Characterization of the ROK-family transcriptional regulator RokA of Streptococcus pneumoniae D39

Sulman Shafeeq, Tomas G. Kloosterman, Vijayanand Rajendran and Oscar P. Kuipers

Department of Molecular Genetics, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 7, 9747 AG, Groningen, The Netherlands

The Gram-positive human pathogen Streptococcus pneumoniae possesses an unusually high number of gene clusters specific for carbohydrate utilization. This provides it with the ability to use a wide array of sugars, which may aid during infection and survival in different environmental conditions present in the host. In this study, the regulatory mechanism of transcription of a gene cluster, SPD0424-8, putatively encoding a cellobiose/lactose-specific phosphotransferase system is investigated. We demonstrate that this gene cluster is transcribed as one transcriptional unit directed by the promoter of the SPD0424 gene. Upstream of SPD0424, a gene was identified encoding a ROK-family transcriptional regulator (RokA: SPD0423). DNA microarray and transcriptional reporter analyses with a rokA mutant revealed that RokA acts as a transcriptional repressor of the SPD0424-8 operon. Furthermore, we identified a 25 bp AT-rich DNA operator site (5'-TATATTTAATTATAAAAAATAAA-3') in the promoter region of SPD0424, which was validated by promoter truncation studies, DNase I footprinting and electrophoretic mobility-shift assays. We tested a large range of different sugars for their effect on the expression of the SPD0424-8 operon, but only moderate variation in expression was observed in the conditions applied. Therefore, a co-factor for RokA-mediated transcriptional control could not be identified.

INTRODUCTION

Streptococcus pneumoniae is an opportunistic Gram-positive human pathogen that resides as a commensal in the nasopharynx (Bogaert et al., 2004). During favourable conditions it has the ability to spread to different parts of the human body, where it can cause serious infections such as otitis media, pneumonia, meningitis and bacteraemia (Kadioglu et al., 2008). To be able to survive in the host, it must adapt to the different environmental conditions at diverse anatomical sites and tune itself to a fluctuating availability of nutrients in the host, an important one being the carbon/energy source (Bogaert et al., 2004; Kadioglu et al., 2008). This requires regulatory mechanisms to ensure that carbon-utilization systems necessary for growth under certain conditions are expressed properly.

Glucose is the preferred carbon source for S. pneumoniae, as it leads to the shortest doubling time (Bidossi et al., 2012; Carvalho et al., 2011; Iyer et al., 2005). However, S. pneumoniae also has the ability to grow on various other carbon sources when provided as a sole energy source in the medium (Bidossi et al., 2012; Buckwalter & King, 2012; Carvalho et al., 2011; Iyer et al., 2005; Marion et al., 2011; McKessar & Hakenbeck, 2007; Shafeeq et al., 2011b; Tyx et al., 2011). Compared with levels of glucose in human blood (3.57–6.06 mM), glucose levels at other common sites of pneumococcal infection are quite low [saliva 0.02–0.4 mM (Gough et al., 1996), nasal secretions <1.0 mM (Wood et al., 2004), lower airway secretions <0.5 mM (de Prost & Saumon, 2007)], meaning that alternative carbon sources are engaged at these sites to maintain the fitness of S. pneumoniae during infection (Shelburne et al., 2008). Indeed, >30% of the transporters in S. pneumoniae are predicted to be involved in carbohydrate transport. These transport systems include phosphoenolpyruvate-dependent phosphotransferase systems (PTS), ATP-binding cassettes and ion gradient-driven transporters and enable S. pneumoniae to utilize different types of carbohydrate in the host (Bidossi et al., 2012; Buckwalter & King, 2012).

It is important for the cell to strictly control the expression of carbohydrate-utilization systems, in order to channel the
cell’s energy towards the metabolism of only the carbon source that is available at a certain moment. In *S. pneumoniae* and other bacteria, the ability to use the preferred sugar(s) before the non-preferred sugar(s) depends on a regulatory process called carbon-catabolite repression (CCR) (Carvalho et al., 2011; Gorke & Stülke, 2008; Iyer et al., 2005). CCR brings about silencing/repression of genes that are specific for utilization of non-preferred sugars, until the cell has completely consumed the preferred carbon source (Carvalho et al., 2011; Iyer et al., 2005). In addition to CCR, the regulation of some sugar uptake systems is governed by dedicated regulatory factors, which guarantee expression at the right moment (Iyer & Camilli, 2007; McKessar & Hakenbeck, 2007; Nieto et al., 1997; Shafeeq et al., 2011b; Tyx et al., 2011).

