



N-acetylgalactosamine-Mediated Regulation of the *aga* Operon by AgaR in *Streptococcus pneumoniae*

Muhammad Afzal^{1,2}, Sulman Shafeeq^{3*}, Hifza Ahmed¹ and Oscar P. Kuipers^{1*}

¹ Department of Molecular Genetics, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Groningen, Netherlands, ² Department of Bioinformatics and Biotechnology, Government College University Faisalabad, Faisalabad, Pakistan, ³ Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden

Here, we analyze the transcriptomic response of *Streptococcus pneumoniae* D39 to N-acetylgalactosamine (NAGa). Transcriptome comparison of *S. pneumoniae* D39 grown in NAGaM17 (0.5% NAGa + M17) to that grown in GM17 (0.5% Glucose + M17) revealed the elevated expression of various carbon metabolic genes/operons, including a PTS operon (denoted here as the *aga* operon), which is putatively involved in NAGa transport and utilization, in the presence of NAGa. We further studied the role of a GntR-family transcriptional regulator (denoted here as AgaR) in the regulation of *aga* operon. Our transcriptome and RT-PCR data suggest the role of AgaR as a transcriptional repressor of the *aga* operon. We predicted a 20-bp operator site of AagR (5'-ATAATTAATATAACAACAAA-3') in the promoter region of the *aga* operon (*PbgaC*), which was further verified by mutating the AgaR operator site in the respective promoter. The role of CcpA in the additional regulation of the *aga* operon was elucidated by further transcriptome analyses and confirmed by quantitative RT-PCR.

Keywords: N-acetylgalactosamine, AgaR, BgaC, *aga*, CcpA, pneumococcus

OPEN ACCESS

Edited by:

Anthony Baughn,
University of Minnesota, USA

Reviewed by:

Indranil Biswas,
University of Kansas, USA
Brian J. Akerley,
University of Mississippi Medical
Center School of Dentistry, USA

*Correspondence:

Sulman Shafeeq
salmansha@gmail.com
Oscar P. Kuipers
o.p.kuipers@rug.nl

Received: 21 June 2016

Accepted: 29 August 2016

Published: 12 September 2016

Citation:

Afzal M, Shafeeq S, Ahmed H and
Kuipers OP (2016)
N-acetylgalactosamine-Mediated
Regulation of the *aga* Operon by
AgaR in *Streptococcus pneumoniae*.
Front. Cell. Infect. Microbiol. 6:101.
doi: 10.3389/fcimb.2016.00101

INTRODUCTION

The human pathogen *Streptococcus pneumoniae* causes a number of infections like pneumonia, sepsis, meningitis, otitis media, and conjunctivitis, and results in over a million of deaths each year worldwide (Ispahani et al., 2004; O'Brien et al., 2009). The genomic abundance of sugar transport genes suggests the importance of carbohydrates in the lifestyle of *S. pneumoniae*. The ability to utilize different nutrient sources plays a vital role in the life-style of the pneumococcus in addition to the virulence factors it possesses (Phillips et al., 1990; Titgemeyer and Hillen, 2002; Carvalho et al., 2011). The prediction of the involvement of over 30% of the transporters in the *S. pneumoniae* genome in the transport of carbohydrates has been authenticated by a recent functional genomic approach targeting carbohydrate transport (Tettelin et al., 2001; Bidossi et al., 2012).

Free carbohydrates are scarce in the human airway, which makes modification and import of complex glycans much more critical for pneumococci (Buckwalter and King, 2012). The human pathogen *S. pneumoniae* encounters glycoconjugates in the nasopharynx, which are composed of a number of monosaccharides and can be used as nutrients after having been depolymerized by glycosidases. The presence of at least nine surface-associated glycosidases to modify host glycans in pneumococcus enhances the chances of its survival in hosts (King et al., 2006; Burnaugh et al., 2008; Dalia et al., 2010). Amino sugars including NAGa (N-acetylgalactosamine) are part of various cell structures in many biological environments. NAGa is a prominent component of the cell wall

in bacteria (Freymond et al., 2006), being present in lipopolysaccharides (Bernatchez et al., 2005). Furthermore, NAGa links carbohydrate chains in mucins in humans (Carraway and Hull, 1991). These amino sugars are also generally present in the carbohydrate chains of glycosylated proteins, both in prokaryotes and eukaryotes (Barr and Nordin, 1980; Abu-Qarn et al., 2008). Regulatory mechanisms of a number of carbohydrate and amino acid systems that are vital for the life-style and virulence of *S. pneumoniae*, have been studied (Kloosterman et al., 2006a; Carvalho et al., 2011; Afzal et al., 2015a,b). Successful survival and virulence of *S. pneumoniae* depend on its ability to utilize complex glycans existing at the site of its colonization (Buckwalter and King, 2012; Linke et al., 2013). Its cell envelope is comprised of several layers of peptidoglycan with bound teichoic acids and lipoteichoic acids (LTAs), which are anchored in the cell membrane (Stool and Field, 1989). Pneumococcal LTAs contain phosphocholine and NAGa (Behr et al., 1992). The presence of NAGa in the cell wall of *S. pneumoniae* may indicate its importance as a carbon source for the cell as pneumococcus possesses both secreted and surface-associated glycosidases that may modify glycoconjugates present in the host environment.

The current study demonstrates the impact of NAGa on the transcriptome of *S. pneumoniae* D39 and points to the regulatory mechanism of the *aga* operon. Our transcriptome analysis with D39 Δ *agaR* suggests the role of AgaR as a transcriptional repressor of the *aga* operon. The transcriptome data was further confirmed by RT-PCR analysis. A putative operator site of AgaR in the promoter region of the *aga* operon (*PbgaC*) is predicted and further verified by promoter-mutation studies. Moreover, we demonstrate by transcriptome analysis that the regulation of the *aga* operon is also CcpA-dependent in *S. pneumoniae* D39.

MATERIALS AND METHODS

Bacterial Strains, Growth Conditions, and DNA Isolation and Manipulation

Bacterial strains and plasmids used in this study are listed in Table 1. *S. pneumoniae* D39 was grown as described previously (Kloosterman et al., 2006b; Afzal et al., 2014). For selection on antibiotics, the medium was supplemented with the following concentrations of antibiotics: 2.5 μ g/ml tetracycline for *S. pneumoniae*; and 100 μ g/ml ampicillin for *Escherichia coli*. All bacterial strains used in this study were stored in 10% (v/v) glycerol at -80°C . All DNA manipulations in this study were done as described before (Shafeeq et al., 2013). For PCR amplification, chromosomal DNA of *S. pneumoniae* D39 (Lanie et al., 2007) was used. Primers used in this study are based on the sequence of the *S. pneumoniae* D39 genome and listed in Table 2.

Construction of the *agaR* Mutant

A marker-less *agaR* mutant (Δ *agaR*: MA900) was constructed in *S. pneumoniae* D39 using the pORI280 plasmid, as described before (Kloosterman et al., 2006b). The primer pairs, *agaR*-1/*agaR*-2 and *agaR*-3/*agaR*-4, were used to generate PCR fragments of the left and right flanking regions of *agaR* (each of nearly 600 bp), respectively. The left and right

TABLE 1 | List of strains and plasmids used in this study.

Strain/Plasmid	Description	Source
S. PNEUMONIAE		
D39	Serotype 2 strain. <i>cps 2</i>	Laboratory of <i>P. hermans</i>
MA900	D39 Δ <i>agaR</i> , containing unmarked chromosomal deletion of <i>agaR</i>	This study
Δ <i>ccpA</i>	D39 Δ <i>ccpA</i> ; Spec ^R	Carvalho et al., 2011
MA901	D39 Δ <i>bgaA</i> :: <i>PbgaC-lacZ</i> ; Tet ^R	This study
MA902	D39 Δ <i>bgaA</i> :: <i>PbgaC-M-lacZ</i> ; Tet ^R	This study
MA903	D39 Δ <i>bgaA</i> :: <i>PbgaC-cre-M-lacZ</i> ; Tet ^R	This study
MA904	MA900 Δ <i>bgaA</i> :: <i>PbgaC-cre-M-lacZ</i> ; Tet ^R	This study
E. COLI		
EC1000		Laboratory collection
PLASMIDS		
pPPP2	Amp ^R Tet ^R ; promoter-less <i>lacZ</i> . For replacement of <i>bgaA</i> with promoter <i>lacZ</i> fusion. Derivative of pPPP1	Halfmann et al., 2007
pORI280	Erm ^R ; <i>ori</i> ⁺ <i>repA</i> ⁻ ; deletion derivative of pWV01; constitutive <i>lacZ</i> expression from P32 promoter	Leenhouts et al., 1998
pMA901	pPPP2 <i>PbgaC-lacZ</i>	This study
pMA902	pPPP2 <i>PbgaC-M-lacZ</i>	This study
pMA903	pPPP2 <i>PbgaC-cre-M-lacZ</i>	This study
pMA904	pORI280 carrying <i>agaR</i> deletion	This study

flanking regions contain BamHI and EcoRI restriction sites as have the pORI280. These PCR fragments were inserted into pORI280 using these restriction sites, and the construct (pMA904) was used to transform *S. pneumoniae* D39 with selection for erythromycin resistance. The transformation leads to single cross-over integration of the construct into the chromosome, as pORI280 depends on RepA for replication. Several erythromycin-resistant, LacZ-positive integrants, as separate cultures for 30–50 generations (culturing two to four times until stationary phase) without antibiotic selection were plated on X-Gal medium. Due to this, we could screen for clones that had lost the integration due to a second recombination event. 0.5% of the colonies were both white and erythromycin sensitive, signifying excision of the plasmid from the chromosome. Eighty percent of these white erythromycin-sensitive colonies had the desired mutation, as verified by PCR and DNA sequencing.

