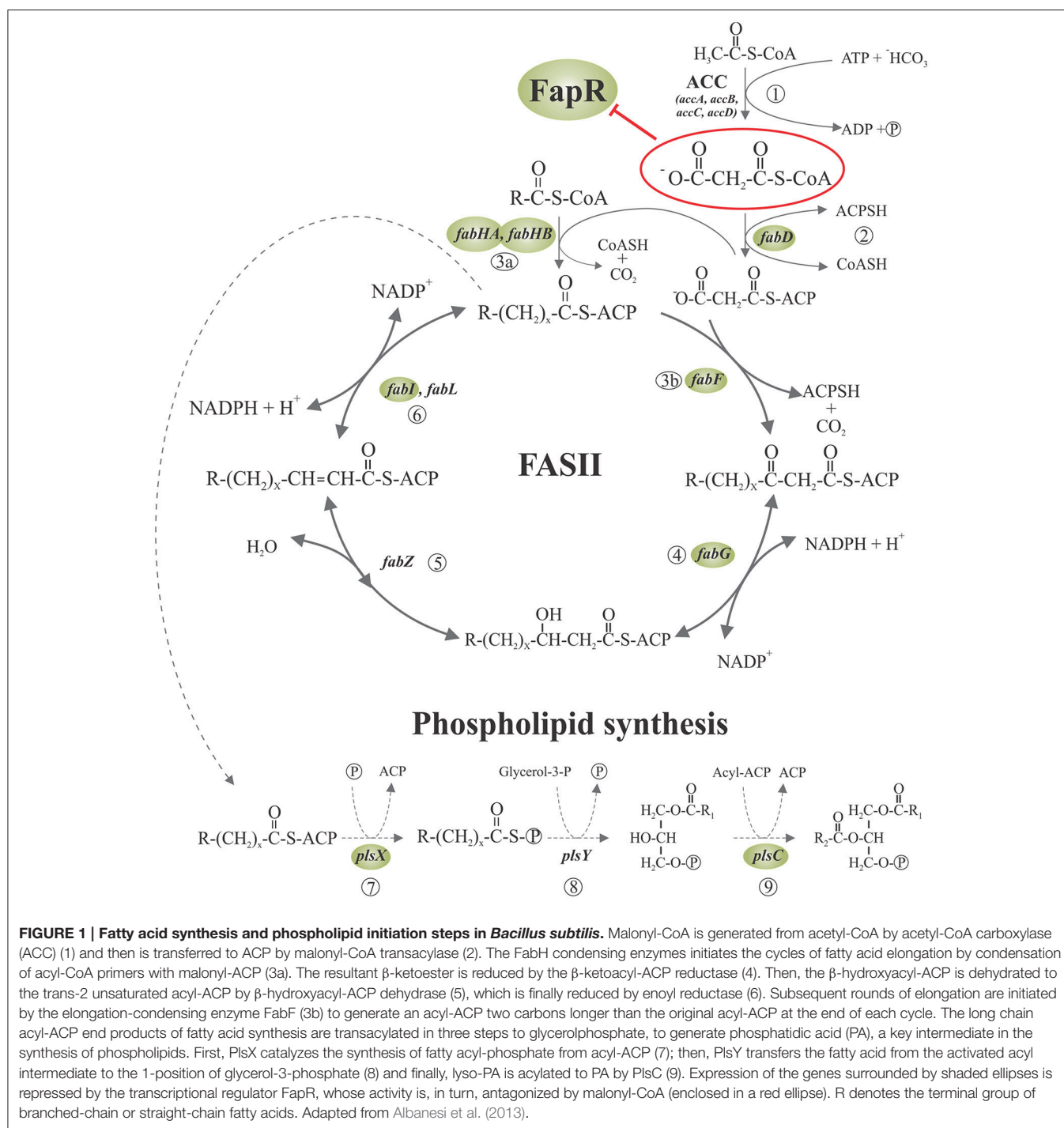
Due to the fact that the membrane lipid bilayer is an essential structure for every living cell and its biogenesis implies a high energetic cost, mainly due to fatty acid biosynthesis, organisms have developed a variety of homeostatic mechanisms to finely adjust the concentration of lipids at particular levels. Bacteria possess regulatory mechanisms acting directly on

the activities of the lipid biosynthetic enzymes, but have also evolved sophisticated mechanisms to exert an exquisite control over the expression of the genes involved in lipid metabolism (Zhang and Rock, 2008; Parsons and Rock, 2013). Six transcriptional regulators controlling the expression of genes involved in fatty acid biosynthesis have been identified to date in bacteria. Among them, FadR (Henry and Cronan, 1991, 1992; Lu et al., 2004), DesR (Aguilar et al., 2001; Mansilla and de Mendoza, 2005), FabR (Zhang et al., 2002), and DesT (Zhu et al., 2006; Zhang et al., 2007), are committed to adjust unsaturated fatty acids to proper levels in membrane phospholipids while FapR (Schujman et al., 2003) and FabT (Lu and Rock, 2006) are global transcriptional repressors in Gram-positive bacteria that simultaneously regulate the expression of a number of genes involved in fatty acid and phospholipid metabolism.

This review focuses on the genetic, biochemical and structural characterization of FapR which paved the way to a major advance in our understanding of the molecular basis of the lipid homeostasis control in bacteria. We will also cover the potential of this regulatory system as a target for new antibacterial compounds as well as emerging biotechnological applications based on it.

## THE DISCOVERY OF THE FapR SYSTEM

FapR from *Bacillus subtilis* was the first global transcriptional regulator of FASII to be discovered in bacteria (Schujman et al., 2003). The initial evidence that fatty acid biosynthesis was transcriptionally regulated came from the study of *lacZ* fusions to the promoter region of the *fabHAF* operon of *B. subtilis*, which codes for two key enzymes involved in the elongation of fatty acids (Schujman et al., 2001). These studies showed that the operon *fabHAF* is transcribed during exponential phase but when the cell culture approaches to stationary phase its transcription is turned off (Schujman et al., 2001). This finding is consistent with the observation that during exponential growth bacteria constantly produce new membrane in order to divide and hence need to actively synthesize fatty acids. Nevertheless, when cell division is completed membrane growth stops and fatty acid synthesis is turned off. An important finding was that when fatty acid synthesis is inhibited the transcription of the *fabHAF* operon is induced with the concomitant increment in protein levels (Schujman et al., 2001). Thus, it was proposed that *B. subtilis* is able to detect a decrease in the activity of FASII and respond accordingly by inducing the production of the condensing enzymes FabHA and FabF (Schujman et al., 2001). Moreover, DNA microarray studies indicated that upon inhibition of fatty acid synthesis the transcription of ten genes was induced (Schujman et al., 2003). These genes coded for proteins involved in fatty acid and phospholipid biosynthesis and belonged to six operons (the *fap* regulon) (Schujman et al., 2003). Furthermore, a conserved 17 bp inverted repeat within, or immediately downstream, of the *fap* predicted promoters, consistent with a putative binding site for a transcriptional repressor, was identified (Schujman et al., 2003). The corresponding binding protein



was isolated from cells extracts using a DNA fragment carrying the promoter region of *fabHA* and identified by N-terminal sequencing (Schujman et al., 2003). The gene encoding the global transcriptional repressor was named *fapR* for *fatty acid and phospholipid regulator* (Schujman et al., 2003). The binding of FapR to the promoter regions of the regulated genes, and its dependence on the 17 inverted repeats was demonstrated *in vitro*. It was also showed that in a *fapR* null mutant the expression

of the *fap* regulon is upregulated and that this expression is not further increased upon inhibition of FASII (Schujman et al., 2003). Therefore, it was established that FapR was a novel global negative regulator of lipid biosynthesis in Gram-positive bacteria and that FapR was involved in the observed induction of transcription in the presence of fatty acids synthesis inhibitors (Schujman et al., 2003). Bioinformatic analyses indicated that FapR is present and highly conserved in all the species of



the *Bacillus*, *Listeria*, and *Staphylococcus* genera (all including important human pathogens like *Bacillus anthracis*, *Bacillus cereus*, *Listeria monocytogenes*, and *Staphylococcus aureus*) as well as in the pathogen *Clostridium difficile* and other related genera. However, *fapR* was not found in Gram-negative bacteria or other Gram-positive genera (Schujman et al., 2003). Furthermore, in the bacterial species bearing FapR, the consensus binding sequence for the repressor is also highly conserved in the putative *fapR* promoter region. Altogether, the observations suggested that the regulatory mechanism identified in *B. subtilis* could be conserved in many other bacteria (Schujman et al., 2003). Indeed, genetic and biochemical assays proved this is the case in *S. aureus* (Albanesi et al., 2013).

## MALONYL-CoA: THE EFFECTOR MOLECULE

A central question in the regulation of the *fap* regulon by FapR was how the status of fatty acids synthesis controlled the activity of the repressor. The fact that (i) the *acc* genes, encoding the subunits of the acetyl-CoA carboxylase (ACC), which catalyzes the synthesis of malonyl-CoA (**Figure 1**), are not under FapR control (Schujman et al., 2003), (ii) malonyl-CoA concentrations are known to increase upon inhibition of fatty acid synthesis (Heath and Rock, 1995), and (iii) the only known fate of malonyl-CoA in *B. subtilis* and most other bacteria is fatty acid synthesis (James and Cronan, 2003), pointed to malonyl-CoA as a reasonable candidate to be the regulatory ligand. Two observations gave experimental support to this hypothesis. First, expression of the *fap* regulon was derepressed by antibiotics that inhibit fatty acid biosynthesis with the concomitant increase in the intracellular levels of malonyl-CoA (Schujman et al., 2001). Second, this upregulation was abolished by precluding the transcription of genes encoding the subunits of the acetyl-CoA carboxylase (ACC) (Schujman et al., 2003).

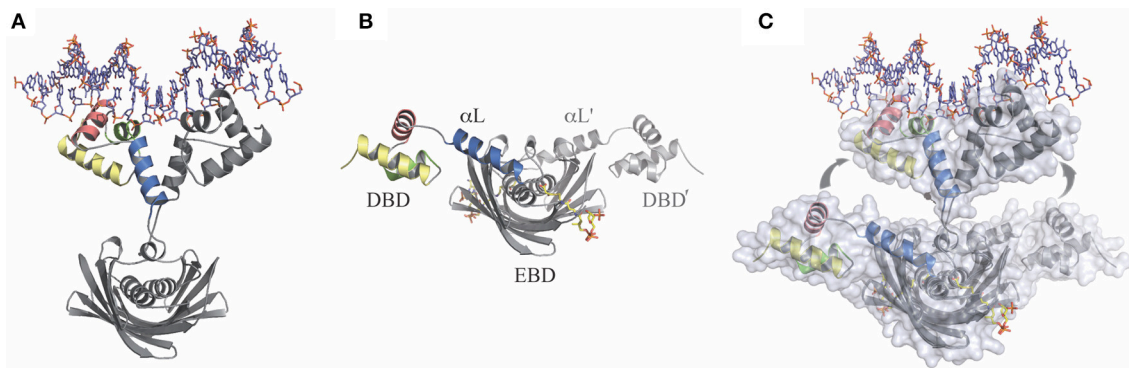
A key issue was to establish if malonyl-CoA bound directly to FapR to regulate its activity or if it was first converted into another product that acted as a signaling molecule. The finding that antibiotics against different steps of FASII led to the transcriptional induction of the *fap* regulon, even when the *B. subtilis fabD* gene (Morbidoni et al., 1996) was not expressed, suggested that malonyl-CoA could be the direct effector of FapR (Schujman et al., 2003). FabD converts malonyl-CoA into malonyl-ACP, which, in turn, is only utilized in the elongation of fatty acid synthesis (de Mendoza et al., 2002). *In vitro* transcription experiments from several promoters of the *fap* regulon, including the *fapR*-operon promoter (*PfapR*), proved that FapR is unable to repress transcription in the presence of malonyl-CoA. Moreover, these assays showed that this molecule operates not only as a direct but also as a specific inducer of the *fap* promoters since different acyl-CoA derivatives related to malonyl-CoA (such as acetyl-CoA, propionyl-CoA, succinyl-CoA, and butyryl-CoA), were not able to prevent FapR transcriptional repression (Schujman et al., 2003). The same direct and specific role of malonyl-CoA as the effector molecule was shown for FapR of *S. aureus* (*SaFapR*) (Albanesi et al., 2013).

