molecular oral microbiology

molecular oral microbiology

INVITED REVIEW

Pneumococcal surface proteins: when the whole is greater than the sum of its parts

I. Pérez-Dorado*, S. Galan-Bartual and J.A. Hermoso

Department of Crystallography and Structural Biology, Instituto de Química-Física Rocasolano, CSIC, Madrid, Spain

Correspondence: Juan A. Hermoso, Department of Crystallography and Structural Biology, Instituto Química-Física Rocasolano, CSIC, Serrano 119, 28006-Madrid, Spain Tel.: +34 917459538; fax: +34 915642431; E-mail: xjuan@iqfr.csic.es
*Present address: Division of Structural Studies, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK.

Keywords: choline-binding proteins; LPXTG proteins; lipoproteins; non-classical surface proteins; *Streptococcus pneumoniae*; surface proteins; three-dimensional structure
Accepted 16 May 2012
DOI: 10.1111/j.2041-1014.2012.00655.x

SUMMARY

Surface-exposed proteins of pathogenic bacteria are considered as potential virulence factors through their direct contribution to host–pathogen interactions. Four families of surface proteins decorate the cell surface of the human pathogen *Streptococcus pneumoniae*. Besides lipoproteins and LPXTG proteins, also present in other gram-positive bacteria, the pneumococcus presents the choline-binding protein (CBP) family and the non-classical surface proteins (NCSPs). The CBPs present specific structural features that allow their anchorage to the cell envelope through non-covalent interaction with choline residues of lipoteichoic acid and teichoic acid. NCSP is an umbrella term for less characterized proteins displaying moonlighting functions on the pneumococcal surface that lack a leader peptide and membrane-anchor motif. Considering the unceasing evolution of microbial species under the selective pressure of antibiotic use, detailed understanding of the interaction between pathogen and the host cells is required for the development of novel therapeutic strategies to combat pneumococcal infections. This article reviews recent progress in the investigation of the three-dimensional structures of surface-exposed pneumococcal proteins. The modular nature of some of them produces a great versatility and sophistication of the virulence functions that, in most cases, cannot be deduced by the structural analysis of the isolated modules.

INTRODUCTION

The pneumococcus *Streptococcus pneumoniae* – a normal component of the microflora of the human upper respiratory tract and an opportunistic pathogen – can, under appropriate conditions, cause serious and life-threatening infections such as pneumonia, meningitis and septicemia (Cartwright, 2002). The disease burden is high both in developed and developing countries, and the high-risk groups include children, elderly persons and immune-compromised patients. Approximately one million children under 5 years of age die from pneumococcal infections every year. In spite of the availability of a large number of antibiotics, mortality and morbidity due to *S. pneumoniae* infections remain high. A prerequisite for developing new control strategies is the detailed understanding of the interaction between pneumococcus and the host cells. This interaction is mainly mediated by the external components of this bacterium: the capsular polysaccharide, the bacterial cell wall and the proteins anchored in the cell wall.

The capsular polysaccharide, which covers the pneumococcal surface, has been recognized as a *sine qua non* of virulence (Austrian, 1981) and is the

target of existing pneumococcal vaccines. To date, 94 serologically distinct capsules have been recognized (Henrichsen, 1995) and despite the wide diversity in their sugar compositions and linkages, all perform the same primary function of reducing opsonophagocytosis by limiting the access of phagocytic receptors to complement bound to the *S. pneumoniae* cell wall (Yother, 2004). The pneumococcal cell wall contains a multi-layered murein (peptidoglycan) with a covalently attached teichoic acid (TA), and a membrane-bound lipoteichoic acid (LTA). The peptides in the murein are cross-linked either directly or via an interpeptide bridge. The repeating units of TA and LTA have an identical chemical structure, and so far this is a unique property of pneumococci (Vollmer, 2007). The structure of the repeating unit in pneumococcal TA and LTA are of unusual chemical complexity and present choline, an amino alcohol that is abundant in eukaryotes and only rarely found in bacteria (Garcia *et al.*, 1998).

Three main groups of surface proteins have been identified in *S. pneumoniae*: (i) around 50 lipoproteins, (ii) up to 18 LPXTG consensus sequence-carrying proteins that are covalently linked via sortase to the cell wall peptidoglycan, and (iii) up to 16 choline-binding proteins (CBPs) (Fig. 1). In addition to the three main groups of surface proteins, the cell envelope of pneumococci is decorated with another cluster of proteins that lack classic leader peptide and membrane-anchoring motifs. These proteins are termed non-classical surface proteins (NCSPs) and recent results indicate that they could also play a relevant role in subverting the physiological function of host-derived proteins (Bergmann & Hammerschmidt, 2006).

MODULARITY AND FUNCTION OF PNEUMOCOCCAL SURFACE PROTEINS

LPXTG proteins

Surface proteins carrying an LPXTG motif are covalently anchored to the bacterial cell wall. To reach their final destination, these surface proteins require both an N-terminal signal peptide and a C-terminal cell-wall sorting signal (CWS) (Bergmann & Hammerschmidt, 2006; Dramsi *et al.*, 2008; Lofling *et al.*, 2011). The N-terminal signal peptide promotes protein translocation across the bacterial membrane. The CWS consists of three indispensable motifs: an LPXTG sequence (where X is any amino acid), followed by a hydrophobic stretch of amino acids, and a short positively charged tail (Dramsi *et al.*, 2008). After translocation through the membrane, the protein harboring the CWS is retained within the membrane via its C-terminal hydrophobic domain and positively charged tail. In this stage, a membrane-associated enzyme with transpeptidase activity, a class A sortase or SrtA, covalently links the surface protein to the membrane (Mazmanian *et al.*, 2000; Dramsi *et al.*, 2008).

The availability of genomic sequence data for pneumococcal strains has facilitated the identification of 18 and 15 proteins with peptidoglycan anchor LPXTG-like motif in TIGR4 and R6 strains, respectively (Table 1) (Tettelin *et al.*, 2001; Frolet *et al.*, 2010; Lofling *et al.*, 2011). The available structural and topological information show that pneumococcal LPXTG proteins are mostly modular proteins with enzymatic or adhesin functions. Adhesins help the pathogen to attach to host cells and tissues by recognizing host saccharide or protein molecules in a step that is critical for successfully establishing an infection (Lofling *et al.*, 2011). Proteins with enzymatic activity are composed of single or multiple domains involved in recognizing host molecules that localize the enzyme to its substrate and module(s) with catalytic activities. There is a total of six glycosyl hydrolases (StrH, NanA, BgaA, EndoD, SpuA and Eng), one enzyme with polysaccharide lyase activity (SpnHL), and four proteases (ZmpA, ZmpB, ZmpC and PtrA). Their substrates are glycoproteins, which are fundamental to most biological processes, including the human immune system; glycosaminoglycans, which are long polysaccharides that form part of the extracellular matrix; and intracellular glycogen, which is a molecule that serves as long-term energy storage (van Bueren *et al.*, 2007; Caines *et al.*, 2008; Lammerts van Bueren *et al.*, 2011; Garbe & Collin, 2012). Hydrolysis of these substrates can therefore have dramatic effects such as compromising the host immune system, promoting adherence and migration events by revealing receptors for adherence, or creating a source of nutrients for the pathogen (Garbe & Collin, 2012).

Lipoproteins

The bacterial lipoproteins (LPs) are modified at their N-terminal with the addition of an *N*-acyl diacylglyceryl group that anchors them to the membrane. The biosynthetic pathway is ubiquitous in bacteria, reflecting

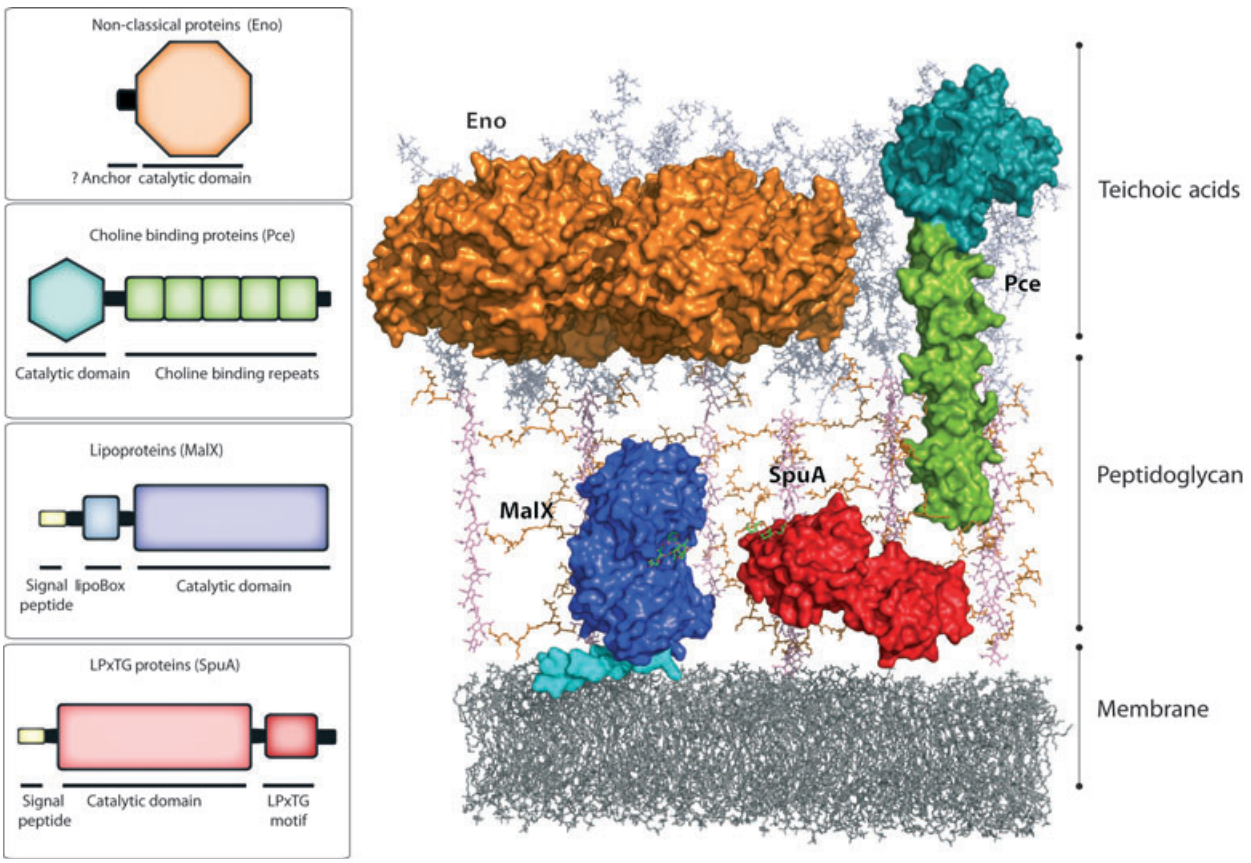

Figure 1 Schematic model of different classes of surface-exposed proteins and their location on pneumococcal cell wall. Left, schematic representation of modular arrangement in pneumococcal surface proteins. Right, over cytoplasmic membrane, multi-layered murein (peptidoglycan) is located with covalently attached teichoic acids (TA), and membrane-bound lipoteichoic acids (LTA). The repeating units of TA and LTA of *Streptococcus pneumoniae* have an identical chemical structure. While lipoproteins (e.g. MalX) remain anchored to membrane, the LPXTG proteins (e.g. SpuA) are covalently bound to peptidoglycan. The pneumococcal cell wall consists of a phospholipid membrane (LM), peptidoglycan (PG), and teichoic and lipoteichoic acids. Choline-binding proteins (e.g. Pce) anchor to the cell envelope through non-covalent interactions with choline residues of teichoic acids. Non-classical surface proteins (e.g. Eno) lack classic leader peptide and membrane-anchoring motifs.

the essential role of these LPs in bacterial survival. The LPs contain an exportation signal peptide followed by a common structural feature, the LipoBox, which is recognized before lipid modification. LipoBox contains a four-amino-acid consensus sequence of: [LVI] [ASTVI] [GAS] [C]. This cysteine is the residue that is modified with a diacylglyceryl group to allow membrane anchoring (Oudega *et al.*, 1993).

Genome-sequencing projects in combination with informatics analyses predict up to 46 LPs out of more than 2000 coding genes reported in *S. pneumoniae* genomes (Table 2) (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001; Babu *et al.*, 2006). Interestingly, most of the LPs seem to have a role in bacterial fitness. In the TIGR4 genome, among the 46 predicted LPs (Tettelin *et al.*, 2001; Babu *et al.*, 2006) 26 are ABC transporters (ATP-binding cassette transporters). These ABC components, together with six other LPs involved in substrate binding and transformation, are essential for maintaining the fitness of *S. pneumoniae*, an obligate commensal that is forced to scavenge host nutrients because of its incomplete biosynthetic pathways (Bergmann & Hammerschmidt, 2006). Six LPs are involved in processes of protein folding or activation of cell surface molecules, such as sortase (StrC-3), which was shown to be specifically involved in pilus formation (Manzano *et al.*, 2008). Six LPs remain uncharacterized without known function, and two defective LPs without catalytic domains are also predicted (Babu *et al.*, 2006).