Construction of Promoter *lacZ*-Fusions and β -Galactosidase Assays

The chromosomal transcriptional *lacZ*-fusion to the *bgaC* promoter was constructed in the integration plasmid pPPP2 (Halfmann et al., 2007) via double crossover in the *bgaA* locus of *S. pneumoniae* D39 with the primer pairs mentioned in Table 2 (*bgaC*-R and *bgaC*-F), resulting in pMA901. *PbgaC-M-lacZ* (mutation in AgaR operator site) was constructed in pPPP2 (Halfmann et al., 2007), using the primer pairs mentioned in Table 2 (*bgaC*-M-F and *bgaC*-R), resulting in plasmid pMA902. These constructs were further introduced into the *S. pneumoniae* D39 wild-type resulting in strains MA901 and

TABLE 2 | List of primers used in this study.

Name	Nucleotide sequence (5'3') ^a	Restriction site
bgaC-R	CATGGGATCCCTGTGATAGAGCTGACATCG	BamHI
bgaC-F	CATGGAATTGATACCAATCCTCTGGAGG	EcoRI
bgaC-M-F	CATGGAATTCTTCTATTGACAATCAAACAGA TTGGTTTATAATTAAGCGCACAAACAATG	EcoRI
bgaC-M-ccpA-F	CATGGAATTCCAGATTGGTTTATAATTAAT ATAACAACAAATGACCGCGCAAACCTTCG	EcoRI
agaR_KO_1	CATGGGATCCGATACCAATCCTCTGGAGG	BamHI
agaR_KO_2	CTGTGATAGACTGACATCG	–
agaR_KO_3	GTCAGCTCTATCACAGCATCTCATGTGA CCGTGATC	–
agaR_KO_4	CATGGAATTCTAGGAGTAACCTCCATCGG	EcoRI
QUANTITATIVE RT-PCR PRIMERS		
63-q1	GCCTTCCAAGCACAGAC	–
63-q2	GGTGTTCCTCAGTCCG	–
65-q1	CGAGCCTCGTGAAGGTGAG	–
65-q2	GCTGGGTCGGATGAGCG	–
66-q1	GGATGCCGATTGATGGAC	–
66-q2	GGTGGTGTGCGAAGTTTC	–
67-q1	GGGAGATGTGACTACTGG	–
67-q2	GCTGTGCGAAGAACCGCACC	–
68-q1	GGTTGGAACACGAACG	–
68-q2	CCAGCGATAATGGTATGG	–
69-q1	CACAGTAGCTCTCTTCC	–
69-q2	CCATTGGAAGATTGATCCG	–
71-q1	GGCGAGGATGTCTTGGC	–
71-q2	CCTACACTTGCTCCATCG	–
RT PCR PRIMERS		
IR-I-1	GGAAGGTGCCAACCGTATC	–
IR-I-2	CGTGTCTACAACCATATGTC	–
IR-II-1	CGTTCTATCAACGTAGTAG	–
IR-II-2	CCTGCAGATGAAACGATCG	–
IR-III-1	GCTGTAGCAGCACCTTCTAC	–
IR-III-2	CCAGAAGCTTGATACGTTCC	–
IR-IV-1	GCTATCGGTATTATTATCG	–
IR-IV-2	CAGCTTCAAATTTAGCAG	–
IR-V-1	GCGGGCTTCGATGATGACG	–
IR-V-2	GCACCCAATTGAGCAAGTC	–
IR-V-1	CCGTGTAGTACAAGGTGTC	–
IR-V-2	GCCAAGACATCCTCGCCCTC	–

^aRestriction enzyme sites are underlined.

MA902, respectively. Similarly, *PbgaC-M-ccpA-lacZ* (mutation in *cre* site) was constructed in pPP2 (Halfmann et al., 2007), using the primer pairs mentioned in **Table 2**, resulting in plasmid pMA903. These constructs were further introduced into the *S. pneumoniae* D39 wild-type and D39 Δ *agaR* (MA900) resulting in strains MA903 and MA904, respectively. All plasmid constructs were checked by PCR and DNA sequencing.

β -galactosidase assays were performed as described before (Israelsen et al., 1995; Kloosterman et al., 2006b) using cells that were harvested in the mid-exponential phase of growth, grown

in M17 medium with appropriate sugars mentioned in Results Section. M17 medium is composed of pancreatic digest of casein, soy peptone, beef extract, yeast extract, and minerals.

RNA Extraction, Reverse Transcription (RT)-PCR, and Purification for Quantitative RT-PCR

Total RNA was isolated from *S. pneumoniae* D39 wild-type and D39 Δ *agaR* (MA900) grown in GM17 (0.5% Glucose + M17) as described (Shafeeq et al., 2011a). Similarly, total RNA was isolated from *S. pneumoniae* D39 wild-type and D39 Δ *ccpA* grown in NAGaM17 (0.5% NAGa + M17) as described (Shafeeq et al., 2011a). The RNA sample was treated with 2U of RNase free Dnase I (Invitrogen, Paisley, UK) to remove any DNA contamination. First, strand cDNA synthesis was performed on RNA (Shafeeq et al., 2011a). cDNA (2 μ l) was amplified in a 20 μ l reaction volume that contained 3 pmol of each primer and the reactions were performed in triplicate (Shafeeq et al., 2011a). The transcription level of specific genes was normalized to *gyrA* transcription, and amplified in parallel with the *gyrA*-F and *gyrA*-R primers. The results were interpreted using the comparative CT method (Schmittgen and Livak, 2008).

To confirm that the *aga* operon is transcribed as a single transcriptional unit, *S. pneumoniae* D39 wild-type was grown in NAGaM17 (0.5% NAGa + M17) and total RNA was isolated as described (Shafeeq et al., 2011a). RT-PCR (reverse transcription PCR) was performed as described before (Afzal et al., 2014) on all possible intergenic regions of the *aga* operon with primer pairs mentioned in **Table 2**. For a fair comparison of the PCR products, 100 ng of RNA and 20 ng of DNA were used.

Microarray Analysis

For DNA microarray analysis in the presence of NAGa, the transcriptome of *S. pneumoniae* D39 wild-type, grown in biological replicates in GM17 (0.5% Glucose + M17) and NAGa (0.5% NAGa + M17) were compared. For DNA microarray analysis of D39 Δ *agaR* (MA900), the transcriptome of *S. pneumoniae* D39 wild-type and D39 Δ *agaR*, grown in biological replicates in GM17 (0.5% Glucose + M17) was compared. Similarly, for DNA microarray analysis of D39 Δ *ccpA*, the transcriptome of *S. pneumoniae* D39 wild-type and D39 Δ *ccpA*, grown in biological replicates in NAGaM17 (0.5% NAGa + M17) was compared. The cells were harvested at their respective mid-exponential growth phases. All other procedures regarding the DNA microarray experiment and data analysis were performed as previously described (Afzal et al., 2015a; Shafeeq et al., 2015). Briefly, microarray slide images were scanned using GenPix Pro 6.1 (MSD analytical technologies). Processing and normalization (LOWESS spotpin-based) of slides were performed with the in-house developed MicroPrep software. DNA microarray data were obtained from independent biological replicates hybridized to glass slides, of which one was a dye swap. Expression ratios were calculated from the measurements of at least five spots. Differential expression tests were performed on expression ratios with a local copy of the Cyber-T implementation of a variant of the *t*-test. For the identification of differentially expressed genes

a Bayesian *p*-value of <0.001 and a fold change cut-off 2 was applied. Microarray data have been submitted to GEO under accession number GSE86008.

RESULTS

NAGa-Dependent Gene Expression in *S. pneumoniae*

To study the response of *S. pneumoniae* D39 to NAGa, we performed transcriptome comparison of *S. pneumoniae* D39 wild-type grown in NAGaM17 (0.5% NAGa + M17) to that in glucose (0.5% Glucose + M17). **Table 3** summarizes the transcriptome changes incurred on *S. pneumoniae* in the presence of NAGa. The presence of NAGa in the medium resulted in the upregulation of a number of carbon metabolic genes/operons after applying the criteria of ≥ 2 -fold difference and a *p*-value of <0.001.

A system for sialic acid transport and utilization is upregulated in the presence of NAGa. This system consists of a neuraminidase (*nanA*) and *nan* operon-I (Afzal et al., 2015c). *nanA* has been demonstrated to be involved in virulence of *S. pneumoniae* (Dalia et al., 2010). Another important gene coding for an N-acetylglucosamine-6-phosphate deacetylase (NagA) is also upregulated in the presence of NAGa. N-acetylglucosamine (NAG) is phosphorylated by the PTS and this phosphorylated N-acetylglucosamine is deacetylated to glucosamine-6-P by NagA (Kanehisa et al., 2014). Another putative operon *spd_1969-72* is highly upregulated in our microarray analysis in the presence of NAGa. This putative operon is supposedly involved in the utilization of amino sugars and in the conversion of chitobiose into NAG (Kanehisa et al., 2014). Further experiments are needed to explore the role of this operon in the utilization of amino sugars.

