

# Structure of the competence pilus major pilin ComGC in *Streptococcus pneumoniae*

Received for publication, March 22, 2017, and in revised form, June 14, 2017. Published, Papers in Press, June 28, 2017, DOI 10.1074/jbc.M117.787671

Sandra Muschiol<sup>†§1</sup>, Simon Erendsson<sup>¶</sup>, Marie-Stephanie Aschtgen<sup>‡§</sup>, Vitor Oliveira<sup>‡§</sup>, Peter Schmieder<sup>||</sup>, Casper de Lichtenberg<sup>¶</sup>, Kaare Teilum<sup>¶</sup>, Thomas Boesen<sup>\*\*</sup>, Umit Akbey<sup>\*\*\*††</sup>, and Birgitta Henriques-Normark<sup>†§§2</sup>

From the <sup>†</sup>Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, 171 77 Stockholm, Sweden, the <sup>§</sup>Department of Clinical Microbiology, Karolinska University Hospital, 171 76 Stockholm, Sweden, the <sup>¶</sup>Structural Biology and NMR Laboratory, Linderstrøm-Lang Center for Protein Science, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen N, Denmark, the <sup>||</sup>Leibniz-Institut für Molekulare Pharmakologie FMP, Robert-Rössle-Strasse 10, 13125 Berlin, Germany, the <sup>\*\*</sup>Interdisciplinary Nanoscience Center (iNANO), Aarhus University, Gustav Wieds Vej 14, 8000 Aarhus C, Denmark, the <sup>††</sup>Aarhus Institute of Advanced Studies (AIAS), Aarhus University, Høegh-Guldbergs Gade 6B, 8000 Aarhus C, Denmark, and the <sup>§§</sup>Singapore Centre on Environmental Life Sciences Engineering (SCELSSE) and Lee Kong Chian School of Medicine (LKC), Nanyang Technological University, Singapore 639798, Singapore

Edited by Chris Whitfield

Type IV pili 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 pili 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  $\alpha$ -helical segments forming a well-defined electronegative cavity and confined electronegative and hydrophobic patches. The structure is particularly flexible between the first and second  $\alpha$ -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.

Type IV pili are important virulence factors on the surface of many pathogenic bacteria. These extracellular appendages can be several microns long and are involved in various functions, including adherence (1, 2), twitching motility (3, 4), biofilm

This work was supported by grants from the Swedish Research Council, the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, the Swedish Foundation for Strategic Research (SSF), the Knut and Alice Wallenberg Foundation, the Carlsberg Foundation, and the Lundbeck Foundation. The authors declare that they have no conflicts of interest with the contents of this article.

✂ Author's Choice—Final version free via Creative Commons CC-BY license.

This article contains Tables S1–S3 and Figs. S1–S3.

The atomic coordinates and structure factors (code 5NCA) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

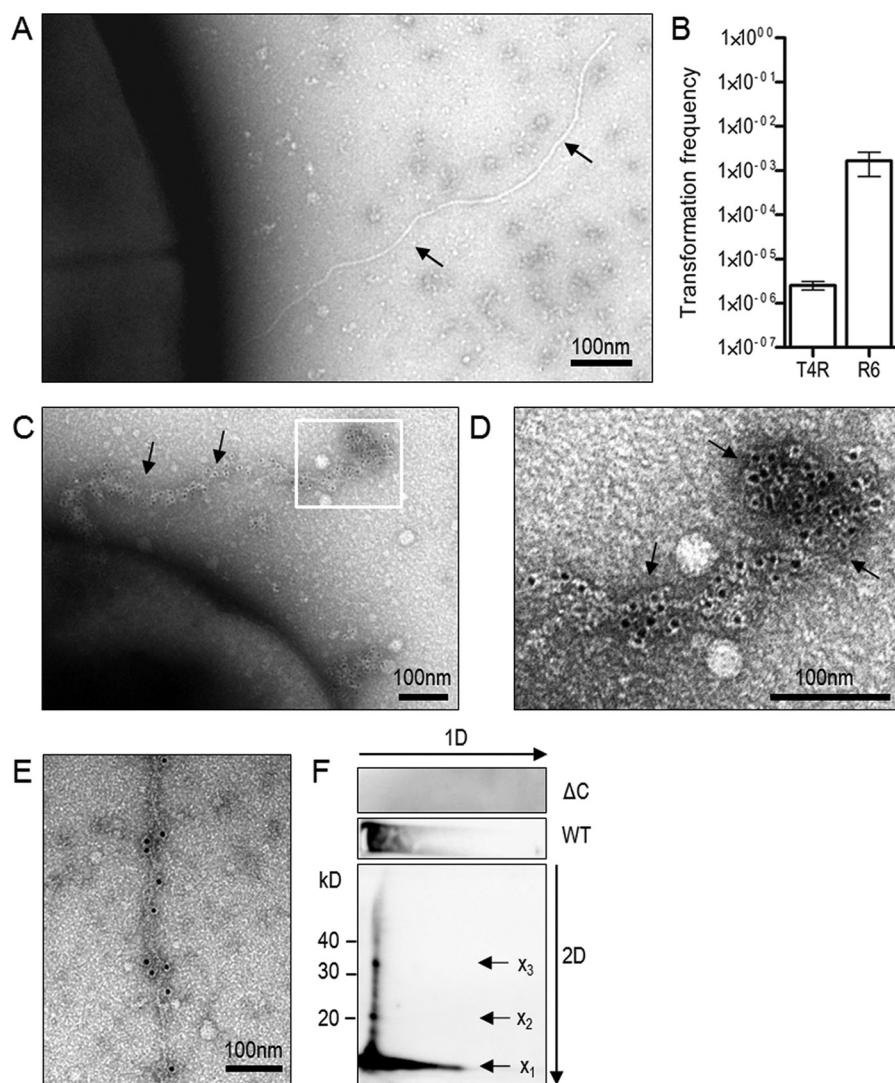
<sup>1</sup> To whom correspondence may be addressed. E-mail: sandra.muschiol@ki.se.

<sup>2</sup> To whom correspondence may be addressed. E-mail: birgitta.henriques@ki.se.

formation (5, 6), and DNA uptake (7–9). Type IV pili are composed of thousands of copies of major pilin protein that are tightly packed in a helical arrangement (10, 11). Pilins are synthesized as prepilins containing a conserved N-terminal prepilin cleavage motif. Once synthesized, prepilins are processed by a membrane-bound prepilin peptidase, often called PilD, which removes the signal peptide. Based on the length of the signal peptide and the length of mature pilin, two subclasses, namely type IVa and type IVb pilins, have been distinguished (12).