## STRUCTURAL SNAPSHOTS OF THE FapR REGULATION CYCLE

Like many transcriptional regulators in bacteria, FapR is a two-domain protein with an N-terminal DNA-binding domain (DBD) connected through a linker  $\alpha$ -helix ( $\alpha_L$ ) to a larger C-terminal effector-binding domain (EBD) (Schujman et al., 2006). The first insights on the molecular mechanism for the control of FapR activity came from the crystal structures of truncated forms of FapR from *B. subtilis* (*BsFapR*) (Schujman et al., 2006). These structures showed that the EBD is a symmetric dimer displaying a “hot-dog” architecture, with two central  $\alpha$ -helices surrounded by an extended twelve-stranded  $\beta$ -sheet (Schujman et al., 2006). This fold is similar to the one observed in many homodimeric acyl-CoA-binding enzymes (Leesong et al., 1996; Li et al., 2000) involved in fatty acid biosynthesis and metabolism (Dillon and Bateman, 2004; Pidugu et al., 2009). Interestingly, FapR, a bacterial transcriptional repressor, seems to be the only well-characterized protein to date with non-enzymatic function that harbors the “hot-dog” fold (Albanesi et al., 2013). On the other hand, the EBD domain of *BsFapR* was crystallized in complex with malonyl-CoA. Comparison of both structures revealed structural changes induced by the effector molecule in some ligand-binding loops of the EBD that were suggested to propagate to the N-terminal DBDs impairing their productive association for DNA binding (Schujman et al., 2006). However, the actual mechanisms involved in the regulation of FapR activity remained largely unknown due to the lack of detailed structural information of the full-length repressor and its complex with DNA. Recently, important mechanistic advances into the mode of action of FapR were done through the structural characterization of the full-length repressor from *S. aureus* (*SaFapR*). The crystal structures of *SaFapR* were obtained for the protein alone (apo-*SaFapR*) as well as in complex with the cognate DNA operator and the effector molecule malonyl-CoA (Albanesi et al., 2013) (**Figure 2**).

### Structure of the *SaFapR*-DNA Complex

The crystal structure of the *SaFapR*-DNA complex was obtained using a 40-bp oligonucleotide comprising the *PfapR* promoter, which, as mentioned above, belongs to the *fap* regulon (Schujman et al., 2006). In the crystal, two *SaFapR* homodimers were observed to bind to each DNA molecule. Interestingly, an inverted repeat covering half of the FapR-protected region in DNaseI footprinting analyses (Schujman et al., 2006), corresponded to the recognition site of one of the homodimers (Albanesi et al., 2013). This suggested a sequential mechanism of binding that was confirmed by isothermal titration calorimetry (ITC) studies of the *SaFapR*-DNA interaction, which also provided the dissociation constants of each binding reaction (Albanesi et al., 2013). In the crystal structure of the *SaFapR*-DNA complex, each protein homodimer exhibited an elongated asymmetric conformation with the two DNA-bound DBDs completely detached from the central dimeric “hot-dog” EBD (**Figure 2A**) (Albanesi et al., 2013). In each homodimer the amphipathic linker  $\alpha$ -helices from the protomers ( $\alpha_L$  and  $\alpha_L'$ ) interact, mainly through their exposed hydrophobic faces,



**FIGURE 2 | The transitional switch between the relaxed and tense states of FapR involves a significant structural rearrangement of the DBDs.**

**(A)** Relaxed state: FapR in complex with DNA in which the amphipathic linker  $\alpha$ -helix ( $\alpha$ L) from each protomer associates with each other. **(B)** Tense state: FapR in complex with malonyl-CoA (shown in stick representation). **(C)** Superposition of the two conformational states of the repressor illustrating the structural transition which involves substantial changes and large ( $\sim 30$  Å) inter-domain movements. Solvent accessible surfaces are shown in transparent to highlight the DNA-induced dissociation of the invariant effector-binding domain (EBD) from the DNA-binding domains (DBDs). The molecules are shown in light (relaxed) and dark (tense) gray, except for the helices from one DBD (colored). Adapted from Albanesi et al. (2013).

playing an important role in the stabilization of SaFapR's molecular architecture in the complex with DNA (**Figure 2A**) (Albanesi et al., 2013). On their hand, both DBDs interact in a similar manner with DNA establishing sequence-specific contacts between the helix-turn-helix motifs with the major and minor grooves of the DNA double helix (Albanesi et al., 2013). Importantly, two arginine residues from each linker  $\alpha$ L (one from  $\alpha$ L and one from  $\alpha$ L') make base-specific interactions in the minor groove promoting its opening and inducing a pronounced local bending of DNA (Albanesi et al., 2013). Notably, the aminoacid residues making key contacts with DNA are highly conserved in FapR from all bacterial species where it was identified indicating the DNA-binding-mode of this transcriptional repressor is conserved (Albanesi et al., 2013).

## Structure of the SaFapR-Malonyl-CoA Complex

The crystal structure of full-length SaFapR in complex with malonyl-CoA showed that in the presence of the effector molecule the repressor adopts a quaternary arrangement that is different and more compact than when bound to DNA (**Figure 2B**) (Albanesi et al., 2013). In this conformation, both amphipathic linker helices  $\alpha$ L bind to either side of the central EBD domain instead of interacting with each other as when binding DNA. Like this, the two DBDs domain are far apart from each other, resulting in a non-productive conformation incompetent to bind DNA. Stabilization of the observed quaternary organization of the protein is principally due to the interaction of the linker  $\alpha$ L with the lateral face of the EBD (Albanesi et al., 2013). Concerning ligand binding, the structure showed that a tunnel is formed at the interface between the two protomers in the SaFapR homodimer into which the phosphopantetheine group is bound, adopting the same conformation as observed in the truncated BsFapR-malonyl-CoA complex structure, as well as in a number of acyl-CoA-binding proteins harboring the "hot-dog" fold (Albanesi et al., 2013).

In this way, the ligand malonate is completely occluded from the bulk solvent. The charged carboxylate group of malonate is neutralized at the bottom of the binding pocket by a specific interaction with an arginine residue. Upon engagement of this arginine in effector binding, a local reorganization is triggered that ultimately leads to surface reshaping and stabilization of the non-productive conformation, thus preventing DNA binding (Albanesi et al., 2013). On the other hand, the adenosine-3'-phosphate moiety of malonyl-CoA is largely exposed to the solvent making no specific contacts with the protein. This implies that SaFapR specifically recognizes the malonyl-phosphopantetheine moiety of the ligand (Albanesi et al., 2013) in agreement with the fact that either malonyl-CoA or malonyl-acyl carrier protein (malonyl-ACP) can both function as effector molecules (Martinez et al., 2010). A detailed comparison of the complexes of full-length SaFapR and the truncated form of BsFapR (lacking the DBDs) with malonyl-CoA revealed a conserved structural arrangement of the EBD core and ligand binding effects. Altogether, the structural alignment indicates an identical mode of malonyl-CoA binding and also the conservation of the DBD- $\alpha$ L-EBD interactions required to stabilize the FapR-malonyl-CoA complex as observed in the SaFapR model (**Figure 2B**) (Albanesi et al., 2013).

## The Structure of Full-Length SaFapR

Full-length SaFapR was also crystallized in the absence of ligands (apo-SaFapR) and two crystals forms were obtained (Albanesi et al., 2013). In the different structures, most of the crystallographic independent repressor protomers exhibited the non-productive quaternary arrangement with helix  $\alpha$ L bound to the lateral face of the EBD, as observed in the structure of SaFapR in complex with malonyl-CoA (**Figure 2B**), strongly suggesting that in solution the apo-protein would also display this conformation (Albanesi et al., 2013). However, in one of the crystal forms, the helix  $\alpha$ L and the associated DBD of one SaFapR protomer could not be modeled due to their high flexibility and

the corresponding first visible residues connecting the helix  $\alpha$ L with the EBD exhibited a similar conformation to that found for one subunit of the repressor in the asymmetric SaFapR-DNA complex (**Figure 2A**) (Albanesi et al., 2013). These facts and other crystal parameters (like the extensive crystal contact engagement, the high temperature factors or even the partial disorder displayed by the helix-turn-helix motifs) suggested that alternative conformational states of SaFapR, marked by flexible DBDs, would coexist in solution (Albanesi et al., 2013).

## Structural Transitions along the FapR Regulation Cycle

The structural snapshots of full-length SaFapR along its regulation cycle revealed distinct quaternary arrangements for the DNA-bound (relaxed) and the malonyl-CoA-bound (tense) forms of the repressor, with the linker  $\alpha$ L involved in different protein-protein interactions in each case, highlighting a functional switch entailing a large-scale structural rearrangement (**Figure 2C**) (Albanesi et al., 2013). Indeed, the amphipathic  $\alpha$ L, that in the tense state binds through its hydrophobic face to the protein EBD (**Figure 2B**), dissociates and moves  $\sim 30$  Å to finally interact with  $\alpha$ L from the second protomer ( $\alpha$ L') and with DNA in the relaxed state (**Figure 2A**) (Albanesi et al., 2013). Furthermore, the structural analysis of apo-SaFapR in two distinct crystal forms also showed that the ligand-free repressor species can populate both, the tense and relaxed conformational states (Albanesi et al., 2013). This suggested that DNA would promote and stabilize the relaxed form of the repressor while an increment in the intracellular concentration of malonyl-CoA would not only trigger the structural changes leading to disruption of the repressor-operator complex but would also drive a shift of the ligand-free SaFapR population toward the tense form (Albanesi et al., 2013).

