

**Investigation of the role of *Streptococcus pneumoniae*  
surface proteins PspA and PspC**

**Dissertation**

**zur Erlangung des akademischen Grades eines**

**„doctor rerum naturalium“ (Dr. rer. nat.)**

**vorgelegt dem Rat der Fakultät für Biowissenschaften**

**der Friedrich-Schiller-Universität Jena**

**Von Shanshan Du**

**geboren am 09. Mai 1990 in Henan, China**

Gutachter 1: Prof. Dr. Peter F. Zipfel  
Department of Infection Biology, Leibniz Institute for Natural Product  
Research and Infection Biology e.V. Hans-Knöll-Institute (HKI),  
Beutenbergstr. 11a, 07745 Jena Germany

Gutachter 2: Prof. Dr. Marc Thilo Figge  
Department of Applied Systems Biology, Leibniz Institute for  
Natural Product Research and Infection Biology e.V. Hans-Knöll-Institute (HKI),  
Beutenbergstr. 11a, 07745 Jena Germany

Gutachter 3: Prof. Dr. Peter Valentin-Weigand  
University of Veterinary Medicine Hannover,  
Bischofsholer Damm 15, 30173 Hannover

Data of defense: April 5th, 2022

## Table of content

<b>1. Abbreviation</b>	<b>5</b>
<b>2. Introduction</b>	<b>7</b>
<b>2.1 An overview of the host immune system</b>	<b>8</b>
<b>2.2 Innate immunity</b>	<b>10</b>
2.2.1 The complement system	10
2.2.2 Complement pathways	11
<b>2.3 Adaptive immunity</b>	<b>16</b>
2.3.1 T lymphocytes and cellular response	17
2.3.2 B lymphocytes and humoral production	17
<b>2.4 Pathogen: <i>Streptococcus pneumoniae</i></b>	<b>17</b>
<b>2.5 Streptococcus virulence factors and immune evasion</b>	<b>19</b>
2.5.1 Capsule	19
2.5.2 Pneumococcal-protein virulence factors	19
2.5.3 Choline binding proteins (CBPs)—PspA and PspC	21
<b>2.6 <i>Streptococcus pneumoniae</i> associated diseases</b>	<b>22</b>
<b>2.7 Focus: <i>Streptococcus pneumoniae</i>-associated hemolytic uremic syndrome</b>	<b>23</b>
<b>2.8 Aim of the study</b>	<b>24</b>
<b>3. Overview of published and submitted manuscripts</b>	<b>26</b>
<b>3.1 Manuscript I</b>	<b>26</b>
<b>Molecular analysis identifies new domains and structural differences among <i>Streptococcus pneumoniae</i> immune evasion proteins PspC and Hic</b>	<b>26</b>
<b>3.2 Manuscript II</b>	<b>27</b>
<b>Modular structure of <i>S. pneumoniae</i> surface protein A: High level of domain-based sequence diversity may qualify for molecular strain typing</b>	<b>27</b>
<b>3.3 Manuscript III</b>	<b>28</b>
<b>Choline-binding proteins of <i>Streptococcus pneumoniae</i> and their role on host cellular adhesion and damage</b>	<b>28</b>
<b>4 Manuscripts</b>	<b>29</b>
<b>5 Discussion</b>	<b>101</b>
<b>6 Summary</b>	<b>111</b>
<b>7 Bibliography</b>	<b>114</b>

<b>8</b>	<b><i>Eigenständigkeitserklärung</i></b>	<b><i>126</i></b>
<b>9</b>	<b><i>Overview of publication list, oral, poster presentations</i></b>	<b><i>127</i></b>
<b>10</b>	<b><i>Curriculum Vitae</i></b>	<b><i>128</i></b>
<b>11</b>	<b><i>Acknowledgments</i></b>	<b><i>129</i></b>
<b>12</b>	<b><i>Appendix</i></b>	<b><i>130</i></b>

## 1. Abbreviation

Ag-Ab	Antigen-antibody
aHUS	Atypical hemolytic uremic syndrome
BCR	Antigen-binding membrane receptor
C2	Complement component 2
C3	Complement component 3
C4BP	C4b-binding protein
C5	Complement component 5
C9	Complement component 9
CBD	Choline binding domain
CBM	Choline binding module
CBM	Choline-binding module
CBPs	Choline binding proteins
CBPs	Choline-binding proteins
CD55	Complement decay-accelerating factor
CD59	Membrane attack complex-inhibitory protein encoded by CD59
CFHR1	Complement Factor H related proteins 1
CFHR2	Complement Factor H related proteins 2
CFHR5	Complement Factor H related proteins 5
CP	Classical pathway
CPS	Capsular polysaccharides
CR1/CD35	Complement receptor 1
CR2/CD21	Complement receptor 2
CR3	Complement receptor 3
CR4/CD11	Complement receptor 4
D4	Domain 4
DEAP-HUS	Deficiency of CFHR (complement factor H-related) plasma proteins and autoantibody Positive form of HUS
EM	Electron microscopy
ExPRD	Extracellular proline rich domain
FDD	Family Defining domain
FHL-1	Factor H like protein
FM	Functional module
GPCRs	G protein-coupled receptors
Hic	Factor H-binding inhibitor of complement
HUS	Hemolytic–uremic syndrome
HVD	Hypervariable domain
iC3b	Inactivate C3b
IFN- $\gamma$	Interferon gamma
IgA	Immunoglobulin A
LP	Lectin pathway
LP	Lactoferrin protein
LytA	N-acetylmuramoyl-l-alanine amidase

## Abbreviation

---

LytB	Glycosaminidase
LytC	Lysozyme
MAC	Membrane attack complex
MAMPs	Microbe-associated molecular patterns
MASP	Mannose binding lectin associated protease
MBL	Mannose-binding lectin
MCP/CD46	Membrane cofactor protein
MHC	Major histocompatibility complex
NanA	Neuraminidases
NHS	Normal human serum
PCV	Pneumococcal Conjugate Vaccine
PCV13	Pneumococcal Conjugate Vaccine, Prevnar 13
PECAM-1	Platelet endothelial cell adhesion molecule
pHUS	<i>Streptococcus pneumoniae</i> -associated HUS (pHUS)
Ply	Pneumolysin
PMNs	Polymorphonuclear leukocytes
PRD	Proline rich domain
PspA	pneumococcal surface protein A
PspC	pneumococcal surface protein C
RBC	Red blood cell
RCD	Random coil domain
RCD-E	Random coil domain-extension
RD	Repeat domain
SCR	Short consensus repeat domain
slgR	Secretory IgA receptor
STEC-HUS	typical HUS caused by Shiga toxin-producing <i>Escherichia coli</i>
T-Ag	Thomsen cryptantigen
TCC	Terminal complement complex
TCR	T cell receptor
TE	Thromboembolism
TNF- $\alpha$	Tumor necrosis factor
UNICEF	United Nations Children's Fund
URT	Upper respiratory tract
VD	Variable domain
WHO	World health organization

## 2. Introduction

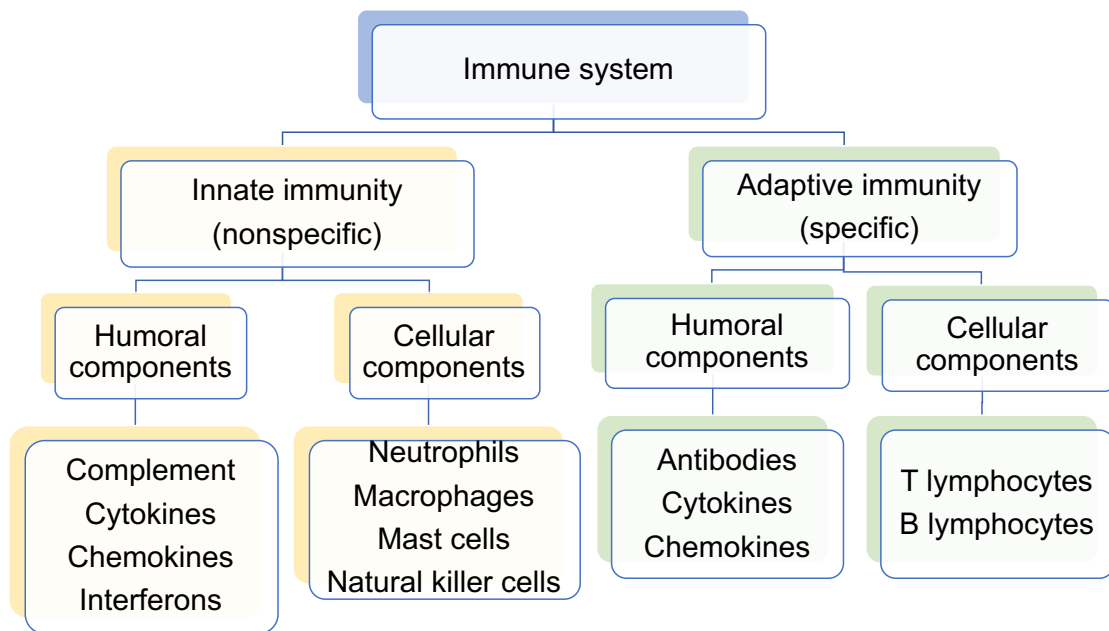
*Streptococcus pneumoniae* (also known as pneumococcus) is a lancet-shaped, Gram-positive, and classic extracellular pathogen that colonizes the mucosal surfaces of the human upper respiratory tract (URT)<sup>1</sup>. In addition to pneumonia, streptococci cause frequent infection of the middle ear and sinusitis and bronchitis by colonizing the airways<sup>2</sup>. Streptococci can invade other niches, causing invasive diseases such as conjunctivitis, meningitis, sepsis, and *Streptococcus pneumoniae*-associated hemolytic uremic syndrome (pHUS) which affects children under the age of 5<sup>3-9</sup>. Pneumonia accounts for nearly 16 percent of the 5.6 million deaths in children less than 5 years old, killing around 800,000 children in 2016 reported according to the United Nations Children's Fund (UNICEF) and the World Health Organization (WHO)<sup>10-12</sup>. *S. pneumoniae* was first isolated from the saliva of a patient with rabies in 1881 by Louis Pasteur. Already in 1911, efforts have been devoted to the vaccine's development, however the first pneumococcal vaccine was not produced and licensed for use until 1977. The first conjugate pneumococcal vaccine was licensed in 2000<sup>13</sup>. Nowadays, the widespread use of pneumococcal conjugate vaccines (PCVs) has reduced invasive disease of serotypes with the capsular polysaccharide (CPS) types<sup>14</sup>. Based on differences of the polysaccharide capsule, so far 97 serotypes are identified<sup>15</sup>. As part of its life cycle, *S. pneumoniae* remodels its genome by uptaking and incorporating of exogenous DNA from other pneumococci or viruses and has thus facilitated the spread of antibiotic resistance and evasion of vaccine-induced immunity<sup>16</sup>. For example, *S. pneumoniae* has been a leading secondary infection and cause of death during influenza pandemics<sup>17,18</sup>. The growing high burden of disease, the increase in microbial resistance to antibiotics, and insufficient vaccine coverage make it necessary to research for novel targets, to understand the diversity of immune evasion proteins, and also immune escape strategies of this relevant pathogenic bacterium<sup>19,20</sup>. As a result of the above situation, the potential new drug or vaccine targets in *S. pneumoniae* should be virulence factors common to all pneumococcal serotypes. Surface-exposed proteins are key players during the infectious process, such novel targets may be found in the bacterial cell wall which is the traditional target for antibiotics<sup>21</sup>.

Pneumococcal surface proteins PspA and PspC are abundant cell surface proteins and important virulence factors. PspA inhibits complement activation in the early phases of infection<sup>22</sup>. PspC binds several human plasma proteins, including factor H, plasminogen, C3, C4BP, secretory IgA, and vitronectin<sup>23</sup>. In a bacterial model, a strain lacking the *pspC* gene alone behaved like the wild type, but the absence of both *pspC* and *pspA* caused accelerated clearance of the bacteria<sup>24</sup>. PspA and PspC are structurally and functionally related, and both proteins combine antigenic diversity, modular composition, and mosaic structure for immune evasion. The surface proteins PspC and PspA play relevant physiological roles in bacterial viability and virulence. Both proteins represent choline binding proteins (CBPs). They are characterized by a structural organization in two modules: a functional N-terminal module (FM), and a C-terminal choline-binding module (CBM) that links the proteins to choline moieties in the cell wall. PspA is constituted of five domains: (1) a signal peptide, (2) an  $\alpha$ -helix charged domain, (3) a proline rich domain, (4) family determining domain, and (5) choline binding repeats. PspC also shows a clear multi-domain pattern, with a (1) a signal peptide, (2) hypervariable region, (3) repeat region I, (4) random coil region, (5) repeat region II, (6) a proline rich region and (7) a CBD surface anchor. Furthermore, the diversity of PspA and PspC suggests that pneumococcal strains and clinical isolates adjust antigen diversity to escape adaptive immunity<sup>25,26</sup>.

## 2.1 An overview of the host immune system

Humans and other mammals live in a world surrounded by both pathogenic and nonpathogenic microbes which are challenging immune homeostasis. Whether these organisms penetrate and cause disease is up to the balance between the virulence factors and the integrity of the host defense mechanisms<sup>27</sup>. The host immune system is divided into two major parts, which can be separated according to the specificity and speed of the reaction: innate immunity and adaptive immunity (**Figure 1**). Both systems are integrated into an interactive network of organs, cells, humoral factors, and cytokines.





**Figure 1. Overview of the immune system.** The immune system is made up of two parts: innate (nonspecific) and adaptive immunity (specific). Both systems include humoral components and cellular components.

Broadly defined, the innate immunity, which is conserved in even simple animals provides immediate host defense against infection which encompasses physical barrier, complement, and immune cells such as neutrophils, macrophages, mast cells, and natural killer cells. Physical barriers, for example, the skin and other epithelial surfaces block the entry of nonself microbes. If microbes cross the barrier, complement would activate immediately and activation fragments can cause the entry of mediate neutrophils, macrophages, and other leukocytes which ultimately kill the invading microbes<sup>28</sup>. Unlike the innate mechanisms, the adaptive immune system consists of exquisite antigen-specific cells such as T lymphocytes and B lymphocytes. Furthermore, the adaptive immune system manifests immunological memory, which is a hallmark of vertebrates<sup>29</sup>. The innate response as the first line of defense is vigorous, rapid, and nonspecific, sometimes damaging normal tissues. In contrast, the adaptive response is precise, takes several days or weeks to fully become prominent, and then antigen-specific T and B cells amplify to

expansion. However, humoral components are involved in both types of immunity as shown in **Figure 1**. Components of the complement confer to activation of the antigen-specific cells. In addition, antigen-specific cells amplify their responses also by recruiting innate components to attack invading microbes. Thus, a synergy between innate and adaptive immunity is imperative for a fully effective immune response<sup>30</sup>.

## **2.2 Innate immunity**

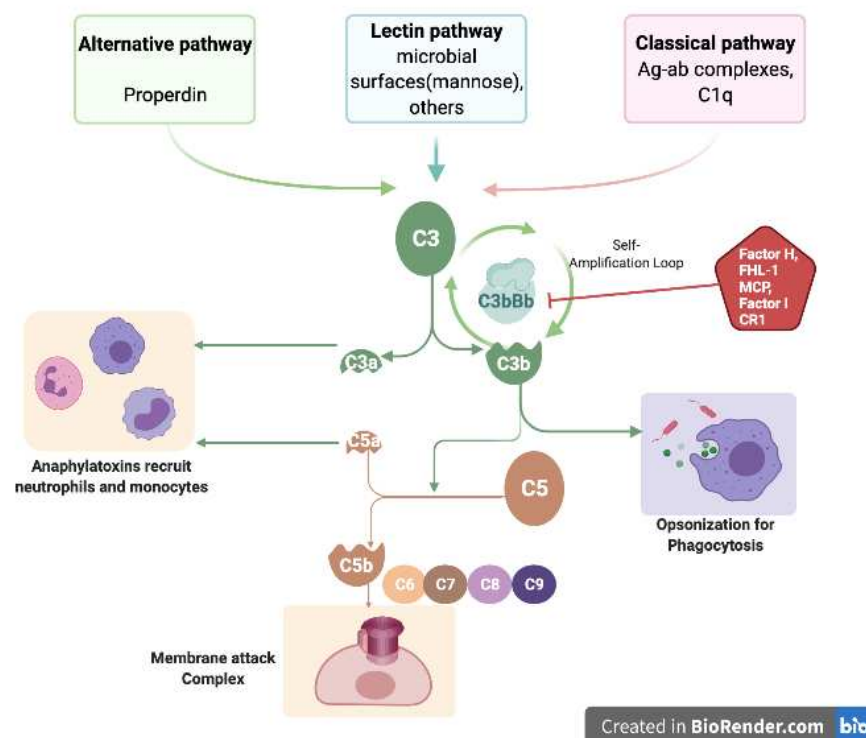
Innate immunity covers a varied series of host defenses to recognize and respond directly and immediately against infectious microbes. These include on the host side: epithelial surfaces preventing infection, secreted mucus clearance, complement activation, neutrophils recruitment, and macrophages. It is quickly activated to attack and destroy invaders by recognizing conserved structures and features which are unique to microorganisms, but which are absent in the host. These structures are generally termed microbe-associated molecular patterns (MAMPs) and represent conserved motifs such as lipopolysaccharides, teichoic acids, nucleic acids, and mannose-rich oligosaccharides<sup>31</sup>. MAMPs stimulate two major types of innate immune responses: complement response as well as induction of humoral mediators and phagocytosis by primary innate immune cells such as neutrophils and macrophages. Both of these responses can happen immediately, even if the infectious microbe has never been intruded into the host.

### **2.2.1 The complement system**

The complement system is an ancient part of immunity, which dates back over 500 million years ago as testified by genome analysis<sup>32</sup>. Complement was first recognized as a heat-labile part of human plasma, which “complemented” the action of antibodies in defense against bacteria. It consists of more than 30 distinct serum and membrane bound components which circulate in plasma, extracellular fluid and which are also expressed on the surface of the host<sup>33</sup>. Although complement as an enhancer of the antibacterial activity of antibody, lacking antibodies, it can also be triggered in the early infection stage. What is certain is that complement, on one hand, evolved as part of the innate immune system, on the other hand, it also bridges innate and adaptive immunity.

## 2.2.2 Complement pathways

Complement is also known as complement cascade. There are three complement pathways that are initiated by different reactions or media and named as follows: alternative pathway, classical pathway, and lectin pathway. The activation of three pathways produces related protease C3 convertase which represents the central stage of complement. In the end, all the three pathways trigger the terminal complement pathway (Figure 2).



**Figure 2. Complement cascades.** Figure adapted from Peter Zipfel and Skerka (2009)<sup>34</sup>. Complement is a group of soluble proteins undergoing a complex series of cleavage and combination reactions. The complement system is initiated by three pathways (alternative, lectin, and classical pathway). The major products resulting from the cleavage of center complement proteins: C3 and C5 each of which plays important roles in the complement system. C3a and C5a act as anaphylatoxins, recruiting neutrophils and monocytes; C3b serves as an important opsonin for phagocytosis and in the end, C5b combines with C6, C7, C8, C9 proteins into the membrane to form the membrane attack complex.

**Figure 2** shows an overview of complement activation and regulation. The alternative pathway is initiated by spontaneous hydrolysis of C3 or via properdin binding to the cell surface. The activation of the lectin pathway (LP) occurs by mannose-binding lectin (MBL) or other molecules. The classical pathway (CP) is initiated by C1q binding to the Ag-Ab complexes or other molecules. Subsequently, each of the three pathways forms a C3

convertase which cleaves C3 in C3a and C3b. C3a is a powerful anaphylatoxin and C3b acts as an opsonin and can be further processed into fragments which have many characteristics. Then surface deposited C3b can initiate a C5 convertase, which cleaves C5 in C5a, also acting as a strong anaphylatoxin and C5b. In the next step, C5b binds on the surface and interacts with C6-C9, forming the membrane attack complexes.

The alternative pathway is a natural defense that opsonizes and kills infectious microbes. This pathway will spontaneously be triggered either directly by C3b protein binding to a microbe, or to foreign materials, infectious agents, and also to some tumor cell lines<sup>35</sup>. The initial surface attached to C3b attracts plasma protein Factor B to form a complex. The next step is the cleavage of C3b attached to Factor B into Ba and Bb by Factor D. In the next step, Bb remains bound to the combined complex and forms activate C3 convertase which can cleave multiple C3 onto C3a and C3b. On the one hand, a large number of C3b bind additional Factor B from plasma Bb and form new C3 convertases and induce the amplification loop of the alternative pathway cascade. On the other hand, such surface attached C3b carry out other functions: (i) C3b molecules deposited on the surface of infectious particles promote the further adhesion and phagocytosis by human phagocytes; (ii) bind C3bBb to form a C3bBbC3b complex, which is a C5 convertase that cleaves C5 and can initiate the terminal complement pathway.

The lectin pathway starts when mannose binds lectin which is produced by the liver or serum ficolin binds to certain sugars exposed on the surface of microorganisms. Mannose-binding lectin-associated serine protease has several isoforms that are zymogens, MASP1, MASP2, and MASP3 which can form multimers with MBL (Mannan-binding lectin). The MASPs have similar structures to the serine proteases C1r and C1s molecules of the classical complement pathway<sup>36</sup>. MBL binding to the surface of infectious microbes induces MASP1 to cleave MASP2. Such activated MASP2 cleaves C4 into C4a and C4b, then MASP2 cleaves C2 attached to C4b into C2a and C2b. In the end, C3 convertase-C4bC2b is constituted, and induces an amplification loop of complement cascades as described for the alternative pathway (**Figure 2**).

The most notable feature of the classical pathway is that the classical pathway is mainly initiated by antigen-antibody (Ag-Ab) complexes with the antibodies of the isotypes IgG and

IgM binding to C1q protein<sup>37</sup>. The intact C1 protein complex is composed of one C1q molecule and two serine proteases C1s and two C1r molecules<sup>38</sup>. These molecules of C1 contain a C-terminal immunoglobulin receptor binding site that binds to the Fc region of the antibody. The globular C1q binds to microbial surface proteins, apoptotic cells, and other activation factors leading to a conformational change of the C1 complex. Then C1s get activation, which cleaves the C4 into C4a and C4b. Surface attached C4b will gather and bind C2, then C2 is cleaved into C2a and C2b by C1r, which allowed C3 convertase-C4bC2b complex to form. The originated C3 convertase cleaves the component C3 into C3a and C3b. The role of newly formed C3b is similar to the other two pathways and it either triggers an amplification loop of complement activity, forms a C3bBb convertase, or subsequently forms a C3bBbC3b complex, which is a C5 convertase to initiate the terminal complement pathway, or deposited on the surfaces to contribute to the phagocytosis **(Figure 2)**.

Based on the newly generated products upon complement activities, such as C3a, C3b, iC3b (inactive C3b), C3dg, C5a, C5b, and TCC (terminal complement complex), the complement as a protein-based defense system, in general, has multiple functions. The produced 10 kDa anaphylatoxins C3a and C5a interact with their specific G protein-coupled receptors (GPCRs) to recruit leukocytes and moderate leukocyte chemotaxis. C3a and C5a maintain chronic inflammation, promote an immunosuppressive microenvironment, induce angiogenesis and increase the motility and metastatic potential of cancer cells<sup>39</sup>. C3a also has anti-microbial activity<sup>40</sup>. C3b and C5b contribute to downstream complement activities. C3b can be further degraded fragments iC3b, C3dg are recognized by specific receptors that are expressed on different types of host cell surfaces, for example, complement receptor type 1 (CR1, also named CD35), CR2 (CD21), CR3 (CD11b), CR4 (CD11c). CR1 interaction with C3b inhibits B cells to secrete antibodies and thus down-regulation of B cells response<sup>41</sup>. But iC3b, C3dg bind their receptors CR2 which is mainly expressed on the surface of B cells and dendritic cells, allowing the complement system to play a role in B cell activation and maturation<sup>42</sup>. During the final phase of the complement cascade, TCC is formed and moderates the cytolysis and death of infectious particles and targeted cells. TCC was reported to induce endothelial cells to

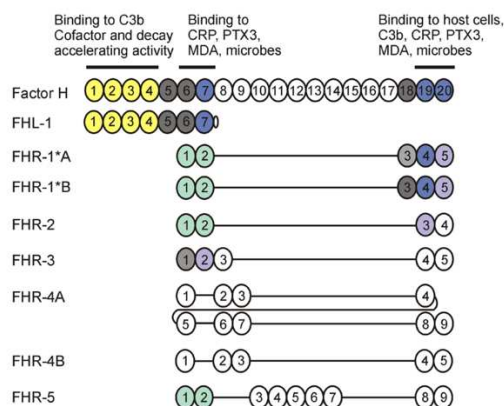
release cytokine, which indirectly mediated the migration of polymorphonuclear leukocytes (PMNs)<sup>43</sup>. TCC depositing on the microbial surfaces is suggested to cause phagocytosis of microbes by PMNs<sup>44</sup>. Overall, on the one hand, the complement system leads to elimination of invading microbes, on the other hand, these complement activities efficiently collaborated with adaptive immunity to retain host homeostasis.

### 2.2.3 Complement regulation

As a self-amplifying cascade system, complement is supervised by a cluster of complement regulators. Uncontrolled complement activation can cause many inflammatory and life-threatening conditions<sup>45</sup>. It can result in red blood cell (RBC) hemolysis, platelet activation, and thromboembolism (TE), subsequently organ impairment, and early mortality<sup>46</sup>. The host cells and extracellular matrices are protected from complement attack by complement regulators.

Factor H is a pivotal controller to inhibit the alternative pathway and the amplification loop and is best characterized member of the Factor H protein family. It is a soluble glycoprotein that circulates in the human plasma with a concentration of 200-300 µg/ml and is also localized on the cell surface<sup>47,48</sup>. The Factor H protein family is composed of nine homologous and immunologically-related proteins: Factor H, Factor H like protein (FHL-1) that is an alternative transcript of *CFH* gene, and the complement related proteins (CFHR1-CFHR5) derived from *CFHR* genes which are adjacent to *CFH* gene (**Figure 3**)<sup>49,50</sup>. The individual member comprises a different number of short consensus repeat domains (SCRs) that connect by short linkers. These SCRs are approximately 60 amino acids long with four conserved cysteine residues. As indicated in **Figure 3**, The head of Factor H and FHL-1 includes the seven N-terminal SCRs with associated ligand binding capacities. FHL-1 shares SCR1-4 domains with Factor H to bind ligand C3b that facilitating the decay of C3 convertase by displacing the bound Bb. Factor H and FHL-1 as cofactors for Factor I to mediate C3b inactivation by cleaving C3b into inactivate C3b (iC3b) and C3f<sup>51</sup>. But the C-terminal of Factor H SCR19-20 domains with the binding sites for C3b/C3d, heparin, host cells, and certain ligands limit and restrict complement attack. Mutations of the C-

terminus of Factor H and FHL-1 result in uncontrolled activation of the alternative pathway and cause autoimmune diseases such as aHUS (atypical hemolytic uremic syndrome)<sup>52</sup>.



**Figure 3. The human factor H (FH) protein family.** Figure from Józsi (2017)<sup>50</sup>. Complement-regulatory functions and binding domains within Factor H and F Factor H-related (FHR) proteins. FH-like protein 1 (FHL-1) is a cleaved variant of Factor H. Colors indicate domains identical between Factor H and FHRs; High sequence similarity (>80% identity) but not equal to 100% identity indicated by light shades. The green domains are closely related to each other but distant from Factor H and mediate dimerization of FHR-1, FHR-2, and FHR-5. Interaction domains or other protein binding sites in Factor H are shown by horizontal lines.

The five *CFHR* genes code seven variant FHR proteins with structural homology to Factor H but lack the domains SCR1-4 which are responsible for inhibiting complement. Each protein has distinct functions to adjust the complement activities. FHR-1 was present in two forms differing in the third SCR with a different number of carbohydrate side-chain<sup>49</sup>. There are several studies on FHRs to inspect their functions in complement. FHR-1 binds C5 to inhibit C5 convertase activity and prevent assembly of membrane attack complex<sup>53</sup>. FHR2 was reported to inhibit the C3 convertase and TCC assembly<sup>54</sup>. FHR-3 and FHR-4 as cofactors bind the C3d region of C3b to enhance the activity of Factor H<sup>55</sup>. FHR-5 also has cofactor activity with Factor H to inhibit C3 convertase activity<sup>56</sup>. FHR-5 interacts with extracellular matrix to compete with Factor H binding, but FHR-5 also allowed C3 convertase to form and promote complement activation<sup>57</sup>. Moreover, FHR-1, FHR-2, and FHR-5 were recognized to form homodimerize, and FHR-1 combines FHR-2 to form heterodimers, consequently increasing their ability for binding C3b and increasing competition with Factor H<sup>58</sup>. Mutations, genetic deletions, duplications, or rearrangements

in a *CFHR* gene are related to diseases. Mutations of FHR-5 are associated with familial C3 glomerulopathy<sup>59</sup>. FHR-1 and FHR-3 deficiency increase the risk of causing aHUS<sup>60</sup>.

Besides Factor H family proteins, there are kinds of complement regulators were discovered: C1 inhibitor, C4b-binding protein, membrane bound regulators, vitronectin, and clusterin. Both C1 inhibitor and C4b-binding proteins are fluid-phase regulators that circulate in the blood to mediate the lectin and classical pathway. C1 inhibitor binding to C1 complex inactivates C1r and C1s proteases. Since the similarity between MASPs and C1r, C1s molecules, C1 inhibitor also prevents MASPs proteases spontaneous activation<sup>61</sup>. C4b-binding protein (C4BP) can bind C4b and C3b to accelerate the decay of C3 convertase<sup>62</sup>. As the complement inhibitor, C4BP allows Factor I to cleave and inactivate C4b. A variety of pathogens capture human C4BP to establish infections<sup>63</sup>.

Membrane-bound regulators include membrane cofactor protein (MCP/CD46), complement receptor 1 (CR1/CD35), CD55, and CD59. MCP as a cofactor for Factor I inactivate the C3b and C4b deposited on the target membranes and it is widely distributed on different types of human cells. As Factor H and FHL-1 do, mutations in CD46 also predispose to develop aHUS<sup>64</sup>. CD55 also known as C3b/C4b receptor regulates the assembly of C3 convertases. Moreover, CD59 is known as MAC-inhibitory protein by blocking the binding of C9 polymerization and MAC formation<sup>65</sup>.

Vitronectin and clusterin are terminal complement pathway inhibitors that are circulating in the human blood. Vitronectin blocks the C5b-C7 complex to insert into the cell membranes, thereby stopping the membrane-damaging clusterin effect of the terminal complement pathway<sup>66,67</sup>.

### **2.3 Adaptive immunity**

The adaptive immune system of the vertebrate host is also defined as the acquired immune system. Unlike innate immunity, adaptive immunity provides a highly specific and broader recognition system to differentiate and eliminate nonself-antigens. After initial exposure to the infectious microbes, the adaptive immunity generates antigen-specific cells combined with immunological memory some of which can persist in the host for over decades. The second response is quicker and more effective. The cellular response of adaptive immune



involves two groups of cells: the T lymphocytes and the B lymphocytes, which produce in the thymus and the bone marrow respectively. Lymphocytes accounting for 20-40% of white blood cells, mainly perform two activities: antibody generation and cell-mediated response.

### **2.3.1 T lymphocytes and cellular response**

T cells are produced by stem cells in the bone marrow and circulate to the thymus. Then T cells develop further by expressing specific membrane receptors for antigens which are called T cell receptors (TCR). TCR recognizes the antigenic peptide in combination with major histocompatibility complex (MHC) molecules which locate on the cell surface of antigen presenting cells. Then T cells proliferate and can phagocyte quickly to clear infectious microbes away, a small part of them survive and differentiate into memory T cells. The cells also produce or secrete cytokines and chemokines.

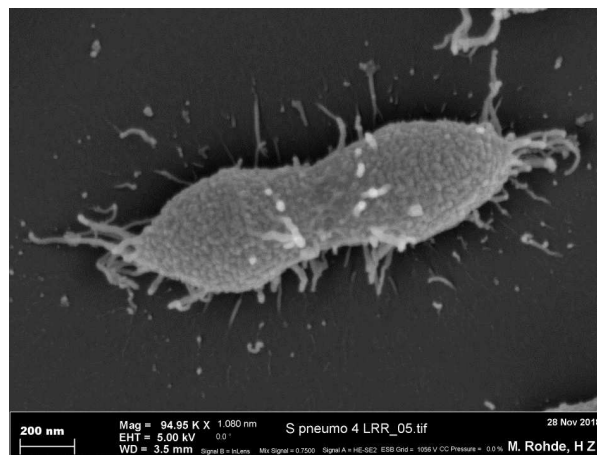
### **2.3.2 B lymphocytes and humoral production**

Once the signal is received from T cells and other cells, B cells are primarily responsible for the production of antibodies that when released circulate in the blood and lymph system. Similar to T cells, B cells express a unique antigen-binding membrane receptor (BCR) and secrete antigen specific antibodies. Generally speaking, each clone of B cells expresses a specific BCR that can recognize and bind to a distinct antigen. Then antigen-antibody complex can be the target for phagocytes and trigger the classical complement pathway. About 10% of these antibodies secreted B cells survive and develop long-lived antigen-specific memory B cells. Such memory B cells can respond quickly when microbes with the same antigen infect the host again<sup>68,69</sup>.

## **2.4 Pathogen: *Streptococcus pneumoniae***

*Streptococcus pneumoniae* (*S. pneumoniae*) is a lancet-shaped, Gram-positive, and extracellular diplococci, that can also form as single cocci or in short chains of cocci. *S. pneumoniae* is classified as alpha-hemolytic bacteria, according to a greenish halo that surrounds colonies when grown aerobically on blood agar plates. Single pneumococcus is between 0.5 and 1.25 micrometers in diameter. *S. pneumoniae* does not form spores and

are non-motile, though they sometimes have pili used for adherence<sup>70</sup> (**Figure 4**). The outmost layer of *S. pneumoniae* is a negative charge polysaccharide capsule that completely encloses the cell, followed by a few layers of peptidoglycan which form the cell wall and which consist of teichoic acid attached to every third N-acetylmuramic acid, and is around 6 layers thick. Lipoteichoic acid is attached to the cell membrane via a lipid moiety, and both teichoic and lipoteichoic acid contains phosphorylcholine. *S. pneumoniae* D39, the virulent reference strain which is a historically important serotype 2 strain that was used in experiments by Avery to demonstrate that DNA is the genetic material. D39 was reported that codes for 1914 proteins and 73 RNAs<sup>71,72</sup>. More than 500 of these are different surface proteins. A notable group is the family of choline-binding proteins (CBPs) which represent pneumococcal virulence attributes, that share common anchor. Choline binding proteins attach to the choline residues present in the cell wall through non-covalent interactions. Pneumococcal CBPs include cell wall hydrolases, adhesins, and other virulence factors, all playing relevant physiological roles for bacterial viability and important determinants virulence, including PspA, PspC, LytA, LytB, and LytC (autolysins)<sup>73</sup>.



**Figure 4** *S. pneumoniae* observed by raster electron microscopy (EM). One representative figure of *S. pneumoniae* derived from a patient with HUS is shown. Photos were taken by our collaborator Dr. Manfred Rohde, Helmholtz center for infection research in Braunschweig. Scale as 200nm.

*S. pneumoniae* has shown a significant increase in antibiotic resistance over the past 20 years. This is likely due to its natural transformation system used for genetic exchange. *S. pneumoniae* can also develop resistance to antibiotics through mutation and natural selection.

## 2.5 Streptococcus virulence factors and immune evasion

To conquer and evade immune systems such as the complement system, *S. pneumoniae* has evolved a wide variety of virulence factors to colonize the host, invade tissues, and impair complement activity<sup>74,75</sup>.

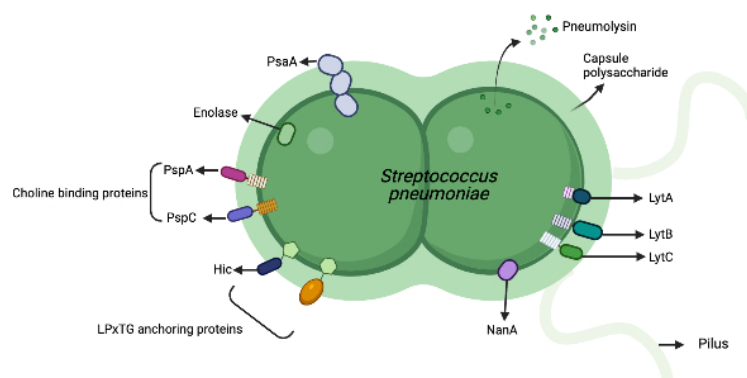
### 2.5.1 Capsule

The first important virulence factor for *S. pneumoniae* is the extracellular capsule, a layer consisting of chains of monosaccharides that surrounds the bacteria. Serotypes are dictated by the order and type of monosaccharide units within the polysaccharide chain and by different side branches. *S. pneumoniae* expresses a thick polysaccharide capsule to protect the pathogen from opsonophagocytosis and prevent entrapment in the nasal mucus. Thereby allowing access to epithelial surfaces, of which there are at least 97 antigenically distinct serotypes<sup>15,20,76</sup>. In addition, the polysaccharide capsule in conjunction with microbial surface molecules binds to human complement components and thus fixes them with a functionally activated state<sup>21,77</sup>.

### 2.5.2 Pneumococcal-protein virulence factors

Secondly, *S. pneumoniae* expresses more than 500 different surface proteins and toxins that drive pathogenesis. The major virulence factors are highlighted in **figure 5**. Major pneumococcal-protein virulence factors, such as pneumolysin (Ply), neuraminidases (NanA), the metal-ion-binding proteins, LPsTG anchoring proteins, and choline-binding proteins (CBPs) have specific roles in respiratory colonization and disease<sup>4</sup>. Ply is localized in the cytoplasm and is secreted into the environment of the bacteria by the activity of LytA. It belongs to the family of cholesterol-dependent cytolysins. Ply has lytic effects on many mammalian cell types, including complement activation<sup>78</sup>. The neuraminidase of *S. pneumoniae*, which catalyzes the release of terminal sialic acid residues from glycoconjugates, is involved in biofilm formation<sup>79</sup>. Immunization with native or recombinant NanA has also been demonstrated to afford protection against nasopharyngeal colonization and chinchilla otitis media mode<sup>80</sup>.

Choline binding proteins (CBPs) represent a notable group of pneumococcal virulent determinants, that is characterized by a structural composition in two separate regions: an N-term functional module (FM), and a C-term choline-binding module (CBM) that noncovalently attaches to the phosphorylcholine moiety of the cell wall<sup>73</sup>. Pneumococcal CBPs include cell wall hydrolases, adhesins, and other virulence factors, all playing relevant physiological roles for bacterial viability and virulence, which are anti-complement and essential to the progression of diseases such as PspA, PspC, LytA, LytB, and LytC. *S. pneumoniae* is a very fragile bacterium, the enzymatic (autolysin- autolytic enzyme, LytA, LytB, and LytC) within itself has the ability to enzymatically disrupt and autolysis itself. LytA the major pneumococcal autolysin, activates during the stationary phase (the phase in which growth slows due to exhaustion of available nutrients and the buildup of toxins) or under penicillin treatment<sup>81</sup>. Autolysis usually begins within 18-24 hours, with colonies collapsing in the centers and cell wall remodeling in optimal conditions<sup>70</sup>. Moreover, LytA is essential in the evasion of the complement-mediated immunity by inhibiting C3 convertase formation and reducing iC3b opsonization on the bacterial surface<sup>82</sup>. LytB is a pneumococcal glucosaminidase, that separates daughter cells, the final event of cell-division cycle<sup>83</sup> which was reported specifically located at the septum of the dividing cells. LytB and LytC are involved in fratricide<sup>84</sup>, biofilm formation<sup>85</sup>, and LytC behaves as an autolysin at 30°C and the fact that it has its maximum enzymatic activity at this temperature suggests that it might be more crucial in the upper respiratory tract.



**Figure 5 *S. pneumoniae* protein virulence factors.** Figure modified from Aras Kadioglu<sup>4</sup>. Major pneumococcal virulence factors include the thick capsule, choline-binding proteins (PspA, PspC LytA, LytB, and LytC), enolase (Eno), the LPXTG-anchored proteins (neuraminidase and others), pneumolysin, and the metal-binding proteins pneumococcal surface antigen A (PsaA).

### 2.5.3 Choline binding proteins (CBPs)—PspA and PspC

Pneumococcal surface proteins A (PspA) and C (PspC) as important virulence factors are the most abundant cell surface proteins of *S. pneumoniae*. Based on sequence homology, PspA is a paralogue of PspC. PspA and PspC are structurally related and are comprised of an N-terminal  $\alpha$ -helical domain, a proline-rich domain, and a choline-binding domain<sup>86</sup>. PspA and PspC both have a domain composition, the proteins are highly polymorphic, which promotes immune evasion and produces a benefit for the bacteria<sup>87</sup>. PspA and PspC are promising candidates for the generation of more effective vaccines to overcome the limitations of polysaccharide-based vaccines. PspA and PspC will be studied and discussed individually in this work.

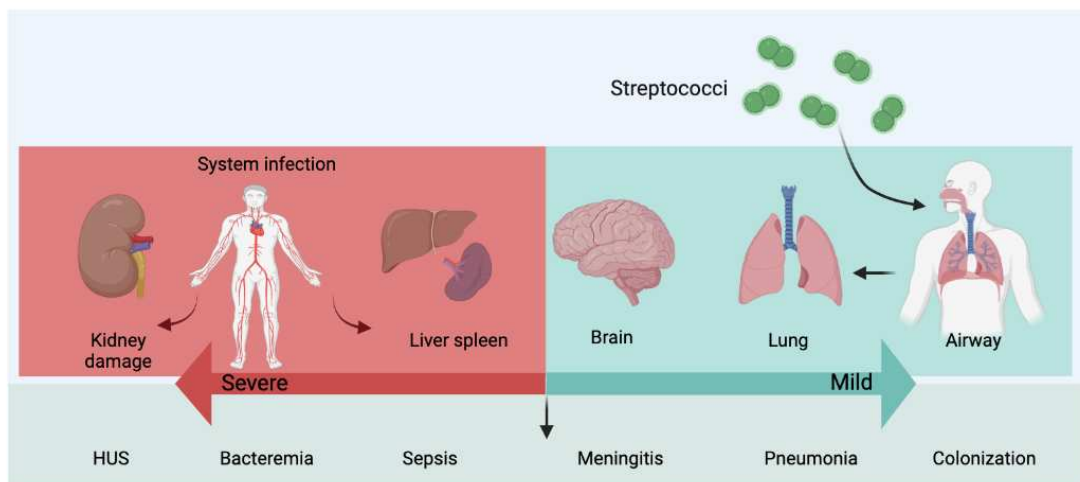
The most extensively studied molecule is pneumococcal surface protein A (PspA). PspA is shown to be essential for pneumococcal evasion and to elicit protection of *S. pneumoniae*<sup>88–90</sup>. Mutagenesis of *pspA* in *S. pneumoniae* D39 significantly reduce virulence, and PspA elicits protection in mice against fatal bacteremia and sepsis caused by genetically diverse pneumococci and protects against carriage and lung infection indicating that *pspA* contributes to pathogenesis<sup>91–93</sup>. PspA binding human lactoferrin offers protection against killing by apolactoferrin, and antibody to PspA enhance the killing of pneumococci by apolactoferrin<sup>89,94</sup>. PspA of *S. pneumoniae* WU2 inhibits complement C3 deposition via the classical pathway<sup>95</sup>. With high genetic variability, PspA is a serologically variable protein and is strain-specific. Nearly all pneumococcal strains contain at least one other locus with sequence homology to *pspA*<sup>96</sup>. Susan K. Hollingshead reported that 24 *pspA* gene sequences from unrelated strains separate into two major families and that the proteins show a high-order divergence<sup>25</sup>. Domain organization of PspA shows an N-terminal  $\alpha$  helical domain, a clade-defining region, a proline rich, region and choline-binding repeats. In addition, a human lactoferrin binding domain is located at the N-terminal end or is contained within the clade-defining region<sup>94</sup>.