A number of pilin structures are available for both subclasses mainly for Gram-negative bacteria (13). They suggest an overall conserved architecture, with each pilin having an extended N-terminal domain ( $\alpha$ 1-N and  $\alpha$ 1-C) and a C-terminal globular head domain.  $\alpha$ 1-N is primarily hydrophobic and retains the pilin subunits in the inner membrane until assembly, whereas the  $\alpha$ 1-C is tightly packed against the head domain composed of several  $\beta$ -strands. The  $\alpha/\beta$  loop connects the N-terminal helix to the  $\beta$ -sheet and is important for interactions between individual pilin subunits (10). Upon pilus assembly,  $\alpha$ 1-N forms the core of the assembled pilus, and  $\alpha$ 1-C is buried in the C-terminal head domain that forms the pilus surface. Characteristic for most pilins is also a disulfide-bonded loop (D-region) in the C-terminal domain, which is essential for pilus assembly (10). Most of the structural diversity among different pilins lies in the  $\alpha/\beta$  loop, and the number and topology of  $\beta$ -strands are in the C-terminal domain. Notably, many of the available pilin structures are lacking the highly hydrophobic N-terminal domain ( $\alpha$ 1-N) making the truncated protein more soluble and easier to purify for later structural characterization.

Type IV pili are also produced by Gram-positive bacteria, including several *Clostridium* species (14), *Ruminococcus albus* (15), and *Streptococcus* species (9, 16, 17), but many molecular and structural aspects of pilus biogenesis in Gram-positive species remain unclear. Recently, DNA uptake in *Streptococcus pneumoniae* was shown to rely on the formation of a type IV pilus that is able to directly bind to DNA (9). This transformation pilus is assembled on the surface of competent bacteria and composed of the major pilin ComGC. Pneumococcal comGC is encoded in the comG operon that also encodes a putative



**Figure 1. *S. pneumoniae* assembles competence type IV pili composed of ComGC.** *A*, electron micrograph of negatively stained competent *S. pneumoniae* T4R $\Delta$ rrgA-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. pneumoniae* 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  $\Delta$ C 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.

ATPase (ComGA), which powers pilus assembly (9), a membrane-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  $\alpha$ -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 (T4R $\Delta$ rrgA-srtD). The *rlrA* operon encodes an adhesive pneumococcal pilus that is assembled by pilus-associated 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 T4R $\Delta$ rrgA-srtD cultures induced with the competence-stimulating peptide (CSP)<sup>3</sup> and control cultures without CSP addition. As shown in Fig. 1*A*, a type IV pilus could be visualized by transmission electron microscopy in negatively stained *S. pneumoniae*

<sup>3</sup> The abbreviations used are: CSP, competence stimulating peptide; BACTH, bacterial adenylate cyclase two-hybrid; hetNOE, heteronuclear <sup>1</sup>H-<sup>15</sup>N-NOE; O/N, overnight; RDC, residual dipolar coupling; Ni-NTA, nickel-nitrilotriacetic acid.

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

T4R $\Delta$ rrgA-srtD. When we compared electron micrographs of negatively stained competent T4R $\Delta$ rrgA-srtD to R6, pili 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 pili preparations by two-dimensional (2D) PAGE. First, pili preparations of wild-type T4 (WT) or a comGC knock-out mutant ( $\Delta$ C) 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 immunoblotted 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 ( $x_1$ ); when only partially denatured, distinct ComGC building blocks ( $x_2$  and  $x_3$ ) 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 pili, 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  $0.0252 \text{ \AA}^{-1}$  from the equator corresponding to a helical pitch of  $\sim 40 \text{ \AA}$  (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  $64 \pm 1.6 \text{ \AA}$  (Fig. 2D). A representative class average can be seen in Fig. 2E.

### ComGC processing and dimerization in the membrane

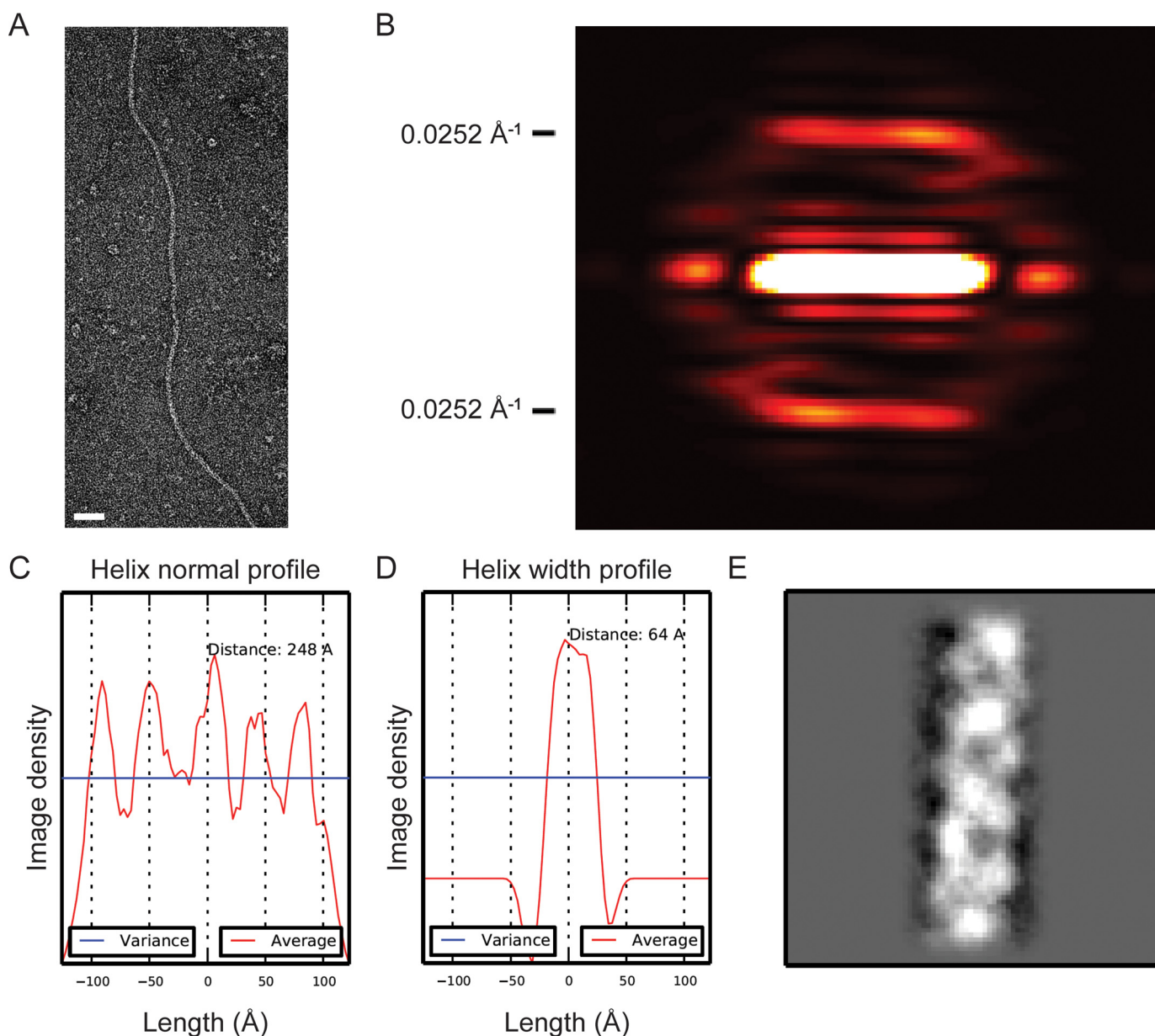
One characteristic of proteins belonging to the type IV pilin family is the presence of a well-conserved prepilin cleavage motif Gly-Phe-Xaa-Xaa-Xaa-Glu (20). Pneumococcal ComGC

