



OPEN ACCESS

EDITED BY

Grzegorz Wegrzyn,
University of Gdansk, Poland

REVIEWED BY

Davide Roncarati,
University of Bologna, Italy
Nicolas Sauvageot,
Université de Caen Normandie, France

*CORRESPONDENCE

Sofía Ruiz-Cruz,
✉ sofia.ruizcruz@ehu.eus
Alicia Bravo,
✉ abravo@cib.csic.es

†PRESENT ADDRESS

Ana Moreno-Blanco, Servicio de Microbiología del Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS) and CIBERINFEC, Madrid, Spain; Sofía Ruiz-Cruz, Facultad de Ciencia y Tecnología, Departamento de Bioquímica y Biología Molecular, Universidad del País Vasco, Leioa, Spain

RECEIVED 15 September 2023

ACCEPTED 30 November 2023

PUBLISHED 13 December 2023

CITATION

Moreno-Blanco A, Pluta R, Espinosa M, Ruiz-Cruz S and Bravo A (2023), Promoter DNA recognition by the *Enterococcus faecalis* global regulator MafR. *Front. Mol. Biosci.* 10:1294974. doi: 10.3389/fmolb.2023.1294974

COPYRIGHT

© 2023 Moreno-Blanco, Pluta, Espinosa, Ruiz-Cruz and Bravo. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) 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.

Promoter DNA recognition by the *Enterococcus faecalis* global regulator MafR

Ana Moreno-Blanco^{1†}, Radoslaw Pluta², Manuel Espinosa¹, Sofía Ruiz-Cruz^{1*†} and Alicia Bravo^{1*}

¹Centro de Investigaciones Biológicas Margarita Salas, Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain, ²Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain

When *Enterococcus faecalis* is exposed to changing environmental conditions, the expression of many genes is regulated at the transcriptional level. We reported previously that the enterococcal MafR protein causes genome-wide changes in the transcriptome. Here we show that MafR activates directly the transcription of the *OG1RF_10478* gene, which encodes a hypothetical protein of 111 amino acid residues. We have identified the *P10478* promoter and demonstrated that MafR enhances the efficiency of this promoter by binding to a DNA site that contains the -35 element. Moreover, our analysis of the *OG1RF_10478* protein AlphaFold model indicates high similarity to 1) structures of EIIb components of the bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system, and 2) structures of receiver domains that are found in response regulators of two-component signal transduction systems. However, unlike typical EIIb components, *OG1RF_10478* lacks a Cys or His residue at the conserved phosphorylation site, and, unlike typical receiver domains, *OG1RF_10478* lacks a conserved Asp residue at the position usually required for phosphorylation. Different from EIIb components and receiver domains, *OG1RF_10478* contains an insertion between residues 10 and 30 that, according to ColabFold prediction, may serve as a dimerization interface. We propose that *OG1RF_10478* could participate in regulatory functions by protein-protein interactions.

KEYWORDS

AlphaFold, *Enterococcus faecalis*, global regulators, phosphotransferase systems, protein-DNA interactions, two-component systems

1 Introduction

The Gram-positive bacterium *Enterococcus faecalis* is a member of the healthy human gut microbiota. Nevertheless, if it gains access to extra-intestinal niches (e.g., urinary tract, heart, or blood), it can cause deadly or serious diseases, especially in susceptible hosts (Banla et al., 2019; Kao and Kline, 2019). The opportunistic behaviour of *E. faecalis* (harmless commensal that may turn pathogenic) reflects its ability to modulate gene expression in response to specific environmental signals and host-imposed stresses (adaptive responses). This ability partially relies on proteins that act as global transcriptional regulators. In some

Abbreviations: BHI, Brain Heart Infusion; OD₆₅₀, optical density at 650 nm; PRD, PTS regulation domain; PTS, phosphotransferase system; REC domain, receiver domain; TCS, two-component signal transduction system; TY, tryptone-yeast extract.

cases, these regulators increase the activity of multiple promoters by binding to specific DNA sites that are located either upstream of or overlapping the main promoter elements (Browning and Busby, 2016; Bervoets and Charlier, 2019). Characterization of the DNA binding preferences of particular proteins has shown that some of them recognize intrinsic DNA structural characteristics (shape readout mechanism) rather than particular nucleotide sequences (base readout mechanism) (Rohs et al., 2010; Abe et al., 2015). Moreover, examples, where both direct DNA sequence readout and DNA shape recognition are required for protein binding, have also been reported (Deng et al., 2015; Ding et al., 2015; Al-Zyoud et al., 2016; Freda et al., 2023).

E. faecalis shows high levels of genetic diversity (Bakshi et al., 2016; He et al., 2018). The genome sequence of the well-studied OG1RF strain was published in 2008 (Bourgogne et al., 2008). Using genome-wide microarrays designed for this strain, we found that the enterococcal protein MafR (482 amino acid residues) functions as a global transcriptional regulator (Ruiz-Cruz et al., 2016). It influences positively the transcription of numerous genes, including genes that encode enzymes and transporters required for the utilization of carbon sources (e.g., mannitol, glycerol, gluconate, maltose, and citrate). Later, combining *in vivo* and *in vitro* approaches, we demonstrated that MafR acts as a transcription activator (Ruiz-Cruz et al., 2019). It activates directly the transcription of two genes: *OG1RF_12294*, which is predicted to encode a calcium-transporting P-type ATPase, and *OG1RF_11486*, which encodes a putative QueT transporter family protein. In both cases, MafR stimulates transcription by binding to a DNA site that overlaps promoter sequences. Because of these results, we proposed that MafR could have a regulatory role in calcium homeostasis and queuosine synthesis (Ruiz-Cruz et al., 2019). Through bioinformatics analyses, we have shown that MafR is highly conserved among many strains whose genomes have been sequenced (Ruiz-Cruz et al., 2018). Furthermore, gel filtration chromatography and analytical ultracentrifugation studies demonstrated that MafR behaves as a dimer in solution (Ruiz-Cruz et al., 2018). *In vitro* analyses of protein-DNA interactions established that 1) MafR binds to linear double-stranded DNAs in a nonsequence-specific manner (Ruiz-Cruz et al., 2018), 2) MafR binds to DNA sites that contain regions of potential bendability (Ruiz-Cruz et al., 2019), and 3) MafR generates multimeric complexes on linear double-stranded DNAs (Ruiz-Cruz et al., 2018).

The bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS) transports and phosphorylates numerous carbohydrates. It is usually composed of four cytoplasmic proteins (EI, HPr, EIIA, EIIB) and one membrane-spanning protein (EIIC). Unlike EI and HPr, the EIIA, EIIB, and EIIC components are normally specific for one substrate. To phosphorylate the carbohydrates during their transport, the cytoplasmic PTS proteins form a phosphorylation cascade starting with the phosphoenolpyruvate-mediated autophosphorylation of the EI protein (Deutscher et al., 2014). The PTS not only functions as a carbohydrate transporter but also regulates numerous cellular processes either by phosphorylating its target proteins or by interacting with them. For instance, some proteins have integrated a specific PTS-recognized phosphorylation domain known as the PTS regulation

domain (PRD). PRDs have been found in both transcription antiterminators and transcription activators that control the expression of genes involved in the utilization of carbohydrates (Deutscher et al., 2014; Galinier and Deutscher, 2017). Examples where unphosphorylated or phosphorylated PTS components (HPr, EIIA, and EIIB) interact with particular proteins, including transcription regulators, have also been reported. Such interactions can either stimulate or inhibit the function of the target proteins (Galinier and Deutscher, 2017).

MafR belongs to a family of transcriptional regulators that includes Mga from *Streptococcus pyogenes* (Hondorp et al., 2013), AtxA from *Bacillus anthracis* (Hammerstrom et al., 2015), and MgaSpn and PclR from *S. pneumoniae* (Solano-Collado et al., 2012; Moreno-Blanco et al., 2022). This family has been defined as a new class of PRD-containing regulators that have two N-terminal helix-turn-helix DNA-binding motifs (Hondorp and McIver, 2007; Tsvetanova et al., 2007; Hondorp et al., 2013; Hammerstrom et al., 2015). In the case of Mga, both motifs were found to be required for DNA-binding and transcriptional activation (McIver and Myles, 2002; Vahling and McIver, 2006). Moreover, the C-terminal region of the members of this family is predicted to have structural homology to PTS-EIIB components (EIIB-like domain). In AtxA and Mga, the C-terminal region was shown to be involved in self-association (Hammerstrom et al., 2011; Hondorp et al., 2012).

The two-component signal transduction systems (TCSs) play important roles in bacterial colonization and adaptation to new niches. In a prototypical TCS, the response regulator generates a specific cellular response to the signal detected by the cognate histidine kinase (Gao et al., 2019). Usually, the response regulators have a receiver (REC) domain and an effector domain. The histidine kinase phosphorylates the REC domain on a conserved Asp residue. In a majority of response regulators, the effector domain is a DNA-binding domain, however, response regulators having an RNA-binding domain or a protein-binding domain have also been reported. Still more, there are examples of response regulators that lack an effector domain (known as single-domain response regulators) (Galperin, 2006; Jenal and Galperin, 2009). Some of the latter regulators are unclassified because their encoding genes are not located in the vicinity of histidine kinase-encoding genes (Nguyen et al., 2015).

The enterococcal *OG1RF_10478* gene (new locus_tag *OG1RF_RS02540*) encodes a hypothetical protein of 111 amino acid residues (NCBI Reference Sequence WP_002358719.1). In this study, we demonstrate that MafR activates directly the transcription of this gene. We show that, in exponentially growing bacteria, the RNA polymerase recognizes the *P10478* promoter to initiate transcription of the *OG1RF_10478* gene. The activity of this promoter is enhanced in the presence of the MafR regulator. We also show that MafR interacts with a region that contains both the -35 element of the *P10478* promoter and a site that is needed for its activation by MafR. Furthermore, we have obtained an AlphaFold three-dimensional model of the *OG1RF_10478* protein and found that it has similarity to 1) structures of PTS-EIIB components, and 2) structures of REC domains present in response regulators of TCSs. We discuss that protein *OG1RF_10478* could have a regulatory role.

TABLE 1 Oligonucleotides used in this work.

Name	Sequence (5' to 3') ^a
F10478	CACTAATACAG <u>GAGCTCA</u> ATGTTGTCA
R10478	GCATACGCTTAG <u>GAGCTC</u> TAGAACT
F10478Δ122	CTGTTTCCAGTGAG <u>GAGCTC</u> TATTGTACCT
F10478Δ210	GATAGTTTAAAAAC <u>GAGCTC</u> CAAGCGTTGT
F10478Δ228	GATAGTTTAAAAACTATCACCAAGCGTTGTT <u>GAGCTC</u> GTCACCTGT
F10478Δ234	GATAGTTTAAAACTATCACCAAGCGTTGTTTACTAG <u>GAGCTC</u> GTAATAA
F10478Δ243	CTAGTCACTGTAG <u>GAGCTC</u> ATTTGTTTTT
R10478Δ-10	GAACTAAGTATAC <u>GAGCTC</u> AGCTTA
FrecA-q	GCAACGAAATGGTGGAACAG
RrecA-q	AAGGCATCGGCAATCTCTAAG
F10478-q	GAGCAGCAAAACATCAAAGCCT
R10478-q	AGATAGGGAGCGAGCATTTC
F10478-D	GCAAACGTGTTCCAGTGATAGT
R10478-D	CCACAAACAACGTCGCTCCATCA
F10478-S	GTCATTTTAGTTCCTCCCTATGT
R10478-S	GGCTTTGATGTTTTGCTGCTCT

^aRestriction sites are underlined, and the base changes that generate restriction sites are in bold.