One class of transcriptional regulators involved in carbohydrate-dependent transcriptional control is the ROK (repressor, ORF and kinase) protein family, which also contains sugar kinases and many functionally uncharacterized proteins (Titgemeyer et al., 1994). The transcriptional regulation mediated by ROK-family repressors and their role in carbohydrate utilization has been well elucidated in many bacteria (Decker et al., 1998; Dubeau et al., 2011; Kimata et al., 1998; Kreuzer et al., 1989; Plumbridge, 1995; Plumbridge & Pellegrini, 2004; Plumbridge, 1991). In *Escherichia coli*, the ROK-family transcription factor Mic is a repressor protein of genes and operons (*ptsG*, *ptsHI* and *manXYZ*) involved in glucose utilization (Kimata et al., 1998) and maltose metabolism (Decker et al., 1998). Another example is NagC, a repressor of the N-acetylgalactosamine (GlcNAc)-utilization operon (*nagE-nagRACD*) (Plumbridge, 1991), the chb operon involved in the utilization of chitobiose (Plumbridge & Pellegrini, 2004) and an activator of the glmUS operon involved in synthesis of UDP-GlcNAc, an essential precursor for cell-wall components (Plumbridge, 1995). In *Bacillus subtilis*, a ROK-family repressor, XylR, was characterized as a repressor of xylose-utilization genes (Kreuzer et al., 1989).

Recently, a ROK-family transcriptional repressor has been demonstrated to be involved in the repression of a putative cellobiose/lactose PTS operon and a sulfatase operon. Moreover, it was found that these genes contribute to the pathogenesis of *S. pneumoniae* strain WCH206 (McAllister et al., 2012). This particular ROK-family protein is encoded on a genomic island, which is, however, absent in many pneumococcal strains, including D39. Despite the absence of this genomic island, the genome of *S. pneumoniae* D39 encodes four ROK-family proteins, one of which is a putative ROK-family transcriptional repressor protein. In this study, we report the functional characterization of RokA in strain D39 and show that it is involved directly in the regulation of a putative cellobiose/lactose PTS operon.

### METHODS

**General procedures.** Strains and plasmids used in this study were stored in 10% (v/v) glycerol at −80 °C and are listed in Table 1. *S. pneumoniae* (Lanier et al., 2007) was grown in M17 broth (Terzaghi & Sandine, 1975) supplemented with 0.5% (w/v) glucose, on blood agar plates supplemented with 1% (v/v) defibrinated sheep blood in micro-aerobic conditions at 37 °C. Luria–Bertani broth was used to grow *E. coli* in a shaking incubator at 37 °C. *Lactococcus lactis* strain NZ9000 was used for overexpression of the RokA protein using the nisin-inducible system and was grown at 30 °C in GM17 (0.5% glucose + M17) broth as described before (Kuijpers et al., 1998). When appropriate, media were supplemented with the following concentrations of antibiotics: 150 μg spectinomycin ml⁻¹, 2.5 μg tetracycline ml⁻¹ for *S. pneumoniae*, 100 μg ampicillin ml⁻¹ for *E. coli* and 4 μg chloramphenicol ml⁻¹ for *L. lactis*. Chromosomal DNA of *S. pneumoniae* wild-type strain D39 was used as a template for PCR amplification (Avery et al., 1989), and all DNA manipulations were done as described previously (Kloosterman et al., 2006). Primers used in this study are based on the sequence of the D39 genome (Lanier et al., 2007) and are listed in Table S1, available with the online version of this paper.

**Construction of the rokA mutant.** The rokA (SPD0423) mutant was made by means of allelic replacement with a spectinomycin–resistance cassette. In short, primers SPD0423-1/SPD0423-2 and SPD0423-3/SPD0423-4 were used to generate PCR products of the left and right flanking regions of rokA, which were, by means of ligation using *AscI/*NolI restriction sites, fused to a spectinomycin–resistance gene, PCR-amplified with primers Spec-F and Spec-R from plasmid pORI38. The resulting ligation mixture was transformed to *S. pneumoniae* strain D39, yielding strain SS500, and the mutation was verified by PCR.

**Construction of transcriptional lacZ fusions.** Chromosomal transcriptional lacZ fusions to SPD0424 and the intergenic region (IR) between SPD0424-5 and SPD0424-6 (SPD0426) were constructed in the integration plasmid pPP2 (Halfmann et al., 2007) with the primer pairs P5PD0424-F/P5PD0424-R1, P5PD0424-6/FSPD0424-R1, P5PD0424-F2/P5PD0424-R2 and P5PD0424-6/FSPD0424-R2, respectively (Table S1), resulting in plasmids pSS501–02. These plasmids were further introduced into wild-type D39 and the rokA mutant (SS500), resulting in strains SS501–04. All plasmid constructs were checked by DNA sequencing.

**Subcloning of the SPD0424 promoter.** A scheme of SPD0424 subclones is shown in Fig. 2(a). Transcriptional lacZ fusions of subclones P1, P2, P3 and P4 were constructed in plasmid pPP2 (Halfmann et al., 2007) with the primer pairs SPD0424-F1/SPD0424-R1, SPD0424-6/FSPD0424-R1, SPD0424-F2/SPD0424-R2 and SPD0424-6/FSPD0424-R2, respectively, resulting in plasmids pSS503–06. These plasmids were further transformed into *S. pneumoniae* wild-type D39 and the rokA mutant (SS500), giving strains SS505–12. All plasmid constructs were checked by DNA sequencing.