The pneumococcal surface protein C gene (*pspC*) of *S. pneumoniae* encodes PspC (also known as CbpA, PbcA, and SpsA), which is present in approximately 75% of all *S. pneumoniae* strains. This immune evasion protein binds human complement factor H to

prevent alternative complement pathway activation and opsonophagocytosis<sup>97,98</sup>. Additionally, PspC binds secretory immunoglobulin A (IgA), via a hexapeptide motif located in the N-terminal region to provide a role in adhesion<sup>99</sup>. And PspC interacts with the C-terminal heparin-binding domain of vitronectin which also required R domain located in the N-terminal region of PspC. Consequently, secretory IgA competitively inhibits binding of vitronectin to PspC<sup>100</sup>. PspC binds human complement C3 as one of its specific substrates on epithelial cells, and induces IL-8 release from pulmonary epithelial cells, thus showing that PspC acts as two sites, as surface bound and secreted by pneumococci<sup>101, 102</sup>. Clinical isolates of *S. pneumoniae* bind plasminogen via PspC and the strains use activate plasmin to damage human endothelial cells and thereby induce exposure to underlying matrix<sup>103</sup>. Compared with PspA, PspC is a more polymorphic and strain-specific immune evasion protein identified to date<sup>104</sup>. PspCs vary in size and mass, ranging from 503-929 aa and a molecular mass of 65-110 kDa. PspC comprises distinct regions that consist of multiple domains. Furthermore, a substantial overlap of domains exists between PspC and Hic (Factor H binding inhibitor of complement) variants<sup>105</sup>.

## **2.6 *Streptococcus pneumoniae* associated diseases**

Pneumococcal disease occurs around the world and it represents mortal danger, especially in developing countries where fewer people have access to pneumococcal vaccine. Most of these deaths occur in countries in Africa and Asia. The World Health Organization (WHO) estimates that *S. pneumoniae* kills close to half a million children under 5 years old worldwide per year. Besides causing severe illness in children, *S. pneumoniae* also infected the elderly and other people with weakened immune systems. Upon infection *S. pneumoniae* disseminate to human organs such as: sinuses, lungs, ears, blood, brain and other normally sterile sites (sites where it is not commonly found) to cause pneumonia, otitis media, meningitis (inflammation of the coverings of the brain and spinal cord), arthritis, sepsis and hemolytic uremic syndrome (**Figure 6**)<sup>103,106–108</sup>.



**Figure 6** *S. pneumoniae* infection and diseases.

*S. pneumoniae* transmits and colonizes the mucosa of the upper respiratory tract (URT). *S. pneumoniae* invades and infects the host organ (lung, brain, liver, and spleen), causing pneumonia, and meningitis. When *S. pneumoniae* invades the blood system and causes systemic infections (bacteremia, sepsis, and hemolytic uremic system-HUS).

## 2.7 Focus: *Streptococcus pneumoniae*-associated hemolytic uremic syndrome

Hemolytic uremic syndrome (HUS) is defined by acute kidney injury, microangiopathic hemolytic anemia, and thrombocytopenia occurring after infections. HUS can also be genetic dysregulation called aHUS, and autoimmune in origin (DEAP-HUS) which are associated with defective complement regulation<sup>109,110</sup>. HUS in children is usually caused by Shiga-like toxin-producing *E. coli* (STEC-HUS). However, HUS is also a complication of invasive pneumococcal infection known as *streptococcus pneumoniae*-associated HUS (pHUS) with high morbidity and mortality<sup>6,8</sup>. 5 to 15% of HUS cases are related to *S. pneumoniae* infection, most often meningitis or pneumonia. Although the introduction of PCV13 and an overall decrease in the incidence of invasive pneumococcal disease in children, the incidence of pHUS cases is rising<sup>111–113</sup>. The pathogenesis of pHUS is not yet completely resolved. Pneumococcal neuraminidase transiently desialylated Factor H which can reduce its capacity to control complement activation<sup>114</sup>. Pneumococcal neuraminidase is also thought to splits off neuraminic acid from the glycoproteins present on the surface of red cells, platelets, thrombocytes, and endothelial cells, and thus exposes the hidden Thomsen cryptantigen (T-Ag)<sup>115</sup>. The T-Ag can then react with a complement-

fixing antibody of the IgM class which is present in all human plasmas after the age of 6 months<sup>116</sup>. This specific IgM may activate the complement system and coagulation, in the end, leading to damage of endothelial cells in the blood vessels and disturbing local complement homeostasis. Under certain conditions, pneumococci are able to cross the epithelial barrier, leading to dissemination of the bacteria into the bloodstream and potential tissues. Thereby pHUS promote a thrombogenic state that drives HUS pathology<sup>6,113</sup>.

## **2.8 Aim of the study**

*S. pneumoniae* has a relatively fast growth rate and can reach high cell densities in infections environments and can cause severe diseases, like hemolytic uremic syndrome (HUS). As part of its life cycle, *S. pneumoniae* remodel the genome by taking up and incorporating of exogenous DNA from other pneumococci or viruses. This can facilitate the spread of antibiotic resistance and evasion of vaccine-induced immunity and makes it necessary to search for novel targets, to understand the diversity, as well as the immune escape strategies of this pathogenic bacterium. *S. pneumoniae* has developed different strategies to evade or limit complement mediated opsonization and subsequent phagocytosis. Furthermore, sequence variation suggests that the two immune evasion proteins PspA and PspC are important for interaction of pneumococci with the host. Given the multifunctional characteristics and mosaic structure of PspA and PspC, it is important to investigate the domain composition of the proteins among different strains and in particular among clinical isolates.

*S. pneumoniae* can induce pneumococcal hemolytic uremic syndrome (HUS). To characterize the role of HUS inducing strains, we evaluated 48 *S. pneumoniae* strains isolated from patients. These diseases associated isolates, Sp-HUS show strong complement resistance when challenged with complement active human serum. Sp-HUS strains show lower levels of surface C3 deposition, as compared to a pathogenic, reference strain D39. Consequently Sp-HUS strains evade host complement rather efficiently. In addition, I show that Sp-HUS strains have specific PspA and PspC variants which include unique domain profiles. By evaluating complement resistance of Sp-HUS, PspA interacting



with the human complement regulator C3 and PspC binding with human Factor H together assisted Sp-HUS to resist and evade from the complement. During this procedure, I studied the location of PspA and PspC on the bacterial surface and compared the surface distribution of the proteins with bound human complement regulators and deposited C3b. By identifying such unique features of clinical *S. pneumoniae* strains derived from Sp- HUS strains, I can show their efficient role in complement evasion and disease pathology.

### 3. Overview of published and submitted manuscripts

#### 3.1 Manuscript I

##### **Molecular analysis identifies new domains and structural differences among *Streptococcus pneumoniae* immune evasion proteins PspC and Hic**

**Shanshan Du**, Cludiá Vilhena, Samantha J. King, Alfredo Sahagún-Ruiz, Sven Hammerschmidt, Christine Skerka, Peter F. Zipfel. Scientific reports. 2021 Jan 11(1071):1-15.

Major aspects of the manuscript

In this manuscript, we combined sequence comparison and domain structure evaluation. This novel strategy improved understanding of individual PspC and Hic proteins variability and modular domain composition, enabled a structural and functional characterization at the domain level and revealed substantial structural differences between PspC and Hic proteins. We identified nine new domains and new domain alternates. Several domains are unique to PspC and Hic variants, while other domains are also present in other virulence factors encoded by pneumococci and other bacterial pathogens. The strategy we used to define the diversity of PspC and Hic could also apply to other pneumococcal virulence factors, which will increase understanding of their roles in immune evasion and provide important information for molecular strain typing and vaccine design.

Own contribution and contribution of the coauthors to the manuscript

**Shanshan Du** has planned, performed, and interpreted the following analyses: homology and sequence analysis of PspC and Hic; amino acid identity of the full-length selected cluster variants; secondary structural and domain composition analysis of PspC and Hic; manuscript preparation.

Cludiá Vilhena, Samantha J. King, Alfredo Sahagún-Ruiz, Sven Hammerschmidt and Christine Skerka interpret the data and correct the language.

Peter F. Zipfel designed the study, interpreted the results and wrote the manuscript.

### 3.2 Manuscript II

#### **Modular structure of *S. pneumoniae* surface protein A: High level of domain-based sequence diversity may qualify for molecular strain typing**

**Shanshan Du**, Cludiá Vilhena, Dorit Fuest, Monika von der Heide, Sven Hammerschmidt, Christine Skerka, Peter F. Zipfel. (Manuscript in preparation)

Major aspects of the manuscript

In this study, we evaluated the domain pattern and sequence analyses among PspA proteins derived from 48 different *S. pneumoniae* strains. Then we defined three groups of PspA proteins and based on domain patterning further subgroups were identified in the family I and III. PspA proteins have three separated regions: an N-terminal  $\alpha$ -helical region, a coiled-coil structured middle segment followed by a  $\beta$ -sheet structured region. We applied family specific domain pattern features to 9 PspA proteins encode by clinical *S. pneumoniae* strains which derived from young infant patients with pneumococcal hemolytic uremic syndrome. We found that clinical PspA proteins identified in PspA family II and III but not family I, which suggests PspA<sub>HUS</sub> are clade predominate. This combined pattern based and sequence variability seems to extend to the strain level. It suggests that PspA analyses may qualify for molecular strain typing.

Own contribution and contribution of the coauthors to the manuscript

**Shanshan Du** has designed, performed, and interpreted the following analyses: PspA sequences collection and removal of redundant sequences; homology and sequence analysis of PspA; amino acid identity of the full-length selected cluster variants; secondary structural and domain composition analysis of PspA; human lactoferrin binding assay; manuscript preparation.

Cludiá Vilhena interpreted the results.

Monika von der Heide has performed PspA sequencing experiment.

Dorit Fuest collected the target sequence from the genomic data.

Sven Hammerschmidt and Christine Skerka interpreted the data and corrected the language.

Peter F. Zipfel has designed the study, interpreted the results, and wrote the manuscript.

### 3.3 Manuscript III

#### **Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage**

Cláudia Vilhena, **Shanshan Du**, Miriana Battista, Martin Westermann, Thomas Kohler, Sven Hammerschmidt, Peter F. Zipfel. (Manuscript in revision)

Major aspects of the manuscript

In this manuscript, we investigated the role of three choline-binding proteins, PspC, PspA, and *lytA* on adhesion and pathogenicity and their interaction with the cell wall of *S. pneumoniae*. Compared to *S. pneumoniae* D39 as reference strain, *lytA*, *pspC* or *pspA* genes knockout mutant strains were evaluated for testing their capacity to adhere to surfaces by performing in vitro biofilm formation assays. We observed a key role of LytA as a robust synthesis of the biofilm matrix formed in *lytA* mutant. The CBPs PspA and PspC were crucial for the hemolysis capacity of *S. pneumoniae* towards human red blood cells. Furthermore, damage to endothelial cells was decreased in all three tested knockout strains, suggesting their significant relevance for the overall pathogenicity of *S. pneumoniae*.

Own contribution and contribution of the coauthors to the manuscript

**Shanshan Du** has planned, performed, and interpreted the following experiment: set up the static biofilm model, biofilm imaging by confocal laser scanning microscopy assays, phagocytosis assay and, biofilm quantification.

Cláudia Vilhena has designed, performed, and interpreted the following experiment: hemolysis assays, biofilm imaging by scanning electron microscopy, and bacteria adhesion with endothelial and epithelial cell assay. Wrote the manuscript.

Miriana Battista has performed cytotoxicity assays and interpreted the data.

Thomas Kohler has built the mutant strains.

Martin Westermann and Sven Hammerschmidt have written the manuscript.

Peter F. Zipfel has interpreted the data and written the manuscript.

## 4 Manuscripts

1 Molecular analysis identifies new domains and structural differences among *Streptococcus pneumoniae* immune evasion proteins PspC and Hic

www.nature.com/scientificreports

## scientific reports



OPEN

Molecular analyses identifies new domains and structural differences among *Streptococcus pneumoniae* immune evasion proteins PspC and HicShanshan Du<sup>1</sup>, Cláudia Vilhena<sup>1</sup>, Samantha King<sup>2,3</sup>, Alfredo Sahagún-Ruiz<sup>1,4</sup>, Sven Hammerschmidt<sup>5</sup>, Christine Skerka<sup>1</sup> & Peter F. Zipfel<sup>1,6</sup>✉

The PspC and Hic proteins of *Streptococcus pneumoniae* are some of the most variable microbial immune evasion proteins identified to date. Due to structural similarities and conserved binding profiles, it was assumed for a long time that these pneumococcal surface proteins represent a protein family comprised of eleven subgroups. Recently, however, the evaluation of more proteins revealed a greater diversity of individual proteins. In contrast to previous assumptions a pattern evaluation of six PspC and five Hic variants, each representing one of the previously defined subgroups, revealed distinct structural and likely functionally regions of the proteins, and identified nine new domains and new domain alternates. Several domains are unique to PspC and Hic variants, while other domains are also present in other virulence factors encoded by pneumococci and other bacterial pathogens. This knowledge improved pattern evaluation at the level of full-length proteins, allowed a sequence comparison at the domain level and identified domains with a modular composition. This novel strategy increased understanding of individual proteins variability and modular domain composition, enabled a structural and functional characterization at the domain level and furthermore revealed substantial structural differences between PspC and Hic proteins. Given the exceptional genomic diversity of the multifunctional PspC and Hic proteins a detailed structural and functional evaluation need to be performed at the strain level. Such knowledge will also be useful for molecular strain typing and characterizing PspC and Hic proteins from new clinical *S. pneumoniae* strains.

**The pathobiont *Streptococcus pneumoniae*.** *S. pneumoniae* (the pneumococcus) is the leading cause of community-acquired pneumonia. In addition, this Gram-positive pathogen can cause otitis media and may also cause acute life-threatening invasive infections such as sepsis and meningitis<sup>1-4</sup>. Malnutrition and *S. pneumoniae* infections are the major cause of childhood mortality worldwide. Pneumonia accounts for approximately 16 percent of the 5.6 millions of deaths among children under five years of age, killing around 808,000 children in 2016 according to the United Nations Children's Fund (UNICEF) and the World Health Organization (WHO)<sup>5-7</sup>. At any point in time pneumococci can reside asymptotically in the upper respiratory tract of about 50% of children, from where they can spread to other sites and cause disease or be transmitted to other individuals<sup>8</sup>. Based on the differences in the polysaccharide capsule 100 *S. pneumoniae* serotypes have been identified so far<sup>9,10</sup>.

<sup>1</sup>Department of Infection Biology, Leibniz Institute for Natural Product Research and Infection Biology, Jena, Germany. <sup>2</sup>Center for Microbial Pathogenesis, Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, OH, USA. <sup>3</sup>Department of Pediatrics, The Ohio State University, Columbus, OH, USA. <sup>4</sup>Molecular Immunology Laboratory, Department of Microbiology and Immunology, Faculty of Veterinary Medicine and Animal Husbandry, National Autonomous University of Mexico, Mexico City, Mexico. <sup>5</sup>Department of Molecular Genetics and Infection Biology, Interfaculty Institute for Genetics and Functional Genomics, Center for Functional Genomics of Microbes, University of Greifswald, Greifswald, Germany. <sup>6</sup>Institute of Microbiology, Friedrich-Schiller-University, Jena, Germany. ✉email: peter.zipfel@leibniz-hki.de

Pneumococcal diseases are widespread and antibiotic resistant strains are constantly emerging resulting in a need for new therapeutics. In addition, currently available vaccines are based on the capsular polysaccharide and only provide protection against the limited number of serotypes included. Vaccines protecting against a higher number of serotypes or a serotype-independent vaccine is needed to combat the pathogen efficiently. These limitations make it important to identify new virulence determinants that may serve as novel vaccine or therapeutic targets, to understand the diversity of these determinants and also to define the immune escape strategies of this pathogenic bacterium<sup>1,11,12</sup>.

Immune and in particular complement evasion is critical for all pathogenic microbes, including *S. pneumoniae*. Common mechanisms of complement evasion are emerging as a large list of pathogenic microbes bind and exploit the same human complement regulators<sup>13–17</sup>. Thus, it is important to understand the exact role of individual pneumococcal virulence determinants in complement and immune evasion. Furthermore, it is important to establish whether the virulence determinants are localized to the surface and, if so, the specific regions of the protein exposed<sup>18–21</sup>.

**PspC and Hic proteins as central pneumococcal immune evasion proteins.** The PspC and Hic proteins are important pneumococcal immune evasion proteins and adhesins that represent promising vaccine candidates<sup>22</sup>. The majority of virulent *S. pneumoniae* strains express at least one PspC or Hic variant, and strains that have the *pspC/hic* genes deleted show significant amelioration of lung infection, nasopharyngeal colonization, and bacteremia in mice<sup>23</sup>.

Based on overall sequence similarities PspC and Hic variants are considered to represent one group of pneumococcal immune evasion proteins. Initial analyses by Brooks Walter in 1999 and Iannelli et al. in 2001 revealed both sequence similarity and diversity among PspC and Hic proteins<sup>24,25</sup>. Iannelli et al. identified several domains within the 43 PspC and Hic proteins evaluated including, the leader peptide,  $\alpha$ -helical regions with a seven-amino acid periodicity, repeat domains and a proline-rich stretch followed by either a choline-binding or sortase-dependent anchor<sup>26</sup>. At that time, the cell wall anchors were used as the criterion to differentiate between PspC and Hic family members and based on sequence differences six PspC-type and five Hic-type clusters were defined. However, today there are still no precise criteria regarding cluster specific domain composition or domain characteristics. Because the patterns of domains are not exactly known and the borders of individual domains are not well-defined, a straightforward system of variant designation is at present difficult to achieve. This makes assignment of existing and newly identified *pspC* and *hic* genes, including those from novel clinical pneumococcal isolates, difficult or even impossible<sup>27</sup>.

Initially, PspC was identified as an adhesin, which targets the secretory component of secretory Immunoglobulin A (sIgA) and the polymeric IgA receptor (pIgR)<sup>28</sup>. Because *pspC* and *hic* genes were identified independently by multiple groups, different names were given, including CbpA (choline-binding protein A), SpsA (secretory IgA binding protein), PbcA (C3-binding protein A), or Hic (Factor H binding inhibitor of complement) (Table 1)<sup>29–39</sup>. Over time *pspC* and *hic* have become the favored nomenclature.

PspC and Hic proteins are attached to the bacterial cell wall. PspC proteins attach non-covalently to the phosphorylcholine (PCho) moiety of teichoic acids (TAs) via their C-terminal choline binding domains and Hic proteins, are covalently linked to the peptidoglycan via an LPsTG motif. The fact that both proteins are anchored via their C-terminal regions suggests that the N-terminal region of the protein spans the capsular polysaccharides and extends beyond the capsule into the external environment. However, the different mechanisms of localization suggest that there might be differences between PspC and Hic in the strengths of interaction with the bacterial surface. Furthermore, attachment of Hic to the peptidoglycan will result in the protein being attached closer to the cell membrane.

PspC and Hic proteins bind several human plasma proteins including Factor H, C3, C4BP, Plasminogen, thrombospondin-1, and vitronectin<sup>26,28–41</sup>. These multifunctional proteins represent one of the most diverse group of immune evasion proteins<sup>26,41</sup>. PspC and Hic proteins have a mosaic structure, comprised of distinct regions that consist of multiple domains. Furthermore, a substantial overlap of domains exists between PspC and Hic variants. Standard domain or sequence-based comparisons between members of this protein family are complex due to structural differences and variable domain composition. Currently, the protein NCBI databank lists 54,852 entries for PspC or Hic and 12,193 entries for CbpA, including both full-length proteins and partial protein sequences (October 13, 2020; NCBI [www.ncbi.nlm.nih.gov/protein](http://www.ncbi.nlm.nih.gov/protein)). The individual entries show homology, but also exhibit considerable variation in structure and sequence. Examination of several PspC and Hic proteins revealed proteins composed with variable domain patterns, different combinations of domains, and novel domains.

**Mosaic-structured PspC and Hic proteins.** Our understanding of these important pneumococcal immune evasion proteins is currently incomplete. Thus, our ability to understand the function of single domains, know the binding sites for host ligands, determine how the proteins of different strains vary in structure, and correlate these properties with disease states is limited. To achieve these goals it is essential to define the exact domain composition of individual PspC and Hic variants.

**Aim of the study.** Thus far, the domain organization of individual PspC and Hic variants, whether each domain is likely within or extending beyond the cell wall and precise borders of the domains is unclear. Furthermore we do not know exactly which domain(s) are integrated into the bacterial cell wall, which domain(s) span the capsule and which domains are externally positioned. Given these limitations, and the heterogeneity among these important immune evasion proteins, we aimed to evaluate the structure and domain composition of six PspC and five Hic variants, each representing one of the clusters defined by Iannelli et al.<sup>40</sup>. We further

Host regulators	Function	Binding site
Factor H	Complement regulation	HVD
slgA/plgR	Adhesion	Repeat domains
C3	C3 inactivation	Not mapped
C4BP	CP inhibition	Not mapped
Plasminogen	Proenzyme; plasmin cleaves inactivates C3, C3b and fibrin	Not mapped
Thrombospondin-1	Adhesive glycoprotein, cell-cell and cell-matrix interaction	Not mapped
Vitronectin	Complement control and adhesion	Not mapped
Lactoferrin	Fe metabolism	Proposed by homology
IgA	IgA inactivation?	Proposed by homology

**Table 1.** Host regulators binding to *S. pneumoniae* PspC and Hic proteins. The domains are listed in order of their location starting from the N-terminus. Known domains and new domains are included. The table includes domains which are found in both PspC and Hic variants, domains which are specific for either PspC or Hic, and those which are found in other bacterial proteins. *SP* signal peptide, *HVD* hypervariable Domain, *RD* Repeat Domain, *RCD* random coil domain, *S<sub>n</sub>D/GS*, Serine Rich segment, *RCE* random coil extension, *R-type* repeat related Domain; *EPRD* extracellular proline rich domain, *VS* variant specific, *IgA* IgA binding domain, *PRD* proline-rich domain, *CBP* choline-binding domain.

aimed to define domain composition and position. Our studies illustrate structural and compositional differences between the full-length PspC and Hic proteins, within the PspC or Hic group and between the N and C-terminal regions. Furthermore, this comparison also identified nine new domains and several subvariants.

## Results

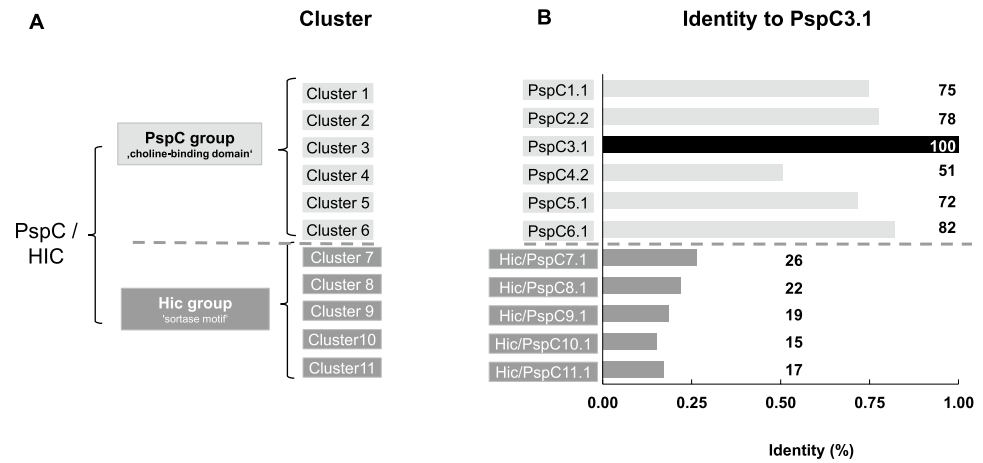
**Global similarity of PspC and Hic variant proteins.** *Selection of PspC and Hic variants.* One protein from each variant cluster as defined by Ianelli et al. was selected<sup>40</sup>. These are the six PspC variant clusters, i.e. PspC1.1, PspC2.2, PspC3.1, PspC4.2, PspC5.1, PspC6.1, and the five Hic variant clusters, Hic/PspC7.1, Hic/PspC8.1, Hic/PspC9.1, Hic/PspC10.1, Hic/PspC11.1. At the date of the cluster designation Ianelli et al. considered the PspC and Hic variants as one protein family and used a PspC nomenclature for both protein groups<sup>37</sup>. To preserve the differentiation between Hic and PspC families and at the same time follow the nomenclature suggested by Ianelli et al. we combined the Hic and PspC designations (Fig. 1A). The selected proteins vary in size and mass, with PspC1.1 being the largest protein with a length of 929 aa and a molecular mass of 110 kDa, while Hic/PspC8.1 is the smallest protein with a length of 503 aa and a mass of 65 kDa (Supplementary Table 1). When compared to the well-characterized PspC3.1 protein (strain D39), the overall amino acid identity of the six PspC proteins ranged from 51 to 82%. In contrast, the five Hic variants were less conserved, with aa sequence identity ranging from 15 to 26%. These high levels of sequence diversity also suggest functional differences between the PspC and Hic variants (Fig. 1B).

*PspC3.1 as a prototype PspC.* PspC3.1 was selected as a prototype and used for analyzing structure and domain composition. PspC3.1 has a signal peptide that directs the protein to export. The N-terminal region of the protein extends beyond the cell surface, while the C-terminal region interacts with the teichoic acids of the bacterial cell wall via the C-terminal Choline-Binding Domain. Because some regions of these proteins are within the cell wall while others extend beyond, we hypothesized, that hydrophilic and hydrophobic surroundings, could influence protein structure and composition.

*Structure and residue composition of PspC3.1.* In silico analysis of PspC3.1 revealed three different structural regions. The N-terminal 410 residues form mostly  $\alpha$ -helices, this region is followed by a 70 aa predominately coiled-coil region and a 221 aa region composed mainly of  $\beta$ -sheets (Fig. 2A). Given these structural differences the 410 aa mainly  $\alpha$ -helical region was designated as the N-terminal region and the remainder of the protein containing the coiled-coil and  $\beta$ -sheet segments was designated the C-terminal region.

For the purpose of this study, the terminology region is used to describe longer protein elements which have related structural or compositional features. Domains are considered to represent separate, individual folding units which display specific functions. Single domains can be further subelements, including modular elements or repeat units which are assembled in repetitive manner and which can vary in sequence and in aa length.

When the structural regions were aligned with the previously identified domains of PspC3.1, the N-terminal  $\alpha$ -helical region included the signal peptide, the Hypervariable Domain, the two Repeat Domains, and the Random Coil Domain. The Hypervariable Domain includes the binding sites for human Factor H and each Repeat Domain includes a binding site for slgA/polymeric Ig receptor, which is in agreement with the concept that these domains extend beyond the cell wall. In contrast the C-terminal region consist of domains expected to be within the cell wall and in the membrane. The mostly coiled-coil region represented the Proline-Rich Domain (aa 411–482), which is considered a cell wall-spanning and flexible domain and the  $\beta$ -sheet region represented the Choline-Binding Domain (aa 483–701) which mediates attachment to the cell wall (Fig. 2B)<sup>41,42,55,56</sup>.



**Figure 1.** Diversity of PspC and Hic cluster variants. PspC and Hic proteins were initially considered to represent one protein class that, based on the different surface anchors, can be divided into two major groups. (A) PspC variants with choline-binding domains representing the PspC group, and Hic variants with sortase dependent LPsTG motifs for cell wall anchoring representing the Hic group. For each group additional clusters were identified. For the analysis one variant from each cluster was selected, i.e. for the PspC group: PspC1.1, PspC2.2, PspC3.1, PspC4.2, PspC5.1, PspC6.1; and for the Hic group: Hic/PspC7.1, Hic/PspC8.1, Hic/PspC9.1, Hic/PspC10.1, Hic/PspC11.1. (B) Amino acid identity of the full-length selected cluster variants with PspC3.1. The variation identified for the six PspC and the five Hic variants selected is indicative of compositional variation among the two major protein groups.

**Amino acid composition.** Next we evaluated if the proposed cell wall integration and external environments influence the protein make up. Of the aa residues within the N-terminal region of PspC3.1 45.3% are charged, 18.0% are polar and amphipathic residues and a low proportion are Tyr (1.7%). In contrast, the C-terminal region contains a lower percentage of charged residues (15.0%), an increased percentage of polar and amphipathic amino acids (9.5%) and a high level of Tyr residues (8.9%) (Fig. 2C). Thus, the N-terminal and C-terminal regions of PspC3.1 differ in domain structure, and amino acid composition.

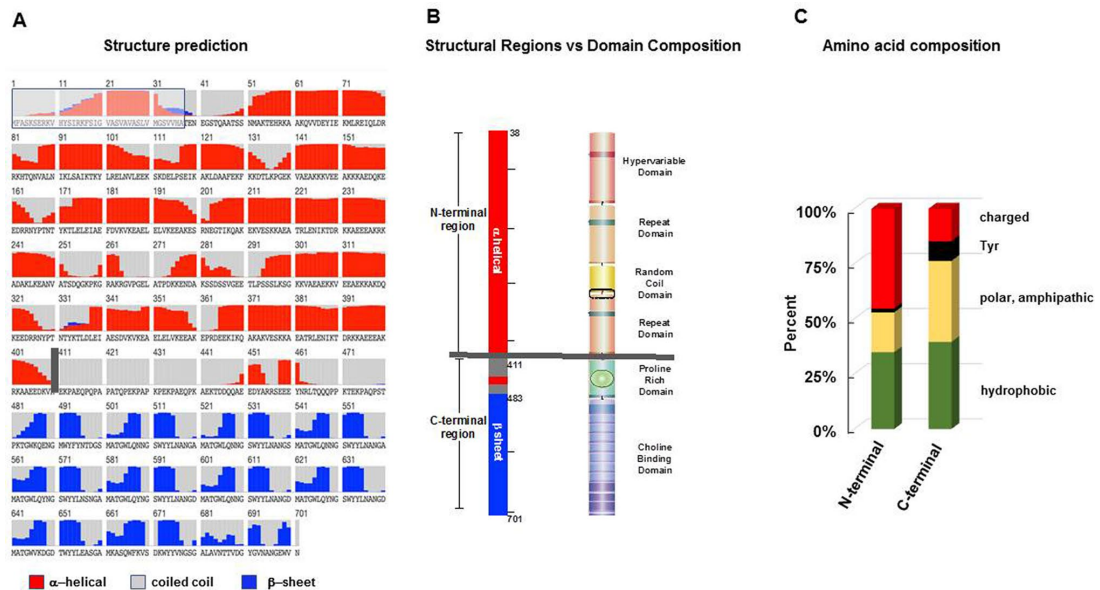
**The differences between the N and C-terminal regions are conserved in the other PspC and Hic variants.** Next we evaluated if the structural composition, as outlined for PspC3.1, is conserved in the other PspC and Hic variants. The N-terminal region of all analyzed PspC and Hic variants consists mainly of  $\alpha$ -helices, and the C-terminal Proline-Rich Domains are predominantly coiled-coil structures. The Choline-Binding Domains within the C-terminal PspC variants consist mainly of  $\beta$ -sheets, while the Hic specific LPsTG anchors consist of a coiled-coil stretch followed by an  $\alpha$ -helical segment (Supplementary Figs. 1 and 2).

In addition, the amino acid composition was determined. Thirty-five to forty-five percent of the aa residues in the N-terminal regions of the six PspC variants are charged. In contrast only 16% of residues in their C-terminal regions were charged. The C-terminal regions of the PspC variants also contained more polar and amphipathic amino acids (32–36%), and were rich in Tyr residues (8.3–9.8%) (Fig. 3A). Charged residues were common in both the N-terminal (28–37%) and C-terminal (28–41%) regions of the Hic variants. Furthermore, the C-terminal region of Hic variants contained less polar and amphipathic residues (15–21%) than the PspC variants (Fig. 3A). Thus, the N and C-terminal regions of the proteins differ in structure and amino acid composition, and the C-terminal regions of the PspC and Hic proteins show differences in amino acid composition.

The N-terminal regions of the different variants ranged in length from 146 (Hic/PspC8.1) to 633 (PspC5.1) residues. A homology alignment of the N-terminal regions showed two distinct clusters. One N-terminal cluster included five PspC variants (PspC1.1, PspC6.1, PspC2.2, PspC5.1, PspC3.1) and the Hic/PspC11.1 variant, while the second N-terminal panel included PspC4.2 and four Hic variants (Hic/PspC7.1, Hic/PspC9.1, Hic/PspC10.1, Hic/PspC8.1) (Fig. 3B, upper panel). The C-terminal regions were more conserved in length, ranging from 236 (PspC5.1) to 348 aa (Hic/PspC8.1) and by sequences clearly separated into distinct PspC and Hic groups. The level of diversity between the C-terminal regions of variants within each group was low indicating that these domains are more highly conserved (Fig. 3B, lower panel).

**Domain analyses of PspC and Hic variants.** Using PspC3.1 with its five known domains as a blueprint, a sequence based comparison was followed to determine the presence and organization of domains within the other ten cluster variants was evaluated. This approach identified three domains of PspC3.1, the signal peptide, the N-terminal Hypervariable Domain and the C-terminal Proline-Rich Domains, present in all PspC and Hic variants. All PspC variants use a Choline-Binding Domain, while Hic/PspC proteins have an LPsTG anchor (Figs. 1 and 4). The Repeat Domains and the Random Coil Domain are found mainly in PspC proteins, but not





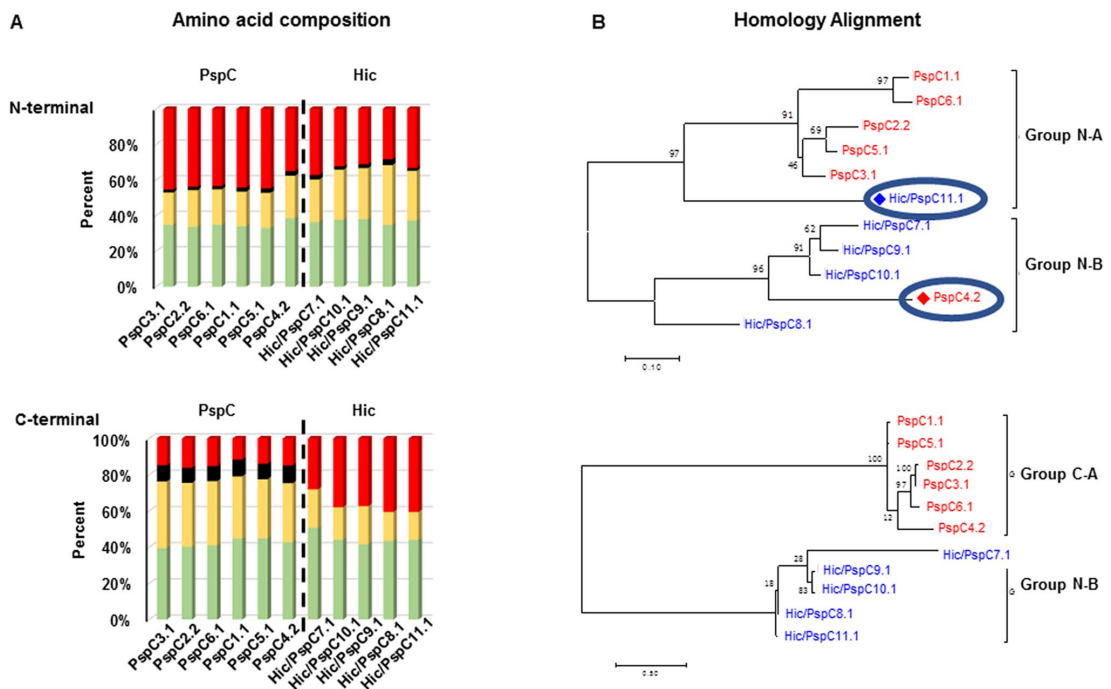
**Figure 2.** Structural regions and domain position of PspC3.1. *Dissection of PspC3.1 into distinct structural regions using in silico analyses.* (A) Secondary structure of the well-characterized PspC3.1 variant (strain D39). The N-terminal part of the molecule shows a long stretch composed mainly of  $\alpha$ -helices (red bars) (aa 1–410) followed by a 72 aa coiled-coil segment (grey area) and a 219 aa region consisting mainly of  $\beta$ -sheet folds (blue bars). The numbers represent the amino acid position. The signal peptide (positions 1–37) which is cleaved upon processing is shown by the box with grey background and blue lines. The vertical grey bar separating the N-terminal  $\alpha$ -helical from the coiled-coil region may represent the boundary to the bacterial cell wall. (B) Structural regions and domain composition of PspC3.1. The mainly  $\alpha$ -helical region (positions 38 to 410) is termed the N-terminal region. The remainder of the protein includes the 72 aa coiled-coil and the 219 aa mainly  $\beta$ -sheet segments is termed the C-terminal region (left panel). To correlate structural regions with the domain composition, the known domains of PspC3.1 were included (right panel). The Hypervariable Domain, Repeat Domain I, Random Coil Domain and Repeat Domain II aligned with the N-terminal, mainly  $\alpha$ -helical region. In the C-terminal region of the protein the coiled-coil segment consisted of the Proline Rich Domain and the  $\beta$ -sheet segment with the Choline-Binding Domain. The grey horizontal line separates the N and C-terminal regions and likely marks the border of the cell wall and capsule facing the outside environment. (C) Amino acid composition of N and C-terminal regions. The amino acid composition was evaluated separately for each region. The N-terminal region is rich in charged residues (48%), has a low number of both polar and amphipathic residues (24%), and Tyr residues (left panel). In contrast, the C-terminal region contained a lower percentage of charged residues (22%), had more polar and amphipathic amino acids (38%) and more Tyr residues (8%).

in all variants. Additional sequences were identified in some variants that did not match known domains of PspC3.1. These domains were evaluated to determine whether they are present in other PspC and Hic variants or whether homologs exist in the protein data bank. This approach identified nine new domains, including one new domain in PspC3.1, and three new sub-variants of the Proline Rich Domain. Including these new domains in an examination of the PspC and Hic variants revealed that the individual proteins harbour between four (Hic/PspC8.1) and ten different domains (PspC4.2) (Fig. 4).

**Known domains of the N-terminal region.** The known domains identified in the N-terminal region include:

**Signal peptide.** A highly-conserved 37 aa N-terminal signal sequence which directs the proteins for export and is cleaved upon processing, is present in all PspC and Hic/PspC variants (Supplementary Fig. 3A).

**Hypervariable domains.** At the N-terminus of the mature Hic and PspC proteins are the Factor H binding Hypervariable Domains<sup>26,28</sup>. These domains are rich in charged residues and vary in length from 91 (PspC4.2) to 113 aa (PspC2.2). As their name suggests, they were highly variable in sequence with each PspC and Hic variant examined encoding a distinct variant (Supplementary Fig. 3B). Only five residues, T<sub>11</sub>, S<sub>12</sub>, I<sub>59</sub>, Y<sub>63</sub>, K<sub>96</sub> (numbering based on PspC3.1) present in all variants; although, additional residues are conserved in several variants. Factor H binding by PspC3.1 is mediated by a 12 amino acid region<sup>28</sup>, we identified diversity in this region of differ-



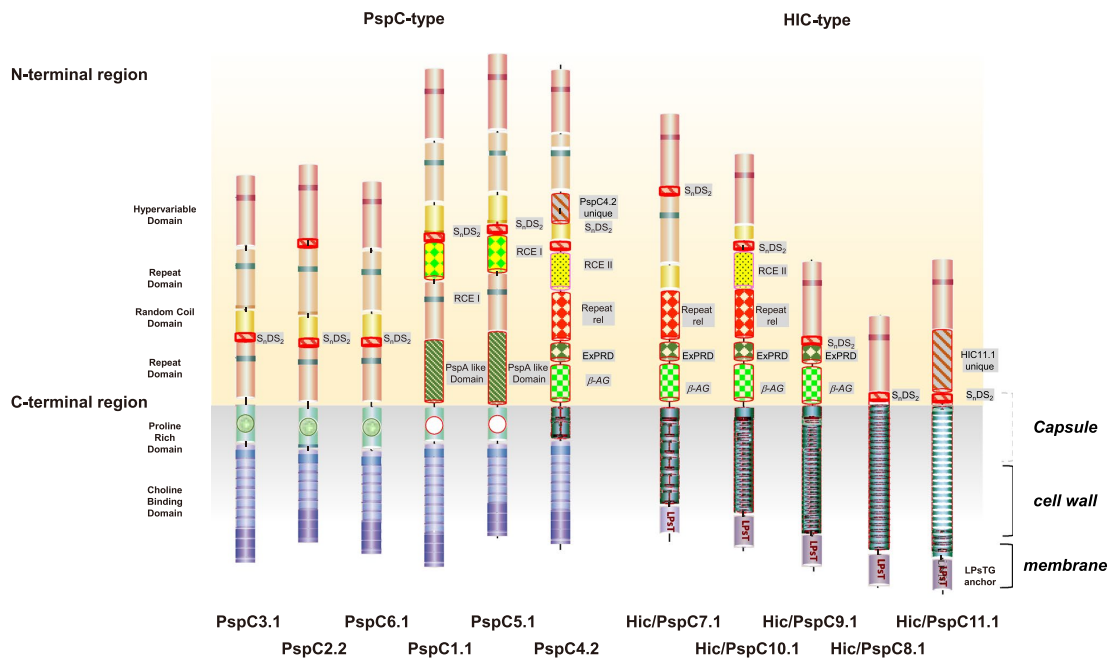
**Figure 3.** Differences in the N and C-terminal regions of the PspC and Hic variants. **(A)** The N and C-terminal regions of PspC and Hic type proteins differ in amino acid composition. The amino acid composition of the N and C-terminal regions was evaluated for each selected variant. The N-terminal regions of the six PspC and the five Hic variants are rich in charged residues (35–45%) and have a low number of both polar and amphipathic, and Tyr residues. The PspC variants had also a high proportion of charged residues (28–42%) (upper panel). The C-terminal regions of the PspC variants had a lower percentage of charged residues (16% or less) and more polar and hydrophilic (32–36%) and Tyr residues (8.3–9.1%). The composition of the C-terminal region of Hic variants differed from that of PspC variants. The C-terminal regions of Hic variants contained more charged residues, lower levels of Tyr and polar and amphipathic residues (lower panel). **(B)** Phylogenetic trees of the N and C-terminal regions of PspC and Hic type proteins. The homology alignment of the N and C-terminal regions identifies two groups. For the N-terminal regions group A is dominated by PspC type proteins, but also includes Hic/PspC11.1. Group B is dominated by Hic type proteins, but also includes the PspC4.2 variant. The C-terminal regions show a clear separation between the PspC and Hic variants.

ent variants and whether these domains all bind Factor H remains to be determined (Fig. 5A, Supplementary Fig. 3C).

Relationship analysis using a dendrogram identified three subtypes of the hypervariable domains. Subtype A (HVD-A) is present in PspC3.1, PspC5.1, and Hic/PspC11.1, HVD-B is present in PspC2.2, PspC1.1, and PspC4.2, and HVD-C is present in PspC6.1, Hic/PspC7.1, Hic/PspC10.1, Hic/PspC9.1, and Hic/PspC8.1 (Supplementary Fig. 3C).

**Repeat domains.** All PspC-type proteins and Hic/PspC7.1 possess a repeat domain of approximately 110 aas (Repeat Domain). Five PspC variants (i.e. PspC3.1, PspC2.2, PspC6.1, PspC1.1, PspC5.1) contain a second Repeat Domain. These Repeat Domains are rich in charged residues, and include conserved RNYPT motifs, which are binding sites for sIgA/pIgR (Fig. 5B, Supplementary Fig. 4). Related repeat domains identified in pneumococcal PspK (H2BJK8) share 55% aa identity with Repeat Domain I and 71.6% identity with Repeat Domain II. The solution structure of the Repeat Domain of PspC3.4 from strain TIGR4 has been solved<sup>42</sup>. This domain folds into three antiparallel  $\alpha$ -helices and the YPT residues, representing the core sIgA/pIgR binding motif, are positioned in a coiled-coil loop, which separates the first and second helices. This experimentally determined structure confirms and validates our in vitro structure prediction (Fig. 2A).

**Random coil domain.** The Random Coil Domains are typically positioned downstream of the first Repeat Domain. They are approximately 30 aas in length, have a coiled-coil structure and are relatively conserved in sequence. No homologous sequences were identified in the sequence database (Supplementary Fig. 5).



**Figure 4.** Domain structure of the six PspC and five Hic variants. The domain architecture of PspC3.1 is shown on the left-hand side. The PspC and Hic variants differ in length and domain number. The proteins are arranged based on their overall homology. To reflect the different lengths of the regions proposed to be within cell wall and the external environment the proteins are centered along the axis which separates the N-terminal,  $\alpha$ -helical region from the C-terminal region. The N-terminal and C-terminal regions are shown on yellow and grey backgrounds, respectively. Proteins are drawn to scale. The signal peptides and for the Hic-cluster the C-terminal region which is cleaved upon anchoring are not included. Domains previously identified within PspC3.1 are shown in solid colors. New domains are patterned, and their names are given alongside the domain on a grey background. The predicted binding sites for the human plasma protein Factor H within the Hypervariable Domain are shown by the purple bar and those of *slgA/plgR* within the Repeat Domains by green bars. The PspA like domain and the  $\beta$ -AG binding domains were identified by homology with the binding domains within *S. pneumoniae* protein PspA and the IBC protein from *S. agalactiae*.