2 Materials and Methods

2.1 Oligonucleotides, bacterial strains, and plasmids

The oligonucleotides used in this work are listed in Table 1. *E. faecalis* strains OG1RF (Bourgogne et al., 2008) and OG1RFΔmafR (Ruiz-Cruz et al., 2016) were used. OG1RFΔmafR lacks the mafR regulatory gene. Plasmid pDLF is a constitutive expression vector based on the enterococcal promoter P2493 (Ruiz-Cruz et al., 2010; Ruiz-Cruz et al., 2016). It carries a kanamycin resistance gene. Plasmid pDLFmafR is a pDLF derivative that carries the P2493:mafR fusion gene (Ruiz-Cruz et al., 2016). Plasmid pASTT is a promoter-probe vector based on the gfp reporter gene (Ruiz-Cruz et al., 2019). It carries a tetracycline resistance gene. The following pASTT-derivatives were constructed in this work. In all cases, a region of the OG1RF chromosome was amplified by PCR using the indicated primers. Then, the PCR product was digested with SacI, and the restriction fragment was ligated to the SacI-linearized pASTT vector: (a) pASTT-P10478 (primers F10478 and R10478, 301-bp restriction fragment), (b) pASTT-P10478Δ122 (primers F10478Δ122 and R10478, 177-bp restriction fragment), (c) pASTT-P10478Δ210 (primers F10478Δ210 and R10478, 90-bp restriction fragment), (d) pASTT-P10478Δ228 (primers F10478Δ228 and R10478, 73-bp restriction fragment), (e) pASTT-P10478Δ234 (primers F10478Δ234 and R10478, 67-bp restriction fragment), (f) pASTT-P10478Δ243 (primers F10478Δ243 and R10478, 58-bp restriction fragment), and (g) pASTT-P10478Δ-10 (primers F10478 and R10478Δ-10, 279-bp restriction fragment). For overproduction of MafR-His, an inducible expression system based

on the *Escherichia coli* strain BL21 (DE3) (a gift of F. W. Studier) and the pET24b-mafR-His plasmid (Ruiz-Cruz et al., 2018) was used. This plasmid encodes the MafR-His protein, which carries the Leu-Glu-6xHis peptide fused to its C-terminus.

2.2 Growth and transformation of bacteria

The *E. faecalis* strains were grown in Bacto™ Brain Heart Infusion (BHI) medium, at 37°C in a static water bath. For plasmid-harboring strains, the medium was supplemented with kanamycin (250 μg/mL; pDLF and its derivatives) and/or tetracycline (4 μg/mL; pASTT and its derivatives). The *E. coli* strain BL21 (DE3) harbouring plasmid pET24b-mafR-His was grown in tryptone-yeast extract (TY) medium supplemented with kanamycin (30 μg/mL), at 37°C in a shaking water bath. The protocol used to transform *E. faecalis* by electroporation was described (Shepard and Gilmore, 1995).

2.3 DNA and RNA isolation

Genomic DNA was prepared using the Bacterial Genomic Isolation Kit (Norgen Biotek Corporation). Plasmid DNA was prepared using the High Pure Plasmid Isolation Kit (Roche Applied Science) as described (Ruiz-Cruz et al., 2016). Total RNA was isolated using the RNeasy Mini Kit (QIAGEN) as reported (Ruiz-Cruz et al., 2016). The integrity of rRNAs was analysed by agarose gel electrophoresis. RNA concentration was determined using a NanoDrop ND-2000 Spectrophotometer.

2.4 Polymerase chain reaction

The Phusion High-Fidelity DNA polymerase (Thermo Scientific) and the Phusion HF buffer were used as described (Ruiz-Cruz et al., 2016). PCR products were purified using the QIAquick PCR Purification Kit (QIAGEN).

2.5 Quantitative reverse transcription PCR

For each strain, total RNA was isolated from three independent bacterial cultures. Then, from each RNA preparation, cDNA was synthesized using random primers and the iScript Select cDNA Synthesis Kit (Bio-Rad) as described (Ruiz-Cruz et al., 2016). To rule out the presence of genomic DNA in the RNA preparations, reactions without adding reverse transcriptase were performed. Quantitative PCRs were carried out using the iQ SYBR Green Supermix (Bio-Rad) and an iCycler Thermal Cycler (Bio-Rad) as reported (Ruiz-Cruz et al., 2016). From each cDNA sample, three PCRs per gene (gene of interest and internal control gene) were performed. Data were analysed with the iQTM5 Optical System Software. Relative quantification of gene expression was performed using the comparative C_T method (Schmittgen and Livak, 2008) as described (Moreno-Blanco et al., 2022). The *recA* gene (*OG1RF_12439*; recombination protein RecA) was used as the internal control gene (oligonucleotides *RecA-q* and *RrecA-q*). The oligonucleotides *F10478-q* and *R10478-q* (Table 1) were used to determine the relative expression of the *OG1RF_10478* gene. The threshold cycle values (C_T) of *OG1RF_10478* and *recA* were used to calculate $2^{-\Delta C_T}$, where $\Delta C_T = C_T \text{ } OG1RF_10478 - C_T \text{ } recA$. For the *OG1RF_10478* gene, the fold change in expression (FC) in one strain compared to another was obtained by dividing the corresponding mean $2^{-\Delta C_T}$ values as reported (Moreno-Blanco et al., 2022).

2.6 Primer extension

The oligonucleotide *R10478-D* was radioactively labelled at the 5'-end using [γ -³²P]-ATP (PerkinElmer) and T4 polynucleotide kinase (New England Biolabs) as reported (Solano-Collado et al., 2013). Primer extension reactions were performed using total RNA from the indicated bacterial strain, the ³²P-labelled oligonucleotide *R10478-D*, and the ThermoScript Reverse Transcriptase enzyme (Invitrogen) as described (Ruiz-Cruz et al., 2019). cDNA products were analysed by sequencing gel (8 M urea, 6% polyacrylamide) electrophoresis. Dideoxy-sequencing reactions (Sequenase Version 2.0 DNA Sequencing Kit; USB Corporation) were run in the same gel as DNA size markers. These reactions were carried out using a 422-bp PCR-amplified DNA fragment (*OG1RF* DNA and oligonucleotides *F10478-S* and *R10478-S*) as DNA template and the ³²P-labelled oligonucleotide *R10478-D* as primer. Labelled products were visualized using a Fujifilm Image Analyser FLA-3000.

2.7 Fluorescence assays

Plasmid-harboring enterococcal strains were grown as indicated above to an optical density at 650 nm (OD₆₅₀) of 0.4

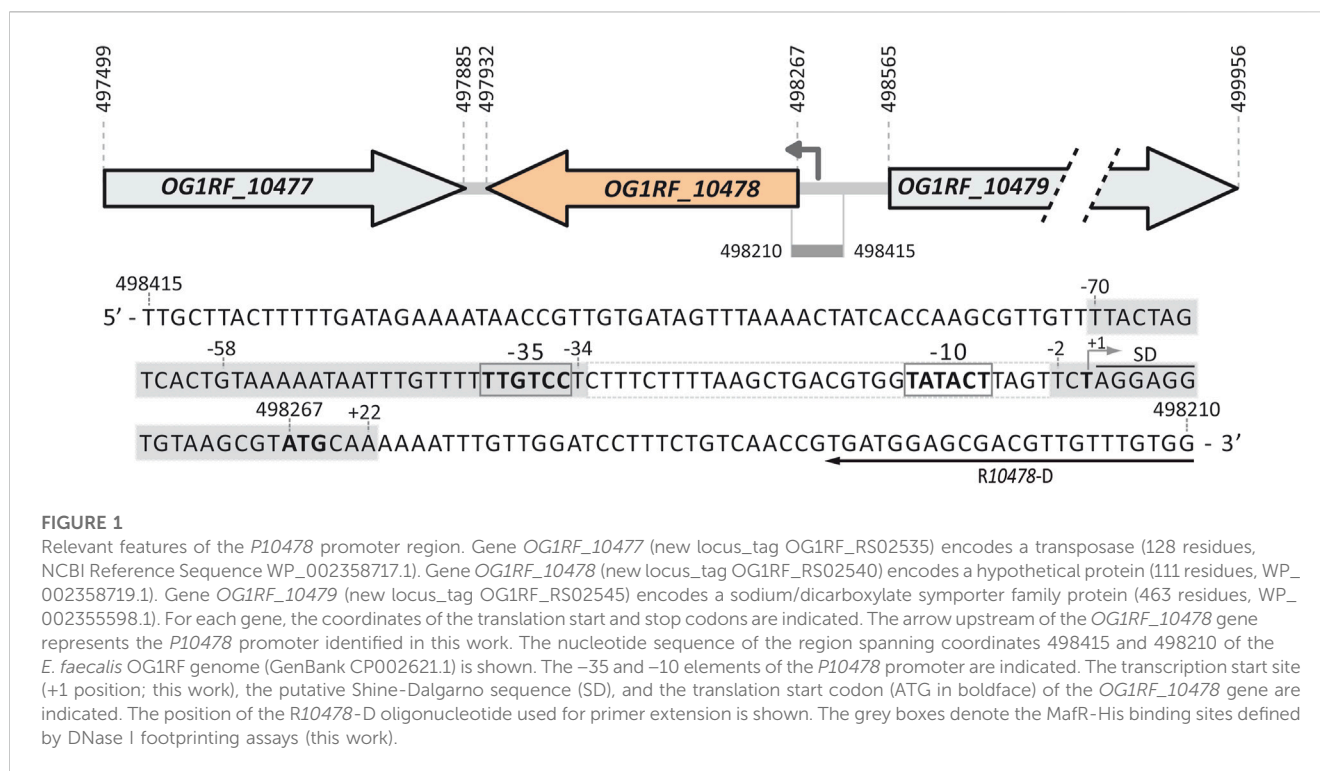
(exponential phase). Then, different volumes of culture (0.4–1 mL) were centrifuged. Cells were resuspended in 200 μ L of phosphate-buffered saline (PBS). For each strain, three independent cultures were analysed. Fluorescence intensity was measured using a Thermo Scientific Varioskan Flash instrument (excitation at 488 nm and emission at 515 nm).

2.8 Purification of MafR-His

The protocols used to overproduce and purify the MafR-His protein were described previously (Ruiz-Cruz et al., 2018). Purification of this protein involved the use of fast-pressure liquid chromatography (Biologic Duoflow, Bio-Rad) on a nickel affinity column (HisTrap HP column, GE Healthcare). Protein concentration was determined using a NanoDrop ND-2000 Spectrophotometer (Thermo Scientific).

