<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Structure of the competence pilus major pilin ComGC in Streptococcus pneumoniae</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Muschiol, Sandra; Erlendsson, Simon; Aschtgen, Marie-Stephanie; Oliveira, Vitor; Schmieder, Peter; De Lichtenberg, Casper;teilum, Kaare; Boesen, Thomas; Akbey, Umit; Henriques-Normark, Birgitta</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2017</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10220/44346">http://hdl.handle.net/10220/44346</a></td>
</tr>
</tbody>
</table>

© 2017 American Society for Biochemistry and Molecular Biology (ASBMB). This paper was published in Journal of Biological Chemistry and is made available as an electronic reprint (preprint) with permission of American Society for Biochemistry and Molecular Biology (ASBMB). The published version is available at: [http://dx.doi.org/10.1074/jbc.M117.787671]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.
Type IV pilus are important virulence factors on the surface of many pathogenic bacteria and have been implicated in a wide range of diverse functions, including attachment, twitching motility, biofilm formation, and horizontal gene transfer. The respiratory pathogen *Streptococcus pneumoniae* deploys type IV pilus to take up DNA during transformation. These “competence pili” are composed of the major pilin protein ComGC and exclusively assembled during bacterial competence, but their biogenesis remains unclear. Here, we report the high resolution NMR structure of N-terminal truncated ComGC revealing a highly flexible and structurally divergent type IV pilin. It consists of only three α-helical segments forming a well-defined electronegative cavity and confined electronegative and hydrophobic patches. The structure is particularly flexible between the first and second α-helix with the first helical part exhibiting slightly slower dynamics than the rest of the pilin, suggesting that the first helix is involved in forming the pilus structure core and that parts of helices two and three are primarily surface-exposed. Taken together, our results provide the first structure of a type IV pilin protein involved in the formation of competence-induced pili in Gram-positive bacteria and corroborate the remarkable structural diversity among type IV pilin proteins.
ATPase (ComGA), which powers pilus assembly (9), a mem-
brane-spanning protein (ComGB), and four minor pilins
(ComGD, -E, -F, and -G) whose functions remain elusive.

Herein, we characterize the pneumococcal major pilin
ComGC and its ability to assemble into type IV pili. We also
present the NMR structure of N-terminally truncated ComGC,
which exclusively consists of 3H9251-helical segments and a variable
C-terminal domain with no sequence similarity to previously
characterized type IV pilin proteins.

Results

ComGC is the major pilin in S. pneumoniae
competence-induced pili

Previously it was reported that S. pneumoniae produces type
IV pili composed of ComGC in S. pneumoniae strain R6 and the
clinical isolates G54 and CP strains (9). To detect competence-
induced pili in the S. pneumoniae TIGR4 (T4) background, we
have used the un-encapsulated T4 strain (T4R) deficient in the
rlrA operon (T4RΔrgA-srtD). The rlrA operon encodes an
adhesive pneumococcal pilus that is assembled by pilus-associ-
ated sortases (18, 19). By using this mutant strain, we were able
to rule out other pilus structures expressed on the bacteria. We
then looked at the formation of type IV pili in T4RΔrgA-srtD
cultures induced with the competence-stimulating peptide
(CSP)3 and control cultures without CSP addition. As shown in
Fig. 1A, a type IV pilus could be visualized by transmission
electron microscopy in negatively stained S. pneumoniae

Figure 1. S. pneumoniae assembles competence type IV pili composed of ComGC. A, electron micrograph of negatively stained competent S. pneumoniae
T4RΔrgA-srtD induced with CSP. Black arrows indicate the pilus. B, transformation frequency of S. pneumoniae T4R and R6 strain. The error bars represent
standard deviation (S.D.) of a minimum of three independent experiments. C and D, immunogold electron microscopy to visualize pili on competent S. pneu-
moniae R6 using primary antibody specific to ComGC and secondary antibody conjugated to 6-nm gold particles. D, enlargement of the immunogold-labeled
pilus. Black arrows indicate the pilus. E, electron micrograph of a competence pilus in strain R6 stained with anti-ComGC antibody and protein A coupled to
10-nm gold particles. F, two-dimensional PAGE to assess multimerization of mature ComGC. A pilus preparation of T4 WT or ΔC strain was run on a 12% native
gel (1D, first dimension). A piece of gel corresponding to one lane of the gel was cut and placed horizontally on top of a second SDS-PAGE (2D, second
dimension). After migration, gels were immunoblotted with anti-ComGC antibody. Arrows indicate ComGC and protein multimers.

3 The abbreviations used are: CSP, competence stimulating peptide; BACTH, bacteri
al adenylate cyclase two-hybrid; hetNOE, heteronuclear 1H-15N-
NOE; O/N, overnight; RDC, residual dipolar coupling; Ni-NTA, nickel-nitri-
lactric acid.
Solution structure of ComGC from S. pneumoniae

T4RΔrgA-srtD. When we compared electron micrographs of negatively stained competent T4RΔrgA-srtD to R6, pilis were less frequently observed in the T4R background, which likely provides an explanation as to why the transformation frequency is almost three orders of magnitude lower in T4R than in R6 (Fig. 1B).

For that reason, we decided to do immunogold labeling of ComGC in competent R6 bacteria and used primary polyclonal ComGC antibodies, raised against the purified protein or an anti-peptide antibody, and secondary antibody labeled with 6-nm gold particles. We frequently found gold particles labeling the entire type IV pilus suggesting that ComGC is the major pilin protein (Fig. 1, C and D). We also stained competent R6 bacteria with primary polyclonal ComGC antibody followed by incubation with protein A coupled to 10-nm gold particles. In this way the pilus is less frequently labeled with gold particles; however, the underlying pilus filament is clearly visible (Fig. 1E).