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  $\beta$ -strand-rich head domain. Indeed, secondary structure predictions using Jpred4 (23) and Agadir (24) show that the ComGC sequence has several segments with high  $\alpha$ -helix propensity and no  $\beta$ -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<sup>s</sup>, see Fig. 5A). It was previously shown for the major pilin PAK in *Pseudomonas aeruginosa* that deletion of  $\alpha$ 1-N does not perturb the structural fold, with full-length and truncated protein essentially being identical (25). To determine the ComGC<sup>s</sup> structure, we used NMR spectroscopy. ComGC<sup>s</sup> 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<sup>s</sup> in solution (Fig. 5B). A summary of the structural statistics and constraints is provided in Table 1. ComGC<sup>s</sup> consists of three flexible helical segments:  $\alpha$ 1-C, involving residues 54–69; a shorter  $\alpha$ 2 helix involving residues 75–81; and finally, a C-terminal  $\alpha$ 3 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-



**Figure 2. Structural features of ComGC filaments.** *A*, representative electron micrograph of a competence pilus in strain R6 used for class averages. *Scale bar*, 50 nm. *B*, averaged power spectrum of a representative class average based on straight pilus regions. Predominant layer lines at 0.0252 Å<sup>-1</sup> from the equator are visible corresponding to a helical pitch of ~40 Å. *C*, helix normal profile along the helical axis showing a distance between peaks corresponding to the pitch derived from the power spectrum. *D*, helix width profile of class averages showing a width of 64 Å. The mean diameter of 16 class averages was 63 ± 1.6 Å. *E*, representative 2D class average of the competence pilus.

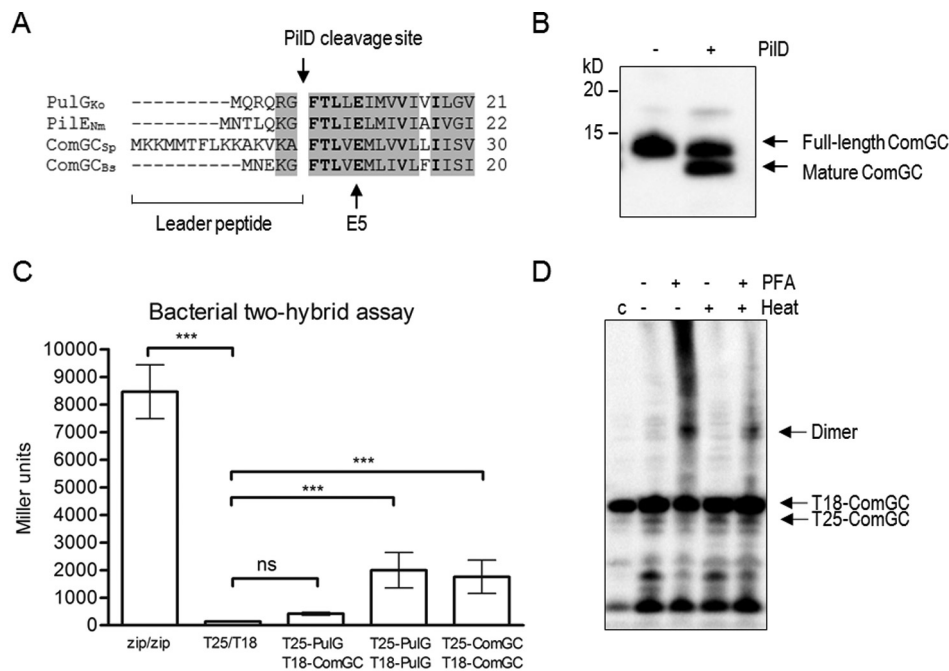
cule (supplemental Fig. S2). This observation is in good agreement with the secondary chemical shifts (supplemental Fig. S1A).  $\alpha 1$ -C seems to be only loosely attached to  $\alpha 2$  and  $\alpha 3$ , and a general lack of long-range distance restraints between these two “domains” indicates 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 Å<sup>2</sup>). Visualizing the electrostatic potential of the solvent-accessible surface revealed a well-defined elec-

tropositive cavity formed between the helices, and two highly electronegative areas (denoted  $\delta^-$ ) in the top of  $\alpha 1$ -C and the opposite side of  $\alpha 2$  (Fig. 5E). Similarly, visualizing the hydrophobicity of the surface also showed two well-defined patches. The first hydrophobic patch ( $\phi 1$ ) is situated at the back of  $\alpha 1$ -C, opposite to the electropositive patch ( $\phi 1$ ) and the other ( $\phi 2$ ) in  $\alpha 2$  (Fig. 5F).

#### **ComGC<sup>s</sup> 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 ( $R_1$ ) and the transverse ( $R_2$ ) <sup>15</sup>N relaxation rates in both (relative to the external magnetic field) as well as he-