2.9 DNase I footprinting assays

A 266-bp region of the *OG1RF* genome (coordinates 498475 to 498210) was amplified by PCR using the oligonucleotides *F10478-D* and *R10478-D* (Table 1). To ³²P-label the amplified DNA fragment at either the 5'-end of the coding strand or the 5'-end of the non-coding strand, either the oligonucleotide *F10478-D* or the oligonucleotide *R10478-D* was ³²P-labelled at the 5'-end as described (Solano-Collado et al., 2013). Binding reactions and DNase I digestion were performed as reported (Ruiz-Cruz et al., 2019). Briefly, as shown by our previous results (Ruiz-Cruz et al., 2018), MafR (and MafR-His) binds to its primary DNA site and then additional MafR units bind sequentially to the same DNA molecule, generating multiple complexes. Therefore, in the binding reactions, and previous to DNA digestion with DNase I, free DNA molecules and various types of MafR-DNA complexes coexist. Their proportion will depend on the concentration of MafR. To favour the presence of complexes constituted by MafR bound to its primary site versus higher-order MafR-DNA complexes, we have to work with low protein concentrations and, under such conditions, the amount of free DNA molecules in the reaction mixtures can be high. Samples were analysed by sequencing gel (8 M urea, 6% polyacrylamide) electrophoresis. Dideoxy-sequencing reactions (Sequenase Version 2.0 DNA Sequencing Kit; USB Corporation) were run in the same gel. These reactions were carried out using a 422-bp PCR-amplified DNA fragment (*OG1RF* DNA and oligonucleotides *F10478-S* and *R10478-S*) as DNA template and either the ³²P-labelled oligonucleotide *F10478-D* (sequence of the coding-strand relative to the *P10478* promoter) or the ³²P-labelled oligonucleotide *R10478-D* (sequence of the non-coding-strand) as primer. Labelled products were visualized using a Fujifilm Image Analyser FLA-3000 and the intensity of the bands was quantified using the Quantity One software (Bio-Rad). Briefly, to compare the lane corresponding to “300 nM MafR-His” with the lane corresponding to “no MafR-His,” we determined the relative intensity of each band detected at “300 nM MafR-His” with respect to the same band in the “no MafR-His” lane. Then, we compared the relative intensities of all the bands corresponding to “300 nM MafR-His” and we detected that some of the bands had a



lower relative intensity. Such bands define regions with changes in the relative sensitivity to DNase I (diminished cleavages; regions protected against DNase I digestion). This is a quantitative measurement that eliminates the differences in loading and defines the primary binding site of MafR.

2.10 Electrophoretic mobility shift assays

Binding reactions were performed as described (Ruiz-Cruz et al., 2018). The ³²P-labelled 266-bp DNA fragment (4 nM) was incubated with different amounts of MafR-His. Reaction mixtures were analysed by electrophoresis on native polyacrylamide (6%) gels. Labelled DNA was visualized using a Fujifilm Image Analyser FLA-3000.

2.11 Bioinformatics analyses

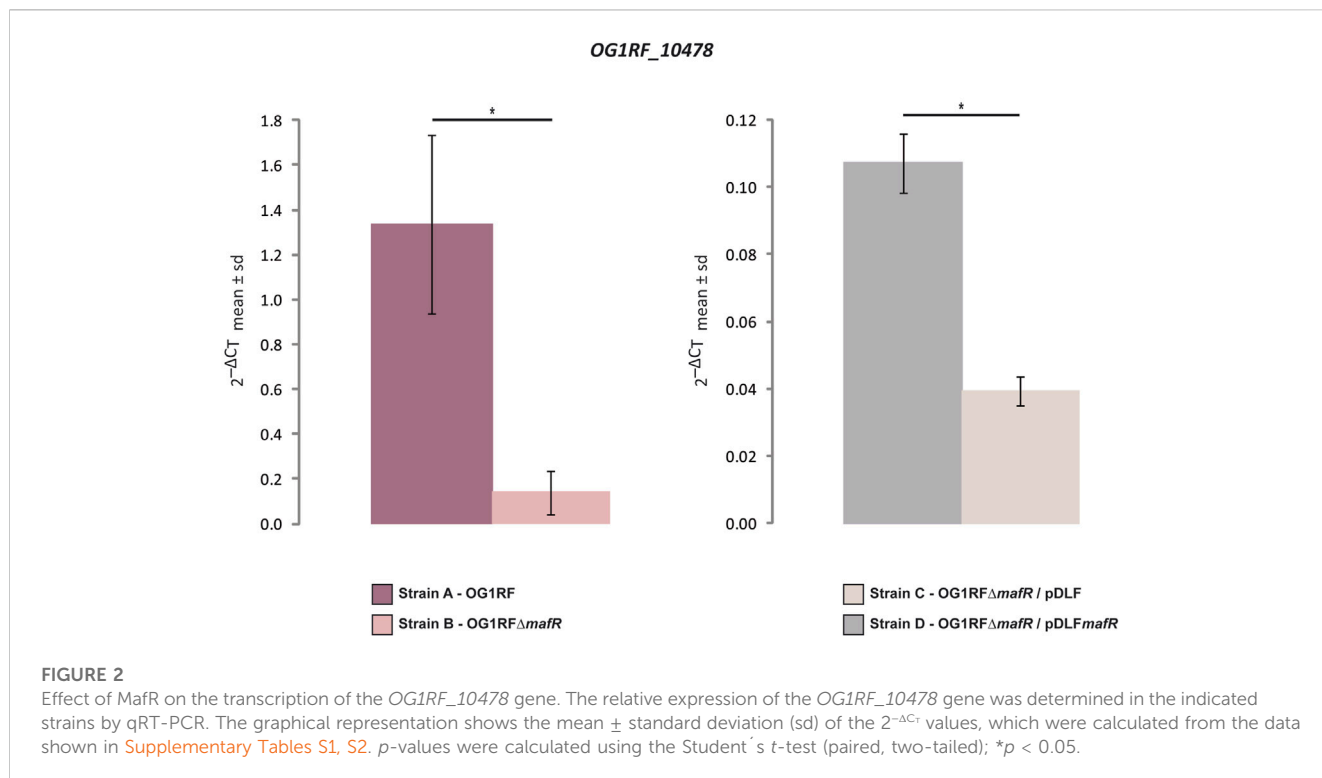
The bendability/curvature propensity plot was calculated with the bend it server (http://pongor.itk.ppke.hu/dna/bend_it.html) (Vlahovicek et al., 2003). The BPROM bacterial σ 70-dependent promoter prediction program (Softberry, Inc.) (<http://www.softberry.com>) was used. AlphaFold Protein Structure Database (Jumper et al., 2021; Varadi et al., 2022) and RoseTTAFold (Baek et al., 2021) were used to obtain three-dimensional models of the OG1RF_10478 protein (Uniprot Q837T7). The Dali server was employed for a search of similar structures (Holm, 2022). The ColabFold accelerated prediction of protein structures and complexes software (<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>) (Mirdita et al., 2022) was used to explore the dimerization potential of

OG1RF_10478. Figures, representing structural findings, were either directly taken from the abovementioned resources or generated using UCSF Chimera (Pettersen et al., 2004) and PyMol (Schrödinger, Inc.).

3 Results

3.1 MafR has a positive effect on the transcription of the *OG1RF_10478* gene

Our previous work showed that the MafR protein of *E. faecalis* causes genome-wide changes in the transcriptome (Ruiz-Cruz et al., 2016). Specifically, by DNA microarray studies, we compared the transcriptional profiles of the *E. faecalis* strains OG1RF (wild-type) and OG1RF Δ mafR (*mafR* deletion mutant) grown to mid-log phase under standard laboratory conditions (BHI broth, 37°C, without aeration). This analysis revealed that MafR activates the transcription of numerous genes, either directly or indirectly (Ruiz-Cruz et al., 2016). Among them, transcription of the *OG1RF_10478* gene (Figure 1), which encodes a hypothetical protein of 111 amino acid residues, was found to be higher (~7-fold) in the wild-type strain. In this work, we addressed the validation of such a finding by quantitative RT-PCR (qRT-PCR) assays using the comparative C_T method (Schmittgen and Livak, 2008). First, we determined the relative expression of the *OG1RF_10478* gene in the strains OG1RF and OG1RF Δ mafR using the *recA* gene as a reference gene (Figure 2 and Supplementary Table S1). The amount of *OG1RF_10478* transcripts was higher (~9.6-fold) in the wild-type strain. Next, we determined the relative expression of the *OG1RF_10478* gene in two plasmid-harboring strains that lack the chromosomal *mafR* gene: strain OG1RF Δ mafR/pDLF (absence of MafR) and strain OG1RF Δ mafR/pDLFmafR (plasmid-encoded



MafR) (Figure 2 and Supplementary Table S2). The amount of *OG1RF_10478* transcripts was higher (~2.7-fold) in the strain *OG1RFΔmafR/pDLFmafR*, indicating that plasmid-encoded MafR has a positive effect on the transcription of the *OG1RF_10478* gene under laboratory conditions.

3.2 MafR enhances the activity of the *P10478* promoter

The bacterial RNA polymerase core enzyme is a multi-subunit complex. It minimally consists of five subunits ($\alpha_2\beta\beta'$). However, a large number of bacteria, and particularly the Firmicutes, contain additional small subunits (δ and ϵ) (Weiss and Shaw, 2015). The RNA polymerase core enzyme interacts transitorily with one of a set of sigma factors forming the RNA polymerase holoenzyme. The promoter specificity of the RNA polymerase depends on its sigma factor. In exponentially growing bacterial cells, the majority of promoters are recognized by the housekeeping sigma factor (known as σ_{70} in *E. coli* and SigA in *E. faecalis*) (for a review see Browning and Busby, 2016). According to the BPROM bacterial σ_{70} -dependent promoter prediction program (Softberry, Inc.), the region of 650 nucleotides (nt) immediately upstream of the translation start site (coordinate 498267) of the *OG1RF_10478* gene has two potential σ_{70} -dependent promoters, here named *P10478* (proximal promoter) and *P2* (distal promoter) (Figure 3A). Compared to the consensus sequence of the promoters recognized by the *E. coli* σ_{70} factor, the proximal promoter *P10478* has a near-consensus -10 element (5'-TATACT-3') and shows a 4/6 match at the -35 element (5'-TTGtCc-3'). The spacer length between the two promoter elements is 22-nt, the optimal length being 17-nt. Moreover,

the *P10478* promoter has a 5'-TG-3' motif positioned one base upstream of the -10 element (the extended -10 element). In Gram-positive bacteria, numerous promoters have the 5'-TRTG-3' motif (R = purine) located at such a position (Voskuil and Chambliss, 1998; Voskuil and Chambliss, 2002; Solano-Collado et al., 2021). The distal promoter *P2* has also a near-consensus -10 element (5'-TACAAT-3') and shows a 3/6 match at the -35 element (5'-TTcAtt-3'). This predicted promoter has also non-optimal spacing (14-nt) between the two elements.