To further study pilus polymerization also in an encapsulated T4 background, we analyzed pilis preparations by two-dimensional (2D) PAGE. First, pilis preparations of wild-type T4 (WT) or a comGC knock-out mutant (ΔC) were run on a 12% native gel, which resulted in a local concentration of ComGC on the top of the gel. One lane of each sample was then cut and placed horizontally on SDS-PAGE. After migration, the gel was immunobotted and probed with ComGC antibodies. When entering the SDS-containing gel, high-molecular-weight structures will be denatured, which is why we observe monomeric ComGC (x1); when only partially denatured, distinct ComGC building blocks (x2 and x3) can be detected (Fig. 1F) suggesting that ComGC also forms the pilus backbone in S. pneumoniae T4.

Structural features of pneumococcal ComGC filaments

To further assess the structural features of native competence pilis, micrographs of uranyl acetate-stained samples of S. pneumoniae strain R6 were inspected. The filaments showed a pronounced degree of flexibility and only short straight regions were observed (Fig. 2A). A suitable amount of straight filament regions were identified by inspecting a large number of micrographs and were used for analysis of potential helical symmetry in the filaments. No helical diffraction pattern was evident from raw micrographs, but after averaging filament segments layer lines became visible in the power spectra. Based on a class average power spectrum the first prominent layer line was observed at a distance of 0.0252 Å–1 from the equator corresponding to a helical pitch of ~40 Å (Fig. 2B). This was in accordance with a helix normal profile plot in real space showing similar distances between peaks along the helical axis (Fig. 2C). The average diameter of the filaments was derived from helix width profiles of 16 class averages and found to be 64 ± 1.6 Å (Fig. 2D). A representative class average can be seen in Fig. 2E.

ComGC processing and dimerization in the membrane

One characteristic of proteins belonging to the type IV pilin family is the presence of a well-conserved prepilin cleavage motif Gly-Phe-Xaa-Xaa-Xaa-Glu (20). Pneumococcal ComGC is synthesized with a 15-residue leader sequence and shares a highly conserved PilD cleavage site with other known major pilins (Fig. 3A). It can also be processed in vitro by co-expressing full-length ComGC and PilD (Fig. 3B).

To test whether two full-length membrane-embedded ComGC monomers can directly interact with each other, we used the bacterial adenylate cyclase two-hybrid (BACTH) system (21). Mature ComGC was fused to the C-terminal end of T25 and T18 fragments of Bordetella pertussis adenylate cyclase (CyaA), and lacZ expression was measured. Compared with the negative control (T25/T18), T25-comGC/T18-comGC showed a statistically significant increase in CyaA activity (Fig. 3C). Because the positive control, in which T25 and T18 are fused to the leucine zipper domain of GCN4 (21), showed much higher activity, we included another functional control protein, PulG. PulG is the major pilin protein of the type II secretion system in Klebsiella oxytoca and is known to form type IV-like pili when overproduced (22). The level of CyaA activation in T25-PulG/T18-PulG was similar to T25-comGC/T18-comGC, indicating efficient dimerization of ComGC in the membrane (Fig. 3C). We also tested a strain expressing T25-PulG/T18-ComGC, which showed very low CyaA activity similar to the negative control (T25/T18), suggesting that these two functional major pilins cannot interact in the membrane. To validate our interaction between two ComGC monomers, we also performed chemical cross-linking of Escherichia coli BTH101 expressing T25-comGC/T18-comGC and were able to detect ComGC dimerization by immunoblotting with ComGC antiserum (Fig. 3D).

Structure of soluble ComGC

ComGC has very little sequence similarity to other pilins of which the three-dimensional structure has been solved. Specifically, ComGC has only few hydrophobic amino acid residues in the C terminus, which in other pilins form the β-strand-rich head domain. Indeed, secondary structure predictions using Jpred4 (23) and Agadir (24) show that the ComGC sequence has several segments with high α-helix propensity and no β-strand propensity (supplemental Fig. S1). This suggests that the three-dimensional structure of pneumococcal ComGC may differ significantly from known structures of type IV pilins. To determine the structure of ComGC, we prepared a truncated construct lacking the predicted N-terminal transmembrane helical domain (ComGC*, see Fig. 5A). It was previously shown for the major pilin PAK in Pseudomonas aeruginosa that deletion of α1-N does not perturb the structural fold, with full-length and truncated protein essentially being identical (25). To determine the ComGC* structure, we used NMR spectroscopy. ComGC* provided well-resolved spectra and remained largely homogeneous and stable at 10 °C (Fig. 4), which enabled us to solve the atomic resolution structure of ComGC* in solution (Fig. 5B). A summary of the structural statistics and constraints is provided in Table 1. ComGC* consists of three flexible helical segments: α1-C, involving residues 54–69; a shorter α2 helix involving residues 75–81; and finally, a C-terminal α3 spanning residues 86–99. The first 14 N-terminal and 10 C-terminal residues remain unfolded in our in-solution structure, and only few inter-residual NOEs are observed in this part of the mole-
Solution structure of ComGC from S. pneumoniae

cule (supplemental Fig. S2). This observation is in good agreement with the secondary chemical shifts (supplemental Fig. S1A). α1-C seems to be only loosely attached to α2 and α3, and a general lack of long-range distance restraints between these two “domains” indicates that the structure is less constrained and rather flexible in this hinge. The relative orientation of the helices was restrained by measuring residual dipolar couplings (Fig. 5, B and C). The overall tertiary fold appears almost two-dimensional being ~60 Å tall in the vertical plane, ~50 Å wide in the horizontal plane, but only ~10 Å broad in the profile plane (Fig. 5D), which essentially caused all three helices to be mostly solvent-exposed, providing a very large solvent-accessible surface (~8000 Å²). Visualizing the electrostatic potential of the solvent-accessible surface revealed a well-defined electronegative cavity formed between the helices, and two highly electronegative areas (denoted δ~) in the top of α1-C and the opposite side of α2 (Fig. 5E). Similarly, visualizing the hydrophobicity of the surface also showed two well-defined patches. The first hydrophobic patch (ϕ1) is situated at the back of α1-C, opposite to the electropositive patch (ϕ1) and the other (ϕ2) in α2 (Fig. 5F).