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



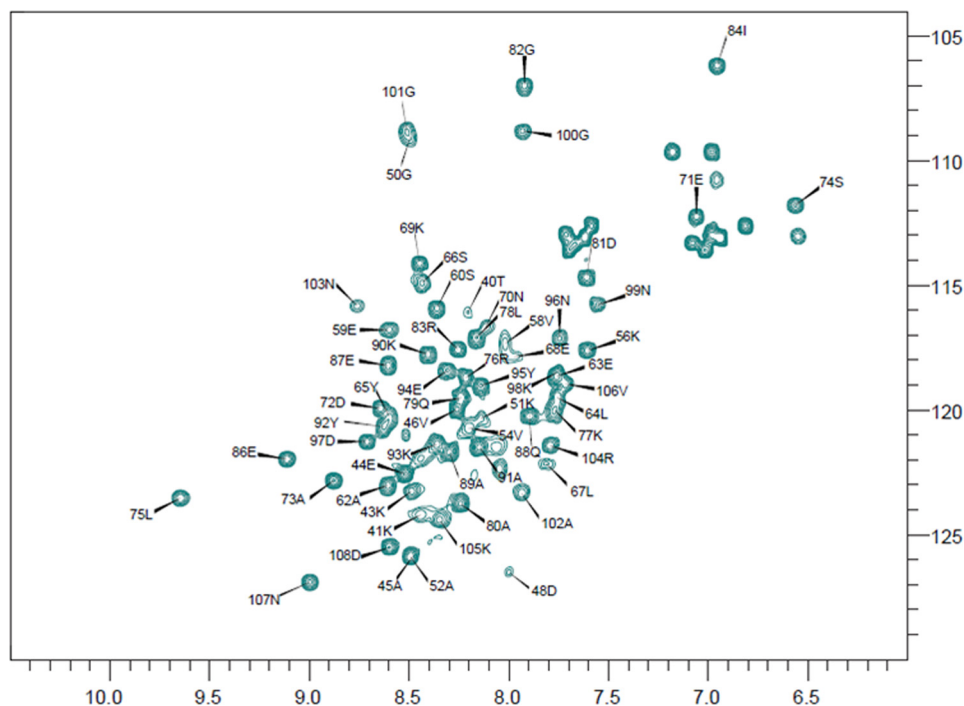
**Figure 3. Pneumococcal ComGC is processed by PiID and can interact with itself.** *A*, partial sequence alignment of the N-terminal domains of the major pilin proteins in *K. oxytoca* (PulG), *N. meningitidis* (PilE), *S. pneumoniae* (ComGC<sub>Sp</sub>), and *B. subtilis* (ComGC<sub>Bs</sub>). The characteristic prepilin cleavage site recognized by the prepilin peptidase PiID and the invariant Glu residues at position 5 (E5) after the cleavage site are indicated by an arrow. Conserved residues are shaded, and identical residues are shown in bold. *B*, *in vitro* ComGC cleavage analyzed by Western blotting with antibodies specific to ComGC. ComGC is partially processed when co-expressed with PiID in *E. coli*. *C*, quantification of ComGC–ComGC interaction identified by bacterial two-hybrid assay. The graph shows mean values of  $\beta$ -galactosidase activity expressed in Miller units between the indicated hybrid proteins T25–ComGC/T18–ComGC. A strain expressing Zip–T18 and Zip–T25, in which the hybrid proteins interact through a leucine zipper motif, was used as positive control. *K. oxytoca* PulG, T25–PulG/T18–PulG was included as functional positive control. *E. coli* BTH101 co-transformed with pUT18C and pKT25 empty plasmid or a strain expressing T25–PulG/T18–ComGC was used as negative control. The error bars represent S.D. of a minimum of three independent experiments with three different clones. A one-way analysis of variance test followed by Dunnett's post-test to compare each interaction pair to the negative control (T25/T18) was used for statistical analysis: \*\*\*,  $p < 0.001$ ; ns, no significant difference. *D*, paraformaldehyde cross-linking of *E. coli* BTH101 expressing T25–ComGC/T18–ComGC. Cell extracts were analyzed by immunoblotting with anti-ComGC antibody. An untreated sample was used as expression control (indicated as *C*, left lane). Hybrid proteins and dimerization are indicated on the right side of the panel.

teronuclear  $^1\text{H}$ - $^{15}\text{N}$ -NOEs (hetNOEs), at two different field strengths (Fig. 6A). As expected, and in support of the ComGC<sup>s</sup> structure, the unstructured region 40–52 displayed generally longer  $R_1$  rates ( $>1.5 \text{ s}^{-1}$ ), short  $R_2$  rates ( $<10 \text{ s}^{-1}$ ), 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 ( $20.3 \text{ s}^{-1}$  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  $J(0.87\omega_H)$ , 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 indicates 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, and 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  $\alpha$ C-1 helix (Fig. 7). This suggests that the hypothesized hydrophobic pilin–pilin interface involving the



**Figure 4. Assigned  $^1\text{H}$ - $^{15}\text{N}$  HSQC of ComGC recorded at 283K.** The HSQC spectrum displays signals from all the backbone and side chain N-H correlations. The  $^1\text{H}$  chemical shift is plotted along the x axis and the  $^{15}\text{N}$  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.

transmembrane  $\alpha$ 1-N and  $\alpha$ 1-C helical domains has been largely conserved throughout evolution and that the  $\alpha$ 2- $\alpha$ 3 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  $\sim$ 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*,  $\sim$ 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.

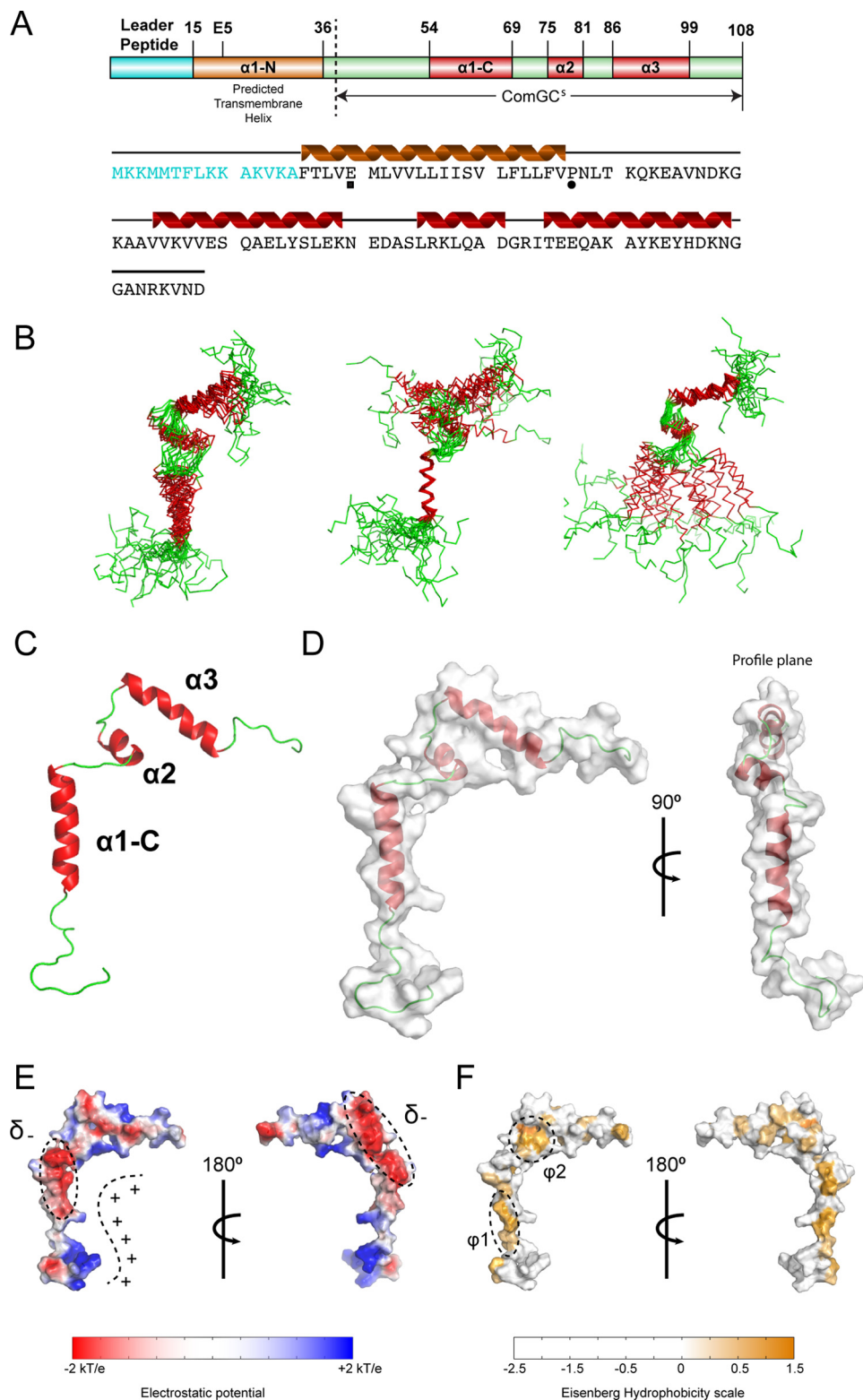
Pneumococcal ComGC shares many features of canonical type IV pilins. Full-length ComGC has a well-defined conserved prepilin cleavage motif, an invariant Glu residue at position 5 after the cleavage site, and it is processed by PilD. The structure of soluble ComGC, provided here, is the first example 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  $\alpha$ -helix but differs otherwise significantly as follows: 1) ComGC is exclusively  $\alpha$ -helical; 2) the head group is much smaller; 3) the  $\alpha$ 1-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  $\alpha$ -helices ( $\alpha$ 2 and  $\alpha$ 3) 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