To identify transcription initiation sites of the *OG1RF_10478* gene, we performed primer extension assays using the oligonucleotide *R10478-D* as a primer (Figure 3A). To favour the presence of *OG1RF_10478* transcripts in the RNA preparations, we isolated total RNA from enterococcal cells that synthesize the MafR regulator: strain *OG1RF* (chromosome-encoded MafR) (Figure 3B, lane 2) and strain *OG1RFΔmafR/pDLFmafR* (plasmid-encoded MafR) (Figure 3B, lane 1). In both primer extension reactions, 73-, 74-, 126-, and 127-nt cDNA products were detected. The 73- and 74-nt cDNA products could be generated by reverse transcriptase running off at 5' ends of transcripts initiated at coordinates 498282 and 498283, respectively (Figure 3A). These coordinates are located 8- and 7-nt downstream of the -10 element of the *P10478* promoter. The 126- and 127-nt cDNA products could correspond to transcription initiation events at coordinates 498335 and 498336, respectively (Figure 3A). However, a σ_{70} -dependent promoter just upstream of such coordinates was not predicted, and could not be detected by visual inspection of the sequence. Alternatively, since there is an imperfect inverted repeat between coordinates 498336 and 498413 (Figure 3A), reverse transcriptase pausing at potential RNA structures could generate the 126- and 127-nt cDNAs. Thus, these results suggested that the enterococcal

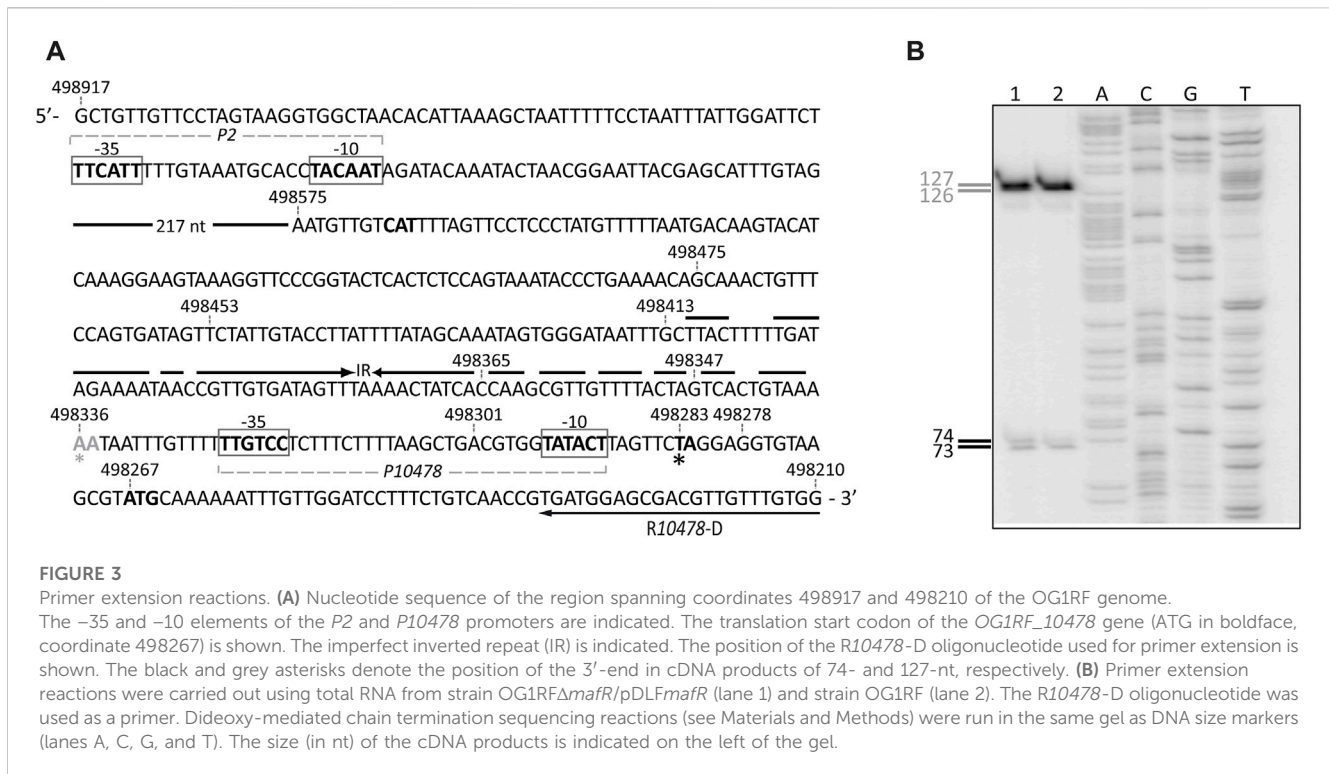


FIGURE 3

Primer extension reactions. **(A)** Nucleotide sequence of the region spanning coordinates 498917 and 498210 of the OG1RF genome. The -35 and -10 elements of the *P2* and *P10478* promoters are indicated. The translation start codon of the *OG1RF_10478* gene (ATG in boldface, coordinate 498267) is shown. The imperfect inverted repeat (IR) is indicated. The position of the *R10478-D* oligonucleotide used for primer extension is shown. The black and grey asterisks denote the position of the 3'-end in cDNA products of 74- and 127-nt, respectively. **(B)** Primer extension reactions were carried out using total RNA from strain OG1RF Δ *mafR*/pDLF*mafR* (lane 1) and strain OG1RF (lane 2). The *R10478-D* oligonucleotide was used as a primer. Dideoxy-mediated chain termination sequencing reactions (see Materials and Methods) were run in the same gel as DNA size markers (lanes A, C, G, and T). The size (in nt) of the cDNA products is indicated on the left of the gel.

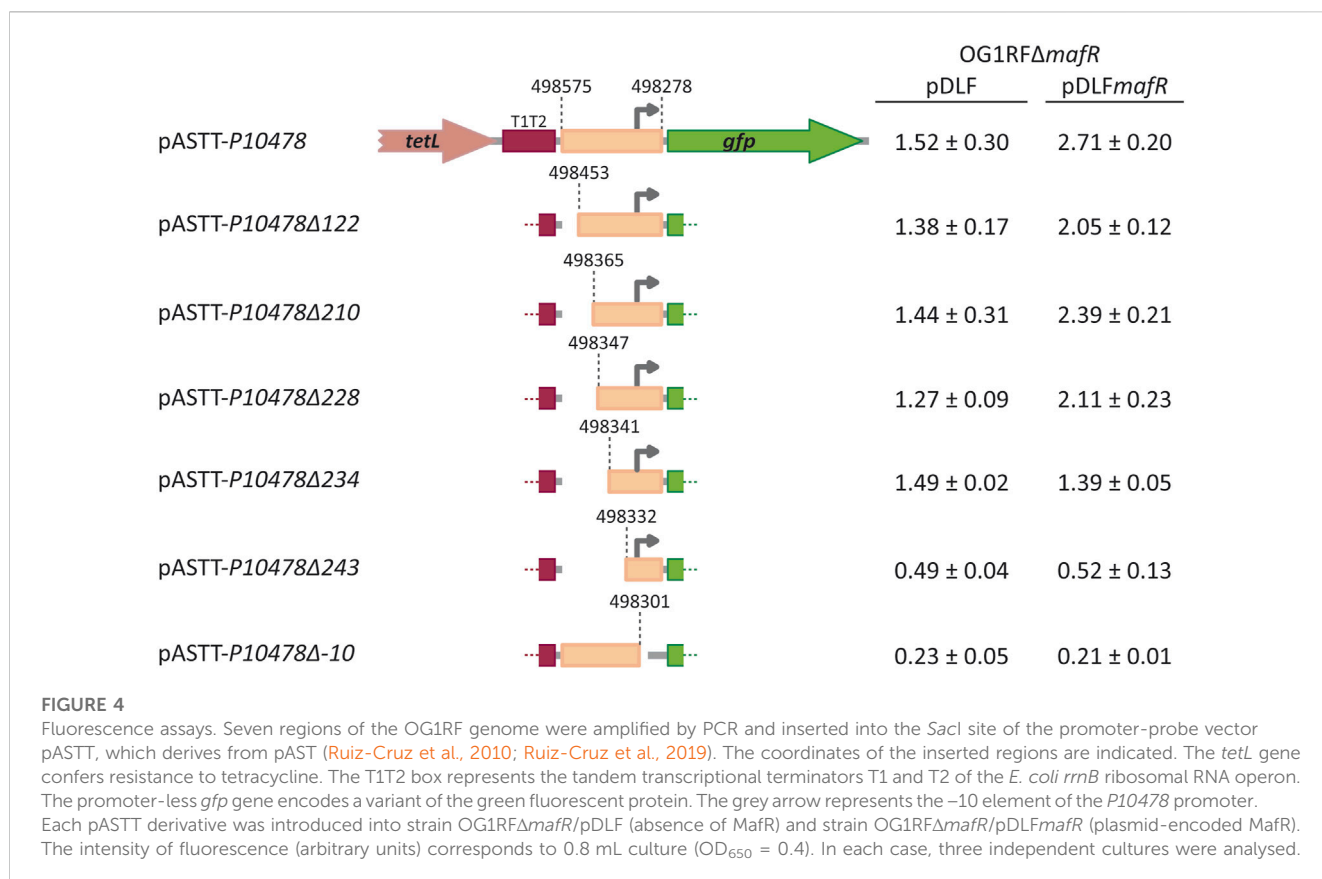
RNA polymerase could recognize two promoters to initiate transcription of the *OG1RF_10478* gene: the *P10478* promoter and a promoter located upstream of the coordinate 498336. The latter promoter could be either *P2* or another undetected promoter. This work has been focused on the characterization of the *P10478* promoter (see below). The transcription start site mapped at coordinate 498282 of the OG1RF genome (promoter *P10478*) coincides with the transcription start site mapped at coordinate 699386 of the *E. faecalis* V583 genome (Michaux et al., 2020). The latter transcription start site was mapped using the dRNA-seq technique in combination with ANNOgesic analysis and was associated with the *EF0743* gene (*OG1RF_10478* in strain OG1RF).