Table 1 Identified LPXTG proteins in *Streptococcus pneumoniae*

Protein name	UniProtKB-locus	Function/bibliography	PDB entry codes
StrH	P49610-SP_0057-StrH (1312 aa) Q8DRL6-spr0057-StrH (1312 aa)	β-*N*-acetylglucosaminidase. Hydrolysis of glycoconjugates of IgA1, human lactoferrin, human secretory component, and a1- acid glycoprotein (Clarke *et al.*, 1995; Jiang *et al.*, 2011; Pluvinage *et al.*, 2011; Garbe & Collin, 2012)	2YL5, 2YL6, 2YL8, 2YL9, 2YLA, 2YLL, 3RPM
ZmpC	Q97T80-SP_0071-ZmpC-MMP-9 protease (1856 aa)	Zinc metalloproteinase. Cleaves human matrix metalloproteinase 9 (Tettelin *et al.*, 2001; Oggioni *et al.*, 2003)	
PavB	Q97T70-SP_0082-PavB (857 aa) Q8DRK2-spr0075 (1161 aa)	Adhesin involved in nasopharyngeal colonization (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001; Jensch *et al.*, 2010; Paterson & Orihuela, 2010)	
SpuA	Q97SQ7-SP_0268 (1280 aa) Q8DRA6-spr0247-PulA (1256 aa)	Hydrolysis of intracellular glycogen (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001; van Bueren *et al.*, 2007; Frolet *et al.*, 2010; Lammerts van Bueren *et al.*, 2011)	2J44, 2YA0, 2YA1, 2YA2
SpnHL	Q54873-SP_0314 (1066 aa) Q8CWU3-spr0286-HysA (1078 aa)	Polysaccharide lyase. Degradation of connective tissue (Berry *et al.*, 1994; Jedrzejas *et al.*, 1998; Mitchell, 2000; Ponnuraj & Jedrzejas, 2000; Botzki *et al.*, 2004; Rigden *et al.*, 2006)	1C82, 1EGU, 1F1S, 1BRP, 1N7N, 1N70, 1N7Q, 1N7R, 1N7P, 1OJM, 1OJN, 1OJO, 1OJP, 1F9G, 1LOH, 1W3Y, 1BRP, 2BRV, 2BRW
Eng	Q2MGH6-SP_0368-SpGH101 (1767 aa) Q8DR60-spr0328-SpGH101 (1767 aa)	Endo-α-*N*-acetylgalactosaminidase. Hydrolyzes host glycoproteins (the T-antigen or Core-1-Type O-glycan). Possible adhesin (Umemoto *et al.*, 1977; Caines *et al.*, 2008; Willis *et al.*, 2009; Frolet *et al.*, 2010)	3ECQ
Endo D	Q97S90-SP_0498-EndoD (1646 aa) Q8CZ52-spr0440 (1659 aa)	Endo-β *N*-acetylglucosaminidase. Hydrolysis of glycoconjugates of transferrin, fetuin, and IgG (Tettelin *et al.*, 2001) (Hoskins *et al.*, 2001; Muramatsu *et al.*, 2001; Abbott *et al.*, 2009; Jiang *et al.*, 2011; Garbe & Collin, 2012)	2W91[1], 2W92[1], 2XQX, 3GDB
PrtA	Q97RY6-SP_0641-PrtA (2140 aa) Q8DQP7-spr0561 (2144 aa)	Cell wall associated serine proteinase. Proteolysis of human lactoferrin. Possible adhesin (Bethe *et al.*, 2001; Hoskins *et al.*, 2001; Frolet *et al.*, 2010; Mirza *et al.*, 2011)	
BgaA	Q9KGU0-SP_0648-BgaA (2233 aa) Q8DQP4-spr0565 (2228 aa)	β-Galactosidase. Hydrolysis of glycoconjugates of IgA1, human lactoferrin, human secretory component, and a1- acid glycoprotein (Zahner & Hakenbeck, 2000; Hoskins *et al.*, 2001; Garbe & Collin, 2012)	
ZmpB	Q9L7Q2-SP_0664-ZmpB (1906 aa) Q8DQN5-spr0581 (1876 aa)	Zinc metalloprotease. Possible adhesin (Hoskins *et al.*, 2001; Frolet *et al.*, 2010)	
ZmpA	Q97QP7-SP_1154-ZmpA-IgA1 protease (2004 aa) Q59947-spr1042-IgA1 protease (1963 aa)	Inmunoglobulin A1 protease (Wani *et al.*, 1996; Tettelin *et al.*, 2001; Mistry & Stockley, 2006)	
MucB	Q97PV0-SP_1492-MucB (202 aa) Q8CYK3-spr1345 (202 aa)	Mucin-binding domain (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001; Du *et al.*, 2011)	3NZ3
PclA	Q8CYI8-spr1403-PclA (2551 aa)	Adhesin (Hoskins *et al.*, 2001; Paterson *et al.*, 2008)	
NanA	P62575-SP_0463-nanA (1035 aa) P62576-spr1536-NanA (1035 aa)	Neuraminidase or sialidase. Hydrolyses of glycoconjugates of IgA1, human lactoferrin, human secretory component, and a1- acid glycoprotein. Possible adhesin (Camara *et al.*, 1994; Hoskins *et al.*, 2001; Xu *et al.*, 2008; Hsiao *et al.*, 2009; Frolet *et al.*, 2010; Gut *et al.*, 2011; Garbe & Collin, 2012)	2VVZ, 2YA4, 2YA5, 2YA6, 2YA7, 2YA8 2W20, 3H71, 3H72, 3H73

Table 1 (Continued)

Protein name	UniProtKB-locus	Function/bibliography	PDB entry codes
PsrP	Q97P71-SP_1772-PspR (4776 aa)	Adhesion and invasion of epithelial cells. Antiphagocytic activity (Tettelin *et al.*, 2001; Shivshankar *et al.*, 2009)	
PfbA	Q97P11-SP_1833 (708 aa) Q8CYC9-spr1652-PfbA (719 aa)	Plasmin and fibronectin-binding protein A. Adhesin (Tettelin *et al.*, 2001; Yamaguchi *et al.*, 2008)	
Spr1806	Q97NM8-SP_1992 (221 aa) Q8CY93-spr1806 (221 aa)	Possible adhesin (Hoskins *et al.*, 2001; Frolet *et al.*, 2010)	

[1]PDB entries 2W91 and 2W92 referrer to sequence with UniProtKB entry code Q93HW0 that also corresponds to an EndoD sequence from TIGR4 strain. Both EndoD sequences from TIGR4 strain (Q93HW0 and Q97S90) exhibit 98.2% of sequence identity. Not ordered locus name was found for Q93HW0 entry.

Non-classical surface proteins

A subset of the pneumococcal surface exposed proteins lacks conventional anchoring or secretory signals. These proteins, the NCSPs (Table 3), are described primarily as cytoplasmic proteins with intracellular roles that are not involved in host–pathogen interactions. However, once they are located by unknown mechanisms on the cell surface, NCSPs display moonlighting functions and frequently act as adhesins. Binding of NCSP to host molecules, like fibronectin (PavA) or plasminogen (Enolase), promotes pneumococcal invasion and the spread of the infection (Holmes *et al.*, 2001; Ehinger *et al.*, 2004).

To date, only six NCSPs have been identified (Table 3). However, the lack of secretory signals or cell-wall anchoring domains makes difficult for bioinformatics tools to detect these NCSPs. It is, therefore, believed that their number is severely underestimated. Future research and technological advances in direct surface protein detection will uncover new NCSPs displaying different moonlighting roles.

Choline-binding proteins

A unique characteristic of *S. pneumoniae* is its nutritional requirement for choline (Tomasz, 1967), which is taken up from the growth medium (Bean & Tomasz, 1977) and incorporated into the repeating units of TA and LTA (Behr *et al.*, 1992). Replacement of choline in the growth medium with ethanolamine, although satisfying the nutritional requirement, causes a variety of functional and morphological alterations. Phosphorylcholine residues of pneumococcal TA and LTA are essential for the optimal activity of murein hydrolases (Lopez & Garcia, 2004), are involved in many physiological functions of *S. pneumoniae* (Lopez *et al.*, 1982; López *et al.*, 2004), and serve as an anchor for surface-located CBPs (Swiatlo *et al.*, 2004).

Choline-binding protein family members share a modular organization (Fig. 2) consisting of a biologically active module and a choline-binding module, which anchors these proteins to the cell envelope through a non-covalent interaction with choline residues of teichoic acids. The choline-binding module, placed at the N-terminus or C-terminus, is made up of homologous repeats of about 20 amino-acid residues (Garcia *et al.*, 1988, 1998), the CBP family includes some virulence factors involved in cellular adhesion and colonization (see Table 4). Among the approximately 15 CBP members (depending on the strain) the X-ray crystal structure of the complete protein (carrying both the catalytic and the choline-binding modules) is available only for three of them (Pce, CbpF and LytC). Three-dimensional structures of different domains have been reported for four other CBPs.

THREE-DIMENSIONAL STRUCTURE OF LPXTG PROTEINS

Glycosyl hydrolases StrH, NanA, EndoD, SpGH101 and SpuA

Structural information has been reported for five LPXTG proteins with glycosyl hydrolase activities. They are StrH (exo-β-*N*-acetylhexosaminidase), NanA (neuramidase or sialase), EndoD (endo-β-D-glucosaminidase), SpGH101 (endo-α-*N*-acetylgalactosaminidase) and SpuA (pullulanase), which belong to

Table 2 Identified lipoproteins in *Streptococcus pneumoniae*

Protein name	UniProtKB-locus	Function/bibliography	PDB entry codes
SP_0092	Q97T63-SP_0092 (491 aa) Q8DRJ6-spr0083-ABC-SBP (514 aa)	ABC transporter (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
SP_0112	Q97T43-SP_0112 (268 aa) Q8DRI6-spr0101-ABC-SBP (265 aa)	Amino acid ABC transporter (Tettelin *et al.*, 2001)	
SP_0148	Q97T12-SP_0148 (276 aa) Q8DRG2-spr0146-ABC-SBP (276 aa)	ABC transporter substrate-binding protein-amino acid transport (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
SP_0149	Q97T11-SP_0149 (284 aa) Q8DRG1-spr0147-ABC-SBP (284 aa)	Lipoprotein (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
SP_0191	Q97SX2-SP_0191 (189 aa)	Putative uncharacterized protein (Tettelin *et al.*, 2001)	
SP_0198	Q97SW6-SP_0198 (152 aa)	Putative D-stereospecific aminopeptidase (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	3GE2
SP_0366	P35592-SP_0366-aliA-exp1-plpA (661 aa)	ABC transport system. Oligopeptide-binding protein AliA (Tettelin *et al.*, 2001; Babu *et al.*, 2006)	
SrtC-3; SrtD	Q97SB7-SP_0468-srtD (283 aa)	Sortase (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001; Manzano *et al.*, 2008; Moschioni *et al.*, 2008)	2W1K
SP_0620	Q97S05-SP_0620 (266 aa) Q8CWT0-spr0545-ABC-SBP (271 aa)	Amino acid ABC transporter (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
SP_0629	Q97RZ7-SP_0629 (238 aa) Q8DQQ1-spr0554 (238 aa)	Putative uncharacterized protein (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
SP_0659	Q97RX4-SP_0659 (188 aa) Q8DQN9-spr0576 (197 aa)	Thioredoxin family protein (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
LivJ	Q97RQ0-SP_0749-livJ (386 aa)	ABC transport system (Tettelin *et al.*, 2001)	
PpiA	Q97RN2-SP_0771 (267 aa) Q8DQG5-spr0679-ppiA (267 aa)	Peptidyl-propyl cis-tras isomerase, cyclophilin-type (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
SP_0845	Q97RH0-SP_0845 (350 aa) Q8DQC2-spr0747 (374 aa)	Substrate binding lipoprotein (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
SP_0859	Q97RG2-SP_0859 (307 aa)	DUF979 superfamily. Unknown function	
SP_0899	Q97RC4-SP_0899 (290 aa) Q8CYX1-spr0799 (290 aa)	Putative uncharacterized protein (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
PrsA	Q97R51-SP_0981 (313 aa) Q8DQ24-spr0884-prsA-ppmA (313 aa)	Protease maturation protein, putative. Foldase protein prsA (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
SP_1000	Q97R36-SP_1000 (185 aa) Q8DQ10-spr0904 (191 aa)	Thioredoxin family protein (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
AdcII	Q97R34-SP_1002 (305 aa) Q8DQ09-spr0906-lmb (311 aa)	Adhesion lipoprotein. Putative laminin-binding protein. Glutamine-binding protein (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001; Loisel *et al.*, 2008)	3CX3
SP_1032	Q97R09-SP_1032 (341aa) Q8DPY6-spr0934 (341aa)	Iron-compound ABC transporter ion (Tettelin *et al.*, 2001)	
NanE	Q97Q95-SP_1330-nanE (233 aa)	*N*-acetylmannosamine-6-phosphate epimerase (Tettelin *et al.*, 2001)	
GlnH	Q97Q37-SP_1394 (271 aa) Q8DPB7-spr1251-glnH (271 aa)	Amino acid ABC transporter (Tettelin *et al.*, 2001)	
PstS 1	Q97Q31-SP_1400-pstS1 (292 aa) Q8DPB1-spr1257-pstS1 (292 aa)	Phosphate ABC transporter, phosphate-binding protein, putative Phosphate-binding protein pstS 1 (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
AatB	Q97PU3-SP_1500-aatB (278 aa) Q8DP50-spr1353-glnH (278 aa)	Amino acid ABC transporter (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
AliB	P0A4G0-SP_1527-aliB (652 aa) P0A4G1-spr1382-aliB (652 aa)	Oligopeptide ABC transporter (Alloing *et al.*, 1994; Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
PsaA	P0A4G2-SP_1650-psaA (309 aa) P0A4G3-spr1494-psaA (309 aa)	Manganeso ABC transporter, manganeso-binding adhesión lipoprotein or Pneumococcal surface adhesin A (Berry & Paton, 1996; Dintilhac *et al.*, 1997; Lawrence *et al.*, 1998; Novak *et al.*, 1998)	1PSZ, 3ZTT

Table 2 (Continued)

Protein name	UniProtKB-locus	Function/bibliography	PDB entry codes
SP_1683	Q97PE6-SP_1683 (442 aa) Q8DNU8-spr1527-ABC-SBP (442 aa)	Sugar ABC transporter (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
nanE2	P65520-SP_1685-nanE2 (232 aa)	*N*-acetylmannosidase-6-phosphate epimerase (NanE) (Tettelin *et al.*, 2001)	
SP_1690	Q97PE1-SP_1690 (445 aa) Q8DNU2-spr1534-ABC-SBP (445 aa)	ABC transporter (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
SP_1796	Q97P48-SP_1796 (538 aa)	ABC transporter (Tettelin *et al.*, 2001)	
SP_1826	Q97P18-SP_1826 (355 aa)	ABC transporter (Tettelin *et al.*, 2001)	
SP_1870	Q97NY1-SP_1870 (318 aa) Q8DNJ4-spr1685-fatC (318 aa)	Iron-compound ABC transporter, pernease protein (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
SP_1872	Q97NX9-SP_1872 (321 aa) Q8DNJ2-spr1687-fatB (321 aa)	Iron-compound ABC transporter, iron-compound binding protein (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
AmiA	P18791-SP_1891-amiA (659 aa) Q8DNI1-spr1707-amiA (659 aa)	Oligopeptide ABC transporter (Alloing *et al.*, 1990, 1994; Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
RafE	Q97NW2-SP_1897-rafE (419 aa) Q8DNH8-spr1712-msmE (419 aa)	Sugar transporter, sugar-binding protein (Tettelin *et al.*, 2001)	2HEU, 2HFB, 2HQ0, 2I58
SP_1916	Q97NU6-SP_1916 (167 aa)	PAP2 family protein (Babu *et al.*, 2006)	
SP_1945	Q97NS2-SP_1945 (206 aa)	Hypothetical protein (Babu *et al.*, 2006)	
OxaA2	Q97NP5-SP_1975-oxaA2 (308 aa) Q8DNE1-spr1790-oxaA2 (308 aa)	SpoIIIJ family protein (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
OxaA1	Q97NI6-SP_2041-oxaA1 (274 aa) Q8DN93-spr1852-oxaA1 (276 aa)	SpoIIIJ family protein (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
PstS	P0C2M5-SP_2084-pstS 2 (291 aa) Q8DN64-spr1895-pstS2 (291 aa)	Phosphate ABC transporter (Novak *et al.*, 1999; Orihuela *et al.*, 2001)	
MalX	P59213-SP_2108-malX (423 aa) P59214-spr1918-malX (423 aa)	Maltose/maltodextrin ABC transporter, maltose/maltodextrin-binding protein (Puyet & Espinosa, 1993; Abbott *et al.*, 2010)	2XD2, 2XD3
AdcA	O05703-SP_2169-adcA (501 aa) Q8CWN2-spr1975-adcA (501 aa)	Zinc ABC transporter (Dintilhac *et al.*, 1997)	
SP_2197	Q97N69-SP_2197 (335 aa) Q8DMZ7-spr2003 (335 aa)	ABC transporter (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	