**β-Galactosidase assays.** Cells were grown in triplicate in M17 broth (Terzaghi & Sandine, 1975) at 37 °C in the presence of 0.5% (w/v) of one of the following sugars: glucose, fructose, mannose, arabinose, mannotol, cellobiose, maltose, sucrose, raffinose, lactose, trehalose, N-acetylgalactosamine (GlcNAc) or N-acetylgalactosaminate (GalNAc). Cells were harvested in the mid-exponential phase of growth and β-galactosidase activity was measured as described by Israelsen et al. (1995), except that cells were permeabilized with a final concentration of 0.06 mg CTAB (cetyltrimethyl ammonium bromide) ml⁻¹.

**Overexpression and purification of RokA.** To overexpress C-terminally Streptagged RokA (RokA-Strep), primers SPD0423-nco/SPD0423-Cstreph-hin were used for PCR amplification of the rokA gene. The rokA PCR product was restricted with NcoI/HindIII and cloned into the NcoI/HindIII sites of pNZ8048 under the nisin-inducible promoter, resulting in plasmid pSS507. Overexpression of C-terminally Streptagged RokA was achieved in strain *L. lactis* NZ9000 as described before (Kuijpers et al., 1998; Shafeeq et al., 2011).
Characterization of pneumococcal RokA

Table 1. Strains and plasmids used in this study.

<table>
<thead>
<tr>
<th>Strain/plasmid</th>
<th>Description</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>S. pneumoniae</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D39</td>
<td>Serotype 2 strain, cpS 2</td>
<td>Laboratory of P. Hermans</td>
</tr>
<tr>
<td>SS500</td>
<td>D39 ΔrokA; SpecR</td>
<td>This study</td>
</tr>
<tr>
<td>SS501</td>
<td>D39 ΔbgaA::Pspd0424-lacZ; TetR</td>
<td>This study</td>
</tr>
<tr>
<td>SS502</td>
<td>D39 ΔbgaA::Pspd0426-lacZ; TetR</td>
<td>This study</td>
</tr>
<tr>
<td>SS503</td>
<td>SS500 ΔbgaA::Pspd0424-lacZ, TetR</td>
<td>This study</td>
</tr>
<tr>
<td>SS504</td>
<td>SS500 ΔbgaA::Pspd0426-lacZ, TetR</td>
<td>This study</td>
</tr>
<tr>
<td>SS505</td>
<td>D39 ΔbgaA::Pspd0424-P1-lacZ; TetR</td>
<td>This study</td>
</tr>
<tr>
<td>SS506</td>
<td>D39 ΔbgaA::Pspd0424-P2-lacZ; TetR</td>
<td>This study</td>
</tr>
<tr>
<td>SS507</td>
<td>D39 ΔbgaA::Pspd0424-P3-lacZ; TetR</td>
<td>This study</td>
</tr>
<tr>
<td>SS508</td>
<td>D39 ΔbgaA::Pspd0424-P4-lacZ, TetR</td>
<td>This study</td>
</tr>
<tr>
<td>SS509</td>
<td>SS500 ΔbgaA::Pspd0424-P1-lacZ, TetR</td>
<td>This study</td>
</tr>
<tr>
<td>SS510</td>
<td>SS500 ΔbgaA::Pspd0424-P2-lacZ, TetR</td>
<td>This study</td>
</tr>
<tr>
<td>SS511</td>
<td>SS500 ΔbgaA::Pspd0424-P3-lacZ, TetR</td>
<td>This study</td>
</tr>
<tr>
<td>SS512</td>
<td>SS500 ΔbgaA::Pspd0424-P4-lacZ, TetR</td>
<td>This study</td>
</tr>
<tr>
<td><strong>L. lactis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NZ29000</td>
<td>MG1363 ΔpcpN::nisRK</td>
<td>Kuipers et al. (1998)</td>
</tr>
<tr>
<td>SS513</td>
<td>N29000::pSS507</td>
<td>This study</td>
</tr>
<tr>
<td><strong>E. coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EC1000</td>
<td>KmR; MC1000 derivative carrying a single copy of</td>
<td>Laboratory collection</td>
</tr>
<tr>
<td></td>
<td>the pWV1 repA gene in glgB</td>
<td></td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pPP2</td>
<td>AmpR TetR; promoterless lacZ.</td>
<td>Halfmann et al. (2007)</td>
</tr>
<tr>
<td></td>
<td>Derivative of pTP1</td>
<td></td>
</tr>
<tr>
<td>pNZ8048</td>
<td>CmR; nisin-inducible PnisA</td>
<td>de Ruyter et al. (1996)</td>
</tr>
<tr>
<td>pSS501</td>
<td>pPP2 Pspd0424-lacZ</td>
<td>This study</td>
</tr>
<tr>
<td>pSS502</td>
<td>pPP2 Pspd0426-lacZ</td>
<td>This study</td>
</tr>
<tr>
<td>pSS503</td>
<td>pPP2 Pspd0424-P1-lacZ</td>
<td>This study</td>
</tr>
<tr>
<td>pSS504</td>
<td>pPP2 Pspd0424-P2-lacZ</td>
<td>This study</td>
</tr>
<tr>
<td>pSS505</td>
<td>pPP2 Pspd0424-P3-lacZ</td>
<td>This study</td>
</tr>
<tr>
<td>pSS506</td>
<td>pPP2 Pspd0424-P4-lacZ</td>
<td>This study</td>
</tr>
<tr>
<td>pSS507</td>
<td>pNZ8048 carrying Strept-tagged RokA downstream of</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>PnisA</td>
<td></td>
</tr>
</tbody>
</table>