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



**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<sup>s</sup> 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),  $\alpha 1$ -C only (middle), and  $\alpha 2$ - $\alpha 3$  only (right) are shown. The structure is particularly flexible between  $\alpha 1$ -C and  $\alpha 2$ - $\alpha 3$ , which is illustrated by alignment of the  $\alpha 1$ -C only and  $\alpha 2$  and  $\alpha 3$ . Root mean square deviation values are listed in supplemental Table S1. *C*, schematic of the ComGC<sup>s</sup> NMR structure showing the three helical segments as well as the flexible regions. *D*, calculated solvent-accessible surface of ComGC<sup>s</sup>. *E*, APBS calculated electrostatics from  $\pm 2$  kT/e display well defined electropositive (+) and electronegative ( $\delta^-$ ) solvent-accessible patches. *F*, solvent-accessible hydrophobic patches  $\phi 1$  and  $\phi 2$  colored using the Eisenberg hydrophobicity scale ranging from  $-2.5$  to  $1.5$ .

**Table 1**  
NMR structural constraints and structure statistics for ComGC<sup>5</sup>

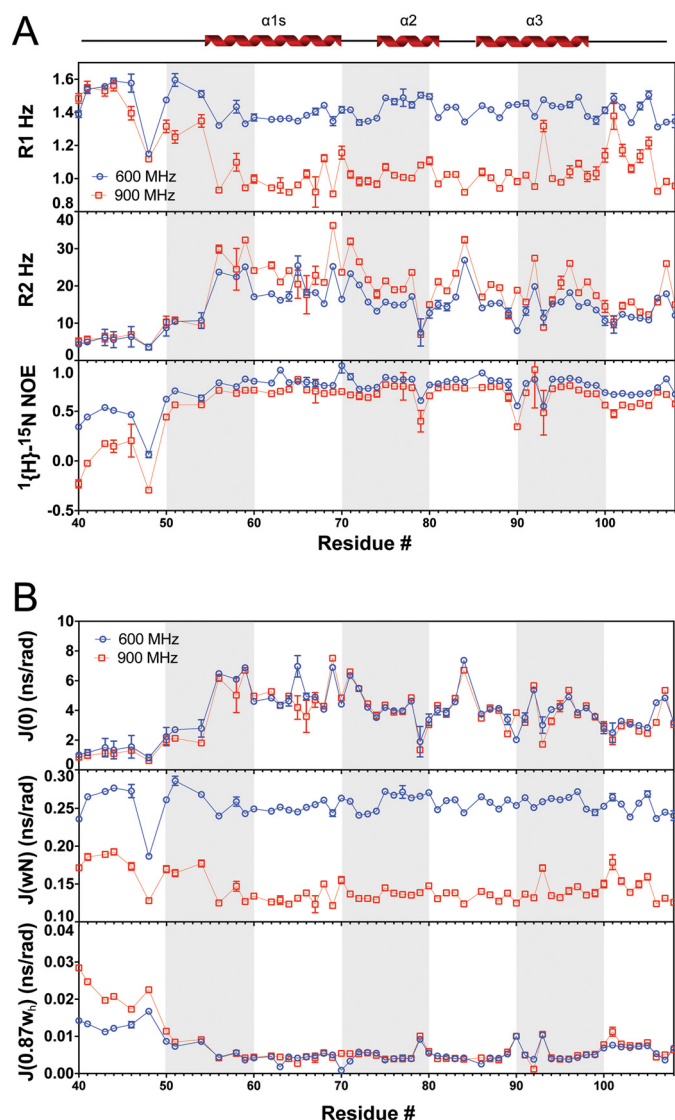
None of the structures exhibit distance violations of &gt;0.3 Å or dihedral angle violations &gt;4°.

Restrictions	
NOE-based restraints	
Intraresidual ( $ i - j  = 0$ )	177
Sequential ( $ i - j  = 1$ )	127
Medium range ( $2 \leq  i - j  \leq 4$ )	63
Long ( $ i - j  > 5$ )	19
Total	386
Hydrogen bond restraints	
Dihedral angle restraints	106
Residual dipolar couplings	
Magnitude ( $D_n$ )	-15
Rhombicity ( $R$ )	0.362
Restrictions statistics	
Mean r.m.s.d from experimental restraints	
NOE-based distances, Å	0.022 ± 0.007
Dihedrals, °	0.415 ± 0.009
RDC, Q-factor	0.078 ± 0.006
Structure statistics <sup>a</sup>	
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 <sup>b</sup>	
Backbone heavy atoms (N, C <sup>α</sup> , and C')	3.48 ± 0.76
Heavy atoms	4.40 ± 0.86
H1s only (16–30)	
Backbone heavy atoms (N, C <sup>α</sup> , and C')	0.93 ± 0.32
Heavy atoms	1.94 ± 0.42
H2 and H3 (36–60)	
Backbone heavy atoms (N, C <sup>α</sup> , and C')	2.41 ± 1.16
Heavy atoms	3.49 ± 1.40

<sup>a</sup> PROCHECK, structured regions 16–30, 36–42, and 47–60 are shown.<sup>b</sup> 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.