**New domains of the N-terminal region.** Sequences in the PspC and Hic variants that did not match known domains of PspC3.1 were also identified. A data base search for counterparts identified nine new domains, including one new domain in PspC3.1 and also three new variants of the Proline Rich Domain.

**Serine-rich elements.** Serine-Rich Elements with the overall motif  $S_2D/GS_2$  were detected in all PspC and Hic variants with the exception of PspC4.2. Nine variants harbored one serine-rich element, whereas PspC2.2 contained two. These Serine-rich elements share a coiled-coil structure; but differ in their sequence and position within the protein. Serine-rich elements following the Hypervariable Domain (PspC2.2, Hic/PspC7.1, Hic/PspC9.1, Hic/PspC8.1) or the unique Hic/PspC11.1 domain have the consensus  $S_2D/GS_2$  and are up to 24 aa in length. The serine-rich elements following the Random Coil Domain (PspC3.1, PspC2.2, PspC6.1, PspC1.1, PspC5.1, Hic/PspC10.1) are comprised of  $S_2DS_2$  units and can be up to 18 aa long. The domain of Hic/PspC10.1 shows a variation to these common features (Supplementary Fig. 5A). The biological role(s) of these elements are as yet unknown. However, in engineered proteins, related poly-serine-rich elements are integrated as flexible linkers that separate functional, individually folding domains<sup>43</sup>. Interestingly the TKPET motif at the end of  $S_2DS_2$  domains following the Hypervariable domains are related to the first seven residue long units found in Proline Rich Domains III and IV (see below).

**Random coil extension domains.** Two new domains were identified downstream of the Random Coil Domain- $S_2DS_2$  combination of domains.

**Random coil extension domain 1.** Two proteins, PspC1.1 and PspC5.1, contain an almost identical new 83 aa domain. This domain includes several charged residues, and shares homology with RICH type domains in other proteins, including PspC Q9KK19, SpsA O33742 and IgA Fc receptor binding protein P27951 from *Streptococcus agalactiae*. These domains are predicted to be involved in bacterial adherence or cell wall binding<sup>44</sup>.

**Figure 5.** Sequence Variation and Conservation of Binding Domains and Surface Orientation of PspC1.1 and Hic/PspC8.1. (A) Sequence variation of the Factor H binding motif within the Hypervariable Domains of the six PspC and five Hic variants. WebLogo was used to evaluate amino acid variation. (B) Sequence conservation of the binding sites for human sIgA/pIgR in Repeat Domains I and II. (C) WebLogo was used to evaluate sequence variations in the second and third choline-binding modules of the PspC variants. Sequence variation among the Choline-Binding Modules 2 and 3 of the PspC variants. Residues relevant for the interaction with choline are indicated by the box arrows and include Trp at position #3, i.e.  $W_3$  and  $W_{10}$  of module  $n$ , as well as  $Y_{11}$  of module  $n + 1$ . (D) Sequence conservation of the sortase recognition motifs LPsTG in the C-termini of Hic-type proteins. (E) Structure of and proposed orientation of the phosphorylcholine (PCho) associated PspC1.1, and sortase A dependent covalently linked Hic/PspC8.1 variant. The arrangement is based on the concept that PspC1.1 is non-covalently associated to the teichoic acids via its interaction with PCho. In contrast the Hic/PspC8.1 variant is covalently linked via the sortase anchor to peptidoglycan Penicillin binding protein (PBP). This attachment and orientation suggests that the Proline-Rich Domains may represent a flexible cell wall and capsule spanning segment. The grey line represents the bacterial membrane and cell wall, and the capsule is indicated by the shaded grey region. The domains proposed to extend beyond the cell wall and capsule exodomains are shown in yellow or red. The binding domains for human plasma regulator Factor H within the Hypervariable Domains (PspC1.1 and Hic/PspC8.1) and the sIgA or cell surface receptor pIgR binding domains within the Repeat Domains I and II (PspC1.1) are indicated by purple and green bars, respectively. Attached Factor H mediates complement evasion and blocks complement mediated opsonophagocytosis and release of the anaphylatoxins C3a and C5a. sIgA or pIgR bind to two sites in PspC1.1 and block opsonization by sIgA or mediate adhesion to human epithelial cells. The binding sites of vitronectin and other human plasma proteins remain to be mapped. The C-terminal regions, with a proposed location within the cell wall or capsule are shown in green, blue or purple and include the Proline Rich Domains followed by Choline-Binding Domains (PspC1.1) or LPsTG mediated anchor (Hic/PspC8.1).

**Random coil extension domain 2.** PspC4.2 and Hic/PspC10.1 have 114 and 126 aa domains that follow the Random Coil Domain and which share moderate sequence identity. The N-terminal domain of Hic/PspC10.1 has a 37 aa extension, with the remainder of the domain being sequence similar with the PspC4.2 domain. The biological role of this unique segment is unclear. In PspC4.2 this domain includes a long  $\alpha$ -helical stretch that is followed by a 30 aa coiled-coil region.

**PspA-like domain.** PspC1.1 and PspC5.1 have related, new domains following Repeat Domain II. These 130 or 131 aa domains are rich in charged residues, and share 84.5% sequence identity with the A\*/B element of PspA from pneumococcal strain DBL6A. The A\*/B element includes a lactoferrin-binding region<sup>45,46</sup>, suggesting that the newly identified domains in PspC1.1 and PspC5.1 bind lactoferrin<sup>47,48</sup>.

**PspC4.2 specific element.** Domain pattern analysis identified an element in PspC4.2 which is positioned between the Hypervariable Domain and the Random Coil Domain. This 33 aa  $\alpha$ -helical structured element, lacks homology to other proteins in the databank, thus its role remains unclear.

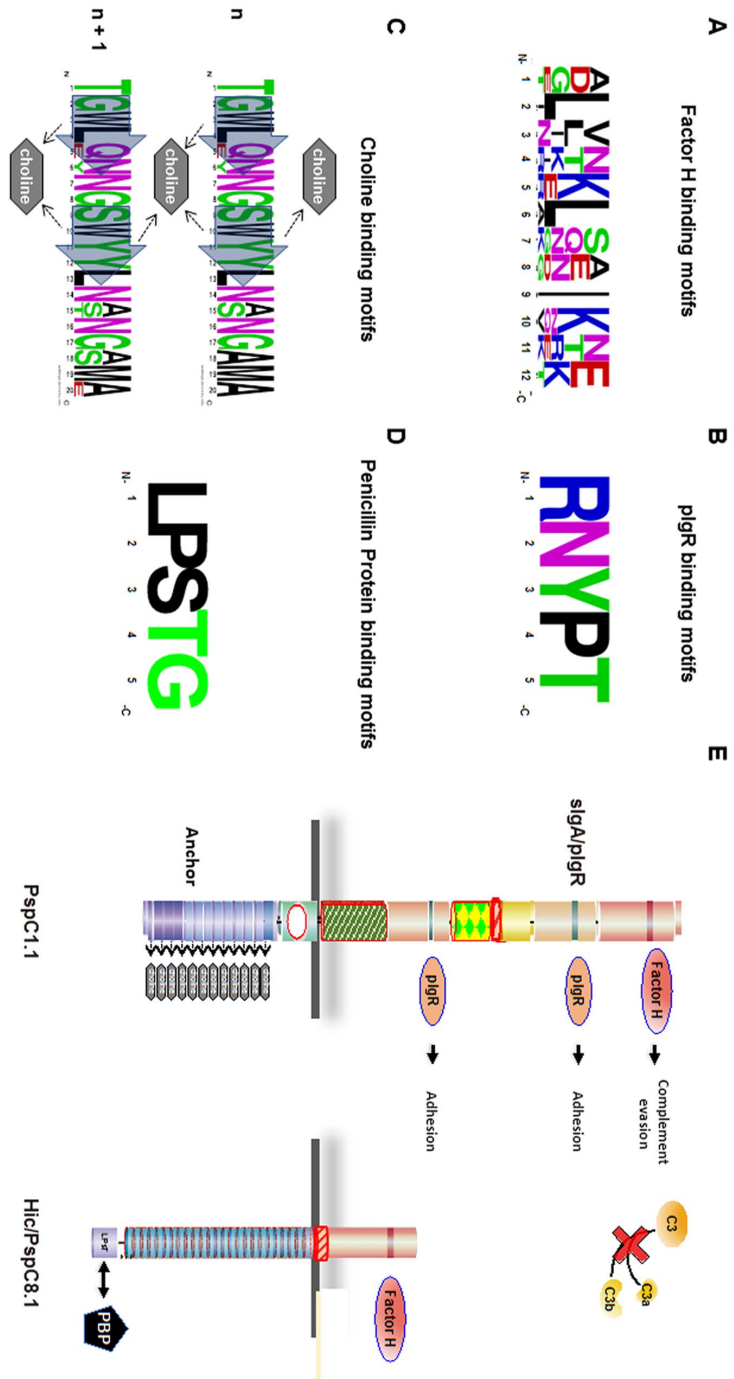
**Repeat type domain.** PspC4.2, Hic/PspC7.1, and Hic/PspC10.1 contain related 92, 82, and 68 aa domains, respectively. These mostly  $\alpha$ -helical domains are distantly related (41.6% aa identity) to the Repeat Domains, but lack the sIgA/pIgR binding motif (RNYPT) binding motif and seem to be specific to PspC and Hic proteins.

**A new two-segmented domain.** A new two-domain segment was identified in PspC4.2 and the three Hic proteins, Hic/PspC7.1, Hic/PspC10.1, Hic/PspC9.1.

**The upstream segment.** The 24–40 aa upstream segments of this domain are rich in proline residues, have a predicted coiled-coil structure, and due to their location in the N-terminal region of PspC are termed *Extracellular Proline Rich Segments*. The high Proline content may suggest a function as linker separating domains<sup>49</sup>. These External Proline Rich Segments lack homology to other bacterial proteins, and thus seem unique to PspC proteins.

**The downstream elements share sequence similarity with the Fc binding domain of protein C from *S. agalactiae*.** The 78 or 89 aa elements are rich in charged residues, lack proline residues, and have an  $\alpha$ -helical structure. A blast search revealed 51.1% aa identity with an IgA binding domain within the trypsin sensitive beta-antigen of *Streptococcus agalactiae* (strain P27951/Uniprot). This protein binds the Fc region of human IgA, likely via two putative binding sequences<sup>50</sup> which are also found in several other bacterial immune evasion proteins including SpsA from *S. pneumoniae*. Based on the many charged residues this IgA binding domain (pfam05062) is also named RICH (Rich In Charged residues) the proposed function of which is bacterial adherence or cell wall binding.

**Hic/PspC11 specific element.** Following the Hypervariable Domain, Hic/PspC11.1 contains a unique 102 aa  $\alpha$ -helical domain. Related domains were identified in most Hic/PspC11 variants, but not in other bacterial proteins. Thus far, the function of this domain is unknown.



**Domain composition of the C-terminal region.** The C-terminal region of each variant contains a modular Proline-Rich Domain with a Choline-Binding Domain for PspC variants or an LPsTG anchor for Hic variants<sup>46–49</sup>. The C-terminal regions of the PspC and Hic proteins analyzed are relatively conserved in length (ranging from 237 aa (PspC5.1) to 348 aa (Hic/PspC8.1)). A general pattern is emerging: PspC proteins link shorter Proline-Rich Domains (57 to 77 aa) to longer Choline-Binding Domains (179 to 219 aa), while Hic proteins combine longer, Proline-Rich Domains (186 to 286 aa) with shorter LPsTG anchors (50 to 62 aa).

*Proline-rich domains.* Proline-Rich Domains have a modular structure and connect the N-terminal region to the cell wall anchor<sup>51</sup>. The proposed role of these domains as spanning the bacterial cell wall-spanning is consistent with the position proximal to the anchor<sup>51,52</sup>. Our in-silico analysis identified a modular composition and further distinct proline-rich domains, which differ in length (57 to 286 aa), modular composition, and sequence.

*Proline-rich domain I.* Five PspC variants have highly related 59 to 77 aa domains, termed Proline Rich domain I. This modular domain can consist of two (PspC1.1, PspC5.1) or three (PspC3.1, PspC6.1, PspC2.2) segments (Supplementary Fig. 7A). The N-terminal segments have Proline dominated PAPA- and PAPAP motifs and can be up to 46 aa long. The C-terminal segments include PAPAP or PAPTP motifs, are up to 19 aa long, and have a coiled-coil structure. The middle segment, present only in the domains with three segments is conserved in length (23 aa), sequence, exhibits characteristic flanking Q-residues, and is rich in charged residues. In contrast to the other two segments this segment has a predicted  $\alpha$ -helical structure and lacks Prolines. Such Proline-Rich segments are also found in PspA<sup>52–54</sup>.

*Proline-rich domain II.* PspC4.2 has a unique 57 aa-long Proline-Rich Domain. This new domain includes 19 Prolines and has an internal repeated segment with the sequence TPQVPKPEAPK. To date, this new domain has been identified only in PspC proteins) (Supplementary Fig. 7B).

*Proline-rich domain III.* Hic/PspC7.1 contains a unique 186 aa-long Proline-Rich Domain which includes an N-terminal 7 aa element followed by five almost identical 31 aa repeats (KKPSAPKP(G/D)MQPSQPPEGK-KPSVPAQPGTED). Each repeat contains nine proline residues and two KKPS(A/V)P motifs. The repeats are followed by a truncated 24 aa repeat element (Supplementary Fig. 7C, D).

*Proline-rich domain IV.* Four Hic variants harbor 247 to 286 aa, Proline-Rich Domains containing 19, 23 or 26 modules. The modules vary in type and sequence, but all include multiple 11 aa repeats, (Supplementary Fig. 8A–C). Hic/PspC10.1 and Hic/PspC9.1 contain 14 and 16 (L/P)EKPKPEVKPQ repeats, respectively. Both Hic/PspC8.1 and Hic/PspC11.1 contain 23 copies of a (L/P)ETPKPEVKPE repeats (variant residues are displayed as white letters on a black background). In each case, these repeats are followed by one shortened repeat and a nearly identical 16 aa-long C-terminal module, which varies only at position 15 (T/P variation) (Supplementary Fig. 8D, E, F).

*Cell wall attachment.* Both PspC and Hic/PspC variants have modular domains within their C-terminal regions that we propose span the cell wall. PspC proteins bind the cell wall via modular Choline-Binding Domains in contrast, Hic proteins have shorter, 50–62 aa- anchors that include a sortase-dependent LPsTG cell wall attachment motif<sup>55,56</sup>.

PspC-type protein variants possess choline-binding anchors. PspC type variants have C-terminal Choline-Binding Domains that range in length from 178 (PspC5.1) to 248 aa (PspC1.1) and consist of modules most of which are 20 aa in length (Fig. 5C)<sup>57</sup>. Related Choline-Binding Domains are found in up to 15 other pneumococcal proteins, including the immune evasion protein PspA, the autolysins LytA, LytB, LytC, and CbpL<sup>57</sup>. In the literature these modular Choline-Binding Domains are sometimes termed choline-binding modules. However, given the domain composition of full length PspC and Hic variants we prefer to term such smaller, repetitively assembled subunits as modules.

Hic variants have C-terminal sortase signals. The five Hic variants analyzed share C-terminal 50–62 aa anchors which contain a pentapeptide LPsTG motif. The transpeptidase, sortase A cleaves this conserved motif between the Thr and Gly residues. Subsequently the protein is covalently linked via the Thr residue to lipid II (P3 precursor) and a penicillin binding protein<sup>58,59</sup> (Fig. 5D).

## Discussion

The mature PspC and the Hic/PspC proteins are heterogeneous in structural composition and in sequence. Our analysis of domains within one member of the six PspC and five Hic variants identified 13 N-terminal and three C-terminal domains, including nine new domains and three new variants of the Proline-Rich Domain. The extensive diversity is the result of different combinations of domains, several of which are present in different numbers. Domain variability is increased by distinct variants of some domains, differences in the assembly of modular elements within domains and sequence variation. This diversity results in antigenic variation, functional specialization and mechanisms of cell wall anchoring<sup>18,20</sup>. Three domains, the Signal peptides, the Hypervariable Domains and Proline-Rich Domains are found in all analyzed variants (Table 2). Eleven domains are found in some variants, and two domains are unique to single variants. This extensive characterization shows a different

#	Region		Domain	Sub domains	Class	n	Module	Structure		Comment host ligand	
1			Known	SP		11					
2	N-term		Known	HVD	HVD-A, HVD-B, HVD-C			α helix	PspC/Hic specific	Factor H	
3			Known	RD	RD-I, RD-II, RD-III			α helix		sIgA/plgR	
					RD-II	PspC		α helix			
4			Known	RCD				α helix			
5		1	New	S <sub>2</sub> D/Gs <sub>2</sub>	3 Positions			Coiled coil			
6		2	New	RCE1		PspC		α helix		Lactoferrin	
7		3	New	RCE2				A helix			
8		4	New	PspA related		PspC		α helix	In PspA		
9		5	New	R-type				α helix		IgA	
10		6	New	EPRD				α helix			
11		7	New	IgA				α helix	<i>S. agalactiae</i>		
12		8	New	VS4.2		PspC		α helix	Specific		
13		9	New	VS11.1		Hic		α helix	Specific		
14	C-term		Known	PRD	PRD-IA, PRD-1B	PspC	5	Modular	Coiled coil	Also in PspA	Cell wall spanning
			New		PRD-II	PspC	1	Modular	Coiled coil	?	
			New		PRD-III	Hic	1	Modular	Coiled coil	?	
			New		PRD-IV	Hic	4	Modular	Coiled coil	?	
15			Known	Anchor	CBD	PspC	6	Modular	β sheets	Several	Anchor
16			Known		LPsTG	Hic	5		Coiled coil	Many	Anchor

**Table 2.** Domain used by *S. pneumoniae* PspC and Hic proteins. The binding sites for Factor H has been mapped within the Hypervariable Domain of PspC3.1 and that of sIgA and the extracellular domain of plgR to the RNYPT motif of Repeat Domains I and II. C3, C4BP, Plasminogen, Thrombospondin 1, vitronectin have been shown to bind intact *S. pneumoniae* and full length PspC and Hic proteins, but their binding sites have not been mapped to specific domains. Binding of Lactoferrin and IgA is proposed based on the homology between PspC and Hic variants and the *S. pneumoniae* immune escape protein PspA and the sIgA binding protein of *S. agalactiae*.

composition of the N and C-terminal regions, reveals differences between PspC and Hic variants, as well as differences in the distribution, order, number and sequence variants of domains and repeats present.

**Variability among PspC and Hic/PspC-variants.** PspC, and Hic-type variants have related domains in their N-terminal regions but differ more in their C-terminal regions. The proteins have different C-terminal anchors. PspC proteins with the Choline-Binding Domains contact multiple choline-moieties in a non-covalent manner. In contrast the LPsTG anchors attach the proteins covalently to the peptidoglycan<sup>56</sup>. The type of C-terminal anchor not only influences cell wall attachment, but the length and composition of the Proline-Rich Domains. Furthermore the cell wall anchors seems to influence selection, composition, and number of the N-terminal domains. These differences in structure likely alter the role of the proteins in immune evasion and may result in different domains extending beyond the cell wall.

**Variability of N vs C-terminal regions.** Broadly speaking, each PspC and Hic protein is divided into two major parts: the N-terminal region that extends beyond the cell wall and includes immune evasion and adhesion domains, and the C-terminal anchoring region.

The N-terminal regions of the PspC and Hic proteins analyzed vary in length, and domain number, ranging from 155 aa containing two domains (Hic/PspC8.1) to 610 aa containing eight domains (PspC4.2). These regions share structural features, including long α-helical structures, and a high proportion of charged residues. The Hypervariable Regions are most likely located most distant from the cell surface and show the highest degree of sequence variation. This diversity can reflect differences in antigenic variability, which is relevant for evading immune recognition by antibodies. Six of the N-terminal domains are unique to PspC and Hic variants, others like the PspA Related Domain and the region with homology to the IgA binding β antigen are found in other bacterial immune evasion proteins (Supplementary Fig. 6C).

The C-terminal regions are more conserved in length, have more polar and amphipathic residues and in the case of PspC variants also have more Tyr residues. The Proline-Rich Domains, preceding the PspC and Hic-specific anchors, are of variable length, have a modular composition, consist mostly of coiled-coil structures. Proline-Rich Domains of PspC proteins are shorter than those of Hic proteins. Given the proposed location at the interface between cell wall and capsule, such diversity could result in different binding dynamics, strength of cell wall integration, morphological differences or capsule thickness<sup>53–62</sup>. Similarly, the anchor domains in the C-terminus differ in length, composition, and type of cell wall attachment.

**Protein orientation, and cell wall integration.** PspC and Hic are cell wall associated surface proteins and we are starting to understand which regions of the proteins are spanning the cell wall and capsule, and which might be extended into the environment. The N-terminal region, by extending beyond the capsule, is exposed to the external environment and can interact with human proteins. The C-terminal region includes a capsule spanning region and an internal cell wall anchor.

Cell wall attachment via the C-terminal anchor orients the N-terminus to the external environment allowing interactions with host plasma proteins and cell receptors. An illustration of the orientation, spatial organization and known binding sites for human plasma regulators of one PspC and one Hic/PspC variant is presented in Fig. 5E. PspC1.1 is an eight domain variant that binds choline and the short four domain Hic/PspC8.1 variant have different compositions both in the N- and C-terminal regions. The variable lengths of the N-terminal regions mean these domains extend with different distances into the external environment. In a linear model, for example, Factor H, when bound via the hypervariable domain inhibits C3b formation and assists in C3b inactivation remote from the bacterial surface. Similarly, the variable length of the Proline-Rich Domains and the type of cell wall anchors encoded can result in differences in the strength of interactions and different localizations within the cell wall.

**Tactical positioning and immune evasion.** The two distinct anchors have different structures. Choline-Binding Domains are composed mainly of  $\beta$ -sheets, whereas sortase A dependent LPSTG anchors mainly consist of coiled-coil and  $\alpha$ -helical structures. This not only dictates whether cell wall attachment is non-covalent or covalent, but is also indicative of a more flexible vs. fixed cell wall interaction. These distinctions in cell-wall attachment may result in a different surface distribution and likely the extent to which the protein is exposed to the external environment. Indeed, different spatial localization of the PspC and Hic/PspC variants both expressed by *S. pneumoniae* strain BNH418, was shown by super resolution microscopy<sup>63</sup>. The PspC-protein, with the Choline-Binding Domain localized to the division septum and Factor H, when bound to this protein, controlled C3b opsonization. In contrast, the LPSTG anchored Hic protein was localized to the bacterial poles. Such differences in surface localization could influence the site on the bacteria where complement control and adhesion to host cells occurs. Therefore, these differences in distribution can influence the biological function of these important immune evasion proteins.

When comparing prevalence and distribution of PspC and Hic variants among 349 pneumococcal isolates from adult patients with invasive pneumococcal disease, 298 isolates (85.4%) had a single *pspC*-variant, 22 isolates had a (6.3%) a *hic*-variant, 19 isolates (5.4%) had *pspC* and *hic* gene and only 10 isolates (2.9%) did not possess either gene<sup>64</sup>. In addition, invasive, PspC expressing strains bound more Factor H, and Factor H binding and immune control was more effective in encapsulated as compared to unencapsulated strains. Similarly, the PspC variants PspC2 and PspC6 were more efficient in Factor H binding and complement inhibition on the bacterial surface as compared to the Hic variants, Hic/PspC9 and Hic/PspC11<sup>65,66</sup>.

**Conclusions and perspectives.** Evaluating the domain composition of selected PspC and Hic variants and an in-depth characterization of the domain composition advanced our understanding of the structure of these virulence determinants. Our approach identified differences between PspC and Hic proteins beyond their distinct membrane anchors. Such knowledge allows a comparison of full-length proteins based on domain patterns, numbers and can result in a better comparison of different PspC and Hic/Hic variants. Similarly, individual domains can be compared based on structure, modular composition and sequence.

Analyzing the additional > 60,000 PspC and Hic proteins deposited in the NCBI protein database or gene products from additional clinical isolates, will likely identify additional variants due to the discovery of new domains and subdomains, and novel domain combinations. Defining the diversity within these pneumococcal virulence factors will increase understanding of their role in immune evasion and provide important information for molecular strain typing and vaccine design. Finally, this may also allow a correlation between PspC or Hic type variants with invasive pneumococcal infections and with clinical outcome.

## Materials and methods

**Selection of PspC and Hic variant proteins.** Each of the selected six PspC and five Hic proteins represent one of the two clusters as initially defined by Iannelli et al.<sup>40</sup>. The sequences were derived from the NCBI database (status: Feb 2018). The PspC/Hic designation is based on Iannelli et al.<sup>40</sup>. The protein names, corresponding bacterial strain, protein size, GenBank Accession number and protein ID are shown in (Supplementary Table 1).

**Secondary structure evaluation.** The structure ( $\alpha$ -helical, coiled-coil and  $\beta$ -sheet) of each selected PspC and Hic protein was evaluated using RaptorX (<http://raptorx.uchicago.edu/http://raptorx.uchicago.edu/>). PspC3.1 most predicted structural similarity 2vyuA ( $p$  value:  $3.39e-10$  and secondary structure: 42%  $\alpha$ -helical, 43% coiled-coil and, 14%  $\beta$ -sheets). Analysis of the other ten PspC / Hic variants revealed a similar secondary structure (Supplementary Figs. 1–2). Each of the six PpsC variants was most similar to 2vyuA. Hic/PspC7.1, Hic/PspC8.1, Hic/PspC9.1, Hic/PspC10.1, Hic/PspC11.1) were most similar to 1w9rA, 4k12B, 2m6uA, 6iaA, 2m6uA, respectively. The secondary structure prediction are shown in the form of histograms which were constructed using ggplot2 from the R/Bioconductor.

**Phylogenetic analysis.** The PspC and Hic amino acid sequences and composition were evaluated using MEGA7 ([www.megasoftware.net](http://www.megasoftware.net)). There was a total of 976 positions in the final dataset<sup>67</sup>. The CLUSTALW program and the BLOSUM amino acid matrix were used to compare the allelic variants of PspC, following which phylograms



were generated using the Neighbor-Joining method (Bootstrap value: 100). The phylogram for each domain was generated using the same method. Phylogenetic trees are modified in MEGA7.

**Domain homology searches.** BLASTp was used to identify related proteins or protein segments within the GenBank database available at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Furthermore, BLAST targeting database UnipRotKB reference proteomes plus Swiss-Prot was used to find regions of local similarity between sequences (<https://www.uniprot.org/blast/>). All the domains in this work have been done a blast.

Received: 14 August 2020; Accepted: 7 December 2020

Published online: 18 January 2021

## References

- Weiser, J. N., Ferreira, D. M. & Paton, J. C. *Streptococcus pneumoniae*: transmission, colonization and invasion. *Nat. Rev. Microbiol.* **16**, 355–367 (2018).
- Seth-Smith, H. Pneu tricks. *Nat. Rev. Microbiol.* **9**, 230 (2011).
- Henriques-Normark, B., Blomberg, C., Dagerhamn, J., Bättig, P. & Normark, S. The rise and fall of bacterial clones: *Streptococcus pneumoniae*. *Nat. Rev. Microbiol.* **6**, 827–837 (2008).
- Kadioglu, A., Weiser, J. N., Paton, J. C. & Andrew, P. W. The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat. Rev. Microbiol.* **6**, 288–301 (2008).
- UNICEF. *Pneumonia*. (2018). [https://www.unicef.org/publications/files/Pneumonia\\_The\\_Forgotten\\_Killer\\_of\\_Children.pdf](https://www.unicef.org/publications/files/Pneumonia_The_Forgotten_Killer_of_Children.pdf)
- WHO Int; Home - Newsroom - Fact sheets - Detail - Pneumonia <https://www.who.int/biologicals/areas/vaccines/pneumo/en/>
- Martinón-Torres, F. et al. EUCLIDS Consortium. Life-threatening infections in children in Europe (the EUCLIDS Project): a prospective cohort study. *Lancet Child Adolesc Health.* **2**(6), 404–414 (2018).
- Brooks, L. R. K. & Mias, G. I. *Streptococcus pneumoniae*'s virulence and host immunity: aging, diagnostics, and prevention. *Front Immunol.* **9**, 1366 (2018).
- Geno, K. A. et al. Pneumococcal capsules and their types: past, present, and future. *Clin. Microbiol. Rev.* **28**, 871–899 (2015).
- Ganaie, F. et al. New pneumococcal capsule type, 10D, is the 100th serotype and has a large cps fragment from an oral streptococcus. *mBio.* **11**(3), e00937–20 (2020).
- Subramanian, K., Henriques-Normark, B. & Normark, S. Emerging concepts in the pathogenesis of the *Streptococcus pneumoniae*: From nasopharyngeal colonizer to intracellular pathogen. *Cell. Microbiol.* **21**, e13077 (2019).
- Keller, L. E., Robinson, D. A. & McDaniel, L. S. Nonencapsulated *streptococcus pneumoniae*: emergence and pathogenesis. *MBio.* **7**, e01792 (2016).
- Zipfel, P. F., Hallström, T., Hammerschmidt, S. & Skerka, S. The complement fitness factor H: role in human diseases and for immune escape of pathogens, like pneumococci. *Vaccine.* **26**(Suppl 8), 167–74 (2008).
- Fernie-King, B., Seilly, D. J., Davies, A., Lachmann, P. J. Subversion of the innate immune response by micro-organisms. *Ann. Rheum. Dis.* **61**, Suppl 2:ii8–12 (2002).
- Zipfel, P. F., Hallström, T. & Riesbeck, K. Human complement control and complement evasion by pathogenic microbes—tipping the balance. *Mol. Immunol.* **56**, 152–160 (2013).
- Roosjakkars, S. H. & van Strijp, J. A. Bacterial complement evasion. *Mol. Immunol.* **44**, 23–32 (2007).
- Lambris, J. D., Ricklin, D. & Geisbrecht, B. V. Complement evasion by human pathogens. *Nat. Rev. Microbiol.* **6**, 132–142 (2008).
- Engholm, D. H., Kilian, M., Goodsell, D. S., Andersen, E. S. & Kjærgaard, R. S. A visual review of the human pathogen *Streptococcus pneumoniae*. *FEMS Microbiol. Rev.* **41**, 854–879 (2017).
- Andre, G. O. et al. Role of *Streptococcus pneumoniae* proteins in evasion of complement-mediated immunity. *Front Microbiol.* **8**, 224 (2017).
- Jedrzejewski, M. J. Pneumococcal virulence factors: structure and function. *Microbiol. Mol. Biol. Rev.* **65**, 187–207 (2001).
- Pérez-Dorado, I., Galan-Bartual, S. & Hermoso, J. A. Pneumococcal surface proteins: when the whole is greater than the sum of its parts. *Mol. Oral Microbiol.* **27**, 221–245 (2012).
- Chen, A. et al. Multivalent pneumococcal protein vaccines comprising pneumolysinoid with epitopes/fragments of CbpA and/or PspA elicit strong and broad protection. *Clin. Vacc. Immunol.* **22**, 1079–1089 (2015).
- Yuste, J. et al. The effects of PspC on complement-mediated immunity to *Streptococcus pneumoniae* vary with strain background. *J. Infect. Immun.* **78**, 283–292 (2010).
- Hammerschmidt, S., Talay, S. R., Brandtzaeg, P. & Chhatwal, G. S. PspA, a novel pneumococcal surface protein with specific binding to secretory immunoglobulin A and secretory component. *Mol. Microbiol.* **25**, 1113–1124 (1997).
- Rosenow, C. et al. Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of *Streptococcus pneumoniae*. *Mol. Microbiol.* **25**, 819–829 (1997).
- Dave, S., Brooks-Walter, A., Pangburn, M. K. & McDaniel, L. S. PspC, a pneumococcal surface protein, binds human factor H. *J. Infect. Immun.* **69**, 3435–3437 (2001).
- Jarva, H. et al. *Streptococcus pneumoniae* evades complement attack and opsonophagocytosis by expressing the pspC locus-encoded Hic protein that binds to short consensus repeats 8–11 of factor H. *J. Immunol.* **168**, 1886–1894 (2002).
- Zhang, J. R. et al. The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. *Cell* **102**, 827–837 (2000).
- Cheng, Q., Finkel, D. & Hostetter, M. K. Novel purification scheme and functions for a C3-binding protein from *Streptococcus pneumoniae*. *Biochem.* **39**, 5450–5457 (2000).
- Janulczyk, R., Iannelli, F., Sjöholm, A. G., Pozzi, G. & Björck, L. Hic, a novel surface protein of *Streptococcus pneumoniae* that interferes with complement function. *J. Biol. Chem.* **275**, 37257–37263 (2000).
- Dave, S., Carmicle, S., Hammerschmidt, S., Pangburn, M. K. & McDaniel, L. S. Dual roles of PspC, a surface protein of *Streptococcus pneumoniae*, in binding human secretory IgA and factor H. *J. Immunol.* **173**, 471–477 (2014).
- Binsker, U. et al. Serotype 3 pneumococci sequester platelet-derived human thrombospondin-1 via the adhesin and immune evasion protein Hic. *J. Biol. Chem.* **292**, 5770–5783 (2017).
- Lu, L. et al. Species-specific interaction of *Streptococcus pneumoniae* with human complement factor H. *J. Immunol.* **181**, 7138–7146 (2008).
- Lu, L., Ma, Y. & Zhang, J. R. *Streptococcus pneumoniae* recruits complement factor H through the amino terminus of CbpA. *J. Biol. Chem.* **281**, 15464–15474 (2006).

35. Hyams, C. *et al.* *Streptococcus pneumoniae* capsular serotype invasiveness correlates with the degree of factor H binding and opsonization with C3b/iC3b. *J. Infect. Immunol.* **81**, 354–363 (2013).
36. Köhler, S. *et al.* Binding of vitronectin and factor H to hic contributes to immune evasion of *Streptococcus pneumoniae* serotype 3. *Thromb. Haemost.* **113**, 125–142 (2015).
37. Haleem, K. S. *et al.* The pneumococcal surface proteins PspA and PspC sequester host C4-binding protein to inactivate complement C4b on the bacterial surface. *J. Infect. Immunol.* **87**, pii: e00742–18 (2018).
38. Dieudonné-Vatran, A. *et al.* Clinical isolates of *Streptococcus pneumoniae* bind the complement inhibitor C4b-binding protein in a PspC allele-dependent fashion. *J. Immunol.* **182**, 7865–7877 (2009).
39. Brooks-Walter, A., Briles, D. E. & Hollingshead, S. K. The PspC gene of *Streptococcus pneumoniae* encodes a polymorphic protein, PspC, which elicits cross-reactive antibodies to PspA and provides immunity to pneumococcal bacteremia. *J. Infect. Immunol.* **67**, 6533–6542 (1999).
40. Iannelli, F., Oggioni, M. R. & Pozzi, G. Allelic variation in the highly polymorphic locus pspC of *Streptococcus pneumoniae*. *Gene* **284**, 63–71 (2002).
41. Meinel, C. *et al.* *Streptococcus pneumoniae* from patients with hemolytic uremic syndrome binds human plasminogen via the surface protein PspC and uses plasmin to damage human endothelial cells. *J. Infect. Dis.* **217**, 358–370 (2018).
42. Luo, R. *et al.* Solution structure of choline-binding protein A, the major adhesin of *Streptococcus pneumoniae*. *EMBO J.* **24**, 34–43 (2005).
43. Van Rosmalen, M., Krom, M. & Merckx, M. Tuning the flexibility of glycine-serine linkers to allow rational design of multidomain proteins. *Biochemistry* **56**, 6565–6574 (2017).
44. Kim, H. K., Thammavongsa, V., Schneewind, O. & Missiakas, D. Recurrent infections and immune evasion strategies of *Staphylococcus aureus*. *Curr. Opin. Microbiol.* **15**, 92–99 (2012).
45. Håkansson, A. *et al.* Characterization of binding of human lactoferrin to Pneumococcal Surface protein A. *J. Infect. Immunol.* **69**, 3372–3381 (2001).
46. Hammerschmidt, S., Bethe, G., Remane, P. H. & Chhatwal, G. S. Identification of pneumococcal surface protein A as a lactoferrin-binding protein of *Streptococcus pneumoniae*. *J. Infect. Immunol.* **67**, 1683–1687 (1999).
47. Senkovich, O. *et al.* 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 (2007).
48. Xu, Q., Zhang, J. W., Chen, Y., Li, Q. & Jiang, Y. L. Crystal structure of the choline-binding protein CbpJ from *Streptococcus pneumoniae*. *Biochem Biophys Res Commun.* **514**, 1192–1197 (2019).
49. Kanchi, P. K. & Dasmahapatra, A. K. Polyproline chains destabilize the Alzheimer's amyloid- $\beta$  protofibrils: a molecular dynamics simulation study. *J. Mol. Graph. Model.* **93**, 107456 (2019).
50. Jerlström, P. G., Chhatwal, G. S. & Timrnis, K. N. The IgA-binding  $\beta$  antigen of the c protein complex of Group B streptococci: sequence determination of its gene and detection of two binding regions. *Mol. Microbiol.* **5**, 843–849 (1991).
51. Girgis, M. M., Abd El-Aziz, A. M., Hassan, R. & Ali, Y. M. Immunization with proline rich region of Pneumococcal surface protein A has no role in protection against *Streptococcus Pneumoniae* serotype 19F. *Microbiol Pathog.* **138**, 103761 (2020).
52. Mukerji, R. *et al.* The diversity of the proline-rich domain of pneumococcal surface protein A (PspA): potential relevance to a broad-spectrum vaccine. *Vaccine* **36**, 6834–6843 (2018).
53. McDaniel, L. S., Ralph, B. A., McDaniel, D. O. & Briles, D. E. Localization of protection-eliciting epitopes on PspA of *Streptococcus pneumoniae* between amino acid residues 192 and 260. *Microbiol. Pathog.* **17**(5), 323–337 (1994).
54. Hollingshead, S. K., Becker, R. & Briles, D. E. Diversity of PspA: Mosaic genes and evidence for past recombination in *Streptococcus pneumoniae*. *Infect. Immunol.* **68**(10), 5889–5900 (2000).
55. Maestro, B. & Sanz, J. M. Choline binding proteins from *Streptococcus pneumoniae*: a dual role as enzybiotics and targets for the design of new antimicrobial. *Antibiotics* **5**(2), 21 (2016).
56. Hakenbeck, R., Madhour, A., Denapaite, D. & Brückner, R. Versatility of choline metabolism and choline-binding proteins in *Streptococcus pneumoniae* and commensal streptococci. *FEMS Microbiol.* **3**(3), 572–586 (2009).
57. Marraffini, L. A., Dedent, A. C. & Schneewind, O. Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol Mol Biol Rev.* **70**, 192–221 (2006).
58. Pallen, M. J., Lam, A. C., Antonio, M. & Dunbar, K. An embarrassment of sortases—A richness of substrates?. *Trends Microbiol.* **9**, 97–101 (2001).
59. Daniels, C. C. *et al.* The proline-rich region of pneumococcal surface proteins A and C contains surface-accessible epitopes common to all pneumococci and elicits antibody-mediated protection against Sepsis. *J. Infect. Immunol.* **78**, 2163–2172 (2010).
60. McDaniel, L. S., McDaniel, D. O., Hollingshead, S. K. & Briles, D. E. Comparison of the PspA sequence from *Streptococcus pneumoniae* EF5668 to the previously identified PspA sequence from strain Rk1 and ability of PspA from EF5668 to elicit protection against pneumococci of different capsular types. *J. Infect. Immunol.* **66**(10), 4748–4754 (1998).
61. Georgieva, M., Kagedan, L., Lu, Y. J., Thompson, C. M. & Lipsitch, M. Antigenic variation in *Streptococcus pneumoniae* PspC promotes immune escape in the presence of variant-specific immunity. *mBio* **9**, 2.pii: e00264–18 (2018).
62. Desvaux, M., Dumas, E., Chafsey, I. & Hébraud, M. Protein cell surface display in Gram-positive bacteria: from single protein to macromolecular protein structure. *FEMS Microbiol Lett.* **256**, 1–15 (2006).
63. Pathak, A. *et al.* Factor H binding proteins protect division septa on encapsulated *Streptococcus pneumoniae* against complement C3b deposition and amplification. *Nat. Commun.* **9**, 3398 (2018).
64. van der Maten, E. *et al.* *Streptococcus pneumoniae* PspC subgroup prevalence in invasive disease and differences in contribution to complement evasion. *J. Infect. Immunol.* **86**, pii:e00010–18 (2018).
65. Chang, B. *et al.* Capsule switching and antimicrobial resistance acquired during repeated *Streptococcus pneumoniae* pneumonia episodes. *J. Clin. Microbiol.* **53**, 3318–3324 (2015).
66. Anderson, D. *et al.* Genome-wide association study of IgG1 responses to the choline-binding protein PspC of *Streptococcus pneumoniae*. *Genes Immunol.* **16**, 289–296 (2015).
67. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetic analyses version 7. *Mol. Biochem. Evol.* **33**(7), 1870–1874 (2016).

## Acknowledgements

The work of the authors is supported by the Collaborative Research Center, FungiNet (projects C6 (PFZ) and C4 (CS)) Deutsche Forschungsgemeinschaft (DFG). SL acknowledges a fellowship from the German Academic Exchange Service (DAAD) and from the International Leibniz Research School for Biomolecular Interaction (ILRS), Jena, Germany. Alfredo Sahagún-Ruiz was funded by a scholarship from PASPA-DGAPA, National Autonomous University of Mexico (UNAM), and from Mexican National Science and Technology Council (CONACYT) for a sabbatical at the Department of Infection Biology, Leibniz Institute for Natural Product Research and Infection Biology—Hans Knöll Institute, Jena, Germany. SH received funding by the Deutsche Forschungsgemeinschaft DFG HA 3125/5-2.

## 2 Modular structure of *S. pneumoniae* surface protein A: high level of domain-based sequence diversity may allow molecular strain typing

1  
2 **Modular Structure of *S. pneumoniae* Surface Protein A:**  
3 **High Level of Domain-Based Sequence Diversity May Allow Molecular**  
4 **Strain Typing**

5  
6 Shanshan Du<sup>1</sup>, Cláudia Vilhena<sup>1</sup>, Monika von Heide<sup>1</sup>, Dorit Fuest<sup>1</sup>, Sven Hammerschmidt<sup>2</sup>,  
7 Christine Skerka<sup>1</sup>, and Peter F. Zipfel<sup>1,3\*</sup>

8 <sup>1</sup>Department of Infection Biology, Leibniz Institute for Natural Product Research and  
9 Infection Biology, Jena, Germany.

10 <sup>2</sup>Department of Molecular Genetics and Infection Biology, Interfaculty Institute for Genetics  
11 and Functional Genomics, Center for Functional Genomics of Microbes, University of  
12 Greifswald, Greifswald, Germany.

13 <sup>3</sup>Institute of Microbiology, Friedrich-Schiller-University, Jena, Germany.

14 \*Corresponding author, e-mail: peter.zipfel@leibniz-jena.de

15 **Running title:** variations in *S. pneumoniae* PspA proteins

16 **Keywords:** immune evasion, complement escape, lactoferrin-binding.

17

### 18 **Abstract**

19 PspA proteins are mosaic structured immune evasion proteins expressed by the human  
20 pathobiont *Streptococcus pneumoniae*. In this study, we compared 48 PspA proteins derived  
21 from *S. pneumoniae* strains, as well as nine proteins from clinical isolates obtained from  
22 infants with hemolytic uremic syndrome. PspA proteins have three clearly separated regions:  
23 an N-terminal  $\alpha$ -helical region, a coiled-coil structured middle segment, and a  $\beta$ -sheet-  
24 structured region. Furthermore, we show that PspA separate into three families, and that  
25 each PspA protein comprises six domains. Sequence comparisons identified conserved,  
26 semi-conserved, variable, and even hypervariable domains, and several domains show a  
27 modular composition. Full-length PspA proteins display extreme sequence variability, and  
28 this sequence variability differs between domains. These analyses identified PspA-specific  
29 domains, but at the same time identified up to three domains also used by PspA and by PspC,  
30 including a lactoferrin-binding domain, proline-rich domains, and choline-binding domains.  
31 When this domain pattern analysis was applied to nine PspA variants derived from clinical *S.*  
32 *pneumoniae* strains from infant patients with pneumococcal hemolytic uremic syndrome, we

33 found that the new variants use domain patterns specific to PspA family II and III, but not  
34 family I. The variability appears to extend to the strain level, suggesting that this type of  
35 analysis, can be used as a form of molecular typing. Furthermore, the data may allow  
36 selection of conserved regions within specific surface-exposed, variable protein domains,  
37 e.g., for vaccine design.