2011a). Nisin (Sigma) was used to induce nisin-dependent overexpression at a concentration of 10 ng ml⁻¹ in L. lactis. RokA-Strep was purified from L. lactis by using a Streptactin column from IBA according to the supplier instructions (http://www.iba-go.com). Purified protein was stored with 10% glycerol at −80 °C.

RT-PCR analysis. To confirm that SPD0424-5 and SPD0426-8 are transcribed together on one transcript, total RNA was isolated from the rokA mutant (which expresses these genes strongly) grown in GM17 medium. Primers PSPD0426-F and PSPD0426-R were used to amplify the IR between SPD0424-5 and SPD0426-8. PCs were performed with 1/100 of the reverse transcription reactions, and 200 ng RNA and 50 ng DNA as controls.

Electrophoretic mobility-shift assays (EMSAs) and DNase I footprinting analysis. EMSAs were performed as described before (Shafeeq et al., 2011a). PCR products of PSPD0424, PSPD0426, PSPD1830 and truncated fragments of PSPD0424 (P1, P2 and P3) were amplified with primer pairs PSPD0424-F/PSPD0424-R, PSPD0426-F/PSPD0426-R, PSPD0424-F1/PSPD0424-R1 (P1), PSPD0424-F2/PSPD0424-R1 (P2) and PSPD0424-F1/PSPD0424-R2 (P3), respectively, and were labelled with [γ-³²P]ATP. A PCR fragment of PSPD1830 that was amplified with primer pair PSPD1830-F/PSPD1830-R was used as a negative control. [γ-³²P]ATP-labelled PCR products [3000 c.p.m. per reaction] were used to perform EMSAs with increasing concentrations of Strept-tagged RokA.

Experimental procedures for DNase I footprinting were performed essentially as described before (den Hengst et al., 2003b; Kloosterman et al., 2007). For each reaction, 150 000 c.p.m. [γ-³²P]ATP-labelled PCR product of PSPD0424 that was PCR-amplified by either [γ-³²P]ATP-labelled primer SPD0424-F (forward strand) or SPD0424-R (reverse strand) in combination with the unlabelled primer, was used.

DNA microarray experiments and data analysis. For microarray analysis of S. pneumoniae wild-type and its isogenic rokA mutant (SS500), cells were grown as three biological replicates in GM17 medium and harvested at an OD₅₉₅ of approximately 0.25. All other procedures regarding microarrays and data analysis were done as described previously (Shafeeq et al., 2011c; van Hijum et al., 2005). The DNA microarray data have been submitted to GEO with accession number GSE41448.

RESULTS

In silico analysis of ROK-family protein members in S. pneumoniae

The ROK family of proteins consists of transcriptional repressors, kinases and uncharacterized proteins (Conejo et al., 2010; Titgemeyer et al., 1994). The ROK-family
repressor proteins are generally about 400 aa in length, while the kinases are about 100 aa shorter due to the absence of a helix–turn–helix (HTH) DNA-binding motif (Conejo et al., 2010). A BLAST search for ROK-family proteins in the genome of *S. pneumoniae* D39 strain revealed the presence of four ROK-family proteins. Here, these proteins are named RokA (SPD0423), RokB (SPD0580), RokC (SPD1488) and RokD (SPD1970) (Table S2). The rokA gene, which is present upstream of a PTS gene cluster (SPD0424-8) (Fig. 1a), encodes a 407 aa protein with an N-terminal HTH (helix–turn–helix) DNA-binding motif. The second ROK protein, RokB, is a putative glucose kinase and is about 90 aa shorter than RokA, due to the absence of an N-terminal HTH DNA-binding motif. RokC is located in close proximity to a gene (SPD1489) that is involved in sialic acid utilization and metabolism (Marion et al., 2011b). The fourth ROK protein, RokD, is encoded in an operon with three other hypothetical genes. The function of this operon is unknown. However, this operon was highly upregulated in the presence of cellobiose or in the absence of glucose (unpublished data). Thus, RokA seems to be the only ROK protein in *S. pneumoniae* that has an N-terminal HTH DNA-binding motif. It is therefore to be expected that it can act as a transcriptional repressor, possibly involved in carbohydrate metabolism. As the fitness of *S. pneumoniae* probably requires proper ways to ensure optimal expression of its sugar metabolic genes, we started to investigate the role of RokA and the gene cluster with which we hypothesized it to be associated.