ComGC* dynamics on different time scales might be important for pilus assembly

The structural flexibility hypothesized above led us to study the dynamics in greater detail, and we therefore measured the longitudinal (R₁) and the transverse (R₂) ¹⁵N relaxation rates in both (relative to the external magnetic field) as well as he-
teronuclear ¹H-¹⁵N-NOEs (hetNOEs), at two different field strengths (Fig. 6A). As expected, and in support of the ComGC structure, the unstructured region 40–52 displayed generally longer \( R_1 \) rates (>1.5 s⁻¹), short \( R_2 \) rates (<10 s⁻¹), and relatively lower hetNOE values (<0.5), compared with the structured regions. Unlike \( R_2 \), \( R_1 \) rates are strongly field-dependent, and therefore the relative \( R_1 \) difference is mostly visible at the higher field strength. The \( R_2 \) rates are generally high (20.3 s⁻¹ on average for residues 56–108) in all of the structured regions as would be expected. However, several residues in \( \alpha_1-C \) exhibit high \( R_2 \) rates, which could suggest conformational exchange with one or more additional states (Fig. 6B). To gain insight into site-specific internal motion, we used the measured \( R_1 \), \( R_2 \), and hetNOE values to calculate the reduced spectral density functions, \( J(0) \), \( J(\omega_n) \) and \( J(0.87\omega_n) \), reporting on dynamics on three different time scales (Fig. 6B). \( J(0) \) represents protein mobility in the nano-second time scale, thus low \( J(0) \) values normally indicate higher flexibility as observed for residues 40–52, showing also higher internal motion in the \( J(\omega_n) \) pico-second time scale supporting that this region remains largely unstructured in solution. The more structured regions, and especially the flexible hinges between the helical segments, displayed much higher \( J(0) \) values indicating that these regions have mobility on the nano-second time scale. Interestingly, Asp-48 just before \( \alpha_1-C \) appeared highly dynamic from the hetNOE experiment; however, on a time scale (lower nano-second) different from all other residues.

Also, the hydrophobic residues Leu-78 and the stretch from Ala-89 to Lys-93, located in the interface between \( \alpha_2 \) and \( \alpha_3 \) showed dynamics on a faster time scale than observed for other nearby residues.

**Sequence variation in ComGC**

A total of 14 polymorphic sites differing in the number of variations from the reference sequence TIGR4 were identified in 23 publicly available *S. pneumoniae* genomes suggesting that ComGC is well conserved. A phylogenetic tree of strains clustered according to their ComGC sequence and the corresponding multiple sequence alignment are shown in supplemental Fig. S3. The most divergent strains exhibit a sequence identity of 91%. All strains can be grouped into two main clusters. The strains belonging to cluster 1 are identical with exception of strain NT11058 that carries one additional variation (N107Y) than the other strains present in this group. The second cluster, containing our reference sequence, is more diverse and can be sub-grouped into five sub-clusters. The virulent strain D39 and the avirulent, un-encapsulated laboratory strain R6, a derivative of D39, are closely related to TIGR4 with only one sequence variation (N96H). The majority of polymorphisms are localized to the interface between helices \( \alpha_2 \) and \( \alpha_3 \) and only 1 out of 14 is localized to the \( \alpha_1 \)-helix (Fig. 7). This suggests that the hypothesized hydrophobic pilin–pilin interface involving the
transmembrane α1-N and α1-C helical domains has been largely conserved throughout evolution and that the α2-α3 head group has undergone significantly larger changes primarily affecting the exposed electrostatic regions depicted in Fig. 5E.

Discussion

The pneumococcal competence pilus was first visualized recently (9). It is morphologically similar to other type IV pili described displaying filament diameters between 6–9 nm (11). Competence pili in *S. pneumoniae* have a mean diameter of 64 Å, which is comparable with the 60 Å diameter of type IV pili in *Neisseria gonorrhoeae* (26). They are helical assemblies, and the observed pitch of ~40 Å is somewhat larger than the 37 Å pitch of *N. gonorrhoeae* pili, suggesting different assembly and stabilization strategies in ComGC competence pili. In comparison, the type IV pilus in *Thermus thermophilus*, ~3 nm in diameter, shows a helical pitch of 49 Å forming a less compact pilus than type IV pili in *N. gonorrhoeae* (27). The observed differences in pilus diameter and helical pitch are likely explained by structural features of the major pilin subunit, which can vary considerably in sequence and size among bacteria expressing type IV pili.

Pneumococcal ComGC shares many features of canonical type IV pilins. Full-length ComGC has a well-defined conserved prepilin cleavage motif, an invariant Glu residue at position 5 after the cleavage site, and it is processed by PilD. The structure of soluble ComGC, provided here, is the first example of a type IV pilin protein involved in the formation of competence-induced pili in Gram-positive bacteria and reveals new structural features. Similar to previously described type IV pilins, ComGC has a predicted extended N-terminal α-helix but differs otherwise significantly as follows: 1) ComGC is exclusively α-helical; 2) the head group is much smaller; 3) the α1-C helix is separated from the transmembrane helix by a flexible linker that is largely unfolded in solution; and 4) ComGC contains no cysteines. Overall, ComGC is shorter than other type IV pilins and highly dynamic in solution, which may be an important feature for pilus assembly and function.

