A comparative structure/function analysis of two type IV pilin DNA receptors defines a novel mode of DNA-binding

Jamie-Lee Berry[1], Yingqi Xu[2], Philip N. Ward[3], Susan M. Lea[3], Stephen J. Matthews[2,*] and Vladimir Pelicic[1,*]

[1]MRC Centre for Molecular Bacteriology and Infection, Imperial College London, London SW7 2AZ, United Kingdom

[2]Centre for Structural Biology, Imperial College London, London SW7 2AZ, United Kingdom

[3]Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, United Kingdom

*Correspondence: s.j.matthews@imperial.ac.uk (S.J.M.), v.pelicic@imperial.ac.uk (V.P.)

ABSTRACT

DNA transformation is a widespread process allowing bacteria to capture free DNA by using filamentous nano-machines composed of type IV pilins. These proteins can act as DNA receptors as demonstrated by the finding that *Neisseria meningitidis* ComP minor pilin has intrinsic DNA-binding ability. ComP binds DNA better when it contains the DNA uptake sequence (DUS) motif abundant in this species genome, playing a role in its trademark ability to selectively take up its own DNA. Here, we report high-resolution structures for meningococcal ComP and *Neisseria subflava* $ComP_{sub}$, which recognize different DUS motifs. We show that they are structurally identical type IV pilins that pack readily into filament models and display a unique DD-region delimited by two disulfide bonds. Functional analysis of $ComP_{sub}$ defines a new mode of DNA-binding involving the DD-region, adapted for exported DNA receptors.

INTRODUCTION

Numerous bacterial species, defined as naturally competent, are able to capture free DNA from the environment and import it into their cytoplasm across formidable permeability barriers (Chen and Dubnau, 2004). Although, imported DNA can be used as a source of food or as a template for the repair of DNA damage (Chen and Dubnau, 2004), when the DNA is new and it is stably acquired, this process is called transformation since the bacteria are "transformed" by exhibiting novel phenotypic traits. Transformation allows competent bacteria to evolve rapidly by promoting transfer of DNA between different species, an important evolutionary process known as horizontal gene transfer (HGT) (Thomas and Nielsen, 2005).

Most naturally competent species import extracellular DNA using the same two-step process (Chen and Dubnau, 2004). The first step involves DNA uptake mediated by type IV filamentous (Tff) nano-machines composed of type IV pilins (Berry and Pelicic, 2015), either long filaments known as type IV pili (Tfp) or elusive competence (pseudo)pili. It is widely accepted (Chen and Dubnau, 2004; Maier et al., 2004) that filaments bind free DNA and pull it across the outer membrane (in Gram-negative bacteria) and/or peptidoglycan (in Gram-positive species). This scenario is supported by the finding that in species where *bona fide* retractable Tfp are involved, as in the human pathogens *Neisseria meningitidis* and *Neisseria gonorrhoeae*, there is no DNA uptake when pilus retraction is abolished (Brown et al., 2010; Wolfgang et al., 1998). In the second step, once DNA is in the periplasm/pseudo-periplasm it is bound by the ComE/ComEA receptor (Chen and Gotschlich, 2001; Provvedi and Dubnau, 1999; Seitz et al., 2014) and translocated across the cytoplasmic membrane by the Com machinery (Chen and Dubnau, 2004). Eventually, imported DNA is integrated in the chromosome in a RecA-dependent manner.

Until recently, no filament-localized DNA receptor had been identified and it was unknown how Tff nano-machines bound extracellular DNA. However, recent findings clearly showed that some type IV pilins can act as DNA receptors. We

discovered that the minor Tfp component ComP, which is identical in meningococci and gonococci and is crucial for their competence (Brown et al., 2010; Wolfgang et al., 1999), is the only one of four *N. meningitidis* type IV pilins with intrinsic DNA-binding ability (Cehovin et al., 2013). Furthermore, quantitative DNA-binding experiments (Berry et al., 2013; Cehovin et al., 2013) showed that although purified ComP can interact with any DNA sequence, it displays a preference for the 12 bp DNA uptake sequence (DUS) motif (Ambur et al., 2007). This motif is highly repeated in meningococcal and gonococcal genomes and enhances uptake by these species of DNA containing it (Goodman and Scocca, 1988). Selective uptake of their own DNA is a characteristic feature of Neisseriaceae (Frye et al., 2013) and Pasteurellaceae in which a different motif (termed USS) is found (Danner et al., 1980). This property is thought to protect these competent species from indiscriminate transformation by foreign DNA. Identification of ComP as the DUS receptor shed light on the long-standing mystery of how pathogenic *Neisseria* species manage to recognize and import their own DNA during transformation. The presence of ComP homologs in all other *Neisseria* species and most Neisseriaceae, some of which were shown to exhibit preferential uptake of their own DNA containing DUS variants (Frye et al., 2013), suggests that co-evolution of DUS variants and cognate ComP receptors is an elegant mechanism for modulating HGT between competent species sharing the same environmental niche (Berry et al., 2013). This scenario is further supported by the finding that DUS variants present in other *Neisseria* species, which differ from meningococcal DUS by as little as one bp, show sub-optimal transformation of *N. meningitidis* (Berry et al., 2013).

High-resolution structural information is necessary to advance our understanding of how the ComP family of type IV pilins recognize DNA. Although a NMR analysis generated a low-resolution global fold for meningococcal ComP and highlighted an electropositive surface important for DNA-binding and transformation (Cehovin et al., 2013), the finer details were lacking. We therefore embarked upon a

comparative structural analysis of two ComP orthologs. In the present study, we shine new light on this poorly understood and fascinating phenomenon (i) by reporting high-resolution 3D structures for ComP from *N. meningitidis* and $ComP_{sub}$ from the non-pathogenic species *Neisseria subflava*, which is a common inhabitant of the human upper respiratory tract, and (ii) by performing an in-depth functional analysis of $ComP_{sub}$ DNA-binding ability, which shows specificity for DUS_{var1} differing from meningococcal DUS by one bp.

RESULTS

DNA-binding with a preference for their cognate DUS is a conserved property in ComP homologs

Although purified $ComP_{sub}$ was previously shown to bind DNA (Berry et al., 2013), quantitative DNA-binding data was needed to strengthen the notion that all ComP homologs bind their cognate DUS specifically. To overcome previous protein stability problems, we fused the 118 aa-long soluble portion of $ComP_{sub}$ (Fig. 1) to non-cleavable N-terminal maltose-binding protein (MBP) or hexahistidine tag (His_6). This allowed us to purify well-folded proteins, which we used to test whether $ComP_{sub}$ has a higher affinity for its cognate DUS_{var1}. First, we used acrylamide electrophoretic mobility shift assays (EMSA) to perform competition reactions. We assessed the effect of an excess of unlabelled double-stranded (ds) primer on a pre-formed complex between purified MBP-$ComP_{sub}$ and a biotinylated DUS_{var1} ds primer, which produces a characteristic shift on gel (Fig. 2A). While unlabelled DUS_{var1} efficiently out-competed bound biotinylated DUS_{var1} in a dose-dependent manner, as demonstrated by the gradual and eventually complete disappearance of the biotinylated complex, a scrambled SUD primer (in which every second base is altered) had no effect (Fig. 2A). Next, the affinity of His_6-$ComP_{sub}$ for DUS_{var1} and two different scrambled primers was quantified in real-time using surface plasmon resonance (SPR), as previously done for meningococcal ComP (Cehovin et al., 2013). In brief, equivalent amounts of ds biotinylated DUS_{var1}, SUD and SDU primers were coupled to adjacent channels on a neutravidin-coated sensor chip. Increasing amounts of purified His_6-$ComP_{sub}$ (10, 25, 50, 100 or 200 μM) were then injected and the responses at equilibrium (R_{eq}) were measured for each protein concentration (Fig. 2B). This clearly showed that $ComP_{sub}$ binds DNA in a dose-dependent fashion, and that the affinity for DUS_{var1} is much higher than the affinity for SUD or SDU, as indicated by the higher R_{eq} at each protein concentration. Unlike binding to the

scrambled primer, binding to DUS_{var1} was approaching saturation, allowing us to estimate (using a non linear regression - least squares fit) a dissociation constant (Kd) of 52.7 ± 2.2 µM. Taken together, these findings confirm that DNA binding, which is tighter in the presence of their cognate DUS, is a conserved property in ComP orthologs.

