DNA RESEARCH **16**, 177–186, (2009) doi:10.1093/dnares/dsp007

SHORT COMMUNICATION

Transcriptional Regulation of the Capsular Polysaccharide Biosynthesis Locus of *Streptococcus Pneumoniae*: a Bioinformatic Analysis

MIRIAM MOSCOSO and ERNESTO GARCÍA*

Centro de Investigaciones Biológicas, (CSIC) and CIBER de Enfermedades Respiratorias (CIBERES), Ramiro de Maeztu, 9 28040, Madrid, Spain

(Received 10 December 2008; accepted 10 April 2009; published online 8 May 2009)

Abstract

The polysaccharide capsule of *Streptococcus pneumoniae* is the main virulence factor, which makes the bacterium resistant to phagocytosis. Expression of capsular polysaccharide must be adjusted at different stages of pneumococcal infection, thus, their transcriptional regulation appears to be crucial. To get insight into the existence of regulatory mechanisms common to most serotypes, a bioinformatic analysis of the DNA region located upstream of the capsular locus was performed. With the exception of serotype 37, the capsular locus is located between *dexB* and *aliA* on the pneumococcal chromosome. Up to 26 different sequence organizations were found among pneumococci synthesizing their capsule through a Wzy-polymerase-dependent mechanism, mostly varying according to the presence/absence of distinct insertion elements. As a consequence, only ~250 bp (including a 107 bp RUP_A element) was conserved in 86 sequences, although only a short (ca. 87 bp) region located immediately upstream of *cpsA* was strictly conserved in all the sequences analyzed. An exhaustive search for possible operator sequences was done. Interestingly, although the promoter region of serotype 3 isolates completely differs from that of other serotypes, most of the proteins proposed to regulate transcription in serotype 3 pneumococci were also predicted to function as possible regulators in non-serotype 3 *S. pneumoniae* isolates.

Key words: capsular polysaccharide; *Streptococcus pneumoniae*; transcriptional regulation; bioinformatic analysis; operator sequences

Streptococcus pneumoniae, or pneumococcus, is a significant human pathogen causing both mucosal, such as otitis media and pneumonia, and systemic diseases, including septicemia and meningitis. Pneumococcal capsular polysaccharide (CPS) is immunogenic and induces type-specific protective immunity.[1] Although a 23-valent CPS vaccine and a heptavalent protein–polysaccharide conjugate vaccine (designed for pediatric use) are currently available, they are far from being satisfactory. At least 91 different CPS have been described to date in *S. pneumoniae*.[2] The *cap* (or *cps*) cluster of *S. pneumoniae* (Supplementary Fig. S1), which appears to be organized as a single transcriptional unit (see below), is located between *dexB* and *aliA* (two genes that do not participate in capsule biosynthesis),[3] with the notable exception of the serotype 37 CPS whose synthesis depends on a single protein encoded by a gene located far from the *cps* locus on the *S. pneumoniae* chromosome.[4] At least 89 of the 91 pneumococcal CPS known to date appear to be synthesized by a Wzy-polymerase-dependent mechanism in which individual repeat units assembled on undecaprenyl phosphate on the inner face of the bacterial membrane are polymerized on the outer membrane surface.[5] The synthesis of CPS of serotypes 3 and 37 is catalyzed by a single, membrane-bound glycosyltransferase (synthase) referred

Edited by Katsumi Isono

* To whom correspondence should be addressed. Tel. +34 91-837-3112. Fax. +34 91-536-0432. E-mail: e.garcia@cib.csic.es

© The Author 2009. Kazusa DNA Research Institute.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

to as Cap3B/Cps3S, and Tts, respectively. In these two serotypes, the common sequences located at the 5′ end of all the other loci and that code for regulatory proteins either are not present (type 37) or are mutated and not transcribed (type 3) (Supplementary Fig. S1).[4]

One of the most striking features of the pneumococcal *cps* locus is its huge genetic divergence, since only a few genes are conserved among different clusters.[6] These genes are located at the 5′ end of the *cps* locus and are known to be involved in the processing, regulation and export of CPS and, possibly in the attachment of the CPS to the cell wall.[7] Remarkably, only the first gene of the cluster (*cap*/*cpsA*) is over 90% identical in all the gene clusters. A near consensus, functional promoter sequence (5′-TAGACA-17 nucleotides-TATAAT3′) (*cpsp*) has been identified 30 nucleotides upstream of the initiation codon of the *cap*/*cpsA* gene, and the transcription start point of the *cap*/*cps* operon has also been determined.[8]

Since the capsule makes the bacterium resistant to phagocytosis, maximal expression of CPS is essential for systemic virulence, although the capacity to regulate the amount of CPS also appears to be crucial, e.g. a reduced level of CPS is an absolute requirement for efficient nasopharyngeal colonization.[9] It has been reported that the expression of some capsular genes was reduced when pneumococcal cells were treated with penicillin or vancomycin.[10,11] Nevertheless, the existence of possible regulatory pathways for CPS biosynthesis, however, is basically unknown and controversial.[12–17]