An important question raised by this structure regards the stabilization of ComGC. Most type IV pilins in Gram-negative bacteria and the major pilin ComGC in *Bacillus subtilis* have two cysteine residues in the C-terminal part of the protein that are important for protein stability and polymerization (13, 28), but there is no disulfide bond to stabilize ComGC. The major pilin, PilA1, in the Gram-positive bacterium *Clostridium difficile* also lacks cysteines (29). However, it is structurally much more compact in its C terminus than pneumococcal ComGC. In ComGC only two α-helices (α2 and α3) are forming the head domain, and the absence of other stabilizing structural elements might explain the observed flexibility in this region. In fact, PilA of *Geobacter sulfurreducens* (only 66 amino acids) is essentially lacking any globular head domain, and the NMR structure also showed a highly dynamic C-terminal region (30).

The assembly of pilin monomers into a model of the fully formed pilus has primarily been based on negative staining and electron cryo-micrographs, where individual monomers are fixed in a favorable multimeric organization and fitted into the obtained electron density (26, 31). These structural models

---

**Figure 4. Assigned ¹H-¹⁵N HSQC of ComGC recorded at 283K.** The HSQC spectrum displays signals from all the backbone and side chain N-H correlations. The ¹H chemical shift is plotted along the x axis and the ¹⁵N chemical shifts along the y axis. The chemical shifts report on the local chemical environment, and thus very small changes in structure will cause changes in chemical shifts. We observed 65 well defined amide peaks in the ComGC HSQC of which we were able to assign 61 residue-specific resonances (black arrows). Unassigned resonances in the upper right corner are side-chain N-H correlations. Importantly, we do not see any clear indication of more than one conformational state as this would give rise to more chemical shifts in the spectrum.
Figure 5. NMR structure of soluble ComGC is composed of three flexible helical segments. A, schematic and secondary structure overview of the full-length ComGC protein. The leader peptide is colored in cyan; the membrane-spanning helix is in brown; soluble helices are in red, and flexible regions are in green. The ComGC^C^ construct used for structural determination by NMR is indicated. In the protein sequence Glu-5 is marked by ■ and the helix-breaking residue Pro-22 by ●. B, alignment of the 20 lowest energy structures returned from the structure calculations. Ensemble alignments for all three helices (left), α1-C only (middle), and α2-α3 only (right) are shown. The structure is particularly flexible between α1-C and α2-α3, which is illustrated by alignment of the α1-C only and α2 and α3. Root mean square deviation values are listed in supplemental Table S1. C, schematic of the ComGC^C^ NMR structure showing the three helical segments as well as the flexible regions. D, calculated solvent-accessible surface of ComGC^C^. E, APBS calculated electrostatics from ±2 kT/e display well-defined electropositive (+) and electronegative (−) solvent-accessible patches. F, solvent-accessible hydrophobic patches ϕ1 and ϕ2 colored using the Eisenberg hydrophobicity scale ranging from −2.5 to 1.5.
serve as important frameworks for understanding pilus dimensions, appearance, and surface, but as a consequence of the low structural similarity of ComGC, primarily in the head group, we were not able to reliably predict ComGC assembly. Our data suggest that soluble monomeric ComGC will not adopt secondary or tertiary folds similar to other type IV pilins, but we cannot rule out that additional conformational states (i.e. helical rearrangements) will be favored during assembly or in the mature pilus structure.

The structure of ComGC itself provides initial information on the assembly and function of competence pilin in *S. pneumoniae*. The electrostatic potential as well as the hydrophobicity of the accessible surface in ComGC reveals highly defined patches, which might restrict or guide pilus formation. Based on the hydrophobicity profile, we propose that α1-N and α1-C are involved in forming the core of the pilus structure, with parts of α2 and α3 being primarily surface-exposed. The residues Leu-78, Ile-84, and Tyr-92, involved in the hydrophobic patch, β2, formed between α2 and α3 seem functionally distinct and may contribute to pilin flexibility during pilus assembly. Additionally, they could provide better resistance to shear forces in the environment by increasing the flexibility of the assembled pilus. It is also notable that the proposed α1-N and α1-C helices in ComGC seem to be separated by a larger stretch of residues, including the helix-breaking residue Pro-22, with no or less helical propensity. Interestingly, cryo-electron microscopy reconstruction of the *Neisseria meningitidis* type IV pilus recently revealed a similar non-helical portion in α1-N, between the residues Gly-14 and Pro-22, of the major pilin pilE (31). This stretch was proposed to function as a spring providing the filament additional flexibility in response to external forces, and it may have a similar purpose in pneumococcal competence pilin.