ComP orthologs share similar 3D structures with a highly distinctive DD-region stabilized by two disulfide bonds

Since there is no high-resolution structural data for this class of DNA-binding pilins, we first endeavoured to solve the 3D structure of its defining member, *N. meningitidis* ComP. To facilitate purification and crystallization, we used a synthetic *comP* gene, codon-optimized for expression in *E. coli*, and fused the 115 aa-long soluble portion of ComP to a MBP modified to promote crystallisation by surface entropy reduction (Moon et al., 2010). This soluble portion of ComP excludes the 6 aa-long leader peptide (Fig. 1A), which is processed by the dedicated prepilin peptidase PilD (Brown et al., 2010), and the first 28 residues of the mature protein, most of which correspond to the hydrophobic residues that form the protruding part (α1N) of N-terminal α1-helix in type IV pilins (Berry and Pelicic, 2015). The MBP-ComP protein crystallized readily in multiple conditions. After optimizing the best diffracting crystals, we collected a complete dataset on crystals formed in the spacegroup $P2_12_12_1$, which diffracted to a resolution of 1.43 Å (Table S1), and solved the structure of the fusion protein (Fig. S1). As can be seen in Fig. 3, the ComP moiety adopts the classical type IV pilin fold (Giltner et al., 2012) with the C-terminal part of the long α1-helix (α1C), packed against a β-sheet of four anti-parallel β-strands. This conserved core is highly similar to that of the major pilin PilE (Parge et al., 1995) and the minor pilin PilX (Helaine et al., 2007) (Fig. 3B). Subtle differences in ComP's conserved core include the lack of curvature in α1C (as previously observed for PilX), and the rather long loop (14 residues) connecting β1 and β2, the first two β-strands of the β-sheet.

In contrast, major differences exist in the structurally variable "edges" that usually distinguish type IV pilins (Giltner et al., 2012), the α1β1-loop connecting α1 and β1 and the C-terminal D-region delimited by a disulfide bond (hence its name). These regions in ComP are unique (Fig. 3). While ComP's α1β1-loop is an extended unstructured region, its D-region is particularly striking. Unlike in (most) other type IV pilins, where one disulfide bond stabilizes the C-terminus of the protein by stapling it to β4, in ComP the C-terminus forms a long unstructured loop held in place across the face of the β-sheet by two disulfide bonds between the four Cys residues in the protein (C_{76}-C_{125} and C_{118}-C_{139}). The C_{118}-C_{139} bond delimits the loop by stapling its C-terminus back to β4 and is therefore equivalent to the single disulfide bond found in other pilins (Fig. 3B). In contrast, C_{76}-C_{125} is unique and pins the middle of the loop back to β1. To highlight the unique nature of this region, we termed it the DD-region. Finally, three of the last five residues of the protein form a short β-strand (β5), which is incorporated into the β-sheet (Fig. 3A).

To determine the structural relationship between ComP orthologs that recognize different DUS, we also solved the structure of $ComP_{sub}$ from *N. subflava*. The 118 aa-long soluble portion of this protein, which has a binding preference for DUS_{var1} (see Fig. 2), displays ~52% sequence identity to ComP (Fig. 1). Since we could not obtain crystals for the MBP-$ComP_{sub}$ fusion because the protein was too soluble, we decided to explore whether a solution structure determination by NMR would be possible for His_6-$ComP_{sub}$. We isotopically labelled our His_6-$ComP_{sub}$ with ^{13}C and ^{15}N for NMR assignment and obtained a high-resolution NOE-derived structure in solution (Table S1). The $ComP_{sub}$ structures within the NMR ensemble superpose well onto each other, with a root mean square deviation (rmsd) of 0.28 Å for all backbone atoms, which suggests that there is no significant flexibility in the structure (Fig. 4A). As can be seen in Fig. 4B, $ComP_{sub}$ adopts a 3D structure highly similar to ComP. The two structures align over their whole length with a rmsd of 2.41

Å for all backbone atoms (Fig. 4C). Importantly, the distinctive features highlighted in ComP are also present in $ComP_{sub}$ - namely, the unstructured α1β1-loop, the long β1-β2 loop and the DD-region that is held in place across the face of the β-sheet by two disulfide bonds between the four Cys residues (C_{76}-C_{127} and C_{118}-C_{141}) in $ComP_{sub}$.

Taken together, these structural findings show that ComP orthologs adopt very similar 3D structures and present a unique structural feature previously unreported in type IV pilins, *i.e.* a DD-region that sits on top of the β-sheet and is delimited by two disulfide bonds between four conserved Cys residues.

ComP orthologs bind DNA mainly via their DD-region and β1-β2 loop, which are predicted to be exposed on the surface of Tfp

As already mentioned, ComP's DNA-binding activity has been characterized to some extent by NMR (Cehovin et al., 2013). However, this analysis was limited by relatively poor NMR data with many key resonances broadened through conformational exchange, which rendered them unassignable and precluded the calculation of a high-resolution solution structure for ComP. In contrast, the NMR data for $ComP_{sub}$ is of much higher quality, leading to complete assignments of ^{1}H,^{15}N and ^{13}C nuclei and eventually to the calculation of a high-resolution solution structure (Fig. 4). We therefore characterized by NMR the atomic resolution features of DNA-binding by His_6-$ComP_{sub}$. We titrated increasing amounts of an unlabelled DUS_{var1} ds primer into isotopically labelled His_6-$ComP_{sub}$ and monitored chemical shift perturbations (CSP). This analysis revealed significant and specific CSP (Fig. 5A), which can be attributed to the interaction between the two molecules. Detailed inspection of ^{1}H-^{15}N HSQC spectra revealed that the $ComP_{sub}$ residues undergoing CSP in the presence of DUS_{var1} coalesce into three main contiguous "patches" (Fig. 5B). The first patch corresponds to the large β1-β2 loop together with β2, the second patch corresponds

to the first half of the DD-region, while the last patch corresponds to the second half of the DD-region. A small part of the α1β1-loop is also involved. Strikingly, these regions of $ComP_{sub}$ form an almost vertical "stack" along one face of the protein (Fig. 5C). The residues experiencing CSP with low concentrations of DNA, and are thus likely to contact DNA first, lie almost exclusively in the DD-region and the large β1-β2 loop (Fig. 5). The residues experiencing CSP at higher concentrations of DNA are part of the β-sheet, which is initially covered by the DD-region and β1-β2 loop in the 3D structure. This suggests that there might be conformational changes in the protein upon DNA-binding leading to multiple binding modes.

The interaction between $ComP_{sub}$ and DUS_{var1} led to the broadening of many signals even under saturating DNA concentrations, suggesting that even in the bound state there are significant conformational dynamics. Since this precluded the direct measurement of intermolecular NOEs, we used the HADDOCK software (de Vries et al., 2007; Dominguez et al., 2003) to generate a structural model for the $ComP_{sub}$-DUS_{var1} complex. All the bases of DUS_{var1} were considered as important, because the mutagenesis of even a single base significantly impairs DNA-binding and/or transformation (Berry et al., 2013; Frye et al., 2013). In contrast, active residues in $ComP_{sub}$ were defined as those experiencing CSP at DNA concentrations below 20 µM (1/5 ratio of DNA to protein), to avoid giving weight to residues showing CSP at later titration points. The lowest overall energy cluster contained ten structures with a rmsd of 2.5 Å and an overall HADDOCK score of 70.7 ± 14.5 (Fig. 6A). Complex formation in this cluster resulted in an average buried surface area of 1,558.1 ± 137.3 $Å^2$. Closer inspection of this complex reveals that the DNA docks onto the vertical stack of residues on $ComP_{sub}$ just above the ledge feature introduced by the DD-region. The DD-region, β1-β2 loop and α1-β1 loop intercalate with successive grooves of the DNA establishing contacts with multiple bases of DUS_{var1} (Fig. 6A). Finally, using the MODELLER software (Webb and Sali, 2014), we

produced a full-length $ComP_{sub}$ model using the gonococcal PilE structure (Craig et al., 2006) as a template for the missing N-terminal α1N helix and we were able to model its packing within a Tfp structural model (Craig et al., 2006). This revealed that $ComP_{sub}$ not only fits readily into the filaments, but that the regions of the protein involved in DNA-binding are clearly exposed on the surface of Tfp (Fig. 6B).