## THE FapR SYSTEM AS A TARGET FOR NEW ANTIBACTERIAL DRUGS

As discussed above, bacterial fatty acid biosynthesis is essential for the formation of biological membranes. Indeed, the importance of the pathway in bacterial physiology is highlighted by the existence of multiple natural products that target different points in this biosynthetic route (Parsons and Rock, 2011). The emergence of resistance to most clinically deployed antibiotic has stimulated considerable interest in finding new therapeutics, leading to a significant effort in academia and industry to develop antibiotic that target individual proteins in fatty acid biosynthesis. One concern about such drugs is that fatty acids are abundant in the mammalian host, raising the possibility that fatty acid synthesis inhibitors would be bypassed *in vivo* (Brinster et al., 2009). Although all bacteria studied to date are capable of incorporating extracellular fatty acids into their membranes, recent research shows that, opposite to what happens in *Streptococcus pneumoniae* (Parsons et al., 2011), exogenous fatty acids cannot circumvent the inhibition of FASII in *S. aureus* and many major human pathogens (Yao and Rock, 2015).

Notably, disruption of FapR-malonyl-CoA interactions by structure-based amino acid substitutions in *S. aureus* leads to permanent repression of fatty acid and phospholipid synthesis, which is lethal and cannot be overcome by addition of exogenous fatty acids (Albanesi et al., 2013), as observed with antibiotics targeting FASII (Parsons et al., 2011). Thus, the distinctive mode of action of FapR together with the promising *in vivo* results highlight lipid homeostasis and the FapR system as a propitious target for the development of new drugs against Gram-positive bacteria.

## THE FapR SYSTEM AS A BIOTECHNOLOGICAL TOOL

In the last few years, a number of research groups have taken advantage of the unique properties of FapR to design and construct malonyl-CoA biosensors. Recently, a FapR-based malonyl-CoA sensor has been developed to detect changes of malonyl-CoA flux in living mammalian cells (Ellis and Wolfgang, 2012). After codon optimization, FapR from *B. subtilis* was fused to VP16, a viral transcriptional activator. The VP16 fusion converted FapR from a bacterial transcriptional repressor into a transcriptional activator in the absence of malonyl-CoA. The FapR operator sequence (*fapO*) was then multimerized and cloned upstream of a minimal promoter driving a reporter gene. This FapR-based malonyl-CoA biosensor was proven to be transcriptionally regulated by malonyl-CoA in mammalian cells and the reporter gene activity was demonstrated to be correlated with the intracellular levels of this effector molecule (Ellis and Wolfgang, 2012). This biosensor was then used to identify several novel kinases that when expressed in COS1 cells (a fibroblast-like cell line derived from monkey kidney tissue) promoted an increment of malonyl-CoA concentrations. In particular, it was shown that the expression of one of these kinases, LIMK1, altered both fatty acid synthesis and fatty acid oxidation rates. Thus, this simple malonyl-CoA responsive biosensor proved to be useful for the study of lipid metabolism in live mammalian cells and the identification of a novel metabolic regulator (Ellis and Wolfgang, 2012).

Two independent groups reported the development of a malonyl-CoA biosensor based on the FapR system of *B. subtilis* in the yeast *Saccharomyces cerevisiae* (Li et al., 2015; David et al., 2016). In both cases FapR was directed to the nucleus where it acted as a repressor on a synthetic promoter containing the FapR-operator site in optimized positions. The biosensors were validated and showed to reflect the change of intracellular malonyl-CoA concentrations. Both groups then used the malonyl-CoA biosensor to improve the production of the biotechnological valuable intermediate 3-hydroxypropionic acid (3-HP), which serves as the precursor to a series of chemicals, such as acrylates. Each group followed a different strategy to achieve this goal. Li et al. (2015) used the malonyl-CoA biosensor to screen a genome-wide overexpression library resulting in the identification of two novel gene targets that raised the intracellular malonyl-CoA concentration. Furthermore, they overexpressed the identified genes in a yeast



strain carrying a bifunctional enzyme, *caMCR*, from *Chloroflexus aurantiacus* that acts both, as an NADPH-dependent malonyl-CoA reductase and as a 3-hydroxypropionate dehydrogenase, converting malonyl-CoA to malonic-semialdehyde first and then to 3-HP. Interestingly, the authors found that the recombinant yeast strains producing higher amounts of malonyl-CoA showed over 100% improvement of 3-HP production (Li et al., 2015). Using a different approach, David et al. (2016) expressed the gene coding for *caMCR* (*mcrCa*) under the control of the FapR-based biosensor. This self-regulated system gradually expressed the *mcrCa* gene depending on the available concentration of malonyl-CoA. Subsequently, in order to increase the malonyl-CoA supply for 3-HP production, the authors (David et al., 2016) implemented a hierarchical dynamic control system using the PHXT1 promoter to render FAS1 expression dependent on the concentration of glucose. FAS1 codes for the  $\beta$ -subunit of the fatty acid synthase complex in *S. cerevisiae*, while the  $\alpha$ -subunit is encoded by FAS2. The expression of FAS1 and FAS2 is co-regulated, implying a coordinated up—or downregulation of the entire FAS system. Hence, when the external glucose concentration is low the PHXT1 promoter is repressed and FAS1 gene expression is downregulated, decreasing the consumption of malonyl-CoA in fatty acid biosynthesis. As a consequence, there is an increment in the intracellular malonyl-CoA concentration available for 3-HP production. Using this hierarchical two-level control and the fine-tuning of *mcrCa* gene expression, a 10-fold increase in 3-HP production was obtained (David et al., 2016).

Aliphatic hydrocarbons produced by microorganisms constitute a valuable source of renewable fuel so, in order to satisfy the global energy demand, high productivity and yields become essential parameters to achieve. Nowadays big efforts in microbial biofuel production are dedicated to build efficient metabolic pathways for the production of a variety of fatty acid-based fuels. In this regard, two studies were reported on the implementation of the FapR system in *E. coli*, which originally lacks the *fap* regulon, for the improvement of fatty acid production (Xu et al., 2014a; Liu et al., 2015). Malonyl-CoA, produced by ACC (**Figure 1**), is the rate limiting precursor for the synthesis of fatty acids. The *E. coli* ACC is composed of four subunits: a biotin carboxyl carrier protein, a biotin carboxylase, and two carboxyltransferase subunits. The overexpression of the genes coding for the ACC subunits improves fatty acids production but at the same time is toxic to the cells (Davis et al., 2000; Zha et al., 2009). To overcome this drawback, Liu et al. designed a strategy for increasing malonyl-CoA synthesis reducing the toxicity provoked by the concomitant *acc* overexpression (Liu et al., 2015). To this end, they built a negative regulatory system for the *acc* genes based on the ability of FapR to respond to the level of malonyl-CoA. Their goal was to promote a reduction in *acc* expression when malonyl-CoA levels were high and induce it when the malonyl-CoA levels were low. This required the design of a rewired system to create a negative feedback circuit. To this end, the *B. subtilis* *fapR* gene was cloned into *E. coli* using a low copy number plasmid under the control of a  $P_{BAD}$  promoter responding to arabinose. A FapR-regulated synthetic promoter (PFR1) was also constructed by inserting the 17-bp FapR operator

sequence into two regions flanking the  $-10$  region of a phage PA1 promoter. PFR1 was validated as a FapR-regulated promoter by analyzing the expression of a fluorescent protein under its control in response to different concentrations of malonyl-CoA (Liu et al., 2015). To complete the circuit, the *acc* genes were placed under the control of a LacI-repressive T7 promoter, PT7, and the *lacI* gene was placed under the control of PFR1. Hence, *acc* expression is initiated upon IPTG induction, producing malonyl-CoA. When malonyl-CoA is accumulated in this strain, the expression from PFR1 will turn on producing LacI, which in turn down-regulates *acc*, decreasing the malonyl-CoA synthesis rate. Using this approach, it was demonstrated that the negative feed-back circuit alleviated growth inhibition caused by either ACC overexpression or malonyl-CoA accumulation (Liu et al., 2015). In addition, this method was used for improving fatty acid titers and productivity and, in principle, could be extended to the production of other chemicals that use malonyl-CoA as precursor (Liu et al., 2015). Xu et al. (2014b) also constructed a malonyl-CoA sensing device by incorporating *fapO* into a hybrid T7 promoter that was shown to be able to respond to a broad range of intracellular malonyl-CoA concentrations, inducing the expression from the T7 promoter at increasing concentrations of the effector molecule. Interestingly, this group then discovered that the FapR protein could activate gene expression from the native *E. coli* promoter  $P_{GAP}$  in the absence of malonyl-CoA, that malonyl-CoA inhibits this activation, and that the dynamic range (in response to malonyl-CoA) can be tuned by incorporating *fapO* sites within the  $P_{GAP}$  promoter (Xu et al., 2014a). In order to improve fatty acid production, the genes coding for the ACC were then put under the control of the  $P_{GAP}$  promoter and the fatty acid synthase (*fabADGI* genes) and the soluble thioesterase *tesA'* were placed under the control of the T7-based malonyl-CoA sensor promoter. Upon constitutive FapR expression, the resulting genetic circuit provided dynamic pathway control that improved fatty acid production relative to the “uncontrolled” strains (Xu et al., 2014a). Taken together, these studies highlight FapR as a powerful responsive regulator for optimization and efficient production of malonyl-CoA-derived compounds.

## CONCLUSIONS AND PERSPECTIVES

FapR is a global transcriptional repressor of lipid synthesis highly conserved in Gram-positive bacteria. Notably, the activity of this repressor is controlled by malonyl-CoA, the product of the first dedicated step of fatty acid biosynthesis, converting FapR into a paradigm of a feed-forward-modulated regulator of lipid metabolism. The activity of other well-characterized bacterial lipid regulators, like FadR of *E. coli* (van Aalten et al., 2001) or the TetR-like *P. aeruginosa* DesT (Miller et al., 2010), is feedback controlled by the long-acyl chain-end products of the FASII pathway (Zhang and Rock, 2009; Parsons and Rock, 2013). The EBDs of these proteins, frequently exhibit an  $\alpha$ -helical structure with a relaxed specificity for long-chain acyl-CoA molecules, possibly because helix-helix interactions are permissive enough to constitute a platform for the evolution of a binding site for fatty acids of diverse chain lengths (Albanesi et al., 2013). In



contrast, the feed-forward regulation mechanism of the FapR repressor family, which implies the recognition of the upstream biosynthetic intermediate malonyl-CoA, requires a high effector-binding specificity. In FapR, this high specificity is achieved by confining the charged malonyl group into a quite rigid internal binding pocket, and may be the reason why the “hot-dog” fold was recruited for this function (Albanesi et al., 2013). It is important to note that organisms using the FapR pathway could also count on a complementary feed-back regulatory loop operating at a biochemical level, for instance by controlling the synthesis of malonyl-CoA (Paoletti et al., 2007). If this is proven, it would imply that FapR-containing bacteria finely tune lipid homeostasis by feed-back and feed-forward mechanisms, as it indeed happens in higher organisms ranging from the nematode *Caenorhabditis elegans* to humans (Raghow et al., 2008).