An ABC transporter (encoded by *spd-0088-90*) putatively involved in galactose transport (Bidossi et al., 2012) is among the ones upregulated in our NAGa transcriptome analysis. Furthermore, Leloir pathway genes (*galKT*) (Afzal et al., 2015d) are highly upregulated in the presence of NAGa. Galactose can also be utilized through the Tagatose pathway and we could observe significant upregulation of the Tagatose pathway genes (*lacABCD*) in the presence of NAGa. The Tagatose pathway genes are present in an operon (*lac* operon-I) and are involved in the utilization of lactose and galactose (Afzal et al., 2014). The expression of a number of other genes putatively involved in the transport and utilization of carbohydrates are also altered in the presence of NAGa. The altered expression of these genes might be due to the absence of CCR (carbon catabolite repression) in the presence of NAGa. We have analyzed the promoter regions of these genes and we could find putative *cre* boxes in the promoter regions of these genes.

glnAR genes that are part of the glutamine regulon were downregulated in the presence of NAGa. This regulon consists of genes involved in glutamine synthesis and uptake (*glnA* and *glnPQ*), glutamate synthesis (*gdhA*), and the gene coding for the pentose phosphate pathway enzyme *Zwf*, which forms an operon with *glnPQ* (Kloosterman et al., 2006a). The glutamine regulon is shown to be repressed in the presence of a nitrogen

TABLE 3 | Summary of transcriptome comparison of *S. pneumoniae* D39 wild-type grown in NAGaM17 (0.5% NAGa + M17) to that grown in GM17 (0.5% Glucose + M17).

D39 tag ^a	Function ^b	Ratio ^c
<i>spd_0030</i>	Hypothetical protein	-7.6
<i>spd_0031</i>	Hypothetical protein	-4.1
<i>spd_0032</i>	Hypothetical protein	-3.1
<i>spd_0065</i>	Beta-galactosidase, BgaC	2.0
<i>spd_0066</i>	PTS system, IIB component, GadV	3.6
<i>spd_0067</i>	PTS system, IIC component, GadW	3.5
<i>spd_0068</i>	PTS system, IID component, GadE	3.7
<i>spd_0069</i>	PTS system, IIA component, GadF	3.3
<i>spd_0070</i>	Sugar isomerase domain protein, AgaS	6.4
<i>spd_0071</i>	Aldose 1-epimerase, GalM	2.0
<i>spd_0088</i>	ABC transporter, permease protein	5.9
<i>spd_0089</i>	ABC transporter, permease protein	6.4
<i>spd_0090</i>	ABC transporter, substrate-binding protein	76.8
<i>spd_0091</i>	Hypothetical protein	-5.4
<i>spd_0092</i>	Hypothetical protein	-6.5
<i>spd_0447</i>	Transcriptional regulator, MerR family protein, GlnR	-4.1
<i>spd_0448</i>	Glutamine synthetase, type I, GlnA	-2.0
<i>spd_0490</i>	Hypothetical protein	-4.3
<i>spd_0553</i>	Hypothetical protein	-7.1
<i>spd_0559</i>	PTS system IIA component, putative	5.0
<i>spd_0560</i>	PTS system, IIB component, putative	5.7
<i>spd_0561</i>	PTS system, IIC component, putative	5.4
<i>spd_0608</i>	Orotidine 5'-phosphate decarboxylase, PyrF	-48.0
<i>spd_0609</i>	Orotate phosphoribosyltransferase, PyrE	-39.5
<i>spd_0675</i>	Hypothetical protein	-5.5
<i>spd_0731</i>	DNA topology modulation protein FlaR, putative	-31.3
<i>spd_0850</i>	Lactoylglutathione lyase	-19.0
<i>spd_0851</i>	Dihydroorotate dehydrogenase electron transfer subunit, PyrK	-66.2
<i>spd_0852</i>	Dihydroorotate dehydrogenase, catalytic subunit, PyrDb	-32.5
<i>spd_1050</i>	Tagatose 1,6-diphosphate aldolase, LacD	29.6
<i>spd_1051</i>	Tagatose-6-phosphate kinase, LacC	15.3
<i>spd_1052</i>	Galactose-6-phosphate isomerase, LacB subunit	20.2
<i>spd_1053</i>	Galactose-6-phosphate isomerase, LacA subunit	19.4
<i>spd_1133</i>	Aspartate carbamoyltransferase, PyrB	-18.2
<i>spd_1134</i>	Pyrimidine operon regulatory protein/uracil phosphoribosyltransferase, PyrR	-26.0
<i>spd_1427</i>	PhnA protein	-11.4
<i>spd_1489</i>	N-acetylneuraminase lyase, NanA2	3.7
<i>spd_1490</i>	Hypothetical protein	3.5
<i>spd_1491</i>	Hypothetical protein	7.4
<i>spd_1492</i>	Hypothetical protein	3.9
<i>spd_1493</i>	Sugar ABC transporter, permease protein, NanW	4.8
<i>spd_1494</i>	Sugar ABC transporter, permease protein, NanV	2.5
<i>spd_1495</i>	Sugar ABC transporter, sugar-binding protein, NanU	6.4
<i>spd_1496</i>	PTS system, IIBC components, NanP	3.2
<i>spd_1497</i>	N-acetylmannosamine-6-phosphate 2-epimerase 2, NanE	3.1
<i>spd_1504</i>	Sialidase A, NanA	16.6
<i>spd_1633</i>	Galactose-1-phosphate uridylyltransferase, GalT2	21.6
<i>spd_1634</i>	Galactokinase, GalK	27.6

(Continued)

TABLE 3 | Continued

D39 tag ^a	Function ^b	Ratio ^c
<i>spd_1635</i>	Galactose operon repressor, GalR	5.5
<i>spd_1726</i>	Pneumolysin, Ply	-4.0
<i>spd_1866</i>	N-acetylglucosamine-6-phosphate deacetylase, NagA	3.7
<i>spd_1898</i>	Hypothetical protein	-3.0
<i>spd_1899</i>	Glutamine amidotransferase, class 1, YvdE	-3.4
<i>spd_1969</i>	Glycosyl hydrolase-related protein	12.3
<i>spd_1970</i>	ROK family protein	31.2
<i>spd_1971</i>	Glycosyl hydrolase-related protein	5.1
<i>spd_1972</i>	Hypothetical protein	6.8
<i>spd_1973</i>	Alpha-1,2-mannosidase, putative	2.7
<i>spd_1974</i>	Hypothetical protein	4.0

^aGene numbers refer to D39 locus tags.

^bD39 annotation (Lanie et al., 2007).

^cRatio represents the fold increase/decrease in the expression of genes in NAGa compared to glucose. All *p*-values were <0.001.

source (Kloosterman et al., 2006a). The down-regulation of the glutamine regulon might be due to the presence of nitrogen in NAGa.

The expression of genes that are putatively involved in NAGa transport and utilization (the putative *aga* gene cluster) was also altered in our transcriptome analysis. The upregulation of the *aga* gene cluster might indicate that the system for putative transport and utilization of NAGa is functional and responds to NAGa in *S. pneumoniae*. Therefore, we decided to further characterize the *aga* gene cluster in *S. pneumoniae* D39.

Organization of the *aga* Gene Cluster in *S. pneumoniae*

The putative *aga* gene cluster consists of seven genes encoding a beta-galactosidase (*bgaC*), a predicted galactosamine disaccharide-specific PTS system (IIBCDA components: *gadVWVEF*), a gene coding for a sugar isomerase (*agaS*) and a gene coding for an aldose 1-epimerase (*galM*) (Figure 1A). To confirm whether the putative *aga* gene cluster transcribes as a single transcriptional unit, we performed RT-PCR on all possible intergenic regions present in the *aga* gene cluster with the primer pairs mentioned in Table 2. RT-PCR data revealed that the putative *aga* gene cluster transcribes as a single transcriptional unit (Figure 1B) and from here on, we call it the *aga* operon.

NAGa Induces the Expression of the *aga* Operon

To further confirm our microarray results and study the role of NAGa in the regulation of the *aga* operon, we grew the *S. pneumoniae* D39 wild-type in GM17 (0.5% Glucose + M17) and NAGaM17 (0.5% NAGa + M17) and performed quantitative RT-PCR on the genes of the *aga* operon. The results of quantitative RT-PCR showed that the expression of the *aga* operon increased significantly when grown in the presence of NAGa (Figure 2A). These results further confirm our microarray data mentioned above and demonstrate that the *aga* operon is functional and responds to NAGa in *S. pneumoniae* D39.

Microarray Analysis of D39 Δ *agaR*

AgaR, a GntR-family transcriptional regulator, is encoded by the gene present upstream of the *aga* operon (Figure 1). The presence of the *agaR* gene next to the *aga* operon might indicate its role in the regulation of the *aga* operon. To study whether AgaR is involved in the regulation of the *aga* operon, we constructed a marker-less mutant of *agaR* and performed microarray analysis of D39 Δ *agaR* against D39 wild-type grown in GM17 (0.5% Glucose + M17). GM17 was used as a growth medium as we got repression of the *aga* operon in the presence of glucose. Table 4 summarizes the results of the transcriptomic changes induced in *S. pneumoniae* due to the deletion of *agaR*. *agaR* was downregulated 26-fold in our transcriptome analysis, confirming *agaR* deletion in Δ *agaR*. After choosing the criterion of ≥ 2 -fold difference as the threshold change and a *p*-value <0.001, the *aga* operon was upregulated significantly in Δ *agaR* and no other larger responses were observed in the transcriptome. This data further suggests that AgaR is a negative transcriptional regulator of the *aga* operon in the absence of NAGa.