Table 3 Identified non-classical surface proteins in *Streptococcus pneumoniae*

Protein name	UniProtKB-locus	Function/bibliography	PDB entry codes
PavA	Q97R64-SP_0966-pavA (551 aa) Q8DQ36-spr0868-flpA (560 aa)	Pneumococcal adherente and virulence factor A. Adhesin (Holmes *et al.*, 2001; Bergmann & Hammerschmidt, 2006; Nobbs *et al.*, 2009)	
Eno	Q97QS2–SP_1128-eno (434 aa) Q8DPS0-spr1036- eno (434aa)	Enolase. Binds plasmin(ogen) (Bergmann *et al.*, 2001; Ehinger *et al.*, 2004; Bergmann & Hammerschmidt, 2006; Nobbs *et al.*, 2009)	1W6T
GAPDH	Q97NL1-SP_2012-gap-gapdh (335 aa) Q8CWN6- spr1825– apA (359 aa)	Glyceraldehyde 3-phosphate dehydrogenase (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
6PGD	Q97SI6-SP_0375-gnd (474 aa) Q8DR54-spr0335-gnd (474 aa)	6-phosphogluconate dehydrogenase. Putative adhesin. Binds epithelial cells (Daniely *et al.*, 2006; Nobbs *et al.*, 2009)	
HtrA	Q97N37-SP_2239 (393 aa) Q8DMW2-spr2045-sphtra (397 aa)	Serine protease involved in nasopharyngeal colonization (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001; Sebert *et al.*, 2002; Nobbs *et al.*, 2009)	
PGK	Q97S89-SP_0499 (398aa)	Phosphoglycerate kinase (Bernardo-Garcia *et al.*, 2011)	

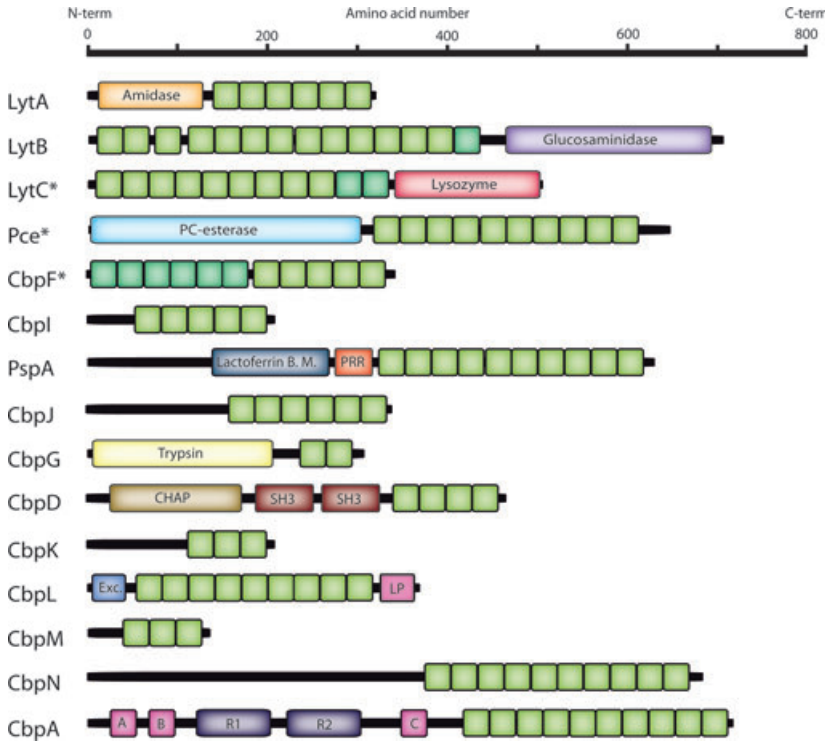

Figure 2 The modular nature of choline-binding proteins (CBP). CBP family members share an N-terminal or C-terminal choline-binding module made up of homologous repeats of about 20 amino acid residues (green boxes), which facilitates the anchorage of these proteins to the cell envelope. In addition, CBPs present a catalytic or functional module displaying different activities. The asterisks indicate proteins for which the three-dimensional structure for the complete protein exists.

the families of glycosyl hydrolases 20, 33, 85, 101 and 13, respectively.

StrH and NanA sequentially hydrolyse *N*-glycan structures present in several molecules of host defense in concert with another glycosyl hydrolase, BgaA (a β-galactosidase). Substrates of StrH, NanA and BgaA identified so far are: IgA1, human lactoferrin, human secretory component and α1-acid glycoprotein (Garbe & Collin, 2012). StrH is a proved virulence factor in multiple animal model screens of pneumonia and otitis media (Pluvinage *et al.*, 2011). NanA has been shown to play a critical role in pneumococcal endothelial brain invasion through its lectin domain, and its catalytic activity is involved in biofilm formation and colonization (Gut *et al.*, 2011). Moreover, the combined action of SrtH, NanA and BgaA has been recently shown to play an important role in evasion from opsonophagocytic killing mediated by neutrophils (Dalia *et al.*, 2010).

StrH is a modular enzyme formed by two tandem N-terminal GH20 catalytic domains (GH20A and GH20B) and two C-terminal G5-domains. GH20 domains hydrolyse the β-(1,2) bond between *N*-acetyl-D-glucosamine (GlcNAc) and mannose (Man) from various glycoconjugates present in the α-(1,3) or α-(1,6) arms of glycans (Clarke *et al.*, 1995). Each GH20 domain consists of an $(\alpha/\beta)_8$ triose-phosphate isomeraseb (TIM) barrel followed by a three-helix bundle domain (Fig. 3A) (Jiang *et al.*, 2011; Pluvinage *et al.*, 2011). Structural analysis of GH20A and GH20B domains in complex with substrate molecules shows that they bind GlcNAc-Man disaccharide in the same manner, but there are remarkable differences in glycan-binding sites beyond position +1. These differences provide them with distinctive substrate specificities: GH20A domain recognizes the α-(1,3) arms of an *N*-glycan, whereas GH20B recognizes bisected glycans (Clarke *et al.*, 1995; Pluvinage *et al.*, 2011). Conservation of both GH20 domains with different specificities suggests that one domain could be providing the substrate for the other, or that combination of both specificities could provide the required working pair to hydrolyse the necessary substrate range (Pluvinage *et al.*, 2011).

Table 4 Identified choline-binding proteins in *Streptococcus pneumoniae*

Protein name	UniProtKB-locus	Function/bibliography	PDB entry codes
CbpI	Q9KGY8-SP_0069-CbpI (211 aa)	Putative adhesin (Gosink *et al.*, 2000; Tettelin *et al.*, 2001; Frolet *et al.*, 2010)	
PspA	Q97T39-SP_0117-PspA (744 aa) Q8DRI0-spr0121-PspA (619 aa)	Virulence factor. It inhibits complement activation (Tu *et al.*, 1999; Mitchell, 2000; Hoskins *et al.*, 2001; Tettelin *et al.*, 2001; Senkovich *et al.*, 2007)	2PMS[1]
CpbF	Q97SI4-SP_0377-CpCp-CbpC (340 aa) Q8DR52-spr0337-CbpF (338 aa)	Regulatory function for pneumococcal autolysis by inhibiting the autolytic LytC muramidase (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001; Molina *et al.*, 2009a,b)	2V04, 2V05, 2VYU, 2X8M, 2X8O, 2X8P
CbpJ	Q9KGY7-SP_0378-CbpJ (332 aa)	Putative adhesin (Gosink *et al.*, 2000; Tettelin *et al.*, 2001; Frolet *et al.*, 2010)	
CbpG	Q97SH5-SP_0390-CbpG (285 aa)	Unknown function (Tettelin *et al.*, 2001)	
CbpK	Q9KGY9-SP_0391-CbpF (340 aa) Q8DR39-spr0351-PcpC (294 aa)	Putative adhesion (Gosink *et al.*, 2000; Hoskins *et al.*, 2001; Tettelin *et al.*, 2001)	
CbpL	Q97RW9-SP_0667 (332 aa) Q8CZ16-spr0583 (329 aa)	Putative adhesion (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001; Frolet *et al.*, 2010)	
CbpE	Q9KGZ1-SP_0930-CbpE (627 aa) Q8DQ62-spr0831-CbpE-LytD-Pce (627 aa)	Critical function in pneumococcal adherence and invasiveness. Adhesin (Gosink *et al.*, 2000; Hoskins *et al.*, 2001; Tettelin *et al.*, 2001; Hermoso *et al.*, 2005; Frolet *et al.*, 2010)	2BIB, 1WRA
LytB	P59205-SP_0965-LytB (658 aa) P59206-spr0867-LytB (702 aa)	*N*-acetylglucosaminidase activity involved in separation of daughter cells (Garcia *et al.*, 1999; Hoskins *et al.*, 2001; Tettelin *et al.*, 2001; De Las Rivas *et al.*, 2002)	
PspC-Like	SP_1417	PspC-related protein (degenerate) (Tettelin *et al.*, 2001)	
LytC	Q2MGF6-SP_1573-LytC (490 aa) Q8DP07-spr1431-LytC (501 aa)	Muramidase activity. Role in colonization. Virulence factor involved in fratricide (competence-programmed mechanism) (Garcia *et al.*, 1999; Gosink *et al.*, 2000; Claverys & Havarstein, 2007; Perez-Dorado *et al.*, 2010)	2WW5, 2WWC, 2WWD
LytA	P06653-SP_1937-LytA (318 aa) Q7ZAK4-spr1754-LytA (318 aa)	*N*-acetylmuramoyl-l-alanine amidase. Virulence. Release of Ply and inflamatory PG and TA from bacterial cell wall. Cell wall grow and turnover. Participates in fratricide (Garcia *et al.*, 1985; Fernandez-Tornero *et al.*, 2001, 2002; Claverys & Havarstein, 2007; Kadioglu *et al.*, 2008)	2BML, 1GVM, 1H8G, 1HCX
CbpA	Q97N74-SP_2190-PspC-SpsA-CbpA Q8DN05-spr1995-CbpA-PcpA	Major adhesin of streptococcus pneumonia. Immunoglobulin A inactivation (Mitchell, 2000; Luo *et al.*, 2005; Hammerschmidt *et al.*, 2007; Kadioglu *et al.*, 2008; Agarwal *et al.*, 2010; Frolet *et al.*, 2010)	1W9R
CbpM	Q8DP99-spr1274 (129aa)	Putative adhesin (Hoskins *et al.*, 2001; Zhang *et al.*, 2009; Frolet *et al.*, 2010)	3HIA
PcpA	Q97NB5-SP_2136-Pcpa (621 aa) Q8DN38-spr1945-Pcpa (690 aa)	Putative adhesin (Sanchez-Beato *et al.*, 1998; Hakenbeck *et al.*, 2009)	
CbpD	Q9KGZ2-SP_2201-CbpD (448 aa) Q8DMZ4-spr2006-CbpD (448 aa)	Adherence and nasopharingeal colonization (Gosink *et al.*, 2000)	

[1]PDB entry 2PMS referrers to UnitProtKB entry Q54972 corresponding to PspA from pneumococcal Rx1 strain (Senkovich *et al.*, 2007), which exhibits 100% sequence identity with PspA from R6 strain (UniProtKB entry Q8DRI0).

NanA is formed by an N-terminal laminin G-like lectin-binding domain and a C-terminal catalytic domain that is responsible for cleaving α-(2,3)-, α-(2,6)-, α-(2,8)-linked sialic acid in *N*-glycans (Gut *et al.*, 2011). The structure of the NanA catalytic domain has been reported alone, and in complex with substrate

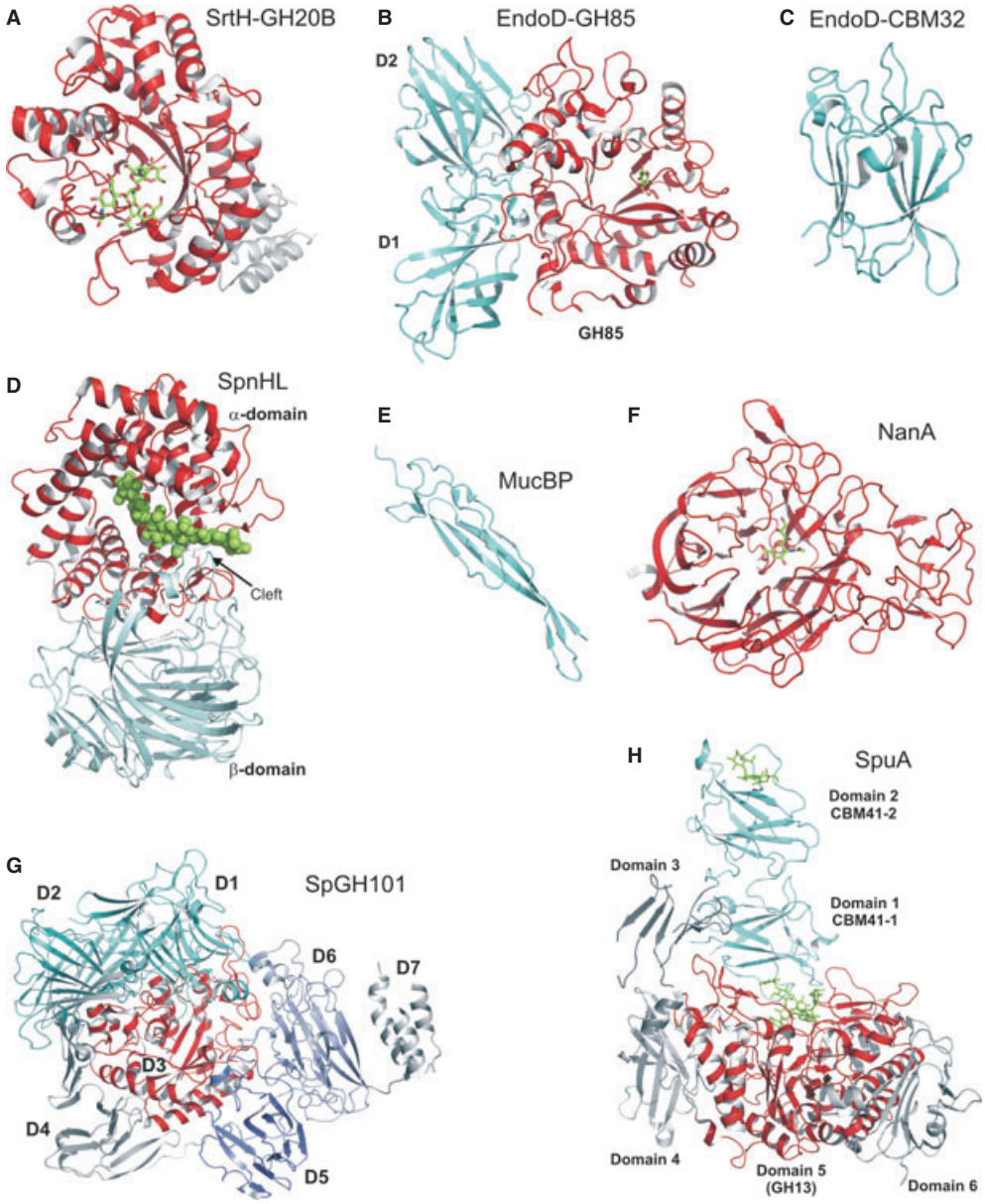

Figure 3 Three-dimensional structures of pneumococcal LPXTG proteins. (A) StrH(E361Q)_GH20 mutant in complex with NGA2B [Protein Data Bank (PDB) entry 2YLA]. (B) Catalytic domain of EndoD in complex with GlcNAc-thiazoline (PDB entry 2W92). (C) EndoD-CBM32 domain (PDB entry 2XQX). (D) SpnHL in complex with hyaluronic disaccharide (PDB entry 1C82). (E) Mucin-binding domain of MucBP (PDB entry 3NZ3). (F) Catalytic domain of NanA in complex with the inhibitor DANA (PDB entry 2VVZ). (G) SpGH101 structure (PDB entry 3ECQ). (H) SpuA in complex with maltotetraose (PDB entry 2YA1). Domains/modules with catalytic activity are in red, domains/modules with known/putative carbohydrate-binding function are in blue, and ligand molecules are in green.

analogues and inhibitors. Its structure (residues 322–791) consists of a canonical six-bladed β-propeller fold, with a 200-residue insertion between the second and third β-strands of the second blade (Fig. 3F) (Gut *et al.*, 2011). The structural information enabled understanding of the mechanism of action of inhibitors targeting the catalytic domain of NanA (Gut *et al.*, 2011).