**Organization of gene cluster SPD0424-8**

The organization of the gene cluster lying adjacent to rokA, SPD0424-8, is shown in Fig. 1(a). It consists of five genes and is most likely to be organized in one transcriptional unit. These genes putatively encode the following proteins: SPD0424, a cellobiose/lactose-specific PTS IIC component; SPD0425, a hypothetical protein; SPD0426, a cellobiose/lactose-specific PTS IIA component; SPD0427, an enzyme 6-phospho-β-galactosidase; and SPD0428, cellobiose/lactose-specific PTS IIIC components.

To investigate whether this gene cluster is transcribed as one transcriptional unit, transcriptional lacZ fusions of both IRs (SPD0424 and SPD0426) were constructed in the pPP2 plasmid and introduced into D39 wild-type. β-Galactosidase assays showed that, of the strains containing these transcriptional lacZ fusions, only the one with the SPD0424-lacZ construct showed expression, whilst no expression of SPD0426-lacZ was observed in GM17 medium (Fig. 1b). Indeed, no clear promoter sequences could be identified in SPD0424-lacZ, whereas we could find obvious −35 and −10 sites in SPD0424 (Fig. 1a). In addition, a terminator structure is present downstream of SPD0428, but not before SPD0426. RT-PCR with a primer bridging the IR between SPD0424-5 and SPD0426-8 further confirmed that these genes are transcribed as a single transcriptional unit (Fig. 1c). Thus, from this point forward, this gene cluster will be designated the SPD0424-8 operon.

**RokA represses the expression of the SPD0424-8 operon**

To explore whether RokA is involved in the regulation of the SPD0424-8 operon, the rokA gene was replaced by a spectinomycin-resistance marker, PSD0424-lacZ and PSD0426-lacZ transcriptional fusions were transformed into the rokA mutant strain and β-galactosidase assays were performed with the strains containing these lacZ transcriptional fusions grown in GM17 medium. Deletion of rokA led to strong expression of PSD0424-lacZ, which

---

**Fig. 1.** (a) Organization of the rokA gene and its upstream operon (SPD0424-8). Arrows indicate putative promoters, while oval indicates putative terminators. The sequence of PSD0424 is indicated above. The translation initiation codon (ATG) is in bold type and the predicted −35 and −10 core promoter sequences are shown in boxes. IR, Intergenic region between SPD0424-5 and SPD0426-8. (b) Specific β-galactosidase activity of D39 wild-type (WT) and rokA mutant strains containing the PSD0424-lacZ and PSD0426-lacZ transcriptional fusions grown in GM17 medium. (c) RT-PCR analysis of the IR between SPD0424 and SPD0426. RT-PCR was performed on total RNA isolated from the rokA mutant grown in GM17 medium, without (RNA) and with (RT) reverse transcriptase treatment using the IR primer pair. Chromosomal DNA of D39 was used as a positive control.
suggests a role of RokA as a transcriptional repressor of the SPD0424-8 operon (Fig. 1b). No expression was observed with the P\textsuperscript{SPD0426-lacZ} fusion in the \textit{rokA} mutant strain. This also supports our interpretation that SPD0424-8 constitutes one transcriptional unit arising from P\textsuperscript{SPD0424} (Fig. 1b).

**DNA microarray analysis with the \textit{ΔrokA} mutant strain**

To investigate the effect of rokA deletion on the transcriptome of \textit{S. pneumoniae} D39, the wild-type was compared with its isogenic \textit{rokA} deletion strain grown in GM17 medium. Table 2 summarizes the transcriptome changes induced upon deletion of \textit{rokA}. The \textit{rokA} deletion seems to have a very specific effect on the transcriptome of \textit{S. pneumoniae} D39, since after applying the criteria of at least threefold difference as the threshold change and a \textit{P}-value < 0.001, the SPD0424-8 operon was the only cluster of genes significantly upregulated in the \textit{rokA} deletion strain. This confirms the function of RokA as a repressor of the SPD0424-8 operon and suggests that the SPD0424-8 operon is the only target of RokA.

**Identification of a RokA operator site in P\textit{SPD0424}**

To find the putative operator site for RokA, a series of truncations from the 5’ and 3’ ends of P\textit{SPD0424} (P1, P2, P3 and P4) was constructed (Fig. 2a). These truncated promoter fragments were transcriptionally fused to \textit{lacZ} and transformed to D39 wild-type and the \textit{rokA} mutant. β-Galactosidase activity measurements revealed that deletion of the region upstream of the predicted core promoter of the \textit{SPD0424} gene [from bp –240 to –188 (P1) and from bp –188 to –139 (P2) relative to the start codon], did not show a significant change in transcriptional activity compared with the full promoter of \textit{SPD0424} (P0) in wild-type D39 (Fig. 2b). This demonstrates that there is no promoter activity residing in this region (bp –240 to –139 relative to the start codon), and also indicates the absence of an operator site that mediates RokA-dependent repression. However, deletion of the –34 to +1 (P3 and P4) region relative to the start codon of \textit{SPD0424} gene did lead to derepression of P\textit{SPD0424} expression in wild-type D39. In the \textit{rokA} mutant, derepression of P0, P1 and P2 was observed compared with the wild-type, whereas P3 and P4 were not.