We identified 115 different entries in the databases fulfilling the requirements established, i.e. the complete nucleotide sequence was available from the termination codon of *dexB* to the initiation codon of *cpsA*. As indicated in Supplementary Table S1, 26 different sequence organizations (SOs) were identified in this region. Sequence organization 10 (37 sequences) was by far the most frequent, and together with SO_1 (22 sequences), SO_2 (8 sequences) and SO_22 (18 sequences) accounted for close to 75% of all the SOs. Although, in most cases, the nucleotide sequence of only one isolate per serotype was available, it was noted that different strains with identical CPS may have different SOs, mostly varying according to the presence (or absence) of distinct insertion sequences (ISs) (Supplementary Fig. S2). Despite the appreciable polymorphism in the *dexB*–*cpsA* region, two major SO groups could be recognized: those containing an intact or truncated copy of IS*630*_Spn*1* (from SO_1 to SO_21) (designated group I) and SO_22 to SO_26 (group II) lacking this IS and some additional fragments. IS*630*_Spn*1* is a ~0.9 kb element firstly reported by Oggioni and Claverys.[18] When examining group I sequences, with the exception of SO_9, SO_12 and SO_21 that contain one or more IS in this region, it is evident that similarity in the vicinity of the *cpsp* region was restricted to a ca. 250 bp sequence (Supplementary Fig. S3). Two different regions can be distinguished in this sequence: (i) a 140 bp fragment that includes a 107 bp RUP_A sequence and a ~34 bp sequence that resembles the insertion site of IS*1381* (Fig. 1A) and (ii) a short (~87 bp) region embracing the *cpsp* region (Fig. 1B) (indicated by a green rectangle in Supplementary Fig. S2). RUP_A is a highly repeated extragenic element of *S. pneumoniae* that is very similar to the inverted terminal repeats of IS*630*-Spn*1* and might be *trans*-activated by transposase and promote sequence rearrangements.[18] The possibility that RUP elements may serve as binding sites for regulatory proteins has been proposed.[19] Interestingly, transcriptional start sites have been mapped in a RUP_C element located upstream of *tts*, the gene encoding the polysaccharide synthase responsible for the synthesis of serotype 37 CPS.[20]

In sharp contrast with most of the group I sequences, group II sequences either lack RUP_A or it is separated from the *cpsp* region by intervening ISs (Supplementary Fig. S2). Consequently, the ~87 bp region that contains the promoter of the *cps* gene cluster (Fig. 1B) turned out to be the only conserved sequence in all the *S. pneumoniae* isolates that synthesize their CPS through a Wzy-polymerase-dependent mechanism. However, on closer examination of the full alignment (Supplementary Fig. S3), the existence of potentially significant polymorphisms in this region was revealed. In addition to the T to C transition at position −8 in the unencapsulated R6 strain (AE008412), an identical mutation was found at position −11 in strains E294 (CR931699; serotype 33B), CSF/79 (CR931701; serotype 33D), SP18-BS74 (NZ_ABAE01000002; serotype 18C) and WCH18 (CR9316640; serotype 6B). All these mutations cause a change in the −10 region of *cpsp* from the consensus sequence TATAAT to TATAAC (in R6)[21] or TACAAT (in the latter four strains) thereby potentially reducing the corresponding promoter strength. In particular, the reported reduced transcription of the *cps* genes associated with capsular polysaccharide formation in the unencapsulated strain R6[21,22] is likely attributable to a mutation in the −10 box of the R6 *cps* promoter from the consensus TATAAT to TATAAC (see above). Moreover, only the last four nucleotides of the −35 promoter box (TAGACA) were conserved among the 115 sequences examined (Supplementary Fig. S3). It should be noted that in all the sequences examined, 17 perfectly conserved nucleotides separate the −35 and −10 promoter boxes. The transcription initiation site (indicated as +1 in Fig. 1B) is also conserved in all the sequences analyzed.

```
A
                                                                                     *******************
     8888888888888888888888888888888888888888888888888888888888888888888888777777777777777777888888888888888
     6446445666666666666666466656534346246666666666456664666666666656455554555555545455454566665777766666566666666666
-203 WWWAKTAKATTGAAACTAGARTAGTASACMTCTGCTTCTAAAAHA-TTKTTAGAAAWCGATTTRACTGTCCYGAWYRWTTTRTYMTRTTCTTRTTTCAWTYTAYTATA -97
     -200                                            -150                                              -100

     8888888888888888888888888888888888
     6666666666645666666666666066664666
-96  TTTTTGKTTYGCGGGAAGTCTACTWA-GATACTTA -63

B
     111111111111111111111     1111111111111111111111111111111111111111111111111111111111111
     1111111111111111111106777000111111111111111111111111111111111111111111111111111111111111
     333332333333455043509560276645555555555555555555555515545543553555455555555555555544145
-62  AAGATGCAGATAGTRRAAAWAA-GRTGTAGACATTACCGTAAAAAAGTGATATAATYRTAHRATRTTSAAKGTATAGGTRTTAATCATG +26
             -50                                              +1                          └─► cpsA

C
-121 GGAAACAGAGGTTAGGAAAGTAATYAG/TTAACGGGATATCTTTTCAAAGCTGATACTAAGGCAC/AAAAAAAGTTTGATATTCCCC -35
                 -100                                                 -50

-34  TTGACAATAGATAAAATTATTATATAATTAAACTATTGCTTTTTAAATAAAGTGAGAATATTAATAATRCAGAGAAAGAGGACTGTAGTAAAATG +61
                               +1                                                   +50       └─► cps3A
```