EndoD cleaves the chitobiose core present in *N*-glycans (Muramatsu *et al.*, 2001), and its substrates identified so far are transferrin, fetuin and IgG (Garbe & Collin, 2012). Deglycosylation of IgG by EndoD is thought to contribute to virulence by decreasing the ability of these antibodies to interact with complement (Garbe & Collin, 2012). Pneumococcal EndoD is a

multi-domain protein with an N-terminal GH85 domain, followed by a carbohydrate-binding domain of the family 32 (CBM32), whose structures have been solved independently (Abbott *et al.*, 2009; Abbott & Boraston, 2011). The EndoD-GH85 domain consists of an N-terminal $(\beta/\alpha)_8$ barrel (residues 173–522) followed by two β-sandwich domains, D1 (residues 523–702) and D2 (residues 703–808) (Fig. 3B). The substrate-binding site consists of a groove that extends along the protein surface and diverges in two channels. The fact that EndoD has preference by branched glycans (Muramatsu *et al.*, 2001) agrees with the presence of a bisected substrate-binding channel that should be able to accommodate the arms of branched glycans (Abbott *et al.*, 2009; Abbott & Boraston, 2011). The catalytic activity of EndoD is likely assisted by domains D1, D2 and EndoD-CBM32, which have been proposed to bind carbohydrate molecules and might therefore help the enzyme to interact with its substrate (Fig. 3C) (Abbott *et al.*, 2009; Abbott & Boraston, 2011).

Pneumococcal SpGH101 cleaves the bond between Gal-β-1,3-GalNAc moieties and Ser/Thr residues present in host glycoproteins (the T-antigen or Core-1-Type O-glycan). It is a virulence factor that may aid colonization by destroying the mucin barrier and providing a nutrient source (Caines *et al.*, 2008). The structure of SpGH101 was solved for seven of the eight predicted domains for the enzyme (residues 119–1481) (Caines *et al.*, 2008) (Fig. 3G). The catalytic GH101 domain or domain 3 (residues 602–893) folds as a distorted $(\beta/\alpha)_8$ barrel with three helices and strands replaced by a complex arrangement of loops. The putative active site is rich in polar residues, including those proposed to be the catalytic ones, which should be able to interact with the substrate. As occurs with other glycosyl hydrolases described in this section, the catalytic domain is accompanied by other domains with putative carbohydrate-binding ability (domains 1, 2, 5 and 6), which are likely participating in the recognition of the substrate (Caines *et al.*, 2008).

SpuA is a multi-modular enzyme required for full virulence in a mouse-lung model of *S. pneumoniae* infection (van Bueren *et al.*, 2007). It contains two N-terminal carbohydrate-binding domains (CBM41-1 and CBM41-2) and a catalytic GH13 domain that carries out the hydrolysis of α-1,6-branchpoints of glycogen. Interaction with intracellular glycogen has been proposed to take place through the CBM41 domains, including glycogen in the context of type II alveolar cells in lung tissue (van Bueren *et al.*, 2007). By depleting intracellular glycogen, the pathogen may negatively influence the synthesis of surfactant (for which glycogen is a precursor) and secure energy source during invasive lung infections. The structure of almost full-length SpuA in complex with maltotetraose has been reported, which comprises six of the seven domains of the enzyme and three molecules of maltotetraose (Fig. 3H) (Lammerts van Bueren *et al.*, 2011). The catalytic domain (residues 573–1035) consists of an $(\alpha/\beta)_8$ barrel. In the substrate-binding channel two molecules of maltotetraose were found mimicking the substrate (Lammerts van Bueren *et al.*, 2011). The CBM41-1 and CBM41-2 domains (residues 130–350) present a carbohydrate-binding site built by three aromatic residues. The CBM41-1 module orients over the catalytic center, and its carbohydrate-binding site is involved in both extending the active site and substrate recognition. Deletion of the CBM41 domains compromises the ability of the enzyme to hydrolyse its substrate, which reveals that both catalytic and CBM41-1 domains cooperate during catalysis (Lammerts van Bueren *et al.*, 2011). The CBM41-2 is distal from the active site and it likely has no direct effect on catalysis, although it should serve to anchor the glycogen. Results reported by van Bueren *et al.* provide ample information about the catalytic mechanism, in which the interplay between CBM41-1 and GH13 domains is key for the activity of the enzyme.

Hyaluronan lyase SpnHL

SpnHL is one of the major surface proteins among streptococci. It belongs to the polysaccharide lyase family 8, and it degrades essential components of the host's extracellular matrix, hyaluronan (HA), unsulfated chondroitin (CH) and certain chondroitin sulfates (CHS). This process facilitates bacterial spread among the host tissues (Rigden *et al.*, 2006). SpnHL is a four-domain protein formed by a presumed N-terminal HA/CH/CHS-binding domain followed by a spacer domain, a catalytic domain (α-domain) and a C-terminal (β-domain) that modulates the access of the substrate to the catalytic machinery (Fig. 3D) (Li *et al.*, 2000; Ponnuraj & Jedrzejas, 2000). Structural studies yielded 18 structures of the region

containing both α-domain and β-domain, alone and in complex with different ligands (Berry *et al.*, 1994; Jedrzejas *et al.*, 2000; Li *et al.*, 2000; Mitchell, 2000; Ponnuraj & Jedrzejas, 2000; Nukui *et al.*, 2003; Rigden & Jedrzejas, 2003; Botzki *et al.*, 2004; Rigden *et al.*, 2006). Both α-domain (residues 171–531) and β-domain (residues 543–893) build the catalytic cleft, whose properties allow substrate degradation to occur by a processive mechanism: (i) it is highly positively charged, which facilitates binding of the acidic substrate; (ii) it has a negative area at its end that will favor product release; (iii) it has three putative catalytic residues responsible for substrate cleavage; and (iv) it has a hydrophobic core that permits proper substrate orientation for catalysis (Li *et al.*, 2000; Ponnuraj & Jedrzejas, 2000; Rigden & Jedrzejas, 2003; Rigden *et al.*, 2006). In addition, flexibility of its modular architecture seems to permit an opening/closing motion that will favor substrate release and translocation during processive degradation (Rigden *et al.*, 2006).

Mucin-binding protein MucBP

MucBP is a mucin-binding protein that functions as adhesin by directly interacting with the epithelial cells (Bumbaca *et al.*, 2007; Du *et al.*, 2011). It is likely to play an important role in pneumococcal colonization and invasion of the host tissues by binding mucins, which are *O*-linked glycosylated proteins constituting the major part of the mucus. MucBP is formed by one N-terminal mucin-binding domain (MucBD) and one C-terminal proline-rich domain (PRD). Structure of MucBD consists of a single mucin-binding repeat with an immunoglobulin-like β-sandwich fold (Fig. 3E) (Du *et al.*, 2011). The protein presents three hydrophobic residues that are relatively conserved with respect to other MucBPs, and that could be involved in saccharide recognition (Du *et al.*, 2011).

THREE-DIMENSIONAL STRUCTURE OF LIPOPROTEINS

SP_0198

Three-dimensional structure of SP_0198 has been deposited in the Protein Data Bank (PDB) by the Midwest Center for Structural Genomics. The SP_0198 crystal structure is formed by a β-barrel motif, pointed by CATH database as a putative D-stereospecific aminopeptidase (Fig. 4A). SP_0198 has not been functionally characterized but holds the DUF3642 domain, a major virulence factor from gram-negative bacteria (Punta *et al.*, 2012). This fact strongly suggests that this lipoprotein could be involved in host–pathogen interactions in *S. pneumoniae.*

Sortase StrC-3

StrC-3 (also called StrD) is the only one of the three sortases of the pilus *rlrA* operon with a LipoBox signal whereas the other two sortases are probably linked to the membrane via a transmembrane domain. Sortases catalyse the attachment of LPXTG proteins to the peptidoglycan by first recognizing their LPXTG motif and subsequent nucleophilic attack on the Thr-Gly bond of the target protein by the sortase catalytic cysteine. StrC-3 sortase is an α/β structure composed of a β-barrel core surrounded by helices. The α3 helix is reported to act as a 'lid' that might regulate the StrC-3 interaction with substrate (Fig. 4D) (Manzano *et al.*, 2008). Direct mutagenesis experiments provided evidence that *in vivo* StrC could be involved in attaching two proteins of the pilus (RrgA and RrgB) to the peptidoglycan (Neiers *et al.*, 2009).

Metal ion-binding receptors AdcAII and PsaA

Upregulation of metal ion transport is critical for bacterial survival, metal deficiency will induce oxidative stress while excess will generate toxicity (Moore & Helmann, 2005). AdcAII (SP_1002) and PsaA (SP_1650) are two metal ion-binding receptors that belong to the ABC permeases cluster 9 (Dintilhac & Claverys, 1997; Claverys, 2001). AdcAII and PsaA are reported to have affinity for both Zn^{2+} and Mn^{2+} and their homologues are referred to as laminin-binding proteins in other bacterial species. The three-dimensional structures of AdcAII (Fig. 4C) and PsaA (Fig. 4B) have been reported, (Lawrence *et al.*, 1998; Loisel *et al.*, 2008). Both present a similar fold, the main difference being the conformation of a flexible loop that is located at the entrance of the binding site. Its fold presents two $(\beta/\alpha)_4$ domains linked by a rigid α-helix (Lawrence *et al.*, 1998; Loisel *et al.*, 2008). In both structures a Zn^{2+} cation is found in the interdomain cleft, being stabilized by a tetrahedral coordination involving His71, His147, His211 and

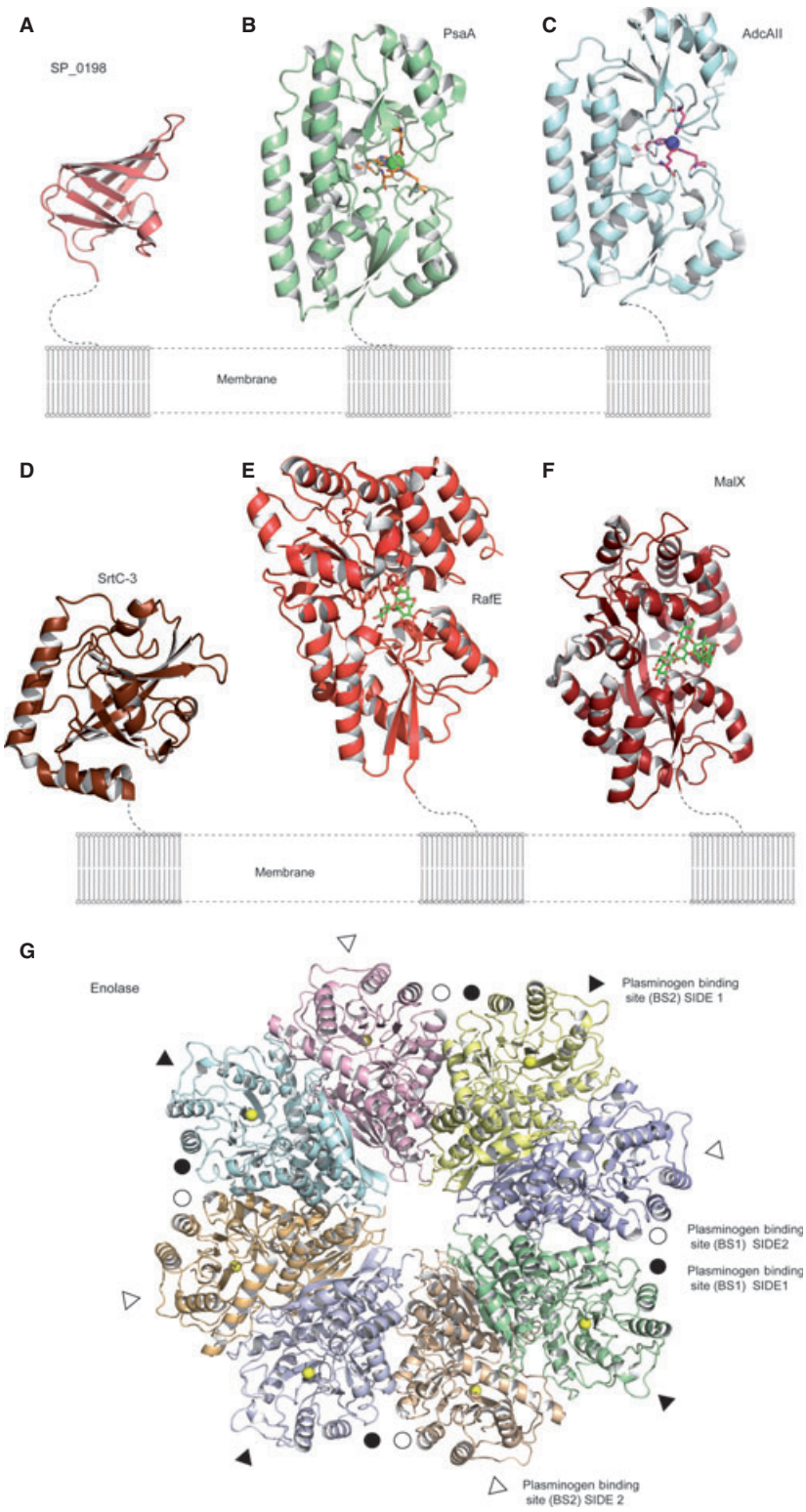

Figure 4 Three-dimensional structures of pneumococcal lipoproteins and non-classical surface proteins. (A) Lipoprotein SP_0198 [Protein Data Bank (PDB) entry 3GE2]. (B) Lipoprotein PsaA in complex with manganese, residues involved in metal coordination are marked in orange (PDB entry 3ZTT). (C) Lipoprotein AdcAII in complex with zinc (PDB entry 3CX3), residues involved in metal coordination are marked in red (D) Lipoprotein StrC-3 (PDB entry 2W1K). (E) Lipoprotein RafE in complex with Raffinose, ligand molecules are colored in green (PDB entry 2I58). (F) Lipoprotein MalX in complex with maltoheptaose, ligand molecules are colored in green (PDB entry 2XD3). (G) Non-classical surface protein Enolase. Protein is shown in its oligomeric state (octamer), monomers are highlighted in different colors. Binding sites are also indicated for reference (PDB entry 1W6T).