**Table 2. Summary of transcriptome comparison of \textit{S. pneumoniae} wild-type strain D39 with the isogenic \textit{rokA} mutant (SS500) grown in GM17 medium**

<table>
<thead>
<tr>
<th>D39 locus tag</th>
<th>R6 locus tag</th>
<th>Function</th>
<th>Ratio*</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPD0423</td>
<td>spr0420</td>
<td>ROK family transcriptional repressor, RokA</td>
<td>–23.6</td>
</tr>
<tr>
<td>SPD0424</td>
<td>spr0421</td>
<td>Cellobiose-specific EIIC PTS component</td>
<td>57.2</td>
</tr>
<tr>
<td>SPD0425</td>
<td>spr0422</td>
<td>Hypothetical protein</td>
<td>84.8</td>
</tr>
<tr>
<td>SPD0426</td>
<td>spr0423</td>
<td>Lactose-specific EIIA PTS component, LacF-1</td>
<td>33.6</td>
</tr>
<tr>
<td>SPD0427</td>
<td>spr0424</td>
<td>6-Phospho-β-galactosidase, LacG-1</td>
<td>112.8</td>
</tr>
<tr>
<td>SPD0428</td>
<td>spr0425</td>
<td>Lactose-specific EIIBC PTS components, LacE-1</td>
<td>48.6</td>
</tr>
</tbody>
</table>

*Ratios > 3.0 or < −3.0 (D39 ΔrokA compared with D39 wild-type).*

---

**Fig. 2.** Subcloning of P\textit{SPD0424}. (a) Schematic drawing of P\textit{SPD0424} truncations from both the 5’ and 3’ ends. The oval indicates the position of the putative RokA operator site. Numbers next to the vertical dotted lines indicate the positions of the truncations with reference to the start codon (ATG= +1). Horizontal solid black lines indicate the length of truncated fragments. ■ indicates the positions of the putative –35 and –10 core promoter sequences. (b) Expression (β-galactosidase activity in Miller units) of the truncated P\textit{SPD0424} promoter fragments in \textit{S. pneumoniae} wild-type D39 and the \textit{rokA} mutant grown in GM17 medium, as measured via transcriptional lacZ fusions.
expression of P3 and P4 was similar to that in the wild-type (Fig. 2b). This suggests the presence of a RokA operator site in the −34/+1 region that mediates RokA-dependent repression of expression of the SPD0424-8 operon.

**Verification of the RokA-binding site by means of EMSAs and DNase I footprinting**

To study the direct interaction of RokA with its target promoters, we performed EMSA experiments with purified Strep-tagged RokA. These assays indicated that RokA binds specifically to PSD0424, but not to PSD0426 (Fig. 3), which is consistent with the above-mentioned transcriptional reporter studies, and suggests that RokA confers its regulatory effect on expression of the SPD0424-8 operon by direct binding to its target promoter. To find out whether the region identified above mediates the repressive effect of RokA (−34 to +1) and indeed contains its binding site, we performed EMSA experiments with truncated fragments of PSD0424 [see Fig. 2(a) for a schematic overview of the truncations]. When the −34 to +1 region of PSD0424 was deleted, binding of RokA was abolished completely (i.e. P3; Fig. 3), even at higher concentrations of RokA-Strep. However, RokA-Strep did retard promoter fragments of PSD0424 that do contain the −34 to +1 region (i.e. P1 and P2; Fig. 3). Thus, these RokA–PSD0424 interaction assays indicate that RokA is directly involved in transcriptional control of PSD0424 and, in addition, suggest the occurrence of an operator site for RokA in the −34 to +1 region relative to the start codon of SPD0424.

We next wanted to identify more exactly the RokA recognition sequence. Therefore, a DNase I footprint experiment was performed. For this purpose, the forward strand of PSD0424 was radiolabelled with [γ-33P]ATP and incubated with increasing concentrations of RokA-Strep. A very clear region of protection was observed that lies within the area spanning bp −34 to +1 of the PSD0424 promoter (Fig. 4). When the DNase I footprinting experiment was performed with the labelled complementary DNA strand, a similar protection pattern was seen (data not shown). Further analysis of this protected region revealed a 25 bp AT-rich possible operator site (5′-TATATTTAATTAT-TTTAAAAATAAA-3′) (Fig. 4), with resemblance to the predicted operator site of XylR (5′-ACTTTAANNNNNTTTAAAAAGT-3′) in Firmicutes (Gu et al., 2010), which we hypothesize to function as the RokA operator.

![Fig. 3.](image-url) **Fig. 3.** *In vitro* interaction of RokA-Strep with the promoter regions of SPD1829, SPD0424 and the full-length (PSD0424 : P0) and truncated PSD0424 promoter fragments (P1, P2 and P3). Purified RokA-Strep was added at concentrations of 0, 100, 250 and 500 nM in lanes 1–4, respectively. Arrows indicate the position of the shifted probes. The presence of weaker bands for some of the DNA fragments that run higher than the free probe in the gel is a phenomenon that has also been seen by others in similar experiments; these bands may represent unspecific PCR products or single-stranded DNA (Albano et al., 2005; den Hengst et al., 2005a).