Taken together, these findings suggest that $ComP_{sub}$ binds DNA using unique structural features conserved in this class of type IV pilins and exposed on the surface of the Tfp. These features, *i.e.* the large β1-β2 loop and DD-region, define a new DNA-binding motif differing dramatically from well-known motifs (Luscombe et al., 2000).

DISCUSSION

Natural transformation is a widespread biological property playing a key role in bacterial physiology. Although, it has been the subject of intense study for almost a century (Griffith, 1928), its very first step, during which Tff nano-machines bind free DNA to subsequently promote its uptake, remains one of the least understood. The discovery that ComP is the only purified meningococcal pilin capable of binding DNA showed that type IV pilins (the Tff subunits) can act as DNA receptors (Cehovin et al., 2013), highlighting a new property for these versatile molecular modules (Giltner et al., 2012). Furthermore, the finding that ComP shows better binding to the 12 bp DUS sequence motif (Ambur et al., 2007), which is highly abundant in pathogenic *Neisseria* species genomes and enhances uptake of their own DNA (Goodman and Scocca, 1988), solved a long-standing mystery. It revealed an elegant mechanism for limiting transformation by foreign DNA that is widespread in competent Neisseriaceae (Frye et al., 2013). A significant barrier to our understanding of how and why the ComP family can bind DNA is the absence of high-resolution structural information for this class of type IV pilins. In this report, we have addressed this limitation, which led to significant findings.

We first provide evidence suggesting that all ComPs are DNA-binding pilins displaying binding preference for their cognate DUS. Our previous finding that meningococcal ComP has a higher affinity for its cognate DUS (Cehovin et al., 2013) could readily be extended to other *Neisseria* species that contain the same DUS and encode ComPs with ~80-100% aa identity (such as *N. gonorrhoeae*, *N. lactamica*, *N. polysaccharea* and *N. cinerea*). However, to generalize our findings to this entire class of proteins, it was still to be demonstrated that a ComP from a more distant Neisseriaceae, which usually display ~30-50% aa identity to meningococcal ComP (Cehovin et al., 2013), would show a similar binding preference to its cognate DUS that differs from meningococcal DUS by 1-4 bases (Frye et al., 2013). A previous competition EMSA with purified $ComP_{sub}$ fell short of this goal, although it showed a

slightly better competition by its cognate DUS_{var1} (Berry et al., 2013). No quantitative DNA-binding data could be obtained because of protein stability/folding issues. These problems were solved here by using novel expression/purification strategies. Using SPR, the affinity of $ComP_{sub}$ for its cognate DUS_{var1} was found to be much higher than its affinity for scrambled primers. The Kd of $ComP_{sub}$ for DUS_{var1} (53 µM) was found to be comparable to the Kd of meningococcal ComP for its cognate DUS (29 µM) measured previously (Cehovin et al., 2013). However, it is worth noting that these affinities, which are significantly lower than those for "classical" sequence-specific DNA-binding proteins such as transcription factors, might be under-estimated due to the use of purified proteins. Indeed, it is not unlikely that ComP's affinity for DNA is higher when this protein is in its natural location within a Tfp, and would be further increased by the expected incorporation of multiple ComP subunits within a single filament. Together, these could cooperate to recognise one target DNA molecule with very high affinity and specificity.

We have also derived high-resolution structures for ComP and $ComP_{sub}$ orthologs. Consistent with the notion that they are minor components of Tfp (Aas et al., 2002; Brown et al., 2010), these structures revealed that ComP and $ComP_{sub}$ are *bona fide* type IV pilins and can pack efficiently within the available model for Tfp (Craig et al., 2006). Interestingly, these proteins also exhibit a distinctive structural feature exposed on the surface of the pilus, the DD-region delimited by two disulfide bonds, which sits across the β-sheet. The fact that this structural motif is not found in the dozens of type IV pilin structures available in the databases (Giltner et al., 2012) is consistent with ComP orthologs being (so far) the only known type IV pilins that bind DNA. Considering that (i) the 3D structures of ComP and $ComP_{sub}$ that share 50% of their residues are highly similar, (ii) there is significant sequence homology even with more distant orthologs (~30% aa are identical) (Cehovin et al., 2013), and (iii) all the DUS motifs share a conserved core, it is likely that all ComP orthologs will display the same 3D structure. This assumption is supported by homology modeling

(using MODELLER and ComP as a template) of the more distant $ComP_{Kor}$ ortholog from *Kingella oralis*, which shares only 27% aa identity with ComP and is expected to recognize a DUS differing from meningococcal DUS by three bases (Fig. S2A). $ComP_{Kor}$ structural model was found to be virtually identical to meningococcal ComP structure since they align over their whole length with a rmsd of 0.26 Å for all backbone atoms (Fig. S2B). Once other DNA-binding pilins are identified (that are not homologous to ComP), it will be interesting to explore whether they share a similar structure to ComP and whether the DD-region is a conserved DNA-binding motif. This will be particularly interesting once the USS receptor in Pasteurellaceae is identified, because no ComP homolog is present in these species and their USS bears no resemblance to DUS motifs.

The last, and perhaps most important, finding in this study is that the mode in which ComP orthologs interact with DNA has not been seen before. Rather than well-known and widespread DNA-binding motifs (helix-turn-helix, zinc finger, leucine zipper *etc.*) (Luscombe et al., 2000), ComP orthologs exploit a series of residues in two distinct domains, *i.e.* the DD-region and the tip of the β1-β2 loop (although a small part of the α1-β1 loop is also involved), to establish contacts with multiples bases of DUS in successive grooves of the ds DNA. These residues form an almost vertical stack on a face of the protein predicted to be exposed on the surface of the filaments, which was previously implicated in meningococcal ComP binding to DUS (Cehovin et al., 2013). As previously noted, this surface is highly positively charged (Fig. S3) and is likely to be involved initially in electrostatic attraction of the negatively charged DNA. Interestingly, the DD-region and the β1-β2 loop, which are flanked by almost invariable residues, are among those regions that differ the most between ComP orthologs (see Fig. 1 and Fig. S2). This is likely to be the reason why these proteins recognize different DUS motifs. It is now possible to propose a model for the mode of action of these DNA receptors in natural transformation (Fig. 7). When

Neisseriaceae encounter free DNA, they might attract it, electrostatically at first, and use ComP subunits in their Tfp to "scan" it for the presence of their cognate DUS. Once a cognate DUS has been recognized (it is unknown at this stage which specific aa differences in different ComPs are responsible for their different DUS specificities), ComP "docks" onto it using the vertical stack residues exposed on the filament surface, which is probably accompanied by conformational changes further increasing the strength and specificity of the interaction. Upon rotation of the pilus during PilT-powered retraction (Nivaskumar et al., 2014), the DNA docked at ComP anchor point(s) might wrap around the Tfp following the previously noted grooves on their corrugated surface (Craig et al., 2006). Eventually, upon translocation across the secretin pore, this would result in DNA uptake.

In conclusion, by providing high-resolution structural information for the ComP class of proteins, this study has shed light on the atomic basis for their DNA-binding ability, which is yet an additional property for the highly versatile type IV pilins. This has led to the discovery of a novel DNA-binding domain, which is consistent with the fact that members of the ComP family are, to best of our knowledge, the only known surface-exposed DNA receptors that bind specific DNA sequences.