38

### 39 **Introduction**

40 *Streptococcus pneumoniae* is a virulent human Gram-positive pathogenic bacterium that  
41 causes human airway infections, as well as otitis media, sinusitis, bronchitis and sepsis(1, 2).  
42 Upon spreading within the human organism and to niches, the pathogen can cause acute  
43 life-threatening conditions such as meningitis, sepsis, and pneumococcus-associated  
44 hemolytic uremic syndrome (pHUS). The latter occurs in children under the age of 5 years,  
45 and accounts for 5–15% of all HUS cases(3, 4). Currently, more than 100 *S. pneumoniae*  
46 serotypes have been identified based on polysaccharide differences(5). Only two  
47 pneumococcal conjugate vaccines, PCV13 and PCV23, are licensed for adults; these vaccines  
48 provide protection against 13 and 23 serotypes, respectively, of pneumococci(6). Work to  
49 improve current vaccination paradigms either by enlarging the coverage of serotype-  
50 independent vaccines or introducing new protein-based and whole cell vaccines is ongoing.

51 Pneumococcal surface protein A (PspA) is a central and highly variable, surface-  
52 exposed, immune evasion protein that mediates interaction with host plasma and immune  
53 cells, for example, by binding to the central human complement component C3(7, 8). PspA  
54 can also function as an adhesin for dying host cells by binding to host-derived GAPDH(9).  
55 PspA also binds lactoferrin, which may help *S. pneumoniae* overcome extreme conditions  
56 such as iron limitation(10). Different PspA alleles affect resistance of *S. pneumoniae* to killing  
57 by apolactoferrin(11). Due to its importance for pneumococcal virulence, PspA is one of the  
58 most promising vaccine targets on *S. pneumoniae*(12).

59 The known structures of PspA comprise an  $\alpha$  helical N-terminal region and a C-  
60 terminal choline-binding domain. PspA proteins are integrated into the cell wall by  
61 noncovalent interaction of the C-terminal repeat region with the choline residues of teichoic  
62 or lipoteichoic acids present in the pneumococcal cell wall(13). PspA variants can share  
63 domains with PspC and Hic two other pneumococcal immune evasion proteins. Variants of  
64 PspA and PspC can share up to three domains: a lactoferrin-binding segment, a proline-rich

65 domain, and a choline-binding domain(14). PspA and PspC cross-reaction antibodies are  
66 generated by infected organisms and which protect against sepsis(15).

67 The N-terminus of PspA contains  $\alpha$ -helical regions harboring charged residues, while the  
68 C-terminus contains a clade-specific segment, a proline-rich segment and a choline-binding  
69 domain (CBD). PspA proteins have a mosaic composition and are diverse(16, 17). However,  
70 our understanding of the structure and orientation of the PspA protein is incomplete. Many  
71 PspA variants do exist, the exact domain composition of single variants is unknown, and the  
72 sequence among strains shows variations and different environmental prevalence. Also, it is  
73 unclear how PspA is integrated into the bacterial cell wall, and how sequence variations affect  
74 domain composition and function.

75 The diversity of PspA proteins, and the apparent differences in domain composition,  
76 mean that the exact domain organization of single PspA proteins and structural composition  
77 are unclear. Based on sequence diversity, three separate clades have been identified(16).  
78 However, neither a clear domain composition of the members of each clade, nor differences  
79 in domain usage between different clades, have been defined. In addition, the position of the  
80 single protein regions are not clear, nor is the precise border between the domains. Also, we  
81 do not know which part of the protein is facing the outside, or exactly which part of the protein  
82 is integrated into the cell wall. Given the heterogeneity and our limited knowledge of the  
83 domain structure of PspA proteins, we analyzed the structural composition and domain  
84 composition of 48 PspA variants selected from the NCBI data bank, including the pathogenic  
85 reference strains D39 and TIGR4. In addition, we characterized PspA proteins from nine  
86 clinical strains isolated from infants with pHUS. The results suggest that PspA has a clear  
87 domain structure, and that the proteins separate into three distinct clades or families. We  
88 identified five new domains and several subdomains.

89 Taken together, the results presented herein increase our understanding of the  
90 structure and function of pneumococcal immune evasion proteins, and have implications with  
91 respect to (cross)reactivity of antibodies and vaccine design.

92

## 93 **Results**

94 **Selection of PspA variants.** PspA proteins from 48 *S. pneumoniae* strains were selected  
95 from the NCBI data bank, and redundant and identical proteins were removed. PspA<sub>GA07914</sub>  
96 from strain GA07914 was the largest protein, comprising 781 amino acids with a mass of

97 85.9 kDa. The smallest was PspA<sub>SP3-BS71</sub> from strain SP3-BS71, comprising 596 aa with a  
98 mass of 66.3 kDa (**Table 1**).

99

100 **PspA variants separate into three families.** Homology evaluation revealed that 47 full-  
101 length proteins and one truncated PspA<sub>WU2</sub> protein separated into three major groups, which  
102 we term families (**Figure 1A**). Family I comprises 21 members, including PspA<sub>D39</sub> from the  
103 pathogenic reference strain D39; family II comprises 11 members, including PspA<sub>G54</sub> from  
104 strain G54; and family III comprises 16 members, including PspA<sub>TIGR4</sub> from the pathogenic  
105 reference strain TIGR4.

106 A pairwise sequence-based similarity evaluation of all proteins confirmed the  
107 existence of these three groups, and identified overall similarities among the proteins, with  
108 scores ranging from 0.4 to 1.0 (**Figure 1B**). The members of each group, as highlighted in  
109 Figure 1B, showed the closest relationships (denoted by blue in the heat map), reaching  
110 similarity scores > 0.7. Homology with members of the other two families was less  
111 pronounced. One exception was PspA<sub>WU2</sub> (Protein Number 4) in family I; this may be because  
112 PspA<sub>WU2</sub> comprised only a partial sequence representing the N-terminal region (it lacked the  
113 C-terminal region).

114 **2D Structure of PspA proteins.** To better understand the structural organization and to  
115 identify potential differences among PspA proteins, we chose a representative from of each  
116 family and evaluated its length, secondary structure, and amino acid composition. The three  
117 selected proteins, PspA<sub>D39</sub>, PspA<sub>G54</sub>, and PspA<sub>TIGR4</sub>, differed in terms of length (619, 709, and  
118 744 aa long, respectively), although they shared related structures, as identified by secondary  
119 structure prediction programs. Each protein has a relatively long (ca. 300 aa), N-terminal  $\alpha$ -  
120 helical structured region, followed by a shorter (ca. 80 aa) coiled-coil structured region, and  
121 a (ca. 220 aa) C-terminal  $\beta$ -sheet structured region (**Figure 2A, Supplementary Figure 2A–**  
122 **C**). Because secondary structures are determined by aa composition, we analyzed the aa  
123 composition of each region. The N-terminal  $\alpha$ -helical regions included >40% charged amino  
124 acids, with a low fraction of polar/amphipathic residues (ca. 20%) and <2% tyrosine residues  
125 (i.e., 1.9%, 1.8%, and 2.3% for PspA<sub>D39</sub>, PspA<sub>G54</sub>, and PspA<sub>TIGR4</sub>, respectively) (**Figure 2B,**  
126 **left panel**). By contrast, the coiled-coil segment included more hydrophobic residues  
127 (ca. >46%), mostly proline residues (ca. >27%) (**Figure 2B, middle panel**). Whereas the C-  
128 terminal  $\beta$ -sheet region comprised fewer charged residues (ca. 15–17%), but had more

129 polar/amphipathic (>30%) and more tyrosine and tryptophan residues (ca. 9–11%) (**Figure**  
130 **2B, right panel**) than the N-terminal  $\alpha$ -helical regions.

131

132 **The N- and the C-terminal regions differ with respect to homology.** Given the differences  
133 in structure and aa composition, we performed a pairwise sequence comparison between the  
134 N- and C-terminal regions of the 48 proteins. Sequence homology evaluation of the N-  
135 terminal region separated the three families, and again revealed the same family separation  
136 as the full-length proteins (**Figure 3A**). The C-terminal regions in contrast to the N-terminal  
137 regions, lacked family separation. In this case, the members of the three families could not  
138 be clearly distinguished from each other (**Figure 3B**). These results highlight that the N-, but  
139 not C-terminal, region determines family separation. This conclusion is also supported by  
140 sequencing of the N- and C-terminal regions. A sequence comparison performed separately  
141 showed that the N-terminal regions are diverse, and show a high level of sequence variability  
142 (denoted by red on the heat map) (**Figure 3C, upper-right panel**). By contrast, the C-  
143 terminal regions are more closely related and more conserved, with sequence similarities of  
144 up to 99% (shown in blue) (**Figure 3C, lower-left panel**). Again, PspA<sub>wuz</sub> (entry:  
145 AAF27710.1), of which we had only a partial sequence, was the exception as it lacks the  $\beta$ -  
146 sheet region. The difference between the N- and C-terminal regions may be indicative of the  
147 spatial separation and functional specialization.

148

149 **PspA domain organization, composition, and variability.** Next, we asked whether PspA  
150 proteins have a domain organization. For this purpose, we defined regions as protein  
151 segments that are related with respect to structure. Such long regions (80–300 aa) can be  
152 dissected into domains. Domains shared the same structure, but varied in length from ca.  
153 20–259 aa. Individual domains also varied with respect to amino acid composition and  
154 showed different degrees of sequence homology. Furthermore, some domains had a  
155 modular composition. Modular elements appeared to be domain-specific, and single modules  
156 differed in length and sequence (see supplementary **file X – alignment**). This domain  
157 composition analysis also revealed that each family has a preferred domain pattern, and that  
158 each shows a domain preference.

159

160 **PspA has a domain organization.** To analyze PspA domain organization in detail, the full-  
161 length PspA sequences were arranged into family groups. Subsequent sequence alignment

162 revealed the domain composition of all proteins and showed that individual domains differ in  
163 terms of both length and homology (**Figure 4**). However, the length of the signal peptides is  
164 identical, and the sequence is well conserved (93.5% conserved residues) (**Supplementary**  
165 **file X – alignment**). The first domain of the mature protein (D39: 112 aa), which is highly  
166 variable in sequence (26.5–30.3% conserved residues among the three families), was  
167 therefore termed the hypervariable domain (HVD) (**Supplementary Figure 1**). The second  
168 domain (19.1–70.0% conserved residues) is also variable and was named the variable  
169 domain. There are three variants of domain 3. Each belongs to one of the three family groups.  
170 Domain 3 was named the **f**amily **d**etermining **d**omain, FDD. Domain 4 is the domain sited  
171 furthest downstream of the  $\alpha$ -helical region (D39: 20 aa). Domain 5 is a **p**roline-**r**ich **d**omain  
172 (PRD), which has a modular composition and separates into two subvariants (gray box in  
173 **Figure 5**) (D39: 84 aa). **Domain 6**, the most C-terminal domain, is the CBD (D39: 218 aa).  
174 These six domains vary in length. Domain 4 (20–29 aa) is the shortest, and Domain 6 is the  
175 longest (200–260 aa) (**Figure 4**). The individual domains show different degrees of sequence  
176 variation, ranging from hypervariable, variable, semi-conserved, to conserved  
177 (**Supplementary Figure 1**). In addition, the FDD<sub>III</sub>, the PRDs, and the CBDs have a modular  
178 composition and show modular variability. FDD<sub>III</sub> comprises either two or three modules,  
179 whereas the number of modules in the CBDs varies from 10 (PspA<sub>GA13430</sub>) to 13 (PspA<sub>GA07914</sub>).  
180

181 **Domain structure.** After signal peptide cleavage, mature PspA proteins comprise six  
182 domains (**Figure 5A**). These domains, from N-terminal to C-terminal, are as follows:

183 *Domain 1 (HVD):* The first domain of the N-terminus is specific to PspA proteins and shows  
184 a high degree of sequence diversity (**Table 2**). The length of the HVD is rather similar  
185 between families, around 110 aa long (although it can be up to 130 aa)(**Figure 4**). The HVDs  
186 of the three families harbor 16 conserved residues, located mainly at the N-terminus  
187 (**Supplementary file X – alignment**).

188 *Domain 2 (VD):* A blast search revealed that the VD is unique to PspA. The most striking  
189 feature of the VD is the high fraction of both negatively and positively charged amino acids.  
190 VD split into two main clusters (VD-A and VD-B). Their length is clearly different between the  
191 three families, ranging from 36–83 aa. As shown in **Figure 4**, the VD-A of family I is up to 44  
192 aa long, whereas the VD-B of family II and III is 83 aa long. The VDs share five conserved  
193 residues located at the N-terminus. The VD of family I proteins has three conserved residues,  
194 that of family II has 33, and that of family III has 51. The VDs of the three families differ in



195 terms of length and aa composition. In particular, VD-B contains characteristic conserved  
196 residues. The homology of VDs within each family is more pronounced than that of the HVDs  
197 (see Supplementary **file X – alignment**).

198 *Domain 3 (FDD)*: The third domain has three distinct variants that clearly segregate with the  
199 families identified by homology tree analyses. A homology tree based specifically on Domain  
200 3 shows the separation of the three families, as do trees which are based on full-length  
201 proteins or the N-terminal  $\alpha$ -helical regions (**Supplementary Figure 3C**). The three variants  
202 of Domain 3 qualify as FDDs, and they include the following:

203 FDD<sub>I</sub> (family I), a lactoferrin-binding domain. FDD<sub>I</sub> of family I PspA proteins (i.e., PspA<sub>D39</sub>) are  
204 up to 110 aa long and have 52 conserved residues. This domain corresponds to the  
205 lactoferrin-binding region identified previously for some PspA variants(18). FDD<sub>I</sub>-related  
206 domains are also found in *S. pneumoniae* surface proteins PspC1.1 and PspC5.1. The FDD<sub>I</sub>  
207 of PspA<sub>DBL6A</sub>, and the corresponding region in PspC1.1, share 84.5% sequence identity(14).

208 FDD<sub>II</sub> (family II), which has a conserved core element is common to all eleven members of  
209 family II. These domains can be up to 134 aa long and are rich in charged amino acids. Those  
210 in family II harbor 64 conserved residues, with a prominent, conserved 18 aa long stretch  
211 (**T<sub>70</sub>LDPE-EAAE<sub>98</sub>**; numbering based on PspA<sub>D39</sub>) in the middle region. FDD<sub>III</sub> (family III), is  
212 a modular FDD which is common to all members of family III. The composition with two or  
213 three modules identifies further subvariants in this family. All members share the first 88 aa  
214 long, and the last 14 aa long, modules. Nine variants have an extra 81 aa-long module  
215 inserted in the middle. The three modules have related N-terminal sequence motifs of the  
216 type **G<sub>1</sub>(A/V)<sub>2</sub>D<sub>3</sub>(P/S)<sub>4</sub>(-/E)<sub>5</sub>D<sub>5/6</sub>D<sub>6/7</sub>**. The subgroups found in family III are defined according  
217 to the two or three modules in FDD<sub>III</sub>, and on the type of PRD.

218 *Domain 4 (D4)*. D4 of PspA proteins is relatively short, comprising 20, 22, or 29 residues. D4  
219 is specific to PspA, and lacks homology with other *S. pneumoniae* proteins in the database.  
220 This domain has three essential conserved residues (i.e., E<sub>7</sub>, L<sub>15</sub>, and A<sub>18</sub> (PspA<sub>D39</sub>)). The  
221 domains segregate with the PspA subfamilies, with each family harboring more conserved  
222 residues. D4 of family I proteins is 20 aa long and has nine conserved residues. D4 of family  
223 II proteins comprise either 22 aa (four proteins) or 29 aa residues (four proteins), with a total  
224 of 11 conserved residues. D4 of family III proteins has 20 conserved residues out of 22  
225 residues. There are 13 variants of D4.

226 *Domain 5 (PRD)*. Domain 5 has a coiled-coil structure and is about 80 aa long. The coiled-  
227 coil structure provides structural flexibility for the entire protein(19). The PRD connects the

228 N-terminal  $\alpha$ -helical region with the C-terminal CBD. PRDs have multiple proline (Pro)  
229 residues, but only Pro<sub>2</sub> is conserved. Despite a low degree of sequence homology, the PRDs  
230 have a characteristic modular composition. The PRDs comprise multiple related modules,  
231 many of which include palindromic **PAPAP** core motifs(20, 21). These modules vary in size,  
232 assembly order, and spacing. Domain 5 splits into two major subtypes: PRD-1<sub>QQ-QQQ</sub> is found  
233 in 29 PspA variants and is about 118 aa long. PRD-2 comprises multiple modules and is up  
234 to 84 aa long. PRD-1 is composed of three model-type segments. The first is up to 57 aa  
235 long and includes smaller modules with **PAPAP** motifs. The second is a 22 aa long conserved  
236 “QQ-QQQ” segment, and the third is up to 23 aa long and also includes modular elements  
237 containing common **PAPAP** motifs.

238 The first segment includes modules with a common core PAPAP motif. The second  
239 (middle) segment has characteristic flanking glutamine (Q) residues and has a conserved  
240 length. This middle segment with characteristic QQ-QQQ borders has 20 conserved residues  
241 and has alternating residues only at two positions. Interestingly, this segment lacks Pro, and  
242 is therefore referred in the literature also as ‘non-proline stretch’. This segment is rich in  
243 charged residues, and adopts a  $\alpha$ -helical structure. The third segment includes different  
244 numbers of modules containing common **PAPAP** motifs. The modules are 11, 22, or 23 aa  
245 long, include up to ten Pro residues, most have 12 conserved residues, and all have a C-  
246 terminal lysine (K). The first and third modules include sections containing **PAPAP** motifs.  
247 PRD-1 has been found in the pneumococcal proteins PspC3.1, PspC2.2, and PspC6.1(14).  
248 PRD-2 can be up to 84 aa long and comprises multiple small modules or units, most of which  
249 include the palindromic **PAPAP** core. These five aa motifs can represent a module in itself.  
250 However also longer PAPAP-containing module variants with N-terminal and C-terminal  
251 extensions exist, including upstream or downstream flanking residues. By assembling related,  
252 but clearly different, modules in a variable number or combinations, PRD-2 (as a PAPAP-  
253 based modular domain) provides a new format for how pattern diversity is generated.

254 *Domain 6 (CBD)*. The most C-terminal region of PspA proteins has a  $\beta$ -sheet structure and  
255 contains the CBD. This common cell wall anchor is used by PspA and by 13 other  
256 pneumococcal proteins, including LytA, LytB, LytC, and PspC(22). Studies show that the W<sub>3</sub>  
257 and W<sub>10</sub> of one module ( $n$ ), and the Tyr<sub>11</sub> of the next module ( $n + 1$ ), attach to the same choline  
258 moiety and form  $\pi$  contacts (23, 24). CBDs are modular units, and PspA variants comprise a  
259 different number of modules (ranging from 9 to 13). A single module is formed by apparent  
260 combination of conserved and variable aa at specific positions. This suggests that aa at a

261 given position affect structure and folding, whereas residues at other positions mediate  
262 choline binding.

263

#### 264 **Typing of clinical HUS strains**

265 To determine whether this domain-based analysis is useful for evaluating new PspA proteins,  
266 this domain-based analysis was applied to nine PspA<sub>HUS</sub> proteins. The sequence were  
267 derived from three strains isolated from infants with pneumococcal hemolytic uremic  
268 syndrome (pHUS) (PspA<sub>HUSA</sub>-PspA<sub>HUSC</sub>). In addition six PspA<sub>HUS</sub> genes were extracted from  
269 genome data from clinical pHUS strains (PspA<sub>HUSd</sub>-PspA<sub>HUSi</sub>). The encoded PspA<sub>HUS</sub>  
270 proteins are between 686–774 aa in length, and all represent new, unique variants. Two  
271 PspA<sub>HUS</sub> proteins matched the pattern profile of family II; four matched that of family IIIB  
272 (harboring PRD1); and three matched that of family IIIC (harboring PRD2) (**Figure 5B**). None  
273 of the PspA<sub>HUS</sub> variants matched the domain pattern of family I. Thus, domain pattern  
274 evaluation is useful for analyzing new PspA variants, and suggests that PspA<sub>HUS</sub> proteins  
275 from clinical HUS strains have a family as well as domain preference.

276

#### 277 **The diversity of PspA affects binding to human lactoferrin.**

278 Because the FDD of the three families have different sequences, and family I and family II  
279 PspAs bind to lactoferrin(18), we wondered whether this was the case for family III.  
280 Representative strains (D39, G54, and TIGR4) from each family were tested for their ability  
281 to bind human lactoferrin using a solid-phase binding assay. As shown in **Figure 6A**, the  
282 three strains express PspA<sub>D39</sub>, PspA<sub>G54</sub>, and PspA<sub>TIGR4</sub>, in line with their theoretical molecular  
283 mass (65.4 kDa, 75.5 kDa, and 79.5 kDa, respectively). Immobilized lactoferrin bound  
284 significantly more PspA<sub>G54</sub> and PspA<sub>TIGR4</sub> than PspA<sub>D39</sub>; there was no difference between  
285 PspA<sub>G54</sub> and PspA<sub>TIGR4</sub> (**Figure 6B**). Our hypothesis, therefore, is that PspAs from the three  
286 families bind human lactoferrin with different affinities. Two strains derived from HUS patients  
287 were tested at the same time. Figure 6B shows that human lactoferrin bound more PspA<sub>HUS1</sub>  
288 and PspA<sub>HUS2</sub> from family III than PspA<sub>D39</sub> from family I.

289

#### 290 **Discussion**

291 Most strains of the human pathobiont *S. pneumoniae* have *pspA* genes that encode variable  
292 surface-exposed immune evasion proteins. Here, we show that PspA proteins have a mosaic  
293 structure, and that they are diverse in terms of structure and sequence. These central

294 pneumococcal immune evasion proteins combine domain and extreme sequence diversity  
295 with a conserved structure, immune evasion functions, and cell wall anchoring. By comparing  
296 48 PspA proteins, we show that each comprises six domains and belongs to one of three  
297 families. The pattern and sequence variability of PspA extend to the strain level, suggesting  
298 that PspA sequence analyses may qualify as a kind of molecular strain typing. Such  
299 information can allow identification of conserved regions in certain proteins, particularly within  
300 the surface-exposed, hypervariable domains, which are relevant to vaccine design.

301

302 **PspA have three structural regions.** By analyzing the secondary structure of PspA proteins,  
303 we show that PspA proteins comprise three differently structured regions. The N-terminal  $\alpha$ -  
304 helical regions are rich in charged residues; these are followed by coiled-coil structured regions  
305 with many hydrophobic residues. The C-terminal region comprises  $\beta$ -sheets rich in Tyr and  
306 Trp, and in polar amphipathic residues.

307

308 **PspA proteins generate diversity in multiple ways.** Pattern- and sequence-based  
309 inspection of 48 PspA proteins showed that each comprises six domains selected from a  
310 portfolio of nine possible domains. Domain pattern distribution identified three major PspA  
311 families. Each has a preference for certain domain combinations, but exceptions do exist within  
312 subfamilies. Families are separated based on the HVD, VD, and, particularly, the FDD regions.  
313 Five domains (HVD, VD, FDD<sub>II</sub>, FDD<sub>III</sub>, and D4) are specific to PspA. By contrast, the FDD<sub>I</sub>  
314 lactoferrin-binding domain is also found in PspC1.1 and PspC5.1. PRD-1 of PspA and PRD-  
315 2 are shared with other pneumococcal proteins, but have a different 'proline segment core'  
316 (e.g., PspC1.1 and PspC2.2). Due to the remarkable similarity between the PRD-1 of PspA  
317 and that of PspC, the antibodies elicited by PspC containing the first and third segments of  
318 PRD-1 cross-protect mice from infection by a pneumococcal strain lacking a *pspC* gene(25).  
319 Immunity to the proline-rich domain present in PspC can be protective via cross-reactions with  
320 the proline-rich domain of PspA (PRD-1 in this study)(26). In addition, CBDs are found in other  
321 pneumococcal cell wall-integrated proteins, including PspC, LytA, and LytB(27).

322

323 PspA proteins generate diversity by assembling different domains in variable  
324 combinations; four domains (FDD<sub>III</sub>, PRD-1, PRD-2, and CBD) provide modular diversity, and  
325 all tested strains show sequence variation. Furthermore, these domains have a modular  
326 composition. These modules represent conserved, repetitive domain-specific elements. In

326 addition, modules within the same domains are mostly related to each other with respect to  
327 length and sequence. Assembly of different numbers of modules alters domain length.

328 This domain-based characterization confirms and extends information about the clade-  
329 defining region(s) of PspA, which were reported previously as being positioned upstream of  
330 the PRDs(28).

331 The NCBI data bank includes almost 22,106 (2021.08.09) entries for either full-length  
332 PspA proteins or partial PspA sequences. Application of such a domain-based pattern  
333 approach to all entries in the database is expected to improve the characterization of  
334 individual proteins, and may identify new domains.

335 Because individual PspA proteins differ in terms of domain composition and sequence,  
336 we recommend always to combine the gene or protein names with the strain name, resulting  
337 for example in designations such as PspA<sub>D39</sub>, PspA<sub>TIGR4</sub>, or PspA<sub>HUS-A</sub>. This nomenclature  
338 reflects the individuality of single PspA proteins, gives the strain designation, and is more  
339 informative when comparing proteins from different strains.

340 **PspA<sub>HUS</sub>**. Application of this new domain pattern approach to nine PspA<sub>HUS</sub> variants from  
341 clinical pneumococcal strains showed that each isolate encoded a unique protein, and that  
342 PspA<sub>HUS</sub> variants had a preference for family II and III; there was no match with family I.

343 Immobilized lactoferrin bind significantly more PspA<sub>HUSA</sub>, PspA<sub>HUSB</sub>, and PspA<sub>TIGR4</sub>  
344 family III proteins than PspA<sub>D39</sub> (family I), which might have an effect on immunomodulatory  
345 function during infection. Lactoferrin is found in human serum during inflammatory responses  
346 against *Escherichia coli* infection(29, 30). The preference of PspA<sub>HUS</sub> for family II and family  
347 III domain patterns, which bind more lactoferrin, might contribute to the pathogenesis of HUS.  
348 Further studies are needed to ascertain whether the lactoferrin-binding function is related  
349 directly to differences among FDDs (FDD<sub>I</sub>, FDD<sub>II</sub>, and FDD<sub>III</sub>).

350

351 **Topology: Orientation.** Following an N- to C-terminal orientation, the domain sequence  
352 ranges from hypervariable, to variable, to semi-conserved, to conserved. Cell wall attachment  
353 via the C-terminal CBD orients the preceding PRD towards the outside, suggesting a cell wall  
354 and capsule spanning role, or a likely function as a surface-exposed domain. This suggests  
355 that the N-terminal domains are exterior-facing. The different secondary structures are in  
356 agreement with likely exposure to lipophilic and hydrophilic milieus. The  $\alpha$ -helical structure,  
357 the many charged residues, and the overall hydrophilic nature also suggest an exterior  
358 location for the N-terminal region. This topology also supports the high sequence variability

359 of the HVD and VD, which reflects antigenic variation. In a previous study, PRD-recognizing  
360 antibodies were generated upon immunization of mice with recombinant isolated PRD. This  
361 suggests that this domain is exposed to the environment and is accessible to antibodies(21).  
362 PRD-1, which is shared by PspA and PspC, has an internal non-proline block, with  
363 characteristic flanking "QQ-QQQ" residues. This QQ-QQQ module can enhance virulence by  
364 binding to host lactate dehydrogenase(31). PRD can be considered as a flexible linker between  
365 the N-terminal  $\alpha$ -helical and the C-terminal  $\beta$ -sheet structures, and may form an interface  
366 between the outside environment and the bacterial cell wall.

367 **Domain pattern variations exist in other pneumococcal proteins.** The three distinct  
368 structural regions, a modular composition, and sequence variability(as described here for PspA)  
369 also apply to other pneumococcal proteins. PspC and Hic have similar  $\alpha$ -helical structured N-  
370 terminal regions with many charged residues, as well as hypervariable, variable, and  
371 conserved domains. Both have related mosaic structures, and show domain pattern variability  
372 and domain dependent sequence variation. In addition, PspC, but not Hic variants, have PRD-  
373 1 and CBDs that are highly similar to those of PspA. Thus, a combined evaluation of these  
374 three multivariant genes, or the encoded PspA, PspC, and Hic proteins, will improve the  
375 precision of domain- and sequence-based strain typing.

376 **Consequences for pathogen host interactions.** Immune escape by *S. pneumoniae* and of  
377 pathogenic microbes in general involves control of the complement cascade as well as  
378 inhibition of innate and adaptive immunity. PspA, as well as PspC and Hic, contribute to  
379 complement evasion and control of adaptive immune responses. The high diversity of these  
380 three pneumococcal immune evasion proteins shows that the pathogen has developed special  
381 answers or responses that generate variability, which assists with the immunobiological arms  
382 race with the human host.

383 **Pneumococci and the human host generate diversity in different ways.** The human host  
384 generates antibody and T-cell diversity to induce a highly specific, virtually unlimited immune  
385 response. A human organism can probably produce more than  $10^{12}$  different antibodies, some  
386 of which cross-react with a variety of related but different antigenic determinants(32). Diverse  
387 T-cell receptor (TCR)- $\alpha\beta$  gene segments combine randomly to form a large diverse and  
388 polymorphic T-cell repertoire to protect the human host against pathogens(33). The  
389 mechanistic details regarding how *S. pneumoniae* generates such large strain diversity is less  
390 well understood. It will be informative to evaluate the exact molecular mechanism(s) and  
391 selection mechanisms that underlie this enormous diversity.

392 **Perspectives.** *S. pneumoniae* PspA proteins are one of the most variable surface-exposed  
393 pneumococcal proteins identified to date. Here, we show that PspA has a domain  
394 composition with extreme sequence diversity. Evaluating domain pattern and sequence  
395 variations among known and new pneumococcal strains, and of clinical isolates, will likely  
396 identify new domains, new domain combinations, and more subfamilies and/or clades. A  
397 detailed understanding of the domain composition, domain pattern variability, and sequence  
398 diversity among PspAs will increase our understanding of their functions, and facilitate  
399 selection of surface-exposed and conserved protein regions that may be targets for vaccine  
400 design. In addition, evaluating the high variability of PspA, either alone or in combination with  
401 other pneumococcal variant proteins, may allow molecular strain typing.

402 **Materials and Methods**

403 **Selection of PspA proteins.** After redundant PspA sequences were removed, 48  
404 PspA identical sequences were extracted from NCBI protein database. As far as could  
405 be traced, the strains were collected from 2000 to 2018. The protein names,  
406 corresponding bacterial strain, protein size, GenBank Accession number are shown in  
407 (Table 1).

408 **Secondary structure evaluation.** The structure ( $\alpha$ -helical, coiled-coil and  $\beta$ -sheet) of  
409 each selected PspA was evaluated using RaptorX (<http://raptorx.uchicago.edu/>). The  
410 secondary structure prediction of PspA<sub>D39</sub>, PspA<sub>G54</sub>, PspA<sub>TIGR4</sub> are shown by  
411 histograms in the supplementary Figure 2 which were constructed using ggplot2 from  
412 the R/Bioconductor.

413 **Phylogenetic analysis.** The PspA amino acid sequences and composition were  
414 evaluated using MEGA7 ([www.megasoftware.net](http://www.megasoftware.net)). The CLUSTALW program and the  
415 BLOSUM amino acid matrix were used to compare the allelic variants of PspA,  
416 following which phylograms were generated using the Neighbor-Joining method  
417 (Bootstrap value:100). The phylogram for each domain was generated using the same  
418 method. Phylogenetic trees are modified in MEGA7. The amino acid composition of 7  
419 PspA representatives was shown by histograms in GraphPad 8.

420 **Domain homology searches.** BLASTp was used to identify related proteins or protein  
421 segments within the GenBank database available at the National Center for  
422 Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Furthermore, BLAST  
423 targeting database Uniport reference proteomes plus Swiss-Prot was used to find  
424 regions of local similarity between sequences (<https://www.uniprot.org/blast/>). All the  
425 domains in this work have been done a blast.

426  
427 **Pneumococcal strains and growth conditions.** *S. pneumoniae* D39 (NCTC 7466,  
428 serotype 2), G54 (serotype 19F) and TIGR4 (serotype 4) were provided by collaborator.  
429 Three *S. pneumoniae* strains were cultivated either on Columbia 5% sheep blood agar  
430 plates or incubated overnight in Todd-Hewitt broth supplemented with 0.5% yeast  
431 extract (THY) in an incubator at 37 °C with 5% CO<sub>2</sub>.

432 **DNA Sequencing.** The *pspA* genes of *S. pneumoniae* isolated from HUS patients  
433 were sequenced by using a primer walking strategy. Strains genomic DNA were  
434 extracted by using GenElute™ bacterial genomic DNA kit (Sigma-Aldrich). The PCR  
435 fragments containing the *pspA* locus was amplified by flank primers (Table 3).



436 Amplified PCR products were sequenced using a capillary sequencer (3130x/ Genetic  
437 Analyzer, Thermo Fisher Scientific GmbH, Germany). Software Clone manager 9.0  
438 (Sci-Ed Software) was used to analyze the sequencing data. DNA sequences were  
439 assembled and translate to amino acid sequenced, then blasted in the protein  
440 database NCBI BLASTP 2.5.0 (2018).

441 **SDS-PAGE, Western Blot.** Native Choline-binding proteins were extracted by choline  
442 wash as described by Frias(34). Grow *S. pneumoniae* strains in THY media until  
443 OD600 nm reaching to 0.9 (6 hour after inoculation). Take 10 ml cultural media and  
444 harvest cell by centrifugation (3200 g for 10 min at 4 °C). Pellets were washed by 5 ml  
445 DPBS twice. Then resuspend the pellets in 200 µl of 2% choline chloride (w/v)  
446 prepared in 1x DPBS and incubate 30 min at 4 °C without agitation. Choline-binding  
447 proteins mixture were harvested in the supernatant after maximum speed  
448 centrifugation to get rid of cells. Choline-binding proteins mixture were separated by  
449 10% SDS-PAGE and transferred to nitrocellulose membrane. The membranes were  
450 incubated polyclonal rabbit anti-PspA: QP2(10). QP2 were detected using anti-rabbit-  
451 conjugated HRP and a chemiluminescent substrate.

452 **Enzyme-Linked Immunosorbent Assay (ELISA).** 10 ug/ml lactoferrin and BSA were  
453 immobilized onto a high-binding infinity microtiter 96 wells plate and incubated  
454 overnight at 4 °C. Plates were washed three times with ice-cold DPBS, and nonspecific  
455 binding sites were blocked with blocking buffer (Applichem) for 1 hour at room  
456 temperature (RT). After three more washing steps, 50 ul native choline-binding  
457 proteins mixture extracted was added and incubated for 2 hours at room temperature.  
458 Plates were then washed three times with DPBS and incubated with polyclonal rabbit  
459 anti-PspA: QP2 diluted in DPBS for one hour. After washing, primary antibodies were  
460 detected using anti-rabbit HRP-conjugated antibody and detected using TMB  
461 substrate solution (Thermo scientific, cat: 00-4201-56). The reaction was measured at  
462 450nm.

463 **Acknowledgments**

464 The work of the authors is supported by the Deutsche Forschungsgemeinschaft  
465 (DFG). SD acknowledges a fellowship from the German Academic Exchange  
466 Service (DAAD) and from the International Leibniz Research School for  
467 Biomolecular Interaction (ILRS), Jena, Germany. SH received funding by the  
468 Deutsche Forschungsgemeinschaft DFG HA 3125/5-2.

469 **Author contributions**

470 PFZ, SD, CV, and CS designed the research; SD performed research, and  
471 analyzed data together with PFZ and CS; PFZ, SD, CV, SH, AS, SK, and CS  
472 wrote the manuscript.

473 **Reference**

- 474 1. Grevers, G. Challenges in reducing the burden of otitis media disease: An ENT  
475 perspective on improving management and prospects for prevention. *Int. J. Pediatr.*  
476 *Otorhinolaryngol.* **74**, 572–577 (2010).
- 477 2. Bogaert, D., De Groot, R. & Hermans, P. W. M. *Streptococcus pneumoniae* colonisation:  
478 The key to pneumococcal disease. *Lancet Infectious Diseases* (2004)  
479 doi:10.1016/S1473-3099(04)00938-7.
- 480 3. Musher, D. M. Infections caused by *Streptococcus pneumoniae*: clinical spectrum,  
481 pathogenesis, immunity, and treatment. *Clin. Infect. Dis. an Off. Publ. Infect. Dis. Soc.*  
482 *Am.* **14**, 801–807 (1992).
- 483 4. Scobell, R. R., Kaplan, B. S. & Copelovitch, L. New insights into the pathogenesis of  
484 *Streptococcus pneumoniae*–associated hemolytic uremic syndrome. *Pediatr. Nephrol.*  
485 **35**, 1585–1591 (2020).
- 486 5. Ganaie, F. *et al.* A new pneumococcal capsule type, 10D, is the 100th serotype and has  
487 a large cps fragment from an oral streptococcus. *MBio* **11**, 1–15 (2020).
- 488 6. Matanock, A. *et al.* Use of 13-Valent Pneumococcal Conjugate Vaccine and 23-Valent  
489 Pneumococcal Polysaccharide Vaccine Among Adults Aged ≥65 Years: Updated  
490 Recommendations of the Advisory Committee on Immunization Practices. *MMWR.*  
491 *Morb. Mortal. Wkly. Rep.* **68**, 1069–1075 (2019).
- 492 7. Tu, A. H. T., Fulgham, R. L., Mccrory, M. A., Briles, D. E. & Szalai, A. J. Pneumococcal  
493 surface protein A inhibits complement activation by *Streptococcus pneumoniae*. *Infect.*  
494 *Immun.* **67**, 4720–4724 (1999).
- 495 8. Zipfel, P. F. & Skerka, C. Complement regulators and inhibitory proteins. *Nat. Rev.*  
496 *Immunol.* **9**, 729–740 (2009).
- 497 9. Park, S. S. *et al.* *Streptococcus pneumoniae* binds to host GAPDH on dying lung  
498 epithelial cells worsening secondary infection following influenza. *Cell Rep.* **35**, 109267  
499 (2021).

- 500 10. Hammerschmidt, S., Bethe, G., Remane, P. H. & Chhatwal, G. S. Identification of  
501 pneumococcal surface protein A as a lactoferrin- binding protein of *Streptococcus*  
502 *pneumoniae*. *Infect. Immun.* **67**, 1683–1687 (1999).
- 503 11. Mirza, S. *et al.* The effects of differences in *pspA* alleles and capsular types on the  
504 resistance of *Streptococcus pneumoniae* to killing by apolactoferrin. *Microb. Pathog.* **99**,  
505 209–219 (2016).
- 506 12. Briles, D. E. *et al.* Pneumococcal diversity: Considerations for new vaccine strategies  
507 with emphasis on pneumococcal surface protein A (PspA). *Clin. Microbiol. Rev.* **11**,  
508 645–657 (1998).
- 509 13. Yother, J. & White, J. M. Novel surface attachment mechanism of the *Streptococcus*  
510 *pneumoniae* protein PspA. *J. Bacteriol.* **176**, 2976–2985 (1994).
- 511 14. Du, S. *et al.* Molecular analyses identifies new domains and structural differences  
512 among *Streptococcus pneumoniae* immune evasion proteins PspC and Hic. *Sci. Rep.*  
513 1–15 (2021) doi:10.1038/s41598-020-79362-3.
- 514 15. Schachern, P. A. *et al.* Pneumococcal PspA and PspC proteins: Potential vaccine  
515 candidates for experimental otitis media. *Int. J. Pediatr. Otorhinolaryngol.* **78**, 1517–  
516 1521 (2014).
- 517 16. Hollingshead, S. K., Becker, R. & Briles, D. E. Diversity of PspA: Mosaic genes and  
518 evidence for past recombination in *Streptococcus pneumoniae*. *Infect. Immun.* **68**,  
519 5889–5900 (2000).
- 520 17. Jedrzejewski, M. J., Lamani, E. & Becker, R. S. Characterization of Selected Strains of  
521 Pneumococcal Surface Protein A. *J. Biol. Chem.* **276**, 33121–33128 (2001).
- 522 18. Roche, H. *et al.* Characterization of Binding of Human Lactoferrin to Pneumococcal  
523 Surface Protein A. *Infect. Immun.* **69**, 3372–3381 (2002).
- 524 19. Williamson, M. P. The structure and function of proline-rich regions in proteins. *Biochem.*  
525 *J.* **297**, 249–260 (1994).

- 526 20. Mukerji, R. *et al.* The diversity of the proline-rich domain of pneumococcal surface  
527 protein A (PspA): Potential relevance to a broad-spectrum vaccine. *Vaccine* **36**, 6834–  
528 6843 (2018).
- 529 21. King, J. *et al.* The Proline-Rich Region of Pneumococcal Surface Proteins A and C  
530 Contains Surface-Accessible Epitopes Common to All Pneumococci and Elicits  
531 Antibody-Mediated Protection against Sepsis. *Infect. Immun.* **78**, 2163–2172 (2010).
- 532 22. Briles, D. E. *et al.* The potential to use PspA and other pneumococcal proteins to elicit  
533 protection against pneumococcal infection. *Vaccine* **18**, 1707–1711 (2000).
- 534 23. Maestro, B. & Sanz, J. M. Choline Binding Proteins from *Streptococcus pneumoniae* :  
535 A Dual Role as Enzybiotics and Targets for the Design of New Antimicrobials. *antibiotics*  
536 (2016) doi:10.3390/antibiotics5020021.
- 537 24. Cheng, J., Goldstein, R., Gershenson, A., Stec, B. & Roberts, M. F. The cation- $\pi$  box is  
538 a specific phosphatidylcholine membrane targeting motif. *J. Biol. Chem.* **288**, 14863–  
539 14873 (2013).
- 540 25. Galán-Bartual, S., Pérez-Dorado, I., García, P. & Hermoso, J. A. Structure and Function  
541 of Choline-Binding Proteins. *Streptococcus Pneumoniae Mol. Mech. Host-Pathogen*  
542 *Interact.* 207–230 (2015) doi:10.1016/B978-0-12-410530-0.00011-9.
- 543 26. Brooks-Walter, A., Briles, D. E. & Hollingshead, S. K. The *pspC* gene of *Streptococcus*  
544 *pneumoniae* encodes a polymorphic protein, PspC, which elicits cross-reactive  
545 antibodies to PspA and provides immunity to pneumococcal bacteremia. *Infect. Immun.*  
546 **67**, 6533–6542 (1999).
- 547 27. Galán-Bartual, S., Pérez-Dorado, I., García, P. & Hermoso, J. A. Chapter 11 - Structure  
548 and Function of Choline-Binding Proteins. in (eds. Brown, J., Hammerschmidt, S. &  
549 Orihuela, C. B. T.-S. P.) 207–230 (Academic Press, 2015).  
550 doi:https://doi.org/10.1016/B978-0-12-410530-0.00011-9.
- 551 28. Yun, K. W., Choi, E. H. & Lee, H. J. Genetic diversity of pneumococcal surface protein  
552 A in invasive pneumococcal isolates from Korean children, 1991-2016. *PLoS One* **12**,  
553 1–13 (2017).

- 554 29. Masson, P. L., Heremans, J. F. & Dive, C. H. An iron-binding protein common to many  
 555 external secretions. *Clin. Chim. Acta* **14**, 735–739 (1966).
- 556 30. Yen, C.-C. *et al.* Lactoferrin: an iron-binding antimicrobial protein against *Escherichia*  
 557 *coli* infection. *Biometals an Int. J. role Met. ions Biol. Biochem. Med.* **24**, 585–594  
 558 (2011).
- 559 31. Park, S.-S. *et al.* *Streptococcus pneumoniae* Binds to Host Lactate Dehydrogenase via  
 560 PspA and PspC To Enhance Virulence. *MBio* **12**, (2021).
- 561 32. Alberts B, Johnson A, Lewis J, *et al.* *Molecular biology of the cell.* (Garland Science,  
 562 2002).
- 563 33. Nikolich-Žugich, J., Slifka, M. K. & Messaoudi, I. The many important facets of T-cell  
 564 repertoire diversity. *Nat. Rev. Immunol.* **4**, 123–132 (2004).
- 565 34. Frias, M. J., Melo-Cristino, J. & Ramirez, M. Preparation of Pneumococcal Proteins for  
 566 Western Blot Analysis. *Bio-protocol* **3**, e807 (2013).

567

568 **Figure legends**

569 **Figure 1: Diversity among 48 PspA cluster variants.**

570 **A:** Full-length PspA protein sequences from 48 different *S. pneumoniae* isolates randomly  
 571 selected from the Gene Bank were compared for their homology. Homology is represented by  
 572 the heat map with three different groups were identified.

573 **B:** Full-length PspA proteins (n=47) when evaluated in a phylogenetic analysis separated into  
 574 three families: Family I included 21 proteins together with PspA<sub>D39</sub>, a standard pathologic  
 575 reference strain. Family II included 11 proteins and strain PspA<sub>G54</sub> and family III included 16  
 576 proteins and PspA<sub>TIGR4</sub>.