Figure 1. Consensus sequences derived from alignments of strains synthesizing CPS through a Wzy-polymerase- (**A** and **B**) or a synthase-dependent mechanism (**C**). Eighty six (A), 115 (B) and four (C) sequences were aligned. The frequency of each nucleotide is indicated above the sequence, in vertical format. The position of the transcription initiation site is assigned +1 and other positions are numbered accordingly. (A) The RUP_A element is shaded and a sequence possibly related to some ISs is italicized. Asterisks indicate a region deleted in several strains. Abbreviations: H, A or C or T; K, G or T; M, A or C; R, A or G; S, C or G; W, A or T; Y, C or T. (B) The initiation codon of *cpsA* is indicated with an arrow. The −35, −10 and the transcription initiation site are indicated by white lettering on a black background. The underlined sequence corresponds to that tandemly duplicated in some strains. (C) Aligned nucleotide sequence of three serotype 3 strains. At positions −94 and −56, a slash indicates a T (or no nucleotide), or an A (or no nucleotide), respectively.

The effect of intrinsic curvature upstream of a bacterial promoter on the efficiency of transcription was first reported in the early 1980s. To date, there are countless examples indicating the importance of a curved DNA sequence during steps of transcription, mainly in regulating the transcription initiation process.[23] It has been recently shown that global transcription factors as well as several other transcriptional regulators have a significant tendency to regulate operons with curved DNA sequences in their upstream regulatory regions.[24] Bendability/curvature propensity plots were constructed with the help of the BEND.IT server (http://hydra.icgeb.trieste.it/dna/bend_it.html) using defaults parameters with the exception of a 20-nucleotide window size.[25] We used the sequence of the 228 bp fragment (including the ATG initiation codon) located upstream of *cpsA* from the serotype 2 strain D39 (AE026 471) to predict its curvature. The curvature-propensity plot, constructed using DNase I-based trinucleotide parameters, shows four peaks around positions −125, −98, −43 and −17 of magnitudes ≥ 9.0 (data not shown). It should be underscored that the regions showing these potential curvatures occur in the 87 bp fragment containing *cpsp*, which is the only stretch common to all pneumococci that synthesize their CPS through a Wzy-dependent pathway (see above). A computer prediction of bendability of the above mentioned DNA fragment (AE026471) (data not shown) rendered at least seven conspicuous peaks (positions −175, −137, −128, −77, −57, −9, +5) with magnitudes ≥ 5.0. In addition, the bendability plot showed two long troughs (from positions −158 to −151 and −103 to −96; magnitudes ≤ 3.5) indicative of rigid segments. Interestingly, the latter rigid segment embraces a potential static curvature (centered at −98; see above) and is located near the 3′ end of the RUP_A element.

The annotated genome sequence of the *S. pneumoniae* strains TIGR4, R6 and D39 revealed the presence of, at least, 80 putative transcriptional regulatory genes, representing ~4% of the complete genome. Unfortunately, in only few cases was the operator site identified. Putative operators were searched for using FUZZNUC (http://mobyle.pasteur.fr/cgi-bin/MobylePortal/portal.py?form=fuzznuc) and the consensus sequences reported in Fig. 1A and B allowing a maximum of two mismatches. Among the currently known (and putative) pneumococcal operators, only sequences similar to the binding sites of ComX1,[26] CopY,[27,28] MalR,[29] GlnR[30,31] or RitR[32] were detected (Table 1). As a notable exception, it should be mentioned that none of the four potential binding sites for GlnR are located in the region common to all isolates whose CPS is synthesized through a Wzy-dependent pathway.

We also searched conserved regions for additional operator sequences using proven and putative binding sites from related streptococci (from published works; see Supplementary Table S2) and *Bacillus subtilis* (from the database of transcriptional regulation in *B. subtilis*; DBTBS).[33] Among the reported binding sites for transcriptional regulators,

Table 1. Binding sites of transcriptional regulators putatively involved in CPS biosynthesis of *S. pneumoniae*