EXPERIMENTAL PROCEDURES

Protein production and purification of *N. meningitidis* ComP

A synthetic gene, codon-optimized for *E. coli* expression, encoding ComP from *N. meningitidis* 8013 (Rusniok et al., 2009) was synthesized by GeneArt. The portion of optimized *comP* encoding residues 29-143 from the mature protein was PCR-amplified using opt*comP*-F and opt*comP*-R primers (Table S2), cut with *Eco*RI and *Hin*dIII and cloned into the pMALX(E) vector cut with the same enzymes. This resulted in the fusion of the soluble portion of ComP with a modified MBP carrier containing several mutations designed to promote crystallization (Moon et al., 2010). The resulting plasmid was verified by sequencing and transformed into chemically competent *E. coli* SHuffle T7 express cells (New England Biolabs), which enables formation of disulfide bonds in the cytoplasm. A single colony was transferred to 5 ml LB medium (Difco) supplemented with ampicillin (100 $\mu g.ml^{-1}$) and allowed to grow to saturation at 30°C overnight in an orbital shaker. The following day, this pre-culture was then used to inoculate 1 l of fresh LB-ampicillin and grown at 30°C in an orbital shaker until the OD_{600} reached 0.6-0.8. The culture was then placed into a refrigerated orbital shaker set at 16°C and allowed to cool for 30 min, before adding 0.4 mM IPTG (Merck Chemicals) to induce protein production during 16 h. Cells were then harvested by centrifugation at 8,000 *g* for 20 min and subjected to one freeze/thaw cycle in lysis buffer A [50 mM Tris-Cl pH 8, 100 mM NaCl, 1 x SIGMAFAST EDTA-free protease inhibitor cocktail (Sigma)]. This lysate was further disrupted by repeated cycles of sonication, *i.e.* pulses of 5 sec on and 5 sec off during 3-5 min, until the cell suspension was visibly less viscous. The cell lysate was then centrifuged for 30 min at 18,000 *g* to remove cell debris. The clarified lysate was then passed using an Akta Purifier FPLC through a 5 ml MBPTrap HP column (both from GE Healthcare), pre-equilibrated in lysis buffer A, in order to bind MBP-ComP.

The column was then washed extensively with lysis buffer A to remove unbound material before the fusion protein was eluted using TSM buffer (50 mM Tris-Cl pH 8, 100 mM NaCl, 10 mM maltose). The affinity-purified MBP-ComP was further purified by gel-filtration chromatography on a HiLoad 16/60 Superdex 200 column, using TSM buffer for elution.

Protein production and purification of *N.* subflava ComP$_{sub}$

For assessing the DNA-binding activity of ComP$_{sub}$ by EMSA, the portion of *N. subflava comP* encoding residues 29-146 from the mature ComP$_{sub}$ protein was PCR-amplified from *N. subflava* NJ9703 genomic DNA (Marri et al., 2010) using *comP*$_{sub}$-pMalF and *comP*$_{sub}$-pMalR primers (Table S2) and cloned as above into the pMALX(E) vector. This resulted in the fusion of the soluble portion of ComP$_{sub}$ with a non-cleavable MBP carrier. The fusion MBP-ComP$_{sub}$ protein was purified similarly to MBP-ComP.

For determining the 3D structure of ComP$_{sub}$, the above portion of *N. subflava comP* was amplified using his*comP*$_{sub}$-pETF and his*comP*$_{sub}$-pETR primers (Table S2) and cloned into the pET-28b vector (Novagen). The forward primer was designed to fuse a non-cleavable N-terminal His$_6$ tag to ComP$_{sub}$. This resulting plasmid was verified by sequencing and transformed into chemically competent *E. coli* SHuffle T7 express cells. A single colony was transferred to 2 ml of LB supplemented with 50 μg.ml^{-1} kanamycin and grown at 30°C to OD$_{600}$ of ~0.5. This pre-culture was back-diluted 1:50 into 10 ml M9 minimal medium, supplemented with a mixture of vitamins and trace elements. This was grown to saturation overnight at 30°C in an orbital shaker, then back-diluted 1:500 into 1 l of the same medium containing ^{13}C D-glucose and ^{15}N NH_4Cl for isotopic labelling. Cells were grown in an orbital shaker at 30°C until the OD$_{600}$ reached 0.8, before adding 0.4 mM IPTG to induce protein production overnight at 30°C. Cells were then harvested and disrupted as above in

lysis buffer B (50 mM Na_2HPO_4/NaH_2PO_4 pH 7.4, 500 mM NaCl, 20 mM imidazole, 1 x SIGMAFAST EDTA-free protease inhibitor cocktail). Two ml of Ni-NTA agarose resin (Qiagen), pre-washed in lysis buffer B, was then added to the clarified lysate and incubated for 2 h at 4°C with gentle agitation. This chromatography mixture was then filtered through a Poly-Prep gravity-flow column (BioRad) and washed several times with lysis buffer B, before eluting the protein with lysis buffer B containing 500 mM imidazole. The affinity-purified His_6-$ComP_{sub}$ was further purified by gel-filtration chromatography on an Akta Purifier using a Superdex 75 10/300 GL column (GE Healthcare), and simultaneously buffer-exchanged into 50 mM Na_2HPO_4/NaH_2PO_4 pH 6, 50 mM NaCl.

Crystallisation and structure determination of meningococcal ComP

Purified MBP-ComP was concentrated to 7, 14 and 28 $mg.ml^{-1}$ and placed through sitting-drop vapour diffusion crystallisation trials in MRC 2 well crystallization plates (Hampton Research) using a wide range of commercially available kits. Trials were initially conducted using 100 nl of protein and 100 nl of mother liquor, over a reservoir of 80 μl of mother liquor. The trials produced a large number of initial hits, which were varied and optimized to yield larger and better diffracting crystals, by altering the pH, precipitant concentration and drop size. The crystals used for structure determination were obtained when the purified protein at 14 $mg.ml^{-1}$ was mixed 1:1 with crystallisation liquor containing 0.1 M sodium cacodylate pH 6.5 and 25% PEG 4000, and left to equilibrate with the reservoir solution at 20°C. Crystals were cryo-protected using crystallisation liquor containing 20% ethylene glycol and flash-cooled in liquid nitrogen. Data were collected on beamline I02 at Diamond Light Source (Oxfordshire UK), and processed with xia2 (Winter, 2010). Molecular replacement was carried out using Phaser MR (McCoy et al., 2007) with MBP structure (PDB ID 1HSJ) as a search model. ComP was then built manually into the remaining density

using Bucaneer (Cowtan, 2006) and Coot (Emsley et al., 2010). Multiple iterative rounds of refinement and model building were carried out using Phenix (Adams et al., 2010) and Coot, resulting in a final R_{work}/R_{free} of 0.16/0.19.

NMR structure determination of $ComP_{sub}$

Isotopically labelled purified His_6-$ComP_{sub}$ was concentrated to ~750 µM in NMR buffer (50 mM Na_2HPO_4/NaH_2PO_4 pH 6, 50 mM NaCl, 10% D_2O). A full set of triple resonance NMR spectra was recorded on a Bruker Avance III 800MHz spectrometer equipped with triple resonance cryoprobes at 295 K, and processed with NMRPipe (Delaglio et al., 1995). Backbone assignments were completed using a combination of HBHA, HNCACB, HNCO, HN(CA)CO, and CBCA(CO)NH experiments using NMRView (One Moon Scientific) as previously described (Johnson and Blevins, 1994). Side-chain resonance assignments were obtained from a combination of CC(CO)NH, HC(C)H-TOCSY and (H)CCH-TOCSY experiments using an in-house software developed within NMRView (Marchant et al., 2008). Distance restraints were obtained from 3D $^{1}H^{1}H^{15}N$-NOESY and $^{1}H^{1}H^{13}C$-NOESY spectra and used for structure calculations in ARIA 2.3 (Rieping et al., 2007), along with dihedral angle restraints obtained from chemical shift values calculated using the TALOS+ server (Shen et al., 2009). During later rounds of calculations, disulfide restraints for C_{76}-C_{127} and C_{118}-C_{141} were introduced, after these had already formed in the calculated structures. For each round of calculations, 20 structures were calculated over eight iterations. In the final iteration, the 20 lowest energy structures were submitted to a water refinement stage to form the final structural ensemble.