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 pili 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  $\alpha 1$ -N and  $\alpha 1$ -C are involved in forming the core of the pilus structure, with parts of  $\alpha 2$  and  $\alpha 3$  being primarily surface-exposed. The residues Leu-78, Ile-84, and Tyr-92, involved in the hydrophobic patch,  $\phi 2$ , formed between  $\alpha 2$  and  $\alpha 3$  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  $\alpha 1$ -N and  $\alpha 1$ -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  $\alpha 1$ -N, between the residues Gly-14 and Pro-22, of the major pilin pIL



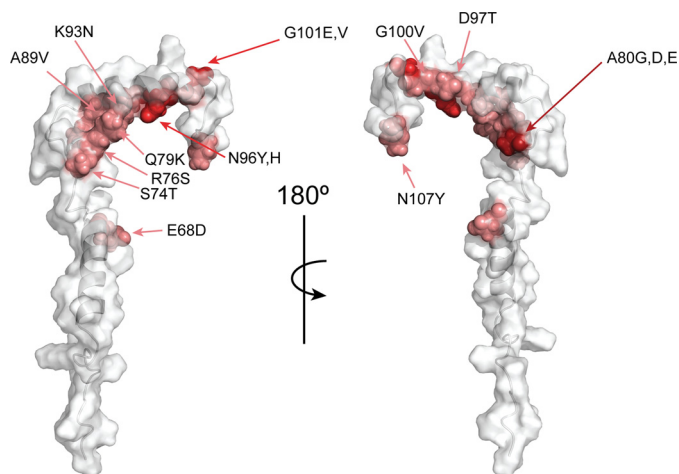
**Figure 6.** ComGC<sup>5</sup> is dynamic on several different time scales. *A*, fitted <sup>15</sup>N  $R_1$  rates (top),  $R_2$  rates (middle), and hetNOEs (bottom), at two different field strengths, 600 and 900 MHz, with respect to protons. The secondary structure of ComGC<sup>5</sup> is shown above. *B*, reduced spectral density functions  $J(0)$  (top),  $J(\omega_N)$  (middle), and  $J(0.87\omega_h)$  (bottom) calculated from the backbone relaxation data.

(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 pili.

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



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



**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.

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% CO<sub>2</sub> overnight (O/N). For competence induction, plate-grown bacteria were used to inoculate C+Y medium, pH 7.9–8.0, at  $A_{620} = 0.05$  and grown without agitation at 37 °C until  $A_{620} = 0.15$ . Competence was induced by addition of competence stimulating peptide (CSP-1 or 2 dependent 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 T4RΔrrgA-srtD, an unencapsulated strain lacking pili encoded by the *rlrA* islet, were grown at 37 °C in C+Y medium until  $A_{620} = 0.15$  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 neg-

ative 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  $A_{620} = 0.15$ . Bacteria were then incubated at 30 °C for 15 min before CSP was added. After 15 min, 1  $\mu\text{g}/\text{ml}$  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  $\mu\text{g}/\text{ml}$  final concentration. Blood plates were incubated O/N at 37 °C and 5%  $\text{CO}_2$  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 NdeI and Xho (*comGC*) and subcloned into pJWV25 and pACYC-Duet-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  $\mu\text{g}/\text{ml}$  ampicillin and 50  $\mu\text{g}/\text{ml}$  chloramphenicol at 37 °C until  $A_{600} = 0.5$ , and induced with 1 mM isopropyl  $\beta$ -D-thiogalactopyranoside for 3 h. Bacteria were spun down, and 1 $\times$  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  $\beta$ -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  $\mu\text{g}/\text{ml}$  ampicillin, 50  $\mu\text{g}/\text{ml}$  kanamycin, and 0.5 mM isopropyl  $\beta$ -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 $\times$  Z buffer (90 mM  $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$ , 40 mM  $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ , 6 mM NaOH,

10 mM  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 50 mM  $\beta$ -mercaptoethanol) and diluted until a final  $A_{600 \text{ nm}} = 0.3$ . Then, 1 ml of bacterial suspension was permeabilized by adding 100  $\mu\text{l}$  of chloroform and 50  $\mu\text{l}$  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  $\mu\text{l}$  of 0.4% *o*-nitrophenyl- $\beta$ -D-galactopyranoside in phosphate buffer (90 mM  $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$ , 40 mM  $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ ). The reaction was stopped by the addition of 500  $\mu\text{l}$  of 1 M  $\text{Na}_2\text{CO}_3$  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  $\beta$ -galactosidase activity was expressed in Miller units by using the following formula:  $1000 \times (A_{420 \text{ nm}} - 1.75 \times A_{550 \text{ nm}}) / (\text{incubation time (minutes)} \times \text{volume (1 ml)} \times A_{600 \text{ nm}})$ .

### 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 $\times$  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 (ComGC $\Delta$ 1–39) was cloned downstream of the His<sub>6</sub> 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 [<sup>13</sup>C]glucose and [<sup>15</sup>N]ammonium sulfate containing 50  $\mu\text{g}/\text{ml}$  kanamycin. The O/N culture was then diluted into fresh medium; cells were grown to  $A_{620} = 0.5$  at 37 °C and induced with 1 mM isopropyl  $\beta$ -D-1 thiogalactopyranoside for 3.5 h. Cells were harvested by centrifugation at 7000  $\times g$  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  $\times g$ , and supernatants were incubated with nickel-nitrilotriacetic 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 His<sub>6</sub> 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*

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  $\mu$ l. Experiments were performed in 5-mm tubes at a temperature of 283 K on Bruker AVIII 600 spectrometer operating at 600 MHz ( $^1\text{H}$  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 ( $^1\text{H}$  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  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC,  $^1\text{H}$ ,  $^{13}\text{C}$ -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  $^{15}\text{N}$ -TOCSY-HSQC and  $^{13}\text{C}$ -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  $^{15}\text{N}$  NOESY-HSQC and  $^{13}\text{C}$  NOESY-HSQC experiments recorded using mixing times of 150 and 80 ms, respectively. Residual dipolar couplings were obtained from a sample aligned in PEG ( $\text{C}_{12}\text{E}_5$ )/hexanol (Sigma: 76437/H13303) liquid crystal medium with a final PEG concentration of 4%. The couplings were measured in in-phase/anti-phase  $^1\text{H}$ ,  $^{15}\text{N}$  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**— $R_1$  and  $R_2$  relaxation rates were measured in using the 2D  $^1\text{H}$ ,  $^{15}\text{N}$  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  $R_1$  experiments and 0.016\*, 0.032, 0.048, 0.08\*, 0.112, 0.144\*, and 0.2 s for  $R_2$  (starred values are recorded in duplicates). Heteronuclear  $^1\text{H}$ - $^{15}\text{N}$  NOE values were determined from peak ratios between saturated/steady-state and a reference spectrum ( $I_{\text{ss}}/I_{\text{ref}}$ ) (57). Reduced spectral densities are calculated using Equations 1–4,