Human health and life quality have significantly improved with the discovery of antibiotics for the treatment of infectious bacterial diseases. However, the emergence of bacterial resistance to all antimicrobials in clinical use (Levy and Marshall, 2004; Davies and Davies, 2010) has caused infectious bacterial diseases to re-emerge as a serious threat to human health. This scenario highlights the need to develop new strategies to combat bacterial pathogens. FapR controls the expression of many essential genes for bacteria not only involved in fatty acids but also in phospholipid synthesis. It has been experimentally shown that the presence of mutant variants of FapR unable to bind malonyl-CoA result lethal for bacteria (even in the presence of exogenous fatty acids), as the regulator remains permanently bound to DNA impeding the expression of its target genes. These results and the existence of FapR in important human pathogens validate FapR and lipid homeostasis as interesting

targets for the search of new antibacterial drugs. With another perspective, the high specificity of FapR for malonyl-CoA has allowed for the development of *in vivo* malonyl-CoA sensors in diverse organisms that originally lack FapR and the *fap* regulon. These sensors have been shown to function in mammalian cells, in yeast and in bacteria responding accurately to the intracellular variations in the concentration of malonyl-CoA. The different FapR-based-malonyl-CoA biosensors were constructed following alternative strategies and used with a broad range of purposes focused on biological processes involving malonyl-CoA, including signaling mechanisms and metabolic engineering. Malonyl-CoA is the precursor of many industrial-valuable compounds like fatty acids, 3-hydroxypropionic acid, polyketides, and flavonoids, since they can be used as or converted to biofuels, commodity chemicals, fine chemicals, and drugs. Due to the success in the implementation of the FapR-based biosensors to improve the productivity and yields of the production of several malonyl-CoA-derived compounds, it is expected that new biotechnological applications of the FapR system emerge in the short term.

## AUTHOR CONTRIBUTIONS

DA and DdM conceived and wrote this review.

## ACKNOWLEDGMENTS

Financial support was provided by Agencia Nacional de Promoción Científica y Tecnológica (awards PICT 2010–2678 and PICT 2014–2474), Argentina. DA and DdM are Career Investigators of CONICET, Argentina.

## REFERENCES

- Aguilar, P. S., Hernandez-Arriaga, A. M., Cybulski, L. E., Erazo, A. C., and de Mendoza, D. (2001). Molecular basis of thermosensing: a two-component signal transduction thermometer in *Bacillus subtilis*. *EMBO J.* 20, 1681–1691. doi: 10.1093/emboj/20.7.1681
- Albanesi, D., Reh, G., Guerin, M. E., Schaeffer, F., Debarbouille, M., Buschiazzo, A., et al. (2013). Structural basis for feed-forward transcriptional regulation of membrane lipid homeostasis in *Staphylococcus aureus*. *PLoS Pathog.* 9:e1003108. doi: 10.1371/journal.ppat.1003108
- Brinster, S., Lamberet, G., Staels, B., Trieu-Cuot, P., Gruss, A., and Poyart, C. (2009). Type II fatty acid synthesis is not a suitable antibiotic target for Gram-positive pathogens. *Nature* 458, 83–86. doi: 10.1038/nature07772
- Campbell, J. W., and Cronan, J. E. Jr. (2001). Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery. *Annu. Rev. Microbiol.* 55, 305–332. doi: 10.1146/annurev.micro.55.1.305
- Coleman, J. (1992). Characterization of the *Escherichia coli* gene for 1-acyl-sn-glycerol-3-phosphate acyltransferase (*plsC*). *Mol. Gen. Genet.* 232, 295–303.
- David, F., Nielsen, J., and Siewers, V. (2016). Flux Control at the Malonyl-CoA Node through hierarchical dynamic pathway regulation in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* 5, 224–233. doi: 10.1021/acssynbio.5b00161
- Davies, J., and Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* 74, 417–433. doi: 10.1128/MMBR.00016-10
- Davis, M. S., Solbiati, J., and Cronan, J. E. Jr. (2000). Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in *Escherichia coli*. *J. Biol. Chem.* 275, 28593–28598. doi: 10.1074/jbc.M004756200
- de Mendoza, D., Aguilar, P., and Schujman, G. E. (2002). “Biosynthesis and function of membrane lipids,” in *Bacillus Subtilis and Its Relatives: From Genes to Cells*, eds J. A. Hoch, R. Losick, and A. Sonenshein (Washington, DC: American Society for Microbiology), 43–55.
- Dillon, S. C., and Bateman, A. (2004). The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. *BMC Bioinformatics* 5:109. doi: 10.1186/1471-2105-5-109
- Ellis, J. M., and Wolfgang, M. J. (2012). A genetically encoded metabolite sensor for malonyl-CoA. *Chem. Biol.* 19, 1333–1339. doi: 10.1016/j.chembiol.2012.08.018
- Heath, R. J., and Rock, C. O. (1995). Regulation of malonyl-CoA metabolism by acyl-acyl carrier protein and beta-ketoacyl-acyl carrier protein synthases in *Escherichia coli*. *J. Biol. Chem.* 270, 15531–15538. doi: 10.1074/jbc.270.26.15531
- Henry, M. F., and Cronan, J. E. Jr. (1991). *Escherichia coli* transcription factor that both activates fatty acid synthesis and represses fatty acid degradation. *J. Mol. Biol.* 222, 843–849. doi: 10.1016/0022-2836(91)90574-P
- Henry, M. F., and Cronan, J. E. Jr. (1992). A new mechanism of transcriptional regulation: release of an activator triggered by small molecule binding. *Cell* 70, 671–679. doi: 10.1016/0092-8674(92)90435-F
- James, E. S., and Cronan, J. E. (2003). Never fat or gaunt. *Dev. Cell* 4, 610–611. doi: 10.1016/S1534-5807(03)00132-1
- Leesong, M., Henderson, B. S., Gillig, J. R., Schwab, J. M., and Smith, J. L. (1996). Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure* 4, 253–264. doi: 10.1016/S0969-2126(96)00030-5
- Levy, S. B., and Marshall, B. (2004). Antibacterial resistance worldwide: causes, challenges and responses. *Nat. Med.* 10, S122–S129. doi: 10.1038/nm1145

- Li, J., Derewenda, U., Dauter, Z., Smith, S., and Derewenda, Z. S. (2000). Crystal structure of the *Escherichia coli* thioesterase II, a homolog of the human Nef binding enzyme. *Nat. Struct. Biol.* 7, 555–559. doi: 10.1038/76776
- Li, S., Si, T., Wang, M., and Zhao, H. (2015). Development of a Synthetic Malonyl-CoA Sensor in *Saccharomyces cerevisiae* for Intracellular Metabolite Monitoring and Genetic Screening. *ACS Synth. Biol.* 4, 1308–1315. doi: 10.1021/acssynbio.5b00069
- Liu, D., Xiao, Y., Evans, B. S., and Zhang, F. (2015). Negative feedback regulation of fatty acid production based on a malonyl-CoA sensor-actuator. *ACS Synth. Biol.* 4, 132–140. doi: 10.1021/sb400158w
- Lu, Y. J., and Rock, C. O. (2006). Transcriptional regulation of fatty acid biosynthesis in *Streptococcus pneumoniae*. *Mol. Microbiol.* 59, 551–566. doi: 10.1111/j.1365-2958.2005.04951.x
- Lu, Y. J., Zhang, Y. M., Grimes, K. D., Qi, J., Lee, R. E., and Rock, C. O. (2006). Acyl-phosphates initiate membrane phospholipid synthesis in Gram-positive pathogens. *Mol. Cell* 23, 765–772. doi: 10.1016/j.molcel.2006.06.030
- Lu, Y. J., Zhang, Y. M., and Rock, C. O. (2004). Product diversity and regulation of type II fatty acid synthases. *Biochem. Cell Biol.* 82, 145–155. doi: 10.1139/03-076
- Mansilla, M. C., Banchio, C. E., and de Mendoza, D. (2008). Signalling pathways controlling fatty acid desaturation. *Subcell. Biochem.* 49, 71–99. doi: 10.1007/978-1-4020-8831-5\_3
- Mansilla, M. C., and de Mendoza, D. (2005). The *Bacillus subtilis* desaturase: a model to understand phospholipid modification and temperature sensing. *Arch. Microbiol.* 183, 229–235. doi: 10.1007/s00203-005-0759-8
- Martinez, M. A., Zaballa, M. E., Schaeffer, F., Bellinzoni, M., Albanesi, D., Schujman, G. E., et al. (2010). A novel role of malonyl-ACP in lipid homeostasis. *Biochemistry* 49, 3161–3167. doi: 10.1021/bi100136n
- Miller, D. J., Zhang, Y. M., Subramanian, C., Rock, C. O., and White, S. W. (2010). Structural basis for the transcriptional regulation of membrane lipid homeostasis. *Nat. Struct. Mol. Biol.* 17, 971–975. doi: 10.1038/nsmb.1847
- Morbidoni, H. R., de Mendoza, D., and Cronan, J. E. Jr. (1996). *Bacillus subtilis* acyl carrier protein is encoded in a cluster of lipid biosynthesis genes. *J. Bacteriol.* 178, 4794–4800.
- Paoletti, L., Lu, Y. J., Schujman, G. E., de Mendoza, D., and Rock, C. O. (2007). Coupling of fatty acid and phospholipid synthesis in *Bacillus subtilis*. *J. Bacteriol.* 189, 5816–5824. doi: 10.1128/JB.00602-07
- Parsons, J. B., Frank, M. W., Subramanian, C., Saenkham, P., and Rock, C. O. (2011). Metabolic basis for the differential susceptibility of Gram-positive pathogens to fatty acid synthesis inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 108, 15378–15383. doi: 10.1073/pnas.1109208108
- Parsons, J. B., and Rock, C. O. (2011). Is bacterial fatty acid synthesis a valid target for antibacterial drug discovery? *Curr. Opin. Microbiol.* 14, 544–549. doi: 10.1016/j.mib.2011.07.029
- Parsons, J. B., and Rock, C. O. (2013). Bacterial lipids: metabolism and membrane homeostasis. *Prog. Lipid Res.* 52, 249–276. doi: 10.1016/j.plipres.2013.02.002
- Pidugu, L. S., Maity, K., Ramaswamy, K., Surolia, N., and Suguna, K. (2009). Analysis of proteins with the “hot dog” fold: prediction of function and identification of catalytic residues of hypothetical proteins. *BMC Struct. Biol.* 9:37. doi: 10.1186/1472-6807-9-37
- Raghow, R., Yellaturu, C., Deng, X., Park, E. A., and Elam, M. B. (2008). SREBPs: the crossroads of physiological and pathological lipid homeostasis. *Trends Endocrinol. Metab.* 19, 65–73. doi: 10.1016/j.tem.2007.10.009
- Rock, C. O., and Cronan, J. E. (1996). *Escherichia coli* as a model for the regulation of dissociable (type II) fatty acid biosynthesis. *Biochim. Biophys. Acta* 1302, 1–16. doi: 10.1016/0005-2760(96)00056-2
- Rock, C. O., and Jackowski, S. (2002). Forty years of bacterial fatty acid synthesis. *Biochem. Biophys. Res. Commun.* 292, 1155–1166. doi: 10.1006/bbrc.2001.2022
- Sastre, D. E., Bisson-Filho, A., de Mendoza, D., and Gueiros-Filho, F. J. (2016). Revisiting the cell biology of the acyl-ACP:phosphate transacylase PlsX suggests that the phospholipid synthesis and cell division machineries are not coupled in *Bacillus subtilis*. *Mol. Microbiol.* 100, 621–634. doi: 10.1111/mmi.13337
- Schujman, G. E., Choi, K. H., Altabe, S., Rock, C. O., and de Mendoza, D. (2001). Response of *Bacillus subtilis* to cerulenin and acquisition of resistance. *J. Bacteriol.* 183, 3032–3040. doi: 10.1128/JB.183.10.3032-3040.2001
- Schujman, G. E., and de Mendoza, D. (2006). Solving an old puzzle in phospholipid biosynthesis. *Nat. Chem. Biol.* 2, 573–574. doi: 10.1038/nchembio1106-573
- Schujman, G. E., Guerin, M., Buschiazzo, A., Schaeffer, F., Llarrull, L. I., Reh, G., et al. (2006). Structural basis of lipid biosynthesis regulation in Gram-positive bacteria. *EMBO J.* 25, 4074–4083. doi: 10.1038/sj.emboj.7601284
- Schujman, G. E., Paoletti, L., Grossman, A. D., and de Mendoza, D. (2003). FapR, a bacterial transcription factor involved in global regulation of membrane lipid biosynthesis. *Dev. Cell* 4, 663–672. doi: 10.1016/S1534-5807(03)00123-0
- van Aalten, D. M., DiRusso, C. C., and Knudsen, J. (2001). The structural basis of acyl coenzyme A-dependent regulation of the transcription factor FadR. *EMBO J.* 20, 2041–2050. doi: 10.1093/emboj/20.8.2041
- Xu, P., Li, L., Zhang, F., Stephanopoulos, G., and Koffas, M. (2014a). Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11299–11304. doi: 10.1073/pnas.1406401111
- Xu, P., Wang, W., Li, L., Bhan, N., Zhang, F., and Koffas, M. A. (2014b). Design and kinetic analysis of a hybrid promoter-regulator system for malonyl-CoA sensing in *Escherichia coli*. *ACS Chem. Biol.* 9, 451–458. doi: 10.1021/cb400623m
- Yao, J., and Rock, C. O. (2015). How bacterial pathogens eat host lipids: implications for the development of fatty acid synthesis therapeutics. *J. Biol. Chem.* 290, 5940–5946. doi: 10.1074/jbc.R114.636241
- Zha, W., Rubin-Pitel, S. B., Shao, Z., and Zhao, H. (2009). Improving cellular malonyl-CoA level in *Escherichia coli* via metabolic engineering. *Metab. Eng.* 11, 192–198. doi: 10.1016/j.ymben.2009.01.005
- Zhang, Y. M., Marrakchi, H., and Rock, C. O. (2002). The FabR (YijC) transcription factor regulates unsaturated fatty acid biosynthesis in *Escherichia coli*. *J. Biol. Chem.* 277, 15558–15565. doi: 10.1074/jbc.M201399200
- Zhang, Y. M., and Rock, C. O. (2008). Membrane lipid homeostasis in bacteria. *Nat. Rev. Microbiol.* 6, 222–233. doi: 10.1038/nrmicro1839
- Zhang, Y. M., and Rock, C. O. (2009). Transcriptional regulation in bacterial membrane lipid synthesis. *J. Lipid Res.* 50, S115–S119. doi: 10.1194/jlr.R800046-JLR200
- Zhang, Y. M., Zhu, K., Frank, M. W., and Rock, C. O. (2007). A *Pseudomonas aeruginosa* transcription factor that senses fatty acid structure. *Mol. Microbiol.* 66, 622–632. doi: 10.1111/j.1365-2958.2007.05934.x
- Zhu, K., Choi, K. H., Schweizer, H. P., Rock, C. O., and Zhang, Y. M. (2006). Two aerobic pathways for the formation of unsaturated fatty acids in *Pseudomonas aeruginosa*. *Mol. Microbiol.* 60, 260–273. doi: 10.1111/j.1365-2958.2006.05088.x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Albanesi and de Mendoza. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Influence of Copy-Number of Targeted Extrachromosomal Genetic Elements on the Outcome of CRISPR-Cas Defense