NMR chemical shift perturbation analysis in $ComP_{sub}$ upon DNA-binding

The residues in $ComP_{sub}$ important for DNA-binding were identified by NMR. Complementary primers corresponing to DUS_{var1} were first dissolved in NMR buffer to a final concentration of 6 mM, combined in equimolar amounts, heated 5 min at

95°C for and left to cool overnight to produce 3 mM ds target DNA. The ds primer was washed with NMR buffer using G-25 spin columns (GE Healthcare) to remove residual buffer components from the oligonucleotide synthesis. The ds target DNA was titrated into a 100 µM sample of His_6-$ComP_{sub}$ in NMR buffer, at increasing concentrations (4, 5, 10, 20, 40, 80 and 120 µM), and two-dimensional $^{1}H^{15}N$-HSQC spectra were recorded at each titration point revealing changes to backbone chemical shifts.

DNA-binding assays

DNA-binding by $ComP_{sub}$ was assessed using EMSA or SPR. EMSA were performed as follows (Berry et al., 2013; Cehovin et al., 2013). Biotin-labelled and non-labelled ds primers corresponding to DUS_{var1} or scrambled sequences (Table S2) were prepared by mixing equimolar amounts of two complementary oligonucleotides in 50 mM Tris pH 8, 100 mM NaCl, heating for 5 min at 95°C and leaving to cool overnight. Purified MBP-$ComP_{sub}$ was prepared in the same buffer. A MBP-$ComP_{sub}$/DUS_{var1} complex was generated by mixing 5 fmol of biotinylated DUS_{var1} ds primer and 0.8 µM MBP-$ComP_{sub}$ (except for a DNA-only control) in 20 µl of 20 mM Tris-Cl pH 8, 50 mM NaCl, 2.5 mM Mg^{2+}, and incubating for 20 min at ambient temperature. Increasing concentrations of unlabelled competitor DNA (0.7, 2.8, 11.25, 45 and 180 pmol) were then added to the DNA-binding reactions, which were further incubated for 20 min at ambient temperature, before being analyzed by native gel electrophoresis as described previously (Cehovin et al., 2013).

SPR was performed as follows (Berry et al., 2013; Cehovin et al., 2013). Equivalent amounts of biotin-labelled ds primers, prepared as above in 20 mM Tris pH 8, 150 mM NaCl, were coupled to neutravidin on the surface of a ProteOn NLC sensor chips resulting in ~265 RU as assessed on a ProteOn XPR36 protein interaction array system instrument (BioRad). SPR was then performed by passing

10, 25, 50, 100 and 200 μM His_6-$ComP_{sub}$ (in 20 mM Tris pH 8, 150 mM NaCl, 0.05% Tween-20) across the six available analyte channels of the chip at 60 $\mu l.min^{-1}$, and the responses at equilibrium (R_{eq}) were recorded. A control trace was also collected using an empty ligand channel and used to normalize for non-specific binding effects. Four independent analyte injections were performed, with a regeneration step performed between each using 0.5 M NaCl. All the experiments were carried out at 25°C.

ACCESSION NUMBERS

The X-ray crystal structure of ComP and NMR structure of $ComP_{sub}$ have been deposited in the Protein Data Bank under ID codes 5HZ7 and 2NBA, respectively.

AUTHOR CONTRIBUTIONS

S.J.M. and V.P. designed and directed the research. All the experiments were done by J.L.B. Y.X helped with NMR, P.N.W. and S.M.L. helped with SPR. J.L.B., S.J.M. and V.P. wrote the paper.

ACKNOWLEDGMENTS

This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) and supported by the Wellcome Trust (via a Senior Investigator Award and a multi-user equipment grant to S.J.M.). We are grateful to Vivianne Goosens and Sophie Helaine (both from Imperial College London) for critical reading ot the manuscript. We are grateful to Nathan J. Weyand (University of Arizona) for the gift of *N. subflava* genomic DNA.

REFERENCES

Aas, F.E., Lovold, C., and Koomey, M. (2002). An inhibitor of DNA binding and uptake events dictates the proficiency of genetic transformation in *Neisseria gonorrhoeae*: mechanism of action and links to type IV pilus expression. Mol. Microbiol. *46*, 1441-1450.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D: Biol. Crystallogr. *66*, 213-221.

Ambur, O.H., Frye, S.A., and Tønjum, T. (2007). New functional identity for the DNA uptake sequence in transformation and its presence in transcriptional terminators. J. Bacteriol. *189*, 2077-2085.

Berry, J.L., Cehovin, A., McDowell, M.A., Lea, S.M., and Pelicic, V. (2013). Functional analysis of the interdependence between DNA uptake sequence and its cognate ComP receptor during natural transformation in *Neisseria* species. PLoS Genet. *9*, e1004014.

Berry, J.L., and Pelicic, V. (2015). Exceptionally widespread nano-machines composed of type IV pilins: the prokaryotic Swiss Army knives. FEMS Microbiol. Rev. *39*, 134-154.

Brown, D., Helaine, S., Carbonnelle, E., and Pelicic, V. (2010). Systematic functional analysis reveals that a set of 7 genes is involved in fine tuning of the multiple functions mediated by type IV pili in *Neisseria meningitidis*. Infect. Immun. *78*, 3053-3063.

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.*, et al.* (2013). Specific DNA recognition mediated by a type IV pilin. Proc. Natl. Acad. Sci. USA *110*, 3065-3070.

Chen, I., and Dubnau, D. (2004). DNA uptake during bacterial transformation. Nat. Rev. Microbiol. *2*, 241-249.

Chen, I., and Gotschlich, E.C. (2001). ComE, a competence protein from *Neisseria gonorrhoeae* with DNA-binding activity. J. Bacteriol. *183*, 3160-3168.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. Sect. D: Biol. Crystallogr. *62*, 1002-1011.

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.

Danner, D.B., Deich, R.A., Sisco, K.L., and Smith, H.O. (1980). An eleven-base-pair sequence determines the specificity of DNA uptake in *Haemophilus* transformation. Gene *11*, 311-318.

de Vries, S.J., van Dijk, A.D., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T., and Bonvin, A.M. (2007). HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. Proteins *69*, 726-733.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277-293.

Dominguez, C., Boelens, R., and Bonvin, A.M. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. *125*, 1731-1737.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. Sect. D: Biol. Crystallogr. *66*, 486-501.

Frye, S.A., Nilsen, M., Tønjum, T., and Ambur, O.H. (2013). Dialects of the DNA uptake sequence in *Neisseriaceae*. PLoS Genet. *9*, e1003458.

Giltner, C.L., Nguyen, Y., and Burrows, L.L. (2012). Type IV pilin proteins: versatile molecular modules. Microbiol. Mol. Biol. Rev. *76*, 740-772.

Goodman, S.D., and Scocca, J.J. (1988). Identification and arrangement of the DNA sequence recognized in specific transformation of *Neisseria gonorrhoeae*. Proc. Natl. Acad. Sci. USA *85*, 6982-6986.

Griffith, F. (1928). The significance of pneumococcal types. J. Hyg. (Lon) *27*, 113-159.

Helaine, S., Dyer, D.H., Nassif, X., Pelicic, V., and Forest, K.T. (2007). 3D structure/function analysis of PilX reveals how minor pilins can modulate the virulence properties of type IV pili. Proc. Natl. Acad. Sci. USA *104*, 15888-15893.

Johnson, B.A., and Blevins, R.A. (1994). NMR View: A computer program for the visualization and analysis of NMR data. J. Biomol. NMR *4*, 603-614.

Luscombe, N.M., Austin, S.E., Berman, H.M., and Thornton, J.M. (2000). An overview of the structures of protein-DNA complexes. Genome Biol. *1*, REVIEWS001.

Maier, B., Chen, I., Dubnau, D., and Sheetz, M.P. (2004). DNA transport into *Bacillus subtilis* requires proton motive force to generate large molecular forces. Nat. Struct. Mol. Biol. *11*, 643-649.

Marchant, J., Sawmynaden, K., Saouros, S., Simpson, P., and Matthews, S. (2008). Complete resonance assignment of the first and second apple domains of MIC4 from *Toxoplasma gondii*, using a new NMRView-based assignment aid. Biomol NMR Assign. *2*, 119-121.