Type IV pili have a conserved role during the process of transformation, and pilus-deficient strains of naturally transformable species have reduced DNA uptake potential (32–34). Interestingly, *Neisseria* species bind DNA, in a sequence-specific manner, through the minor pilin ComP exposed on the type IV pilus surface (35, 36). In many other competent bacteria, the exact mechanisms that govern pilus-DNA interactions remain elusive. It is generally believed that DNA binding is a function of the intact pilus through solvent-exposed surface residues that mediate interactions with the DNA backbone. Laurenceau et al. (9) have previously shown direct DNA binding to the pneumococcal pilus, but monomeric ComGC was unable to bind DNA (37) suggesting that elements in the

### Table 1

<table>
<thead>
<tr>
<th>NMR structural constraints and structure statistics for ComGC³</th>
</tr>
</thead>
<tbody>
<tr>
<td>None of the structures exhibit distance violations of &gt;0.3 Å or dihedral angle violations &gt;4°.</td>
</tr>
</tbody>
</table>

#### Restraints

- **NOE-based restraints**
  - Intraresidual (i–i) = 0: 177
  - Sequential (i–i) = 1: 127
  - Medium range (i–i) = 4: 63
  - Long (i–i) > 5: 19
  - Total: 386
- **Hydrogen bond restraints**: 18
- **Dihedral angle restraints**: 106
- **Residual dipolar couplings**: 41
- **Magnitude (D<sub>2</sub>)**: −15
- **Rhombicity (R)**: 0.362

#### Restraints statistics

- **Mean r.m.s.d from experimental restraints**
  - NOE-based distances, Å: 0.022 ± 0.007
  - Dihedrals, ″: 0.415 ± 0.009
  - DRC, Q-factor: 0.078 ± 0.006

#### Structure statistics

- **Most favored regions**: 98.1%
- **Allowed regions**: 1.9%
- **Generously allowed regions**: 0.0%
- **Disallowed regions**: 0.0%
- **Coordinate precision r.m.s.d, Å**: 0.348 ± 0.76
- **Backbone heavy atoms (N, C<sup>α</sup>, and C′)**: 4.40 ± 0.86
- **H1s only (16–30)**: 0.93 ± 0.32
- **Heavy atoms**: 2.41 ± 1.16
- **H2 and H3 (36–60)**: 3.49 ± 1.40

---

**Solution structure of ComGC from S. pneumoniae**

![Graph](http://www.jbc.org/)

**Figure 6.** ComGC is dynamic on several different time scales. *A*, fitted ¹⁵N R<sub>i</sub> rates (top), R<sub>2</sub> rates (middle), and hetNOEs (bottom), at two different field strengths, 600 and 900 MHz, with respect to protons. The secondary structure of ComGC is shown above. *B*, reduced spectral density functions J(0) (top), J(0.87kHz) (middle), and J(0.87kHz) (bottom) calculated from the backbone relaxation data.

---

³ PROCHECK, structured regions 16–30, 36–42, and 47–60 are shown.
⁴ Residues 16–30, 36–42, and 47–60 are shown. Large root mean square deviations can be assigned to the flexible hinge between H1 and H2.
Solution structure of ComGC from *S. pneumoniae*

![Model of soluble ComGC. Polymorphisms are plotted on the ComGC structure. Variable residues are shown in shades of red depending on their susceptibility to undergo mutations.](image)

Figure 7. Polymorphisms in ComGC. Model of soluble ComGC. Polymorphisms are plotted on the ComGC structure. Variable residues are shown in shades of red depending on their susceptibility to undergo mutations.

Restoration of competence pilus quaternary structure are required for DNA interactions. Visualizing the electrostatic potential of the solvent-accessible surface in ComGC revealed a well-defined electropositive cavity formed between the helices. Along with the flexible N-terminal part, this region displays several solvent-exposed Lys and Arg residues, which are residues that have been found to mediate DNA backbone interactions in other DNA-binding proteins (38). Once a competence pilus model is available, it will be interesting to explore whether and how this electropositive cavity contributes to DNA binding.

In conclusion, the structure of pneumococcal ComGC represents a unique member in the growing family of type IV pilins, and it provides initial structural insights into understanding how competence pili assemble and how DNA is taken up in natural transformation of *S. pneumoniae*.

**Experimental procedures**

**Bacterial strains and growth conditions**

All *S. pneumoniae* strains used in this study are described in supplemental Table S1. Bacteria were grown on blood agar plates at 37 °C and 5% CO₂ overnight (O/N). For competence induction, plate-grown bacteria were used to inoculate C+Y medium, pH 7.9 – 8.0, at A₆₀₀ = 0.05 and grown without agitation at 37 °C until A₆₀₀ = 0.15. Competence was induced by addition of competence stimulating peptide (CSP-1 or 2 depending on the strain used) at a final concentration of 100 ng/ml for 20 min, if not specified otherwise.

**Transmission electron microscopy and immunogold labeling to visualize competence pili**

*S. pneumoniae* R6 or T4RΔrgA-srtD, an unencapsulated strain lacking pili encoded by the rlrA islet, were grown at 37 °C in C+Y medium until A₆₀₀ = 0.15 when competence was induced as described above. Twenty minutes post-induction the cells were centrifuged for 15 min at 5000 × g, 4 °C. The pellet was resuspended in 80 μl of phosphate-buffered saline (PBS). Drops of 10 μl were placed for 1 min on glow-discharged carbon-coated copper grids (Oxford Instruments, UK) for negative staining or carbon-coated gold grids (Aurion, Germany) for immunogold labeling. Negative staining was performed with 2% uranyl acetate in water. For immunogold labeling anti-ComGC antiserum, raised against a synthetic peptide corresponding to residues 95–108 of ComGC, was used. Grids were fixed with 10 μl of 0.2% glutaraldehyde for 2 min, and the reaction was stopped with 10 μl of 1% glycine for 15 min. The grids were then washed three times with PBS, 1% BSA, incubated with ComGC antibodies (1:100) for 1 h, washed three times with PBS, and incubated with secondary goat anti-rabbit antibody conjugated to 6-nm gold particles or protein A coupled to 10-nm gold particles diluted 1:250 for 45 min. Finally, the grids were washed six times with PBS and twice with distilled water before negatively staining with 2% uranyl acetate. Specimens were examined in a Tecnai 12 Spirit Bio TWIN transmission electron microscope (FEI Company, Eindhoven, Netherlands) operated at 100 kV. Digital images were recorded using a Veleta camera (Olympus Soft Imaging Solutions, GmbH, Münster, Germany).