Further studies using promoter-reporter fusions (Figure 4) allowed us to conclude that protein MafR activates the transcription of the *OG1RF_10478* gene from the *P10478* promoter. Specifically, the 298-bp region spanning coordinates 498278 and 498575 (see Figure 3A) was inserted into the multicopy promoter-probe vector pASTT, which is based on the *gfp* reporter gene. The recombinant plasmid (pASTT-*P10478*) was then introduced into OG1RF (chromosome-encoded MafR) and OG1RF Δ *mafR* (absence of MafR). In both strains, the expression of *gfp* (1.54 ± 0.23 and 1.60 ± 0.16 units of fluorescence, respectively) was higher (~ 4 -fold) than the basal level (OG1RF harbouring pASTT, 0.38 ± 0.02 units). These results indicated that 1) the 298-bp region contains a promoter sequence, and 2) the chromosomal copy of *mafR* does not enhance the efficiency of such a promoter when it is located on a multicopy plasmid. However, different results were obtained when plasmid pASTT-*P10478* was introduced into OG1RF Δ *mafR*/pDLF (absence of MafR) and OG1RF Δ *mafR*/pDLF*mafR* (plasmid-encoded MafR) (Figure 4). The expression

of *gfp* was higher (~ 1.8 -fold) in OG1RF Δ *mafR*/pDLF*mafR*, indicating that the amount of MafR provided by pDLF*mafR* increases the activity of the promoter that drives the expression of *gfp* in plasmid pASTT-*P10478*. An additional deletion analysis (Figure 4, see also Figure 3A) showed that 1) deletion of the 24-bp region spanning coordinates 498278 and 498301 (-10 element of the *P10478* promoter) eliminates the promoter activity of the 298-bp region (plasmid pASTT-*P10478* Δ -10), 2) the 70-bp region spanning coordinates 498278 and 498347 contains both the promoter and the site required for its activation by MafR (plasmids pASTT-*P10478* Δ 122, pASTT-*P10478* Δ 210, and pASTT-*P10478* Δ 228), 3) the 7-bp region spanning coordinates 498341 and 498347 is essential for MafR-mediated activation of the promoter (plasmid pASTT-*P10478* Δ 234), and 4) deletion of the 10-bp region spanning coordinates 498332 and 498341 reduces to some extent the promoter activity (plasmid pASTT-*P10478* Δ 243), which suggests that the A+T-rich sequence located upstream of the -35 element of the *P10478* promoter could act as an UP element (see Figure 3A). The UP element is known to stimulate promoter activity by interacting with the RNA polymerase α -subunit (Estrem et al., 1998). Taken together, we conclude that MafR has a positive effect on the activity of the *P10478* promoter in bacteria growing under laboratory conditions.

3.3 MafR binds to the *P10478* promoter region

By EMSA, we have previously analysed the interaction of the MafR regulator with various linear double-stranded DNAs,



including DNAs without apparent sequence similarities (Ruiz-Cruz et al., 2018). On all tested DNAs, and like the MgaSpn regulator (Solano-Collado et al., 2013), MafR generated multimeric complexes. This study indicated that 1) multiple MafR units (likely dimers) bind sequentially to the DNA molecule, and 2) MafR binds to linear double-stranded DNAs in a manner that is not sequence-specific.

To analyse whether MafR recognized preferentially a site (primary binding site) on the *P10478* promoter region, we performed DNase I footprinting experiments. We used a His-tagged version of the MafR protein (MafR-His) and a 266-bp DNA fragment (coordinates 498475 to 498210; see Figure 3A), which contains both the *P10478* promoter and the sequence that is essential for its activation by MafR (coordinates 498341 to 498347; positions -58 to -64 of the *P10478* promoter) (see Figure 1, 4). The presence of a His-tag at the C-terminal end of MafR does not affect its DNA-binding properties (Ruiz-Cruz et al., 2018). The 266-bp DNA fragment was radioactively labelled either at the 5'-end of the coding strand or at the 5'-end of the non-coding strand, and the labelled DNA was then incubated with increasing concentrations of MafR-His (Figure 5). Storage phosphor imaging technology was used for quantitative comparisons of the radioactively labelled products (see Materials and Methods). On the coding strand and at 300 nM of MafR-His, changes in DNase I sensitivity (diminished cleavages) were observed from position -69 to -57 , and from position $+5$ to $+22$ relative to the transcription initiation site of the *P10478* promoter. On the non-coding strand and at 300 nM of MafR-His,

diminished cleavages were observed from position -70 to -59 , from -40 to -34 , and from -2 to $+20$. Therefore, protected sequences against DNase I digestion were observed mainly in two regions: between positions -70 and -34 (region A) and between positions -2 and $+22$ (region B). Region A includes both the -35 element of the *P10478* promoter and the sequence that is essential for its activation by MafR (positions -58 to -64). Region B is located just downstream of the -10 element. On both DNA strands and at 600 nM of MafR-His, regions protected against DNase I digestion were observed along the DNA fragment, which suggested that, upon binding to the primary site, additional MafR-His units interacted with the adjacent DNA regions. This result is consistent with the ability of MafR-His to generate multimeric complexes on linear double-stranded DNAs (Ruiz-Cruz et al., 2018). Specifically, we performed EMSA experiments with the same 266-bp DNA fragment that was used in the DNase I footprinting assays (Supplementary Figure S1). The radioactively labelled DNA was incubated with different concentrations of MafR-His. At 300 nM of MafR-His, free DNA and four protein-DNA complexes were detected. Moreover, as the protein concentration was increased, such complexes disappeared and higher-order complexes appeared.

Using the bend.it server (pongor.itk.ppke.hu/dna/bend_it.html) (Vlahovicek et al., 2003), we calculated the bendability/curvature propensity plot of the 266-bp DNA fragment. As shown in Supplementary Figure S2, the profile contains regions of potential bendability around the -10 element of the *P10478* promoter. Figure 6 shows a nucleotide sequence

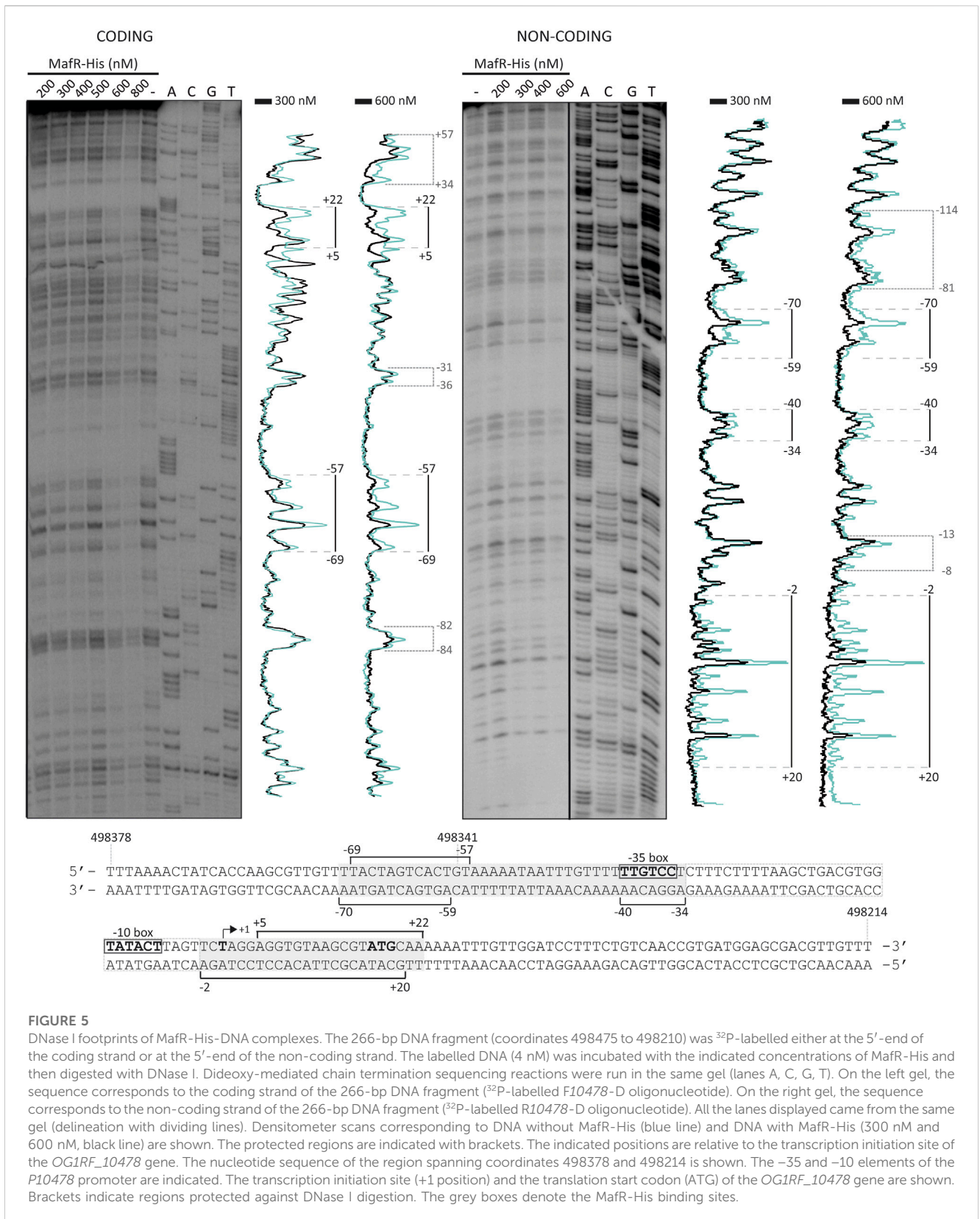


FIGURE 5

DNase I footprints of MafR-His-DNA complexes. The 266-bp DNA fragment (coordinates 498475 to 498210) was ³²P-labelled either at the 5'-end of the coding strand or at the 5'-end of the non-coding strand. The labelled DNA (4 nM) was incubated with the indicated concentrations of MafR-His and then digested with DNase I. Dideoxy-mediated chain termination sequencing reactions were run in the same gel (lanes A, C, G, T). On the left gel, the sequence corresponds to the coding strand of the 266-bp DNA fragment (³²P-labelled F10478-D oligonucleotide). On the right gel, the sequence corresponds to the non-coding strand of the 266-bp DNA fragment (³²P-labelled R10478-D oligonucleotide). All the lanes displayed came from the same gel (delineation with dividing lines). Densitometer scans corresponding to DNA without MafR-His (blue line) and DNA with MafR-His (300 nM and 600 nM, black line) are shown. The protected regions are indicated with brackets. The indicated positions are relative to the transcription initiation site of the *OG1RF_10478* gene. The nucleotide sequence of the region spanning coordinates 498378 and 498214 is shown. The -35 and -10 elements of the *P10478* promoter are indicated. The transcription initiation site (+1 position) and the translation start codon (ATG) of the *OG1RF_10478* gene are shown. Brackets indicate regions protected against DNase I digestion. The grey boxes denote the MafR-His binding sites.

alignment of the regions recognized by MafR on the *P10478* (region A, this work), *P11486*, and *P12294* (Ruiz-Cruz et al., 2019) promoters. The three regions, defined by DNase I

footprinting assays, contain the -35 promoter element. This alignment revealed that the three MafR binding sites have a low sequence identity: they share the motif TG (T/A)AA (A/T)(A/T)A.



FIGURE 6

MafR binding sites. (A) Nucleotide sequence alignment of DNA sites recognized by MafR on three promoter regions. Such MafR binding sites are indicated with grey boxes. They have been defined by DNase I footprinting assays (Ruiz-Cruz et al., 2019; and this work). The -35 and -10 elements of each promoter are shown. Identical nucleotides are highlighted in boldface. (B) Nucleotide sequence of the DNA site recognized by MafR on the *Pma* promoter region (grey box, positions -69 to -104) (Ruiz-Cruz et al., 2018). Nucleotides shared with the MafR binding sites shown in (A) are indicated in boldface.