Marri, P.R., Paniscus, M., Weyand, N.J., Rendon, M.A., Calton, C.M., Hernandez, D.R., Higashi, D.L., Sodergren, E., Weinstock, G.M., Rounsley, S.D*., et al.* (2010). Genome sequencing reveals widespread virulence gene exchange among human *Neisseria* species. PLoS One *5*, e11835.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658-674.

Moon, A.F., Mueller, G.A., Zhong, X., and Pedersen, L.C. (2010). A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. Protein Sci. *19*, 901-913.

Nivaskumar, M., Bouvier, G., Campos, M., Nadeau, N., Yu, X., Egelman, E.H., Nilges, M., and Francetic, O. (2014). Distinct docking and stabilization steps of the pseudopilus conformational transition path suggest rotational assembly of type IV pilus-like fibers. Structure *22*, 685-696.

Parge, H.E., Forest, K.T., Hickey, M.J., Christensen, D.A., Getzoff, E.D., and Tainer, J.A. (1995). Structure of the fibre-forming protein pilin at 2.6 Å resolution. Nature *378*, 32-38.

Provvedi, R., and Dubnau, D. (1999). ComEA is a DNA receptor for transformation of competent *Bacillus subtilis*. Mol. Microbiol. *31*, 271-280.

Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T.E., and Nilges, M. (2007). ARIA2: automated NOE assignment and data integration in NMR structure calculation. Bioinformatics *23*, 381-382.

Rusniok, C., Vallenet, D., Floquet, S., Ewles, H., Mouzé-Soulama, C., Brown, D., Lajus, A., Buchrieser, C., Médigue, C., Glaser, P.*, et al.* (2009). NeMeSys: a resource for narrowing the gap between sequence and function in the human pathogen *Neisseria meningitidis*. Genome Biol. *10*, R110.

Seitz, P., Pezeshgi Modarres, H., Borgeaud, S., Bulushev, R.D., Steinbock, L.J., Radenovic, A., Dal Peraro, M., and Blokesch, M. (2014). ComEA is essential for the transfer of external DNA into the periplasm in naturally transformable *Vibrio cholerae* cells. PLoS Genet. *10*, e1004066.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR *44*, 213-223.

Szabó, Z., Stahl, A.O., Albers, S.V., Kissinger, J.C., Driessen, A.J., and Pohlschröder, M. (2007). Identification of diverse archaeal proteins with class III

signal peptides cleaved by distinct archaeal prepilin peptidases. J. Bacteriol. *189*, 772-778.

Thomas, C.M., and Nielsen, K.M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat. Rev. Microbiol. *3*, 711-721.

Webb, B., and Sali, A. (2014). Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinformatics *47*, 5.6.1-5.6.32.

Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. J. Appl. Crystallogr. *43*, 186-190.

Wolfgang, M., Lauer, P., Park, H.S., Brossay, L., Hébert, J., and Koomey, M. (1998). PilT mutations lead to simultaneous defects in competence for natural transformation and twitching motility in piliated *Neisseria gonorrhoeae*. Mol. Microbiol. *29*, 321-330.

Wolfgang, M., van Putten, J.P., Hayes, S.F., and Koomey, M. (1999). The *comP* locus of *Neisseria gonorrhoeae* encodes a type IV prepilin that is dispensable for pilus biogenesis but essential for natural transformation. Mol. Microbiol. *31*, 1345-1357.

LEGENDS TO FIGURES

Fig. 1. Comparison of ComP orthologs in *N. meningitidis* and *N. subflava,* and of their cognate DUS variants. (A) Sequence alignment of ComP and $ComP_{sub}$ from *N. meningitidis* 8013 and *N. subflava* NJ9703, produced using Clustal Omega. Amino acids were shaded in dark blue (when identical), in light blue (when highly similar) or non-shaded (when non-conserved). Relevant structural and functional features have been highlighted. The proteins start with a conserved N-terminal sequence motif that defines all type IV pilins, the class III signal peptide (Szabó et al., 2007). This motifs consists of a hydrophylic leader peptide, which is cleaved by the prepilin peptidase PilD, followed by a stretch of 21 predominantly hydrophobic residues that forms an extended α-helix, which is the main assembly interface of subunits within filaments (Berry and Pelicic, 2015). To facilitate purification, the recombinant proteins were produced without their 28 N-terminal residues, depicted by an arrow. The four Cys residues that form two crucial disulfide bonds are identified by *. The soluble portions that have been purified in this study, as well as the different structural motifs have also been highlighted. (B) Sequence alignment of DUS and DUS_{var1} found in *N. meningitidis* and *N. subflava* genomes, respectively. These 12 bp motifs (Ambur et al., 2007) differ by just one base, which has been underlined.

Fig. 2. Quantitative analysis of $ComP_{sub}$ DNA-binding propensity. (A) Analysis by competition EMSA. After pre-incubating biotinylated DUS_{var1} ds primer with purified MBP-$ComP_{sub}$, increasing concentrations of unlabelled ds primers (DUS_{var1} or SUD in which every second base was altered) were added to compete with bound DNA. DNA was then resolved by electrophoresis on native acrylamide gel, transferred to a positive nylon membrane and detected using a streptavidin-HRP conjugate (Cehovin et al., 2013). In contrast to the DNA-only control (lane 1), a shift is seen in the

presence of protein indicating the formation of a MBP-$ComP_{sub}$/DUS_{var1} complex (lane 2). When the added unlabelled competitor DNA (lanes 3-7) displaces bound biotinylated DUS_{var1}, the shift disappears. (B) Analysis by SPR. A neutravidin-coated sensor chip was used to immobilize (in different channels) similar amounts of biotinylated DUS_{var1} and SUD ds primers, and increasing concentrations of purified His_6-$ComP_{sub}$ were injected. For each protein concentration, the responses at equilibrium (R_{eq}) were recorded. Results are the mean ± standard deviation of four independent experiments.

Fig. 3. High-resolution 3D structure of the DUS receptor ComP from *N. meningitidis*. (A) Crystal structure of the soluble portion of ComP at 1.43 Å resolution. Two different views are shown as a cartoon drawing. The conserved core in type IV pilins (the N-terminal α-helix and 4-stranded antiparallel β-sheet) is depicted in grey. Distinctive/key structural features such as the are α1β1-loop (red), the large β1-β2 loop (blue) and the DD-region delimited by two disulfide bonds that sits on top of the β-sheet (green) have also been highlighted. (B) Structural similarity/differences between the soluble portions of ComP, major pilin PilE from *N. gonorrhoeae* (Parge et al., 1995) and minor pilin PilX from *N. meningitidis* (Helaine et al., 2007). The above distinctive/key structural features have also been highlighted using the same coloring scheme.

Fig. 4. High-resolution 3D structure of the DUS_{var1}-recognizing $ComP_{sub}$. (A) Ribbon representation of the superposition of the ensemble of 20 $ComP_{sub}$ structures determined by NMR. (B) Cartoon drawing representation of the $ComP_{sub}$ structure. The conserved core in type IV pilins (the N-terminal α-helix and 4-stranded antiparallel β-sheet) is depicted in grey. Distinctive/key structural features such as the are α1β1-loop (red), the large β1-β2 loop (blue) and the DD-region delimited by two

disulfide bonds that sits on top of the β-sheet (green) are also highlighted. (C) Cartoon representation of the superposition of $ComP_{sub}$ (magenta) and ComP (grey) structures. The two structures superpose with a rmsd of 2.41 Å over their entire length.