577 **Figure 2: Domain structure of the PspA family proteins and clade separation.**

578 **A:** A structure analysis of selected PspA variants, i.e. PspA<sub>D39</sub>, PspA<sub>G54</sub> PspA<sub>TIGR4</sub>, each  
 579 representing one of the three families shows related composition. The three proteins have long,  
 580 mostly  $\alpha$ -helical structured N-terminal regions (red), followed by shorter ca 80 aa long mostly  
 581 coiled-coil structured regions (gray) and a  $\beta$ -sheet region at the C-terminal end (blue).

607 terminal regions (red), followed by shorter ca 80 aa long mostly coiled-coil  
 608 structured regions (gray) and a  $\beta$ -sheet region at the C-terminal end (blue).

609 The numbers below show the start residue of the corresponding region for each  
 610 protein. The signal peptide which is cleaved upon processing is not included.

611 The  $\alpha$ -helical region corresponds with the N-terminal region and the C-terminal  
 612 region includes both the coiled-coil structured and the  $\beta$ -sheet region.

613 **B: Amino acid composition of three structural regions of the three PspA**  
 614 **variants.** PspA<sub>D39</sub>, PspA<sub>G54</sub>, PspA<sub>TIGR4</sub> were analyzed as representative for  
 615 each family. The  $\alpha$ -helical, coiled-coil and  $\beta$ -sheet regions of the PspA variants  
 616 have different amino acids composition. However, the three PspA variants have  
 617 rather similar amino acid composition within each region. Among three regions,  
 618 from N- to C-terminal (from  $\alpha$ -helical to  $\beta$ -sheet region), the percentage of  
 619 charged amino acids decreased from 40 % to about around 10 %. The  
 620 percentage of polar/amphipathic increased from 15  
 621 % to 30%. Interestingly, Trp only exists in  $\beta$ -sheet region with a portion of 9.5%.

622 **Figure 3: Homology and relatedness of N and C-terminal regions**

623 **A:** Homology tree of N-terminal domains. Distance scale: 0.2

624 **B:** Homology tree of C-terminal domains. Distance scale: 0.05

625 **C:** Percentage identity of N- and C-terminal regions among 48 PspAs is shown  
 626 by heatmap. Sequence comparison among N-terminal is in the upper-right  
 627 panel, and sequence comparison among C-terminal is in the lower-left panel.  
 628 The value ranges from 0.2 to 1 (red- blue).

629 **Figure 4: Domain length variation of selected PspA clade representatives.**

630 Family I variants (green), family II (purple) and family III (orange). Major  
 631 differences exist for VD with two subsets and FDD with three.

632 The proposed structure of each domain is presented by the color of the column.  
 633 N-terminal  $\alpha$ -helical VD-1, VD-2, the FDDs and Domain 4 are shown in red, the  
 634 coiled-coil structured PRD domain by gray, and the  $\beta$ -sheet CBD region by blue  
 635 columns.

636 **Figure 5: Domain Structure of PspA family and of clade variants show**  
 637 **different structures.**

638 **A:** PspA proteins have a domain structure and the proteins are composed of  
 639 six domains. PspA<sub>GA47976</sub>, PspA<sub>D39</sub>, PspA<sub>SPNA45</sub>, PspA<sub>G54</sub>, PspA<sub>2070531</sub>, PspA<sub>TIGR4</sub>  
 640 and PspA<sub>13430</sub> are chosen as representative from each domain variants group.  
 641 Numbers show the start of each domain for each protein.

642 **Family I** proteins include a N-terminal hypervariable domain 1, a variable  
 643 domain, the family determining domain FDD<sub>I</sub> which includes a lactoferrin-  
 644 binding region and domain 4. They are followed by a PRD which varies and  
 645 allows a clade definition and a CBD. PRD has three types: clade IA uses a  
 646 PRD-1<sub>QQ-QQQ</sub>, clade IB a variant with two internal segments and is therefore  
 647 termed PRD-1<sub>QQ-QQQ::QQ-QQQ</sub> and clade IC uses a multi-modular PRD-2 with a  
 648 common core PAPAP motif . The modules of the PRD domain and CBD are  
 649 shows by the patterns.

650 **Family II** proteins include a related domain structure. These proteins have a  
 651 unique FDD<sub>II</sub> domains and all variants use PRD-1<sub>QQ-QQQ</sub>.

652 **Also, Family III** proteins use a similar domain structure. The FDD<sub>III</sub> with either  
 653 two or three segments and a PRD-2 separates between clade IIIA and clade  
 654 IIIB. Clade IIIC has a three segmented FDD<sub>III</sub> combined with a PRD-1<sub>QQ-QQQ</sub>.

655 **B:** 48 PspA are cladded into the seven domain variants. 9 PspA in HUS are  
 656 cladded into the three out of seven domain variants.

657 **Figure 6: Expression of PspA and Lactoferrin binding of D39, G54, TIGR4,**  
 658 **HUS1 and HUS2.**

659 **A:** PspA proteins of D39, G54 and TIGR4 were assayed by western blot.

660 **B:** D39, G54, TIGR4, HUSA, and HUSB binding lactoferrin were tested by  
 661 ELISA, BSA as a control. Three independent experiments were performed. \*\*P  
 662  $\leq 0.001$ .

663

**Figure 1**



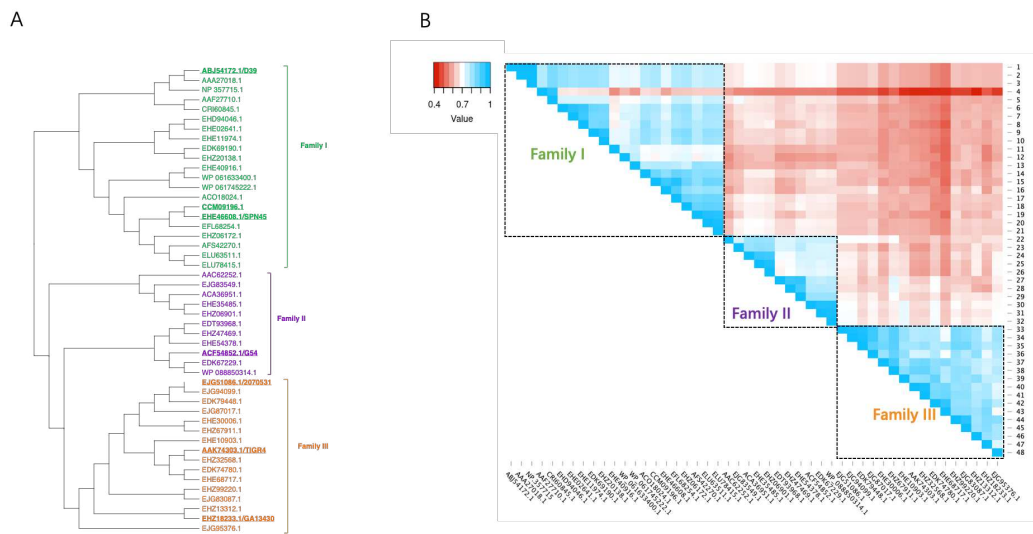


Figure 2

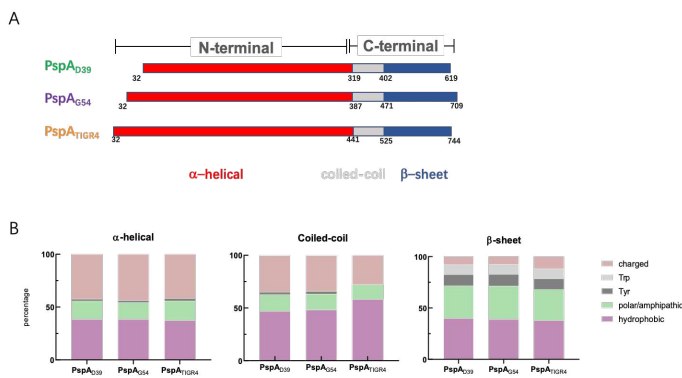


Figure 3

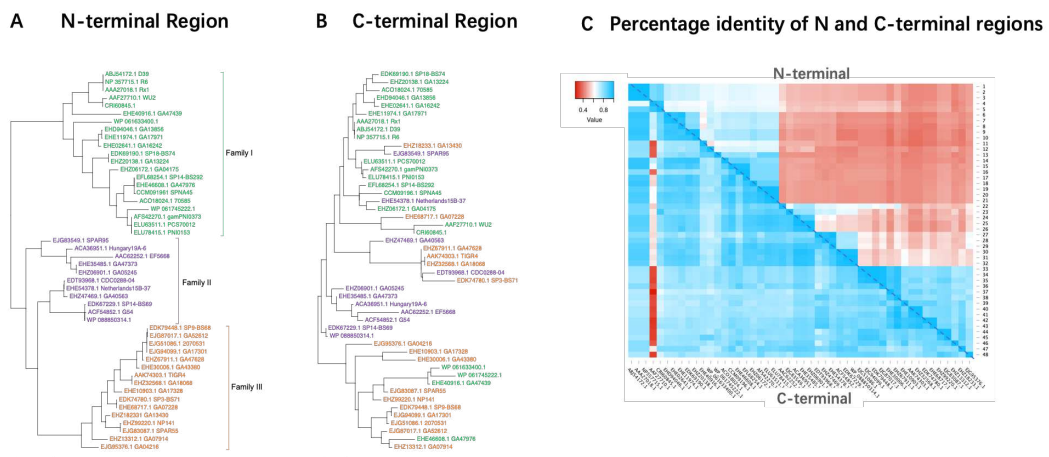


Figure 4

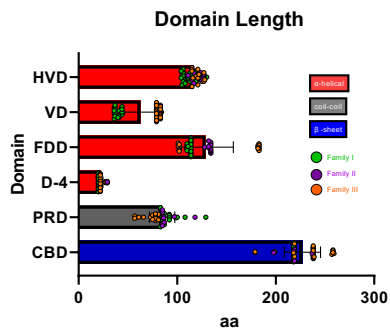


Figure 5

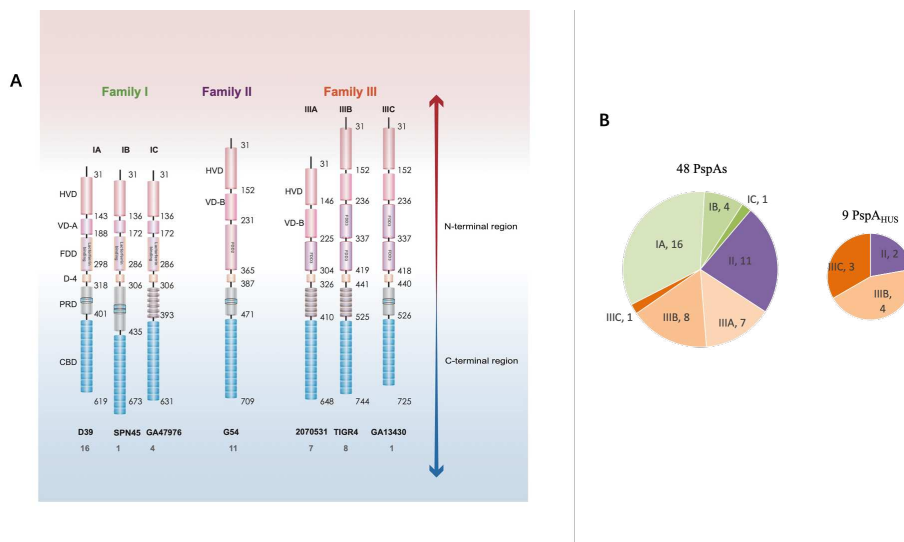
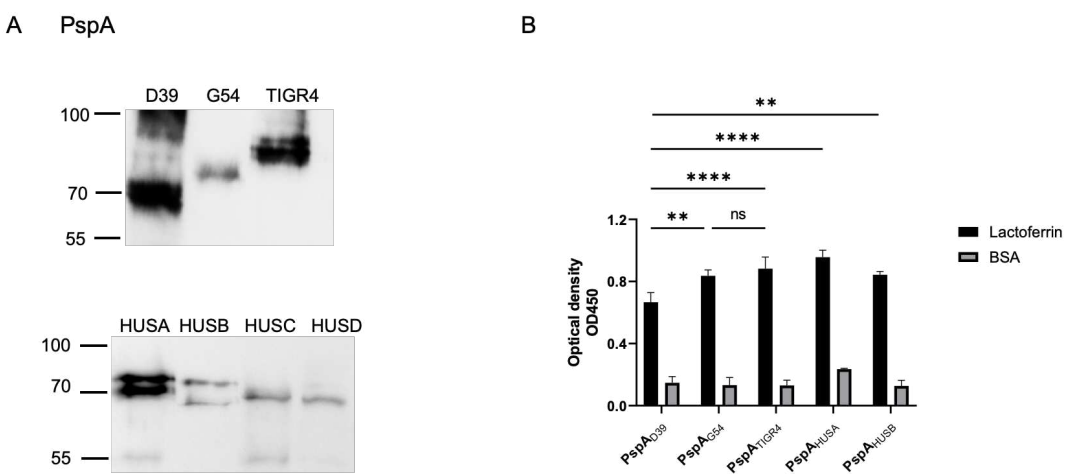




Figure 6



**Table I: PspA variants selected for analyses.**

	Entry ID	Stain	Length	
Family I	<b>ABJ54172.1</b>	<b>D39</b>	619	
	AAA27018.1	Rx1	619	
	NP 357715.	R6	619	
	AAF27710.1	WU2	415	
	CR60845.1		635	
	EHD94046.1	GA13856	635	
	EHE02641.1	GA16242	635	
	EHE11974.1	GA17971	650	
	EDK69190.1	SP18-B574	635	
	EHZ20138.1	GA13224	635	
	EHE40916.1	GA47439	624	
	WP_061633400.1		658	
	WP_061745222.1		610	
	ACC018024.1	70585	638	
	<b>CCM091916.1</b>	<b>SPN445</b>	673	
	<b>EHE46608.1</b>	<b>GA47976</b>	631	
	EFL68254.1	SP14-BS292	609	
	EHZ06172.1	GA04175	603	
	AFS42270.1	gamPNI0373	651	
	ELU63511.1	PC570012	611	
	ELU78415.1	PNI0153	611	
	Family II	AAC02252.1	EF5668	653
		EJG83549.1	SPAR95	706
		ACA36951.1	Hungary19A-6	705
		EHE35485.1	GA47373	721
EHZ06901.1		GA05245	721	
EDT93968.1		CCD0288-04	731	
EHZ47469.1		GA40563	698	
EHE54378.1		Netherlands15B-37	697	
<b>ACF54852.1</b>		<b>G54</b>	709	
EDK67229.1		SP14-B569	722	
WP_088850314.1			711	
Family III		<b>EJG51086.1</b>	<b>2070531</b>	665
		EJG94099.1	GA17301	665
		EDK79448.1	SP9-B568	746
		EJG87017.1	GA52612	688
	EHE30006.1	GA43380	756	
	EHZ67911.1	GA47628	643	
	EHE10903.1	GA17328	678	
	<b>AAK74303.1</b>	<b>TIGR4</b>	744	
	EHZ32568.1	GA18068	735	
	EDK74780.1	SP3-B571	596	
	EHE68717.1	GA07228	677	
	EHZ9220.1	NP141	776	
	EJG83087.1	SPAR55	736	
	EHZ13312.1	GA07914	781	
	<b>EHZ18233.1</b>	<b>GA13430</b>	725	
	EJG95376.1	GA04216	674	

**Table II: Domains used by PspA**

Position	Structure	Domain	Subdomain	Number	Unique for PspA	Match to other proteins	Identity [%]
N-terminal		HVD	HVD-A	21	yes	Unique -no homolog no homolog	
			HVD-B	27	yes		
		VD	VD-A	21	yes	no homolog	
			VD-B	27	yes	no homolog	
		FDD	Lactoferrin	21	no	PspA_D39/ PspC1.1_SRF10 ( <i>S. pneumoniae</i> )	83.3
			FDD II	11	yes	no homolog	
		KADE	16	yes	PspA_2070531 / TATA element modulatory factor 1 DNA binding family protein GA43264 ( <i>S. pneumoniae</i> )??	100	
		D4	48	yes			
C-terminal		Coiled coil	PRD	19	no	PspA_D39 / PspC_R6 ( <i>S. pneumoniae</i> )	75.0
			PRD-II (QQQ_form)	29	no	PspA_GA47976/YALI0_F19030g ( <i>Candida lipolytica</i> )	78.8
		CBD	47	no	PspA_D39 / PspC_R6 ( <i>S. pneumoniae</i> )	91.3	

**Table III: Primers for sequencing**

Primers used for PspA sequencing	
Primer	Sequence
PspA-1F	5'-AAAGATTGTCAGGCTTA-3'
PspA-1R	5'-AAAATGTCAAATGTTCTTAACATGC-3'
PspALSM12F	5'-CCGGATCCAGCGTCGCTATCTTAGGGGCTGGTT-3'
PspASKH63R	5'-TTTCTGGCTCATACTGCTCTT-3'
SKH52F	5'-TCTACTTCTTCTTGAGGTGGGGT-3'
SKH52R	5'-TGGGGTGGAGTTTGTCTTCATCT-3'
GAMKASQ-R	5'-CAGAGCAACCAGCTCCAGTC-3'
SQWFKVS-R	5'-CTCCTGCACAAAACCAGAGC-3'
PspA-2R	5'-TTGAGGGTGTGTGCTTC-3'
PspA-3R	5'-ATCACATCGAGCCCTGCTC-3'

671 **Supplementary material**

672 **Supplementary Figure 1: The percentage of conserved amino acid in each**  
673 **domain.**

674 Column color is shown based on the predicted structure.  $\alpha$ -helical in red, the  
675 coiled-coil in gray, and the  $\beta$ -sheet in blue.

676 **Supplementary Figure 2: Domain structure of selected PspA variants from**  
677 **family.**

678 **A:** Representative PspA variant (D39 from family I, G54 from family II and  
679 TIGR4 from family III) secondary structure prediction are shown by  
680 percentage histogram.  $\alpha$ -helical in red, the coiled-coil in gray, and the  $\beta$ -sheet  
681 in blue.

682 **Supplementary Figure 3: N-terminal domain selects between the three**  
683 **PspA families.**

684 Phylogenetic analysis of three N-terminal domains. A: HVD homology tree,  
685 distance scale: 0.2; B: VD homology tree, distance scale: 1; C: FDD homology  
686 tree, distance scale: 0.2.

687 **Supplementary Figure 4: C-terminal domain selects between the three**  
688 **PspA families.**

689 Phylogenetic analysis of two C-terminal domains. A: PRD homology tree,  
690 distance scale: 0.05; B: CBD homology tree, distance scale: 0.01.

691 **Supplementary Figure 5: Domain-based amino acid composition of**  
692 **PspA representatives.**

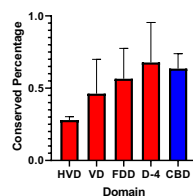
693 **Supplementary Figure 6: 2D structure of selected PspA variants**

694

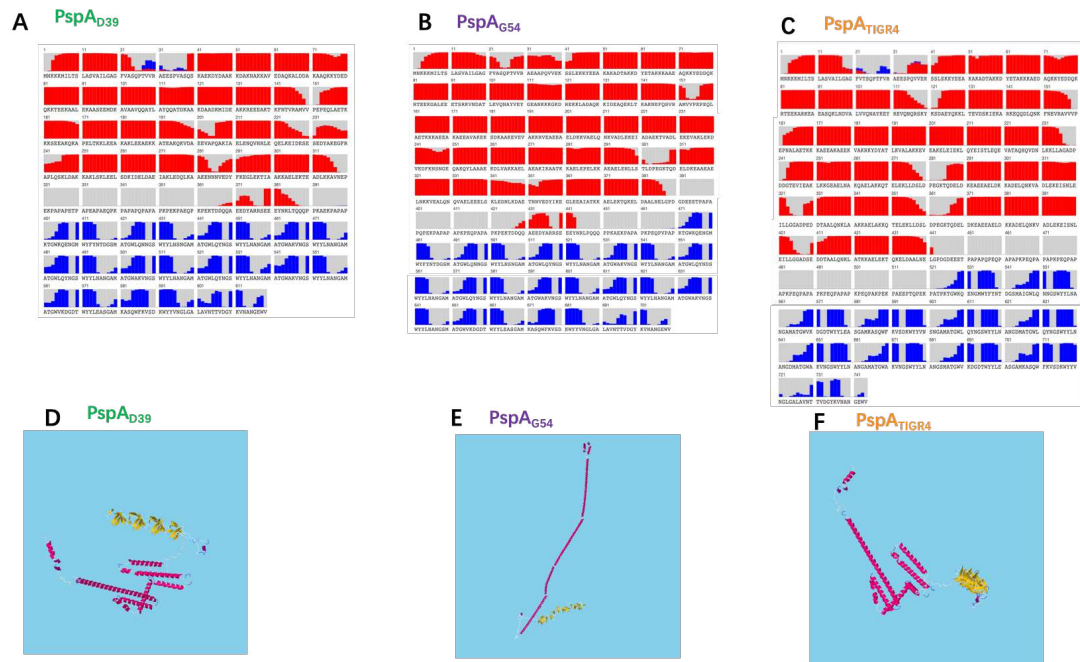
695

696

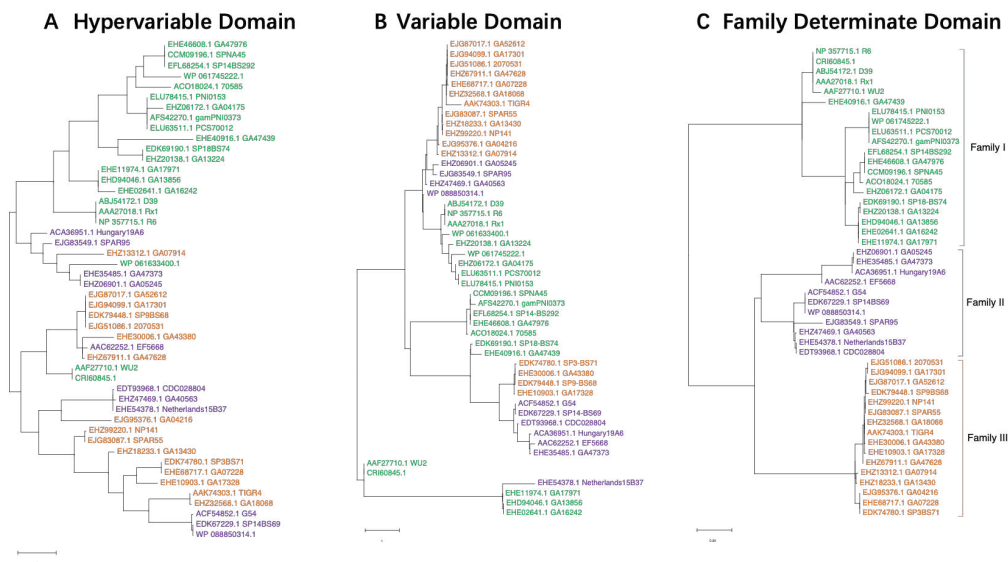
**Supplementary Figure 1**



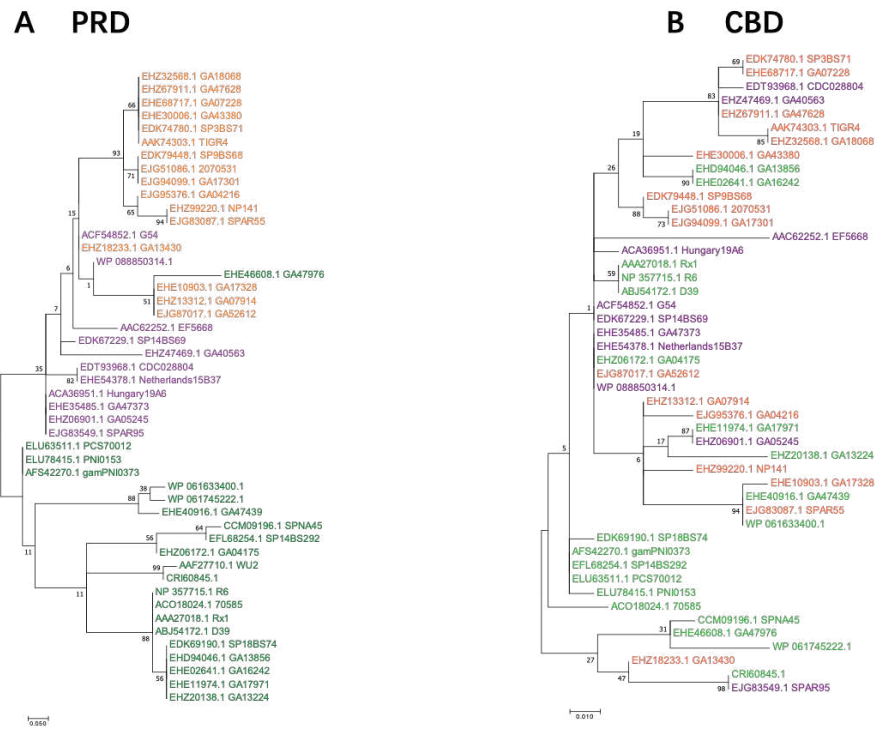
Supplementary Figure 2



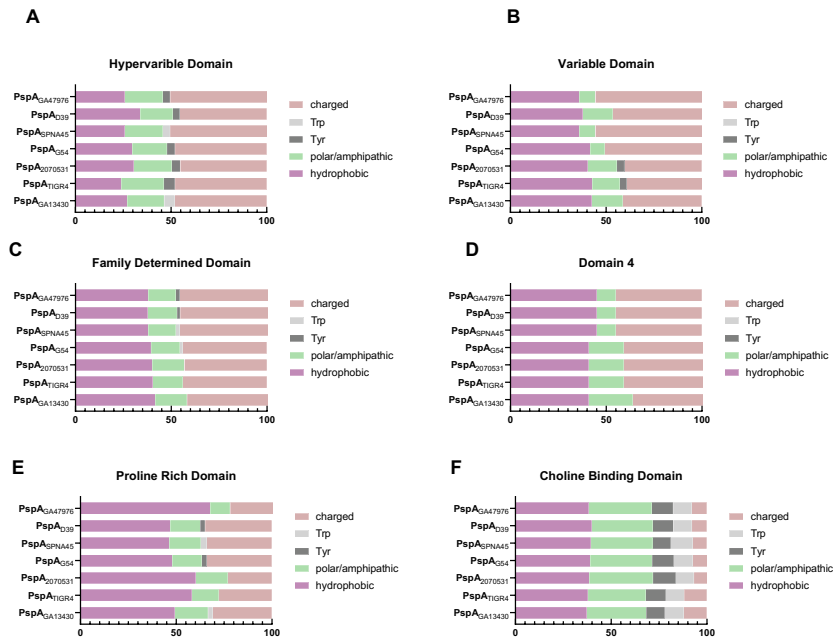
Supplementary Figure 3



Supplementary Figure 4

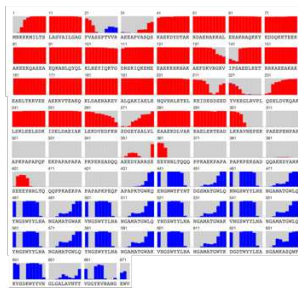


Supplementary Figure 5

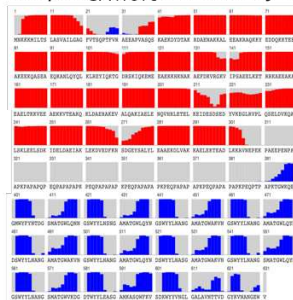


Supplementary Figure 6

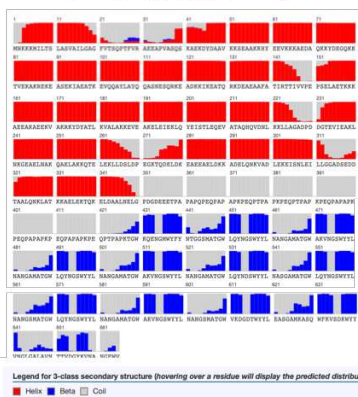
A: PspA<sub>SPN45</sub> Subfamily IB



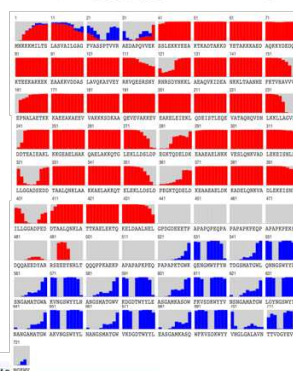
B: PspA<sub>GA47976</sub> Subfamily IC



C: PspA<sub>2070531</sub> family IIIA



D: PspA<sub>GA13430</sub> Subfamily IIIC



### 3 The choline-binding proteins PspA, PspC and LytA of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage



1 **The choline-binding proteins PspA, PspC and LytA of *Streptococcus***  
2 ***pneumoniae* and their role on host cellular adhesion and damage**

3

4 **Cláudia Vilhena<sup>1</sup>, Shanshan Du<sup>1</sup>, Miriana Battista<sup>1</sup>, Martin Westermann<sup>2</sup>, Thomas Kohler<sup>3</sup>,**  
5 **Sven Hammerschmidt<sup>3</sup>, Peter F. Zipfel<sup>1,4\*</sup>**

6

7 <sup>1</sup>Department of Infection Biology, Leibniz Institute for Natural Product Research and Infection  
8 Biology, Jena, Germany

9 <sup>2</sup>Centre for Electron Microscopy, Jena University Hospital, Friedrich-Schiller-University of Jena, Jena,  
10 Germany

11 <sup>3</sup>Department of Molecular Genetics and Infection Biology, Interfaculty Institute for Genetics and  
12 Functional Genomics, Center for Functional Genomics of Microbes, University of Greifswald,  
13 Greifswald, Germany

14 <sup>4</sup>Institute of Microbiology, Friedrich-Schiller-University, Jena, Germany

15

16 **\*Corresponding author:**

17 Prof. Dr. Peter F. Zipfel

18 Department of Infection Biology

19 Leibniz Institute for Natural Product Research and Infection Biology

20 Hans Knöll Institute

21 Beutenbergstrasse 11a

22 07745 Jena

23 Germany

24 E-mail address: peter.zipfel@hki-jena.de

25 **Keywords: biofilms, hemolysis, endothelial, cell wall, immune evasion, hemolytic uremic**  
26 **syndrome**

27





28 **Contribution to the Field Statement**

29 The Gram-positive bacterium *Streptococcus pneumoniae* is a major pathogen that causes acute  
30 respiratory tract infections, which are a major cause of death in children under the age of 5 years.  
31 Socioeconomic conditions in developing countries predispose the population to potentially fatal  
32 bacterial, in particular pneumococcal, infections. Although pneumococcal vaccines are available,  
33 developing countries lack coherent immunization strategies. Moreover, currently available vaccines  
34 are based either on polysaccharides (which provide partial protection against a limited number of  
35 serotypes) or conjugated peptide-based vaccines (which are often expensive). Thus, it is important to  
36 identify new pneumococcal virulence determinants that may qualify as novel vaccine or therapeutic  
37 targets, and also to understand the immune escape strategies used by these human pathogenic bacteria.  
38 The mechanisms by which *S. pneumoniae* (and a growing number of other pathogenic microbes) evade  
39 immune (particularly complement) responses are emerging. Surface-exposed pneumococcal proteins,  
40 such as the family of choline-binding proteins comprising PspA, PspC and LytA, are perfect candidate  
41 targets for vaccination due to their localization at the bacterial surface, which allows interaction with  
42 and binding of human plasma proteins and complement regulators.



43 **Abstract**

44 *Streptococcus pneumoniae* is a Gram-positive opportunistic pathogen that can colonize the upper  
45 respiratory tract. It is a leading cause of a wide range of infectious diseases, including community-  
46 acquired pneumonia, meningitis, otitis media and bacteraemia. Pneumococcal infections cause 1–2  
47 million deaths per year, most of which occur in developing countries, where this bacterial species is  
48 probably the most important pathogen during early infancy. Here, we focused on choline-binding  
49 proteins (CBPs), i.e., PspC, PspA and LytA, and their integration into and interaction with the cell wall  
50 of *S. pneumoniae*. The three pneumococcal proteins have different surface-exposed regions but share  
51 related choline-binding anchors. These surface-exposed pneumococcal proteins are in direct contact  
52 with host cells and have diverse functions. PspC and PspA bind several host plasma proteins, whereas  
53 LytA plays a role in cell division and the lytic phase. We explored the role of the three CBPs on  
54 adhesion and pathogenicity in a human host by performing relevant imaging and functional analyses,  
55 such as electron microscopy, confocal laser scanning microscopy and functional quantitative assays  
56 targeting biofilm formation and the haemolytic capacity of *S. pneumoniae*. *In vitro* biofilm formation  
57 assays and electron microscopy experiments were used to examine the ability of knockout mutant  
58 strains lacking the *lytA*, *pspC* or *pspA* genes to adhere to surfaces. The mutant strains were compared  
59 with the *S. pneumoniae* D39 reference strain. We found that LytA plays an important role in robust  
60 synthesis of the biofilm matrix. PspA and PspC appeared crucial for the haemolytic effects of *S.*  
61 *pneumoniae* on human red blood cells. Furthermore, all knockout mutants caused less damage to  
62 endothelial cells than wild-type bacteria, highlighting the significance of CPBs for the overall  
63 pathogenicity of *S. pneumoniae*. Hence, in addition to their structural function within the cell wall of  
64 *S. pneumoniae*, each of these three surface-exposed CBPs controls or mediates multiple steps during  
65 bacterial pathogenesis.



66 **Introduction**

67 *Streptococcus pneumoniae* is a human pathogen that colonizes the upper respiratory tract and can  
68 cause otitis media, bronchitis, sinusitis, community-acquired pneumonia and sepsis (Kadioglu et al.,  
69 2008; Nobbs et al., 2009). The latest data from the World Health Organization show that pneumonia  
70 kills yearly more than 800,000 children under the age of 5 years (data from 2017), accounting for 15%  
71 of all child deaths at this age. Current *S. pneumoniae* vaccines target mostly the pneumococcal  
72 polysaccharide capsule, which acts as a physical barrier to the outside and protects the bacterium from  
73 recognition by the host immune system (Geno et al., 2015). Below the capsule, the pathogen has a thick  
74 peptidoglycan and teichoic acid-based cell wall, beneath which lies a phospholipid membrane (Pérez-  
75 Dorado et al., 2012). Upon contact with human immune cells, pneumococci shed the capsule and  
76 expose the cell wall to the outside environment (Hammerschmidt et al., 2005; Kietzman et al., 2016).

77 *S. pneumoniae* display several virulence proteins that are integrated into the cell wall; some of  
78 these extend to the outside (Hoskins et al., 2001). In addition to immune escape, several proteins have  
79 structural functions, whereas others mediate carbohydrate and sugar metabolism, or control the cell  
80 division machinery (Pérez-Dorado et al., 2012; Gisch et al., 2013; Heß et al., 2017; Waldow et al.,  
81 2018). One family of these surface proteins shares related anchor domains that bind non-covalently to  
82 phosphoryl choline moieties, which are common constituents of the peptidoglycan layer and are  
83 therefore called choline-binding proteins (CBPs). CBPs are key players in immune evasion, virulence  
84 (Dockrell and Brown, 2015) and host pathogenicity (Roig-Molina et al., 2020; Shizukuishi et al., 2020;  
85 Park et al., 2021). Their pathogenic roles range from proteolytic activities, binding of human plasma  
86 regulators, complement proteins and immune components, and cell-mediated contact to inactivate  
87 human immunoglobulins prior to nasopharyngeal colonization (Agarwal et al., 2010; Galán-Bartual et  
88 al., 2015; Gutiérrez-Fernández et al., 2016; Saumyaa et al., 2016).

**Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage**

89 Two CBPs, PspA and PspC (also termed CbpA), are surface-exposed immune evasion proteins,  
90 which bind human plasminogen, Factor H, secretory IgA, vitronectin, thrombospondin, laminin, C3,  
91 etc. (Jarva et al., 2002; Zipfel et al., 2008; Orihuela et al., 2009; Voss et al., 2013; Binsker et al., 2015;  
92 Meinel et al., 2018; Haleem et al., 2019).

93 PspA and PspC share structural similarities: each have modular, unique and variable N-terminal  
94 regions, a proline rich domain that varies in size and almost identical C-terminal regions (Yother and  
95 Briles, 1992; Du et al., 2020). The role of both proteins in immune evasion has been attributed to their  
96 domain organization, and to sites that bind human plasma proteins. Another CBP is LytA, an intensely  
97 studied pneumococcal autolysin (also called *N*-acetylmuramoyl-l-alanine amidase) (Frolet et al., 2010)  
98 that belongs to a widely distributed group of cell wall-degrading enzymes responsible for  
99 peptidoglycan cleavage; as such, LytA plays a crucial role in cell division (Eldholm et al., 2009).

100 Pneumococcal bloodstream infections start with colonization of the upper respiratory tract,  
101 crossing of the epithelial barrier and culminates with contact to human plasma, where the pathogen is  
102 immediately confronted and attacked by the complement system (Zipfel et al., 2007; Ramos-Sevillano  
103 et al., 2011). During nasopharyngeal colonization and recurrent otitis media in children, *S. pneumoniae*  
104 forms heterogeneous microbial communities embedded in a self-producing polysaccharide matrix  
105 called a biofilm (Hall-Stoodley et al., 2004; Ackermann, 2015; Chao et al., 2019). Pneumococci,  
106 similar to many other pathogenic bacteria, co-exist and colonize the host in their multicellular form  
107 rather than in their planktonic form (Reid et al., 2009; Marks et al., 2012; Shak et al., 2013). Biofilms  
108 are three-dimensional structures formed by agglomerates of bacteria embedded in a self-forming  
109 polysaccharide matrix (Flemming et al., 2016). This complex structure protects the bacteria against  
110 any external dangers, such as host toxins and antibiotics. Members of the *Streptococcus* genus do form  
111 biofilms (Moscoso et al., 2006). However, formation of these multicellular three-dimensional  
112 structures by *S. pneumoniae* has not been studied widely (Cvitkovitch et al., 2003; Donlan et al., 2004).

**Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage**

113 Even though a lot of effort has been put into understanding pneumococcal biofilm formation and its  
114 role in pathogenesis, some questions remain unanswered.  
115 Therefore, the aim of this study was to increase our knowledge of the role of PspA, PspC and LytA in  
116 adhesion and pathogenicity of *S. pneumoniae*. To this end, we generated bacterial mutants each lacking  
117 the gene encoding PspA, PspC or LytA, and compared them with the parental pathogenic reference  
118 strain *S. pneumoniae* D39 (wild-type, WT). By combining state of the art single-cell microscopy,  
119 electron microscopy and infection assays, we systematically examined the relevance of each CBP to  
120 bacterial adhesion, haemolysis and cytotoxicity.

In review



121 **Material and Methods**

122 **Bacterial strains, media, and growth conditions**

123 The pathogenic strain *Streptococcus pneumoniae* D39 was used as reference strain (Lanie et al., 2007).  
124 The D39-derived mutants used comprised:  $\Delta pspA$ ,  $\Delta pspC$  and  $\Delta cps\Delta lytA$ . All *S. pneumoniae* strains  
125 were grown in liquid Todd-Hewitt broth (Roth<sup>®</sup>) supplemented with yeast extract (THY) at 37°C with  
126 5% CO<sub>2</sub>. Blood agar plates were prepared from Blood agar (VWR<sup>®</sup>) with addition of 5% defibrinated  
127 sheep blood (Thermo Scientific<sup>®</sup>). When required, 5 µg/mL erythromycin or 100 µg/mL kanamycin  
128 was used for selection. Growth was monitored by measuring the optical density at 600 nm (OD<sub>600</sub>).

129 **Construction of *S. pneumoniae* D39-derived mutants**

130 The *pspA*, *pspC* and the *lytA* deletion mutants were generated in the genetic background of the  
131 nonencapsulated strain D39 $\Delta cps$  (Pearce et al., 2002; Rennemeier et al., 2007). The construction of the  
132 *pspC* mutant was described earlier (Voss et al., 2013). For the *lytA* mutant, the gene region of *lytA* from  
133 a *S. pneumoniae* D39 $\Delta lytA$  insertion deletion mutant was amplified from genomic DNA (kind gift of  
134 R. Brückner, Kaiserslautern) using primer LytA\_KO\_f (5' GGTGTTATCCTTTGTGAACCTC 3')  
135 and LytA\_KO\_r (5' GCAATCATGCTTTGATTCAAAA 3'). The resulting, 1973 bp fragment contains  
136 498 bp upstream of *lytA*, 38 bp from the beginning of the *lytA* gene, followed by *ermR*, 42 bp from the  
137 end of the *lytA* gene and 486 bp downstream of *lytA*. The amplified PCR fragment was used to  
138 transform D39 $\Delta cps$  using routine protocols (Rennemeier et al., 2007). Resulting colonies were selected  
139 on LB agar plates containing kanamycin and erythromycin and verified by PCR and agarose gel  
140 analysis. The *pspA*-mutant was described earlier (Voß et al., 2018). Briefly, plasmid pQSH29 containing  
141 the full-length *pspA* gene of strain ATCC 11733 (serotype 2) amplified by PCR using the primer  
142 combination SH20 (5' GCGCGCGCGGATCCTTGAATAAGAAAAAATGATTTTAACA 3')  
143 and SH21 (5' CTCAGCTAATT AAGCTTGCTTAAACCCATTACCATTTGGC 3') was used to

**Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage**

144 insert the antibiotic gene cassette. The BamHI/HindIII digested PCR product was then ligated with the  
145 similarly digested vector pQE30 (Qiagen, Hilden, Germany), resulting in a 1.860 *pspA* DNA-insert.  
146 The knockout plasmid was constructed by digestion of the cloned *pspA* gene with SacI and blunt  
147 ligation with the PCR amplified erythromycin gene cassette *ermB*, resulting in plasmid pMSH5.1. This  
148 plasmid was used to transform *S. pneumoniae* strains and knockouts are verified by immunoblot  
149 analysis (Voß et al., 2018).

**150 Cell culture and cell harvesting**

151 Human umbilical vein endothelial cells (HUVEC, CRL-1730) and adenocarcinomic human alveolar  
152 basal epithelial cells (A549, ATCC 107) were cultivated in Dulbecco's modified Eagle's medium,  
153 DMEM (BioWhittaker®) supplemented with 10% fetal calf serum (Biochrom®), 6 mmol/L l-glutamine  
154 (BioWhittaker®) and a mixture of penicillin/streptomycin (100U/100 µg/mL, Sigma®) at 37°C in the  
155 presence of 5% CO<sub>2</sub>. The full supplemented DMEM medium will be referred to as growth medium.  
156 Adherent human cells were washed with pre-warmed Dulbecco's phosphate-buffered saline (DPBS)  
157 (BioWhittaker®) and harvested by incubation for 10 minutes at 37°C with PBS containing  
158 trypsin/EDTA (Gibco®). Cell detachment was stopped by adding 10 mL of growth medium. After  
159 centrifugation, the pellet was resuspended in 1 mL growth medium and the cells were counted using  
160 the cell counter CASY (OLS®CASY).

**161 Static biofilm model**

162 Pneumococci biofilms were grown in either THY or DMEM media to mid-logarithmic phase. Bacteria  
163 were washed and resuspended in the corresponding medium at a concentration of  $1 \times 10^6$  cells/mL).  
164 Bacterial suspensions were incubated on sterile, 18 mm round glass 1.5 H coverslips (Roth®) in the  
165 bottom of 24-well polystyrene plates (Thermo Scientific®). Exceptionally, bacterial suspensions were  
166 incubated on 12-well plates containing 12 mm round glass coverslips. The plates were incubated at  
167 37°C with 5% CO<sub>2</sub> for 48 h. The growth medium was changed every 6 h. Bacterial biofilms were either

**Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage**

168 evaluated by Scanning Electron Microscopy (SEM) or visualized by Confocal Laser Scanning  
169 Microscopy (CLSM).

**170 Biofilm quantification**

171 For quantification of biofilm formation, the Microtiter Dish Biofilm Formation Assay was performed  
172 with small changes (O'Toole, 2010). Pneumococci were grown overnight on solid agar blood plates.  
173 Bacteria were resuspended in PBS and diluted in THY or DMEM media to reach OD<sub>600</sub> of 0.1. Bacteria  
174 were statically grown on 96-well plates to obtain biofilms (Thermo Scientific®), at 37°C with 5% CO<sub>2</sub>.  
175 At the indicated time points, the supernatant was transferred to another plate and OD<sub>600</sub> measured as a  
176 read for planktonic growth. To each well of the original plate, 100 µL of a 1% crystal violet solution  
177 was added and the plate was incubated for 30 min at room temperature. After repeated washing with  
178 water, the plate was left to dry for 1h. Ethanol (95% (v/v)) was added to each well and left for 30 min  
179 at room temperature. The released crystal violet was finally transferred to a new 96-well plate and  
180 absorbance at 620 nm measured. Statistical analysis was performed using Prism version 9 for Windows  
181 (GraphPad Software, La Jolla, CA).