$$J(0) = \frac{1}{3d^2 + 4c^2} \left( 6R_2 - R_1 \left( 3 - \frac{18}{5} \left( \frac{\gamma_{\text{N}}}{\gamma_{\text{H}}} \right) (\text{NOE} - 1) \right) \right) \quad (\text{Eq. 1})$$

$$J(\omega_{\text{N}}) = \frac{4}{3d^2 + 4c^2} \left( R_1 \left( 1 - \frac{7}{5} \left( \frac{\gamma_{\text{N}}}{\gamma_{\text{H}}} \right) (\text{NOE} - 1) \right) \right) \quad (\text{Eq. 2})$$

$$J(0.87\omega_{\text{h}}) = \frac{4}{5d^2} \left( R_1 \left( \frac{\gamma_{\text{N}}}{\gamma_{\text{H}}} \right) (\text{NOE} - 1) \right) \quad (\text{Eq. 3})$$

$$d = \left( \frac{\mu_0 h \gamma_{\text{N}} \gamma_{\text{H}}}{8\pi^2} \right) \frac{1}{r_{\text{NH}}^3}, \quad c = \frac{\omega_{\text{N}} \Delta\sigma}{\sqrt{3}} \quad (\text{Eq. 4})$$

where  $R_1$  and  $R_2$  and NOE are the fitted relaxation rates or intensity ratio;  $\gamma_{\text{N}}$  and  $\gamma_{\text{H}}$  are the gyromagnetic ratios of N and H;  $\mu_0$  is the permeability of free space;  $h$  is Planck's constant;  $r_{\text{NH}}$  is the bond length vector of the amide NH bond (set to 1.02 Å);  $\omega_{\text{N}}$  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., S. E., 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

1. Rudel, T., Scheurerpflug, I., and Meyer, T. F. (1995) *Neisseria* PilC protein identified as type-4 pilus tip-located adhesin. *Nature* **373**, 357–359
2. Winther-Larsen, H. C., Hegge, F. T., Wolfgang, M., Hayes, S. F., van Putten, J. P., and Koomey, M. (2001) *Neisseria gonorrhoeae* PilV, a type IV pilus-associated protein essential to human epithelial cell adherence. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 15276–15281
3. Mattick, J. S. (2002) Type IV pili and twitching motility. *Annu. Rev. Microbiol.* **56**, 289–314
4. Maier, B., and Wong, G. C. (2015) How bacteria use type IV Pili machinery on surfaces. *Trends Microbiol.* **23**, 775–788
5. O'Toole, G. A., and Kolter, R. (1998) Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Mol. Microbiol.* **30**, 295–304
6. Kirn, T. J., Lafferty, M. J., Sandoe, C. M., and Taylor, R. K. (2000) Delineation of pilin domains required for bacterial association into microcolonies and intestinal colonization by *Vibrio cholerae*. *Mol. Microbiol.* **35**, 896–910
7. Aas, F. E., Wolfgang, M., Frye, S., Dunham, S., Løvold, C., and Koomey, M. (2002) Competence for natural transformation in *Neisseria gonorrhoeae*: components of DNA binding and uptake linked to type IV pilus expression. *Mol. Microbiol.* **46**, 749–760
8. Chen, I., and Dubnau, D. (2004) DNA uptake during bacterial transformation. *Nat. Rev. Microbiol.* **2**, 241–249
9. Laurenceau, R., Péhau-Arnaudet, G., Bacconnais, S., Gault, J., Malosse, C., Dujeancourt, A., Campo, N., Chamot-Rooke, J., Le Cam, E., Claverys, J. P.,