Glu286, in AdcAII; and His67, His139, Glu205 and Asp280, in PsaA. Despite the structural results referring to Zn^{2+} coordination, the affinity of PsaA for Mn^{2+} has been reported (Dintilhac *et al.*, 1997), suggesting a probable mechanism for Mn^{2+} stabilization coordinated not only by the amino acid composition

of the binding site but also by modulation of site geometry (Lawrence *et al.*, 1998).

Sugar transporter RafE

RafE is a component of the ABC transport system responsible for the uptake of raffinose, an α-galactosyl sucrose trisaccharide derivate composed of galactose, fructose and glucose (Rosenow *et al.*, 1999). The crystal structure of *S. pneumoniae* RafE has been reported (Fig. 4E) (Paterson *et al.*, 2006) and presents an α/β fold composed of two domains. One of the structures was obtained in complex with raffinose revealing the residues (Tyr65, Trp162, Trp246 and Trp348) that were critical in ligand binding (Paterson *et al.*, 2006). Little is known about the *in vivo* function of this protein, but a RafE-inactive pneumococcal strain showed a polar effect over-expression of its downstream permease genes rafF and rafG (287 times lower expression in the mutant), indicating the presence of a cascade mechanism in the *Raf* operon upon presence of substrate (Tyx *et al.*, 2011).

Maltose/maltodextrin-binding protein MalX

MalX is the only extracellular component of *S. pneumoniae* that is able to transport glycogen degradation products and therefore is critical in using exogenous glycogen as a carbon source. The ability of pneumococcal cells to use host glycogen is not only metabolically advantageous for a pathogen, but also helps pneumococci to invade host by attacking type II alveolar cells. These cells use glycogen as a key component of the innate immune system and as the main precursor for surfactant production (Cundell & Tuomanen, 1994). In that way, degradation and uptake of glycogen would help pneumococci to grow while favoring colonization by lowering the host response. MalX structure (Fig. 4F) (Abbott *et al.*, 2010) presents a typical α/β-fold with two main domains that contain the binding pocket in the central cleft. The structure was obtained in complex with maltoheptaose, allowing the identification of binding sites and their key residues (Tyr197, Trp273 and Trp384), all of which are conserved in maltodextrin-binding proteins. This particular configuration of the binding pocket, with three conserved residues capable of stabilizing three sugar rings, explains the high affinity of MalX for maltotriose. The apo structure revealed that up to 12 residues could be accommodated upon opening the binding site. However, these extra binding sites do not strongly contribute to the stabilization of longer glucose chains; a result that explains similar MalX affinities for substrates ranging from four to seven glucose chains (Abbott *et al.*, 2010).

None of the available *S. pneumoniae* LP structures has interpretable electron density in the first 30-residue region. Sequence homology alignment and motif predictions suggest that this connecting region of the protein to the membrane might be disordered *in vivo*, providing high flexibility of the LP and thereby facilitating orientation towards their targets (Lawrence *et al.*, 1998).

THREE-DIMENSIONAL STRUCTURES OF NON-CLASSICAL SURFACE PROTEINS

Enolase

The pneumococcal enolase is the only NCSP structure reported to date (Ehinger *et al.*, 2004). Analytical gel filtration demonstrated enolase octameric oligomerization, a result further corroborated by the crystal structure (Fig. 4G). The octamer is forming a ring with a 19-Å-wide central pore. The N-terminal part of each monomer is oriented towards the center, while the C-terminus holding the plasminogen-binding sites is facing the outer ring.

The enolase active site acts also as plasminogen-binding site (BS1) and is located at the C-terminus end of the β-barrel including catalytic Glu164, Glu205 and Lys342. Under denaturing conditions, which impede the oligomerization of enolase, the chemical inactivation of Lys342 abolishes plasminogen binding. However, blocking this binding site (BS1) in the oligomer does not prevent plasminogen binding, indicating the presence of another site (BS2). This new site was identified by spot-synthesized peptide analysis and comprises from residues 248 to 256. BS2 is significantly more exposed, in the oligomeric enolase, and is therefore essential in moonlighting functions to recruit human plasminogen during infection (Ehinger *et al.*, 2004).

The moonlighting roles of many NCSPs in virulence remain unknown. However, structural analysis of these proteins might provide clues about their non-classical functions in virulence, providing new targets for the development of novel vaccines and therapeutics.

THREE-DIMENSIONAL STRUCTURES OF CBPS

Autolysin LytA

Two autolysins have been unequivocally identified in pneumococcus so far: the LytA amidase and the LytC lysozyme (Garcia *et al.*, 1999). LytA amidase has been well studied and represents the paradigm of autolytic enzymes (Lopez & Garcia, 2004). LytA is responsible for cellular autolysis, through which it mediates release of toxic substances – such as the pore-forming toxin pneumolysin and cell wall degradation products – that damage endothelial and epithelial barriers and allow pneumococci to gain access to the bloodstream and disseminate through the body (Berry & Paton, 2000). Interestingly, it has been shown that pneumococci are protected from the lytic activity of LytA during exponential growth, but not in the stationary phase (Tomasz *et al.*, 1970). Recent results confirm this point and a model explaining why this occurs has been proposed (Mellroth *et al.*, 2012). LytA is structurally organized as a two-module protein (Fig. 2) with an N-terminal module (residues 1–174) that catalyses the cleavage of the *N*-acetylmuramoyl-L-alanine bond of the pneumococcal peptidoglycan backbone (Mosser & Tomasz, 1970), and a C-terminal choline-binding module. The C-terminal module (residues 175–301), responsible for cell-wall binding, is formed by a tandem of six imperfect 20-residue repeats, known either as P-motifs or choline-binding repeats, characteristics of the CBP family.

Up to now, there is no structural information for the full-length protein but the crystal structure of the choline-binding module from LytA (C-LytA) has been reported (Fernandez-Tornero *et al.*, 2001, 2002). C-LytA presents a single regular fold formed by a superhelical arrangement of the choline-binding repeats and the C-terminal tail (Fig. 5A). Choline is located at the interface of two consecutive repeats in such a way that three structurally conserved aromatic residues (two Trp and one Tyr) form a cavity in which choline methyl groups are placed (Fig. 5B). A cation-π interaction between the electron-rich systems of the aromatic rings and the positive charge of choline enhances the binding. This recognition pattern has been also observed in all the CBPs reported up to now.

Despite the numerous studies on LytA, several basic features of this protein still remain elusive. It is unclear what regulates its lytic activity, why lysis is triggered in a predictable timeframe following entry into stationary phase, how autolysis is connected to penicillin-induced lysis (Tomasz *et al.*, 1970), how conversion of an inactive E-form to the active C-form of LytA (Tomasz & Westphal, 1971) is produced, and how LytA is targeted to the cell wall. In this sense, the elucidation of the three-dimensional structure of the full-length LytA could provide valuable insights to clarify these questions.

Pneumococcal surface protein A

Pneumococcal surface protein A (PspA) is a CBP (Jedrzejas *et al.*, 2000) that, as shown by anti-PspA antibodies, localizes on the surface of the cell wall of pneumococci (McDaniel *et al.*, 1984). The molecular mass of PspA ranges between 67 and 98 kDa in various pneumococcal strains. PspA is structurally organized in four distinct regions: an N-terminal functional module/domain (composed of 288 residues in the Rx1 strain of *S. pneumoniae*), a proline-rich region (83 residues), a choline-binding module built from a stretch of 10 choline-binding repeats, and finally a C-terminal tail of 17 hydrophobic residues (Jedrzejas *et al.*, 2001) (Fig. 2). The N-terminal functional domain is thought to extend from the cell wall and even to protrude outside the capsule. The N-terminal half of mature PspA is predicted to be entirely α-helical (Yother & Briles, 1992). A lactoferrin-binding region of PspA has been localized within residues 168–288 of this α-helical domain (Hakansson *et al.*, 2001).

No structural information has been reported for the full-length protein. However, the crystal structure of a complex of the lactoferrin-binding domain of PspA with the N-lobe of human lactoferrin, a component of the innate immune system, has been reported (Senkovich *et al.*, 2007). The lactoferrin-binding domain of PspA consists of four α-helices connected by mobile loops. At the N-terminus there is a short α-helix followed by three long amphipathic helices with numerous hydrophobic interactions between neighboring antiparallel helices. The complex structure revealed direct and specific interactions between the negatively charged surface of PspA helices and the highly cationic lactoferrin moiety of lactoferrin. Binding of PspA should block the surface accessibility of this bactericidal peptide, preventing it from penetrating the bacterial membrane (Senkovich *et al.*, 2007).

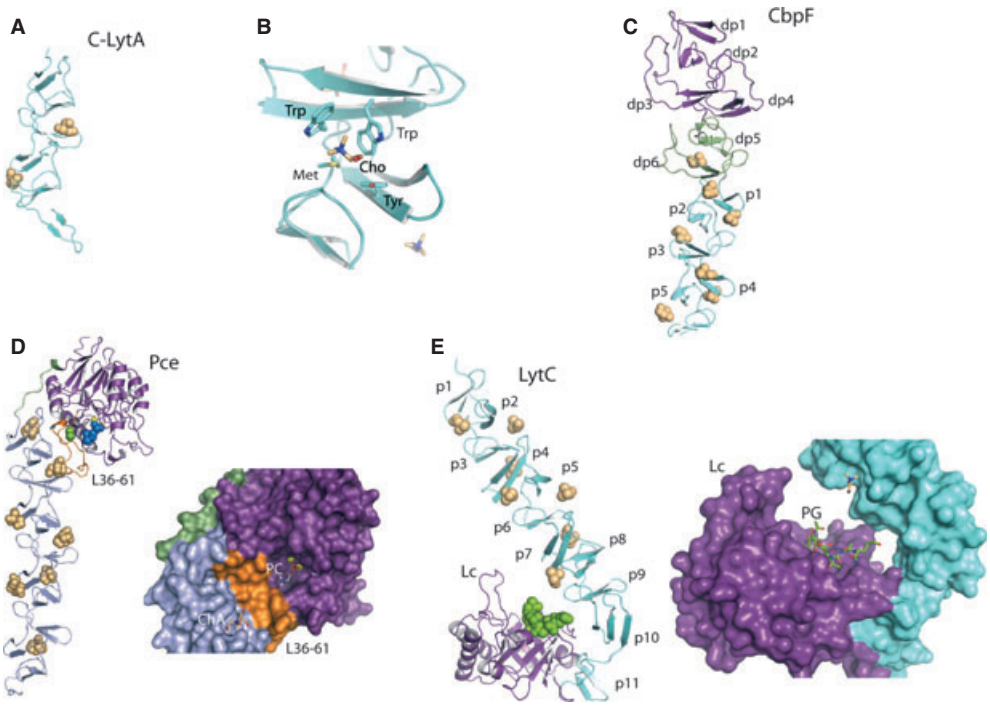

Figure 5 Three-dimensional structures of choline-binding proteins (CBPs). (A) Ribbon diagram of C-LytA (PDB entry 1H8G). The choline analogues are drawn as orange spheres. (B) Choline stabilization in CBPs. Residues forming the choline-binding site cavity are represented in a stick model. Choline molecule (Cho) is labeled. (C) Ribbon diagram of CbpF with choline-binding domain colored in cyan, linker domain colored in green, and N-terminal domain colored in magenta (PDB entry 2V04). Consensus (p1–p5) and non-consensus (dp1–dp6) choline-binding repeats are labeled. (D) Structure of pneumococcal phosphorylcholine esterase, Pce, in complex with phosphorylcholine (PC) and choline analogues (ChA) (PDB entry 2BIB). Modules colored differently: N-terminal catalytic module, magenta; choline-binding module, cyan; linker, green. The loop (residues 36–61) involved in intermodular interactions is colored in orange. The PC molecule at the active site is drawn in blue spheres, while Bis-Tris molecules bound to the choline-binding sites and metal ions (Zn, yellow; Ca, green) are represented as spheres. Right, zoom of the active site of Pce with molecular surface colored according to its domains. The PC molecule (white sticks) and Bis-Tris molecules (orange sticks) are labeled. (E) Structure of pneumococcal autolysin LytC, in complex with peptidoglycan and choline (PDB entry 2WWD). Modules colored differently: C-terminal catalytic module, magenta; choline-binding module, cyan. The peptidoglycan molecule at the active site is drawn in green spheres, while choline molecules are represented as orange spheres. The loop Lc involved in specific activity of LytC is labeled. Right, hook-shape conformation of LytC. Molecular surface is colored according to its domains. The peptidoglycan molecule (green sticks) and choline molecules (orange sticks) are labeled.

Choline-binding protein A

CbpA (also referred to as PspC, SpsA and PbcA) is one of the principal pneumococcal adhesins. The sequence of the CbpA N-terminus (residues 39–514; (Tettelin *et al.*, 2001) exhibits numerous repeats of the leucine zipper motif (Landschulz *et al.*, 1988) that cluster within five domains termed A, B, R1, R2 and C (Hammerschmidt *et al.*, 1997) (Fig. 2). Domains A, B and C are 21–25 amino acids in length and are predicted to form coiled-coil dimers. The 110-residues-long, 'repeated' domains, R1 and R2 (78% identical) (Zhang *et al.*, 2000), are the adhesion domains of CbpA (Lu *et al.*, 2003; Elm *et al.*, 2004b). The C-terminal domain of CbpA is formed by eight choline-binding repeats.