Konstantin Severinov<sup>1,2,3\*</sup>, Iaroslav Ispolatov<sup>4,1</sup> and Ekaterina Semenova<sup>2</sup>

<sup>1</sup> Skolkovo Institute of Science and Technology, Skolkovo, Russia, <sup>2</sup> Waksman Institute of Microbiology, Rutgers, The State University of New Jersey, Piscataway, NJ, USA, <sup>3</sup> Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, Russia, <sup>4</sup> Department of Physics, University of Santiago de Chile, Santiago, Chile

## OPEN ACCESS

### Edited by:

Tatiana Venkova,  
University of Texas Medical  
Branch-Galveston, USA

### Reviewed by:

Marko Djordjevic,  
University of Belgrade, Serbia  
Kurt Henry Piepenbrink,  
The University of Maryland School of  
Medicine, USA  
Peter Fineran,  
University of Otago, New Zealand

### \*Correspondence:

Konstantin Severinov  
severik@waksman.rutgers.edu

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 20 May 2016

**Accepted:** 16 August 2016

**Published:** 31 August 2016

### Citation:

Severinov K, Ispolatov I and  
Semenova E (2016) The Influence of  
Copy-Number of Targeted  
Extrachromosomal Genetic Elements  
on the Outcome of CRISPR-Cas  
Defense. *Front. Mol. Biosci.* 3:45.  
doi: 10.3389/fmolb.2016.00045

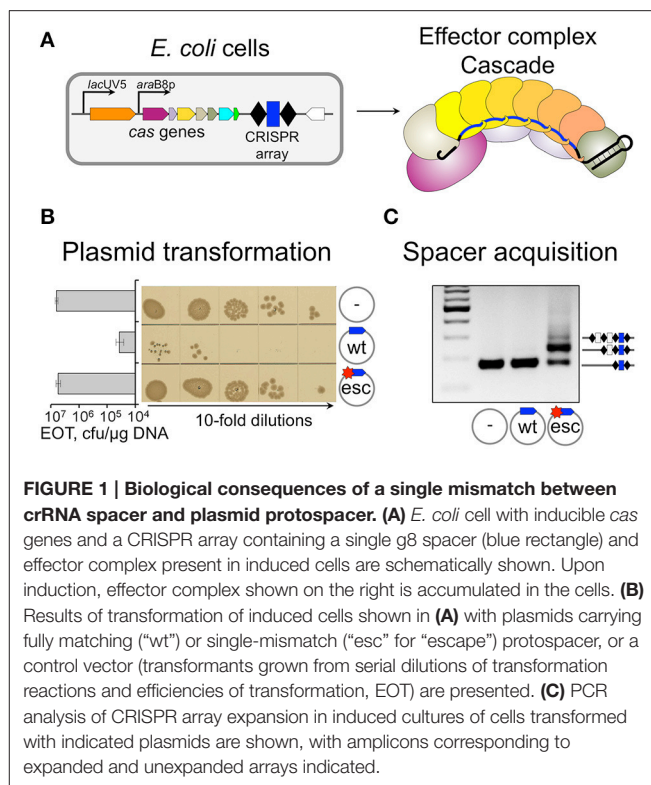
Prokaryotic type I CRISPR-Cas systems respond to the presence of mobile genetic elements such as plasmids and phages in two different ways. CRISPR interference efficiently destroys foreign DNA harboring protospacers fully matching CRISPR RNA spacers. In contrast, even a single mismatch between a spacer and a protospacer can render CRISPR interference ineffective but causes primed adaptation—efficient and specific acquisition of additional spacers from foreign DNA into the CRISPR array of the host. It has been proposed that the interference and primed adaptation pathways are mediated by structurally different complexes formed by the effector Cascade complex on matching and mismatched protospacers. Here, we present experimental evidence and present a simple mathematical model that shows that when plasmid copy number maintenance/phage genome replication is taken into account, the two apparently different outcomes of the CRISPR-Cas response can be accounted for by just one kind of effector complex on both targets. The results underscore the importance of consideration of targeted genome biology when considering consequences of CRISPR-Cas systems action.

**Keywords:** CRISPR-Cas interference, CRISPR-Cas adaptation, plasmid maintenance, bacteriophage infection, primed adaptation

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)—Cas (CRISPR associated proteins) systems provide their prokaryotic hosts with adaptive small-RNA-based immunity against mobile genetic elements such as viruses and plasmids (Barrangou et al., 2007; Brouns et al., 2008; Marraffini and Sontheimer, 2008). While evolutionary and mechanistically diverse, all CRISPR-Cas systems comprise (i) arrays of DNA repeats separated by unique spacers and (ii) *cas* genes (Makarova et al., 2015). Functionally, CRISPR-Cas systems can be divided into two modules. The acquisition module appropriates spacers from foreign DNA into CRISPR arrays and consists of Cas1 and Cas2, which function as stand-alone proteins or as fusions to other Cas proteins. Homologs of Cas1 and Cas2 are present in most functional CRISPR-Cas systems (Makarova et al., 2015). The Cas1 and Cas2 proteins from *Escherichia coli*, alone, are able to perform the spacer acquisition reaction *in vitro* (Nunez et al., 2015), and are also sufficient *in vivo* in the absence of other Cas proteins to incorporate new spacers into a minimal CRISPR array consisting of a single repeat and short upstream leader sequence (Yosef et al., 2012; Arslan et al., 2014). When a

spacer is acquired, a new copy of CRISPR repeat is also generated (Barrangou et al., 2007; Yosef et al., 2012; Arslan et al., 2014). Spacer acquisition catalyzed by Cas1 and Cas2 only is referred to as “naïve CRISPR adaptation” (Datsenko et al., 2012; Fineran and Charpentier, 2012). Acquired spacers become a source of small CRISPR RNAs (crRNAs) programmed against DNA from which they originated. Individual crRNAs are bound to Cas proteins from the interference module and the resulting “effector complex” recognizes foreign nucleic acids through complementary interactions between the targeted sequence (protospacer) and matching crRNA spacer. Unlike Cas1 and Cas2, the interference module proteins are highly diverse and this diversity forms a basis for classification of CRISPR-Cas systems into several types (Makarova et al., 2015). In DNA targeting Type I, Type II, and Type V CRISPR-Cas systems protospacer recognition requires, in addition to crRNA with complementary spacer, a protospacer adjacent motif (PAM) (Deveau et al., 2008; Mojica et al., 2009; Shmakov et al., 2015; Zetsche et al., 2015) recognized by interference module proteins (Sashital et al., 2012; Anders et al., 2014; Hayes et al., 2016). Upon target DNA recognition, a stable R-loop complex containing locally melted protospacer DNA and an RNA-DNA heteroduplex is formed (Jore et al., 2011; Szczelkun et al., 2014). R-loop formation is followed by target DNA destruction either by single-peptide effectors of Type II and V (Sapranauskas et al., 2011; Shmakov et al., 2015; Zetsche et al., 2015) or through recruitment of additional “executor” endonuclease (Cas3) in Type I systems (Sinkunas et al., 2011; Gong et al., 2014; Hochstrasser et al., 2014; Huo et al., 2014). Cas1 and Cas2 are not required for interference either *in vivo* (Brouns et al., 2008) or *in vitro* (Westra et al., 2012; Mulepati and Bailey, 2013).