![Fig. 4.](image-url) **Fig. 4.** DNase I footprinting analysis of RokA binding to PSD0424. (a) Radioactively labelled probe comprising the forward strand of PSD0424 was treated with DNase I alone (lane 2) or in the presence of RokA-Strep at 100 nM (lane 3) or 200 nM (lane 4). Footprints are flanked on the left by a Maxam and Gilbert A+G sequence ladder (lane 1). The protected region is marked with a bar on the right. (b) Sequence of PSD0424. The proposed RokA operator site is shown in bold and underlined. The predicted −35 and −10 core promoter sequences are shown in boxes.
This 25 bp operator sequence was subsequently used to search the entire genome of *S. pneumoniae* D39 with Genome2D software (Baerends et al., 2004). This search revealed that the RokA operator site is exclusively present in *PSPD0424* even with five mismatches allowed, suggesting that the *SPD0424*-8 operon is the only direct target of RokA in *S. pneumoniae*.

RokA seems to mediate its repression by a 25 bp AT-rich operator site present in *PSPD0424*. BLAST searches with RokA revealed that RokA homologues are present in various streptococci. Therefore, we investigated whether the RokA operator site inferred from our experiments is conserved in these streptococci as well. Interestingly, a possible RokA operator site was found in streptococci that contains a similar composition of the *SPD0424*-8 operon as *S. pneumoniae* (Fig. 5). The predicted RokA operator site from these streptococci was aligned with that of *S. pneumoniae* and a 25 bp putative consensus sequence was generated (TATATTTWATTATAAAAAWAAAAA) (Fig. 5). The presence of a similar operon composition and operator site in these streptococci suggests a regulatory role of RokA that is similar to that found in this study.

**Search for a co-factor for RokA-mediated regulation**

Based on the fact that RokA regulates an operon containing genes that encode a PTS, we hypothesized that a specific carbohydrate compound could lead to relief of RokA-mediated repression of *PSPD0424*. To search for a co-factor for RokA, β-galactosidase assays were performed with *PSPD0424*-lacZ in M17 medium with addition of 0.5% (w/v) of different carbon sources (arabinose, cellobiose, galactose, fructose, lactose, glucose, maltose, mannose, mannitol, GlcNAc, GalNAc, raffinose, sorbitol, sucrose, trehalose, xylose and mucin) (Fig. 6). Interestingly, none of the tested sugars was able to induce the expression of *PSPD0424*-lacZ strongly (Fig. 6), although some of the sugars (e.g. raffinose) caused a slightly increased expression of *PSPD0424*-lacZ, while others (e.g. glucose) led to lower expression compared to M17 medium without addition of any carbon source (Fig. 6). It might be that the complex medium M17 contains a certain carbohydrate that acts as a co-repressor of RokA, thereby preventing expression of *PSPD0424*-lacZ in the wild-type. Therefore, we checked the expression of *PSPD0424*-lacZ in chemically defined medium in the presence of different sugars. Interestingly, no difference in expression of *PSPD0424*-lacZ was observed compared with M17 medium in the presence of different sugars (data not shown). It might be that a derivative or analogue of the sugars that were tested is the ‘natural’ ligand for RokA.

**DISCUSSION**

Many previous studies have indicated the ability of *S. pneumoniae* to utilize and respond to different sources of carbon (Bidossi et al., 2012; Buckwalter & King, 2012; Carvalho et al., 2011; Chapuy-Regaud et al., 2003; Iyer et al., 2005; Iyer & Camilli, 2007; Marion et al., 2011a; McKessar & Hakenbeck, 2007; Nieto et al., 2001; Shafeeq et al., 2011b). However, the regulatory mechanism of many carbohydrate-utilization PTSs has not yet been characterized in *S. pneumoniae* strain D39. The ROK family is a group of proteins comprising transcriptional repressors, sugar kinases and uncharacterized ORFs (Conejo et al., 2010; Titgemeyer et al., 1994). The role of ROK-family proteins in carbohydrate utilization has been already investigated in various bacteria, including *B. subtilis*, *E. coli* and *S. pneumoniae* (Decker et al., 1998; Dubeau et al., 2011; Kimata et al., 1998; Kreuzer et al., 1989; Plumbridge, 1995; Plumbridge & Pellegrini, 2004; Plumbridge, 1991). In this study, the role of the ROK-family transcriptional repressor RokA, which we hypothesized to be involved in

---

![Fig. 6. Identification of the RokA operator site in different streptococci. (a) Weight matrix of the identified RokA operator site in different streptococci. (b) Position of operator site in *PSPD0424* of different streptococci, where the predicted RokA operator site is shown in bold, the predicted −35 and −10 core promoter sequences are shown in boxes and the translation initiation codon is shown in italics. SP, *S. pneumoniae*; SS, *S. suis*; SE, *S. epui*; SG, *S. galolyticus*.](http://mic.sgmjournals.org)
carbohydrate utilization in *S. pneumoniae* strain D39, was investigated. We showed that RokA acts as a transcriptional repressor of the *SPD0424-8* operon and that this repression is mediated by direct binding of RokA to a 25 bp AT-rich DNA operator sequence present in *SPD0424*. Despite the fact that the *SPD0424-8* operon has high sequence similarity to other streptococcal lactose/cellobiose-specific PTSs, we were not able to identify a possible co-factor that can cause derepression of the *SPD0424-8* operon.