Organism	Transcriptional regulator (Accession no.)	Binding site	*S. pneumoniae* ortholog (Accession no.)	Identity/ similarity ($\log_{10}E$ value)	Sequence (5′ position)[a]	
					Wzy-dependent[b]	Synthase-dependent
S. pneumoniae						
	ComX1 (Q8CM18)	TACGAATA	—	—	TACtAAgA (−75f)	TgaGAATA (+19f)
					TAacAATA (−152r)	TAttAATA (+25f)
					TAaGAAcA (−111r)	TAatAATR (+28f)
					TACGAtTA (−4r)	TAttAATA (+25r)
	CopY (Q8DQJ7)	KACAN$_2$TGTA	—	—	TAgARTaGTA (−186f)	GAggACTGTA (+44f)
					TACACATcTg (−178f)	GACtGTaGTA (+47f)
					TAaAGATGcA (−63f)	
					aAaAGGTGTA (−44f)	
					TAtAATcGTA (−13f)	
					TAtAGGTGTt (+10f)	
					TACtAYTcTA (−177r)	
					cAgATGTGTA (−169r)	
					GACAAATtgA (−120r)	
					TgCATCTtTA (−54r)	
					TACACCTtTt (−35r)	
					TACgATTaTA (−4r)	
					aACACCTaTA (+19r)	
	MalR (P0A4T2)	AAACGTTT[c]	—	—	AAWCGaTT (−149f)	AAAaGTTT (−52f)
					AAAgGTgT (−43f)	AAACtaTT (−5f)
					AAtCGWTT (−142r)	AAACtTTT (−45r)
					AcACcTTT (−36r)	AAtaGTTT (+3r)
	GlnR (Q8DQX7)	TGTNAN$_7$TNACA	—	—	TcTAAAAHATTKTTAgA (−166f)	TGaTATTCCCCTTGACA (−45f)
					TcTTATTTCATTTTACt (−115f)	TtTTAAATAAAGTGAgA (+7f)
					TcTAAMAATATTTTAgA (−150r)	TGaGAATATTAATAAtR (+19f)
					aGTAAAATGAAATAAgA (−99r)	TGTCAAGGGGAATAtCA (−29r)
						TcTCACTTTATTTAAaA (+23r)
						YaTTATTAATATTCtCA (+35r)
	RitR (Q04M91)	WNATTANW$_3$RWYRR	—	—	ATATTATATTGAaAc (−201f)	TAATcAGTTTAACGG (−101f)
					TGATcAATTTGTCAt (−132f)	AGATaAAATTATTAt (−26f)
					TTAcTATATTtTTGG (−103f)	AAATTATTATATaAt (−21f)
					AAtaTAGTAAAATGA (−94r)	TTATTATATAATTAA (−18f)
						TAATTAAAcTATTGc (−10f)
						AAgTgAGAATATTAA (+16f)

						TGAgaATATTAATAA (+19f)
						ATATTAATAAtgCAG (+24f)
						TGATTACTTTccTAA (−96r)
						TAATaATTTTATCtA (−13r)
						TAATTATATAATaAt (−5r)
						TTAaTtATATAATAA (−4r)
						AGtTTAATTAtATAA (+1r)
						TTATTtAAAAAgCAA (+16r)
						ACtTTATTTAAAaAG (+19r)
	AdcR (Q04I02)	TAACYRGTTAA	—	—	—	TAAtCAGTTtA (−91f)
						TAtCCcGTTAA (−83r)
S. pyogenes						
	CovR/CsrR (Q8P2J8)	DDHHATTARAR	CsrR (Q8DR53)	45/68 (−49)	TTATATTgAAA (−198f)	GGTTAggAAAG (−112f)
					TGAAAcTAGAR (−192f)	AGTAATYAGtt (−103f)
					TGCTtcTAAAA (−170f)	ATCTtTTcAAA (−83f)
					HATTkTTAGAA (−159f)	TGATAcTaAGG (−70f)
					GTCTAcTAAgA (−78f)	TGACAaTAGAt (−33f)
					AGATAcTtAAA (−70f)	AATAgaTAAAA (−29f)
					GATAcTTAAAG (−69f)	TAAAATTAttA (−23f)
					AGATAgTgAAA (−54f)	AATTATTAtAt (−20f)
					GATAgTgAAAA (−53f)	ATATAaTtAAA (−13f)
					AGACATTAccG (−35f)	TATAATTAAAc (−12f)
					TTACcgTAAAA (−30f)	TGCTtTTtAAA (+3f)
					TACCgTaAAAA (−29f)	TTTAAaTAAAG (+8f)
					TTTCAaTAtAA (−188r)	TAAAgTgAGAA (+14f)
					TACTATTctAG (−177r)	GAATATTAAtA (+22f)
					ATDTtTTAGAA (−157r)	TATTAaTAAtR (+25f)
					GGACAgTyAAA (−133r)	TAATAaTRcAG (+28f)
					GAACATgAcAA (−114r)	
					TGAAAYaAGAA (−106r)	
					TAAAATgAAAt (−101r)	
					GTAAAaTgAAA (−100r)	
					ATATAgTAAAA (−95r)	
					GTATcTTAGtA (−65r)	
					ATACATTgAAc (+12r)	