**182 Confocal Laser Scanning Microscopy (CLSM)**

183 Bacterial viability within the biofilms was evaluated using the Bacterial Viability Stain kit (Biotium®)  
184 according to the manufacturer's description. Bacteria were treated as described on static biofilm  
185 section. Planktonic bacteria were removed and the remaining biofilm layer was washed with DPBS and  
186 stained with a fluorescence red marker for dead cells, Ethidium Homodimer III (EthD-III) and a green  
187 peptidoglycan dye, wheat germ agglutinin (WGA) conjugated to CF®488A. After staining, biofilms  
188 were washed to remove unbound dyes and mounted using SlowFade Diamond® (Invitrogen®)  
189 mounting oil. Then the coverslip was sealed with nail polish and biofilms were evaluated by confocal  
190 laser scanning microscopy using a LSM 710 fitted with ZEN 2011 software (Zeiss GmbH).

**191 Scanning Electron Microscopy (SEM)**



**Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage**

192 For SEM, biofilms were grown on 12-well plates containing 12 mm coverslips (Roth®), as described  
193 above for the static biofilm model. At designated time points, medium was aspirated. Then cells were  
194 fixed for 1h in 2.5% glutaraldehyde in sodium cacodylate buffer (0.1 M, pH 7.0) and washed three  
195 times with sodium cacodylate buffer for 20 min each. Samples were dehydrated in rising ethanol  
196 concentrations followed by critical point drying, using a Leica EM CPD300 Automated Critical Point  
197 Dryer (Leica) and finally coated with gold (25 nm) in a Safematic CCU-010 HV Sputter Coating  
198 System (Safematic). SEM images were acquired at different magnifications in a Zeiss-LEO 1530  
199 Gemini field-emission scanning electron microscope (Carl Zeiss) at 6-8kV acceleration voltage and a  
200 working distance of 5-7 mm using an InLense secondary electron detector for secondary electron  
201 imaging.

**202 Haemolysis assays**

203 *S. pneumoniae* strains were grown at 37°C with 5% CO<sub>2</sub> until reaching mid-logarithmic phase. Bacteria  
204 were washed and a 100 µL suspension was combined with 100 µL red blood cells (isolated from buffy  
205 coat as previously described (Repnik et al., 2003; Schmidt et al., 2019; Luo et al., 2020) and the mixture  
206 was incubated in a 96-well plate (Thermo Scientific®) at 37°C with slight agitation (300 rpm) for 30  
207 min (positive control was only added 10 min prior to the end of the incubation). PBS was used as  
208 negative control and bi-distilled water as positive control for erythrocyte lysis. Erythrocytes derived  
209 from 9 different volunteers were tested. Then the plates were centrifuged (400g, 15min, 4°C), the  
210 supernatant was transferred to a new 96-well plate (Thermo Scientific®) and hemoglobin release was  
211 quantified at OD<sub>540nm</sub>. Statistical analysis was performed using Prism version 9 for Windows  
212 (GraphPad Software, La Jolla, CA).

**213 Bacterial incubation with human endothelial and epithelial cells**

214 Endothelial HUVEC cells and epithelial A549 cells were seeded on a 18 mm diameter glass coverslips  
215 in a 12-well plate (Thermo Scientific®), at a concentration of 200,000 cells/well and they were grown

**Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage**

216 at 37°C with 5% CO<sub>2</sub> until confluence was reached. The confluent cells were then transferred to  
217 antibiotic-free growth medium and infected with pneumococci using a multiplicity of infection (MOI)  
218 of 50 bacteria per cell. The mixture was incubated at 37°C with 5% CO<sub>2</sub>. After washing with PBS,  
219 human cells and bacteria were fixed with 4% paraformaldehyde for 10 min at 4°C followed by blocking  
220 with 1% BSA for 1h at room temperature. A rabbit anti-*S. pneumoniae* antibody (Abcam®) was added,  
221 in order to visualize attached extracellular bacteria and also proliferating bacteria for 16h at 4°C. For  
222 HUVEC cells, concomitant incubation of a secondary anti-rabbit antibody with 4',6-Diamidino-2-  
223 Phenylindole, dihydrochloride (DAPI, Biotium®) and Platelet endothelial cell adhesion molecule 1  
224 (PECAM-1) conjugated with FITC was carried out for 1h at room temperature. After final washing  
225 steps with PBS, the coverslips were embedded in SlowFade Diamond (Thermo Fisher®), sealed with  
226 nail polish and stored at 4°C for subsequent imaging. Images were taken on a confocal laser scanning  
227 microscope (LSM710, Zeiss®).

**228 Cytotoxicity assay**

229 Cytotoxicity of *S. pneumoniae* D39 or the isogenic mutants towards human epithelial cells was  
230 assessed using a CellTiter-Blue® (CTB) Cell Viability Assay (Promega), according to manufacturer  
231 instructions. Human cells were seeded on a 96-well plate (Thermo Scientific®) with a concentration of  
232 15,000 cells/well. Cells were cultivated at 37°C with 5% CO<sub>2</sub> until confluence was reached. Then  
233 bacteria were added and the mixture was incubated for 1h under the same growth conditions.  
234 Subsequently unbound bacteria were removed by washing with DPBS. The extracellular and human  
235 cell bound pneumococci were killed by treatment of the cells with gentamicin (500 µg/mL) for 1h at  
236 37°C under 5% CO<sub>2</sub>. CTB (100 µl) was added to each well. Following incubation for 16h at 37 °C in  
237 5% CO<sub>2</sub>, the absorbance was measured using a Tecan® Safire 2 microplate reader at an absorption of  
238 570 nm. In this assay, intact metabolically active endothelial cells can convert the redox dye (resazurin)

**Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage**

239 into a fluorescent end product (resorufin). Statistical analysis was performed using Prism version 9 for  
240 Windows (GraphPad Software, La Jolla, CA).

**241 Construction of protein-protein interaction (PPI) network**

242 The Search Tool for Retrieval of Interacting Genes (STRING) (<https://string-db.org>) database, which  
243 integrates both known and predicted PPIs, was applied to predict functional interactions of *S.*  
244 *pneumoniae* proteins (Szklarczyk et al., 2019). First, this interaction tool was used to evaluate the  
245 interaction of LytA, PspA and PspC. Second, active interaction sources, including text mining,  
246 experiments, databases, co-expression, neighborhood, gene fusion and co-occurrence and an  
247 interaction score > 0.4 were applied to construct the PPI networks. STRING is a database of known  
248 and predicted protein-protein interactions. Given a list of the proteins as input, STRING can search for  
249 their neighbor interactors and generate the PPI network consisting of all these proteins and all the  
250 interactions between them. The interactions include direct (physical) and indirect (functional)  
251 associations; they stem from computational prediction, from knowledge transfer between organisms,  
252 and from interactions aggregated from other (primary) databases (von Mering et al., 2005; Szklarczyk  
253 et al., 2021).



254 **Results**

255 **LytA prevents bacterial survival within mature-biofilm structures**

256 To explore whether and how PspA, PspC and LytA contribute to biofilm formation as the first step of  
257 adhesion, *S. pneumoniae* strains lacking the *pspA*, *pspC* or *lytA* genes were grown in polystyrene multi-  
258 well plates, and biofilm formation was quantified. LytA-lacking bacteria produced significantly more  
259 biofilm mass, whereas biofilm formation by the *pspA* and *pspC* mutants was comparable to that by the  
260 reference strain (**Fig. 1A**). To evaluate the structure of the newly formed biofilms, and bacterial  
261 survival within the tri-dimensional structure, we quantified bacterial viability using CLSM. Within the  
262 biofilms, most *lytA*-lacking bacteria were viable, as shown by green fluorescence, whereas most  
263 bacteria derived from the *pspA* and *pspC* knockout strains and WT bacteria were dead, as revealed by  
264 red fluorescence (**Fig. 1B**), which goes in accordance with the absence of lytic phase of the *lytA*-mutant.  
265 Thus, absence of *lytA* initially affects bacterial lysis and then biofilm formation is altered, whereas  
266 absence of either *pspA* or *pspC* does not influence either biofilm formation or bacterial viability.  
267 Next, we used SEM to evaluate matrix formation by the *lytA* mutant and the reference strain D39 in  
268 more detail. The *lytA* mutant produced less extracellular matrix than the parental strain D39 (**Fig. 1C**).  
269 At both low (**Fig. 1C, panel 1 and 2**) and high magnifications (**Fig. 1C, panel 3 and 4**), the WT strain  
270 (but not the *lytA* mutant) generated a prominent granular matrix. Furthermore, the *lytA* mutant  
271 generated filament-like biofilms (**Fig. 1D, panel 1 and 2**) that allowed dense bacterial agglomeration  
272 (**Fig. 1D, panel 3 and 4**). To exclude an effect of growth medium, we compared growth in THY and  
273 DMEM (**Supplementary Fig. S1**). When grown in DMEM, the  $\Delta$ *lytA* strain did not produce  
274 extracellular matrix; the biofilms were less dense; and bacterial cells were rounder, suggesting  
275 deregulated cell division. Taking into consideration the differential morphology and growth profile of  
276  $\Delta$ *lytA* strain, these results suggest an impaired extracellular matrix production on *lytA* deletion  
277 background.

**Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage**

278

**279 PspA, PspC and LytA reduce metabolic activity of epithelial cells**

280 To examine the role of the three bacterial surface proteins in host cell damage, we first asked whether  
281 the mutants affect the metabolism of human alveolar epithelial cells (A549). To this end, A549 cells  
282 were co-cultivated with either knockout or WT D39 bacteria, and cell metabolism was evaluated by  
283 measuring conversion of resazurin to the fluorescent product resorufin, which only occurs in  
284 metabolically active cells. Upon contact with each of the three mutants ( $\Delta$ *pspA*,  $\Delta$ *pspC* and  $\Delta$ *lytA*),  
285 metabolic activity in cells increased, and was higher in cells challenged with the reference strain D39  
286 (Fig. 2A). This shows that the pathogenic reference strain can damage human epithelial cells, and that  
287 deletion of a single CBP gene affects the ability to cause cell damage. Each mutant showed different  
288 effects. The  $\Delta$ *lytA* strain affected cell metabolism more strongly than  $\Delta$ *pspA* or  $\Delta$ *pspC*. Thus, deletion  
289 of one single CBP affects the metabolic turnover of human epithelial cells, thereby confirming that  
290 each CBP contributes to pathogenicity.

291

**292 The three CBPs mediate pneumococcal haemolytic activity**

293 To further define the role of the three CBPs on interactions with host cells, we evaluated the haemolytic  
294 capacity of the knockout mutants. The *S. pneumoniae* mutants were added to human red blood cells  
295 and, following incubation, erythrocyte lysis was evaluated. Erythrocyte lysis induced by each knockout  
296 was less than that induced by the pathogenic reference strain D39 (Fig. 2B). Thus, each CBP  
297 contributes to the ability of the bacteria to induce erythrocyte lysis.

298

**299 The three CBPs induce expression of PECAM-1 by endothelial cells**

300 After addressing adhesion (biofilm formation) to pulmonary epithelial cells and the effects on red blood  
301 cells, we next examined the effects of the bacterial mutants to examine their effects on human

**Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage**

302 endothelial cells (HUVECs). Expression of the human endothelial cell surface marker platelet-  
303 endothelial cell adhesion molecule-1 (PECAM-1) (Iovino et al., 2014) was monitored by CLSM.  
304 HUVEC cells were cultivated with either the CBP mutants or the WT strain. Each knockout mutant  
305 strongly reduced surface expression of PECAM-1 (**Fig. 3A**). Moreover,  $\Delta$ lytA strain sustained bacterial  
306 growth when in contact with human endothelial cells, as seen by the intense red fluorescence signal  
307 (bacteria) in the  $\Delta$ lytA panel; however, it did not induce expression of this surface marker. Quantitative  
308 analysis of the microscopy data showed significant upregulation of PECAM-1 expression only in the  
309 presence of the reference D39 strain (**Fig. 3B**), corroborating the observed absence of cellular damage  
310 upon incubation with any of the CBP mutants.

311 **Table 1** summarizes the effects of each pneumococcal mutant and the reference D39 strain. Apart from  
312 the effect of  $\Delta$ lytA on biofilm formation, all other effects suggest that mutants show impaired  
313 pathogenic capacity, suggesting that the corresponding proteins play a crucial role in evading host  
314 immune responses.

315

**PspA, PspC and LytA are part of a complex network that regulates host–pathogen interactions**

316 So far, the results show that each pneumococcal protein plays multiple roles during host–pathogen  
317 interactions, i.e., biofilm formation, haemolysis, and epithelial and endothelial cell damage. These  
318 effects are distinct and involve diverse pneumococcal cellular subsystems, such as the cell division  
319 machinery (to regulate growth rate), expression of toxins (to induce host cell damage), and  
320 peptidoglycan synthesis (to release haemolytic enzymes). Therefore, we evaluated the connections  
321 between the three proteins within protein networks.  
322

323 We constructed protein-protein interaction (PPI) networks for each pneumococcal protein using the  
324 STRING database. The initial network focused on PspA, PspC and LytA, and showed that LytA  
325 mediates the interaction between PspA and PspC (**Fig. 4A**). An enlarged and more complex network

**Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage**

326 map was then constructed based on several criteria; strong connections are represented by thick, dark-  
327 grey lines whereas weaker connections (with fewer matching criteria) are represented by shaded, light-  
328 grey lines (**Fig. 4B**). The annotated proteins are presented next to the circular intersection node icon.  
329 After PPI construction, a Markov Cluster Algorithm (MCL) was used to identify cluster-specific  
330 groups (Enright et al., 2002). Four protein clusters appeared: a very densely interconnected cluster I  
331 (red), a small yet strongly interconnected cluster II (brown), cluster III comprising connections with  
332 different degrees of strength (green), and cluster IV (yellow) representing fewer interacting partners.  
333 Each cluster contained (predominantly) members related to a specific cellular machinery: cluster I =  
334 cell division; cluster II = chromosome replication; cluster III = peptidoglycan biosynthesis; and cluster  
335 IV = unannotated proteins.  
336 Each protein (PspA, PspC or LytA) is embedded in a complex network of cross-interactions, and each  
337 is integrated at a very prominent position, suggesting that targeting the proteins therapeutically can  
338 destabilize intracellular bacterial homeostasis. Additionally, this prominent positioning contextualizes  
339 and connects all of the data, suggesting that these CBPs have a multifactorial effect on host invasion  
340 and cell damage.



341 **Discussion**

342 *S. pneumoniae* expresses a family of CBPs that use related cell wall anchors and act as central virulence  
343 proteins that interact with the soluble host complement regulators and with human immune system.  
344 Here, we evaluate and compare three important CBPs, i.e., PspA, PspC and LytA, with respect to their  
345 effects on bacterial biofilm formation and host cell damage, and their integration into bacterial protein  
346 networks.

347 Pneumococci lacking LytA form more biofilm extracellular matrix. By contrast, PspA and PspC have  
348 no such effects on biofilm matrix formation. Previous studies show that CBPs play roles in invasion of  
349 the host under various environmental conditions (Moscoso et al., 2006; Domenech et al., 2015). Efforts  
350 have been made to evaluate the role of extracellular DNA-CBPs complexes during establishment of  
351 biofilms; interestingly, LytA-mediated DNA complexes are relevant to biofilm formation (Moscoso et  
352 al., 2006). Here, we show that PspA and PspC mediate biofilm formation, which supports current  
353 knowledge; however, our observations regarding LytA are rather different (**Fig. 1**). We found that  
354 LytA prevents, rather than promotes, biofilm formation, as bacteria lacking the gene encoding autolysin  
355 were more likely to form biofilms. Similar to the results presented here, another study reported an anti-  
356 biofilm effect of LytA (Domenech et al., 2011); however, few studies have examined the specific and  
357 peculiar role of LytA in pneumococcal biofilms. Protein network analysis revealed that LytA is linked  
358 and integrated into the bacterial cell division machinery and teichoic acid biosynthesis (**Fig. 4**).  
359 Triggering of the bacterial SOS response cascade, which is required for establishment of biofilms,  
360 might affect transcription of the *lytA* operon, most likely leading to repression.

361 This study also identifies a novel role for PspA, PspC and LytA in haemolysis. Bacteria lacking *pspA*,  
362 *pspC* or *lytA* are less efficient at inducing lysis of human erythrocytes (**Fig. 2B**). The PPI network map  
363 (**Fig. 4B**) shows that pneumolysin (ply), a lytic pore forming pneumococcal enzyme (Rai et al., 2016),



**Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage**

364 interacts closely with LytA (and even LytC, another pneumococcal CBP). However, PspA and PspC  
365 do not interact with ply. Instead, these two proteins interact with neuraminidase (NanA), a *S.*  
366 *pneumoniae* enzyme that is upregulated upon contact with the host cells; secreted NanA cleaves host  
367 glycoconjugates (Parker et al., 2009). This relationship between PspA, PspC and NanA might connect  
368 haemolysis and the two surface-bound pneumococcal proteins, and provide another link to biofilm  
369 formation. A recent study proposes that pneumococcus haemolysis activity is mediated mainly by  
370 hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) rather than by excreted pneumolysin (Luo et al., 2020; McDevitt et al.,  
371 2020; Ulrych et al., 2021). In terms of the haemolytic profile, the pneumococcal mutants presented in  
372 that study, which do not produce H<sub>2</sub>O<sub>2</sub>, show a surprising resemblance to the *pspA*, *pspC* and *lytA*  
373 mutants in the present study. Moreover, H<sub>2</sub>O<sub>2</sub> produced by *S. pneumoniae* damages human lung cells  
374 (Rai et al., 2015). Therefore, the reduced cytotoxicity of the CBPs mutants toward human epithelial  
375 and endothelial cells (**Fig. 2A and 3**) might be a consequence of decreased H<sub>2</sub>O<sub>2</sub> production.

376 To obtain a deeper understanding of the role of CBPs in epithelial and endothelial damage, a thorough  
377 metabolome analysis should be performed to examine the intracellular metabolic status of the mutants.

378 The association between bacterial cell wall- and/or membrane-anchored proteins and the stress  
379 response has been described before, with extensive studies being carried out in Gram-negative bacteria,  
380 particularly *Escherichia coli* (Hews et al., 2019; Vilhena et al., 2019; Sueki et al., 2020). In Gram-  
381 positive bacteria, the thick peptidoglycan layer is responsible for protection against environmental  
382 cues. Yet, other studies show the dynamic nature and diverse structure of the peptidoglycan layer across  
383 the cell wall of even a single bacterium (Hayhurst et al., 2008; Beeby et al., 2013). This diversity may  
384 lead to distinct distributions of bacterial surface-exposed proteins within a certain period of time, which  
385 might dictate the interaction with the human host. The interchange of information between the different  
386 intracellular metabolic pathways, and the arrangement of the cell wall (mainly the CBPs) is a crucial  
387 immune evasion strategy.

**Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage**

388 A thread connecting all of our experimental findings is the intermingled network of interactions  
389 sustained by the three CBPs (**Fig. 4B**). The PPI network map shows four clusters that relate to different  
390 cellular pathways. Cluster I contains mainly proteins related to the cell division machinery (DivIVA,  
391 FtsZ, FtsL, FtsX and GpsB) (Goehring and Beckwith, 2005); cluster II contains proteins related to  
392 chromosome replication (DnaA, DnaN, ParE and PriA) (Van Raaphorst et al., 2017); and cluster III  
393 contains proteins related to peptidoglycan biosynthesis (Pbp1B, Pbp2A, PenA, MurC and MurD)  
394 (Pinho et al., 2013). A relevant interaction partner for each evaluated protein is the serine/threonine-  
395 protein kinase StkP (Beilharz et al., 2012; Fleurie et al., 2014; Grangeasse, 2016; Ulrych et al., 2021).  
396 StkP plays a crucial role in regulating cell shape and division of *S. pneumoniae* through control of  
397 DivIVA activity. StkP is thought to sense intracellular peptidoglycan subunits in the division septum  
398 of actively growing cells and to adjust the regulation of DivIVA. The bond between StkP and the triad  
399 PspA-PspC-LytA suggests a link between each CBP in the cell wall of pneumococci and intracellular  
400 regulation subjacent to the synthesis of the cell wall itself, as well as coordination of cell division.  
401 Hence, *S. pneumoniae* fine-tunes expression of surface-anchored immune evasion proteins in response  
402 to the cell cycle. Cluster II comprises additional CBPs such as CbpC and CbpF (Maestro and Sanz,  
403 2016), and proteins related to teichoic acid biosynthesis (LicB, LicC and LicD) (Denapate et al., 2012).  
404 The large cluster III also contains regulators with different backgrounds and functions, e.g., Ply  
405 (pneumolysin) (Rai et al., 2016), NanA and NanB (both of which are implicated in biofilm formation  
406 and in colonization of the upper airway) (Parker et al., 2009) and PavA (which is critical for the overall  
407 virulence of pneumococci) (Pracht et al., 2005). Cluster IV harbours transcriptional regulators,  
408 enzymes and other uncharacterized proteins.

409 Understanding the interconnectivity of PspA, PspC and LytA with other subcellular systems and their  
410 impact on bacterial homeostasis, is pivotal to mechanistically comprehend immune evasion by these  
411 proteins and to developing effective therapeutics.



412 **Acknowledgements**

413

414 This work was supported by the Collaborative Research Center, FungiNet (project C6), Deutsche  
415 Forschungsgemeinschaft (DFG). SH received funding from Deutsche Forschungsgemeinschaft (DFG  
416 HA 3125/5-2). PFZ received support from the KIDNEEDS Foundation Iowa City USA:

417

418 **Author Contributions**

419 CV conceived and designed the experiments. CV, SD and MB performed the experiments. CV, SD  
420 and MB analysed the data. MW, SH and PZ contributed with reagents, materials, and analysis tools.  
421 CV and PZ wrote the manuscript and TPK and SH edited the manuscript.

422

423 **Conflicts of interest**

424 The authors declare that the research was conducted in the absence of any commercial or financial  
425 relationships that could be construed as a potential conflict of interest.

426 **References**

- 427 Ackermann, M. (2015). A functional perspective on phenotypic heterogeneity in microorganisms. *Nat. Rev. Microbiol.* 13,  
428 497–508. doi:10.1038/nrmicro3491.
- 429 Agarwal, V., Asmat, T. M., Luo, S., Jensch, I., Zipfel, P. F., and Hammerschmidt, S. (2010). Complement regulator factor  
430 H mediates a two-step uptake of *Streptococcus pneumoniae* by human cells. *J. Biol. Chem.* 285, 23486–23495.  
431 doi:10.1074/jbc.M110.142703.
- 432 Beeby, M., Gumbart, J. C., Roux, B., and Jensen, G. J. (2013). Architecture and assembly of the Gram-positive cell wall.  
433 *Mol. Microbiol.* doi:10.1111/mpi.12203.
- 434 Beilharz, K., Novakova, L., Fadda, D., Branny, P., Massidda, O., and Veening, J.-W. (2012). Control of cell division in  
435 *Streptococcus pneumoniae* by the conserved Ser/Thr protein kinase StkP. *Proc. Natl. Acad. Sci.* 109, E905–E913.  
436 doi:10.1073/pnas.1119172109.
- 437 Binsker, U., Kohler, T. P., Krauel, K., Kohler, S., Schwartz, H., and Hammerschmidt, S. (2015). Pneumococcal Adhesins  
438 PavB and PspC Are Important for the Interplay with Human Thrombospondin-1. *J. Biol. Chem.* 290.  
439 doi:10.1074/jbc.M114.623876.
- 440 Chao, Y., Bergenfelz, C., and Hakansson, A. P. (2019). “Growing and Characterizing Biofilms Formed by *Streptococcus*  
441 *pneumoniae*,” in *Methods in Molecular Biology* doi:10.1007/978-1-4939-9199-0\_13.
- 442 Cvitkovitch, D. G., Li, Y. H., and Ellen, R. P. (2003). Quorum sensing and biofilm formation in Streptococcal infections.  
443 *J. Clin. Invest.* doi:10.1172/JCI200320430.
- 444 Denapate, D., Brückner, R., Hakenbeck, R., and Vollmer, W. (2012). Biosynthesis of teichoic acids in *Streptococcus*  
445 *pneumoniae* and closely related species: Lessons from genomes. *Microb. Drug Resist.* doi:10.1089/mdr.2012.0026.
- 446 Dockrell, D. H., and Brown, J. S. (2015). *Streptococcus pneumoniae* Interactions with Macrophages and Mechanisms of  
447 Immune Evasion. Elsevier Inc. doi:10.1016/B978-0-12-410530-0.00021-1.
- 448 Domenech, M., García, E., and Moscoso, M. (2011). In vitro destruction of *Streptococcus pneumoniae* biofilms with  
449 bacterial and phage peptidoglycan hydrolases. *Antimicrob. Agents Chemother.* 55, 4144–4148.  
450 doi:10.1128/AAC.00492-11.
- 451 Domenech, M., Ruiz, S., Moscoso, M., and García, E. (2015). In vitro biofilm development of *Streptococcus pneumoniae*  
452 and formation of choline-binding protein-DNA complexes. *Environ. Microbiol. Rep.* doi:10.1111/1758-2229.12295.
- 453 Donlan, R. M., Priede, J. A., Heyes, C. D., Sanii, L., Murga, R., Edmonds, P., et al. (2004). Model system for growing and  
454 quantifying *Streptococcus pneumoniae* biofilms in situ and in real time. *Appl. Environ. Microbiol.*  
455 doi:10.1128/AEM.70.8.4980-4988.2004.
- 456 Du, S., Vilhena, C., King, S. J., Sahagún-Ruiz, A., Hammerschmidt, S., Skerka, C., et al. (2020). Molecular Analyses of  
457 *Streptococcus pneumoniae* Immune Evasion Proteins Identifies new Domains and Reveals Structural Differences  
458 between PspC and Hic variants. *Sci. Rep.*
- 459 Eldholm, V., Johnsborg, O., Haugen, K., Ohnstad, H. S., and Havastein, L. S. (2009). Fratricide in *Streptococcus*  
460 *pneumoniae*: Contributions and role of the cell wall hydrolases CbpD, LytA and LytC. *Microbiology* 155, 2223–  
461 2234. doi:10.1099/mic.0.026328-0.
- 462 Enright, A. J., Van Dongen, S., and Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein  
463 families. *Nucleic Acids Res.* doi:10.1093/nar/30.7.1575.
- 464 Flemming, H.-C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S. A., and Kjelleberg, S. (2016). Biofilms: an emergent  
465 form of bacterial life. *Nat. Rev. Microbiol.* 14, 563–575. doi:10.1038/nrmicro.2016.94.

**Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage**

- 466 Fleurie, A., Manuse, S., Zhao, C., Campo, N., Cluzel, C., Lavergne, J. P., et al. (2014). Interplay of the Serine/Threonine-  
467 Kinase StkP and the Paralogs DivIVA and GpsB in Pneumococcal Cell Elongation and Division. *PLoS Genet.* 10.  
468 doi:10.1371/journal.pgen.1004275.
- 469 Frolet, C., Beniazza, M., Roux, L., Gallet, B., Noirclerc-Savoie, M., Vernet, T., et al. (2010). New adhesin functions of  
470 surface-exposed pneumococcal proteins. *BMC Microbiol.* 10. doi:10.1186/1471-2180-10-190.
- 471 Galán-Bartual, S., Pérez-Dorado, I., García, P., and Hermoso, J. A. (2015). Structure and Function of Choline-Binding  
472 Proteins. *Streptococcus Pneumoniae Mol. Mech. Host-Pathogen Interact.*, 207–230. doi:10.1016/B978-0-12-  
473 410530-0.00011-9.
- 474 Geno, K. A., Gilbert, G. L., Song, J. Y., Skovsted, I. C., Klugman, K. P., Jones, C., et al. (2015). Pneumococcal capsules  
475 and their types: Past, present, and future. *Clin. Microbiol. Rev.* doi:10.1128/CMR.00024-15.
- 476 Gisch, N., Kohler, T., Ulmer, A. J., Müthing, J., Pribyl, T., Fischer, K., et al. (2013). Structural Reevaluation of  
477 *Streptococcus pneumoniae* Lipoteichoic Acid and New Insights into Its Immunostimulatory Potency. *J. Biol. Chem.*  
478 288. doi:10.1074/jbc.M112.446963.
- 479 Goehring, N. W., and Beckwith, J. (2005). Diverse paths to midcell: Assembly of the bacterial cell division machinery.  
480 *Curr. Biol.* 15, 514–526. doi:10.1016/j.cub.2005.06.038.
- 481 Grangeasse, C. (2016). Rewiring the Pneumococcal Cell Cycle with Serine/Threonine- and Tyrosine-kinases. *Trends*  
482 *Microbiol.* 24, 713–724. doi:10.1016/j.tim.2016.04.004.
- 483 Gutiérrez-Fernández, J., Saleh, M., Alcorlo, M., Gómez-Mejía, A., Pantoja-Uceda, D., Treviño, M. A., et al. (2016).  
484 Modular Architecture and Unique Teichoic Acid Recognition Features of Choline-Binding Protein L (CbpL)  
485 Contributing to Pneumococcal Pathogenesis. *Sci. Rep.* 6, 1–19. doi:10.1038/srep38094.
- 486 Haleem, K. S., Ali, Y. M., Yesilkaya, H., Kohler, T., Hammerschmidt, S., Andrew, P. W., et al. (2019). The pneumococcal  
487 surface proteins PspA and PspC sequester host C4-binding protein to inactivate complement C4B on the bacterial  
488 surface. *Infect. Immun.* 87, 1–10. doi:10.1128/IAI.00742-18.
- 489 Hall-Stoodley, L., Costerton, J. W., and Stoodley, P. (2004). Bacterial biofilms: From the natural environment to infectious  
490 diseases. *Nat. Rev. Microbiol.* 2, 95–108. doi:10.1038/nrmicro821.
- 491 Hammerschmidt, S., Wolff, S., Hocke, A., Rosseau, S., Müller, E., and Rohde, M. (2005). Illustration of pneumococcal  
492 polysaccharide capsule during adherence and invasion of epithelial cells. *Infect. Immun.* doi:10.1128/IAI.73.8.4653-  
493 4667.2005.
- 494 Hayhurst, E. J., Kailas, L., Hobbs, J. K., and Foster, S. J. (2008). Cell wall peptidoglycan architecture in *Bacillus subtilis*.  
495 *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.0804138105.
- 496 Heß, N., Waldow, F., Kohler, T. P., Rohde, M., Kreikemeyer, B., Gómez-Mejía, A., et al. (2017). Lipoteichoic acid  
497 deficiency permits normal growth but impairs virulence of *Streptococcus pneumoniae*. *Nat. Commun.* 8.  
498 doi:10.1038/s41467-017-01720-z.
- 499 Hews, C. L., Cho, T., Rowley, G., and Raivio, T. L. (2019). Maintaining Integrity Under Stress: Envelope Stress Response  
500 Regulation of Pathogenesis in Gram-Negative Bacteria. *Front. Cell. Infect. Microbiol.*  
501 doi:10.3389/fcimb.2019.00313.
- 502 Hoskins, J., Alborn, J., Arnold, J., Blaszcak, L. C., Burgett, S., Dehoff, B. S., et al. (2001). Genome of the bacterium  
503 *Streptococcus pneumoniae* strain R6. *J. Bacteriol.* 183, 5709–5717. doi:10.1128/JB.183.19.5709-5717.2001.
- 504 Iovino, F., Molema, G., and Bijlsma, J. J. E. (2014). Platelet endothelial cell adhesion molecule-1, a putative receptor for  
505 the adhesion of *Streptococcus pneumoniae* to the vascular endothelium of the blood-brain barrier. *Infect. Immun.*  
506 doi:10.1128/IAI.00046-14.

**Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage**

- 507 Jarva, H., Janulezyk, R., Hellwage, J., Zipfel, P. F., Björck, L., and Meri, S. (2002). *Streptococcus pneumoniae* Evades  
 508 Complement Attack and Opsonophagocytosis by Expressing the *pspC* Locus-Encoded Hic Protein That Binds to  
 509 Short Consensus Repeats 8–11 of Factor H. *J. Immunol.* doi:10.4049/jimmunol.168.4.1886.
- 510 Kadioglu, A., Weiser, J. N., Paton, J. C., and Andrew, P. W. (2008). The role of *Streptococcus pneumoniae* virulence  
 511 factors in host respiratory colonization and disease. *Nat. Rev. Microbiol.* doi:10.1038/nrmicro1871.
- 512 Kietzman, C. C., Gao, G., Mann, B., Myers, L., and Tuomanen, E. I. (2016). Dynamic capsule restructuring by the main  
 513 pneumococcal autolysin LytA in response to the epithelium. *Nat. Commun.* 7. doi:10.1038/ncomms10859.
- 514 Lanie, J. A., Ng, W. L., Kazmierczak, K. M., Andrzejewski, T. M., Davidsen, T. M., Wayne, K. J., et al. (2007). Genome  
 515 unencapsulated laboratory strain R6. *J. Bacteriol.* 189, 38–51. doi:10.1128/JB.01148-06.
- 517 Luo, S., Hu, D., Wang, M., Zipfel, P. F., and Hu, Y. (2020). Complement in Hemolysis- and Thrombosis- Related Diseases.  
 518 *Front. Immunol.* doi:10.3389/fimmu.2020.01212.
- 519 Maestro, B., and Sanz, J. M. (2016). Choline binding proteins from *Streptococcus pneumoniae*: A dual role as enzybiotics  
 520 and targets for the design of new antimicrobials. *Antibiotics* 5. doi:10.3390/antibiotics5020021.
- 521 Marks, L. R., Iyer Parameswaran, G., and Hakansson, A. P. (2012). Pneumococcal interactions with epithelial cells are  
 522 crucial for optimal biofilm formation and colonization in vitro and in vivo. *Infect. Immun.* doi:10.1128/IAI.00488-  
 523 12.
- 524 McDevitt, E., Khan, F., Scasny, A., Thompson, C. D., Eichenbaum, Z., McDaniel, L. S., et al. (2020). Hydrogen Peroxide  
 525 Production by *Streptococcus pneumoniae* Results in Alpha-hemolysis by Oxidation of Oxy-hemoglobin to Met-  
 526 hemoglobin. *mSphere.* doi:10.1128/msphere.01117-20.
- 527 Meinel, C., Spartà, G., Dahse, H. M., Hörhold, F., König, R., Westermann, M., et al. (2018). *Streptococcus pneumoniae*  
 528 from Patients with Hemolytic Uremic Syndrome Binds Human Plasminogen via the Surface Protein PspC and Uses  
 529 Plasmin to Damage Human Endothelial Cells. *J. Infect. Dis.* 217, 358–370. doi:10.1093/infdis/jix305.
- 530 Moscoso, M., García, E., and López, R. (2006). Biofilm formation by *Streptococcus pneumoniae*: Role of choline,  
 531 extracellular DNA, and capsular polysaccharide in microbial accretion. *J. Bacteriol.* 188, 7785–7795.  
 532 doi:10.1128/JB.00673-06.
- 533 Nobbs, A. H., Lamont, R. J., and Jenkinson, H. F. (2009). Streptococcus Adherence and Colonization. *Microbiol. Mol.*  
 534 *Biol. Rev.* 73, 407–450. doi:10.1128/mmbr.00014-09.
- 535 Orihuela, C. J., Mahdavi, J., Thornton, J., Mann, B., Wooldridge, K. G., Abouseada, N., et al. (2009). Laminin receptor  
 536 initiates bacterial contact with the blood brain barrier in experimental meningitis models. *J. Clin. Invest.* 119.  
 537 doi:10.1172/JCI36759.
- 538 Park, S. S., Gonzalez-Juarbe, N., Martínez, E., Hale, J. Y., Lin, Y. H., Huffines, J. T., et al. (2021). *Streptococcus*  
 539 *pneumoniae* binds to host lactate dehydrogenase via *pspa* and *pspc* to enhance virulence. *MBio.*  
 540 doi:10.1128/mBio.00673-21.
- 541 Parker, D., Soong, G., Planet, P., Brower, J., Ratner, A. J., and Prince, A. (2009). The NanA neuraminidase of *Streptococcus*  
 542 *pneumoniae* is involved in biofilm formation. *Infect. Immun.* doi:10.1128/IAI.00228-09.
- 543 Pearce, B. J., Iannelli, F., and Pozzi, G. (2002). Construction of new unencapsulated (rough) strains of *Streptococcus*  
 544 *pneumoniae*. *Res. Microbiol.* 153, 243–247. doi:10.1016/S0923-2508(02)01312-8.
- 545 Pérez-Dorado, I., Galan-Bartual, S., and Hermoso, J. A. (2012). Pneumococcal surface proteins: When the whole is greater  
 546 than the sum of its parts. *Mol. Oral Microbiol.* 27, 221–245. doi:10.1111/j.2041-1014.2012.00655.x.
- 547 Pinho, M. G., Kjos, M., and Veening, J. W. (2013). How to get (a)round: Mechanisms controlling growth and division of

Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage

- 548 coccoid bacteria. *Nat. Rev. Microbiol.* 11, 601–614. doi:10.1038/nrmicro3088.
- 549 Pracht, D., Elm, C., Gerber, J., Bergmann, S., Rohde, M., Seiler, M., et al. (2005). PavA of *Streptococcus pneumoniae*  
550 modulates adherence, invasion, and meningeal inflammation. *Infect. Immun.* doi:10.1128/IAI.73.5.2680-2689.2005.
- 551 Rai, P., He, F., Kwang, J., Engelward, B. P., and Chow, V. T. K. (2016). Pneumococcal Pneumolysin Induces DNA Damage  
552 and Cell Cycle Arrest. *Sci. Rep.* doi:10.1038/srep22972.
- 553 Rai, P., Parrish, M., Tay, I. J. J., Li, N., Ackerman, S., He, F., et al. (2015). *Streptococcus pneumoniae* secretes hydrogen  
554 peroxide leading to DNA damage and apoptosis in lung cells. *Proc. Natl. Acad. Sci. U. S. A.*  
555 doi:10.1073/pnas.1424144112.
- 556 Ramos-Sevillano, E., Moscoso, M., García, P., García, E., and Yuste, J. (2011). Nasopharyngeal colonization and invasive  
557 disease are enhanced by the cell wall hydrolases LytB and LytC of *Streptococcus pneumoniae*. *PLoS One* 6.  
558 doi:10.1371/journal.pone.0023626.
- 559 Reid, S. D., Hong, W., Dew, K. E., Winn, D. R., Pang, B., Watt, J., et al. (2009). *Streptococcus pneumoniae* forms surface-  
560 attached communities in the middle ear of experimentally infected chinchillas. *J. Infect. Dis.* doi:10.1086/597042.
- 561 Rennemeier, C., Hammerschmidt, S., Niemann, S., Inamura, S., Zähringer, U., and Kehrel, B. E. (2007). Thrombospondin-  
562 1 promotes cellular adherence of Gram-positive pathogens via recognition of peptidoglycan. *FASEB J.* 21, 3118–  
563 3132. doi:10.1096/fj.06-7992com.
- 564 Repnik, U., Knezevic, M., and Jeras, M. (2003). Simple and cost-effective isolation of monocytes from buffy coats. *J.*  
565 *Immunol. Methods.* doi:10.1016/S0022-1759(03)00231-X.
- 566 Roig-Molina, E., Sánchez-Angulo, M., Seelc, J., Garcia-Asencio, F., Nau, R., Sanz, J. M., et al. (2020). Searching for  
567 Antipneumococcal Targets: Choline-Binding Modules as Phagocytosis Enhancers. *ACS Infect. Dis.*  
568 doi:10.1021/acsinfecdis.9b00344.
- 569 Saumyaa, Pujanauski, L., Colino, J., Flora, M., Torres, R. M., Tuomanen, E., et al. (2016). Pneumococcal Surface Protein  
570 A Plays a Major Role in *Streptococcus pneumoniae*- Induced Immunosuppression. *J. Immunol.*  
571 doi:10.4049/jimmunol.1502709.
- 572 Schmidt, F., Kakar, N., Meyer, T. C., Depke, M., Masouris, I., Burchhardt, G., et al. (2019). *In vivo* proteomics identifies  
573 the competence regulon and AliB oligopeptide transporter as pathogenic factors in pneumococcal meningitis. *PLOS*  
574 *Pathog.* 15. doi:10.1371/journal.ppat.1007987.
- 575 Shak, J. R., Vidal, J. E., and Klugman, K. P. (2013). Influence of bacterial interactions on pneumococcal colonization of  
576 the nasopharynx. *Trends Microbiol.* doi:10.1016/j.tim.2012.11.005.
- 577 Shizukuishi, S., Ogawa, M., Ryo, A., and Ohnishi, M. (2020). *Streptococcus pneumoniae* promotes its own survival via  
578 choline-binding protein CbpC-mediated degradation of ATG14. *Autophagy.* doi:10.1080/15548627.2020.1776475.
- 579 Sueki, A., Stein, F., Savitski, M. M., Selkrig, J., and Typas, A. (2020). Systematic Localization of *Escherichia coli*  
580 Membrane Proteins. *mSystems.* doi:10.1128/msystems.00808-19.
- 581 Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: Protein-protein  
582 association networks with increased coverage, supporting functional discovery in genome-wide experimental  
583 datasets. *Nucleic Acids Res.* doi:10.1093/nar/gky1131.
- 584 Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021:  
585 Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets.  
586 *Nucleic Acids Res.* doi:10.1093/nar/gkaa1074.
- 587 Ulrych, A., Fabrik, I., Kupcik, R., Vajrychova, M., Doubravova, L., and Branny, P. (2021). Cell wall stress stimulates the  
588 activity of the protein kinase StkP of *Streptococcus pneumoniae*, leading to multiple phosphorylation. *JMB Ahead of*

This is a provisional file, not the final typeset article

**Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage**

- 589 publishing.
- 590 Van Raaphorst, R., Kjos, M., and Veening, J. W. (2017). Chromosome segregation drives division site selection in  
591 *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. U. S. A.* 114, E5959–E5968. doi:10.1073/pnas.1620608114.
- 592 Vilhena, C., Kaganovitch, E., Grünberger, A., Motz, M., Forné, I., Kohlheyer, D., et al. (2019). Importance of pyruvate  
593 sensing and transport for the resuscitation of viable but nonculturable *Escherichia coli* K-12. *J. Bacteriol.*  
594 doi:10.1128/JB.00610-18.
- 595 von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., et al. (2005). STRING: Known and  
596 predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.*  
597 doi:10.1093/nar/gki005.
- 598 Voß, F., Kohler, T. P., Meyer, T., Abdullah, M. R., Van Opzeeland, F. J., Saleh, M., et al. (2018). Intranasal vaccination  
599 with lipoproteins confers protection against pneumococcal colonisation. *Front. Immunol.* 9.  
600 doi:10.3389/fimmu.2018.02405.
- 601 Voss, S., Hallström, T., Saleh, M., Burchhardt, G., Pribyl, T., Singh, B., et al. (2013). The Choline-binding Protein PspC  
602 of *Streptococcus pneumoniae* Interacts with the C-terminal Heparin-binding Domain of Vitronectin. *J. Biol. Chem.*  
603 288. doi:10.1074/jbc.M112.443507.
- 604 Waldow, F., Kohler, T. P., Hess, N., Schwudke, D., Hammerschmidt, S., and Gisch, N. (2018). Attachment of  
605 phosphorylcholine residues to pneumococcal teichoic acids and modification of substitution patterns by the  
606 phosphorylcholine esterase. *J. Biol. Chem.* 293, 10620–10629. doi:10.1074/jbc.RA118.003360.
- 607 Yother, J., and Briles, D. E. (1992). Structural properties and evolutionary relationships of PspA, a surface protein of  
608 *Streptococcus pneumoniae*, as revealed by sequence analysis. *J. Bacteriol.* 174, 601–609. doi:10.1128/jb.174.2.601-  
609 609.1992.
- 610 Zipfel, P. F., Hallström, T., Hammerschmidt, S., and Skerka, C. (2008). The complement fitness Factor H: Role in human  
611 diseases and for immune escape of pathogens, like pneumococci. *Vaccine* 26, 67–74.  
612 doi:10.1016/j.vaccine.2008.11.015.
- 613 Zipfel, P. F., Würzner, R., and Skerka, C. (2007). Complement evasion of pathogens: Common strategies are shared by  
614 diverse organisms. *Mol. Immunol.* doi:10.1016/j.molimm.2007.06.149.
- 615
- 616





617 **Figures Caption:**

618 **Figure 1- Comparison of biofilm formation among different *S. pneumoniae* strains.** (A)- Quantification of  
619 biofilms grown on multi-well plates via crystal violet staining. Brown-Forsythe ANOVA test with Dunnett's T3  
620 multiple comparisons test and  $p < 0,0001$ (\*\*\*). (B)- Representative images of live/dead staining on biofilms  
621 grown in glass coverslips. Red color represents dead and green alive bacteria. Scale bar is 10  $\mu\text{m}$ . (C)-  
622 Representative SEM images of WT and *lytA* mutant biofilms formed on glass coverslips, at 10,000x (panel 1  
623 and 2) or 20,000x magnifications (panels 3-4) (D)- Representative SEM images of *lytA* mutant with increasing  
624 magnifications (panel 1 to 4).