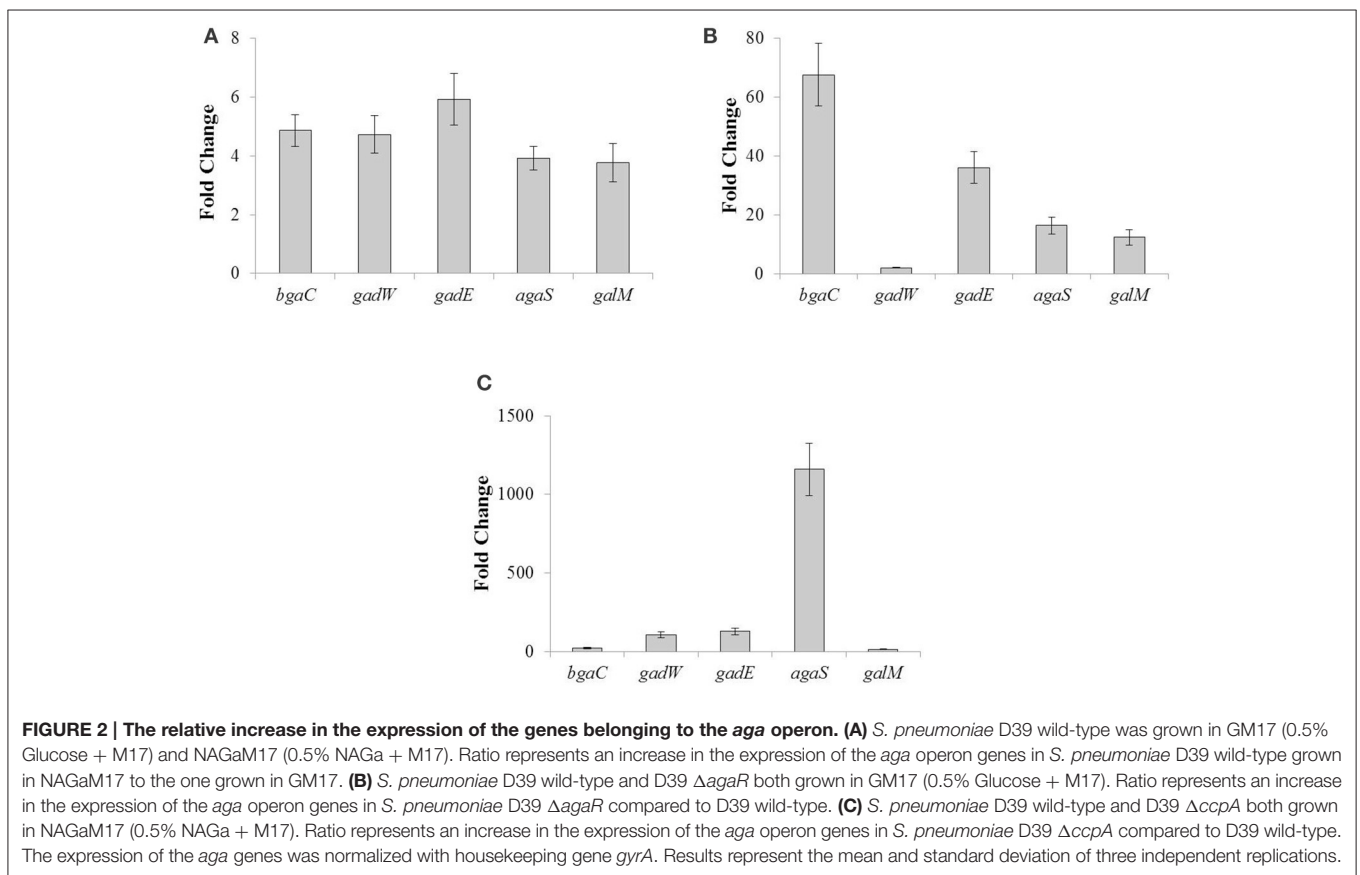
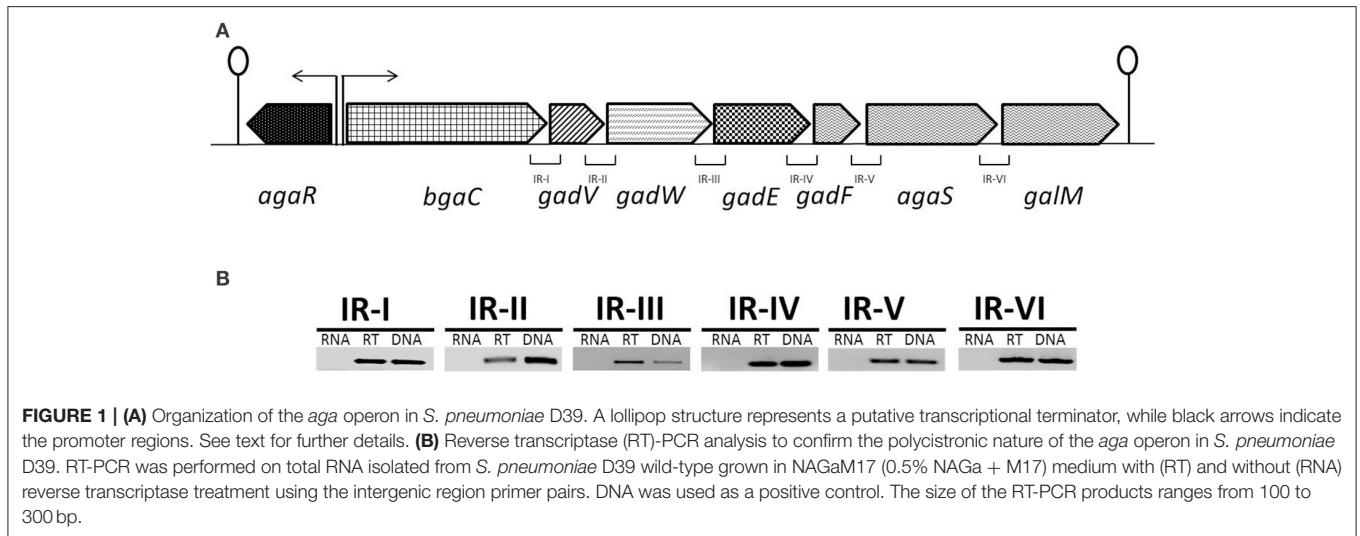
AgaR Acts As a Transcriptional Repressor of the *aga* Operon

To confirm our microarray results of D39 Δ *agaR*, we grew *S. pneumoniae* D39 wild-type and D39 Δ *agaR* in GM17 (0.5% Glucose + M17) and performed quantitative RT-PCR on the genes belonging to the *aga* operon. The results of quantitative RT-PCR showed that the expression of the *aga* operon increased significantly in D39 Δ *agaR* even in the presence of glucose (Figure 2B). These quantitative RT-PCR results further confirm that AgaR represses the expression of the *aga* operon and that this repression is relieved in the absence of *agaR*.

Prediction and Confirmation of the AgaR Operator Site in the Promoter Region of the *aga* Operon (*bgaC*)

AgaR, a putative GntR-family transcriptional regulator, is present next to the *aga* operon in *S. pneumoniae* D39. Using Genome2D tool (Baerends et al., 2004) and a MEME motif sampler search (Bailey and Elkan, 1994), a 20-bp palindromic sequence was found upstream of *bgaC* (5'-ATAATTAATATAACAACAAA-3') in *S. pneumoniae* D39 wild-type (SP) (Figure 3A). This DNA stretch may serve as an AgaR operator site in *S. pneumoniae*. NAGa genes promoters of other streptococcal species were studied to check if the AgaR operator site is also conserved in those streptococci. The outcome of this analysis was the finding that the AgaR operator sequence is highly conserved in these streptococci as well (Figure 3B). A genome-wide search with the pneumococcal AgaR operator site was conducted to locate more putative AgaR targets in the *S. pneumoniae* D39 genome. We could not find any other DNA stretch similar to the AgaR operator site, which suggests that the *aga* operon is the only target of AgaR.

To determine whether the located stretch of DNA mediates the AgaR-dependent transcriptional control of the *aga* operon, we made a *lacZ*-fusion, where conserved bases in the putative



AgaR site were mutated in *PbgaC* (5'-ATAATTAATATAACAACAAA-3' to 5'-ATAATTAAGCGCACAACAAA-3') and performed β -galactosidase assays (Figure 4). The expression of the promoter increased significantly when we mutated a few of the conserved bases of the putative AgaR operator site. These data suggest that the putative AgaR operator site is active and performs the role of AgaR operator site in *S. pneumoniae*. This

operator site might also be active in other streptococcal species as it is highly conserved in those species as well.

The Role of CcpA in the Regulation of the *aga* Operon

The *aga* operon was only around 3-fold upregulated in our transcriptome in the presence of NAGa. The possible reason

could be the involvement of another transcriptional regulator that is repressing the *aga* operon in the presence of NAGa. The other transcriptional regulator could be CcpA. CcpA (Carbon catabolite protein A) is the principal transcriptional regulator in

S. pneumoniae that represses the expression of genes involved in the consumption of non-preferred sugars in the presence of a favored one and has a role in virulence as well (Carvalho et al., 2011). CcpA represses the transcription of genes involved in the utilization non-preferred sugars in the presence of a preferred one by binding to specific DNA sequences called *cre* (Catabolite Repression Element) boxes. To study the role of CcpA in the regulation of the *aga* operon, we analyzed the *PbgaC* for the presence of a putative *cre* box. We could find a *cre* box in the promoter region of *bgaC* (5'-ATGAAAGCGCAAACCTT-3'), which might suggest a role of CcpA in the regulation of the *aga* operon.