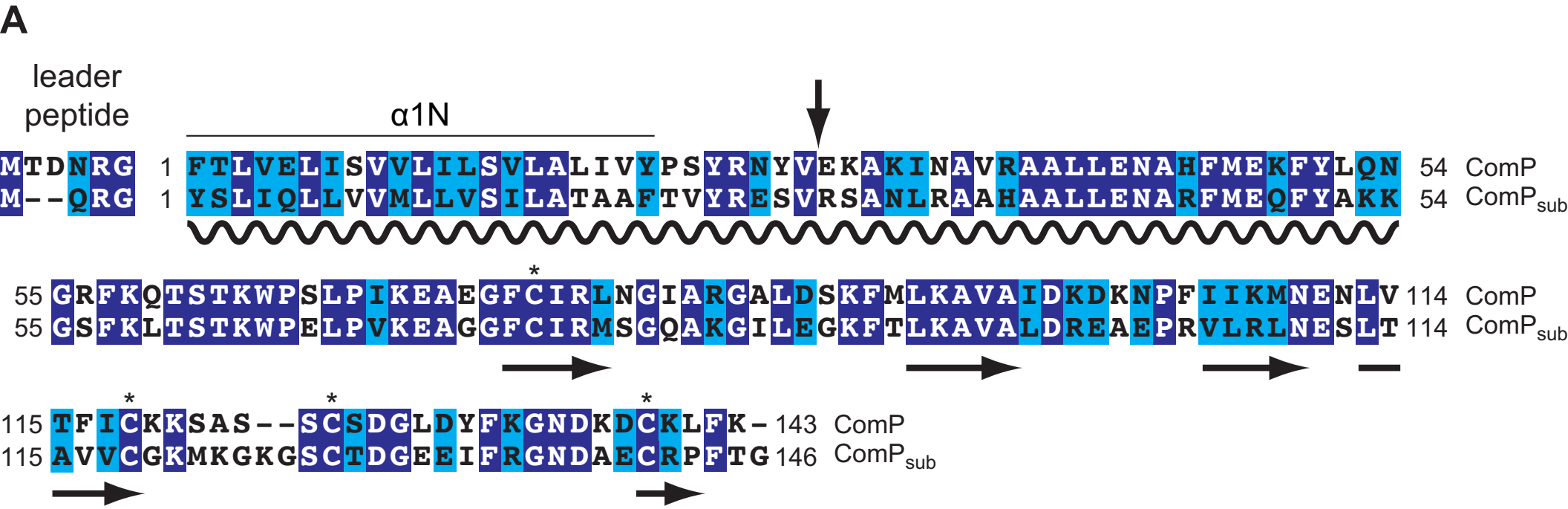
Fig. 5. Detailed NMR analysis of $ComP_{sub}$ binding to DNA. (A) Overlay of representative portions of $^{1}H^{15}N$-HSQC NMR spectra for free His_6-$ComP_{sub}$ and His_6-$ComP_{sub}$ titrated with increasing concentrations of DUS_{var1} ds primer. Labels are placed close to the peaks in the free state. (B). DUS_{var1}-induced chemical shift perturbations (CSP) mapped on the sequence of $ComP_{sub}$. The residues affected by DUS_{var1} titration are coloured in red. The height of the bars above the affected residues is inversely proportional to the concentration of DNA at which CSP were detected. (C) DUS_{var1}-induced CSP mapped on the 3D structure of $ComP_{sub}$ in cartoon and surface representations. The residues affected by DUS_{var1} titration are gradient-coloured according to the concentration of DNA needed to induce from black (lowest) to light green (highest).

Fig. 6. Modelling of the $ComP_{sub}$-DUS_{var1} complex and $ComP_{sub}$ packing within Tfp. (A) HADDOCK model of the interaction between $ComP_{sub}$ (in which the α1β1-loop, β1-β2 loop and the DD-region have been highlighted) and DUS_{var1} (in cyan) in a cartoon representation. Two different views are shown. (B) Packing of full-length $ComP_{sub}$ into Tfp. A full-length $ComP_{sub}$ model was generated using MODELLER. One PilE subunit in the Tfp model was then replaced by this full-length $ComP_{sub}$. The structural features of $ComP_{sub}$ involved in DNA-binding are highlighted on the surface representation.

Fig. 7. Model for the role of ComP DNA-binding pilins in natural transformation in Neisseriaceae. The DUS and its ComP receptor are coloured in red. After ComP

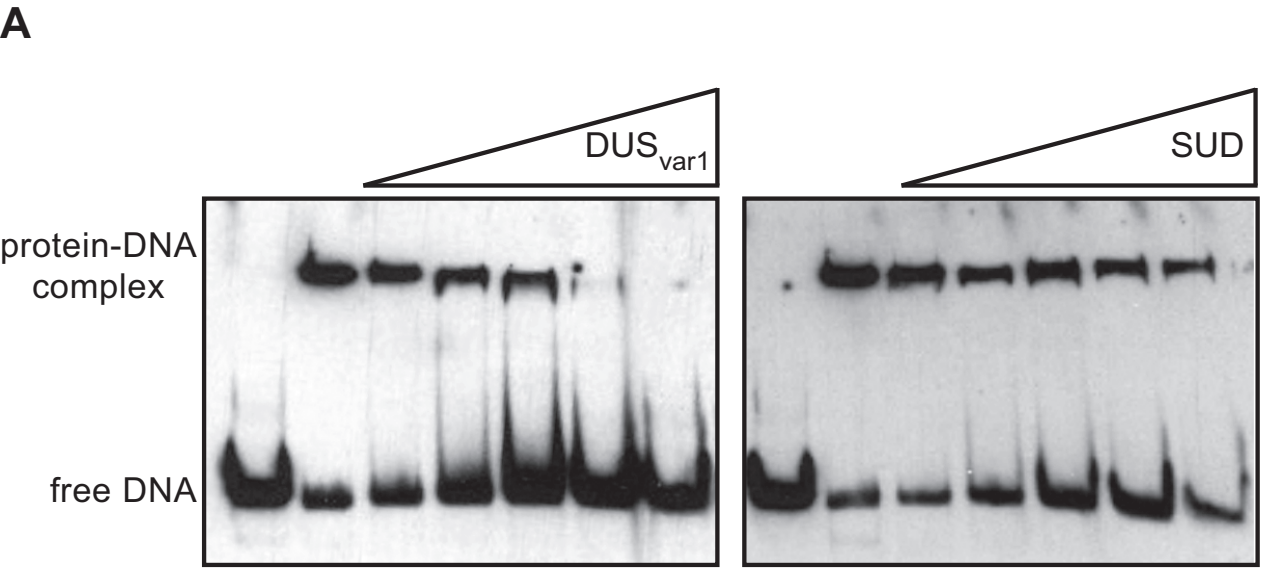
"scans" the DNA for its cognate DUS, it "docks" onto this motif promoting tight binding, which allows DNA to be "pulled and wrapped" upon pilus retraction.

Figure 1


A
leader peptide
α1N
MTDNRG 1 FTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENAHFMEKFYLQN 54 ComP
M--QRG 1 YSLIQLLVVMLLVSILATAAFTVYRESVRSANLRAAHAALLENARFMEQFYAKK 54 ComP$_{sub}$
55 GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIIKMNENLV 114 ComP
55 GSFKLTSTKWPELPVKEAGGFCIRMSGQAKGILEGKFTLKAVALDREAEPRVLRLNESLT 114 ComP$_{sub}$
115 TFICKKSAS--SCSDGLDYFKGNDKDCKLFK- 143 ComP
115 AVVCGKMKGKGSCTDGEEIFRGNDAECRPFTG 146 ComP$_{sub}$


B
DUS ATGCCGTCTGAA (ComP)
DUS$_{var1}$ AGGCCGTCTGAA (ComP$_{sub}$)