Point mutations in protospacer or PAM decrease effector complex affinity (Semenova et al., 2011). *In vitro*, drops in binding affinity (measured as apparent equilibrium association constants) as large as 100-fold were reported (Semenova et al., 2011; Westra et al., 2012). Under pressure from CRISPR-Cas, mobile genetic elements rapidly accumulate such mutations, which allow them to escape CRISPR interference (Deveau et al., 2008; Semenova et al., 2011). In Type I systems spacer acquisition from DNA molecules containing such “escape” protospacers is very strongly stimulated compared to “naïve” adaptation from targets with no matches to crRNA spacers (Datsenko et al., 2012; Fineran et al., 2014; Li et al., 2014; Richter et al., 2014; Westra et al., 2015). This specific version of spacer acquisition is referred to as “primed CRISPR adaptation” (Datsenko et al., 2012). The dramatically different behavior of fully matched (“wild-type,” “wt”) and partially mismatched (“escape,” “esc”) protospacer targets in *E. coli* strains with inducible expression of *cas* genes is shown in **Figure 1**. As can be seen, the presence of protospacer fully matching the crRNA spacer and harboring a functional PAM decreases plasmid transformation efficiency in cells expressing *cas* genes at least two orders of magnitude compared to control plasmid without protospacer (**Figure 1B**). A point mutation in PAM-proximal “seed” region of the protospacer (Semenova et al., 2011) restores the transformation efficiency to control level. The experiment can be modified by transforming plasmids in uninduced cells (when *cas* genes are not expressed all plasmids



are transformed equally well). Next, transformed cells are allowed to grow in the absence of antibiotic that selects for plasmid maintenance and *cas* gene expression is induced. Upon growth in the presence of inducers, robust spacer adaptation is revealed in cultures harboring a plasmid with mismatched protospacer (**Figure 1C**, adaptation is detected by PCR, cells that acquired a spacer—and an additional copy of repeat—result in a longer amplification product). No such product is observed in cultures of cells transformed with a plasmid harboring fully matching protospacer with a functional PAM. Analysis of newly acquired spacers shows that most (90–95%) of them are complementary to the strand where the original priming protospacer was located (Datsenko et al., 2012; Swarts et al., 2012; Fineran et al., 2014; Shmakov et al., 2014). This is a hallmark of primed adaptation, since naïve adaptation reveals no such bias: spacers are chosen from both DNA strands with equal efficiency (Yosef et al., 2012). Primed adaptation requires not just Cas1 and Cas2, but also all other components of the effector complex (the subunits of Cascade: Cse1, Cse2, Cas7, Cas5, Cas6e, and the crRNA), and the Cas3 nuclease/helicase (Datsenko et al., 2012). Primed adaptation is highly beneficial to the host, as it leads to specific acquisition of spacers from genetic parasites that “learned” to evade defenses provided by earlier-acquired spacers.

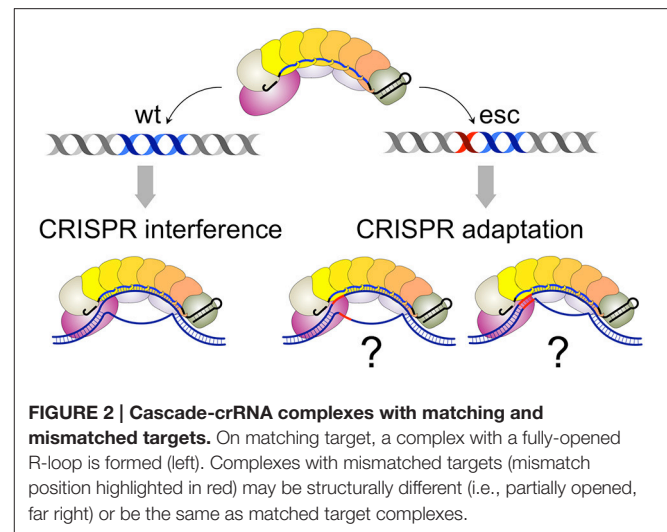
Primed adaptation clearly relies on specific recognition of partially matching protospacers by Cascade-crRNA effector. The dramatic difference in physiological consequences of recognition of fully and partially matching protospacers (interference vs. primed adaptation) raises a question of whether Cascade-crRNA complexes with two kinds of targets are also different from



each other structurally and functionally. Two hypothetical models of primed adaptation based on prior research from several laboratories have been recently put forward. One model summarized by Wright et al. (2016) envisions that effector complex bound to a partially matching target (priming protospacer) asymmetrically recruits Cas3 in the presence of Cas1-Cas2. According to this model, a complex of Cas3, Cas1, and Cas2 then dissociates from the effector complex bound at the priming protospacer and slides along the double-stranded DNA in either direction. As it slides, the Cas3-Cas1-Cas2 complex recognizes PAM sequences located *in cis* with respect to the priming protospacer, excises double-stranded protospacers and channels them for insertion into CRISPR array (Redding et al., 2015; Wright et al., 2016). In the second model, summarized by Amitai and Sorek (2016), binding of the effector complex to a partially matching target causes recruitment of Cas3. The latter directs, through an unspecified mechanism, the Cas1-Cas2 complex to target DNA. In the case of *E. coli* Type I-E system it is proposed that Cas1-Cas2 recognize PAM sequences and excise double-stranded protospacers located at both sides of the bound effector complex (Amitai and Sorek, 2016). Both models envision that naïve adaptation occurs when single-stranded DNA fragments generated by the RecBCD nuclease are bound by Cas1-Cas2, reannealed to form fully or partially double-stranded intermediates, and then processed for insertion into the CRISPR array. In both models Cas3 binding to the effector complex at the fully matching protospacer with a functional PAM causes target destruction without adaptation. In contrast, Cas3 binding to effector complex at partially mismatched priming protospacer leaves DNA bound to the effector complex intact (Wright et al., 2016).

Both models envision that effector complexes bound to fully matching, interference-competent, and partially mismatched, adaptation-competent protospacers are qualitatively different. The structure of Cascade-crRNA complex with a fully matching double-stranded target has been determined (Hayes et al., 2016). The structure reveals an R-loop that is formed as a result of the formation of a perfect 32-bp heteroduplex over the entire length of the crRNA spacer and complementary protospacer. Several studies suggest that the R-loop formation is initiated when the Cse1 subunit of Cascade recognizes the PAM sequence in double-stranded DNA (Sashital et al., 2012; Hochstrasser et al., 2014; Tay et al., 2015; Hayes et al., 2016). One can envision that the presence of mismatches between crRNA spacer and protospacer or imperfect recognition of non-consensus PAM by Cse1 significantly changes the structure, by altering the conformation of protein components, the extent of the R-loop, or both. No structures with mismatched complexes are available at the time of this writing. However, single-molecule analysis has indeed suggested that complexes with mismatched targets may be only partially open (Blosser et al., 2015; Figure 2). Different structures may thus explain the different consequences of effector binding to matched and mismatched protospacers shown in Figures 1B,C.

Recent data from several laboratories (Fineran et al., 2014; Xue et al., 2015) suggest that depending on the spacer-protospacer pair and the kind of mismatch mutation present,



a continuum of phenotypes (from 100% interference with no adaptation to efficient adaptation without visible interference) is observed in plasmid transformation interference/primed adaptation experiments similar to that shown in Figures 1B,C. If effector complexes are able to adopt two different functional conformations, the result would suggest that at different targets the relative proportion of such conformations is correspondingly different. However, there is an alternative view that considers interference and primed adaptation as intimately connected processes involving same complexes (Swarts et al., 2012; Semenova et al., 2016). According to this view interference provides substrates for the primed adaptation. Further, the rates of interference (target destruction) and spacer acquisition reactions can't be considered in isolation: copy number maintenance mechanisms of plasmids or phage genome targeted by CRISPR-Cas should also be considered. The latter become very important when one interprets the results of experiments as that shown in Figure 1. DNA harboring fully matching protospacers can be located by Cascade effector rapidly and then destroyed by Cas3. If copy number maintenance mechanisms are not able to keep up with the rate of interference, foreign DNA (and its degradation products) is rapidly purged from cells. If spacer insertion is a slow reaction (compared to the rate of degradation of target DNA by Cas3 and the rate of degradation of Cas3-generated products by cellular nucleases), no CRISPR array expansion is expected to occur in most cells in the culture.

The outcome becomes different when a mismatch mutation in a protospacer decreases the rate of R-loop complex formation, making CRISPR interference less efficient, giving foreign DNA replication/copy maintenance systems a chance to compensate for the pressure from CRISPR interference over extended periods of time and leading to production of phage progeny from infected cell (during phage infection) or continuous maintenance of plasmid in a clonal population of cells arising from a single transformed founder cell. Both the yield of phage particles and the plasmid copy number are expected to decrease compared to values in unprotected cells. The ongoing, "perpetual" interference process in both cases generates a stream of foreign DNA

degradation products that are maintained in the cells at a constant steady-state level. These products can be acted upon by adaptation proteins Cas1-Cas2. Preferential adaptation of spacers from foreign DNA becomes a default consequence of specific degradation of foreign (as opposed to host) DNA by the interference machinery. The strand bias observed during primed adaptation seems to suggest that at least initially, Cas3 generated products that are acted upon by Cas1-Cas2 are single-stranded. This would be consistent with the known 3'-5' polarity of nucleolytic action of Cas3 *in vitro* (Sinkunas et al., 2011). However, the strand bias may also result from Cas1-Cas2 association with Cas3 moving away from the priming site (Richter et al., 2014; Redding et al., 2015) and as such be independent of its exonuclease activity. Alternatively, the mechanism of spacer insertion into CRISPR array may itself be directional and result in orientation bias that is perceived as an apparent strand bias.