**Analysis of competence pili**

Negative stain grids were prepared as described under immunogold labeling. Micrographs were collected manually at 120 kV using a Tecnai G2 Spirit TWIN electron microscope with a defocus value of 0.5–2.0 μm. Images were collected using a Tietz TemCam-F416 CMOS camera at a nominal magnification of ×67,000 and a pixel size of 1.57 Å employing the EM-Menu software (TVIPS GmbH). Data processing was done using the SPRING suite employing CTFFIND, CTFTILT, EMAN2, and SPARX (39–43). Straight regions of the pilus filaments were extracted, segmented, and averaged to determine outer dimensions and helical parameters. Averaged intensity width profiles were plotted, and the outer diameter was taken as the distance between the two outer minima in the intensity profile. Power spectra were calculated from the averages, and the most prominent layer lines were identified. The pitch was determined in real space from intensity normal profiles along the helical axis as well as from Fourier space based on layer line positions.

**Preparation of pili, two-dimensional PAGE, and immunoblotting to assess ComGC assembly**

Competence pili preparations were obtained from *S. pneumoniae* grown in 500 ml of C+Y medium, and competence was induced as described above. Bacteria were pelleted at 4 °C by centrifugation for 15 min at 6000 × g. The supernatant containing detached/broken pili was filtered and pelleted by ultracentrifugation at 100,000 × g at 4 °C for 1 h. Pellets were resuspended in 100 μl of PBS. Multimerization of mature ComGC was assessed by two-dimensional PAGE, native gel (first dimension) and SDS-PAGE (second dimension). In brief, pili preparations of competent T4 WT and T4ΔcomGC were run on a 12% native gel. Then, one entire lane of each sample was cut and placed perpendicular on top of a second gel. After migration, electroblotting (Bio-Rad, Trans-Blot® Turbo™ Midi PVDF Transfer Packs) and immunodetection with ComGC antibody were performed. Rabbit polyclonal ComGC antibody has been previously described (37). HRP-conjugated goat anti-rabbit
antibody (GE Healthcare) and Amersham Biosciences ECL Prime Western blotting detection reagent (GE Healthcare) were used to visualize the blots.

**Transformation frequency assay**

Genomic DNA of *S. pneumoniae* carrying a streptomycin resistance mutation in the *rpsl* gene (44) was used to transform competent bacteria. In brief, *S. pneumoniae* was grown in C+Y medium at 37 °C until A620 = 0.15. Bacteria were then incubated at 30 °C for 15 min before CSP was added. After 15 min, 1 µg/ml DNA was added. Bacteria were then incubated for 30 min at 30 °C and another 60 min at 37 °C before plating in the presence and absence of streptomycin at 100 µg/ml final concentration. Blood plates were incubated O/N at 37 °C and 5% CO2 before being counted.

**In vitro ComGC processing**

The plasmids pJWV25-*PilD* and pACYCDuet-1-*flcomGC* were constructed as follows. Full-length *PilD* and full-length *comGC* were amplified from *S. pneumoniae* TIGR4 genomic DNA using Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher Scientific) and suitable primers (supplemental Table S2). PCR products were digested with NotI (*pilD*) or Ndel and Xho (*comGC*) and subcloned into pJWV25 and pACYCDuet-1, respectively. The correct insertion was confirmed by PCR and sequencing (Eurofins MWG Operon). The resulting plasmids pJWV25-*PilD* and pACYCDuet-1-*flcomGC* (supplemental Table S3) were then co-transformed into competent T7 express *Escherichia coli* (New England Biolabs). Bacteria were grown in LB, pH 7.5, supplemented with 100 µg/ml ampicillin and 50 µg/ml chloramphenicol at 37 °C until A620 = 0.5, and induced with 1 mm isopropyl-ß-d-thiogalactopyranoside for 3 h. Bacteria were spun down, and 1× sample buffer was added to the pellet. Samples were incubated at 100 °C for 5 min before analysis by SDS-PAGE and immunoblotting with ComGC antibody as described above.

**BACTH**

All plasmids used for BACTH are listed in supplemental Table S3. The gene encoding mature ComGC and mature PuLG were PCR-amplified using suitable primers (supplemental Table S2) and cloned into pUT18C and pKT25. *E. coli* Top10 (Invitrogen) was used for all clonings. *E. coli* BTH101 (Euromedex) was co-transformed with respective BACTH plasmids (supplemental Table S3) and used for BACTH assay. The efficiency of the functional complementation between the recombinant plasmids encoding fusions to T18 (pUT18C) and T25 (pKT25) was quantified by measuring β-galactosidase activity in liquid culture as described previously with some modifications (21). Co-transformed BTH101 strains were grown in 5 ml of LB medium, supplemented with 100 µg/ml ampicillin, 50 µg/ml kanamycin, and 0.5 µM isopropyl-ß-d-thiogalactopyranoside, O/N at 30 °C. Three individual clones of each co-transformation were tested, and at least three independent cultures were performed. Subsequently, cultures were incubated for 20 min on ice and pelleted by centrifugation for 10 min at 4 °C. Next, cells were resuspended in the same volume of 1× X buffer (90 mM Na2HPO4·2H2O, 40 mM NaH2PO4·H2O, 6 mM NaOH, 10 mM MgSO4·7H2O, 50 mM β-mercaptoethanol) and diluted until a final A600 nm = 0.3. Then, 1 ml of bacterial suspension was permeabilized by adding 100 µl of chloroform and 50 µl of 0.1% SDS. Tubes were then vortexed and incubated at 28 °C for 10 min before the enzymatic reaction was started by adding 200 µl of 0.4% 0-nitrophenyl-β-d-galactopyranoside in phosphate buffer (90 mM Na2HPO4·2H2O, 40 mM NaH2PO4·H2O). The reaction was stopped by the addition of 500 µl of 1 mM Na2CO3 when the samples became noticeably yellow, and the time of incubation with the substrate was recorded. The reaction mixtures were centrifuged for 5 min, and the supernatants were transferred into a cuvette. Then, the absorbance was recorded at 420 and 550 nm for each sample. The β-galactosidase activity was expressed in Miller units by using the following formula: 1000 × (A420 nm - 1.75 × A550 nm)/(incubation time (minutes) × volume (1 ml) × A600 nm).