Continued

Table 1. Continued

Organism	Transcriptional regulator (Accession no.)	Binding site	*S. pneumoniae* ortholog (Accession no.)	Identity/ similarity ($\log_{10}E$ value)	Sequence (5′ position)[a]	
					Wzy-dependent[b]	Synthase-dependent
S. agalactiae						
	RovS (Q8E447)	AWAAWVHTDAWN$_{6/7}$ WTKWWAMDWAK	SPD_0939	52/73 (−76)	—	ATAtAATTAAACTATTGCTTTTTAAATAa (−13f)
B. subtilis						
	CitT (O34534)	WWCAAA	RpsI (A5MLP7)[d]	27/50 (−17)	TTgAAA (−193f)	AACAgA (−118f)
					TACAcA (−178f)	TTCAAA (−78f)
					TTCtAA (−167f)	TACtAA (−67f)
					TAgAAA (−153f)	cACAAA (−59f)
					ATCAAt (−130f)	AAaAAA (−55f)
					TACtAA (−75f)	AAaAAA (−54f)
					AAaAAA (−46f)	ATaAAA (−24f)
					TAaAAA (−24f)	ATtAAA (−8f)
					AAaAAA (−23f)	TTtAAA (+8f)
					TTCAAt (+3f)	AAtAAA (+12f)
					TTCAAt (−189r)	TRCAgA (+34f)
					AACAAt (−153r)	*AACtRA* (−93r)
					TTCtAA (−149r)	*TTgAAA* (−74r)
					gTCAAA (−138r)	*ATCAAA* (−42r)
					gACAAA (−120r)	*TAaAAA* (+11r)
					ATgAAA (−105r)	*TTaAAA* (+12r)
					AcCAAA (−88r)	*TTtAAA* (+13r)
					AACcAA (−87r)	
	DeoR (P39140)	TTCAAT	Spr0228 (Q8DRC3)[d]	28/50 (−30)	aTCAAT (−130f)	TTCAAa (−78f)
					TTCAWT (−109f)	TTaAAT (+9f)
					TTCAAT (+3f)	TTtAAT (−3r)
					TTCAAT (−189r)	
					TTCAcT (−45r)	
	GerE (P11470)	RWWTRGGYN$_2$YY	RR03 (Q8DR45)	49/67 (−6)[e]	GATTtGaCTGTC (−145f)	AATTAaaCTATT (−9f)
					ATTTGacTGTCC (−144f)	GAAaAGaTATCC (−76r)
					GTATAGGTRTTa (+10f)	
					AAATcGWTTTCT (−141r)	
					cTTTAaGTATCT (−59r)	
					GATTAtaTCACT (−7r)	
	CcpA (P25144)	WTGNAANCGNWN$_2$CW	CcpA (Q97NM1)	54/74 (−96)	TaGAAAWCGATTTrA (−152f)	TTcAAAGCtGATACT (−78f)
					ATaTAATCGTAAGaT (−14f)	
					gTyAAATCGWTTTCT (−138r)	

Spo0A (P06 534)	TGTCGAA	RR09 (Q8DQN8)	38/62 (−10)	TaTtGAA (−195f)	TtTCaAA (−79f)
				TGTaGAc (−38f)	TGTaGtA (+50f)
				TGTtcAA (+1f)	TaTCaAA (−41r)
				TtTaGAA (−161r)	TGTCaAg (−29r)
				TtTCtAA (−148r)	
				aGTCaAA (−137r)	
				TGaCaAA (−119r)	
				TGTCtAc (−31r)	

[a] f and r indicate whether the sequence corresponds to the forward or reverse sequence of that included in the EMBL database (AF026471 or Z47210), respectively. Unless otherwise stated, a maximum of two mismatches were allowed.
[b] Sequences corresponding to those common to all strains are indicated with a gray background. Mismatches are indicated by lowercase lettering.
[c] This sequence is a subset of the CcpA binding site.
[d] Only one mismatch was allowed.
[e] Similarity restricted to part of the protein (from residue 145 to 197 of the pneumococcal protein and 12 to 64 of that of *B. subtilis*).

only those with a clear pneumococcal ortholog were considered. Many potential *Streptococcus pyogenes* CovR/CsrR binding sites were found[34] and the locations of potential operators for five different *B. subtilis* transcriptional regulators were also determined (Table 1). Furthermore, we also examined the conserved regions for the presence of direct and/or inverted repeats but, although inconclusive at the moment, these searches suggested that other binding sequences are present in this promoter.

As reviewed elsewhere,[4] the promoter region of serotype 3 *S. pneumoniae* isolates completely differs from that of other serotypes. Three different serotype 3 isolates of *S. pneumoniae* were completely sequenced between *dexB* and the ATG initiation codon of *cps3A* (Supplementary Table S1) and two SOs were found, differing only in terms of the length of the IS*630*_Spn*1* element (Supplementary Fig. S4). The three isolates showed $\geq$ 95% nucleotide sequence identity between the deleted copy of *cpsD* and the ATG initiation codon of *cps3A*. This was also true for the serotype 3 strain WU2 (Accession no. U66846 and U15171). Downstream of the deleted copy of *cpsD*, three pseudogenes were detected: an internal fragment (83 bp) of *wchA*, a gene putatively encoding the initial sugar transferase from a serotype 20 *S. pneumoniae* strain; the so-called *orf5* (491 bp) putatively coding for a membrane protein (corresponding to Spr1830 in the genome of strain R6) and a 403 bp fragment showing 86% identity to IS*1548*. A ca. 180 bp DNA fragment containing the serotype 3 promoter (*cps3p*) was located between IS*1548* and *cps3A*. Polymorphism was only found at four positions in the three serotype 3 isolates that were aligned (Supplementary Fig. S3C).

The curvature-propensity plot of the region containing *cps3p* showed two prominent peaks around positions −61 and −18. Further, two peaks of bendability $\geq$ 5.0 were predicted at positions −35 and −66 whereas rigid segments (bendability $\leq$ 3.5) appeared to span positions −49 to −45, −14 to −11, and −2 to +8 (data not shown).

We also searched the ca. 180 bp sequence upstream of *cps3A* (Fig. 1C) for potential operators as for the other serotypes (Table 1). Interestingly, with the exceptions of AdcR and RovS, the proteins proposed to regulate *cps* transcription in serotype 3 pneumococci were also predicted to function as possible regulators in non-serotype 3 *S. pneumoniae* isolates, which might indicate the existence of common regulatory mechanism in otherwise divergent sequences. Nevertheless, the potential relevance of this finding remains to be determined.