To determine the role of CcpA in the regulation the *aga* operon, we performed microarray analysis of D39 Δ *ccpA* against the D39 wild-type grown in NAGaM17 (0.5% NAGa + M17). A number of genes were significantly affected in D39 Δ *ccpA* compared to the D39 wild-type in the presence of NAGa. We could observe a significant upregulation in the expression of the *aga* operon in our transcriptome analysis (Table 4), which suggests that CcpA represses the expression of the *aga* operon in the presence of NAGa. These results are in accordance with the data presented in a previous study, where they performed microarray analysis of D39 Δ *ccpA* in the presence of glucose and galactose (Carvalho et al., 2011). There were also a number of other genes/operons that were differentially expressed in D39 Δ *ccpA* in the presence of NAGa. These genes have been grouped

TABLE 4 | Summary of transcriptome comparison of *S. pneumoniae* D39 wild-type with D39 Δ *agaR* grown in GM17 (0.5% Glucose + M17) and *S. pneumoniae* strain D39 wild-type with Δ *ccpA* grown in NAGaM17 (0.5% NAGa + M17).

D39 tag ^a	Function ^b	Ratio ^c	Ratio ^d
<i>spd_0064</i>	GntR-family transcriptional regulator, AgaR	-26.0	-
<i>spd_0065</i>	Beta-galactosidase, BgaC	10.9	20.5
<i>spd_0066</i>	PTS system, IIB component, GadV	10.8	13.9
<i>spd_0067</i>	PTS system, IIC component, GadW	5.9	-
<i>spd_0068</i>	PTS system, IID component, GadE	8.0	25.2
<i>spd_0069</i>	PTS system, IIA component, GadF	8.5	13.5
<i>spd_0070</i>	Sugar isomerase domain protein, AgaS	20.5	5.6
<i>spd_0071</i>	Aldose 1-epimerase, GalM	3.4	21.8

^aGene numbers refer to D39 locus tags.

^bD39 annotation (Lanie et al., 2007).

^cRatio represents the fold increase/decrease in the expression of genes in D39 Δ *agaR* compared to the D39 wild-type.

^dRatio represents the fold increase/decrease in the expression of genes in D39 Δ *ccpA* compared to the D39 wild-type. The ratios are obtained through microarrays analysis.

A

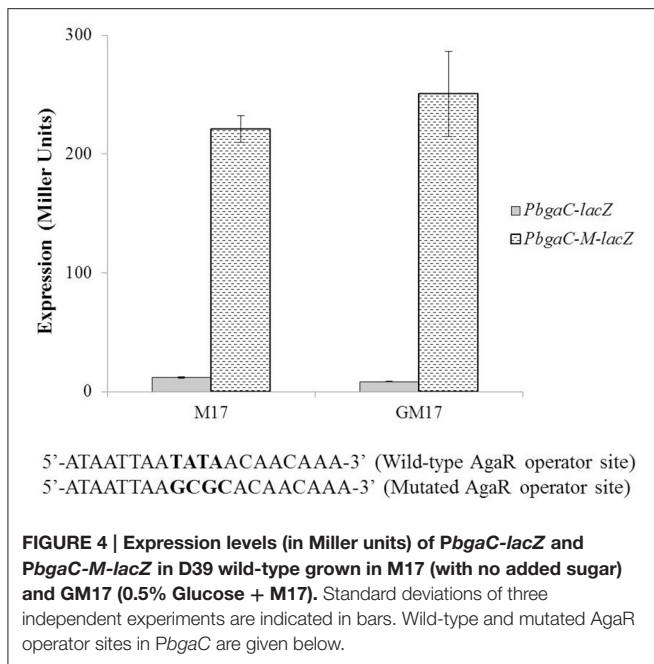
TTACAAGTTATTTATTTAGCTATTATAACAAAAAGCCAGAAGATA-N45-AAAGGAGGAAAAGTG --- *PbgaC* SD
 TCAATAATTGTTTAATACGATTATAGCAAAAATAGCAAGCTGAT-N73-AAAGGAGATAATATG --- *PbgaC* SE
 GACTTTTTTTCAAAATGTTATTATATTAATTAATTGAAAGAGTTT-N24-GAAAGATTTGAAGTG --- *PbgaC* SG
 AACAGATTGGTTTATAATTAATATAACAACAAATGAAAGCGCAA-N13-GAAGGTAGTTTTATG --- *PbgaC* SP
 ATAGAATATTATTAGAACTATTATAACAGAAATAAAAAAGATAT-N59-AGGGAGCTTTGTATG --- *PbgaC* SB
 ATAACCACAAAGAAGTGTTAGTATATTAATAGAAAGGTAAGCGGTTGCATCGGAGGAAGAAAATG --- *PbgaC* SU
 GACATTTATGCAAAAATTTATTATATTAATATTAACAAGAGATG-N18-AGGGAACATCAGATG --- *PbgaC* SS
 AAAATATTGAGATATATTTAATATATCAACAAATGAAAGCGTAA-N14-AAAGGTAGTAAAATG --- *PbgaC* SM

AWWRYTATTATAWYAAWAAW

B



FIGURE 3 | Identification of the AgaR operator site in *PbgaC*. (A) Position of the AgaR operator site in *PbgaC* of different streptococci. Translational start sites are italicized and putative AgaR operator sites are bold-underlined. **(B)** Weight matrix of the identified AgaR operator site in the *PbgaC* of different streptococci. SD, *S. dysgalactiae*; SE, *S. equi*; SG, *S. gordonii*; SP, *S. pneumoniae*; SB, *S. uberis*; SU, *S. suis*; SS, *S. sanguis*; and SM, *S. mitis*.



into COG functional categories according to the putative function of respective proteins (Table 5). Genes belonging to the category G are mostly carbohydrate transport and metabolism genes and the repression on genes caused by CcpA is relieved in the absence of CcpA.

To further confirm our microarray results, we performed quantitative RT-PCR on the genes of the *aga* operon in D39 $\Delta ccpA$ in NAGaM17 (0.5% NAGa + M17). The results of quantitative RT-PCR show that the expression of the *aga* operon increased significantly in D39 $\Delta ccpA$ when grown in the presence of NAGa (Figure 2C). These results further confirm that the *aga* operon is repressed by CcpA in *S. pneumoniae* D39.

To check the functionality of the *cre* site present in promoter region of the *aga* operon, we made a promoter *lacZ*-fusion of *bgaC* (*PbgaC-cre-M-lacZ*) with mutations in the conserved bases of the putative *cre* box (5'-ATGAAAGCGCAAACCTT-3' to 5'-ATGACCGCGCAAACCTT-3') and transformed it into D39 wild-type and D39 $\Delta agaR$, and performed β -galactosidase assays (Figure 5). The expression of the mutated promoter was significantly higher in the *agaR* mutant, which also confirms the functionality of the predicted *cre* box in the promoter region of *bgaC* and the role of CcpA in regulation of *aga* operon.

DISCUSSION AND CONCLUSIONS

Extensive studies regarding regulatory mechanisms of different dedicated systems for carbon sources in *S. pneumoniae* emphasize the importance of carbohydrates in the life style of pneumococcus, which confers an extra advantage in its survival in ever changing nutritional environment (Tettelin et al., 2001; Iyer and Camilli, 2007; Afzal et al., 2014, 2015c,d,e). In bacteria, NAGa is a vital constituent of the cell wall as it is present in lipopolysaccharides (Bernatchez et al., 2005; Freymond

TABLE 5 | Number of genes significantly affected in D39 $\Delta ccpA$ compared to the D39 wild-type grown in NAGaM17 (0.5% NAGa + M17).

Functional categories	Total	Up	Down
C: Energy production and conversion	21	5	16
E: Amino acid transport and metabolism	25	3	22
F: Nucleotide transport and metabolism	13	3	10
G: Carbohydrate transport and metabolism	45	34	11
H: Coenzyme transport and metabolism	6	1	5
I: Lipid transport and metabolism	4	0	4
J: Translation, ribosomal structure and biogenesis	41	3	38
K: Transcription	10	6	4
L: Replication, recombination and repair	7	0	7
M: Cell wall/membrane/envelope biogenesis	12	5	7
O: Posttranslational modification, protein turnover, chaperones	7	2	5
P: Inorganic ion transport and metabolism	9	1	8
Q: Secondary metabolites biosynthesis, transport and catabolism	2	1	1
R: General function prediction only	25	6	19
S: Function unknown	22	14	8
T: Signal transduction mechanisms	12	5	7
U: Intracellular trafficking, secretion, and vesicular transport	2	0	2
V: Defense mechanisms	6	2	4
Others	45	20	25
Total number of genes	314	111	203

Genes affected more than 2-fold in D39 $\Delta ccpA$ as compared to the D39 wild-type are shown in COG functional categories.

et al., 2006). NAGa links carbohydrate chains in mucins in humans (Carraway and Hull, 1991). An important aspect of the interaction between the pneumococcus and its human host is the ability of this bacterium to process host glycans. The current study demonstrates the NAGa-dependent gene expression and regulatory mechanism of the *aga* operon in *S. pneumoniae*.