**Chemical cross-linking**

*In vitro* cross-linking experiments were essentially performed as described previously (46). In brief, exponentially grown bacteria were pelleted by centrifugation, washed with 10 mM sodium phosphate buffer, pH 6.8, and incubated with 1% paraformaldehyde (Sigma) in 10 mM sodium phosphate buffer, pH 6.8, for 30 min. Cross-linking was stopped by addition of 3 M Tris, pH 8.8 (final concentration 300 mM Tris). Bacteria were washed, and pellets were resuspended in 1× NuPAGE sample buffer (Thermo Fisher Scientific) without reducing agent. Each sample was split into two tubes. One tube was kept at room temperature, and one tube was heated at 96 °C for 15 min before further analysis by SDS-PAGE and Western blotting with ComGC antibody.

**Expression and purification of labeled ComGC for NMR**

The DNA sequence of ComGC lacking the signal peptide and codons for the N-terminal hydrophobic domain (ComGCΔ1–39) was cloned downstream of the His6 tag sequence into pet28a vector (Novagen). Constructs were confirmed by sequencing and transformed into *E. coli* Rosetta (DE3). Cells were grown O/N with shaking at 37 °C in M9 minimal media supplemented with [13C]glucose and [15N]ammonium sulfate containing 50 µg/ml kanamycin. The O/N culture was then diluted into fresh medium; cells were grown to A620 = 0.5 at 37 °C and induced with 1 µM isopropyl-β-d-thiogalactopyranoside for 3.5 h. Cells were harvested by centrifugation at 7000 × g for 20 min at 4 °C, and pellets were stored at −20 °C. For affinity purification of ComGC, cell pellets were resuspended in buffer containing 50 mM Tris, 50 mM NaCl, pH 7.5, and protease inhibitor (Roche Applied Science) and lysed in a Stansted cell disrupter. Unbroken cells were pelleted by centrifugation at 35,000 × g, and supernatants were incubated with nickel-nitrotriacetic acid (Ni-NTA, Qiagen)-agarose with rotation at 4 °C O/N. After washing the resin, protein was eluted with buffer containing 50 mM Tris, 50 mM NaCl, and 250 mM imidazole, pH 7.5. Imidazole was removed using PD-10 desalting columns (GE Healthcare). The N-terminal His6 tag was cleaved with thrombin (Sigma) for 2 h at room temperature and removed by incubation with Ni-NTA resin. ComGC was

---

### Solution structure of ComGC from *S. pneumoniae*

The DNA sequence of ComGC lacking the signal peptide and codons for the N-terminal hydrophobic domain (ComGCΔ1–39) was cloned downstream of the His6 tag sequence into pet28a vector (Novagen). Constructs were confirmed by sequencing and transformed into *E. coli* Rosetta (DE3). Cells were grown O/N with shaking at 37 °C in M9 minimal media supplemented with [13C]glucose and [15N]ammonium sulfate containing 50 µg/ml kanamycin. The O/N culture was then diluted into fresh medium; cells were grown to A620 = 0.5 at 37 °C and induced with 1 µM isopropyl-β-d-thiogalactopyranoside for 3.5 h. Cells were harvested by centrifugation at 7000 × g for 20 min at 4 °C, and pellets were stored at −20 °C. For affinity purification of ComGC, cell pellets were resuspended in buffer containing 50 mM Tris, 50 mM NaCl, pH 7.5, and protease inhibitor (Roche Applied Science) and lysed in a Stansted cell disrupter. Unbroken cells were pelleted by centrifugation at 35,000 × g, and supernatants were incubated with nickel-nitrotriacetic acid (Ni-NTA, Qiagen)-agarose with rotation at 4 °C O/N. After washing the resin, protein was eluted with buffer containing 50 mM Tris, 50 mM NaCl, and 250 mM imidazole, pH 7.5. Imidazole was removed using PD-10 desalting columns (GE Healthcare). The N-terminal His6 tag was cleaved with thrombin (Sigma) for 2 h at room temperature and removed by incubation with Ni-NTA resin. ComGC was

---

### References

[14143](http://www.jbc.org/content/292/34/14134)
**Solution structure of ComGC from S. pneumoniae**

further purified by size-exclusion chromatography on a Superdex 75 gel filtration column.

**Nuclear magnetic resonance (NMR)**

**NMR data collection**—The NMR samples contained ComGC at a concentration of 0.9 mM in a buffer containing 50 mM Tris, 50 mM NaCl, pH 7.5, and had a volume of 800 μl. Experiments were performed in 5-mm tubes at a temperature of 283 K on Bruker AVIII 600 spectrometer operating at 600 MHz (1H frequency). They were equipped with either a TXI cryoprobe or a TCI cryoprobe both equipped with a self-shielding z-gradient, and a Bruker AVII 900 spectrometer operating at 900 MHz (1H frequency), also at 283 K, and using a TCI cryoprobe equipped with a self-shielding z-gradient. Residual dipolar coupling (RDC) and HCCH-TOCSY spectra were recorded on an 800 MHz Agilent DD2 spectrometer with a room temperature probe.