As already mentioned, experiments aimed to demonstrate that CPS biosynthesis is transcriptionally regulated have yielded conflicting results.

Notwithstanding, in an elegant electron microscopy study on cultured epithelial cells neither serotype 3 pneumococci in close contact with the host cell membrane nor invading pneumococci exhibited a visible capsular structure, whereas pneumococci not in close contact with the host membrane had a typical capsule. Moreover, *S. pneumoniae* cells expressed CPS in the lungs of infected mice, whereas bacteria in contact with lung epithelial tissue showed a drastic reduction in the density of the CPS layer.[35]

Recently, *in vitro* serotype-dependent expression of *cpsA* in transparent variants of *S. pneumoniae* has been observed, i.e. the serotypes/groups associated with invasive infections tend to express more *cpsA* than those frequently isolated from carriers.[17] Unfortunately, in that report, the number of isolates examined was insufficient for a reliable comparison of *cpsA* expression among different clones of the same serotype (or serogroup).

We should point out that the catabolite repressor protein CcpA (also denoted RegM) appears to be involved in transcriptional activation of the *cps* operon in the serotype 2 strain D39.[36] Besides, it has been reported that a *ccpA* mutant of the serotype 4 strain TIGR4 was drastically attenuated for infection of the lung and colonization of the nasopharynx.[37] Extracellular glucose concentrations might positively regulate the level of CPS biosynthesis. Thus, the glucose concentration is normally very low ($<$ 1 mM) in healthy nasopharyngeal secretions,[38] and a reduced amount of CPS is required for optimal attachment of the pneumococcus to epithelial cells.[9] In contrast, invading pneumococci encounter high glucose concentrations (5.4 mM) in the blood stream of healthy individuals[39] where maximum CPS biosynthesis is most needed. Although no data are available on serotype 3 pneumococci, it should be underscored that the *hasABC* operon, which is involved in the synthesis of the hyaluronic acid capsule of *S. pyogenes* and is phylogenetically related to the pneumococcal *cps3ABC* operon,[40] is significantly down-regulated in a *ΔccpA* mutant.[41]

Recent studies have also examined the role in virulence of several proven pneumococcal transcriptional regulators using microarrays. Kloosterman et al.[31] used a *glnR* mutant and did not observe a differential expression of the capsular locus, suggesting that the binding sites identified here (Table 1) might not be relevant. However, it has been reported that RitR represses the expression of *cps2N*, a gene involved in the biosynthesis of dTDP-rhamnose, one of the sugar nucleotides required for the biosynthesis of type 2 CPS.[32] Interestingly, binding sites for RitR have been found in the *cpsp* region (Table 1). It should be kept in mind, that the *S. pneumoniae* orphan response regulator RitR is very similar to the streptococcal global regulator CovR (also designated as CsrR) and there is conclusive evidence showing that CovR is a global regulator that either represses (*S. pyogenes*, *Streptococcus suis*) or upregulates (*Streptococcus agalactiae*) CPS biosynthesis.[42–44] Putative CovR-like operator sequences are located in the region containing the capsular promoter.

Although there is no experimental data on the existence of functional promoters (different of *cpsp*) located in the intergenic regions of the capsular locus, their possible presence and function should be mentioned. For example, an enhanced biosynthesis of the UDP-glucose dehydrogenase Cps2K has been observed in a D39 *ΔcodY* mutant.[45] A close examination of the 129 bp long region located between *cps2H* and *cps2I* (both genes located upstream of *cps2K*) revealed the existence of a putative promoter (5′-TAGTTG-18 nucleotides-TATTTT-3′) and, further upstream, of an imperfect palindrome (AATTTTtAGAgAATT) quite similar to the consensus binding site of CodY (AATTTTCWGAAAATT).[46] Whether these sequences are relevant *in vivo* requires further studies.

In summary, this study provides some of the sequence data needed to pave the way for systematically identifying the regulatory pathways of CPS biosynthesis in *S. pneumoniae*.

Acknowledgements: The authors wish to thank P. García and R. López for helpful comments and critical reading of the manuscript, A. Burton for revising the English version and E. Cano for skilful technical assistance.

Supplementary Data: Supplemental data are available online at www.dnaresearch.oxfordjournals.org

Funding

This work was supported by a grant from the Dirección General de Investigación Científica y Técnica (SAF2006-00390). CIBER de Enfermedades Respiratorias (CIBERES) is an initiative of ISCIII.