NAGa can support growth of different bacteria, including *E. coli*, acting as a carbon and nitrogen source (Reizer et al., 1996). A NAGa utilization system (AgaBCD and AgaVWEF) was identified in *E. coli* (Reizer et al., 1996). *agaBCD* and *agaVWEF* encode two PTSs that mediate the transport and phosphorylation of galactosamine and NAGa, respectively, in *E. coli* (Brinkkötter et al., 2000). The proposed catabolic pathway of NAGa in *E. coli* involves the transport and subsequent phosphorylation of NAGa, the deacetylation of NAGa-6-P, the deamination/isomerization of GalN-6-P, the phosphorylation of Tag-6-P, and the cleavage of Tag-1,6-PP to produce glyceraldehyde 3-phosphate and glycerone phosphate (Brinkkötter et al., 2000). Moreover, a wide-ranging diversity in galactosamine/NAGa utilization pathways in Proteobacteria such as *Shewanella* has been proposed (Leyn et al., 2012). Particularly, there is a lot of variability among the first steps of the pathway, whereas the concluding three steps are mostly conserved (Leyn et al., 2012). In *E. coli*, AgaR belonging to the DeoR family of transcriptional factors negatively regulates the expression of *agaZ* and *agaS*

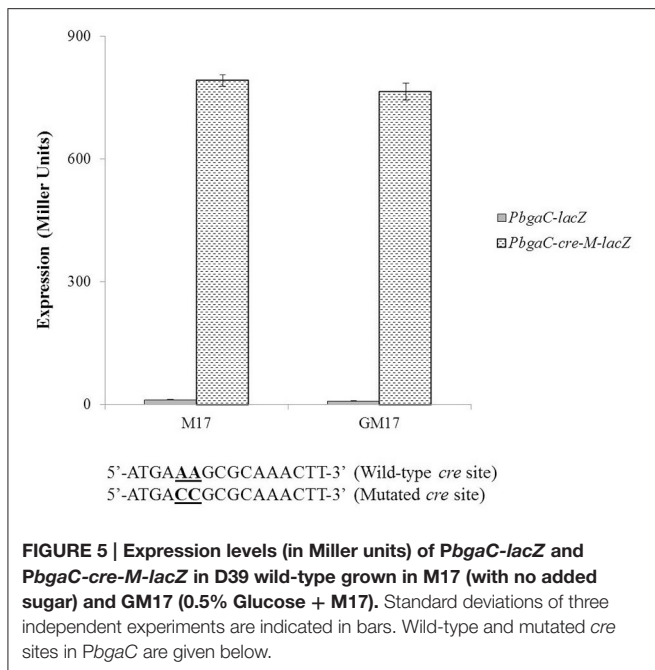


FIGURE 5 | Expression levels (in Miller units) of *Pbgac-lacZ* and *Pbgac-cre-M-lacZ* in D39 wild-type grown in M17 (with no added sugar) and GM17 (0.5% Glucose + M17). Standard deviations of three independent experiments are indicated in bars. Wild-type and mutated *cre* sites in *Pbgac* are given below.

involved in the galactosamine/NAGa catabolism pathways (Ray and Larson, 2004; Leyn et al., 2012). The genes encoding galactosamine/NAGa pathways were recognized in 16 genomes signifying four families; *Streptococcaceae*, *Lactobacillaceae*, *Enterobacteraceae*, and *Carnobacteriaceae*. *Lactobacillus helveticus* and *Streptococcus pyogenes* possess the minimal number of genes in the reconstructed regulons (Zhang et al., 2015). It is very likely that these organisms cannot utilize NAGa as only one or two genes from the NAGa utilization pathway were found in these organisms. *Streptococcus gordonii* and *Streptococcus mitis* have the minimal gene set that allows them to utilize NAGa. These genes include transcriptional regulator (*agaR*), galactosamine-6-phosphate deaminase/isomerase (*agaS*), glycoside hydrolase (*bgaC*) and PTS (*gadVWEF*) (Zhang et al., 2015). The gene for the tagatose-1,6-diphosphate aldolase (*agaY*) is present in some other well-studied *Streptococcal* genomes. On the contrary, all the investigated *Lactobacilli* lack *agaY*, but have the gene for N-acetylgalactosamine-phosphate deacetylase (*agaA*) (Zhang et al., 2015). *agaA* was recognized only in the *Streptococcus suis* genome among *Streptococcaceae*. The gene for tagatose-6-phosphate kinase (*agaZ*) was found only in *Enterococcus faecalis* and *Carnobacterium* (Zhang et al., 2015). The *aga* operon is also annotated as part of the amino sugar metabolism pathways in *S. pneumoniae* (Kanehisa et al., 2014). The *aga* operon consists of a gene encoding a beta-galactosidase (*bgaC*), a predicted galactosamine disaccharide-specific PTS system (*agaVWEF*), a gene coding for a sugar isomerase (*agaS*) and a gene coding for an aldose 1-epimerase (*galM*). *agaVWEF* is the annotated PTS for the transport of NAGa, whereas the deacetylation of NAGa-6-P and deamination/isomerization of GalN-6-P may be performed by NagA and NagB, respectively. Most of the genes involved in amino sugar metabolism are regulated in our transcriptome analysis in the presence of NAGa,

suggesting that the growth conditions used in our studies are adequate and that the NAGa pathway is functional.

One of the important enzymes regulated in our study is BgaC, which is a novel surface-exposed glycohydrolase, which has been shown to have an effect on *S. pneumoniae* adhesion and virulence (Jeong et al., 2009; Terra et al., 2010). A classic β -galactosidase (EC 3.2.1.23) BgaC displays explicit hydrolysis activity toward the terminal Gal β (1,3) NAG moiety of oligosaccharides (Jeong et al., 2009; Terra et al., 2010). Sequence comparison puts BgaC into the glycosidase family 35 (GH-35), which mainly occurs in higher eukaryotes (Henrissat and Davies, 1997). BgaC receives a sequence and structural organization arrangement similar to β -galactosidases from higher eukaryotes and microbial pathogens instead of typical prokaryotic β -galactosidases (Henrissat and Davies, 1997). Several crystal structures of β -galactosidases have been submitted in the Protein Data Bank (PDB) database. These include *E. coli* β -gal (Jacobson et al., 1994), *Arthrobacter* sp. C2-2 C221 β -gal (Skálová et al., 2005), *Kluyveromyces lactis* β -gal (Pereira-Rodríguez et al., 2012), *Thermus thermophilus* A4 β -gal (Hidaka et al., 2002), *Bacillus circulans* sp. *alkalophilus* β -gal (Maksimainen et al., 2012), *Sulfolobus solfataricus* β -gal (Aguilar et al., 1997), *Penicillium* sp. β -gal (Rojas et al., 2004), *Bacteroides thetaiotaomicron* β -gal, 5 *Trichoderma reesei* β -gal (Maksimainen et al., 2011), and *Homo sapiens* β -gal (Ohto et al., 2012). Most of these enzymes possess specific hydrolysis activity toward β (1,4)-galactosyl bond, whereas *H. sapiens* β -gal exhibits hydrolysis activity toward both β (1,3)- and β (1,4)-galactosyl bonds (Alpers, 1969; Asp and Dahlqvist, 1972; Distler and Jourdan, 1973), and the substrate specificity of *B. thetaiotaomicron* β -gal remains unknown. *E. coli* β -gal, as a member of GH-2, is one of the most extensively studied β -galactosidases. The presence of BgaC in *S. pneumoniae* suggests that pneumococcus has the ability to make use of the galactose moieties present in its environment.

In *E. coli*, the *aga* genes are regulated by the transcriptional regulator AgaR, which is a DeoR family transcriptional repressor (Ray and Larson, 2004). AgaR binds in tandem to several repeat sequences in the intergenic regions of *agaZ*, *agaR*, and *agaS* to repress the transcription (Leyn et al., 2012). In *S. pneumoniae*, the *aga* operon is also regulated by transcriptional factor AgaR. The AgaR present in *S. pneumoniae* belongs to the GntR family of transcriptional repressor. It has a winged helix-turn-helix (WHTH) DNA-binding domain and an UbiC transcription regulator-associated (UTRA) domain. Several bacterial and archaeal genomes possess representatives of GntR family of transcriptional regulators. Numerous biological processes, including antibiotic production, sensing of nutritional status, growth, proliferation, development, diverse metabolic processes (fatty acid metabolism, amino acid metabolism, acetoin utilization, etc.) are controlled by the GntR family regulators (Wiethaus et al., 2008; Resch et al., 2010). In this study, we have identified an AgaR operator site in the *Pbgac*, which was further verified by promoter mutational analyses. To explore more putative AgaR operator sites in the D39 genome, we conducted a genome-wide search with the putative pneumococcal AgaR operator site. AgaR operator site was only found in the promoter region of *bgaC* suggesting the *aga* operon as the only target

of AgaR in *S. pneumoniae* D39. This predicted AgaR operator site is also found highly conserved in other streptococci as well (Novichkov et al., 2010), suggesting a similar function of AgaR in other streptococci as well.

The master regulator, CcpA, is involved in the regulation of genes involved in sugar metabolism and has a role in the pathogenesis (Lulko et al., 2007; Zomer et al., 2007; Carvalho et al., 2011). A number of other non-preferred sugar systems are also regulated independently of CcpA like CelR in *S. pneumoniae* (Shafeeq et al., 2011b). The expression of the *aga* operon did not go very high in our transcriptome in the presence of NAGa (Table 3). The moderate upregulation of the *aga* operon in the presence of NAGa suggested the involvement of another transcriptional regulator in the regulation of the *aga* operon. The presence of a *cre* box in the *PbgaC* indicates that CcpA could have a role in the regulation of the *aga* operon, which is

further confirmed by our microarray analysis of *S. pneumoniae* D39 Δ *ccpA* against the D39 wild-type. These results are also consistent with the study performed by Carvalho et al. (2011).

AUTHOR CONTRIBUTIONS

MA designed the project, performed experiments, and wrote manuscript. SS designed the project and wrote the manuscript. HA performed experiments. OK designed the project and wrote manuscript.