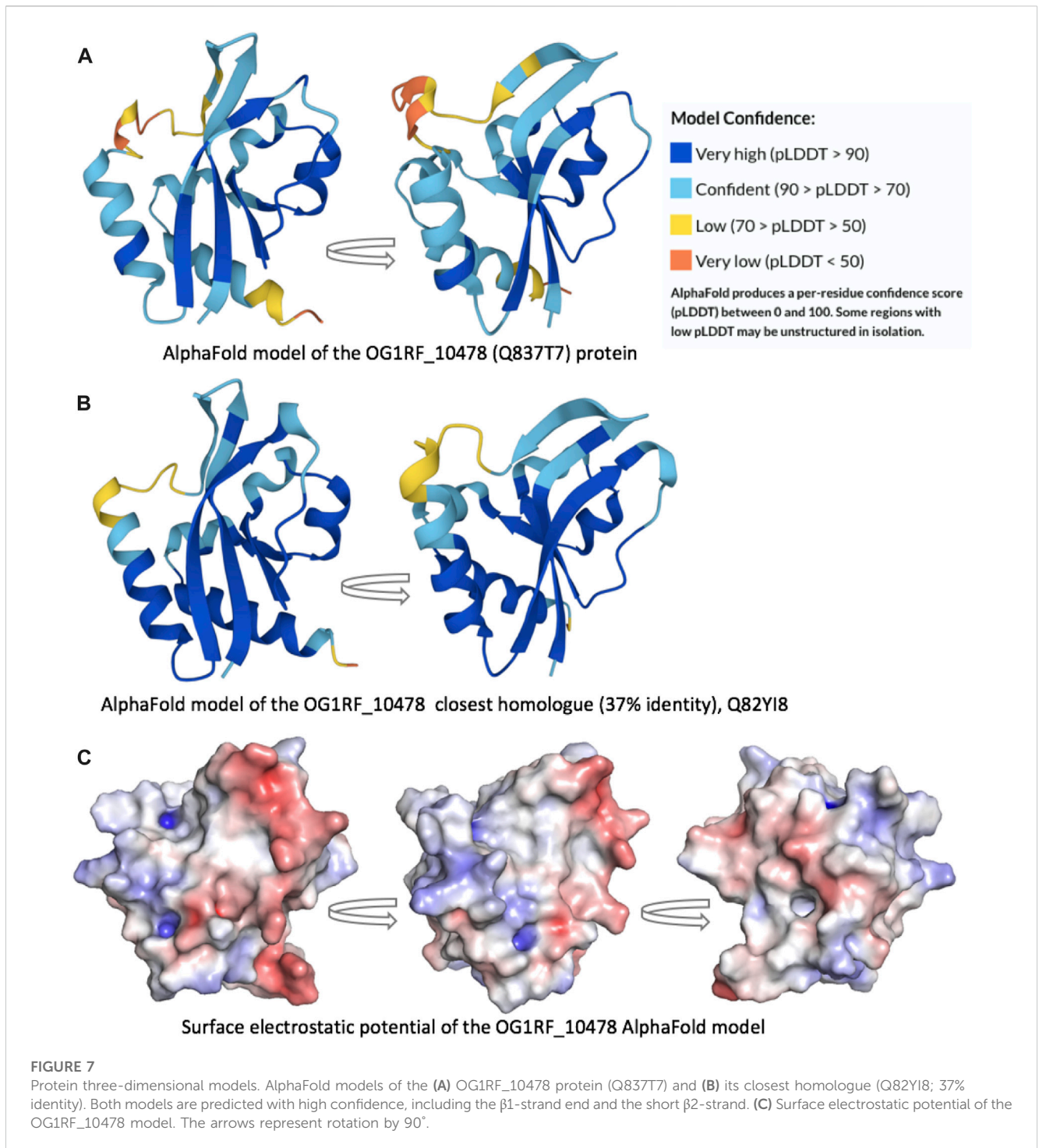
3.4 Bioinformatics analysis of protein OG1RF_10478

In the OG1RF genome (Bourgogne et al., 2008), the ATG codon at coordinate 498267 is likely the translation start site of the *OG1RF_10478* gene (Figure 1). Translation from this codon, which is preceded by a canonical ribosome-binding site (AGGAGG), would generate a protein of 111 amino acids (NCBI Reference Sequence WP_002358719.1). Using the BLASTP protein sequence alignment program (Altschul et al., 1997), we were unable to predict the biological function of the OG1RF_10478 protein. However, the use of additional bioinformatics tools (see below) has allowed us to establish that the predicted three-dimensional structure of the OG1RF_10478 protein is highly similar to 1) structures of PTS-EIIB components, and 2) structures of REC domains present in response regulators of TCSs.

The EIIB component of a PTS is usually specific for one substrate or, in a few cases, for a small group of closely related carbohydrates (Deutscher et al., 2014). The PTS-EIIB components are phosphorylated at a conserved residue located at the end of or right after the β 1-strand (residue around position number 10). In most cases, the role of phosphoryl group acceptor is played by a Cys residue, except for the EIIB components of the mannose PTS family, in which this role is played by a His residue. The phosphorylated EIIB component transfers its phosphoryl group to a carbohydrate molecule bound to the cognate EIIC component (Deutscher et al., 2014). According to the HHpred server for protein homology detection and structure prediction (Söding et al., 2005), the *OG1RF_10478* gene could encode a PTS-EIIB component. The three hits with the lowest E-values corresponded to 1) the lactose-specific PTS-EIIB component structure (PDB 3NBM) from *S. pneumoniae* (E-value = 0.015), 2) the cellobiose-specific PTS-EIIB component structure (PDB 4MGE) from *B. anthracis* (E-value = 0.017), and 3) the N,N'-diacetylchitobiose (Chb)-specific PTS-EIIB component structure (PDB 1IIB) from *E. coli* (E-value = 0.054). Nevertheless, unlike most of the PTS-EIIB components, the OG1RF_10478 protein lacks a Cys or His residue at the conserved position usually required for phosphorylation (Leu in OG1RF_10478), suggesting that the OG1RF_10478 protein could be an atypical PTS-EIIB component.

To obtain three-dimensional models of the OG1RF_10478 protein (Uniprot Q837T7), we used the AlphaFold Protein Structure Database (Jumper et al., 2021; Varadi et al., 2022) and RoseTTAFold (Baek et al., 2021) due to their ability to accurately predict protein structures. Both computational methods predicted similar OG1RF_10478 structures (2.7 Å RMSD) of the α/β fold class (Figure 7A and Supplementary Figure S3). The main difference between both models was found in the region spanning amino acid residues 10 and 30. Such a region had a low confidence score in the RoseTTAFold model. In the AlphaFold model, the region spanning residues 20 and 30 had also a low confidence score, while the confidently predicted part (residues 10-19) formed an extension of the β 1-strand and a short β 2-strand, which together with the top part of the β 3-strand established a small β -sheet. Noteworthy, the AlphaFold model of an OG1RF_10478 close homologue (37% identity; Uniprot Q82YI8; NCBI Reference Sequence WP_002382829.1), which is an uncharacterized protein from *E. faecalis*, exhibits a nearly identical structure (1.1 Å RMSD) and even higher prediction confidence level for the above-mentioned region and in general (Figure 7B). Analysis of the electrostatic surface potential of the OG1RF_10478 model shows that the protein has more negatively than positively charged patches on its surface (Figure 7C; Figure 8), which agrees with its low theoretical isoelectric point (pI: 4.9).

Using the Dali server (Holm, 2022), which performs searches of similar three-dimensional protein structures, we found that the AlphaFold model of OG1RF_10478 is highly similar to 1) structures of PTS-EIIB components (in agreement with the HHpred predictions, see above), and 2) structures of REC domains present in response regulators of TCSs. Specifically, the top ten Dali server hits include three PTS-EIIB components and seven response regulators (Table 2 and Supplementary Figure S4). The REC domain of the response regulators normally includes a conserved Asp residue at the end of the β 3-strand that can be phosphorylated by the cognate histidine kinase (Gao et al., 2019). Nevertheless, some REC domains lack such a conserved Asp residue (Maule et al., 2015; Desai and Kenney, 2017). Moreover, although most of the TCS response regulators have an effector domain involved in DNA binding, response regulators made only of the REC domain have been identified (Nguyen et al., 2015). The



OG1RF_10478 protein lacks not only the conserved Asp residue (having an Asn residue instead) (Supplementary Figure S5) but also an effector domain, suggesting that OG1RF_10478 might function as a non-canonical response regulator. Nonetheless, based on the sequence identity level between 1) OG1RF_10478 and the top EIIB hit (Chb-specific IIB component from *E. coli*; 19% identity), and 2) OG1RF_10478 and the top response regulator hit (atypical response regulator BaeR from *E. coli*; 15% identity), we hypothesize that protein OG1RF_10478 probably evolved from an EIIB protein rather than from a response regulator.

When we compared the AlphaFold model of OG1RF_10478 to the structures of the proteins shown in Table 2, we found that the region between residues 10 and 30 in OG1RF_10478 is absent from both the PTS-EIIB components and the response regulators (Supplementary Figure S6). Interestingly, using the ColabFold accelerated prediction of protein structures and complexes software (<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>) (Mirdita et al., 2022), we found that such a region of OG1RF_10478 may function as a homo-dimerization interface (Figure 8). In the predicted dimer

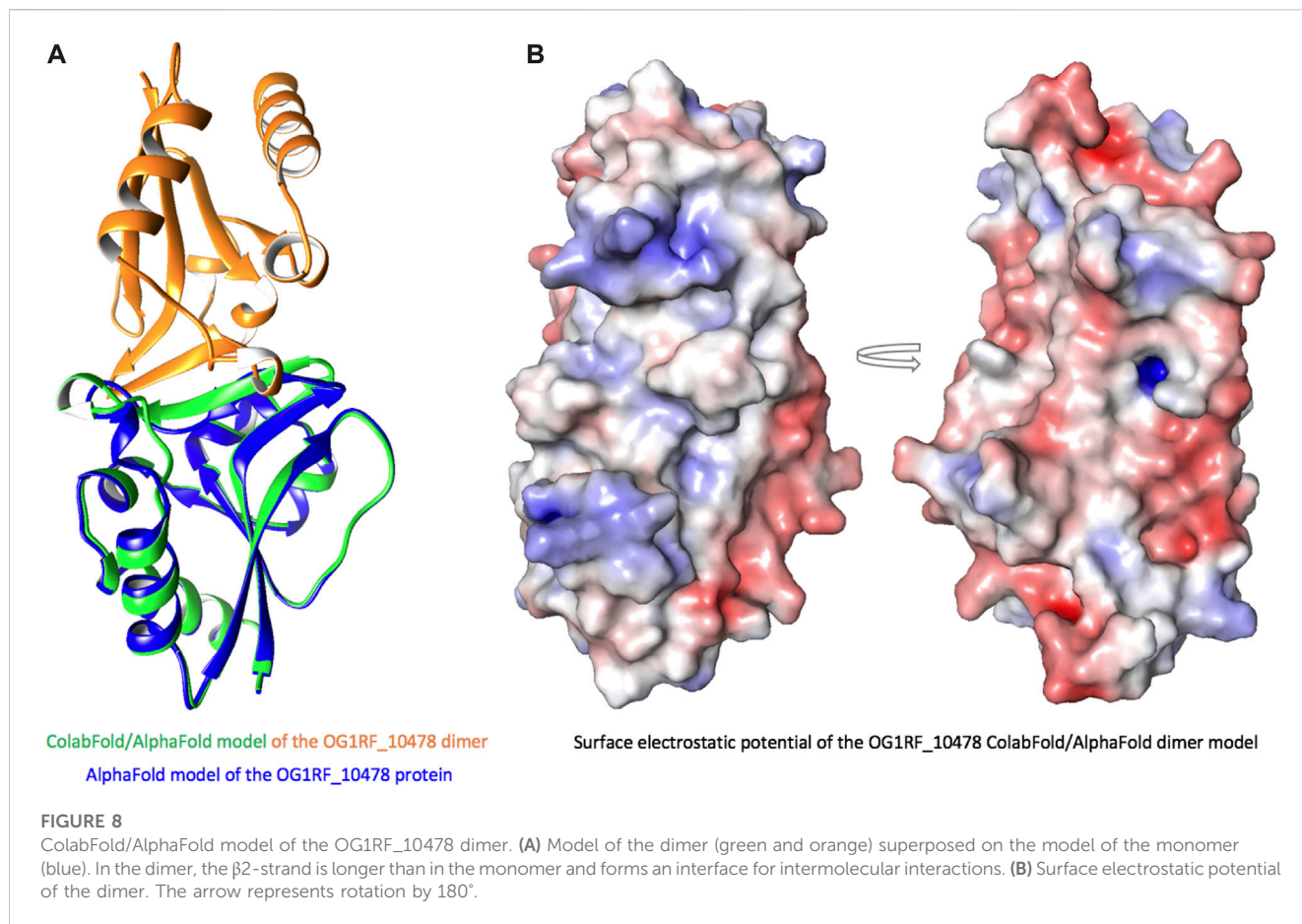


TABLE 2 Top 10 Dali server hits.