References

1. López, R. 2006, Pneumococcus: the sugar-coated bacteria, *Int. Microbiol.*, **9**, 179–190.
2. Park, I. H., Pritchard, D. G., Cartee, R., Brandao, A., Brandileone, M. C. C. and Nahm, M. H. 2007, Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*, *J. Clin. Microbiol.*, **45**, 1225–1233.
3. Bentley, S. D., Aanensen, D., Mavroidi, A., *et al.* 2006, Genetic analysis of the capsular biosynthetic locus

from all 90 pneumococcal serotypes, *PLoS Genet.*, **2**, e31.
4. López, R. and García, E. 2004, Recent trends on the molecular biology of pneumococcal capsules, lytic enzymes, and bacteriophage, *FEMS Microbiol. Rev.*, **28**, 553–580.
5. Yother, J. 2004, In: Tuomanen, E. I., Mitchell, T. J., Morrison, D. A. and Spratt, B. G. (eds.), *The Pneumococcus*, American Society for Microbiology Press: Washington, DC, pp. 30–48.
6. Aanensen, D. M., Mavroidi, A., Bentley, S. D., Reeves, P. R. and Spratt, B. G. 2007, Predicted functions and linkage specificities of the products of the *Streptococcus pneumoniae* capsular biosynthetic loci, *J. Bacteriol.*, **189**, 7856–7876.
7. Kadioglu, A., Weiser, J., Paton, J. C. and Andre, P. W. 2008, The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease, *Nat. Rev. Microbiol.*, **6**, 288–301.
8. Muñoz, R., Mollerach, M., López, R. and García, E. 1997, Molecular organization of the genes required for the synthesis of type 1 capsular polysaccharide of *Streptococcus pneumoniae*: formation of binary encapsulated pneumococci and identification of cryptic dTDP-rhamnose biosynthesis genes, *Mol. Microbiol.*, **25**, 79–92.
9. Magee, A. D. and Yother, J. 2001, Requirement for capsule in colonization by *Streptococcus pneumoniae*, *Infect. Immun.*, **69**, 3755–3761.
10. Haas, W., Kaushal, D., Sublett, J., Obert, C. and Tuomanen, E. I. 2005, Vancomycin stress response in a sensitive and a tolerant strain of *Streptococcus pneumoniae*, *J. Bacteriol.*, **187**, 8205–8210.
11. Rogers, P. D., Liu, T. T., Barker, K. S., *et al.* 2007, Gene expression profiling of the response of *Streptococcus pneumoniae* to penicillin, *J. Antimicrob. Chemother.*, **59**, 616–626.
12. Ogunniyi, A. D., Giammarinaro, P. and Paton, J. C. 2002, The genes encoding virulence-associated proteins and the capsule of *Streptococcus pneumoniae* are upregulated and differentially expressed *in vivo*, *Microbiology*, **148**, 2045–2053.
13. Orihuela, C. J., Radin, J. N., Sublett, J. E., Gao, G., Kaushal, D. and Tuomanen, E. I. 2004, Microarray analysis of pneumococcal gene expression during invasive disease, *Infect. Immun.*, **72**, 5582–5596.
14. LeMessurier, K. S., Ogunniyi, A. D. and Paton, J. C. 2006, Differential expression of key pneumococcal virulence genes *in vivo*, *Microbiology*, **152**, 305–311.
15. Mahdi, L. K., Ogunniyi, A. D., LeMessurier, K. S. and Paton, J. C. 2008, Pneumococcal virulence gene expression and host cytokine profiles during pathogenesis of invasive disease, *Infect. Immun.*, **76**, 646–657.
16. Hall-Stoodley, L., Nistico, L., Sambanthamoorthy, K., *et al.* 2008, Characterization of biofilm matrix, degradation by DNase treatment and evidence of capsule downregulation in *Streptococcus pneumoniae* clinical isolates, *BMC Microbiol.*, **8**, 173.
17. Hathaway, L. J., Bättig, P. and Mühlemann, K. 2007, *In vitro* expression of the first capsule gene of *Streptococcus pneumoniae*, *cpsA*, is associated with serotype-specific colonization prevalence and invasiveness, *Microbiology*, **153**, 2465–2471.
18. Oggioni, M. R. and Claverys, J. P. 1999, Repeated extragenic sequences in procaryotic genomes: a proposal for the origin and dynamics of the RUP element in *Streptococcus pneumoniae*, *Microbiology*, **145**, 2647–2653.
19. Hoskins, J., Alborn, W. E. Jr, Arnold, J., *et al.* 2001, Genome of the bacterium *Streptococcus pneumoniae* strain R6, *J. Bacteriol.*, **183**, 5709–5717.
20. Llull, D., García, E. and López, R. 2001, Tts, a processive β-glucosyltransferase of *Streptococcus pneumoniae*, directs the synthesis of the branched type 37 capsular polysaccharide in pneumococcus and other Gram-positive species, *J. Biol. Chem.*, **276**, 21053–21061.
21. Lanie, J. A., Ng, W. -L., Kazmierczak, K. M., *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.
22. Ko, K. S., Park, S., Oh, W. S., *et al.* 2006, Comparative analysis of growth-phase-dependent gene expression in virulent and avirulent *Streptococcus pneumoniae* using a high-density DNA microarray, *Mol. Cells*, **21**, 82–88.
23. Pérez-Martín, J., Rojo, F. and de Lorenzo, V. 1994, Promoters responsive to DNA bending: a common theme in prokaryotic gene expression, *Microbiol. Rev*, **58**, 268–290.
24. Olivares-Zavaleta, N., Jáuregui, R. and Merino, E. 2006, Genome analysis of *Escherichia coli* promoter sequences evidences that DNA static curvature plays a more important role in gene transcription than has previously been anticipated, *Genomics*, **87**, 329–337.
25. Vlahoviček, K., Kajan, L. and Pongor, S. 2003, DNA analysis servers: plot.it, bend.it, model.it and IS, *Nucleic Acids Res.*, **31**, 3686–3687.
26. Luo, P. and Morrison, D. A. 2003, Transient association of an alternative sigma factor, ComX, with RNA polymerase during the period of competence for genetic transformation in *Streptococcus pneumoniae*, *J. Bacteriol.*, **185**, 349–358.
27. Portmann, R., Poulsen, K. R., Wimmer, R. and Solioz, M. 2006, CopY-like copper inducible repressors are putative 'winged helix' proteins, *Biometals*, **19**, 61–70.
28. Reyes, A., Leiva, A., Cambiazo, V., Méndez, M. A. and González, M. 2006, *Cop*-like operon: structure and organization in species of the *lactobacillale* order, *Biol. Res.*, **39**, 87–93.
29. Nieto, C., Espinosa, M. and Puyet, A. 1997, The maltose/maltodextrin regulon of *Streptococcus pneumoniae*. Differential promoter regulation by the transcriptional repressor MalR, *J. Biol. Chem.*, **272**, 30860–30865.
30. Doroshchuk, N. A., Gel'fand, M. S. and Rodionov, D. A. 2006, Regulation of nitrogen metabolism in gram-positive bacteria, *Mol. Biol.* (Mosk.), **40**, 919–926.
31. Kloosterman, T. G., Hendriksen, W. T., Bijlsma, J. J. E., *et al.* 2006, Regulation of glutamine and glutamate metabolism by GlnR and GlnA in *Streptococcus pneumoniae*, *J. Biol. Chem.*, **281**, 25097–25109.