To avoid complement-mediated bacterial lysis, pneumococci recruit the central complement regulators Factor H and C4b-binding protein. The major Factor H-binding protein of *S. pneumoniae* is CbpA (PspC) and is termed Hic (factor H-binding inhibitor of complement) in another subset of strains (Agarwal *et al.*, 2010). Interestingly, while the C-terminal choline-binding domain of PspC anchors the protein non-covalently to the phosphorylcholine of the cell wall, the PspC-like Hic (PspC11.4) is covalently anchored to the peptidoglycan of pneumococci after transpeptidase cleavage of the LPXTG motif. Both PspC and

Hic share in their N-terminal regions a binding domain for the human complement inhibitor Factor H (Agarwal *et al.*, 2010).

CbpA R domains bind the extracellular immunoglobulin-like domains of polymeric immunoglobulin receptor (pIgR) during invasion of human nasopharyngeal epithelial cells (Hammerschmidt *et al.*, 2000; Zhang *et al.*, 2000). CbpA binds specifically to an extracellular domain of pIgR (Dave *et al.*, 2004; Elm *et al.*, 2004a) and hijacks the endocytosis machinery to translocate pneumococci across nasopharyngeal epithelial cells into the bloodstream (Luo *et al.*, 2005). Despite a lack of information about the complete CbpA structure; the solution structure of domain R2 of CbpA has been determined using nuclear magnetic resonance spectroscopy (Luo *et al.*, 2005). This structure was also used to model that for domain R1. The R domains, comprised of 12 imperfect copies of the leucine zipper heptad motif, adopt a unique 3-α-helix, raft-like structure. Each pair of α-helices is antiparallel and conserved residues in the loop between Helices 1 and 2 exhibit a novel 'tyrosine fork' structure that is involved in binding pIgR.

Phosphorylcholine esterase Pce

The TA phosphorylcholine esterase, Pce (or CbpE) was first described in 1974 (Holtje & Tomasz, 1974), showing that the enzyme is capable of removing a limited number of phosphorylcholine (PC) residues from pneumococcal cell walls. The molecular architecture of Pce includes a catalytic module localized at the N-terminal part of the protein (312 residues), a C-terminal choline-binding module with 10 homologous repeating units (205 residues), and a long C-terminal tail of 85 residues (Fig. 2).

Structural information for the catalytic domain (Garau *et al.*, 2005) and for the complete Pce lacking only the C-terminal tail (Hermoso *et al.*, 2005) has been reported. The crystal structure of full-length protein (Fig. 5D) comprises the catalytic module and the choline-binding module, which are joined by a small linker. The catalytic module folds into an $\alpha\beta/\beta\alpha$ sandwich similar to that of metallo-β-lactamases. The choline-binding module is formed by 10 repeats (p1–p10) following a left-handed superhelical fold, similar to that observed in C-LytA. The active site is located at the interface of the β-sheets in the N-terminal module, where a long groove of highly negative electrostatic potential contains two Zn^{2+} ions placed at the bottom of a deep hole (Fig. 5D).

Pce shows a novel structural arrangement of its constituent modules. Despite strong differences in the sizes and shapes of the modules, the overall structure seems to be quite rigid. The Pce structural framework is determined by three main factors: (i) a short linker on the surface of the catalytic module; (ii) the presence of a very long loop in the catalytic module (residues 36–61), L36–61, that strongly interacts with the first three repeating units of the choline-binding module, and (iii) two structural Ca^{2+} ions reinforcing the L36–61 conformation. The major interactions mediated by this loop comprise a zipper-like system in which aromatic residues from both modules alternate. This modular arrangement produces an extended active site that connects the catalytic site with one choline-binding site (Fig. 5D).

One of the most striking features concerning Pce was the earlier observation, both *in vivo* (Holtje & Tomasz, 1974) and *in vitro* (de las Rivas *et al.*, 2001; Vollmer & Tomasz, 2001), that the enzyme is able to liberate only a limited number of choline residues from the pneumococcal cell wall, representing about 30% of the total PC content. The modular arrangement in Pce configures the active site of the full-length protein in such a way that only residues located at the end of the TA chains are accessible to the catalytic center. Therefore, Pce would be involved in specifically releasing only those PC terminal residues relevant for cell–cell interactions but retaining choline residues of the cell wall that are important for normal cell growth and maintenance of CBPs attached to the bacterial envelope (Hermoso *et al.*, 2005).

Selectively modifying the distribution of the PC moieties on the bacterial surface by Pce should impair the ability of components of the host response targeting PC, such as human C-reactive protein, to efficiently bind the bacteria, and would provide a mechanism for pneumococci to escape the attack by the host defense system. Moreover, the remaining exposed PC residues would be used to bind to platelet-activating factor receptors for promoting invasion. Hence, fine regulation of PC decoration on the bacterial surface, mediated by Pce activity, may favor both infection and colonization by *S. pneumoniae*.

Choline-binding protein F

CbpF is one of the most abundant proteins in the pneumococcal cell wall. The crystal structure of CbpF in complex with choline has been recently reported (Molina *et al.*, 2009a). CbpF displays a novel modular structure comprised of both consensus (p1–p5) and non-consensus (dp1–dp6) choline-binding repeats (Fig. 5C), distributed along its length, which dramatically alters its shape and binding ability, organizing the protein in two well-defined modules. The N-terminal module (N-CbpF) displays a disk-shape conformation while the C-terminal module (N-CbpF) follows the superhelical fold previously found in other CBPs (Fig. 5C). Seven choline molecules were found attached to the C-terminal domain. Remarkably, the repeats building the N-terminal module are highly modified, by both additional amino acids and mutations on different positions of the consensus motif, and do not fulfill the choline-binding requirements. Variations in the sequence of the choline-binding repeats were also responsible for the modular arrangement in CbpF (Molina *et al.*, 2009a).

Experimental assays proved that CbpF can inhibit the activity of pneumococcal autolysin LytC both *in vitro* and *in vivo*. Interestingly, C-CbpF did not alter the enzymatic activity of LytC, strongly suggesting that the N-CbpF might play a critical role in this specific inhibition. Remarkably, the inhibitory effect is not limited to LytC lysozyme but also affects other pneumococcal phage lysozymes such as Cpl-1 and Cpl-7 sharing the same catalytic module as LytC. This inhibitory effect should somehow be related to cell wall binding through its N-terminal domain as probed by surface plasmon resonance experiments (Molina *et al.*, 2009a).

A study of the CbpF orthologues in other pneumococcal strains revealed the existence of proteins CbpC and CbpJ in the TIGR4 strain. Further analyses of known pneumococcal genomes revealed the existence of several proteins that are likely to show the same architecture as CbpF. These proteins might constitute a new CbpF-like subfamily within the large CBP family of proteins having a typical choline-binding module and an N-terminal region formed by a series of non-consensus repeats. Despite the minor differences in amino acid composition among all the members of the CbpF-like subfamily, most of the residues involved in both the N-terminal structural framework and in the intermodular interactions are preserved, suggesting that these proteins could exhibit similar regulatory functions to those of CbpF.

Autolysin LytC

LytC lysozyme is gaining more attention mainly because of its recently described significant role in cellular fratricide (Claverys & Havarstein, 2007). It has been demonstrated that during the competent state pneumococci turn on the expression of proteinaceous toxins (LytA, LytC and CbpD), which kills and lyses non-competent pneumococcal sister cells, as well as other bacteria from closely related species (Johnsborg *et al.*, 2008). Fratricide contributes to virulence by exacerbating an infection through the release of virulence factors and inflammatory mediators. CbpD causes some lysis of target cells on its own, but its effect is multiplied several fold by its activation of LytA and LytC (Eldholm *et al.*, 2009). As lysis coincides with competence, DNA released from target cells is taken up by competent attacker cells, resulting in increased efficiency of gene transfer and therefore also in antibiotic resistance (Johnsborg & Havarstein, 2009). One of the most striking features concerning LytC was the recent observation that the relatively high concentration of LytC detected in the medium of non-competent pneumococcal cultures in the exponential phase of growth was not harmful to the cells, demonstrating that LytC is inactive or highly regulated under these circumstances (Eldholm *et al.*, 2009). It was also shown that the presence of CbpD, which harbors a CHAP (cysteine, histidine-dependent amidohydrolases/peptidases) domain involved in murein stem-peptide cleavage, was required to activate the LytC lysozyme during fratricide. However, the nature of this activation mechanism remained elusive.

The crystal structure of full-length LytC in a ternary complex with choline and a peptidoglycan fragment has been recently reported (Perez-Dorado *et al.*, 2010) (Fig. 5E). LytC comprises an N-terminal choline-binding module (residues 1–267) and a C-terminal catalytic module (residues 268–468) formed by a single structural domain showing the irregular $(\beta/\alpha)_5\ \beta_3$ barrel typical of the glycosyl hydrolase family 25 (GH25) (Hermoso *et al.*, 2003). Eleven choline-binding repeats (p1–p11) build the choline-binding module. While the first nine repeats (p1–p9) are

arranged in the typical super-helical left-handed fold involved in choline binding (NI domain), the last two repeats (p10, p11) (NII domain) are critical in the modular arrangement of LytC. The three-dimensional structure of these repeats differs from that of the preceding repeats; this breaks the linear orientation of the choline-binding module and orients the LytC active site towards the anchoring module. The modular arrangement is in the base of the specific activities observed for this autolysin. LytC and Cpl-1, an endolysin from the Cp-1 phage, both belong to the GH-25 family (Hermoso *et al.*, 2003); they cleave the same glycosidic linkage of pneumococcal peptidoglycan. However, their specific activities are strikingly disparate: 6000 U mg^{-1} for LytC vs. 100,000 U mg^{-1} for Cpl-1. Their optimal temperatures for *in vitro* activity are also different, since LytC is more active at 30°C, whereas Cpl-1 exhibits more activity at 37°C. Production of a chimera retaining the modular arrangement of LytC but with the catalytic center of Cpl-1 showed enzymatic properties that correlated perfectly with those of LytC (Perez-Dorado *et al.*, 2010); that is to say, the specific activity was low and similar to that of LytC and its optimal temperature was 30°C. This result suggested that the specific catalytic properties of LytC could not only be ascribed to the nature of the given active site, but also to its specific modular arrangement.

Furthermore, the unusual hook-shaped conformation of LytC with the active site is oriented towards the choline-binding module (Fig. 5E), which can explain its activation by CbpD in fratricide. The crystal structure of the LytC–choline–peptidoglycan complex indicates that long peptide stems and crosslinked peptidoglycan chains should be deleterious to the hydrolytic activity by steric hindrance with both a LytC-specific loop, Lc, and by the entire choline-binding module. Therefore, previous cleavage of the peptide stems performed by the CHAP domain of CbpD should facilitate hydrolysis of the uncrosslinked peptidoglycan chains by LytC, thereby explaining the activation of LytC observed in fratricide (Perez-Dorado *et al.*, 2010).

CONCLUDING REMARKS

Pneumococcal surface proteins are involved in critical aspects of pathogen life such as bacterial fitness, cell–cell interactions, elimination of competitors and pathogen–host interactions (e.g. by recruiting host components or escaping immune system). Four families of surface proteins are found in the pneumococcus, some of which are unique to this pathogen and its close relatives. During the various stages of the infection, these bacteria have to adapt to different host niches. This implies that the expression of virulence factors, which occurs in a highly coordinated and regulated manner, is essential for successful bacterial persistence in a defined host niche or its dissemination throughout the host. Structural studies have unraveled some of these roles and provide a rich picture of the great diversity and unexpected level of complexity underlying the virulence mechanisms that mediate infection.

As most of the structures of pneumococcal surface proteins remain uncharacterized, it is expected that the structural information of the remaining surface proteins will permit us to decipher interactions orchestrating the union of different proteins and the assembly of modular protein, and so revealing potential targets for new drugs to handle pneumococcal infections. Most of pneumococcal surface proteins display a modular nature. The structural information reported up to now, reveals that the combination of catalytic modules with cell-wall binding or adhesin modules does not yield a new set of activities that are the arithmetic sum of its components. In fact, as nicely observed in LPXTG proteins and especially in CBPs, the final result of the multimodular protein is a novel macromolecular machine displaying unique capacities. These novel activities include, for example, selective hydrolysis of the substrate (e.g. SpuA or Pce) or regulation of (auto)lytic activity (e.g. LytC) and in general produce a dual-function protein that combines bacterial fitness functions with host–pathogen interactions. In this sense, structural information of full-length pneumococcal surface proteins is capital to discover their functional complexity. As evidenced by the few examples reported to date, it is clear, at least in the pneumococcal surface proteins, that the whole is greater than the sum of its parts.

ACKNOWLEDGEMENTS

We are indebted to D. Laurents for his critical reading of the manuscript. This work was supported by a grant from the Spanish Ministry of Economy and

Competitiveness (BFU2011-25326), EU-CP223111 (CAREPNEUMO, European Union), and the biomedicine program of the government of the autonomous community of Madrid (S2010/BMD-2457). IP-D is funded by a FEBS long-term fellowship.

REFERENCES

Abbott, D.W. and Boraston, A. (2011) Structural analysis of a putative family 32 carbohydrate-binding module from the *Streptococcus pneumoniae* enzyme EndoD. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **67**: 429–433.

Abbott, D.W., Macauley, M.S., Vocadlo, D.J. and Boraston, A.B. (2009) *Streptococcus pneumoniae* endohexosaminidase D, structural and mechanistic insight into substrate-assisted catalysis in family 85 glycoside hydrolases. *J Biol Chem* **284**: 11676–11689.

Abbott, D.W., Higgins, M.A., Hyrnuik, S., Pluvinage, B., Lammerts van Bueren, A. and Boraston, A.B. (2010) The molecular basis of glycogen breakdown and transport in *Streptococcus pneumoniae*. *Mol Microbiol* **77**: 183–199.

Agarwal, V., Asmat, T.M., Luo, S., Jensch, I., Zipfel, P.F. and Hammerschmidt, S. (2010) Complement regulator Factor H mediates a two-step uptake of *Streptococcus pneumoniae* by human cells. *J Biol Chem* **285**: 23486–23495.

Alloing, G., Trombe, M.C. and Claverys, J.P. (1990) The ami locus of the gram-positive bacterium *Streptococcus pneumoniae* is similar to binding protein-dependent transport operons of gram-negative bacteria. *Mol Microbiol* **4**: 633–644.

Alloing, G., de Philip, P. and Claverys, J.P. (1994) Three highly homologous membrane-bound lipoproteins participate in oligopeptide transport by the Ami system of the gram-positive *Streptococcus pneumoniae*. *J Mol Biol* **241**: 44–58.

Austrian, R. (1981) The development of pneumococcal vaccine. *Proc Am Philos Soc* **125**: 46–51.

Babu, M.M., Priya, M.L., Selvan, A.T. *et al.* (2006) A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins. *J Bacteriol* **188**: 2761–2773.

Bean, B. and Tomasz, A. (1977) Choline metabolism in pneumococci. *J Bacteriol* **130**: 571–574.