Protein	Type	PDB	Z-score	RMSD (Å)	Seq. iden. (%)
N,N'-diacetylchitobiose (Chb)-specific IIB component from <i>Escherichia coli</i>	EIIB	2WY2, 1H9C, 1IIB	7.1, 7.0, 6.8	2.7, 2.7, 2.4	19, 19, 19
Antibiotic resistance-associated atypical unphosphorylated dimeric BaeR from <i>Escherichia coli</i>	TCS	4B09	6.7	2.7	15
Cellobiose-specific IIB component from <i>Bacillus anthracis</i>	EIIB	4MGE	6.5	3.0	13
Biofilm-controlling BfmR from <i>Acinetobacter baumannii</i>	TCS	6VBF, 5E3J, 5HM6	6.4, 5.9, 5.9	2.7, 2.8, 2.8	7, 9, 9
Adhesion, biofilm, and virulence-controlling ArlR from <i>Staphylococcus aureus</i>	TCS	6IS1	6.4	2.8	12
Fructose-specific EIIB component from <i>Escherichia coli</i>	EIIB	4TN5	6.4	2.8	6
Lantibiotic resistance-associated NsTCS from <i>Streptococcus agalactiae</i>	TCS	5DCL	6.3	2.8	13
Virulence and antibiotic resistance-controlling VbTCS from <i>Vibrio parahaemolyticus</i>	TCS	7E1H	6.2	2.7	10
Colistin resistance-controlling PmrA from <i>Acinetobacter baumannii</i>	TCS	7M0S	6.2	2.8	15
Diguanylate cyclase TCS from <i>Syntrophotalea carbinolica</i>	TCS	3N53	6.2	2.9	10

model, the dimerization interface consists of the $\beta 2$ -strands from both protomers (the region that is absent in EIIBs and response regulators) and the $\beta 2$ - $\beta 2'$ interactions promote elongation of the $\beta 2$ -strands in both protomers. In addition to self-dimerization, this region could potentially also serve as a platform for interactions with other proteins.

4 Discussion

To generate an adaptive response to fluctuating environmental conditions, the expression of numerous bacterial genes is regulated at the transcriptional level. The promoter sequence of a gene determines significantly the basal transcription initiation

frequency. Nevertheless, this frequency may be enhanced and/or reduced by proteins that bind to specific DNA sites, usually close to or even overlapping the binding site of the RNA polymerase (Browning and Busby, 2016; Bervoets and Charlier, 2019). Our previous research established that the *E. faecalis* MafR protein 1) causes genome-wide changes in the transcriptome (Ruiz-Cruz et al., 2016), and 2) functions as a transcription activator (Ruiz-Cruz et al., 2019). Moreover, we have reported that MafR (482 residues) forms dimers in solution and, most likely, binds to its target DNAs as a dimer (Ruiz-Cruz et al., 2018). Now, by qRT-PCR experiments, we have shown that MafR influences positively the transcription of the *OG1RF_10478* gene in bacteria growing exponentially under standard laboratory conditions. Moreover, by primer extension, promoter-reporter fusions, and DNase I footprinting assays, we have identified the *P10478* promoter and demonstrated that MafR stimulates transcription from such a promoter by binding to a DNA site that contains the -35 element (positions -70 to -34 ; region A). MafR also interacts with a region that is located downstream of the -10 element (positions -2 to $+22$; region B). These results suggest that MafR might enhance the activity of the *P10478* promoter by recruitment of the RNA polymerase through direct contacts with the sigma factor, as it has been reported for some activators that bind to a site that overlaps the -35 element (Browning and Busby, 2016; Bervoets and Charlier, 2019). Additionally, we speculate that a dimer of MafR might contact the two DNA regions (A and B) and induce conformational changes in the *P10478* promoter, thereby enabling the recognition of the promoter by the RNA polymerase. This mechanism usually involves the binding of the activator to promoters that have non-optimal spacing between the -35 and -10 elements (Browning and Busby, 2016; Bervoets and Charlier, 2019), which is a characteristic of the *P10478* promoter. Regulatory proteins able to repress and activate transcription by differentially modulating local DNA structure within the promoter have been described (Philips et al., 2015). Different from MafR, the pneumococcal *MgaSpn* and *PclR* activators (also members of the *Mga/AtxA* family) bind to a DNA site that is located upstream of the main promoter elements. *MgaSpn* recognizes a site located between positions -99 and -60 of the *P1623B* promoter (Solano-Collado et al., 2013) and *PclR* recognizes a site located between positions -169 and -68 of the *PpclA* promoter (Moreno-Blanco et al., 2022). Activators that bind to a DNA site located upstream of the promoter typically recruit the RNA polymerase by interacting with the α -subunit (Browning and Busby, 2016; Bervoets and Charlier, 2019).

MafR-mediated transcriptional activation has also been reported for the *P11486* and *P12294* promoters (Ruiz-Cruz et al., 2019). In both cases, MafR binds to a DNA site that contains the -35 promoter element, as well as regions of potential bendability. Furthermore, studies on the *Pma* promoter of the *mafR* gene have shown that MafR interacts with a site that is located 1) upstream of the main promoter elements (positions -104 to -69), and 2) adjacent to the peak of a potential intrinsic curvature (Ruiz-Cruz et al., 2018). However, the function (if any) of this interaction remains unknown. As shown in Figure 6, the DNA sites recognized by MafR on the *P10478*, *P11486*, *P12294*, and *Pma* promoters have a low sequence identity, which suggests that MafR could recognize structural

features in its target DNAs rather than a specific nucleotide sequence. Several studies support that recognition of local DNA conformations (shape readout mechanism) contributes to the DNA-binding specificity of the regulators that constitute the *Mga/AtxA* family (Hadjifrangiskou and Koehler, 2008; Hause and McIver, 2012; Solano-Collado et al., 2013; Ruiz-Cruz et al., 2018). In the case of the pneumococcal *MgaSpn* regulator, *in vitro* DNA binding studies have shown that it binds to linear double-stranded DNAs with little or no sequence specificity (Solano-Collado et al., 2013). Moreover, *MgaSpn* has a preference for AT-rich DNA sites (Solano-Collado et al., 2016) and for DNA regions that contain a potential intrinsic curvature (Solano-Collado et al., 2013).

What could be the role of the *OG1RF_10478* protein? We postulate that it could have a regulatory function. This suggestion is based on the function of proteins whose three-dimensional structure is similar to the *OG1RF_10478* model predicted with AlphaFold (this work). First, we have found significant structural similarity between *OG1RF_10478* and some PTS-EIIB components. However, different from such components, *OG1RF_10478* lacks a conserved Cys or His residue at the position usually required for PTS-mediated phosphorylation. This finding suggests that the putative regulatory role of *OG1RF_10478* could be achieved by protein-protein interactions, as it has been shown for some unphosphorylated PTS-EIIB components that, either as distinct proteins or fused to another PTS component, control the activity of particular transcription regulators by interacting with them (Deutscher et al., 2014; Galinier and Deutscher, 2017). In line with our hypothesis, the ColabFold software (Mirdita et al., 2022) predicts that a region of *OG1RF_10478* could serve as a platform for interactions with other proteins (this work). Moreover, it has been pointed out that PTS-mediated regulation by protein-protein interactions is likely more frequent than by phosphorylation (Deutscher et al., 2014). Second, we have found structural similarity between *OG1RF_10478* and the REC domain of some TCS response regulators. Nevertheless, unlike the majority of such regulators (Gao et al., 2019), *OG1RF_10478* lacks an effector domain. Moreover, the REC domain typically has a conserved Asp residue that is missing in the *OG1RF_10478* protein. Phosphorylation of this residue by the cognate histidine kinase is critical for the activity of the response regulator (Gao et al., 2019). Therefore, *OG1RF_10478* could be considered a non-canonical response regulator that consists of a stand-alone REC domain without the conserved Asp residue. Numerous studies on TCSs have identified different types of non-canonical response regulators. For instance, there are regulators that only have a REC domain with the conserved Asp residue, the so-called single-domain response regulators (Jenal and Galperin, 2009). The phosphorylated forms of some of these regulators play key regulatory roles when they interact with other proteins. In these cases, the phosphoryl group can stay on the REC domain or be transferred to another protein (Galperin, 2006; Jenal and Galperin, 2009; Nguyen et al., 2015). Furthermore, there are response regulators that lack the conserved Asp residue of the REC domain and then regulation of their activity is not via phosphorylation. To regulate the activity of this class of regulators, different strategies have been identified (Maule et al., 2015; Gao et al., 2019). Examples of response regulators

whose unphosphorylated and phosphorylated forms have distinct regulatory roles have also been reported (Desai and Kenney, 2017).

In conclusion, our present work demonstrates that MafR activates the transcription of the *OG1RF_10478* gene by binding to a DNA site that contains the -35 element of the *P10478* promoter. This DNA site has a low sequence identity with the MafR binding sites identified previously. Moreover, our study shows that the predicted three-dimensional model of the *OG1RF_10478* protein is similar to the structure of some PTS-EIIB components and to the structure of the REC domain of some TCS response regulators. Based on these similarities, we propose that *OG1RF_10478* could be a regulatory protein.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Author contributions

AM-B: Investigation, Writing–review and editing. RP: Formal Analysis, Investigation, Writing–review and editing. ME: Conceptualization, Funding acquisition, Writing–review and editing. SR-C: Conceptualization, Investigation, Supervision, Writing–review and editing. AB: Conceptualization, Funding acquisition, Project administration, Supervision, Writing–original draft, Writing–review and editing.