To show that an interplay between the rate of target recognition/destruction, spacer adaptation from products generated by target degradation, and target DNA copy number maintenance mechanisms can generate outcomes similar to the ones shown in **Figure 1** with just one kind of effector complexes, a simple numerical mathematical model of CRISPR-plasmid dynamics was elaborated. The model assumes that spacers are constantly acquired in CRISPR arrays provided that products of interference generated by effector complex binding to intact plasmids are present. The plasmid copy number  $P$  in CRISPR-free cells is assumed to be controlled by the logistic dynamics: the growth term is linear in  $P$  and depends on replication rate  $\alpha$  and the decay term, manifesting a feedback mechanism, which limits the growth when the plasmid number approaches its target value  $P_0$ , is quadratic in  $P$ :

$$\frac{dP}{dt} = \alpha P \left( 1 - \frac{P}{P_0} \right).$$

In the presence of Cas proteins and crRNA recognizing a plasmid protospacer, an additional decay term  $-\beta P$  that depends on CRISPR interference rate  $\beta$  is introduced:

$$\frac{dP}{dt} = \alpha P \left( 1 - \frac{P}{P_0} \right) - \beta P.$$

Thus, ongoing CRISPR interference changes the equilibrium number of plasmids from  $P_0$  to  $(\alpha - \beta)P_0/\alpha$  for  $\alpha > \beta$  and to 0 for  $\alpha < \beta$ . The number of plasmid fragments  $F$  which appear as a result of CRISPR interference is controlled by one gain term, which depends on the efficiency of interference, and two loss terms: the first one describes the intrinsic degradation of such fragments with the rate  $\delta$  and the second one accounts for acquisition of such fragments as spacers into the bacterial genome with the rate  $\chi$

$$\frac{dF}{dt} = C\beta P - (\delta + \chi)F.$$

The conversion factor  $C$  is the number of protospacers produced from one plasmid once it is recognized by the effector complex

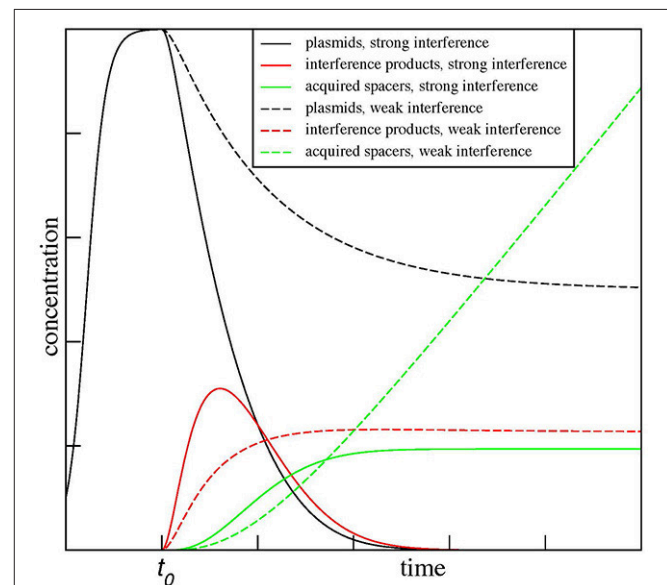
and degraded by Cas3. Finally, the number of spacers  $S$  acquired into an array is given by

$$\frac{dS}{dt} = \chi F.$$

To mimic a realistic induction scenario (**Figure 1C**), we first allow plasmids to reach their CRISPR-free stationary copy number  $P_0$ . Then at time  $t_0$  the CRISPR-Cas system is turned on by increasing  $\beta(t)$  and  $\chi(t)$  from zero to their stationary values  $\beta$  and  $\chi$  via a simple transitory regime with rate  $\rho$ , which mimics their zero-order production and first-order decay kinetics.

$$\begin{aligned}\beta(t) &= \beta(1 - e^{-\rho(t-t_0)}), \\ \chi(t) &= \chi(1 - e^{-\rho(t-t_0)}).\end{aligned}$$

The results for two values of  $\beta$ ,  $\beta = 0.5\alpha$ , and  $\beta = 1.5\alpha$ , are shown in **Figure 3**. As can be seen the model predicts dramatically different outcomes for the two conditions. When  $\beta < \alpha$  (a situation one can expect it for mismatched protospacer target), the plasmid copy number converges to a steady state. The level of degradation products proceeds through and initial increase and then settles on a plateau of its own. The number of spacer acquisition events (and, therefore, cells



**FIGURE 3 | Kinetic modeling of CRISPR interference and spacer acquisition at different ratios of CRISPR interference and foreign DNA copy number maintenance.** The modeling results show outcomes of CRISPR interference and adaptation depending on the rate of replication of foreign DNA targeted by the CRISPR-Cas system  $\alpha$  and the rate of CRISPR interference  $\beta$ . The results for two ratios of these rates,  $\beta = 0.5\alpha$  and  $\beta = 1.5\alpha$ , are shown. When  $\beta < \alpha$ , the plasmid number (dashed black line) converges to a steady state and spacers (dashed green line) are continuously acquired from plasmid degradation products (dashed red line). In contrast, when  $\beta > \alpha$ , the plasmid population dies out (solid black line) and only a few spacers (solid green line) are acquired by the population during a short time when plasmid degradation products (solid red line) are present. The plots are shown for  $\alpha = C = P_0 = \delta = 1$  and  $\chi = \rho = 0.1$ .

that underwent adaptation) increases linearly with time, with spacers continuously acquired from a constant pool of plasmid degradation products. Eventually, every cell in the population acquires at least one spacer. The newly acquired spacers are characterized by high level of interference rate  $\beta$  and so the plasmid is lost due to “secondary” interference, however, cells with expanded arrays remain in the population. In contrast, when  $\beta > \alpha$  from the very beginning (as expected for a fully matching spacer-protospacer pair), the plasmid population becomes extinct rapidly, while plasmid degradation products, generated by interference, accumulate sharply and then abruptly decline. Only a few cells in the population acquire spacer during a short time window when plasmid degradation products are present. It follows from the steady-state analysis of the equations above when  $\beta = \alpha/2$  the maximal rate of spacer acquisition is achieved. When  $\beta$  is zero and CRISPR-Cas system is inactive, spacers are not acquired at all.

Our simple mathematical model shows that experimentally observed and seemingly mutually exclusive outcomes of the kind shown in **Figure 1** can be achieved with minimal adjustments of parameters and without a requirements for two functionally different kinds of effector complexes at matching and mismatched protospacer targets.

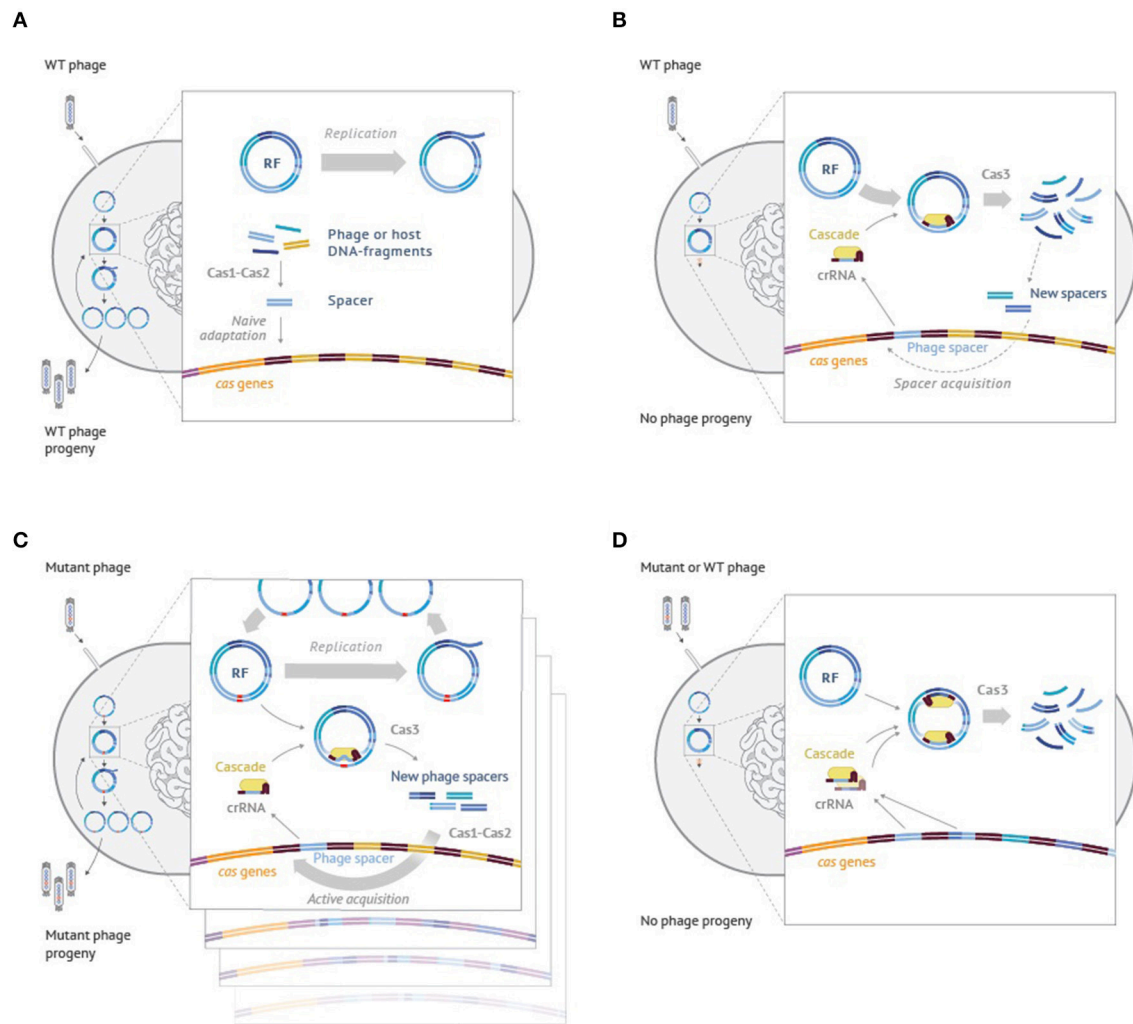
While the model was developed to explain CRISPR-Cas outcomes when targeting plasmids, similar logic can be used to explain the behavior of cells infected with the phage. Though mathematical modeling in this case becomes more complex and will be presented elsewhere, qualitatively, the idea is easy to grasp and is schematically illustrated using the example of M13 phage infection in **Figure 4**. When cells harboring no CRISPR spacers are infected with the phage the infection of most cells proceeds normally, resulting in phage progeny production (**Figure 4A**). Some of the infected cells acquire spacers due to the action of the Cas1-Cas2 adaptation complex, which indiscriminately inserts into the CRISPR array fragments of host or viral DNA that have been generated by the replication and/or recombination processes. Cells that acquired host-derived spacers undergo autoimmune death, which could be beneficial for the population of cells by limiting the spread of infection. Cells that acquired phage-derived spacers are able to destroy intracellular viral DNA and survive. They and their progeny can destroy incoming viral DNA (**Figure 4B**) through effector complex recognition followed by degradation mediated by Cas3. The destruction is rapid and while it generates viral DNA fragments that can be incorporated in the array this does not happened often as these fragments are short-lived. Incorporation of extra spacers can further increase the resistance levels of the host (Brouns et al., 2008; van Houte et al., 2016). Efficient CRISPR interference provides strong selection for phage harboring escape mutations (**Figure 4C**). Two independent processes unfold in cells infected by the mutant phage. On the one hand, the mutant phage genomes replicate. On the other hand, CRISPR interference also happens, which, however, is not efficient to allow full curing from the phage. As a result, a situation that is similar to the one described in the model of **Figure 3** is created. The infected cells will contain substantial steady-state levels of phage DNA destruction products, generated by Cas3, which can be used by the Cas1-Cas2 adaptation complex

for array expansion. Unlike the situation shown in **Figure 4A**, the adaptation process now becomes predominantly targeted to phage DNA. As a result, multiple clones containing various “second-generation” spacers in their arrays appear. Such clones become resistant to both the wild-type and the first generation escape phage, as is shown in **Figure 4D**.

In our model there is no need for the adaptation complex to slide away from the priming site and excise protospacers along its way as envisioned by existing models of primed adaptation (reviewed and summarized in Amitai and Sorek, 2016; Wright et al., 2016). Instead, spacers are selected from a common pool of independent, freely diffusible substrates, which are generated by the interference machinery and channeled for insertion in CRISPR array by the adaptation enzymes. The low level of naïve adaptation is caused by the rarity of appropriate substrates in the absence of CRISPR interference. The increased level of primed adaptation, and its preference for foreign DNA carrying the priming protospacer is a direct consequence of interference with such DNA by the effector complex and the Cas3 protein. The actual amount of the adaptation substrates (and, therefore, the extent of adaptation) results from an interplay of target protospacer binding by the effector complex, the degradation of DNA molecules bound to effector by Cas3, and the ability of foreign DNA replicons to counter interference by their intrinsic copy maintenance mechanisms. Poor adaptation from fully matching targets is a trivial consequence of their rapid destruction (Semenova et al., 2016).