32. Ulijasz, A. T., Andes, D. R., Glasner, J. D. and Weisblum, B. 2004, Regulation of iron transport in *Streptococcus pneumoniae* by RitR, an orphan response regulator, *J. Bacteriol.*, **186**, 8123–8136.
33. Sierro, N., Makita, Y., de Hoon, M. and Nakai, K. 2008, DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information, *Nucleic Acids Res.*, **36**, D93–D96.
34. Federle, M. J. and Scott, J. R. 2002, Identification of binding sites for the group A streptococcal global regulator CovR, *Mol. Microbiol.*, **43**, 1161–1172.
35. Hammerschmidt, S., Wolff, S., Hocke, A., Rosseau, S., Müller, E. and Rohde, M. 2005, Illustration of pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells, *Infect. Immun.*, **73**, 4653–4667.
36. Giammarinaro, P. and Paton, J. C. 2002, Role of RegM, a homologue of the catabolite repressor protein CcpA, in the virulence of *Streptococcus pneumoniae*, *Infect. Immun.*, **70**, 5454–5461.
37. Iyer, R., Baliga, N. S. and Camilli, A. 2005, Catabolite control protein A (CcpA) contributes to virulence and regulation of sugar metabolism in *Streptococcus pneumoniae*, *J. Bacteriol.*, **187**, 8340–8349.
38. Philips, B., Meguer, J. -X., Redman, J. and Baker, E. 2003, Factors determining the appearance of glucose in upper and lower respiratory tract secretions, *Intensive Care Med.*, **29**, 2204–2210.
39. Baker, E. H., Clark, N., Brennan, A. L., *et al*. 2007, Hyperglycemia and cystic fibrosis alter respiratory fluid glucose concentrations estimated by breath condensate analysis, *J. Appl. Physiol.*, **102**, 1969–1975.
40. Llull, D., López, R. and García, E. 2001, Genetic bases and medical relevance of capsular polysaccharide biosynthesis in pathogenic streptococci, *Curr. Mol. Med.*, **1**, 475–491.
41. Shelburne, S. A. III, Keith, D., Horstmann, N., *et al*. 2008, A direct link between carbohydrate utilization and virulence in the major human pathogen group A *Streptococcus*, *Proc. Natl Acad. Sci. USA*, **105**, 1698–1703.
42. Graham, M. R., Smoot, L. M., Migliaccio, C. A. L., *et al*. 2002, Virulence control in group A *Streptococcus* by a two-component gene regulatory system: global expression profiling and *in vivo* infection modeling, *Proc. Natl Acad. Sci. USA*, **99**, 13855–13860.
43. Lamy, M. -C., Zouine, M., Fert, J., *et al*. 2004, CovS/CovR of group B streptococcus: a two-component global regulatory system involved in virulence, *Mol. Microbiol.*, **54**, 1250–1268.
44. Pan, X., Ge, J., Li, M., *et al*. 2009, The orphan response regulator CovR: a globally negative modulator of virulence in *Streptococcus suis* serotype 2, *J. Bacteriol.*, doi:10.1128/JB.01309–01308.
45. Hendriksen, W. T., Bootsma, H. J., Estevão, S., *et al*. 2008, CodY of *Streptococcus pneumoniae*: link between nutritional gene regulation and colonization, *J. Bacteriol.*, **190**, 590–601.
46. den Hengst, C. D., van Hijum, S. A. F. T., Geurts, J. M. W., Nauta, A., Kok, J. and Kuipers, O. P. 2005, The *Lactococcus lactis* CodY regulon: identification of a conserved *cis*-regulatory element, *J. Biol. Chem.*, **280**, 34332–34342.