**Sequential assignment**—ComGC backbone chemical shifts were sequentially assigned using 1H, 15N-HSQC, 1H, 13C-HSQC CBCANH, CBCA(CO)NH, HNCA, HN(CO)CA, HNCO, HN(CA)CO, HBHA(CO)NH CC(CO)NH, and HCC(CO)NH spectra. Side-chain chemical shifts were assigned by using a HCCH-TOCSY spectrum recorded with a mixing time of 20 ms, and 15N-TOCSY-HSQC and 13C-TOCSY-HSQC spectra recorded with mixing times of 80 ms. Chemical shifts are deposited in the BMRB with ID 34112.

**Structural determination**—Distance restraints were obtained from 15N NOESY-HSQC and 13C NOESY-HSQC experiments recorded using mixing times of 150 and 80 ms, respectively. Residual dipolar couplings were obtained from a sample aligned in PEG (C12E5)/hexanol (Sigma: 76437/H13303) liquid crystal medium with a final PEG concentration of 4%. The couplings were measured in in-phase/anti-phase 1H, 15N HSQC spectra (47–49). Backbone dihedral angle restraints were calculated using DANGLE (50). Automated NOE assignment was performed using Cyana (51), and XPLOR-NIH (52) was subsequently used for including the RDC restraints and refining the structure. In total, 200 structures were calculated from which the 20 lowest energy structures were selected. Structures were aligned using Theseus (53) and visualized in PyMOL (DeLano Scientific). Electrostatics were calculated using APBS (54). Ramachandran plot statistics for the structural ensemble were calculated with PROCHECK (55). The coordinates are deposited in the Protein Data Bank with code 5NCA.

**Dynamics**—R1 and R2 relaxation rates were measured in using the 2D 1H, 15N HSQC-based pulse sequences by Farrow et al. (56) using delays of 0.080, 0.240, 0.400, 0.640, 0.880, 1.280, and 1.600 s for R1 experiments and 0.016, 0.032, 0.048, 0.080, 0.112, 0.144, and 0.2 s for R2 (starred values are recorded in duplicates). Heteronuclear 1H, 15N NOE values were determined from peak ratios between saturated/steady-state and a reference spectrum (Iref/Inoe) (57). Reduced spectral densities are calculated using Equations 1–4:

\[ J(\omega_n) = \frac{4}{3d^2 + 4c} \left( R_1 \left( 6R_2 - R_1 \left( 3 - \frac{18}{5} \frac{\gamma_n}{\gamma_H}(\text{NOE}) - 1 \right) \right) \right) \quad (\text{Eq. 1}) \]

\[ J(0.87\omega_n) = \frac{4}{5d^2} \left( R_1 \left( \frac{\gamma_n}{\gamma_H}(\text{NOE}) - 1 \right) \right) \quad (\text{Eq. 3}) \]

\[ d = \frac{(\mu_0 h \gamma_n \gamma_H)}{8\pi^2} \frac{1}{\tau} \frac{\omega_n \Delta \sigma}{\sqrt{3}} \quad (\text{Eq. 4}) \]

where \( R_1 \) and \( R_2 \) and NOE are the fitted relaxation rates or intensity ratios; \( \gamma_n \) and \( \gamma_H \) are the gyromagnetic ratios of N and H; \( \mu_0 \) is the permeability of free space; \( h \) is Planck’s constant; \( R_H \) is the bond length vector of the amide NH bond (set to 1.02 Å); \( \omega_n \) is the larmor frequency of N, and \( \Delta \sigma \) is the chemical shift anisotropy of N (set to −172 ppm) (45).

**Statistical analysis**

Data were statistically analyzed as indicated in the figure legends, using GraphPad Prism 5.04. If not stated otherwise, asterisks in the figures indicate groups of statistically different means (**, p < 0.001), determined by one-way analysis of variance with subsequent Dunnett’s post hoc test.

**Author contributions**—S. M., U. A., and B. H. N. designed the study, and S. M., M. S. A., and V. O. performed the experiments. S. M. purified ComGC, and S. E., P. S., C. D. L., K. T., and U. A. determined the NMR structure and dynamics. T. B. performed class averages on pili. S. M., M. S. A., and V. O. performed the experiments. S. M. purified ComGC, and S. E., P. S., C. D. L., K. T., and U. A. determined the NMR structure and dynamics. T. B. performed class averages on pili. S. M., E. S., M. S. A., and B. H. N. wrote the paper. All authors reviewed the results and approved the final version of the manuscript.

**Acknowledgments**—We thank Sergej Masich at the electron tomography unit at the Department of Cell and Molecular Biology, Karolinska Institutet, for advice and support and Miriam Bertram and Richard Böhme for excellent technical assistance.

**References**


Solution structure of ComGC from *S. pneumoniae*


Structure of the competence pilus major pilin ComGC in Streptococcus pneumoniae
Sandra Muschiol, Simon Erlendsson, Marie-Stephanie Aschtgen, Vitor Oliveira, Peter Schmieder, Casper de Lichtenberg, Kaare Teilum, Thomas Boesen, Umit Akbey and Birgitta Henriques-Normark

doi: 10.1074/jbc.M117.787671 originally published online June 28, 2017

Access the most updated version of this article at doi: 10.1074/jbc.M117.787671

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2017/06/28/M117.787671.DC1

This article cites 57 references, 18 of which can be accessed free at http://www.jbc.org/content/292/34/14134.full.html#ref-list-1