In *S. pneumoniae*, regulation of carbohydrate utilization is important not only for proper metabolic functioning of the cell, but also for pathogenesis. As in other low-GC, Gram-positive bacteria, CcpA plays a major role in regulation of carbohydrate-utilization and virulence genes in *S. pneumoniae* (Carvalho et al., 2011). There are some carbohydrate-utilization systems known to be regulated in a CcpA-independent fashion (Iyer & Camilli, 2007; Nieto et al., 1997; Shafeeq et al., 2011b). Also, the *SPD0424-8* operon seems not to be regulated by CcpA, based on DNA microarray analyses with a *ccpA* mutant in the presence of glucose and galactose (Carvalho et al., 2011). However, *SPD0424* does contain a putative cre box (5′-CACAAAAGCTTGTCAA-3′) (Novichkov et al., 2010). It might be that this cre site is not functional or that the CcpA effect is only evident when RokA is not repressing the operon.

The main differences between ROK kinases and repressor proteins are the presence of a conserved N-terminal ATP-binding motif DxGxT and the absence of an N-terminal HTH DNA-binding motif in the kinases (Conejo et al., 2010). Protein sequence alignment of RokA, B, C and D of *S. pneumoniae* strain D39 revealed that an N-terminal ATP-binding motif, DxGxT, is present in RokB, C and D, while the N-terminal HTH DNA-binding motif is only present in RokA. The presence of ATP-binding motif DxGxT might suggest a kinase function of RokB, C and D proteins. Therefore, further investigation of these proteins might help to understand their role in carbohydrate utilization. Moreover, ROK-family proteins also have a conserved metal-binding site [CxCCGxGxCx(E/D)] that coordinates a single atom of zinc (Conejo et al., 2010). Protein sequence alignment of RokA, B, C and D revealed that this metal-binding site is conserved only in RokA and RokB, while absent in RokC and RokD. This suggests a role of zinc, which is an important factor in the lifestyle of *S. pneumoniae* (Kloosterman et al., 2008; Shafeeq et al., 2011a), in the structure and proper functioning of ROK proteins.

The number of PTSs varies from 15 to 20 between different pneumococcal strains and many of the PTSs are not conserved in all strains of *S. pneumoniae* (Bidossi et al., 2012). BLAST searches revealed that the *SPD0424-8* operon is present in most pneumococcal strains available on the KEGG website (http://www.genome.jp/kegg/) and therefore seems to be an important component of the arsenal of pneumococcal carbohydrate-acquisition systems that together determine its fitness. Recently, a gene cluster comprising a gene encoding a ROK-family transcriptional regulator and a putative cellobiose/lactose PTS operon, located on a genomic island, has been described in strain WCH206 (McAllister et al., 2012). This operon was found to affect virulence in a murine model of pneumonia/sepsis (McAllister et al., 2012). It has a similar structure to the *SPD0424-8* operon, it is around 30% identical on the protein level, and is also regulated by the ROK-family transcriptional regulator. Therefore, in future studies it would be interesting to determine the role of the *SPD0424-8* operon in pneumococcal virulence as well.

BLAST searches of *SPD0424-8* revealed 30% sequence identity to proteins involved in cellobiose and lactose utilization. However, data of a transcriptome comparison of cellobiose with glucose and lactose with glucose (unpublished data) indicate no effect of these sugars on the expression of *SPD0424-8*, consistent with *SPD0424-lacZ* expression data in the presence of different sugar sources (Fig. 6). This suggests that this operon is probably not involved in the metabolism of these sugars. Moreover, to determine the conservation of RokA in other streptococcal species, we performed BLAST searches of the sequence of RokA against different streptococci. This revealed that RokA is absent from many streptococci, including *Streptococcus mutans*, *Streptococcus gordonii* and *Streptococcus pyogenes*. However, in other streptococci, such as *Streptococcus suis*, *Streptococcus galolyticus* and *Streptococcus equi*, a homologue can be found. In these streptococci, rokA is annotated as XylR, a xylose-dependent repressor. However, our β-galactosidase assays with *SPD0424-lacZ* showed no effect of xylose on the expression of the *SPD0424-8* operon.

Fig. 6. Specific β-galactosidase activity of wild-type D39 containing *SPD0424-lacZ* grown in M17 medium with added carbon sources (0.5%, w/v).
(Fig. 6). It might be that structural analogues of one of the above-mentioned or different sugars are present in the host and are the substrate for RokA.

ACKNOWLEDGEMENTS

S.S. is supported by HEC (Higher Education Commission) Pakistan through the HEC-NUFFIC overseas scholarship programme and T.G.K. was supported by the EU-funded Pneumopath project HEALTH-F3-2009-222983.

REFERENCES


Edited by: M. Kilian