625 **Figure 2- Evaluation of the effect of different *S pneumoniae* strains on red blood and epithelial cells.** (A)-  
626 Assessment of A549 cell viability by cell cytotoxicity assay. Human epithelial cells were challenged with either  
627 WT (*S. pneumoniae* D39) or the indicated mutant strains. DMEM, THY and dead A549 cells were used as  
628 controls. One-way ANOVA with Dunnett's multiple comparisons test and  $p < 0,0001$ (\*\*\*),  $p < 0,0005$ (\*\*\*). (B)-  
629 Hemolysis of human red blood cells by *S. pneumoniae* strains. One -way ANOVA with Tukey's multiple  
630 comparisons test and  $p < 0,05$ (\*).

631 **Figure 3- Evaluation of the effect of different *S pneumoniae* strains on human endothelial cell surface**  
632 **marker expression.** (A)- Representative images of HUVEC-bacteria co-incubation visualized by CLSM.  
633 PECAM-1 FITC signal is represented in green, nuclei are stained in DAPI (blue) and bacteria are stained with  
634 anti-pneumococcus antiserum (red fluorescence). Scale bar 10 $\mu\text{m}$ . (B)- Quantification of PECAM-1 expression.  
635 One-way ANOVA with Dunnett's multiple comparisons test and  $p < 0,05$ (\*).

636 **Figure 4- STRING Protein-Protein Interaction (PPI) Networks.** (A)- Linear interaction network of the three  
637 studied CBPs, LytA, PspA and PspC (CbpA). Colorful circles or nodes represent the input proteins. (B)- PPI of  
638 interaction partners of PspA, PspC (CbpA) and LytA. Strength (confidence) of interaction is represented by  
639 thicker grey lines. Dashed lines represent interactions based in fewer criterions. The different colors are specific  
640 of different clusters (MCL clustering).



641 **Tables:**

642

643 **Table 1- Summarizing table of the influence of each strain on pneumococcal pathogenic features.**  
 644 Upwards arrow indicates an increase activity, downwards arrow indicate a decrease activity.

645

646

647

648

649

650

651

652

653

654

655

656

Feature	D39	$\Delta$ <i>pspA</i>	$\Delta$ <i>pspC</i>	$\Delta$ <i>lytA</i>
Biofilms	↓	↓	↓	↑
Epithelial metabolism	↓	↑	↑	↑
Haemolysis	↑	↓	↓	↓
Endothelial PECAM-1 expression	↑	↓	↓	↓

**Figure 1**

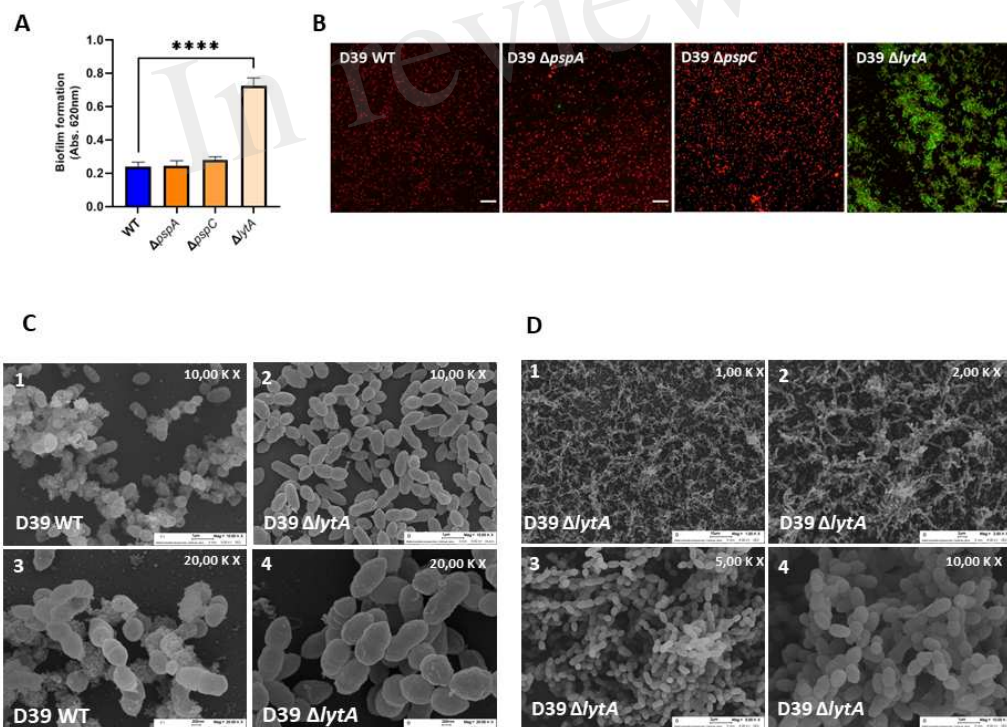




Figure 5

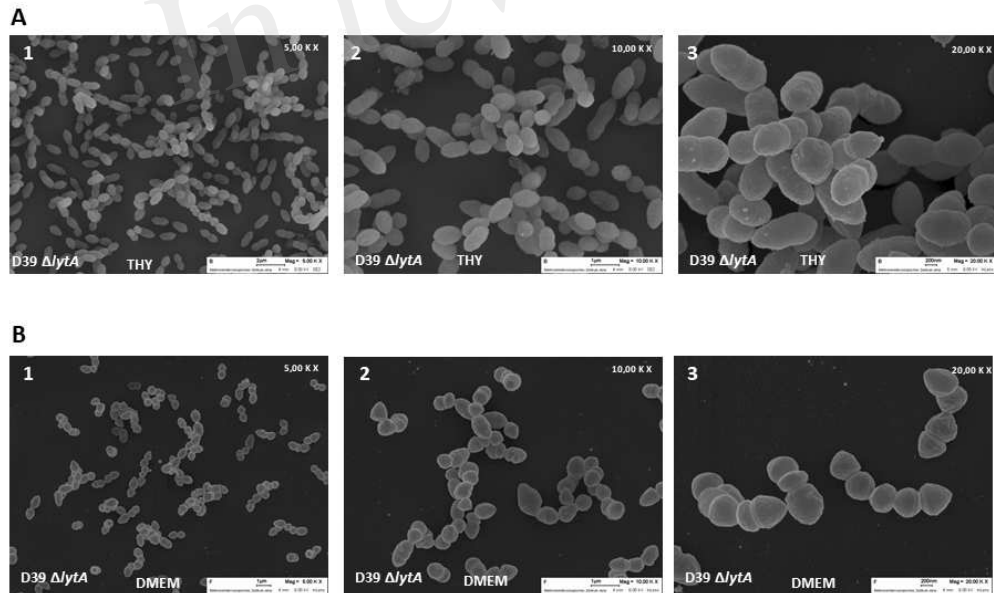


Table 1

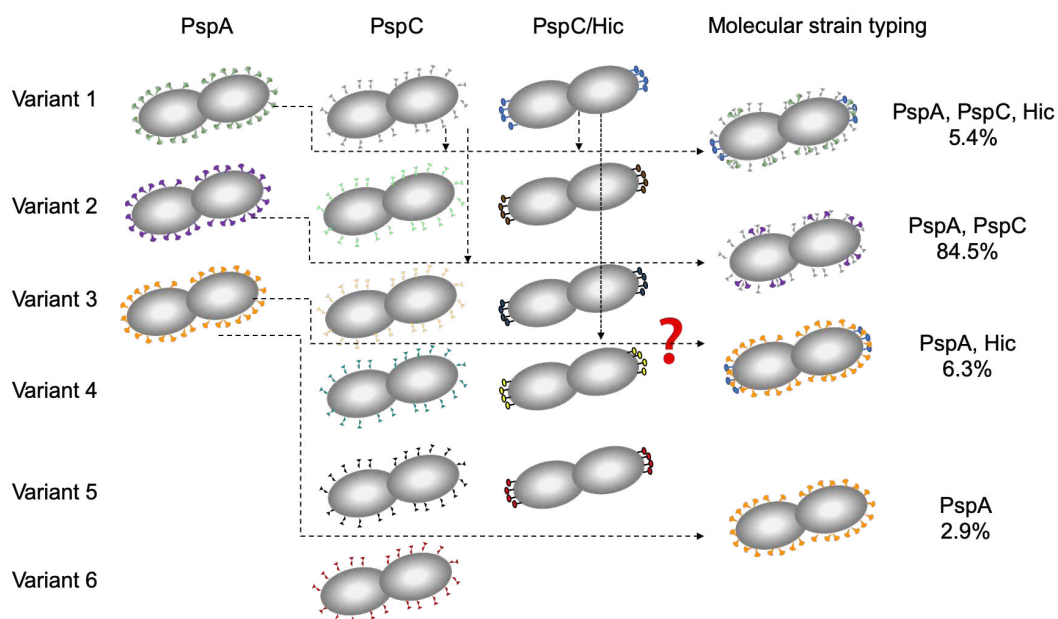
Feature	D39	$\Delta$ pspA	$\Delta$ pspC	$\Delta$ lytA
Biofilms	↓	↓	↓	↑
Epithelial metabolism	↓	↑	↑	↑
Haemolysis	↑	↓	↓	↓
Endothelial PECAM-1 expression	↑	↓	↓	↓

## 5 Discussion

The human pathobiont *S. pneumoniae*, to survive in a human, uses diverse surface proteins to interact with and control human immune attack. Immune evasion of *S. pneumoniae* includes control of the complement as well as innate and adaptive immunity. As a surface exposed protein, PspA influences the interaction of the bacterium with the host by interfering with the fixation of complement C3 which plays a central role in the activation of the complement system<sup>22</sup>. In addition, two other pneumococcal surface proteins, PspC and Hic contribute to complement evasion by binding Factor H<sup>105</sup>. In this work, I studied three pneumococcal immune evasion proteins: PspA, PspC, and PspC/Hic regarding sequence variation, domain composition, and functional analyses. Furthermore, the analytical methodology in this study has been applied to pHUS isolates' PspA and PspC to understand the pathogenesis of Sp-HUS disease and vaccine design.

### Molecular strain typing

By comparing 48 PspA proteins encoded by various *S. pneumoniae* strains, the proteins split into three families and the pattern and sequence variability of PspA seems to extend to the strain level. All strains of the human pathobiont *S. pneumoniae* have *pspA* genes and are serologically variable<sup>117</sup>. PspC and Hic were assumed for a long time that these pneumococcal surface proteins represent a protein family comprising eleven subgroups. However, sequence and domain evaluation of PspC and Hic proteins revealed a greater diversity among individual proteins. In contrast to previous assumptions, Hic, a highly atypical *pspC* allele is distinct from PspC. 85.4% of invasive disease associated *S. pneumoniae* strains present PspC genes; 5.4% of strains that contain both PspC and PspC/Hic with LPsTG anchor; 6.3% of isolates contain only Hic; and the remaining isolates have neither PspC nor Hic<sup>118</sup>. The extreme diversity of PspA, PspC, and Hic is much larger than that of serotypes, ultimately allowing molecular strain typing based on sequence fingerprints. These two studies on PspA, PspC and Hic allow us to have an overview of these three complement evasion proteins and make it possible for molecular strain typing (**Manuscript 1, Manuscript 2**). Four possible example strain types are shown in **Figure 7**. This methodology is a new way to type the strain independent of serotypes.



**Figure 7 Molecular strain typing of PspA, PspC and Hic.** Representative molecular typing of four strains. 3 PspA variants, 6 PspC variants, and 5 Hic variants are shown in different color and items.

**Pneumococcal complement evasion proteins: PspA, PspC and Hic generate diversity in multiple ways.**

PspA variants are separated into three families. Due to structural similarities and conserved binding profiles, PspC and Hic were assumed for a long time that both proteins represent the same protein family comprised of eleven subgroups. All three pneumococcal evasion proteins generate diversity by assembling different domains, using domains in different combinations, and forming different domain patterns. In my study, domain composition inspection was applied to three proteins PspA, PspC, and Hic, which offers a new approach to analyze the three pneumococcal strain diversity (**manuscript 1 and 2**). The three pneumococcal evasion proteins display mosaic structures, and show domain pattern variability and sequence variation between single strains. The mature PspA is composed of six domains (HVD, VD, FDD, D4, PRD, CBD), being selected from a portfolio of nine domains. The mature PspC and the Hic/PspC proteins are more heterogeneous and show a complex domain composition and a high degree of sequence variability. PspC and Hic variants are composed of variable number of domains ranging from 4 to 11 (HVD, RD, RCD, S<sub>n</sub>D/GS<sub>2</sub>, RCD-E, PspA-like, ExPRD, β-AG, unique domain, PRD, CBD/LPsTG). Each domain shows the sequence variation among the tested strains. At the next level, some domains have a modular composition. The repetitive composition specific elements

are mostly related to each other both in the number of modules and in sequence. Repetitive modules exist in three domains of PspA: FDD, PRD, and CBD, in the two domains of PspC: PRD and CBD. The repetitive composition also occurs at the domain level, for example, the repeat domain repeats itself in some PspC and Hic variants. Such domain duplication could evolve a new or modified function either by sequence divergence or by combination with other domains to form new domain architecture<sup>119</sup>. I also identified new domains and PspA's new domain variants (HVD, VD, FDD, PRD), PspC's nine new domains, and also three new alternates of the PRD. This diversity of three proteins results in antigenic variation, functional specialization, and mechanisms of cell wall anchoring<sup>120,121</sup>.

The NCBI protein database includes over 20,000 entries of PspA proteins and over 60,000 PspC and Hic proteins. The application of the domain-based pattern approach is expected to improve the characterization of individual proteins and identify new domains in particular from clinical pneumococcal isolates. A combined evaluation of all three multivariant genes or proteins can clearly improve the precision of domain and sequence-based strain typing. As these three multidomain proteins differ in each strain, we recommend combining the name of protein and strain, for example, PspA<sub>D39</sub>, PspC<sub>TIGR4</sub>.

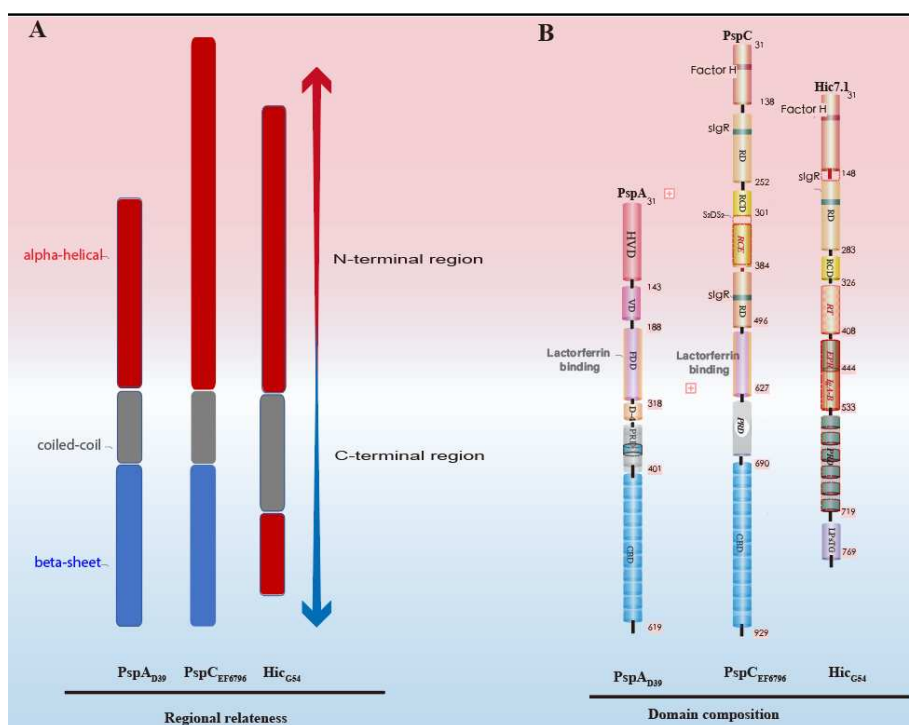
### **Domain and structural relatedness of PspA, PspC, and Hic**

By analyzing the secondary structure of PspA, PspC and Hic proteins, I provide in my thesis project a new way to study pneumococcal multidomain immune evasion proteins. PspA and PspC have related structural compositions with three different regions, including an N-terminal  $\alpha$ -helical region, a coiled-coil structured middle region followed by  $\beta$ -sheet structured regions (**Figure 8A**). Similarly, Hic variants have related structural with PspA and PspC in the N-terminal region. However, Hic proteins use a distinct  $\alpha$ -helical region as the anchor in the C-terminus. Given the secondary structure, amino acid composition, and homology differences between the N-terminal and C-terminal, PspA's family separation is determined by the N-terminal (**manuscript 2**); the separation of PspC and Hic is mainly determined by the C-terminal (**manuscript 1**).

According to the domain pattern inspection of PspA, PspC, and Hic, here I show that each PspA is composed of six domains (HVD, VD, FDD, D4, PRD, CBD) out of a 9-domain reservoir (HVD, VD, FDD<sub>I</sub>, FDD<sub>II</sub>, FDD<sub>III</sub>, D4, PRD<sub>1</sub>, PRD<sub>2</sub>, CBD). PspC and Hic are

composed of a variable number of domains ranging from 3 to 10 (HVD, RD, RCD, RCD-E, PspA-like, ExPRD,  $\beta$ -AG, unique domain, PRD, CBD/LPsTG). As shown in **Figure 8B**, five PspA domains: the HVD, VD, FDD<sub>II</sub>, and FDD<sub>III</sub> are PspA specific. In contrast, the well-researched domain FDD<sub>I</sub>, which represents a lactoferrin binding domain is also used by some PspC variants (PspC1.1, PspC5.1); PRD<sub>1</sub>, PRD<sub>2</sub> are also shared with some PspC and Hic variants (PspC3.1, PspC10.1), and some of these proteins use different “proline segment core”. Additionally, CBDs are used by both PspA and PspC (**Figure 8**), CBD is also found in other pneumococcal choline binding proteins, for example, LytA, LytB, and LytC. The related domains (HVD with Factor H binding site, RD with slgR binding site, RCD) between PspC and Hic variants are located in their N-terminal, not the C-terminal regions. PspA, PspC, and Hic have different C-terminal anchors. PspA and PspC proteins use the CBDs, which contact multiple choline-moieties in a non-covalent manner. However, Hic with the LPsTG anchors attaches the proteins covalently to the peptidoglycan<sup>122</sup>. The different C-terminal anchor not only influences the strengths of the interaction of the protein to the bacterial surface, but also the length and composition of the PRD influence proteins position and topology. Furthermore, the C-terminal anchors seem to influence the selection, composition, number of the N-terminal domain, and distribution of the protein on the surface. These differences in domain composition likely influence protein function regarding also immune evasion and may result in different domains extending beyond the cell wall. Domain, as the evolutionary unit, their similarities in the N-terminal domains may reflect that they tend to have a common ancestor or have the same function<sup>123</sup>.





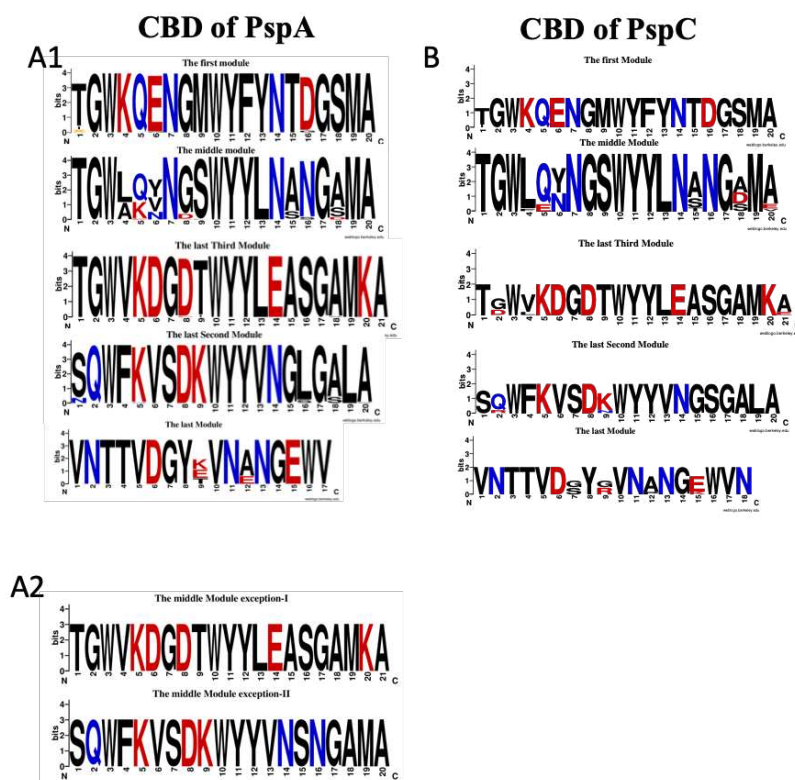
**Figure 8 Structure and domain composition of pneumococcal PspA, PspC, and Hic.** A. Representative structural profile from each protein. PspA from strain D39, PspC from EF6796, and PspC7.1/Hic from G54. PspA and PspC are formed by three different structural regions:  $\alpha$ -helical region (red), a coiled-coil structured middle region (grey), and a  $\beta$ -sheet structured regions (blue). B. Domain composition of three proteins. PspA<sub>D39</sub> is composed of HVD, VD, FDD, D-4, PRD<sub>1</sub>, and CBD. PspC<sub>EF6796</sub> is composed of HVD, 2 copies of RD, RCD, RCE, Lactoferrin binding domain (FDD<sub>I</sub> of PspA), PRD<sub>1</sub>, and CBD. Hic<sub>G54</sub> is composed of HVD, RD, RCD, RCE, Lactoferrin binding domain (FDD<sub>I</sub> of PspA, PspA like), PRD<sub>1</sub>, R-type, IgA binding, PRD<sub>2</sub>, and LPsTG anchor.

### Module diversity of CBD of PspA and PspC

CBDs are the anchor of pneumococcal choline binding proteins by binding choline in the cell wall. CBDs of choline-binding protein J has also been reported to contribute to the pneumococcal adhesion to human lung epithelial cell A549<sup>124</sup>. The simulated choline binding polypeptides derived from CBDs of LytA show growth inhibitory functions<sup>125</sup>. PspC and PspA both proteins belong to choline binding proteins (CBPs) and the proteins share related choline-binding domains as anchors. Besides, as anchors, CBDs of PspC and PspA's function needs more study to reveal.

Further study on CBDs' sequence analysis in detail reveals that: (1) CBD is composed of 5 different module patterns; (2) PspA and PspC share similar modules in the CBD. As shown in **Figure 9**, both PspA and PspC' CBD consist of 5 specific tandem module patterns: one copy of the first module, several copies (5-10) of the middle module depending on the pneumococcal

strain, and one copy of the last three modules (defined as the last third, the last second and the last module respectively). The first module, middle modules, and the last second module are 20 aa long. The last third modules are 21 aa, and the last modules of PspA' CBD are 17 aa long, but the last modules of PspC' CBD are 18 aa long. Most amino acids in one module at a given position are rather conserved, the amino acids show a specific module pattern that is related to interaction with choline in the cell wall. Only some specific positions are variable and they are composed of two or three variable amino acids. For example, the first amino acid at the first position of the N-terminal module either is T or I. The 5 module patterns in **Figure 9A1 and B** are similar between PspA and PspC. The outstanding feature of PspA CBD is that the middle modules with 20 amino acids have one exception. Some PspA middle modules have one or two copies of two consecutive specific patterns with 21 amino acids (**Figure 9A2**). This module pattern conservation and slight difference at specific positions might relate to the structural plasticity of CBD. The structural plasticity in each CBPs yields specific features regarding aspects such as localization on the surface, and specific catalytic activities<sup>126</sup>. What the exception and variability at specific position meaning need more study to find out.



**Figure 9** Sequence logos of PspA and PspC' CBD. One stack (one amino acid) for each position in the sequence. The scale reflects the certitude of a particular amino acid at a given position. The amino acid

residues at each position are arranged in order of predominance from top to bottom, with the highest frequency residue at the top. Residue colors such as red define charged amino acid (D, E, K, R), blue correspond to neutral polar, and other amino acids are in black.

**The role of PspA, PspC, and LytA on bacterial biofilm formation, host cell damage, and their integration in bacterial protein networks.**

LytA, another choline binding protein, is central in biofilm formation. *S. pneumoniae* D39 mutants that lack LytA protein form more dense filament-like biofilms. In contrast, although using related CBDs as anchors, *pspA* and *pspC* deletion does not affect the biofilm matrix formation. LytA prevents biofilm formation, as bacteria lacking this autolysin form increased biofilms, which was also proved that LytA has an anti-biofilm effect<sup>127</sup>. Lacking the three choline-binding proteins induced endothelial cells to express platelet endothelial cell adhesion molecule (PECAM-1) (**manuscript 3**). PECAM-1, as an efficient signaling molecule, is capable of exhibiting both input and output signaling<sup>128</sup>. PECAM-1 on endothelial cells was studied and reported that certain cytokine combinations, ie, TNF- $\alpha$  and IFN- $\gamma$ , can reduce the expression of PECAM-1 from endothelial cell junctions<sup>129</sup>. Why exactly PECAM-1 is induced by these three choline binding proteins is still unclear and needs more research to find out.

Protein-protein network analysis shows that once the bacterial SOS response cascade is triggered, biofilms need to be established, most likely leading to repression. *S. pneumoniae* lacking either *pspA*, *pspC*, or *lytA* is less efficient in inducing lysis of human erythrocytes. McDevitt proposes that pneumococcus hemolysis activity is mainly mediated by hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) than by excreted pneumolysin<sup>130</sup>. Pneumolysin is closely related to LytA, but PspA and PspC do not directly interact with ply, instead, they interact with neuraminidase (NanA). LytA is integrated into the bacterial cell-division machinery and in teichoic acid biosynthesis (**manuscript 3**). NanA is the enzyme that is upregulated upon contact with the host cells and it cleaves host glycoconjugates when it is secreted outside of the bacteria<sup>79</sup>. H<sub>2</sub>O<sub>2</sub> is produced by *S. pneumoniae* DNA damage and apoptosis in the lung cells<sup>131</sup>. The hemolytic profile of the tested strains indicated that CBPs mutants reduce cytotoxicity of *S. pneumoniae* towards human epithelial and endothelial cells.

**PspA<sub>HUS</sub> and PspC<sub>HUS</sub> are specific clades predominate.**

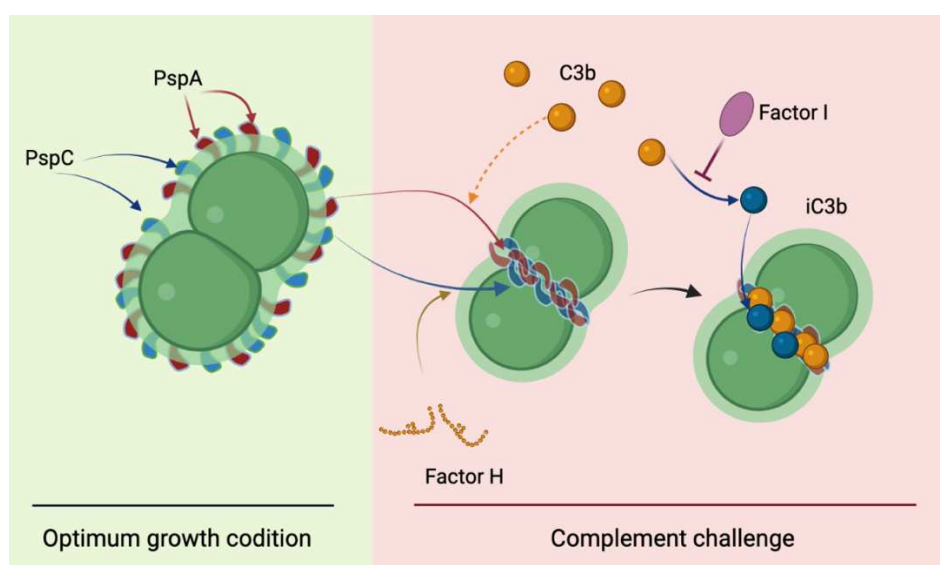
The analytic methodology applied here (**manuscript 1 and manuscript 2**) was further used for characterizing the PspA and PspC variants from HUS clinical strains. Each tested Spn-HUS isolate encodes a novel PspA and PspC variant that shows a unique, specific domain profile. Based on the homology analysis, PspA<sub>HUS</sub> and PspC<sub>HUS</sub> are specific clade determinants. PspA<sub>HUS</sub> have a preference for family II and family III. A functional consequence is that lactoferrin significantly binds more PspA<sub>G54</sub> (family II), PspA<sub>TIGR4</sub> (family III), and PspA<sub>HUS</sub> (family II / family III) than PspA<sub>D39</sub> (family I). Lactoferrin (LF) is an iron-binding glycoprotein, that exists in human serum and has iron metabolism and antimicrobial function during inflammation<sup>132</sup>. LF inhibits the classical complement pathway<sup>133</sup>. The preferable domain pattern of PspA<sub>HUS</sub> to family II and family III which bind more lactoferrin might contribute to the HUS pathogenesis. Whether lactoferrin binding function directly relates to the sequence differences among the three FDDs (FDD<sub>I</sub>, FDD<sub>II</sub>, and FDD<sub>III</sub>), needs to be evaluated by additional research.

According to the inspection of the 38 sequenced PspC<sub>HUS</sub>, 38 PspC<sub>HUS</sub>, on one side, exhibit a considerable diversity but proteins show the same domain pattern. On the other side, FHR1, and FHR5 bind PspC<sub>HUS</sub> from different clades with different intensity. This result highlights a heterogeneity among pHUS strains in pathogen-host interactions. FHRs similar to Factor H share related domain structure, but both FHR1 and FHR5 lack a complement regulatory domain and compete for binding host cell glycans<sup>134</sup>. FHRs, particularly FHR1 and FHR5, are prevalently found in renal biopsies of C3 Glomerulopathy patients<sup>135,136</sup>. Some PspC<sub>HUS</sub> variants bind FHR1 and FHR5 which might relate to HUS pathogenesis.

### **PspA and PspC contribute to immune evasion by interacting with human complement proteins.**

To study the role of PspA and PspC during *S. pneumoniae* evasion from complement, I used strain D39 stains which had the *pspA* or *pspC* genes deleted. PspA<sub>D39</sub> and PspC<sub>D39</sub> switched from overall cell surface distribution (except the septum) and upon serum challenge both PspA<sub>D39</sub> and PspC<sub>D39</sub> moved to the division septum under serum challenge (**Figure 10**). In this setup, D39Δ*pspA* did reduce the amount of C3b deposition at the septum, but not changing the placement of C3b deposition under serum challenge was detectable as

compared with D39. PspA<sub>D39</sub> might help C3b deposited on the septal area, likely as PspC<sub>D39</sub> recruited factor H to the septum. PspA<sub>D39</sub> and PspC<sub>D39</sub> are localization to the cell surface except for septa under normal growth conditions, which allows *S. pneumoniae* to maximize usage of the nutrients to divide. In contrast, once *S. pneumoniae* is challenged by the complement active serum, PspA<sub>D39</sub> and PspC<sub>D39</sub> change their surface localization; will move to the septum and protect these division sites from complement attack. The dynamic localization of PspA and PspC indicates that the cell wall might rearrange under the complement challenge, thus showing dynamic surface distribution upon immune escape.



**Figure 10 Dynamic location of PspA and PspC interacting with complement regulators response upon serum challenge.** PspA and PspC distribute on the cell surface of streptococci and are absent at the septum. Upon complement challenge by human serum, PspA and PspC mainly localized at the septum. Factor H is recruited by PspC and also localized on the septum of bacteria. PspA assists in C3b deposit on the septum.

### **Sp-HUS response to the complement challenge.**

The inspection of complement resistance of Sp-HUS showed that Sp-HUS strains bind more factor H resulting in reduced C3b surface deposition. This makes Sp-HUS strains more resistant to the complement. Upon complement attack, the localization of PspA<sub>HUS</sub> and PspC<sub>HUS</sub> changes, and the proteins move mainly to the septal area; PspA<sub>HUS</sub> by binding C3b and PspC<sub>HUS</sub> by recruiting Factor H also localized on the septum in response to the normal human serum (NHS). Under the high concentration of NHS (50%), Factor H also assembles at the septum and occupies the majority area around the surface of bacteria for Sp-HUS, the distribution is different from D39. In contrast to strain D39, Sp-HUS

bind more factor H and the acquired human complement regulator apparently influences C3b deposition on the surface. This results in a consequent reduction in complement activation. Recently, Delgado reported that pHUS patients show a lower level of C3 and C4 in acute phase samples<sup>114</sup>. This suggests that Sp-HUS strains consume C3 in patient serum and result in less C3b deposition which relates to the interaction between Sp-HUS and complement regulators. Ágnes proved that complement component C3 and activity of complement pathway (the classical and alternative) are decreased in the acute phase of HUS patients<sup>137</sup>. Rodney also hypothesized that transiently dysregulated complement activation may be related to the pathogenesis of HUS by competitive binding of Factor H by PspC reduced binding of factor H to the endothelial cell<sup>138</sup>. All in all, besides the diverse and dynamic localization of surface protein PspA<sub>HUS</sub> and PspC<sub>HUS</sub> interact with complement proteins and complement regulators that contribute to the immune evasion, which may be involved in the HUS pathogenesis process.

## 6 Summary

*S. pneumoniae*, as an important human pathogenic Gram-positive bacterium causes frequent infections associated with airways. In addition, this pathogen can also cause severe invasive diseases: pHUS. In order to evade the immune system, *S. pneumoniae* has evolved multiple mechanisms to control, avoid and inactivate host immune recognition and break host tissue barriers. More importantly, *S. pneumoniae* expresses a group of diverse surface proteins that bind and interact with human surface proteins. The purpose of my study, firstly, was to find a way to study the diverse sequence of three proteins (PspA, PspC, and Hic). I combined sequence homology and domain architecture evaluation of these three proteins which allow us to apply them to clinical protein variants. Secondly, by using super-resolution technology, the surface distribution of PspA and PspC under normal and complement challenge conditions has been inspected. Thirdly, I identified the function of PspA and PspC on bacterial biofilm formation under stress conditions and host cell damage. In the end, we applied the methodology above to study the Sp-HUS response to the complement challenge.

For the first time, I provide a novel way to study the diverse surface pneumococcal proteins. Both PspA and PspC of Sp-HUS are classified into specific subgroups or families, which might be related to the HUS surviving environment in the human body. PspA and PspC play a role in bacterial biofilm formation under stress condition, and they are required factors to damage the host cell. In the normal growth condition, PspA and PspC are distributed at the cell surface except for the septa region. In contrast, under complement challenge, PspA binding C3b and PspC recruiting factor H do condense on the septum: this is likely correlated to protect the division sites from opsonization. In addition, all tested HUS strains bind factor H more efficiently as compared to the reference strain D39 and this can cause less C3b binding on the surface. In consequence, this makes them more resistant to the host complement and more virulent to humans. This work provided detailed information about PspA and PspC, especially from HUS strains help to understand the pathogenesis of Sp-HUS disease and vaccine design.

## Zusammenfassung

*S. pneumoniae*, als ein wichtiges pathogenes Gram-positives Bakterium verursacht häufige Infektionen mit den Atemwegen und kann auch invasive Erkrankungen verursachen: pHUS. Um der Immunimmunität zu entgehen, hat *S. pneumoniae* mehrere Mechanismen entwickelt, um die Erkennung des Immunangriffs des Wirtes zu vermeiden und um physikalische Barrieren zu überwinden und um in das Wirtsgewebe einzudringen. Noch wichtiger ist, dass *S. pneumoniae* eine Gruppe verschiedener Oberflächenproteine exprimiert, um mit der menschlichen Zell-Oberfläche oder humanen Serumproteinen zu interagieren. Der Zweck meiner Studie war es zunächst, einen Weg zu finden, die vielfältige Sequenz von drei Proteinen (PspA, PspC und Hic) zu untersuchen. Ich kombinierte Sequenzhomologie und Domänenarchitektur dieser drei Pneumokokken Proteine, die es uns ermöglichen, auf klinische Proteinvarianten anzuwenden. Zweitens wurde durch den Einsatz der Super-Resolution-Technologie die Oberflächenverteilung von PspA und PspC unter normalen und komplementären Angriff untersucht. Drittens identifizierte ich die Funktion von PspA und PspC auf die bakterielle Biofilmbildung unter Stressbedingungen und Zellschäden des Wirtes. Mit diesen Ansätzen kann ich die Immunevasion von klinischen Sp-HUS Isolaten hinsichtlich Komplementherausforderung zu untersuchen.

Zum ersten Mal ich eine neue Möglichkeit, die verschiedenen Oberflächen-Pneumokokken-Proteine zu untersuchen. Sowohl PspA als auch PspC von Sp-HUS werden in bestimmte Untergruppen oder Familien eingeteilt, die mit dem Krankheitsbild von Pneumokokken HUS-überlebenden Umwelt - dem menschlichen Körper - zusammenhängen könnte. PspA und PspC spielen eine Rolle bei der bakteriellen Biofilmbildung unter Stressbedingungen und sind bakterielle Virulenzfaktoren welche die Wirtszelle schädigen können. Im normalen Wachstumszustand verteilen sich PspA und PspC auf die Zelloberfläche, außer im Bereich des Septums. Im Gegensatz dazu kondensieren PspA-bindende C3b- und PspC-Rekrutierungsfaktor H unter Komplement-Herausforderung im Bereich des bakteriellen Septums, um so möglicherweise die Divisionsstandorte vor Opsonisierung zu schützen. Darüber hinaus binden alle getesteten klinischen *S. pneumoniae* HUS-Stämme mehr Faktor H und zeigen eine verminderte C3b-Beladung auf der Oberfläche. Somit sind diese klinischen Isolate widerstandsfähiger gegen das Wirtskomplement und zeigen eine erhöhte Virulenz für



den Menschen. Diese Arbeit lieferte die detaillierten Informationen über PspA und PspC, insbesondere von HUS-Stämmen, um die Pathogenese der Sp-HUS-Krankheit und des Impfstoffdesigns zu verstehen.

## 7 Bibliography

1. Engholm DH, Kilian M, Goodsell DS, Andersen ES, Kjærgaard RS. A visual review of the human pathogen *Streptococcus pneumoniae*. *FEMS Microbiol Rev.* 2017;41(6):854-879. doi:10.1093/femsre/fux037
2. Bogaert D, De Groot R, Hermans PWM. *Streptococcus pneumoniae* colonisation: The key to pneumococcal disease. *Lancet Infect Dis.* 2004. doi:10.1016/S1473-3099(04)00938-7
3. Caierão J, Hawkins P, Sant'anna FH, et al. Serotypes and genotypes of invasive *Streptococcus pneumoniae* before and after PCV10 implementation in southern Brazil. *PLoS One.* 2014;9(10). doi:10.1371/journal.pone.0111129
4. Kadioglu A, Weiser JN, Paton JC, Andrew PW. The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat Rev Microbiol.* 2008. doi:10.1038/nrmicro1871
5. Sleeman KL, Griffiths D, Shackley F, et al. Capsular serotype-specific attack rates and duration of carriage of *Streptococcus pneumoniae* in a population of children. *J Infect Dis.* 2006;194(5):682-688. doi:10.1086/505710
6. Copelovitch L, Kaplan BS. *Streptococcus pneumoniae*-associated hemolytic uremic syndrome. *Pediatr Nephrol.* 2008;23(11):1951-1956. doi:10.1007/s00467-007-0518-y
7. Klein PJ, Bulla M, Newman RA, et al. THOMSEN FRIEDENREICH ANTIGEN IN HÆMOLYTIC-URÆMIC SYNDROME. *Lancet.* 1977. doi:10.1016/S0140-6736(77)92915-4
8. Scobell RR, Kaplan BS, Copelovitch L. New insights into the pathogenesis of *Streptococcus pneumoniae*-associated hemolytic uremic syndrome. *Pediatr Nephrol.* 2020;35(9):1585-1591. doi:10.1007/s00467-019-04342-3
9. Stuckey-Schrock K, Hayes BL, George CM. Community-acquired pneumonia in children. *Am Fam Physician.* 2012. doi:10.5005/jp/books/12060\_7
10. UNICEF. *Pneumonia*. A child dies of pneumonai every 39 seconds. 2018
11. Martín-Torres F, Salas A, Rivero-Calle I, et al. Life-threatening infections in children in Europe (the EUCLIDS Project): a prospective cohort study. *Lancet Child Adolesc Heal.* 2018;2(6):404-414. doi:10.1016/S2352-4642(18)30113-5
12. Trs R. Recommendations to assure the quality , safety and efficacy of pneumococcal conjugate vaccines. *World Health.* 2009;(October).
13. WHO, Group A. Pneumococcal vaccines : WHO position paper = 2012. *Wkly Epidemiol Rec.* 2003;78(14):110-119.