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.

Bergmann, S. and Hammerschmidt, S. (2006) Versatility of pneumococcal surface proteins. *Microbiology* **152**: 295–303.

Bergmann, S., Rohde, M., Chhatwal, G.S. and Hammerschmidt, S. (2001) Alpha-Enolase of *Streptococcus pneumoniae* is a plasmin(ogen)-binding protein displayed on the bacterial cell surface. *Mol Microbiol* **40**: 1273–1287.

Bernardo-Garcia, N., Bartual, S.G., Fulde, M., Bergmann, S. and Hermoso, J.A. (2011) Crystallization and preliminary X-ray diffraction analysis of phosphoglycerate kinase from *Streptococcus pneumoniae*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **67**: 1285–1289.

Berry, A.M. and Paton, J.C. (1996) Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of *Streptococcus pneumoniae*. *Infect Immun* **64**: 5255–5262.

Berry, A.M. and Paton, J.C. (2000) Additive attenuation of virulence of *Streptococcus pneumoniae* by mutation of the genes encoding pneumolysin and other putative pneumococcal virulence proteins. *Infect Immun* **68**: 133–140.

Berry, A.M., Lock, R.A., Thomas, S.M., Rajan, D.P., Hansman, D. and Paton, J.C. (1994) Cloning and nucleotide sequence of the *Streptococcus pneumoniae* hyaluronidase gene and purification of the enzyme from recombinant *Escherichia coli*. *Infect Immun* **62**: 1101–1108.

Bethe, G., Nau, R., Wellmer, A. *et al.* (2001) The cell wall-associated serine protease PrtA: a highly conserved virulence factor of *Streptococcus pneumoniae*. *FEMS Microbiol Lett* **205**: 99–104.

Botzki, A., Rigden, D.J., Braun, S. *et al.* (2004) L-Ascorbic acid 6-hexadecanoate, a potent hyaluronidase inhibitor. X-ray structure and molecular modeling of enzyme-inhibitor complexes. *J Biol Chem* **279**: 45990–45997.

van Bueren, A.L., Higgins, M., Wang, D., Burke, R.D. and Boraston, A.B. (2007) Identification and structural basis of binding to host lung glycogen by streptococcal virulence factors. *Nat Struct Mol Biol* **14**: 76–84.

Bumbaca, D., Littlejohn, J.E., Nayakanti, H. *et al.* (2007) Genome-based identification and characterization of a putative mucin-binding protein from the surface of *Streptococcus pneumoniae*. *Proteins* **66**: 547–558.

Caines, M.E., Zhu, H., Vuckovic, M. *et al.* (2008) The structural basis for T-antigen hydrolysis by *Streptococcus pneumoniae*: a target for structure-based vaccine design. *J Biol Chem* **283**: 31279–31283.

Camara, M., Boulnois, G.J., Andrew, P.W. and Mitchell, T.J. (1994) A neuraminidase from *Streptococcus pneu-*

moniae has the features of a surface protein. *Infect Immun* **62**: 3688–3695.

Cartwright, K. (2002) Pneumococcal disease in western Europe: burden of disease, antibiotic resistance and management. *Eur J Pediatr* **161**: 188–195.

Clarke, V.A., Platt, N. and Butters, T.D. (1995) Cloning and expression of the beta-*N*-acetylglucosaminidase gene from *Streptococcus pneumoniae.* Generation of truncated enzymes with modified aglycon specificity. *J Biol Chem* **270**: 8805–8814.

Claverys, J.P. (2001) A new family of high-affinity ABC manganese and zinc permeases. *Res Microbiol* **152**: 231–243.

Claverys, J.P. and Havarstein, L.S. (2007) Cannibalism and fratricide: mechanisms and raisons d'etre. *Nat Rev Microbiol* **5**: 219–229.

Cundell, D.R. and Tuomanen, E.I. (1994) Receptor specificity of adherence of *Streptococcus pneumoniae* to human type-II pneumocytes and vascular endothelial cells *in vitro. Microb Pathog* **17**: 361–374.

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.

Daniely, D., Portnoi, M., Shagan, M. *et al.* (2006) Pneumococcal 6-phosphogluconate-dehydrogenase, a putative adhesin, induces protective immune response in mice. *Clin Exp Immunol* **144**: 254–263.

Dave, S., Carmicle, S., Hammerschmidt, S., Pangburn, M.K. and McDaniel, L.S. (2004) Dual roles of PspC, a surface protein of *Streptococcus pneumoniae*, in binding human secretory IgA and factor H. *J Immunol* **173**: 471–477.

De Las Rivas, B., Garcia, J.L., Lopez, R. and Garcia, P. (2002) Purification and polar localization of pneumococcal LytB, a putative endo-beta-N-acetylglucosaminidase: the chain-dispersing murein hydrolase. *J Bacteriol* **184**: 4988–5000.

Dintilhac, A. and Claverys, J.P. (1997) The adc locus, which affects competence for genetic transformation in *Streptococcus pneumoniae*, encodes an ABC transporter with a putative lipoprotein homologous to a family of streptococcal adhesins. *Res Microbiol* **148**: 119–131.

Dintilhac, A., Alloing, G., Granadel, C. and Claverys, J.P. (1997) Competence and virulence of *Streptococcus pneumoniae*: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases. *Mol Microbiol* **25**: 727–739.

Dramsi, S., Magnet, S., Davison, S. and Arthur, M. (2008) Covalent attachment of proteins to peptidoglycan. *FEMS Microbiol Rev* **32**: 307–320.

Du, Y., He, Y.X., Zhang, Z.Y. *et al.* (2011) Crystal structure of the mucin-binding domain of Spr1345 from *Streptococcus pneumoniae. J Struct Biol* **174**: 252–257.

Ehinger, S., Schubert, W.D., Bergmann, S., Hammerschmidt, S. and Heinz, D.W. (2004) Plasmin(ogen)-binding alpha-enolase from *Streptococcus pneumoniae*: crystal structure and evaluation of plasmin(ogen)-binding sites. *J Mol Biol* **343**: 997–1005.

Eldholm, V., Johnsborg, O., Haugen, K., Ohnstad, H.S. and Havarstein, L.S. (2009) Fratricide in *Streptococcus pneumoniae*: contributions and role of the cell wall hydrolases CbpD, LytA and LytC. *Microbiology* **155**: 2223–2234.

Elm, C., Braathen, R., Bergmann, S. *et al.* (2004a) Ectodomains 3 and 4 of human polymeric Immunoglobulin receptor (hpIgR) mediate invasion of *Streptococcus pneumoniae* into the epithelium. *J Biol Chem* **279**: 6296–6304.

Elm, C., Rohde, M., Vaerman, J.P., Chhatwal, G.S. and Hammerschmidt, S. (2004b) Characterization of the interaction of the pneumococcal surface protein SpsA with the human polymeric immunoglobulin receptor (hpIgR). *Indian J Med Res* **119**(Suppl.): 61–65.

Fernandez-Tornero, C., Lopez, R., Garcia, E., Gimenez-Gallego, G. and Romero, A. (2001) A novel solenoid fold in the cell wall anchoring domain of the pneumococcal virulence factor LytA. *Nat Struct Biol* **8**: 1020–1024.

Fernandez-Tornero, C., Garcia, E., Lopez, R., Gimenez-Gallego, G. and Romero, A. (2002) Two new crystal forms of the choline-binding domain of the major pneumococcal autolysin: insights into the dynamics of the active homodimer. *J Mol Biol* **321**: 163–173.

Frolet, C., Beniazza, M., Roux, L. *et al.* (2010) New adhesin functions of surface-exposed pneumococcal proteins. *BMC Microbiol* **10**: 190.

Garau, G., Lemaire, D., Vernet, T., Dideberg, O. and Di Guilmi, A.M. (2005) Crystal structure of phosphorylcholine esterase domain of the virulence factor choline-binding protein e from *Streptococcus pneumoniae*: new structural features among the metallo-beta-lactamase superfamily. *J Biol Chem* **280**: 28591–28600.

Garbe, J. and Collin, M. (2012) Bacterial hydrolysis of host glycoproteins – powerful protein modification and efficient nutrient acquisition. *J Innate Immun* **4**: 121–131.

Garcia, E., Garcia, J.L., Ronda, C., Garcia, P. and Lopez, R. (1985) Cloning and expression of the pneumococcal

autolysin gene in *Escherichia coli*. *Mol Gen Genet* **201**: 225–230.

Garcia, E., Garcia, J.L., Garcia, P., Arraras, A., Sanchez-Puelles, J.M. and Lopez, R. (1988) Molecular evolution of lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. *Proc Natl Acad Sci USA* **85**: 914–918.

Garcia, J.L., Sanchez-Beato, A.R., Medrano, F.J. and Lopez, R. (1998) Versatility of choline-binding domain. *Microb Drug Resist* **4**: 25–36.

Garcia, P., Paz Gonzalez, M., Garcia, E., Garcia, J.L. and Lopez, R. (1999) The molecular characterization of the first autolytic lysozyme of *Streptococcus pneumoniae* reveals evolutionary mobile domains. *Mol Microbiol* **33**: 128–138.

Gosink, K.K., Mann, E.R., Guglielmo, C., Tuomanen, E.I. and Masure, H.R. (2000) Role of novel choline binding proteins in virulence of *Streptococcus pneumoniae*. *Infect Immun* **68**: 5690–5695.

Gut, H., Xu, G., Taylor, G.L. and Walsh, M.A. (2011) Structural basis for *Streptococcus pneumoniae* NanA inhibition by influenza antivirals zanamivir and oseltamivir carboxylate. *J Mol Biol* **409**: 496–503.

Hakansson, A., Roche, H., Mirza, S., McDaniel, L.S., Brooks-Walter, A. and Briles, D.E. (2001) Characterization of binding of human lactoferrin to pneumococcal surface protein A. *Infect Immun* **69**: 3372–3381.

Hakenbeck, R., Madhour, A., Denapaite, D. and Bruckner, R. (2009) Versatility of choline metabolism and choline-binding proteins in *Streptococcus pneumoniae* and commensal streptococci. *FEMS Microbiol Rev* **33**: 572–586.

Hammerschmidt, S., Talay, S.R., Brandtzaeg, P. and Chhatwal, G.S. (1997) SpsA, a novel pneumococcal surface protein with specific binding to secretory immunoglobulin A and secretory component. *Mol Microbiol* **25**: 1113–1124.

Hammerschmidt, S., Tillig, M.P., Wolff, S., Vaerman, J.P. and Chhatwal, G.S. (2000) Species-specific binding of human secretory component to SpsA protein of *Streptococcus pneumoniae* via a hexapeptide motif. *Mol Microbiol* **36**: 726–736.

Hammerschmidt, S., Agarwal, V., Kunert, A., Haelbich, S., Skerka, C. and Zipfel, P.F. (2007) The host immune regulator factor H interacts via two contact sites with the PspC protein of *Streptococcus pneumoniae* and mediates adhesion to host epithelial cells. *J Immunol* **178**: 5848–5858.

Henrichsen, J. (1995) Six newly recognized types of *Streptococcus pneumoniae*. *J Clin Microbiol* **33**: 2759–2762.

Hermoso, J.A., Monterroso, B., Albert, A. *et al.* (2003) Structural basis for selective recognition of pneumococcal cell wall by modular endolysin from phage Cp-1. *Structure* **11**: 1239–1249.

Hermoso, J.A., Lagartera, L., Gonzalez, A. *et al.* (2005) Insights into pneumococcal pathogenesis from the crystal structure of the modular teichoic acid phosphorylcholine esterase Pce. *Nat Struct Mol Biol* **12**: 533–538.

Holmes, A.R., McNab, R., Millsap, K.W. *et al.* (2001) The pavA gene of *Streptococcus pneumoniae* encodes a fibronectin-binding protein that is essential for virulence. *Mol Microbiol* **41**: 1395–1408.

Holtje, J.V. and Tomasz, A. (1974) Teichoic acid phosphorylcholine esterase. A novel enzyme activity in pneumococcus. *J Biol Chem* **249**: 7032–7034.

Hoskins, J., Alborn, W.E. Jr, Arnold, J. *et al.* (2001) Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J Bacteriol* **183**: 5709–5717.

Hsiao, Y.S., Parker, D., Ratner, A.J., Prince, A. and Tong, L. (2009) Crystal structures of respiratory pathogen neuraminidases. *Biochem Biophys Res Commun* **380**: 467–471.

Jedrzejas, M.J., Chantalat, L. and Mewbourne, R.B. (1998) Crystallization and preliminary X-ray analysis of *Streptococcus pneumoniae* hyaluronate lyase. *J Struct Biol* **121**: 73–75.

Jedrzejas, M.J., Hollingshead, S.K., Lebowitz, J., Chantalat, L., Briles, D.E. and Lamani, E. (2000) Production and characterization of the functional fragment of pneumococcal surface protein A. *Arch Biochem Biophys* **373**: 116–125.

Jedrzejas, M.J., Lamani, E. and Becker, R.S. (2001) Characterization of selected strains of pneumococcal surface protein A. *J Biol Chem* **276**: 33121–33128.

Jensch, I., Gamez, G., Rothe, M. *et al.* (2010) PavB is a surface-exposed adhesin of *Streptococcus pneumoniae* contributing to nasopharyngeal colonization and airways infections. *Mol Microbiol* **77**: 22–43.

Jiang, Y.L., Yu, W.L., Zhang, J.W. *et al.* (2011) Structural basis for the substrate specificity of a novel beta-N-acetylhexosaminidase StrH protein from *Streptococcus pneumoniae* R6. *J Biol Chem* **286**: 43004–43012.

Johnsborg, O. and Havarstein, L.S. (2009) Regulation of natural genetic transformation and acquisition of transforming DNA in *Streptococcus pneumoniae*. *FEMS Microbiol Rev* **33**: 627–642.

Johnsborg, O., Eldholm, V., Bjornstad, M.L. and Havarstein, L.S. (2008) A predatory mechanism dramatically increases the efficiency of lateral gene transfer in

Streptococcus pneumoniae and related commensal species. *Mol Microbiol* **69**: 245–253.

Kadioglu, A., Weiser, J.N., Paton, J.C. and Andrew, P.W. (2008) The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat Rev Microbiol* **6**: 288–301.

Lammerts van Bueren, A., Ficko-Blean, E., Pluvinage, B. *et al.* (2011) The conformation and function of a multimodular glycogen-degrading pneumococcal virulence factor. *Structure* **19**: 640–651.

Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**: 1759–1764.

Lawrence, M.C., Pilling, P.A., Epa, V.C., Berry, A.M., Ogunniyi, A.D. and Paton, J.C. (1998) The crystal structure of pneumococcal surface antigen PsaA reveals a metal-binding site and a novel structure for a putative ABC-type binding protein. *Structure* **6**: 1553–1561.