References

- Abe, N., Dror, I., Yang, L., Slattery, M., Zhou, T., Bussemaker, H. J., et al. (2015). Decoupling the recognition of DNA shape from sequence. *Cell* 161, 307–318. doi:10.1016/j.cell.2015.02.008
- 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
- Al-Zyoud, W. A., Hynson, R. M. G., Ganuelas, L. A., Coster, A. C. F., Duff, A. P., Baker, M. A. B., et al. (2016). Binding of transcription factor GabR to DNA requires recognition of DNA shape at a location distinct from its cognate binding site. *Nucleic Acids Res.* 44, 1411–1420. doi:10.1093/nar/gkv1466
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876. doi:10.1126/science.abb3754
- Bakshi, U., Sarkar, M., Paul, S., and Dutta, C. (2016). Assessment of virulence potential of uncharacterized *Enterococcus faecalis* strains using pan genomic approach - identification of pathogen-specific and habitat-specific genes. *Sci. Rep.* 6, 38648. doi:10.1038/srep38648
- Banla, L. I., Salzman, N. H., and Kristich, C. J. (2019). Colonization of the mammalian intestinal tract by enterococci. *Curr. Opin. Microbiol.* 47, 26–31. doi:10.1016/j.mib.2018.10.005
- Bervoets, I., and Charlier, D. (2019). Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology. *FEMS Microbiol. Rev.* 43, 304–339. doi:10.1093/femsre/fuz001
- Bourgogne, A., Garsin, D. A., Qin, X., Singh, K. V., Sillanpää, J., Yerrapragada, S., et al. (2008). Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain OG1RF. *Genome Biol.* 9, R110. doi:10.1186/gb-2008-9-7-r110
- Browning, D. F., and Busby, S. J. W. (2016). Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* 14, 638–650. doi:10.1038/nrmicro.2016.103
- Deng, Z., Wang, Q., Liu, Z., Zhang, M., Machado, A. C. D., Chiu, T.-P., et al. (2015). Mechanistic insights into metal ion activation and operator recognition by the ferric uptake regulator. *Nat. Commun.* 6, 7642. doi:10.1038/ncomms8642
- Desai, S. K., and Kenney, L. J. (2017). To $-P$ or Not to $-P$? Non-canonical activation by two-component response regulators. *Mol. Microbiol.* 103, 203–213. doi:10.1111/mmi.13532
- Deutscher, J., Aké, F. M. D., 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 protein-protein interactions. *Microbiol. Mol. Biol. Rev.* 78, 231–256. doi:10.1128/mbr.00001-14
- Ding, P., McFarland, K. A., Jin, S., Tong, G., Duan, B., Yang, A., et al. (2015). A novel AT-rich DNA recognition mechanism for bacterial xenogeneic silencer MvaT. *PLoS Pathog.* 11, e1004967. doi:10.1371/journal.ppat.1004967
- Estrem, S. T., Gaal, T., Ross, W., and Gourse, R. L. (1998). Identification of an UP element consensus sequence for bacterial promoters. *Proc. Natl. Acad. Sci. U. S. A.* 95, 9761–9766. doi:10.1073/pnas.95.17.9761
- Freda, I., Exertier, C., Barile, A., Chaves-Sanjuan, A., Vega, M. V., Isupov, M. N., et al. (2023). Structural insights into the DNA recognition mechanism by the bacterial transcription factor PdxR. *Nucleic Acids Res.* 51, 8237–8254. doi:10.1093/nar/gkad552
- Galnier, A., and Deutscher, J. (2017). Sophisticated regulation of transcriptional factors by the bacterial phosphoenolpyruvate:sugar phosphotransferase system. *J. Mol. Biol.* 429, 773–789. doi:10.1016/j.jmb.2017.02.006
- Galperin, M. Y. (2006). Structural classification of bacterial response regulators: diversity of output domains and domain combinations. *J. Bacteriol.* 188, 4169–4182. doi:10.1128/jb.01887-05
- Gao, R., Bouillet, S., and Stock, A. M. (2019). Structural basis of response regulator function. *Annu. Rev. Microbiol.* 73, 175–197. doi:10.1146/annurev-micro-020518-115931

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by grant PID 2019-104553RB-C21 (I + D + i) to AB from the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033).

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2023.1294974/full#supplementary-material>

- Hadjiifrangiskou, 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
- Hammerstrom, T. G., Roh, J. H., Nikonowicz, E. P., and Koehler, T. M. (2011). *Bacillus anthracis* virulence regulator AtxA: oligomeric state, function and CO₂-signalling. *Mol. Microbiol.* 82, 634–647. doi:10.1111/j.1365-2958.2011.07843.x
- 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
- He, Q., Hou, Q., Wang, Y., Li, J., Li, W., Kwok, L.-Y., et al. (2018). Comparative genomic analysis of *Enterococcus faecalis*: insights into their environmental adaptations. *BMC Genomics* 19, 527. doi:10.1186/s12864-018-4887-3
- Holm, L. (2022). Dali server: structural unification of protein families. *Nucleic Acids Res.* 50, W210–W215. doi:10.1093/nar/gkac387
- 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
- Hondorp, E. R., Hou, S. C., Hempstead, A. D., Hause, L. L., Beckett, D. M., and McIver, K. S. (2012). Characterization of the Group A *Streptococcus* Mga virulence regulator reveals a role for the C-terminal region in oligomerization and transcriptional activation. *Mol. Microbiol.* 83, 953–967. doi:10.1111/j.1365-2958.2012.07980.x
- Hondorp, E. R., and McIver, K. S. (2007). The Mga virulence regulon: infection where the grass is greener. *Mol. Microbiol.* 66, 1056–1065. doi:10.1111/j.1365-2958.2007.06006.x
- Jenal, U., and Galperin, M. Y. (2009). Single domain response regulators: molecular switches with emerging roles in cell organization and dynamics. *Curr. Opin. Microbiol.* 12, 152–160. doi:10.1016/j.mib.2009.01.010
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2
- Kao, P. H. N., and Kline, K. A. (2019). Dr. Jekyll and Mr. Hide: how *Enterococcus faecalis* subverts the host immune response to cause infection. *J. Mol. Biol.* 431, 2932–2945. doi:10.1016/j.jmb.2019.05.030
- Maule, A. F., Wright, D. P., Weiner, J. J., Han, L., Peterson, F. C., Volkman, B. F., et al. (2015). The aspartate-less receiver (ALR) domains: distribution, structure and function. *PLoS Pathog.* 11, e1004795. doi:10.1371/journal.ppat.1004795
- McIver, K. S., and Myles, R. L. (2002). Two DNA-binding domains of Mga are required for virulence gene activation in the Group A *Streptococcus*. *Mol. Microbiol.* 43, 1591–1601. doi:10.1046/j.1365-2958.2002.02849.x
- Michaux, C., Hansen, E. E., Jenniches, L., Gerovac, M., Barquist, L., and Vogel, J. (2020). Single-nucleotide RNA maps for the two major nosocomial pathogens *Enterococcus faecalis* and *Enterococcus faecium*. *Front. Cell. Infect. Microbiol.* 10, 600325. doi:10.3389/fcimb.2020.600325
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat. Methods* 19, 679–682. doi:10.1038/s41592-022-01488-1
- Moreno-Blanco, A., Solano-Collado, V., Ortuno-Camuñas, A., Espinosa, M., Ruiz-Cruz, S., and Bravo, A. (2022). PclR is a transcriptional activator of the gene that encodes the pneumococcal collagen-like protein PclA. *Sci. Rep.* 12, 11827. doi:10.1038/s41598-022-15758-7
- Nguyen, M.-P., Yoon, J.-M., Cho, M.-H., and Lee, S.-W. (2015). Prokaryotic 2-component systems and the OmpR/PhoB superfamily. *Can. J. Microbiol.* 61, 799–810. doi:10.1139/cjm-2015-0345
- Petersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera-A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi:10.1002/jcc.20084
- Phillips, S. J., Canalizo-Hernández, M., Yildirim, I., Schatz, G. C., Mondragón, A., and O'Halloran, T. V. (2015). TRANSCRIPTION. Allosteric transcriptional regulation via changes in the overall topology of the core promoter. *Science* 349, 877–881. doi:10.1126/science.aaa9809
- 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., Moreno-Blanco, A., Espinosa, M., and Bravo, A. (2018). DNA-binding properties of MafR, a global regulator of *Enterococcus faecalis*. *FEBS Lett.* 592, 1412–1425. doi:10.1002/1873-3468.13032
- Ruiz-Cruz, S., Moreno-Blanco, A., Espinosa, M., and Bravo, A. (2019). Transcriptional activation by MafR, a global regulator of *Enterococcus faecalis*. *Sci. Rep.* 9, 6146. doi:10.1038/s41598-019-42484-4
- 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
- Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C_T method. *Nat. Protoc.* 3, 1101–1108. doi:10.1038/nprot.2008.73
- Shepard, B. D., and Gilmore, M. S. (1995). Electroporation and efficient transformation of *Enterococcus faecalis* grown in high concentrations of glycine. *Methods Mol. Biol.* 47, 217–226. doi:10.1385/0-89603-310-4:217
- 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
- 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., 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
- Solano-Collado, V., Lurz, R., Espinosa, M., and Bravo, A. (2013). The pneumococcal MgaSpn virulence transcriptional regulator generates multimeric complexes on linear double-stranded DNA. *Nucleic Acids Res.* 41, 6975–6991. doi:10.1093/nar/gkt445
- Solano-Collado, V., Ruiz-Cruz, S., Lorenzo-Díaz, F., Pluta, R., Espinosa, M., and Bravo, A. (2021). Recognition of streptococcal promoters by the pneumococcal SigA protein. *Front. Mol. Biosci.* 8, 666504. doi:10.3389/fmolb.2021.666504
- Tsvetanova, B., Wilson, A. C., Bongiorno, C., Chiang, C., Hoch, J. A., and Perego, M. (2007). Opposing effects of histidine phosphorylation regulate the AtxA virulence transcription factor in *Bacillus anthracis*. *Mol. Microbiol.* 63, 644–655. doi:10.1111/j.1365-2958.2006.05543.x
- Vahling, C. M., and McIver, K. S. (2006). Domains required for transcriptional activation show conservation in the Mga family of virulence gene regulators. *J. Bacteriol.* 188, 863–873. doi:10.1128/JB.188.3.863-873.2006
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439–D444. doi:10.1093/nar/gkab1061
- Vlahovicsek, 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
- Voskuil, M. I., and Chambliss, G. H. (1998). The -16 region of *Bacillus subtilis* and other Gram-positive bacterial promoters. *Nucleic Acids Res.* 26, 3584–3590. doi:10.1093/nar/26.15.3584
- Voskuil, M. I., and Chambliss, G. H. (2002). The TRTGn motif stabilizes the transcription initiation open complex. *J. Mol. Biol.* 322, 521–532. doi:10.1016/s0022-2836(02)00802-1
- Weiss, A., and Shaw, L. N. (2015). Small things considered: the small accessory subunits of RNA polymerase in Gram-positive bacteria. *FEMS Microbiol. Rev.* 39, 541–554. doi:10.1093/femsre/fuv005