Our model envisions that R-loop complexes formed on either fully matching or partially mismatched protospacers are either very similar or identical, differing only in times needed for them to form and/or recruit Cas3. Thus, the steady-state amount of such complexes formed on fast-replicating foreign DNA shall also be different. Once formed, open effector complexes recruit Cas3 that proceeds to destroy target DNA, moving progressively away from the protospacer. The lower binding of effector complexes to escape targets slows down the rate of reduction of copy number of plasmids used in standard transformation assays, allowing plasmid copy maintenance mechanism to offset, fully or partially, the CRISPR interference machinery action, essentially creating a condition of perpetual, albeit low-level interference. The products of Cas3 action are then approached by Cas1-Cas2 and channeled for insertion in the CRISPR array. The extent of removal of DNA around the interference site (and stability of Cas3-generated intermediates) may be affected by various non-CRISPR-Cas functions as indeed has been suggested by recent evidence (Ivancic-Bace et al., 2015; Levy et al., 2015). According to this view, the naïve and primed adaptation processes are mechanistically identical and both require only Cas1 and Cas2. The two processes only differ in the way the substrates for Cas1-Cas2 action are generated: through a highly inefficient and random aberrant processes during naïve adaptation or by a highly efficient, directional, and target- and strand-specific Cas3-mediated target DNA destruction during primed adaptation.

The moment a new spacer is acquired, cells raise the ante in this arms race, and the foreign genome is either purged or must respond by the accumulation of mutations in the



**FIGURE 4 | An interplay between interference and adaptation by Type I-E CRISPR-Cas system and phage DNA replication during the M13 phage infection of *Escherichia coli* cells. (A)** Wild-type M13 phage infects a cell lacking CRISPR spacers matching phage sequences. Phage progeny develops; rare cells undergo naïve adaptation, inserting spacers from the phage or host DNA. **(B)** Wild-type phage infects a cell that had earlier acquired a phage-derived spacer. Phage DNA is efficiently recognized by the effector complex and rapidly degraded by Cas3. No phage progeny is produced. On rare occasions, some short-lived Cas3 degradation products are used by the adaptation machinery causing acquisition of additional phage-derived spacers. **(C)** A mutant phage with an escape mutation infects a cell carrying a partially mismatched phage-derived spacer. Phage DNA is inefficiently recognized by the effector complex. While phage genomes replicate, some are degraded by Cas3. Cas3 degradation products are present throughout the infection and are used by the adaptation machinery causing highly efficient acquisition of additional phage-derived spacers (primed adaptation). **(D)** A wild-type or escape phage infect a cell carrying multiple phage-derived spacers. Phage DNA is recognized by the effector complex at multiple protospacers and degraded by Cas3. No phage progeny is produced. See text for more details.

protospacer and/or PAM corresponding to the newly acquired spacer. The indiscriminate nature of the adaptive response by the CRISPR-Cas system must lead to very rapid diversification of the initial clonal population, since individual cells will be acquiring different spacers from foreign DNA in the course of primed adaptation, as indeed observed in the recent study by van Houte et al. (2016). This in turn should drive corresponding diversification of resident plasmids or infecting viruses that infect such cells and evolve escape mutations or curing of bacterial culture. Mathematical modeling coupled to long-term cultivation experiments will be necessary to study the dynamics of such systems.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

## ACKNOWLEDGMENTS

II acknowledges support from FONDECYT grant 1151524. Work in KS laboratories is supported by an NIH grant GM10407, Skoltech and by Russian Science Foundation (No14-14-00988). We thank Visual Science for help with the Figure 4 preparation.



## REFERENCES

- Amitai, G., and Sorek, R. (2016). CRISPR-Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.* 14, 67–76. doi: 10.1038/nrmicro.2015.14
- Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513, 569–573. doi: 10.1038/nature13579
- Arslan, Z., Hermanns, V., Wurm, R., Wagner, R., and Pul, U. (2014). Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. *Nucleic Acids Res.* 42, 7884–7893. doi: 10.1093/nar/gku510
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712. doi: 10.1126/science.1138140
- Blosser, T. R., Loeff, L., Westra, E. R., Vlot, M., Kunne, T., Sobota, M., et al. (2015). Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Mol. Cell* 58, 60–70. doi: 10.1016/j.molcel.2015.01.028
- Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijskuis, R. J., Snijders, A. P., et al. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960–964. doi: 10.1126/science.1159689
- Datsenko, K. A., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* 3, 945. doi: 10.1038/ncomms1937
- Deveau, H., Barrangou, R., Garneau, J. E., Labonte, J., Fremaux, C., Boyaval, P., et al. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* 190, 1390–1400. doi: 10.1128/JB.01412-07
- Fineran, P. C., and Charpentier, E. (2012). Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* 434, 202–209. doi: 10.1016/j.virol.2012.10.003
- Fineran, P. C., Gerritzen, M. J., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S. A., et al. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 111, E1629–E1638. doi: 10.1073/pnas.1400071111
- Gong, B., Shin, M., Sun, J., Jung, C. H., Bolt, E. L., van der Oost, J., et al. (2014). Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16359–16364. doi: 10.1073/pnas.1410806111
- Hayes, R. P., Xiao, Y., Ding, F., van Erp, P. B., Rajashankar, K., Bailey, S., et al. (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from *E. coli*. *Nature* 530, 499–503. doi: 10.1038/nature16995
- Hochstrasser, M. L., Taylor, D. W., Bhat, P., Guegler, C. K., Sternberg, S. H., Nogales, E., et al. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6618–6623. doi: 10.1073/pnas.1405079111
- Huo, Y., Nam, K. H., Ding, F., Lee, H., Wu, L., Xiao, Y., et al. (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat. Struct. Mol. Biol.* 21, 771–777. doi: 10.1038/nsmb.2875
- Ivancic-Bace, I., Cass, S. D., Wearne, S. J., and Bolt, E. L. (2015). Different genome stability proteins underpin primed and naive adaptation in *E. coli* CRISPR-Cas immunity. *Nucleic Acids Res.* 43, 10821–10830. doi: 10.1093/nar/gkv1213
- Jore, M. M., Lundgren, M., van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P., et al. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* 18, 529–536. doi: 10.1038/nsmb.2019
- Levy, A., Goren, M. G., Yosef, I., Auster, O., Manor, M., Amitai, G., et al. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* 520, 505–510. doi: 10.1038/nature14302
- Li, M., Wang, R., Zhao, D., and Xiang, H. (2014). Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res.* 42, 2483–2492. doi: 10.1093/nar/gkt1154
- Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi: 10.1038/nrmicro3569
- Marraffini, L. A., and Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322, 1843–1845. doi: 10.1126/science.1165771
- Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155(Pt 3), 733–740. doi: 10.1099/mic.0.023960-0
- Mulepati, S., and Bailey, S. (2013). *In vitro* reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *J. Biol. Chem.* 288, 22184–22192. doi: 10.1074/jbc.M113.472233
- Nunez, J. K., Lee, A. S., Engelman, A., and Doudna, J. A. (2015). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* 519, 193–198. doi: 10.1038/nature14237
- Redding, S., Sternberg, S. H., Marshall, M., Gibb, B., Bhat, P., Guegler, C. K., et al. (2015). Surveillance and processing of foreign DNA by the *Escherichia coli* CRISPR-Cas System. *Cell* 163, 854–865. doi: 10.1016/j.cell.2015.10.003
- Richter, C., Dy, R. L., McKenzie, R. E., Watson, B. N., Taylor, C., Chang, J. T., et al. (2014). Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res.* 42, 8516–8526. doi: 10.1093/nar/gku527
- Sapranaukas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* 39, 9275–9282. doi: 10.1093/nar/gkr606
- Sashital, D. G., Wiedenheft, B., and Doudna, J. A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol. Cell* 46, 606–615. doi: 10.1016/j.molcel.2012.03.020
- Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., et al. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10098–10103. doi: 10.1073/pnas.1104144108
- Semenova, E., Savitskaya, E., Musharova, O., Strotskaya, A., Vorontsova, D. A., Datsenko, K. A., et al. (2016). Highly efficient primed spacer acquisition from targets destroyed by the *Escherichia coli* Type I-E CRISPR-Cas interfering complex. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7626–7631. doi: 10.1073/pnas.1602639113
- Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., et al. (2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol. Cell* 60, 385–397. doi: 10.1016/j.molcel.2015.10.008
- Shmakov, S., Savitskaya, E., Semenova, E., Logacheva, M. D., Datsenko, K. A., and Severinov, K. (2014). Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Res.* 42, 5907–5916. doi: 10.1093/nar/gku226
- Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* 30, 1335–1342. doi: 10.1038/emboj.2011.41
- Swarts, D. C., Mosterd, C., van Passel, M. W., and Brouns, S. J. (2012). CRISPR interference directs strand specific spacer acquisition. *PLoS ONE* 7:e35888. doi: 10.1371/journal.pone.0035888
- Szczelkun, M. D., Tikhomirova, M. S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., et al. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9798–9803. doi: 10.1073/pnas.1402597111
- Tay, M., Liu, S., and Yuan, Y. A. (2015). Crystal structure of *Thermobifida fusca* Cse1 reveals target DNA binding site. *Protein Sci.* 24, 236–245. doi: 10.1002/pro.2609
- van Houte, S., Ekroth, A. K., Broniewski, J. M., Chabas, H., Ashby, B., Bondy-Denomy, J., et al. (2016). The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* 532, 385–388. doi: 10.1038/nature17436
- Westra, E. R., van Erp, P. B., Kunne, T., Wong, S. P., Staals, R. H., Seegers, C. L., et al. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell* 46, 595–605. doi: 10.1016/j.molcel.2012.03.018
- Westra, E. R., van Houte, S., Oyesiku-Blakemore, S., Makin, B., Broniewski, J. M., Best, A., et al. (2015). Parasite exposure drives selective evolution of constitutive versus inducible defense. *Curr. Biol.* 25, 1043–1049. doi: 10.1016/j.cub.2015.01.065

- Wright, A. V., Nunez, J. K., and Doudna, J. A. (2016). Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* 164, 29–44. doi: 10.1016/j.cell.2015.12.035
- Xue, C., Seetharam, A. S., Musharova, O., Severinov, K., Brouns, S. J., Severin, A. J., et al. (2015). CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res.* 43, 10831–10847. doi: 10.1093/nar/gkv1259
- Yosef, I., Goren, M. G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* 40, 5569–5576. doi: 10.1093/nar/gks216
- Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771. doi: 10.1016/j.cell.2015.09.038

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MD declared a past co-authorship with one of the authors KS to the handling Editor, who ensured that the process met the standards of a fair and objective review.

Copyright © 2016 Severinov, Ispolatov and Semenova. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