ACKNOWLEDGMENTS

MA is supported by Government College University, Faisalabad, Pakistan under the faculty development program of HEC Pakistan.

REFERENCES

- Abu-Qarn, M., Eichler, J., and Sharon, N. (2008). Not just for Eukarya anymore: protein glycosylation in *Bacteria* and *Archaea*. *Curr. Opin. Struct. Biol.* 18, 544–550. doi: 10.1016/j.sbi.2008.06.010
- Afzal, M., Shafeeq, S., Ahmed, H., and Kuipers, O. P. (2015c). Sialic acid-mediated gene expression in *Streptococcus pneumoniae* and role of NanR as a transcriptional activator of the nan gene cluster. *Appl. Environ. Microbiol.* 81, 3121–3131. doi: 10.1128/AEM.00499-15
- Afzal, M., Shafeeq, S., Henriques-Normark, B., and Kuipers, O. P. (2015a). UlaR activates expression of the *ula* operon in *Streptococcus pneumoniae* in the presence of ascorbic acid. *Microbiology* 161, 41–49. doi: 10.1099/mic.0.083899-0
- Afzal, M., Shafeeq, S., and Kuipers, O. P. (2014). LacR is a repressor of lacABCD and LacT is an activator of lacTFEG, constituting the lac gene cluster in *Streptococcus pneumoniae*. *Appl. Environ. Microbiol.* 80, 5349–5358. doi: 10.1128/AEM.01370-14
- Afzal, M., Shafeeq, S., and Kuipers, O. P. (2015b). Ascorbic acid-dependent gene expression in *Streptococcus pneumoniae* and the activator function of the transcriptional regulator UlaR2. *Front. Microbiol.* 6:72. doi: 10.3389/fmicb.2015.00072
- Afzal, M., Shafeeq, S., Manzoor, I., and Kuipers, O. P. (2015d). GalR acts as a transcriptional activator of galKT in the presence of galactose in *Streptococcus pneumoniae*. *J. Mol. Microbiol. Biotechnol.* 25, 363–371. doi: 10.1159/000439429
- Afzal, M., Shafeeq, S., Manzoor, I., and Kuipers, O. P. (2015e). Maltose-dependent transcriptional regulation of the mal regulon by MalR in *Streptococcus pneumoniae*. *PLoS ONE* 10:e0127579. doi: 10.1371/journal.pone.0127579
- Aguilar, C. F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M., et al. (1997). Crystal structure of the beta-glycosidase from the hyperthermophilic archeon *Sulfolobus solfataricus*: resilience as a key factor in thermostability. *J. Mol. Biol.* 271, 789–802. doi: 10.1006/jmbi.1997.1215
- Alpers, D. H. (1969). Separation and isolation of rat and human intestinal beta-galactosidases. *J. Biol. Chem.* 244, 1238–1246.
- Asp, N. G., and Dahlqvist, A. (1972). Human small intestine-galactosidases: specific assay of three different enzymes. *Anal. Biochem.* 47, 527–538. doi: 10.1016/0003-2697(72)90147-9
- Baerends, R. J. S., Smits, W. K., de Jong, A., Hamoen, L. W., Kok, J., and Kuipers, O. P. (2004). Genome2D: a visualization tool for the rapid analysis of bacterial transcriptome data. *Genome Biol.* 5:R37. doi: 10.1186/gb-2004-5-5-r37
- 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.
- Barr, J., and Nordin, P. (1980). Biosynthesis of glycoproteins by membranes of *Acer pseudoplatanus*. Incorporation of mannose and N-acetylglucosamine. *Biochem. J.* 192, 569–577. doi: 10.1042/bj1920569
- Behr, T., Fischer, W., Peter-Katalinic, J., and Egge, H. (1992). The structure of pneumococcal lipoteichoic acid. Improved preparation, chemical and mass spectrometric studies. *Eur. J. Biochem.* 207, 1063–1075. doi: 10.1111/j.1432-1033.1992.tb17143.x
- Bernatchez, S., Szymanski, C. M., Ishiyama, N., Li, J., Jarrell, H. C., Lau, P. C., et al. (2005). A single bifunctional UDP-GlcNAc/Glc 4-epimerase supports the synthesis of three cell surface glycoconjugates in *Campylobacter jejuni*. *J. Biol. Chem.* 280, 4792–4802. doi: 10.1074/jbc.M407767200
- Bidossi, A., Mulas, L., Decorosi, F., Colomba, L., Ricci, S., Pozzi, G., et al. (2012). A functional genomics approach to establish the complement of carbohydrate transporters in *Streptococcus pneumoniae*. *PLoS ONE* 7:e33320. doi: 10.1371/journal.pone.0033320
- Brinkkötter, A., Klöss, H., Alpert, C., and Lengeler, J. W. (2000). Pathways for the utilization of N-acetyl-galactosamine and galactosamine in *Escherichia coli*. *Mol. Microbiol.* 37, 125–135. doi: 10.1046/j.1365-2958.2000.01969.x
- Buckwalter, C. M., and King, S. J. (2012). Pneumococcal carbohydrate transport: food for thought. *Trends Microbiol.* 20, 517–522. doi: 10.1016/j.tim.2012.08.008
- Burnaugh, A. M., Frantz, L. J., and King, S. J. (2008). Growth of *Streptococcus pneumoniae* on human glycoconjugates is dependent upon the sequential activity of bacterial exoglycosidases. *J. Bacteriol.* 190, 221–230. doi: 10.1128/JB.01251-07
- Carraway, K. L., and Hull, S. R. (1991). Cell surface mucin-type glycoproteins and mucin-like domains. *Glycobiology* 1, 131–138. doi: 10.1093/glycob/1.2.131
- Carvalho, S. M., Kloosterman, T. G., Kuipers, O. P., and Neves, A. R. (2011). CcpA ensures optimal metabolic fitness of *Streptococcus pneumoniae*. *PLoS ONE* 6:e26707. doi: 10.1371/journal.pone.0026707
- Dalia, A. B., Standish, A. J., and Weiser, J. N. (2010). Three surface exoglycosidases from *Streptococcus pneumoniae*, NanA, BgaA, and StrH, promote resistance to opsonophagocytic killing by human neutrophils. *Infect. Immun.* 78, 2108–2116. doi: 10.1128/IAI.01125-09
- Distler, J. J., and Jourdain, G. W. (1973). The purification and properties of beta-galactosidase from bovine testes. *J. Biol. Chem.* 248, 6772–6780.
- Freymond, P.-P., Lazarevic, V., Soldo, B., and Karamata, D. (2006). Poly(glucosyl-N-acetylgalactosamine 1-phosphate), a wall teichoic acid of *Bacillus subtilis* 168: its biosynthetic pathway and mode of attachment to peptidoglycan. *Microbiology* 152, 1709–1718. doi: 10.1099/mic.0.28814-0
- Halfmann, A., Hakenbeck, R., and Bruckner, R. (2007). A new integrative reporter plasmid for *Streptococcus pneumoniae*. *FEMS Microbiol. Lett.* 268, 217–224. doi: 10.1111/j.1574-6968.2006.00584.x
- Henrissat, B., and Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7, 637–644. doi: 10.1016/S0959-440X(97)80072-3
- Hidaka, M., Fushinobu, S., Ohtsu, N., Motoshima, H., Matsuzawa, H., Shoun, H., et al. (2002). Trimeric crystal structure of the glycoside hydrolase family 42 beta-galactosidase from *Thermus thermophilus* A4 and the structure of