Li, S., Kelly, S.J., Lamani, E., Ferraroni, M. and Jedrzejas, M.J. (2000) Structural basis of hyaluronan degradation by *Streptococcus pneumoniae* hyaluronate lyase. *EMBO J* **19**: 1228–1240.

Lofling, J., Vimberg, V., Battig, P. and Henriques-Normark, B. (2011) Cellular interactions by LPxTG-anchored pneumococcal adhesins and their streptococcal homologues. *Cell Microbiol* **13**: 186–197.

Loisel, E., Jacquamet, L., Serre, L. *et al.* (2008) AdcAII, a new pneumococcal Zn-binding protein homologous with ABC transporters: biochemical and structural analysis. *J Mol Biol* **381**: 594–606.

Lopez, R. and Garcia, E. (2004) Recent trends on the molecular biology of pneumococcal capsules, lytic enzymes, and bacteriophage. *FEMS Microbiol Rev* **28**: 553–580.

Lopez, R., Garcia, E., Garcia, P., Ronda, C. and Tomasz, A. (1982) Choline-containing bacteriophage receptors in *Streptococcus pneumoniae*. *J Bacteriol* **151**: 1581–1590.

López, R., García, E., García, P. and García, J.L. (2004) Cell wall hydrolases. In: Mitchell T.J., Morrison D.A., Spratt B.G. eds. *The Pneumococcus*. Washington: ASM, 75–88.

Lu, L., Lamm, M.E., Li, H., Corthesy, B. and Zhang, J.R. (2003) The human polymeric immunoglobulin receptor binds to *Streptococcus pneumoniae* via domains 3 and 4. *J Biol Chem* **278**: 48178–48187.

Luo, R., Mann, B., Lewis, W.S. *et al.* (2005) Solution structure of choline binding protein A, the major adhesin of *Streptococcus pneumoniae*. *EMBO J* **24**: 34–43.

Manzano, C., Contreras-Martel, C., El Mortaji, L. *et al.* (2008) Sortase-mediated pilus fiber biogenesis in *Streptococcus pneumoniae*. *Structure* **16**: 1838–1848.

Mazmanian, S.K., Liu, G., Jensen, E.R., Lenoy, E. and Schneewind, O. (2000) *Staphylococcus aureus* sortase mutants defective in the display of surface proteins and in the pathogenesis of animal infections. *Proc Natl Acad Sci USA* **97**: 5510–5515.

McDaniel, L.S., Scott, G., Kearney, J.F. and Briles, D.E. (1984) Monoclonal antibodies against protease-sensitive pneumococcal antigens can protect mice from fatal infection with *Streptococcus pneumoniae*. *J Exp Med* **160**: 386–397.

Mellroth, P., Daniels, R., Eberhardt, A. *et al.* (2012) LytA, the major autolysin of *Streptococcus pneumoniae*, requires access to the nascent peptidoglycan. *J Biol Chem* **287**: 11018–11029.

Mirza, S., Wilson, L., Benjamin, W.H. Jr *et al.* (2011) Serine protease PrtA from *Streptococcus pneumoniae* plays a role in the killing of *S. pneumoniae* by apolactoferrin. *Infect Immun* **79**: 2440–2450.

Mistry, D. and Stockley, R.A. (2006) IgA1 protease. *Int J Biochem Cell Biol* **38**: 1244–1248.

Mitchell, T.J. (2000) Virulence factors and the pathogenesis of disease caused by *Streptococcus pneumoniae*. *Res Microbiol* **151**: 413–419.

Molina, R., Gonzalez, A., Stelter, M. *et al.* (2009a) Crystal structure of CbpF, a bifunctional choline-binding protein and autolysis regulator from *Streptococcus pneumoniae*. *EMBO Rep* **10**: 246–251.

Molina, R., Stelter, M., Kahn, R. and Hermoso, J.A. (2009b) Characterization of gadolinium complexes for SAD phasing in macromolecular crystallography: application to CbpF. *Acta Crystallogr D Biol Crystallogr* **65**: 823–831.

Moore, C.M. and Helmann, J.D. (2005) Metal ion homeostasis in *Bacillus subtilis*. *Curr Opin Microbiol* **8**: 188–195.

Moschioni, M., Donati, C., Muzzi, A. *et al.* (2008) *Streptococcus pneumoniae* contains 3 rlrA pilus variants that are clonally related. *J Infect Dis* **197**: 888–896.

Mosser, J.L. and Tomasz, A. (1970) Choline-containing teichoic acid as a structural component of pneumococcal cell wall and its role in sensitivity to lysis by an autolytic enzyme. *J Biol Chem* **245**: 287–298.

Muramatsu, H., Tachikui, H., Ushida, H. *et al.* (2001) Molecular cloning and expression of endo-beta-N-acetylglucosaminidase D, which acts on the core structure of complex type asparagine-linked oligosaccharides. *J Biochem* **129**: 923–928.

Neiers, F., Madhurantakam, C., Falker, S. *et al.* (2009) Two crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis and substrate specificity. *J Mol Biol* **393**: 704–716.

Nobbs, A.H., Lamont, R.J. and Jenkinson, H.F. (2009) Streptococcus adherence and colonization. *Microbiol Mol Biol Rev* **73**: 407–450, Table of Contents.

Novak, R., Braun, J.S., Charpentier, E. and Tuomanen, E. (1998) Penicillin tolerance genes of *Streptococcus pneumoniae*: the ABC-type manganese permease complex Psa. *Mol Microbiol* **29**: 1285–1296.

Novak, R., Cauwels, A., Charpentier, E. and Tuomanen, E. (1999) Identification of a *Streptococcus pneumoniae* gene locus encoding proteins of an ABC phosphate transporter and a two-component regulatory system. *J Bacteriol* **181**: 1126–1133.

Nukui, M., Taylor, K.B., McPherson, D.T., Shigenaga, M.K. and Jedrzejas, M.J. (2003) The function of hydrophobic residues in the catalytic cleft of *Streptococcus pneumoniae* hyaluronate lyase. Kinetic characterization of mutant enzyme forms. *J Biol Chem* **278**: 3079–3088.

Oggioni, M.R., Memmi, G., Maggi, T., Chiavolini, D., Iannelli, F. and Pozzi, G. (2003) Pneumococcal zinc metalloproteinase ZmpC cleaves human matrix metalloproteinase 9 and is a virulence factor in experimental pneumonia. *Mol Microbiol* **49**: 795–805.

Orihuela, C.J., Mills, J., Robb, C.W., Wilson, C.J., Watson, D.A. and Niesel, D.W. (2001) *Streptococcus pneumoniae* PstS production is phosphate responsive and enhanced during growth in the murine peritoneal cavity. *Infect Immun* **69**: 7565–7571.

Oudega, B., Clark, D., Stegehuis, F., Majoor, M.J. and Luirink, J. (1993) A lipoprotein signal peptide plus a cysteine residue at the amino-terminal end of the periplasmic protein beta-lactamase is sufficient for its lipid modification, processing and membrane localization in *Escherichia coli*. *FEMS Microbiol Lett* **108**: 353–359.

Paterson, G.K. and Orihuela, C.J. (2010) Pneumococcal microbial surface components recognizing adhesive matrix molecules targeting of the extracellular matrix. *Mol Microbiol* **77**: 1–5.

Paterson, N.G., Riboldi-Tunnicliffe, A., Mitchell, T.J. and Isaacs, N.W. (2006) Purification, crystallization and preliminary X-ray diffraction analysis of RafE, a sugar-binding lipoprotein from *Streptococcus pneumoniae*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **62**: 676–679.

Paterson, G.K., Nieminen, L., Jefferies, J.M. and Mitchell, T.J. (2008) PclA, a pneumococcal collagen-like protein with selected strain distribution, contributes to adherence and invasion of host cells. *FEMS Microbiol Lett* **285**: 170–176.

Perez-Dorado, I., Gonzalez, A., Morales, M. *et al.* (2010) Insights into pneumococcal fratricide from the crystal structures of the modular killing factor LytC. *Nat Struct Mol Biol* **17**: 576–581.

Pluvinage, B., Higgins, M.A., Abbott, D.W. *et al.* (2011) Inhibition of the pneumococcal virulence factor StrH and molecular insights into *N*-glycan recognition and hydrolysis. *Structure* **19**: 1603–1614.

Ponnuraj, K. and Jedrzejas, M.J. (2000) Mechanism of hyaluronan binding and degradation: structure of *Streptococcus pneumoniae* hyaluronate lyase in complex with hyaluronic acid disaccharide at 1.7 A resolution. *J Mol Biol* **299**: 885–895.

Punta, M., Coggill, P.C., Eberhardt, R.Y. *et al.* (2012) The Pfam protein families database. *Nucleic Acids Res* **40**: D290–301.

Puyet, A. and Espinosa, M. (1993) Structure of the maltodextrin-uptake locus of *Streptococcus pneumoniae*. Correlation to the Escherichia coli maltose regulon. *J Mol Biol* **230**: 800–811.

Rigden, D.J. and Jedrzejas, M.J. (2003) Structures of *Streptococcus pneumoniae* hyaluronate lyase in complex with chondroitin and chondroitin sulfate disaccharides. Insights into specificity and mechanism of action. *J Biol Chem* **278**: 50596–50606.

Rigden, D.J., Littlejohn, J.E., Joshi, H.V., de Groot, B.L. and Jedrzejas, M.J. (2006) Alternate structural conformations of *Streptococcus pneumoniae* hyaluronan lyase: insights into enzyme flexibility and underlying molecular mechanism of action. *J Mol Biol* **358**: 1165–1178.

de las Rivas, B., Garcia, J.L., Lopez, R. and Garcia, P. (2001) Molecular characterization of the pneumococcal teichoic acid phosphorylcholine esterase. *Microb Drug Resist* **7**: 213–222.

Rosenow, C., Maniar, M. and Trias, J. (1999) Regulation of the alpha-galactosidase activity in *Streptococcus pneumoniae*: characterization of the raffinose utilization system. *Genome Res* **9**: 1189–1197.

Sanchez-Beato, A.R., Lopez, R. and Garcia, J.L. (1998) Molecular characterization of PcpA: a novel choline-binding protein of *Streptococcus pneumoniae*. *FEMS Microbiol Lett* **164**: 207–214.

Sebert, M.E., Palmer, L.M., Rosenberg, M. and Weiser, J.N. (2002) Microarray-based identification of htrA, a *Streptococcus pneumoniae* gene that is regulated by the CiaRH two-component system and contributes to nasopharyngeal colonization. *Infect Immun* **70**: 4059–4067.

Senkovich, O., Cook, W.J., Mirza, S. *et al.* (2007) Structure of a complex of human lactoferrin N-lobe with pneumococcal surface protein a provides insight into microbial defense mechanism. *J Mol Biol* **370**: 701–713.

Shivshankar, P., Sanchez, C., Rose, L.F. and Orihuela, C.J. (2009) The *Streptococcus pneumoniae* adhesin PsrP binds to Keratin 10 on lung cells. *Mol Microbiol* **73**: 663–679.

Swiatlo, E., McDaniel, L.S. and Briles, D.E. (2004) Choline-binding proteins. In: Mitchell T.J., Morrison D.A., Spratt B.G. eds. *The Pneumococcus*. Washington: ASM, 49–60.

Tettelin, H., Nelson, K.E., Paulsen, I.T. *et al.* (2001) Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**: 498–506.

Tomasz, A. (1967) Choline in the cell wall of a bacterium: novel type of polymer-linked choline in Pneumococcus. *Science* **157**: 694–697.

Tomasz, A. and Westphal, M. (1971) Abnormal autolytic enzyme in a pneumococus with altered teichoic acid composition. *Proc Natl Acad Sci USA* **68**: 2627–2630.

Tomasz, A., Albino, A. and Zanati, E. (1970) Multiple antibiotic resistance in a bacterium with suppressed autolytic system. *Nature* **227**: 138–140.

Tu, A.H., Fulgham, R.L., McCrory, M.A., Briles, D.E. and Szalai, A.J. (1999) Pneumococcal surface protein A inhibits complement activation by *Streptococcus pneumoniae*. *Infect Immun* **67**: 4720–4724.

Tyx, R.E., Roche-Hakansson, H. and Hakansson, A.P. (2011) Role of dihydrolipoamide dehydrogenase in regulation of raffinose transport in *Streptococcus pneumoniae*. *J Bacteriol* **193**: 3512–3524.

Umemoto, J., Bhavanandan, V.P. and Davidson, E.A. (1977) Purification and properties of an endo-alpha-N-acetyl-D-galactosaminidase from *Diplococcus pneumoniae*. *J Biol Chem* **252**: 8609–8614.

Vollmer, W. (2007) Structure and biosynthesis of the pneumococcal cell wall. In: Mitchell T.J., Morrison D.A., Spratt B.G. (eds). *Molecular Biology of Streptococci*. Norfolk: Horizon Bioscience, 83–118.

Vollmer, W. and Tomasz, A. (2001) Identification of the teichoic acid phosphorylcholine esterase in *Streptococcus pneumoniae*. *Mol Microbiol* **39**: 1610–1622.

Wani, J.H., Gilbert, J.V., Plaut, A.G. and Weiser, J.N. (1996) Identification, cloning, and sequencing of the immunoglobulin A1 protease gene of *Streptococcus pneumoniae*. *Infect Immun* **64**: 3967–3974.

Willis, L.M., Zhang, R., Reid, A., Withers, S.G. and Wakarchuk, W.W. (2009) Mechanistic investigation of the endo-alpha-*N*-acetylgalactosaminidase from *Streptococcus pneumoniae* R6. *Biochemistry* **48**: 10334–10341.

Xu, G., Li, X., Andrew, P.W. and Taylor, G.L. (2008) Structure of the catalytic domain of *Streptococcus pneumoniae* sialidase NanA. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **64**: 772–775.

Yamaguchi, M., Terao, Y., Mori, Y., Hamada, S. and Kawabata, S. (2008) PfbA, a novel plasmin- and fibronectin-binding protein of *Streptococcus pneumoniae*, contributes to fibronectin-dependent adhesion and antiphagocytosis. *J Biol Chem* **283**: 36272–36279.

Yother, J. (2004) Capsules. In: Mitchell T.J., Morrison D.A., Spratt B.G. eds. *The Pneumococcus*. Washington: ASM, 30–49.

Yother, J. and Briles, D.E. (1992) Structural properties and evolutionary relationships of PspA, a surface protein of *Streptococcus pneumoniae*, as revealed by sequence analysis. *J Bacteriol* **174**: 601–609.

Zahner, D. and Hakenbeck, R. (2000) The *Streptococcus pneumoniae* beta-galactosidase is a surface protein. *J Bacteriol* **182**: 5919–5921.

Zhang, J.R., Mostov, K.E., Lamm, M.E. *et al.* (2000) The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. *Cell* **102**: 827–837.

Zhang, Z., Li, W., Frolet, C. *et al.* (2009) Structure of the choline-binding domain of Spr1274 in *Streptococcus pneumoniae*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **65**: 757–761.