14. Rodgers GL, Whitney CG, Klugman KP. Triumph of Pneumococcal Conjugate Vaccines: Overcoming a Common Foe. *J Infect Dis.* 2021;224(4):S352-S359. doi:10.1093/infdis/jiaa535
15. Geno KA, Gilbert GL, Song Y, et al. Pneumococcal Capsules and Their Types : Past , Present , and Future. *Clin Microbiol Rev.* 2015;28(3):871-899. doi:10.1128/CMR.00024-15
16. Straume D, Stamsås GA, Håvarstein LS. Natural transformation and genome evolution in *Streptococcus pneumoniae*. *Infect Genet Evol.* 2015;33(1432):371-380. doi:10.1016/j.meegid.2014.10.020
17. Weiser JN, Ferreira DM, Paton JC. *Streptococcus pneumoniae*: Transmission, colonization and invasion. *Nat Rev Microbiol.* 2018;16(6):355-367. doi:10.1038/s41579-018-0001-8
18. Evans BA, Rozen DE. Significant variation in transformation frequency in *Streptococcus pneumoniae*. *ISME J.* 2013;791-799. doi:10.1038/ismej.2012.170
19. Subramanian K, Henriques-Normark B, Normark S. Emerging concepts in the pathogenesis of the *Streptococcus pneumoniae*: From nasopharyngeal colonizer to intracellular pathogen. *Cell Microbiol.* 2019;21(11):1-10. doi:10.1111/cmi.13077
20. Keller LE, Robinson DA, McDaniel LS. Nonencapsulated *Streptococcus pneumoniae*: Emergence and pathogenesis. *MBio.* 2016;7(2):1-12. doi:10.1128/mBio.01792-15
21. Bergmann S, Hammerschmidt S. Versatility of pneumococcal surface proteins. *Microbiology.* 2006;152(2):295-303. doi:10.1099/mic.0.28610-0
22. Tu AHT, Fulgham RL, Mccrory MA, Briles DE, Szalai AJ. Pneumococcal surface protein A inhibits complement activation by *Streptococcus pneumoniae*. *Infect Immun.* 1999;67(9):4720-4724. doi:10.1128/iai.67.9.4720-4724.1999
23. Gosink KK, Mann ER, Guglielmo C, Tuomanen EI, Masure HR. Role of novel choline binding proteins in virulence of *Streptococcus pneumoniae*. *Infect Immun.* 2000;68(10):5690-5695. doi:10.1128/IAI.68.10.5690-5695.2000
24. Balachandran P, Brooks-Walter A, Virolainen-Julkunen A, Hollingshead SK, Briles DE. Role of pneumococcal surface protein C in nasopharyngeal carriage and pneumonia and its ability to elicit protection against carriage of *Streptococcus pneumoniae*. *Infect Immun.* 2002;70(5):2526-2534. doi:10.1128/IAI.70.5.2526-2534.2002
25. Hollingshead SK, Becker R, Briles DE. Diversity of PspA: Mosaic genes and evidence for past recombination in *Streptococcus pneumoniae*. *Infect Immun.* 2000;68(10):5889-5900. doi:10.1128/IAI.68.10.5889-5900.2000

26. Mukerji R, Hendrickson C, Genschmer KR, et al. The diversity of the proline-rich domain of pneumococcal surface protein A (PspA): Potential relevance to a broad-spectrum vaccine. *Vaccine*. 2018;36(45):6834-6843. doi:10.1016/j.vaccine.2018.08.045
27. Chaplin DD. Overview of the immune response. *J Allergy Clin Immunol*. 2010;125(2 SUPPL. 2):S3-S23. doi:10.1016/j.jaci.2009.12.980
28. Kapetanovic R, Cavillon JM. Early events in innate immunity in the recognition of microbial pathogens. *Expert Opin Biol Ther*. 2007;7(6):907-918. doi:10.1517/14712598.7.6.907
29. Boehm T. Evolution of vertebrate immunity. *Curr Biol*. 2012;22(17):R722-R732. doi:10.1016/j.cub.2012.07.003
30. Kellie S, Al-Mansour Z. Overview of the Immune System. *Micro- Nanotechnol Vaccine Dev*. 2017;357:63-81. doi:10.1016/B978-0-323-39981-4.00004-X
31. Silva-Gomes S, Decout A, Nigou J. Pathogen-Associated Molecular Patterns (PAMPs). In: Parnham M, ed. Basel: Springer Basel; 2015:1-16. doi:10.1007/978-3-0348-0620-6\_35-1
32. Nonaka M, Kimura A. Genomic view of the evolution of the complement system. *Immunogenetics*. 2006;58(9):701-713. doi:10.1007/s00251-006-0142-1
33. Walport MJ. Complement. *N Engl J Med*. 2001;344(14):1058-1066. doi:10.1056/NEJM200104053441406
34. Zipfel PF, Skerka C. Complement regulators and inhibitory proteins. *Nat Rev Immunol*. 2009;9(10):729-740. doi:10.1038/nri2620
35. Poppelaars F, Faria B, da Costa MG, et al. The complement system in dialysis: A forgotten story? *Front Immunol*. 2018;9(JAN). doi:10.3389/fimmu.2018.00071
36. Gaboriaud C, Teillet F, Gregory LA, Thielens NM, Arlaud GJ. Assembly of C1 and the MBL- and ficolin-MASP complexes: structural insights. *Immunobiology*. 2007;212(4-5):279-288. doi:10.1016/j.imbio.2006.11.007
37. Noris M, Remuzzi G. Overview of complement activation and regulation. *Semin Nephrol*. 2013;33(6):479-492. doi:10.1016/j.semnephrol.2013.08.001
38. Thielens NM, Tedesco F, Bohlsion SS, Gaboriaud C, Tenner AJ. C1q: A fresh look upon an old molecule. *Mol Immunol*. 2017;89:73-83. doi:https://doi.org/10.1016/j.molimm.2017.05.025

39. Ajona D, Ortiz-Espinosa S, Pio R. Complement anaphylatoxins C3a and C5a: Emerging roles in cancer progression and treatment. *Semin Cell Dev Biol.* 2019;85:153-163. doi:10.1016/j.semcd.2017.11.023
40. Sonesson A, Ringstad L, Nordahl EA, Malmsten M, Mörgelin M, Schmidtchen A. Antifungal activity of C3a and C3a-derived peptides against *Candida*. *Biochim Biophys Acta.* 2007;1768(2):346-353. doi:10.1016/j.bbame.2006.10.017
41. Józsi M, Prechl J, Bajtay Z, Erdei A. Complement Receptor Type 1 (CD35) Mediates Inhibitory Signals in Human B Lymphocytes. *J Immunol.* 2002;168(6):2782-2788. doi:10.4049/jimmunol.168.6.2782
42. Goldman AS, Prabhakar BS. Immunology Overview. In: Baron S, ed. Galveston (TX); 1996.
43. Dobrina A, Pausa M, Fischetti F, et al. Cytolytically inactive terminal complement complex causes transendothelial migration of polymorphonuclear leukocytes in vitro and in vivo. *Blood.* 2002;99(1):185-192. doi:10.1182/blood.V99.1.185
44. Egan AM, Gordon DL. Burkholderia pseudomallei activates complement and is ingested but not killed by polymorphonuclear leukocytes. *Infect Immun.* 1996;64(12):4952-4959. doi:10.1128/iai.64.12.4952-4959.1996
45. Moghimi SM, Andersen AJ, Ahmadvand D, Wibroe PP, Andresen TL, Hunter AC. Material properties in complement activation. *Adv Drug Deliv Rev.* 2011;63(12):1000-1007. doi:https://doi.org/10.1016/j.addr.2011.06.002
46. Lee JW, Jang JH, Kim JS, et al. Uncontrolled Complement Activation and the Resulting Chronic Hemolysis As Measured by LDH Serum Level At Diagnosis As Predictor of Thrombotic Complications and Mortality in a Large Cohort of Patients with Paroxysmal Nocturnal Hemoglobinuria (PNH). *Blood.* 2011;118(21):3166. doi:10.1182/blood.V118.21.3166.3166
47. Scholl HPN, Issa PC, Walier M, et al. Systemic complement activation in age-related macular degeneration. *PLoS One.* 2008;3(7):1-7. doi:10.1371/journal.pone.0002593
48. Malhotra V, Sim RB. Expression of complement factor H on the cell surface of the human monocytic cell line U937. *Eur J Immunol.* 1985;15(9):935-941. doi:10.1002/eji.1830150913
49. Zipfel PF, Skerka C, Hellwage J, et al. Factor H family proteins: On complement, microbes and human diseases. *Biochem Soc Trans.* 2002;30(6):971-978. doi:10.1042/BST0300971
50. Józsi M. Factor H family proteins in complement evasion of microorganisms. *Front Immunol.* 2017;8(MAY):1-8. doi:10.3389/fimmu.2017.00571

51. Jokiranta TS, Hellwage J, Koistinen V, Zipfel PF, Meri S. Each of the Three Binding Sites on Complement Factor H Interacts with a Distinct Site on C3b. *J Biol Chem*. 2000;275(36):27657-27662. doi:10.1074/jbc.M002903200
52. Kim S, Park E, Min S II, et al. Kidney transplantation in patients with atypical hemolytic uremic syndrome due to complement factor H deficiency: Impact of liver transplantation. *J Korean Med Sci*. 2018;33(1):1-11. doi:10.3346/jkms.2018.33.e4
53. Heinen S, Hartmann A, Lauer N, et al. Factor H-related protein 1 (CFHR-1) inhibits complement C5 convertase activity and terminal complex formation. *Blood*. 2009;114(12):2439-2447. doi:10.1182/blood-2009-02-205641
54. Eberhardt HU, Buhlmann D, Hortschansky P, et al. Human factor H-related protein 2 (CFHR2) regulates complement activation. *PLoS One*. 2013;8(11):1-11. doi:10.1371/journal.pone.0078617
55. Hellwage J, Jokiranta TS, Koistinen V, Vaarala O, Meri S, Zipfel PF. Functional properties of complement factor H-related proteins FHR-3 and FHR-4: Binding to the C3d region of C3b and differential regulation by heparin. *FEBS Lett*. 1999;462(3):345-352. doi:10.1016/S0014-5793(99)01554-9
56. McRae JL, Duthy TG, Griggs KM, et al. Human Factor H-Related Protein 5 Has Cofactor Activity, Inhibits C3 Convertase Activity, Binds Heparin and C-Reactive Protein, and Associates with Lipoprotein. *J Immunol*. 2005;174(10):6250-6256. doi:10.4049/jimmunol.174.10.6250
57. Csincsi ÁI, Kopp A, Zöldi M, et al. Factor H-Related Protein 5 Interacts with Pentraxin 3 and the Extracellular Matrix and Modulates Complement Activation. *J Immunol*. 2015;194(10):4963-4973. doi:10.4049/jimmunol.1403121
58. Goicoechea De Jorge E, Caesar JJE, Malik TH, et al. Dimerization of complement factor H-related proteins modulates complement activation in vivo. *Proc Natl Acad Sci U S A*. 2013;110(12):4685-4690. doi:10.1073/pnas.1219260110
59. Athanasiou Y, Voskarides K, Gale DP, et al. Familial C3 Glomerulopathy Associated with CFHR5 Mutations: Clinical Characteristics of 91 Patients in 16 Pedigrees. *Clin J Am Soc Nephrol*. 2011;6(6):1436-1446. doi:10.2215/CJN.09541010
60. Zipfel PF, Edey M, Heinen S, et al. Deletion of complement factor H-related genes CFHR1 and CFHR3 is associated with atypical hemolytic uremic syndrome. *PLoS Genet*. 2007;3(3):0387-0392. doi:10.1371/journal.pgen.0030041
61. Davis AE 3rd. Biological effects of C1 inhibitor. *Drug News Perspect*. 2004;17(7):439-446. doi:10.1358/dnp.2004.17.7.863703

62. Blom AM, Villoutreix BO, Dahlbäck B. Complement inhibitor C4b-binding protein-friend or foe in the innate immune system? *Mol Immunol*. 2004;40(18):1333-1346. doi:10.1016/j.molimm.2003.12.002
63. Ermert D, Blom AM. C4b-binding protein: The good, the bad and the deadly. Novel functions of an old friend. *Immunol Lett*. 2016;169:82-92. doi:10.1016/j.imlet.2015.11.014
64. Goodship THJ, Liszewski MK, Kemp EJ, Richards A, Atkinson JP. Mutations in CD46, a complement regulatory protein, predispose to atypical HUS. *Trends Mol Med*. 2004;10(5):226-231. doi:10.1016/j.molmed.2004.03.006
65. Huang Y, Qiao F, Abagyan R, Hazard S, Tomlinson S. Defining the CD59-C9 Binding Interaction. *J Biol Chem*. 2006;281(37):27398-27404. doi:10.1074/jbc.M603690200
66. Preissner KT, Podack ER, Müller-Eberhard HJ. The membrane attack complex of complement: relation of C7 to the metastable membrane binding site of the intermediate complex C5b-7. *J Immunol*. 1985;135(1):445—451. <http://europepmc.org/abstract/MED/3998468>.
67. Jenne DE, Tschopp J. Clusterin: the intriguing guises of a widely expressed glycoprotein. *Trends Biochem Sci*. 1992;17(4):154-159. doi:https://doi.org/10.1016/0968-0004(92)90325-4
68. Janeway CA Jr, Travers P, Walport M et al. *Immunobiology: The Immune System in Health and Disease*. 5th editio. New York: Garland Science; 2001. <https://www.ncbi.nlm.nih.gov/books/NBK10757/>.
69. Bonilla FA, Oettgen HC. Adaptive immunity. *J Allergy Clin Immunol*. 2010;125(2):S33-S40. doi:10.1016/j.jaci.2009.09.017
70. Todar K. *Streptococcus pneumoniae*: Pneumococcal pneumonia. Todar's Online Textbook of Bacteriology.
71. Tettelin H, Nelson KE, Paulsen IT, et al. Complete Genome Sequence of a Virulent Isolate of *Streptococcus pneumoniae*. *Science (80- )*. 2001;293(5529):498 LP - 506. doi:10.1126/science.1061217
72. Lanie JA, Ng WL, Kazmierczak KM, et al. Genome sequence of Avery's virulent serotype 2 strain D39 of *Streptococcus pneumoniae* and comparison with that of unencapsulated laboratory strain R6. *J Bacteriol*. 2007;189(1):38-51. doi:10.1128/JB.01148-06
73. Maestro B, Sanz JM. Choline Binding Proteins from *Streptococcus pneumoniae* : A Dual Role as Enzybiotics and Targets for the Design of New Antimicrobials. *antibiotics*. 2016. doi:10.3390/antibiotics5020021

74. Jonsson S, Musher DM, Chapman A, Goree A, Lawrence EC. Phagocytosis and Killing of Common Bacterial Pathogens of the Lung by Human Alveolar Macrophages. *J Infect Dis.* 1985;152(1):4-13. doi:10.1093/infdis/152.1.4
75. Gamez G, Hammerschmidt S. Combat pneumococcal infections: adhesins as candidates for protein-based vaccine development. *Curr Drug Targets.* 2012;13(3):323-337. doi:10.2174/138945012799424697
76. Kadioglu A, Weiser JN, Paton JC, Andrew PW. The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat Rev Microbiol.* 2008;6:288. <https://doi.org/10.1038/nrmicro1871>.
77. Ribeiro ML, Andre GO, Converso TR, et al. Role of *Streptococcus pneumoniae* Proteins in Evasion of Complement-Mediated Immunity. *Front Microbiol.* 2017;8(February):1-20. doi:10.3389/fmicb.2017.00224
78. Jefferies JMC, Tocheva AS, Rubery H, et al. Identification of novel pneumolysin alleles from paediatric carriage isolates of *Streptococcus pneumoniae*. *J Med Microbiol.* 2010;59(7):808-814. doi:10.1099/jmm.0.018663-0
79. Parker D, Soong G, Planet P, Brower J, Ratner AJ, Prince A. The NanA neuraminidase of *Streptococcus pneumoniae* is involved in biofilm formation. *Infect Immun.* 2009;77(9):3722-3730. doi:10.1128/IAI.00228-09
80. Long JP, Tong HH, DeMaria TF. Immunization with native or recombinant *Streptococcus pneumoniae* neuraminidase affords protection in the chinchilla otitis media model. *Infect Immun.* 2004;72(7):4309-4313. doi:10.1128/IAI.72.7.4309-4313.2004
81. Mitchell TJ, Alexander UE, Morgan PJ, Andrew PW. Molecular analysis of virulence factors of *Streptococcus pneumoniae*. *J Appl Microbiol Symp Suppl.* 1997. doi:10.1046/j.1365-2672.83.s1.7.x
82. Ramos-Sevillano E, Urzainqui A, Campuzano S, et al. Pleiotropic effects of cell wall amidase LytA on *Streptococcus pneumoniae* sensitivity to the host immune response. *Infect Immun.* 2015;83(2):591-603. doi:10.1128/IAI.02811-14
83. De Las Rivas B, García JL, López R, García P. Purification and polar localization of pneumococcal LytB, a putative endo- $\beta$ -N-acetylglucosaminidase: The chain-dispersing murein hydrolase. *J Bacteriol.* 2002;184(18):4988-5000. doi:10.1128/JB.184.18.4988-5000.2002
84. Eldholm V, Johnsborg O, Haugen K, Ohnstad HS, Havastein LS. Fratricide in *Streptococcus pneumoniae*: Contributions and role of the cell wall hydrolases CbpD, LytA and LytC. *Microbiology.* 2009. doi:10.1099/mic.0.026328-0



85. Domenech M, García E, Moscoso M. Biofilm formation in *Streptococcus pneumoniae*. *Microb Biotechnol*. 2012. doi:10.1111/j.1751-7915.2011.00294.x
86. Brooks-Walter A, Briles DE, Hollingshead SK. The *pspC* gene of *Streptococcus pneumoniae* encodes a polymorphic protein, PspC, which elicits cross-reactive antibodies to PspA and provides immunity to pneumococcal bacteremia. *Infect Immun*. 1999;67(12):6533-6542.
87. Variation A, Promotes P, Escape I, Immunity V. Antigenic Variation in *Streptococcus pneumoniae* PspC Promotes Immune Escape in the Presence of Variant-Specific Immunity. 2018;9(2):1-12.
88. Quin LR, Moore QC, McDaniel LS. Pneumolysin, PspA, and PspC contribute to pneumococcal evasion of early innate immune responses during bacteremia in mice. *Infect Immun*. 2007;75(4):2067-2070. doi:10.1128/IAI.01727-06
89. Mirza S, Hollingshead SK, Benjamin WH, Briles DE. PspA protects *Streptococcus pneumoniae* from killing by apolactoferrin, and antibody to PspA enhances killing of pneumococci by apolactoferrin. *Infect Immun*. 2004;72(12):7379. doi:10.1128/IAI.72.12.7379.2004
90. Martinez PJ, Farhan A, Mustafa M, et al. PspA facilitates evasion of pneumococci from bactericidal activity of neutrophil extracellular traps (NETs). *Microb Pathog*. 2019;136. doi:10.1016/j.micpath.2019.103653
91. Berry AM, Paton JC. Additive attenuation of virulence of *Streptococcus pneumoniae* by mutation of the genes encoding pneumolysin and other putative pneumococcal virulence proteins. *Infect Immun*. 2000;68(1):133-140. doi:10.1128/IAI.68.1.133-140.2000
92. Oliveira MLS, Miyaji EN, Ferreira DM, et al. Combination of pneumococcal surface protein a (PspA) with whole cell pertussis vaccine increases protection against pneumococcal challenge in mice. *PLoS One*. 2010. doi:10.1371/journal.pone.0010863
93. Coral MCV, Briles DE, Hollingshead SK, Fonseca N, Castañeda E, Di Fabio JL. Pneumococcal Surface Protein A of Invasive *Streptococcus pneumoniae* Isolates from Colombian Children . *Emerg Infect Dis*. 2012;7(5):832-836. doi:10.3201/eid0705.017510
94. Roche H, Briles DE, Hakansson A, Mirza S, Brooks-Walter A, McDaniel LS. Characterization of Binding of Human Lactoferrin to Pneumococcal Surface Protein A. *Infect Immun*. 2002;69(5):3372-3381. doi:10.1128/iai.69.5.3372-3381.2001
95. Ren B, Szalai AJ, Hollingshead SK, Briles DE. Effects of PspA and Antibodies to PspA on Activation and Deposition of Complement on the Pneumococcal Surface. *Infect Immun*. 2004;72(1):114-122. doi:10.1128/IAI.72.1.114-122.2004

96. Swiatlo E, Brooks-Walter A, Briles DE, McDaniel LS. Oligonucleotides identify conserved and variable regions of *pspA* and *pspA*-like sequences of *Streptococcus pneumoniae*. *Gene*. 1997. doi:10.1016/S0378-1119(96)00823-2
97. Kohler S, Hallström T, Singh B, et al. Binding of vitronectin and factor H to hic contributes to immune evasion of *Streptococcus pneumoniae* serotype 3. *Thromb Haemost*. 2015;113(1):125-142. doi:10.1160/TH14-06-0561
98. Meinel C, Spartà G, Dahse HM, et al. *Streptococcus pneumoniae* from Patients with Hemolytic Uremic Syndrome Binds Human Plasminogen via the Surface Protein PspC and Uses Plasmin to Damage Human Endothelial Cells. *J Infect Dis*. 2018;217(3):358-370. doi:10.1093/infdis/jix305
99. Moreno AT, Oliveira MLS, Ho PL, et al. Cross-reactivity of antipneumococcal surface protein C (PspC) antibodies with different strains and evaluation of inhibition of human complement factor H and secretory IgA binding via PspC. *Clin Vaccine Immunol*. 2012. doi:10.1128/CVI.05706-11
100. Voss S, Hallström T, Saleh M, et al. The choline-binding Protein PspC of *Streptococcus pneumoniae* interacts with the C-terminal heparin-binding domain of vitronectin. *J Biol Chem*. 2013;288(22):15614-15627. doi:10.1074/jbc.M112.443507
101. Smith BL, Hostetter MK. C3 As substrate for adhesion of *Streptococcus pneumoniae*. *J Infect Dis*. 2000;182(2):497-508. doi:10.1086/315722
102. Madsen M, Lebenthal Y, Cheng Q, Smith BL, Hostetter MK. A pneumococcal protein that elicits interleukin-8 from pulmonary epithelial cells. *J Infect Dis*. 2000;181(4):1330-1336. doi:10.1086/315388
103. Meinel C, Spartà G, Dahse HM, et al. *Streptococcus pneumoniae* from Patients with Hemolytic Uremic Syndrome Binds Human Plasminogen via the Surface Protein PspC and Uses Plasmin to Damage Human Endothelial Cells. *J Infect Dis*. 2018;217(3):358-370. doi:10.1093/infdis/jix305
104. Du S, Vilhena C, King SJ, et al. Molecular analyses identifies new domains and structural differences among *Streptococcus pneumoniae* immune evasion proteins PspC and Hic. *Sci Rep*. 2021;1-15. doi:10.1038/s41598-020-79362-3
105. Janulczyk R, Iannelli F, Sjöholm AG, Pozzi G, Björck L. Hic, a novel surface protein of *Streptococcus pneumoniae* that interferes with complement function. *J Biol Chem*. 2000;275(47):37257-37263. doi:10.1074/jbc.M004572200
106. Dasaraju P V, Liu C. Infections of the Respiratory System. *Med Microbiol*. 1996.
107. van der Poll T, Opal SM. Pathogenesis, treatment, and prevention of pneumococcal pneumonia. *Lancet*. 2009. doi:10.1016/S0140-6736(09)61114-4

108. Yau B, Hunt NH, Mitchell AJ, Too LK. Blood-brain barrier pathology and CNS outcomes in *Streptococcus pneumoniae* Meningitis. *Int J Mol Sci.* 2018;19(11). doi:10.3390/ijms19113555
109. Hirt-Minkowski P, Dickenmann M, Schifferli JA. Atypical hemolytic uremic syndrome: Update on the complement system and what is new. *Nephron - Clin Pract.* 2010;114(4):219-235. doi:10.1159/000276545
110. Zipfel PF, Mache C, Müller D, et al. DEAP-HUS: Deficiency of CFHR plasma proteins and autoantibody-positive form of hemolytic uremic syndrome. *Pediatr Nephrol.* 2010;25(10):2009-2019. doi:10.1007/s00467-010-1446-9
111. Veessenmeyer AF, Edmonson MB. Trends in US Hospital Stays for *Streptococcus pneumoniae*-associated Hemolytic Uremic Syndrome. *Pediatr Infect Dis J.* 2013;32(7).  
[https://journals.lww.com/pidj/Fulltext/2013/07000/Trends\\_in\\_US\\_Hospital\\_Stays\\_for\\_Streptococcus.6.aspx](https://journals.lww.com/pidj/Fulltext/2013/07000/Trends_in_US_Hospital_Stays_for_Streptococcus.6.aspx).
112. Bender JM, Ampofo K, Byington CL, et al. Epidemiology of *Streptococcus pneumoniae*-induced hemolytic uremic syndrome in Utah children. *Pediatr Infect Dis J.* 2010;29(8):712-716. doi:10.1097/INF.0b013e3181db03a7
113. Spinale JM, Ruebner RL, Kaplan BS, Copelovitch L. Update on *Streptococcus pneumoniae* associated hemolytic uremic syndrome. *Curr Opin Pediatr.* 2013;25(2).  
[https://journals.lww.com/co-pediatrics/Fulltext/2013/04000/Update\\_on\\_Streptococcus\\_pneumoniae\\_associated.10.aspx](https://journals.lww.com/co-pediatrics/Fulltext/2013/04000/Update_on_Streptococcus_pneumoniae_associated.10.aspx).
114. Gómez Delgado I, Corvillo F, Nozal P, et al. Complement Genetic Variants and FH Desialylation in *S. pneumoniae*-Haemolytic Uraemic Syndrome. *Front Immunol.* 2021;12(March). doi:10.3389/fimmu.2021.641656
115. Kim SH, Kim SY. A Case of *Streptococcus pneumoniae* associated Hemolytic Uremic Syndrome with DIC. *Child Kidney Dis.* 2015;19(1):48-52. doi:10.3339/chikd.2015.19.1.48
116. Seger R, Joller P, Baerlocher K, Hitzig WH. Neuraminidase-producing pneumococci in the pathogenesis of hemolytic-uremic syndrome. *Schweiz Med Wochenschr.* 1980;110(40):1454-1456.
117. Crain MJ, Waltman WD, Turner JS, et al. Pneumococcal surface protein A (PspA) is serologically highly variable and is expressed by all clinically important capsular serotypes of *Streptococcus pneumoniae*. *Infect Immun.* 1990;58(10):3293-3299. doi:10.1128/iai.58.10.3293-3299.1990

118. van der Maten E, van den Broek B, de Jonge MI, et al. *Streptococcus pneumoniae* PspC subgroup prevalence in invasive disease and differences in contribution to complement evasion. *Infect Immun*. 2018;86(4):1-10. doi:10.1128/IAI.00010-18
119. Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA. Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol*. 2004;14(2):208-216. doi:10.1016/j.sbi.2004.03.011
120. Engholm DH, Kilian M, Goodsell DS, Andersen ES, Kjærgaard RS. A visual review of the human pathogen *Streptococcus pneumoniae*. *FEMS Microbiol Rev*. 2017;41(6):854-879. doi:10.1093/femsre/fux037
121. Jedrzejewski MJ. Pneumococcal Virulence Factors: Structure and Function. *Microbiol Mol Biol Rev*. 2001;65(2):187-207. doi:10.1128/mmr.65.2.187-207.2001
122. Hakenbeck R, Madhour A, Denapaite D, Brückner R. Versatility of choline metabolism and choline-binding proteins in *Streptococcus pneumoniae* and commensal streptococci. *FEMS Microbiol Rev*. 2009;33(3):572-586. doi:10.1111/j.1574-6976.2009.00172.x
123. Hegyi H, Gerstein M. Annotation Transfer for Genomics: Measuring Functional Divergence in Multi-Domain Proteins. *Genome Res*. 2001;11(3):313-316. doi:10.1101/gr
124. Xu Q, Zhang JW, Chen Y, Li Q, Jiang YL. Crystal structure of the choline-binding protein CbpJ from *Streptococcus pneumoniae*. *Biochem Biophys Res Commun*. 2019;514(4):1192-1197. doi:10.1016/j.bbrc.2019.05.053
125. Zhang Z, Zhang X, Zhang L, et al. A choline binding polypeptide of LytA inhibits the growth of *Streptococcus pneumoniae* by binding to choline in the cell wall. *J Antibiot (Tokyo)*. 2018;71(12):1025-1030. doi:10.1038/s41429-018-0091-6
126. Galán-Bartual S, Pérez-Dorado I, García P, Hermoso JA. Structure and Function of Choline-Binding Proteins. *Streptococcus pneumoniae Mol Mech Host-Pathogen Interact*. 2015:207-230. doi:10.1016/B978-0-12-410530-0.00011-9
127. Domenech M, García E, Moscoso M. In vitro destruction of *Streptococcus pneumoniae* biofilms with bacterial and phage peptidoglycan hydrolases. *Antimicrob Agents Chemother*. 2011;55(9):4144-4148. doi:10.1128/AAC.00492-11
128. Woodfin A, Voisin MB, Nourshargh S. PECAM-1: A multi-functional molecule in inflammation and vascular biology. *Arterioscler Thromb Vasc Biol*. 2007;27(12):2514-2523. doi:10.1161/ATVBAHA.107.151456
129. Zhu M, Oishi K, Lee SC, Paul H. Vascular Endothelial Platelet Endothelial Cell Adhesion Molecule-1 (PECAM-1) Expression Is Decreased by TNF- $\alpha$  and IFN- $\gamma$ . 1996;156:1221-1228.

- 
130. McDevitt E, Khan F, Scasny A, et al. Hydrogen Peroxide Production by *Streptococcus pneumoniae* Results in Alpha-hemolysis by Oxidation of Oxy-hemoglobin to Met-hemoglobin. *mSphere*. 2020;5(6):1-7. doi:10.1128/msphere.01117-20
131. Rai P, Parrish M, Tay IJJ, et al. *Streptococcus pneumoniae* secretes hydrogen peroxide leading to DNA damage and apoptosis in lung cells. *Proc Natl Acad Sci U S A*. 2015;112(26):E3421-E3430. doi:10.1073/pnas.1424144112
132. Yen C-C, Shen C-J, Hsu W-H, et al. Lactoferrin: an iron-binding antimicrobial protein against *Escherichia coli* infection. *Biometals an Int J role Met ions Biol Biochem Med*. 2011;24(4):585-594. doi:10.1007/s10534-011-9423-8
133. Samuelsen Ø, Haukland HH, Ulvatne H, Vorland LH. Anti-complement effects of lactoferrin-derived peptides. *FEMS Immunol Med Microbiol*. 2004;41(2):141-148. doi:10.1016/j.femsim.2004.02.006
134. Józsi M, Zipfel PF. Factor H family proteins and human diseases. *Trends Immunol*. 2008;29(8):380-387. doi:10.1016/j.it.2008.04.008
135. Loeven MA, Maciej-Hulme ML, Yanginlar C, et al. Selective Binding of Heparin/Heparan Sulfate Oligosaccharides to Factor H and Factor H-Related Proteins: Therapeutic Potential for C3 Glomerulopathies. *Front Immunol*. 2021;12(August):1-11. doi:10.3389/fimmu.2021.676662
136. Medjeral-Thomas NR, Moffitt H, Lomax-Browne HJ, et al. Glomerular Complement Factor H-Related Protein 5 (FHR5) Is Highly Prevalent in C3 Glomerulopathy and Associated With Renal Impairment. *Kidney Int Reports*. 2019;4(10):1387-1400. doi:10.1016/j.ekir.2019.06.008
137. Szilágyi Á, Kiss N, Bereczki C, et al. The role of complement in *Streptococcus pneumoniae*-associated haemolytic uraemic syndrome. *Nephrol Dial Transplant*. 2013;28(9):2237-2245. doi:10.1093/ndt/gft198
138. Gilbert RD, Nagra A, Haq MR. Does dysregulated complement activation contribute to haemolytic uraemic syndrome secondary to *Streptococcus pneumoniae*? *Med Hypotheses*. 2013;81(3):400-403. doi:10.1016/j.mehy.2013.05.030

**8 Eigenständigkeitserklärung**

1 Die geltende Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena ist mir bekannt.

2 Die vorliegende Dissertation wurde von mir selbst angefertigt und alle benutzten Hilfsmittel, persönlichen Mitteilungen und Quellen sind in dieser Arbeit angegeben.

3 Alle Personen, die mich bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts unterstützt haben, habe ich benannt.

4 Die Hilfe eines Promotionsberaters habe ich nicht in Anspruch genommen.

5 Dritte Personen haben weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

6 Diese Arbeit wurde bisher weder an einer anderen Hochschule als Dissertation noch als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht.

Jena, 18.12. 2021

---

Shanshan Du

## 9 Overview of publication list, oral, poster presentations

### List of Publications

**Du, S.**, Vilhena, C., Sahagun, A., King, S., Skerka, C., Hammerschmidt, S., Zipfel, P., Molecular Analyses of *Streptococcus pneumoniae* Immune Evasion Proteins Identifies new Domains and Reveals Structural Differences between PspC and HIC variants. *Sci Rep*, 11, 2021, 1701.

Vilhena, C., **Du, S.**, Battista, M., Westermann, M., Hammerschmidt, S., Zipfel, P. Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage (frontiers in immunology in revision).

**Du, S.**, Vilhena, C., Fuest, D., von der Heide, M., Hammerschmidt, S., Skerka, C., Zipfel, P., Modular structure of *S. pneumoniae* surface protein A: high level of domain-based sequence diversity may qualify for molecular strain typing (Manuscript ready to submit)

**Du, S.**, Zhang, Y., Wang, Q., Protein post translational modifications involved in DNA damage response pathways; *Chemistry of life* (2017)

### Oral and poster presentations

Title: ***Streptococcus pneumoniae* surface protein C acts as guards against host complement.** Joint Meeting ILRS + RTG Greifswald in 2019 Wittenberg, oral presentation.

Title: **PspC from *Streptococcus pneumoniae*: a central and highly variable immune evasion protein.** 11th ILRS Symposium in Jena, 2018, poster presentation.

Title: ***Streptococcus pneumoniae* diverse surface proteins act as guards against complement system.** 14th European Meeting on the Molecular biology of the pneumococcus in Greifswald, 2019, poster presentation.

Title: ***Streptococcus pneumoniae* surface protein C acts as guard against host complement.** 6th Joint Conference of the DGHM & VAAM 2020 in Leipzig, poster presentation.

Title: **Interplay between complement proteins and PspA and PspC contribute to immune evasion of *S. pneumoniae* in hemolytic uremic syndrome.** 28th International Complement Virtual Workshop, ICW2021, poster presentation.

**10 Curriculum Vitae****Name:** Shanshan Du (杜珊珊)

Birth date and place: May 9th, 1990. Henan- China.

**Education:**

Leibniz Institute for Natural product and infection biology-HKI	Jena, Germany
Major in immunobiology, PhD, expected in June 2022	04/2018 - Present
Beijing Institute of Genome Research, Chinese Academy of Sciences	Beijing, China
Major in Proteomics, Master of Bioengineering	09/2014 - 12/2017
Shangqiu Normal University	Shangqiu, China
Major in Bioengineering, Bachelor of Bioengineering GPA: 3.94/4.0	09/2010 - 06/2014

**Research experience:**

Research Program:

Immune evasion of *Streptococcus pneumoniae* in hemolytic uremic syndrome  
(independent Research)Integrative action of cell-wall binding proteins of *Streptococcus pneumoniae* on the  
interaction with the host's immune system (collaborate Research)Quantitative assessment of protein methylation in nuclear during DNA repair (independent  
Research)

Application of THz on protein identification (collaborate Research)

**Honors:**

DAAD scholarship	2018-2022
"Merit-student" honor	2016-2017
National scholarship	2012-2013
"merit-student" honor	2010-2011
Two consecutive years of Shangqiu University Scholarship	2010-2012



## 11 Acknowledgments

This project would not have been possible without the support of many people. To begin with, I would like to acknowledge the great opportunity I had to perform my Ph.D. studies in the Infection Biology department of Hans-knöll-institute, under the supervision of Prof. Dr. Peter F. Zipfel. He has been an ideal teacher, mentor, and supervisor, offering advice and encouragement with a perfect blend of insight and humor. I sincerely give thanks to my supervisor Prof. Dr. Peter F. Zipfel for all his support and friendship during the last four years. His kindness, guidance and support make my life and work easier in Germany.

Thanks to the DAAD for awarding me four years scholarship and a study completion grant, providing me with the financial means to complete this project and offering me a platform to learn and make new friends outside of work.

Many thanks to my colleagues and my friends, Dr. Cludiá Vilhena, Svitlana Babych, Dr. Hans-Martin Dahse and Dr. Luce Perie, Monika von der Heide, Andrés González Delgado for their help in my work and life. I take this opportunity to express gratitude to one and all, who directly or indirectly, have lent their hand in this venture. We had the opportunity to work together for four years which deserve to be remembered forever. They are inspiring colleagues, blazing a trail I followed in writing my thesis. With them, my research in Germany can be both productive and enjoyable.

Thank you to Prof. Dr. Sven Hammerschmidt for offering me the *S. pneumoniae* strains and antibodies used in my experiment. Thank you to Prof. Dr. Sven Hammerschmidt and Prof. Dr. Christine Skerka for your discussion and detailed feedback about my work.

Thank you to our secretaries Heike Gäbler, Ina Löschmann, and ILRS coordinator Dr. Christine Vogler for their assistance in my work and guidance in life in Jena.

I would also like to thank my best friend Chujun Duan I met first in Jena. You accompanied me through my hardest days in Germany. It is important to strike a balance with life outside the dark depths of the lab. Thank you to my friend Lu Zhang, Lucky to make a friend who can always support each other. I will always cherish our friendship.

Most importantly, I am grateful for my family's unconditional, unequivocal, and loving support. My accomplishments and success are because they believed in me and they continuously encouraged me. I owe my deepest gratitude to my parents and my sisters.

Finally, Song Bai who is my love, you were always there, wherever and whenever that was each midnight draft, each early morning rant, each time I cried, screamed and missed home. The love from you and our family makes me strong enough to face life's challenges.

## 12 Appendix

### FORMULAR 1<sup>1</sup>

#### Manuskript Nr. 1

#### Titel des Manuskriptes:

Molecular analyses identifies new domains and structural differences among *Streptococcus pneumoniae* immune evasion proteins PspC and Hic

**Autoren:** Shanshan Du, Cludiá Vilhena, Samantha J. King, Alfredo Sahagún-Ruiz, Sven Hammerschmidt, Christine Skerka, Peter F. Zipfel.

**Bibliographische Informationen:** Scientific reports. 2021 Jan 11(1071):1-15.

#### Der Kandidat / Die Kandidatin ist (bitte ankreuzen)

Erstautor/-in,  Ko-Erstautor/-in,  Korresp. Autor/-in,  Koautor/-in.

**Status :** Published

#### Anteile (in %) der Autoren / der Autorinnen an der Publikation (anzugeben ab 20%)

Autor/-in	Konzeptionell	Datenanalyse	Experimentell	Verfassen des Manuskriptes	Bereitstellung von Material
Shanshan Du	20%	80%	100%	20%	
Cludiá Vilhena	20%				
Samantha J. King					20%
Alfredo Sahagún-Ruiz	20%				
Sven Hammerschmidt					20%
Christine Skerka	20%				
Peter F. Zipfel	20%	20%		80%	60%

\_\_\_\_\_  
Unterschrift Kandidat/-in

\_\_\_\_\_  
Unterschrift Betreuer/-in (Mitglied der Fakultät)

<sup>1</sup> Die Unterschriften müssen nur im separat im Dekanat einzureichenden ausgefüllten Formular im Original vorliegen. In der in die Dissertation eingebundenen Fassung dürfen die Unterschriften und Unterschriftenfelder fehlen.

**FORMULAR 2<sup>2</sup>**

**Manuskript Nr. 1**

**Kurzreferenz** [Du,S. et al (2021),Sci Rep.]

**Beitrag des Doktoranden / der Doktorandin**

Beitrag des Doktoranden / der Doktorandin zu Abbildungen, die experimentelle Daten wiedergeben (nur für Originalartikel):

<b>Abbildung(en) # <u>1, 2,</u> <u>3, 4, 5</u> *</b>	<input type="checkbox"/>	100% (die in dieser Abbildung wiedergegebenen Daten entstammen vollständig experimentellen Arbeiten, die der Kandidat/die Kandidatin durchgeführt hat)
	<input type="checkbox"/>	0% (die in dieser Abbildung wiedergegebenen Daten basieren ausschließlich auf Arbeiten anderer Koautoren)
	<input type="checkbox"/>	Etwaiger Beitrag des Doktoranden / der Doktorandin zur Abbildung: _____% Kurzbeschreibung des Beitrages: (z. B. „Abbildungsteile a, d und f“ oder „Auswertung der Daten“ etc)

*\*Kann sich auf mehrere Abb. beziehen, wenn die Antwort dieselbe ist*

(Anfügen weiterer Tabellenkästen je nach Zahl der Abbildungen)

\_\_\_\_\_  
Unterschrift Kandidat/-in

\_\_\_\_\_  
Unterschrift Betreuer/-in (Mitglied der Fakultät)

<sup>2</sup> Die Unterschriften müssen nur im separat im Dekanat einzureichenden ausgefüllten Formular im Original vorliegen. In der in die Dissertation eingebundenen Fassung dürfen die Unterschriften und Unterschriftenfelder fehlen.

**FORMULAR 1<sup>3</sup>****Manuskript Nr. 2**

**Titel des Manuskriptes:** Modular structure of *S. pneumoniae* surface protein A: high level of domain-based sequence diversity may qualify for molecular strain typing

**Autoren:** Shanshan Du, Cludiá Vilhena, Dorit Fuest, Monika von der Heide, Sven Hammerschmidt, Christine Skerka, Peter F. Zipfel.

**Bibliographische Informationen** (falls publiziert oder zur Publikation angenommen: Zitat):

**Der Kandidat / Die Kandidatin ist** (bitte ankreuzen)

Erstautor/-in,  Ko-Erstautor/-in,  Korresp. Autor/-in,  Koautor/-in.

**Status:** in Vorbereitung

**Anteile (in %) der Autoren / der Autorinnen an der Publikation** (anzugeben ab 20%)

Autor/-in	Konzeptionell	Datenanalyse	Experimentell	Verfassen des Manuskriptes	Bereitstellung von Material
Shanshan Du	30%	80%	80%	15%	
Cludiá Vilhena	20%			3%	
Dorit Fuest		10%			20%
Monika von der Heide			20%		20%
Sven Hammerschmidt	20%			3%	
Christine Skerka				3%	
Peter F. Zipfel	30%	10%		76%	60%

\_\_\_\_\_  
Unterschrift Kandidat/-in

\_\_\_\_\_  
Unterschrift Betreuer/-in (Mitglied der Fakultät)

<sup>3</sup> Die Unterschriften müssen nur im separat im Dekanat einzureichenden ausgefüllten Formular im Original vorliegen. In der in die Dissertation eingebundenen Fassung dürfen die Unterschriften und Unterschriftenfelder fehlen.

**FORMULAR 2<sup>4</sup>**

**Manuskript Nr. 2**

**Kurzreferenz [ ]**

**Beitrag des Doktoranden / der Doktorandin**

Beitrag des Doktoranden / der Doktorandin zu Abbildungen, die experimentelle Daten wiedergeben (nur für Originalartikel):

<b>Abbildung(en) # <u>1, 2,</u> <u>3, 4, 5, 6*</u></b>	<input type="checkbox"/>	100% (die in dieser Abbildung wiedergegebenen Daten entstammen vollständig experimentellen Arbeiten, die der Kandidat/die Kandidatin durchgeführt hat)
	<input type="checkbox"/>	0% (die in dieser Abbildung wiedergegebenen Daten basieren ausschließlich auf Arbeiten anderer Koautoren)
	<input type="checkbox"/>	Etwaiger Beitrag des Doktoranden / der Doktorandin zur Abbildung: _____% Kurzbeschreibung des Beitrages: (z. B. „Abbildungsteile a, d und f“ oder „Auswertung der Daten“ etc)

*\*Kann sich auf mehrere Abb. beziehen, wenn die Antwort dieselbe ist*

(Anfügen weiterer Tabellenkästen je nach Zahl der Abbildungen)

\_\_\_\_\_  
Unterschrift Kandidat/-in

\_\_\_\_\_  
Unterschrift Betreuer/-in (Mitglied der Fakultät)

<sup>4</sup> Die Unterschriften müssen nur im separat im Dekanat einzureichenden ausgefüllten Formular im Original vorliegen. In der in die Dissertation eingebundenen Fassung dürfen die Unterschriften und Unterschriftenfelder fehlen.

**FORMULAR 1<sup>5</sup>****Manuskript Nr. 3**

**Titel des Manuskriptes:** Choline-binding proteins of *Streptococcus pneumoniae* and their role on host cellular adhesion and damage

**Autoren:** Cláudia Vilhena, Shanshan Du, Miriana Battista, Martin Westermann, Thomas P. Kohler, Sven Hammerschmidt, Peter F. Zipfel.

**Bibliographische Informationen** (falls publiziert oder zur Publikation angenommen: Zitat):

**Der Kandidat / Die Kandidatin ist** (bitte ankreuzen)

Erstautor/-in,  Ko-Erstautor/-in,  Korresp. Autor/-in,  Koautor/-in.

**Status:** zur Publikation eingereicht

**Anteile (in %) der Autoren / der Autorinnen an der Publikation** (anzugeben ab 20%)

Autor/-in	Konzeptionell	Datenanalyse	Experimentell	Verfassen des Manuskriptes	Bereitstellung von Material
Cláudia Vilhena	30%	80%	40%	60%	
Shanshan Du	20%	20%	15%	10%	
Miriana Battista			35%	5%	
Martin Westermann			10%		
Thomas P. Kohler					30%
Sven Hammerschmidt	20%			5%	20%
Peter F. Zipfel	30%			20%	50%

\_\_\_\_\_  
Unterschrift Kandidat/-in

\_\_\_\_\_  
Unterschrift Betreuer/-in (Mitglied der Fakultät)

<sup>5</sup> Die Unterschriften müssen nur im separat im Dekanat einzureichenden ausgefüllten Formular im Original vorliegen. In der in die Dissertation eingebundenen Fassung dürfen die Unterschriften und Unterschriftenfelder fehlen.

**FORMULAR 2<sup>6</sup>**

**Manuskript Nr. 3**

**Kurzreferenz [ ]**

**Beitrag des Doktoranden / der Doktorandin**

Beitrag des Doktoranden / der Doktorandin zu Abbildungen, die experimentelle Daten wiedergeben (nur für Originalartikel):

<b>Abbildung(en) #</b> <u>1</u> *	<input type="checkbox"/> 100% (die in dieser Abbildung wiedergegebenen Daten entstammen vollständig experimentellen Arbeiten, die der Kandidat/die Kandidatin durchgeführt hat)
	<input type="checkbox"/> 0% (die in dieser Abbildung wiedergegebenen Daten basieren ausschließlich auf Arbeiten anderer Koautoren)
	<input type="checkbox"/> Etwaiger Beitrag des Doktoranden / der Doktorandin zur Abbildung: <u>80</u> % Kurzbeschreibung des Beitrages: (z. B. „Abbildungsteile a, d und f“ oder „Auswertung der Daten“ etc)

*\*Kann sich auf mehrere Abb. beziehen, wenn die Antwort dieselbe ist*

(Anfügen weiterer Tabellenkästen je nach Zahl der Abbildungen)

\_\_\_\_\_  
Unterschrift Kandidat/-in

\_\_\_\_\_  
Unterschrift Betreuer/-in (Mitglied der Fakultät)

<sup>6</sup> Die Unterschriften müssen nur im separat im Dekanat einzureichenden ausgefüllten Formular im Original vorliegen. In der in die Dissertation eingebundenen Fassung dürfen die Unterschriften und Unterschriftenfelder fehlen.