- its complex with galactose. *J. Mol. Biol.* 322, 79–91. doi: 10.1016/S0022-2836(02)00746-5
- Ispahani, P., Slack, R. C. B., Donald, F. E., Weston, V. C., and Rutter, N. (2004). Twenty year surveillance of invasive pneumococcal disease in Nottingham: serogroups responsible and implications for immunisation. *Arch. Dis. Child.* 89, 757–762. doi: 10.1136/adc.2003.036921
- Israelsen, H., Madsen, S. M., Vrang, A., Hansen, E. B., and Johansen, E. (1995). Cloning and partial characterization of regulated promoters from *Lactococcus lactis* Tn917-lacZ integrants with the new promoter probe vector, pAK80. *Appl. Environ. Microbiol.* 61, 2540–2547.
- Iyer, R., and Camilli, A. (2007). Sucrose metabolism contributes to *in vivo* fitness of *Streptococcus pneumoniae*. *Mol. Microbiol.* 66, 1–13. doi: 10.1111/j.1365-2958.2007.05878.x
- Jacobson, R. H., Zhang, X. J., DuBose, R. F., and Matthews, B. W. (1994). Three-dimensional structure of beta-galactosidase from *E. coli*. *Nature* 369, 761–766. doi: 10.1038/369761a0
- Jeong, J. K., Kwon, O., Lee, Y. M., Oh, D. B., Lee, J. M., Kim, S., et al. (2009). Characterization of the *Streptococcus pneumoniae* BgaC protein as a novel surface beta-galactosidase with specific hydrolysis activity for the Galbeta1-3GlcNAc moiety of oligosaccharides. *J. Bacteriol.* 191, 3011–3023. doi: 10.1128/JB.01601-08
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199–D205. doi: 10.1093/nar/gkt1076
- King, S. J., Hippe, K. R., and Weiser, J. N. (2006). Deglycosylation of human glycoconjugates by the sequential activities of exoglycosidases expressed by *Streptococcus pneumoniae*. *Mol. Microbiol.* 59, 961–974. doi: 10.1111/j.1365-2958.2005.04984.x
- Kloosterman, T. G., Bijlsma, J. J. E., Kok, J., and Kuipers, O. P. (2006b). To have neighbour's fare: extending the molecular toolbox for *Streptococcus pneumoniae*. *Microbiology* 152, 351–359. doi: 10.1099/mic.0.28521-0
- Kloosterman, T. G., Hendriksen, W. T., Bijlsma, J. J., Bootsma, H. J., van Hijum, S. A., Kok, J., et al. (2006a). Regulation of glutamine and glutamate metabolism by GlnR and GlnA in *Streptococcus pneumoniae*. *J. Biol. Chem.* 281, 25097–25109. doi: 10.1074/jbc.M601661200
- Lanie, J. A., Ng, W. L., Kazmierczak, K. M., Andrzejewski, T. M., Davidsen, T. M., Wayne, K. J., et al. (2007). Genome sequence of Avery's virulent serotype 2 strain D39 of *Streptococcus pneumoniae* and comparison with that of unencapsulated laboratory strain R6. *J. Bacteriol.* 189, 38–51. doi: 10.1128/JB.01148-06
- Leenhouts, K., Venema, G., and Kok, J. (1998). A lactococcal pWV01 based integration toolbox for bacteria. *Methods Cell Sci.* 20, 35–50. doi: 10.1023/A:1009862119114
- Leyn, S. A., Gao, F., Yang, C., and Rodionov, D. A. (2012). N-acetylgalactosamine utilization pathway and regulon in proteobacteria: genomic reconstruction and experimental characterization in *Shewanella*. *J. Biol. Chem.* 287, 28047–28056. doi: 10.1074/jbc.M112.382333
- Linke, C. M., Woodiga, S. A., Meyers, D. J., Buckwalter, C. M., Salhi, H. E., and King, S. J. (2013). The ABC transporter encoded at the pneumococcal fructooligosaccharide utilization locus determines the ability to utilize long- and short-chain fructooligosaccharides. *J. Bacteriol.* 195, 1031–1041. doi: 10.1128/JB.01560-12
- Lulko, A. T., Buist, G., Kok, J., and Kuipers, O. P. (2007). Transcriptome analysis of temporal regulation of carbon metabolism by CcpA in *Bacillus subtilis* reveals additional target genes. *J. Mol. Microbiol. Biotechnol.* 12, 82–95. doi: 10.1159/000096463
- Maksimainen, M., Hakulinen, N., Kallio, J. M., Timoharju, T., Turunen, O., and Rouvinen, J. (2011). Crystal structures of *Trichoderma reesei* beta-galactosidase reveal conformational changes in the active site. *J. Struct. Biol.* 174, 156–163. doi: 10.1016/j.jsb.2010.11.024
- Maksimainen, M., Paavilainen, S., Hakulinen, N., and Rouvinen, J. (2012). Structural analysis, enzymatic characterization, and catalytic mechanisms of beta-galactosidase from *Bacillus circulans* sp. alkalophilus. *FEBS J.* 279, 1788–1798. doi: 10.1111/j.1742-4658.2012.08555.x
- Novichkov, P. S., Laikova, O. N., Novichkova, E. S., Gelfand, M. S., Arkin, A. P., Dubchak, I., et al. (2010). RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes. *Nucleic Acids Res.* 38, D111–D118. doi: 10.1093/nar/gkp894
- O'Brien, K. L., Wolfson, L. J., Watt, J. P., Henkle, E., Deloria-Knoll, M., McCall, N., et al. (2009). Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet* 374, 893–902. doi: 10.1016/S0140-6736(09)61204-6
- Ohto, U., Usui, K., Ochi, T., Yuki, K., Satow, Y., and Shimizu, T. (2012). Crystal structure of human beta-galactosidase: structural basis of Gm1 gangliosidosis and morquio B diseases. *J. Biol. Chem.* 287, 1801–1812. doi: 10.1074/jbc.M111.293795
- Pereira-Rodríguez, A., Fernández-Leiro, R., González-Siso, M. I., Cerdán, M. E., Becerra, M., and Sanz-Aparicio, J. (2012). Structural basis of specificity in tetrameric *Kluyveromyces lactis* beta-galactosidase. *J. Struct. Biol.* 177, 392–401. doi: 10.1016/j.jsb.2011.11.031
- Phillips, N. J., John, C. M., Reinders, L. G., Gibson, B. W., Apicella, M. A., and Griffiths, J. M. (1990). Structural models for the cell surface lipooligosaccharides of *Neisseria gonorrhoeae* and *Haemophilus influenzae*. *Biomed. Environ. Mass Spectrom.* 19, 731–745. doi: 10.1002/bms.1200191112
- Ray, W. K., and Larson, T. J. (2004). Application of AgaR repressor and dominant repressor variants for verification of a gene cluster involved in N-acetylgalactosamine metabolism in *Escherichia coli* K-12. *Mol. Microbiol.* 51, 813–826. doi: 10.1046/j.1365-2958.2003.03868.x
- Reizer, J., Ramseier, T. M., Reizer, A., Charbit, A., and Saier, M. H. (1996). Novel phosphotransferase genes revealed by bacterial genome sequencing: a gene cluster encoding a putative N-acetylgalactosamine metabolic pathway in *Escherichia coli*. *Microbiology* 142(Pt 2), 231–250. doi: 10.1099/13500872-142-2-231
- Resch, M., Schiltz, E., Titgemeyer, F., and Muller, Y. A. (2010). Insight into the induction mechanism of the GntR/HutC bacterial transcription regulator YvoA. *Nucleic Acids Res.* 38, 2485–2497. doi: 10.1093/nar/gkp1191
- Rojas, A. L., Nagem, R. A., Neustroev, K. N., Arand, M., Adamska, M., Eneyskaya, E. V., et al. (2004). Crystal structures of beta-galactosidase from *Penicillium* sp. and its complex with galactose. *J. Mol. Biol.* 343, 1281–1292. doi: 10.1016/j.jmb.2004.09.012
- 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
- Shafeeq, S., Afzal, M., Henriques-Normark, B., and Kuipers, O. P. (2015). Transcriptional profiling of UlaR-regulated genes in *Streptococcus pneumoniae*. *Genom. Data* 4, 57–59. doi: 10.1016/j.gdata.2015.02.004
- Shafeeq, S., Kloosterman, T. G., and Kuipers, O. P. (2011b). CelR-mediated activation of the cellobiose-utilization gene cluster in *Streptococcus pneumoniae*. *Microbiology* 157, 2854–2861. doi: 10.1099/mic.0.051359-0
- Shafeeq, S., Kuipers, O. P., and Kloosterman, T. G. (2013). Cellobiose-mediated gene expression in *Streptococcus pneumoniae*: a repressor function of the novel GntR-type regulator BguR. *PLoS ONE* 8:e57586. doi: 10.1371/journal.pone.0057586
- Shafeeq, S., Yesilkaya, H., Kloosterman, T. G., Narayanan, G., Wandel, M., Andrew, P. W., et al. (2011a). The cop operon is required for copper homeostasis and contributes to virulence in *Streptococcus pneumoniae*. *Mol. Microbiol.* 81, 1255–1270. doi: 10.1111/j.1365-2958.2011.07758.x
- Skálová, T., Dohnálek, J., Spiwok, V., Lipová, P., Vondráčková, E., Petroková, H., et al. (2005). Cold-active beta-galactosidase from *Arthrobacter* sp. C2-2 forms compact 660 kDa hexamers: crystal structure at 1.9 Å resolution. *J. Mol. Biol.* 353, 282–294. doi: 10.1016/j.jmb.2005.08.028
- Stool, S. E., and Field, M. J. (1989). The impact of otitis media. *Pediatr. Infect. Dis. J.* 8, S11–S14. doi: 10.1097/00006454-198901001-00006
- Terra, V. S., Homer, K. A., Rao, S. G., Andrew, P. W., and Yesilkaya, H. (2010). Characterization of novel beta-galactosidase activity that contributes to glycoprotein degradation and virulence in *Streptococcus pneumoniae*. *Infect. Immun.* 78, 348–357. doi: 10.1128/IAI.00721-09
- Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., Peterson, S., et al. (2001). Complete genome sequence of a virulent isolate

- of *Streptococcus pneumoniae*. *Science* 293, 498–506. doi: 10.1126/science.1061217
- Titgemeyer, F., and Hillen, W. (2002). Global control of sugar metabolism: a gram-positive solution. *Antonie Van Leeuwenhoek* 82, 59–71. doi: 10.1023/A:1020628909429
- Wiethaus, J., Schubert, B., Pfänder, Y., Narberhaus, F., and Masepohl, B. (2008). The GntR-like regulator TauR activates expression of taurine utilization genes in *Rhodobacter capsulatus*. *J. Bacteriol.* 190, 487–493. doi: 10.1128/JB.01510-07
- Zhang, H., Ravcheev, D. A., Hu, D., Zhang, F., Gong, X., Hao, L., et al. (2015). Two novel regulators of *N*-acetyl-galactosamine utilization pathway and distinct roles in bacterial infections. *Microbiology Open* 4, 983–1000. doi: 10.1002/mbo3.307
- Zomer, A. L., Buist, G., Larsen, R., Kok, J., and Kuipers, O. P. (2007). Time-resolved determination of the CcpA regulon of *Lactococcus lactis* subsp. cremoris MG1363. *J. Bacteriol.* 189, 1366–1381. doi: 10.1128/JB.01013-06
- 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 Afzal, Shafeeq, Ahmed and Kuipers. 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.