- and Fronzes, R. (2013) A type IV pilus mediates DNA binding during natural transformation in *Streptococcus pneumoniae*. *PLoS Pathog.* **9**, e1003473
10. Craig, L., Pique, M. E., and Tainer, J. A. (2004) Type IV pilus structure and bacterial pathogenicity. *Nat. Rev. Microbiol.* **2**, 363–378
  11. Craig, L., and Li, J. (2008) Type IV pili: paradoxes in form and function. *Curr. Opin. Struct. Biol.* **18**, 267–277
  12. Berry, J. L., and Pelicic, V. (2015) Exceptionally widespread nanomachines composed of type IV pilins: the prokaryotic Swiss Army knives. *FEMS Microbiol. Rev.* **39**, 134–154
  13. Giltner, C. L., Nguyen, Y., and Burrows, L. L. (2012) Type IV pilin proteins: versatile molecular modules. *Microbiol. Mol. Biol. Rev.* **76**, 740–772
  14. Melville, S., and Craig, L. (2013) Type IV pili in Gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* **77**, 323–341
  15. Rakotoarivonina, H., Jubelin, G., Hebraud, M., Gaillard-Martinie, B., Forano, E., and Mosoni, P. (2002) Adhesion to cellulose of the Gram-positive bacterium *Ruminococcus albus* involves type IV pili. *Microbiology* **148**, 1871–1880
  16. Gurung, I., Spielman, I., Davies, M. R., Lala, R., Gaustad, P., Biais, N., and Pelicic, V. (2016) Functional analysis of an unusual type IV pilus in the Gram-positive *Streptococcus sanguinis*. *Mol. Microbiol.* **99**, 380–392
  17. Gurung, I., Berry, J. L., Hall, A. M. J., and Pelicic, V. (2017) Cloning-independent markerless gene editing in *Streptococcus sanguinis*: novel insights in type IV pilus biology. *Nucleic Acids Res.* **45**, e40
  18. Barocchi, M. A., Ries, J., Zogaj, X., Hemsley, C., Albiger, B., Kanth, A., Dahlberg, S., Fernebro, J., Moschioni, M., Massignani, V., Hulthenby, K., Taddei, A. R., Beiter, K., Wartha, F., von Euler, A., *et al.* (2006) A pneumococcal pilus influences virulence and host inflammatory responses. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2857–2862
  19. Hilleringmann, M., Giusti, F., Baudner, B. C., Massignani, V., Covacci, A., Rappuoli, R., Barocchi, M. A., and Ferlenghi, I. (2008) Pneumococcal pili are composed of protofilaments exposing adhesive clusters of Rrg A. *PLoS Pathog.* **4**, e1000026
  20. Strom, M. S., Nunn, D. N., and Lory, S. (1993) A single bifunctional enzyme, PilD, catalyzes cleavage and *N*-methylation of proteins belonging to the type IV pilin family. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 2404–2408
  21. Karimova, G., Pidoux, J., Ullmann, A., and Ladant, D. (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5752–5756
  22. Sauvonnnet, N., Vignon, G., Pugsley, A. P., and Gounon, P. (2000) Pilus formation and protein secretion by the same machinery in *Escherichia coli*. *EMBO J.* **19**, 2221–2228
  23. Lacroix, E., Viguera, A. R., and Serrano, L. (1998) Elucidating the folding problem of  $\alpha$ -helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.* **284**, 173–191
  24. Drozdetskiy, A., Cole, C., Procter, J., and Barton, G. J. (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394
  25. Craig, L., Taylor, R. K., Pique, M. E., Adair, B. D., Arvai, A. S., Singh, M., Lloyd, S. J., Shin, D. S., Getzoff, E. D., Yeager, M., Forest, K. T., and Tainer, J. A. (2003) Type IV pilin structure and assembly: X-ray and EM analyses of *Vibrio cholerae* toxin-coregulated pilus and *Pseudomonas aeruginosa* PAK pilin. *Mol. Cell* **11**, 1139–1150
  26. Craig, L., Volkmann, N., Arvai, A. S., Pique, M. E., Yeager, M., Egelman, E. H., and Tainer, J. A. (2006) Type IV pilus structure by cryo-electron microscopy and crystallography: implications for pilus assembly and functions. *Mol. Cell* **23**, 651–662
  27. Gold, V. A., Salzer, R., Averhoff, B., and Kuhlbrandt, W. (2015) Structure of a type IV pilus machinery in the open and closed state. *Elife* **10**, 7554/eLife.07380
  28. Chen, I., Provvedi, R., and Dubnau, D. (2006) A macromolecular complex formed by a pilin-like protein in competent *Bacillus subtilis*. *J. Biol. Chem.* **281**, 21720–21727
  29. Piepenbrink, K. H., Maldarelli, G. A., Martinez de la Peña, C. F., Dingle, T. C., Mulvey, G. L., Lee, A., von Roseninge, E., Armstrong, G. D., Donnenberg, M. S., and Sundberg, E. J. (2015) Structural and evolutionary analyses show unique stabilization strategies in the type IV pili of *Clostridium difficile*. *Structure* **23**, 385–396
  30. Reardon, P. N., and Mueller, K. T. (2013) Structure of the type IVa major pilin from the electrically conductive bacterial nanowires of *Geobacter sulfurreducens*. *J. Biol. Chem.* **288**, 29260–29266
  31. Kolappan, S., Coureuil, M., Yu, X., Nassif, X., Egelman, E. H., and Craig, L. (2016) Structure of the *Neisseria meningitidis* Type IV pilus. *Nat. Commun.* **7**, 13015
  32. Friedrich, A., Prust, C., Hartsch, T., Henne, A., and Averhoff, B. (2002) Molecular analyses of the natural transformation machinery and identification of pilus structures in the extremely thermophilic bacterium *Thermus thermophilus* strain HB27. *Appl. Environ. Microbiol.* **68**, 745–755
  33. Graupner, S., Frey, V., Hashemi, R., Lorenz, M. G., Brandes, G., and Wackernagel, W. (2000) Type IV pilus genes pilA and pilC of *Pseudomonas stutzeri* are required for natural genetic transformation, and pilA can be replaced by corresponding genes from nontransformable species. *J. Bacteriol.* **182**, 2184–2190
  34. Stone, B. J., and Kwai, Y. A. (1999) Natural competence for DNA transformation by *Legionella pneumophila* and its association with expression of type IV pili. *J. Bacteriol.* **181**, 1395–1402
  35. Cehovin, A., Simpson, P. J., McDowell, M. A., Brown, D. R., Noschese, R., Pallett, M., Brady, J., Baldwin, G. S., Lea, S. M., Matthews, S. J., and Pelicic, V. (2013) Specific DNA recognition mediated by a type IV pilin. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 3065–3070
  36. Berry, J. L., Xu, Y., Ward, P. N., Lea, S. M., Matthews, S. J., and Pelicic, V. (2016) A comparative structure/function analysis of two type IV Pilin DNA receptors defines a novel mode of DNA binding. *Structure* **24**, 926–934
  37. Balaban, M., Bättig, P., Muschiol, S., Tirier, S. M., Wartha, F., Normark, S., and Henriques-Normark, B. (2014) Secretion of a pneumococcal type II secretion system pilus correlates with DNA uptake during transformation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E758–E765
  38. Luscombe, N. M., Laskowski, R. A., and Thornton, J. M. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.* **29**, 2860–2874
  39. Mindell, J. A., and Grigorieff, N. (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347
  40. Rohou, A., and Grigorieff, N. (2015) CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221
  41. Hohn, M., Tang, G., Goodyear, G., Baldwin, P. R., Huang, Z., Penczek, P. A., Yang, C., Glaeser, R. M., Adams, P. D., and Ludtke, S. J. (2007) SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* **157**, 47–55
  42. Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., and Ludtke, S. J. (2007) EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46
  43. Desfosses, A., Ciuffa, R., Gutsche, I., and Sachse, C. (2014) SPRING—an image processing package for single-particle based helical reconstruction from electron cryomicrographs. *J. Struct. Biol.* **185**, 15–26
  44. Sung, C. K., Li, H., Claverys, J. P., and Morrison, D. A. (2001) An rpsL cassette, janus, for gene replacement through negative selection in *Streptococcus pneumoniae*. *Appl. Environ. Microbiol.* **67**, 5190–5196
  45. Jarymowycz, V. A., and Stone, M. J. (2006) Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem. Rev.* **106**, 1624–1671
  46. Cascales, E., Bernadac, A., Gavioli, M., Lazzaroni, J. C., and Lloubes, R. (2002) Pal lipoprotein of *Escherichia coli* plays a major role in outer membrane integrity. *J. Bacteriol.* **184**, 754–759
  47. Prestegard, J. H., al-Hashimi, H. M., and Tolman, J. R. (2000) NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q. Rev. Biophys.* **33**, 371–424
  48. Yao, L., Ying, J., and Bax, A. (2009) Improved accuracy of  $^{15}\text{N}$ - $^1\text{H}$  scalar and residual dipolar couplings from gradient-enhanced IPAP-HSQC experiments on protonated proteins. *J. Biomol. NMR* **43**, 161–170

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

49. Ottiger, M., Delaglio, F., and Bax, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373–378
50. Cheung, M. S., Maguire, M. L., Stevens, T. J., and Broadhurst, R. W. (2010) DANGLE: a Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J. Magn. Reson.* **202**, 223–233
51. Güntert, P., and Buchner, L. (2015) Combined automated NOE assignment and structure calculation with CYANA. *J. Biomol. NMR* **62**, 453–471
52. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73
53. Theobald, D. L., and Wuttke, D. S. (2006) THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics* **22**, 2171–2172
54. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037–10041
55. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486
56. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by  $^{15}\text{N}$  NMR relaxation. *Biochemistry* **33**, 5984–6003
57. Ferrage, F., Reichel, A., Battacharya, S., Cowburn, D., and Ghose, R. (2010) On the measurement of  $(1)(5)\text{N}\{-\{1\}\text{H}\}$  nuclear Overhauser effects. 2. Effects of the saturation scheme and water signal suppression. *J. Magn. Reson.* **207**, 294–303

**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

*J. Biol. Chem.* 2017, 292:14134-14146.

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](https://doi.org/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>