Figure 2


A
DUS$_{var1}$
SUD
protein-DNA complex
free DNA

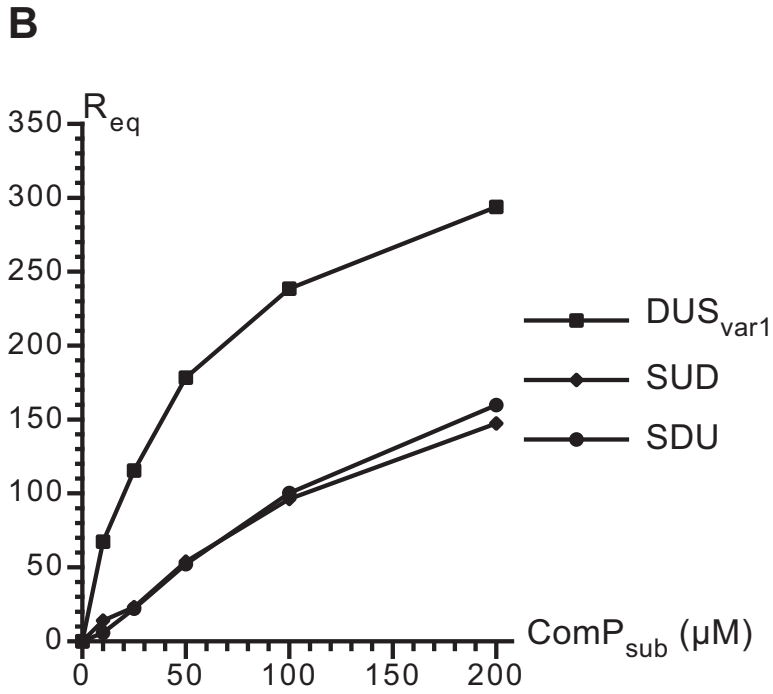

B
R$_{eq}$
350
300
250
200
150
100
50
0
0
50
100
150
200
ComP$_{sub}$ (μM)
DUS$_{var1}$
SUD
SDU

Figure 3

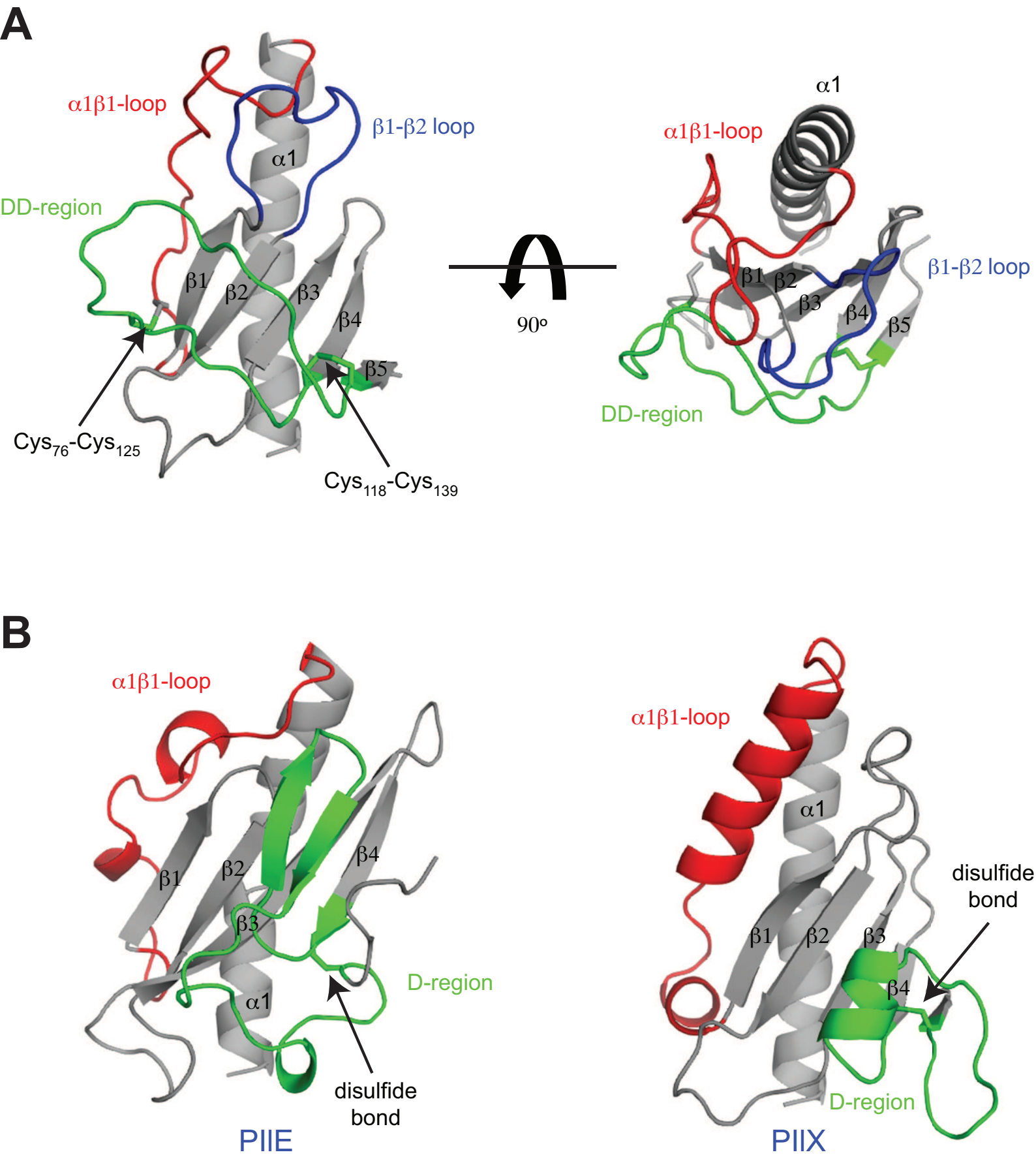

A
α1β1-loop
β1-β2 loop
α1
DD-region
β1
β2
β3
β4
β5
Cys76-Cys125
Cys118-Cys139
90°
α1
α1β1-loop
β1-β2 loop
β1
β2
β3
β4
β5
DD-region
B
α1β1-loop
β1
β2
β4
β3
α1
D-region
disulfide bond
PIIE
α1β1-loop
α1
disulfide bond
β1
β2
β3
β4
D-region
PIIX

Figure 4

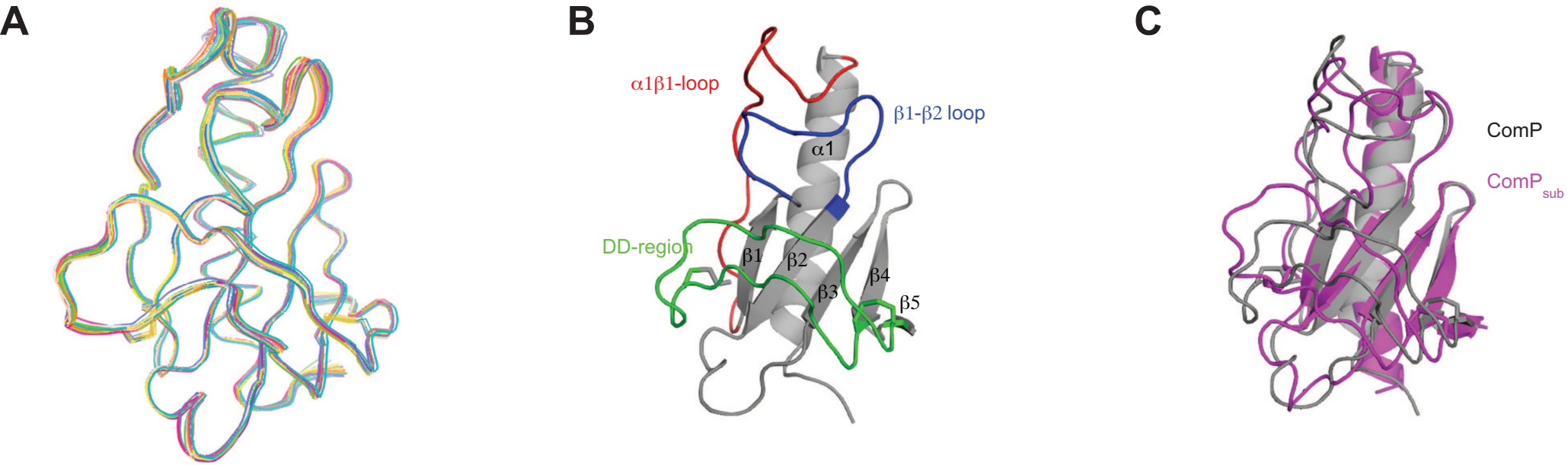

A
B
α1β1-loop
β1-β2 loop
α1
DD-region
β1
β2
β3
β4
β5
C
ComP
ComP sub

Figure 5

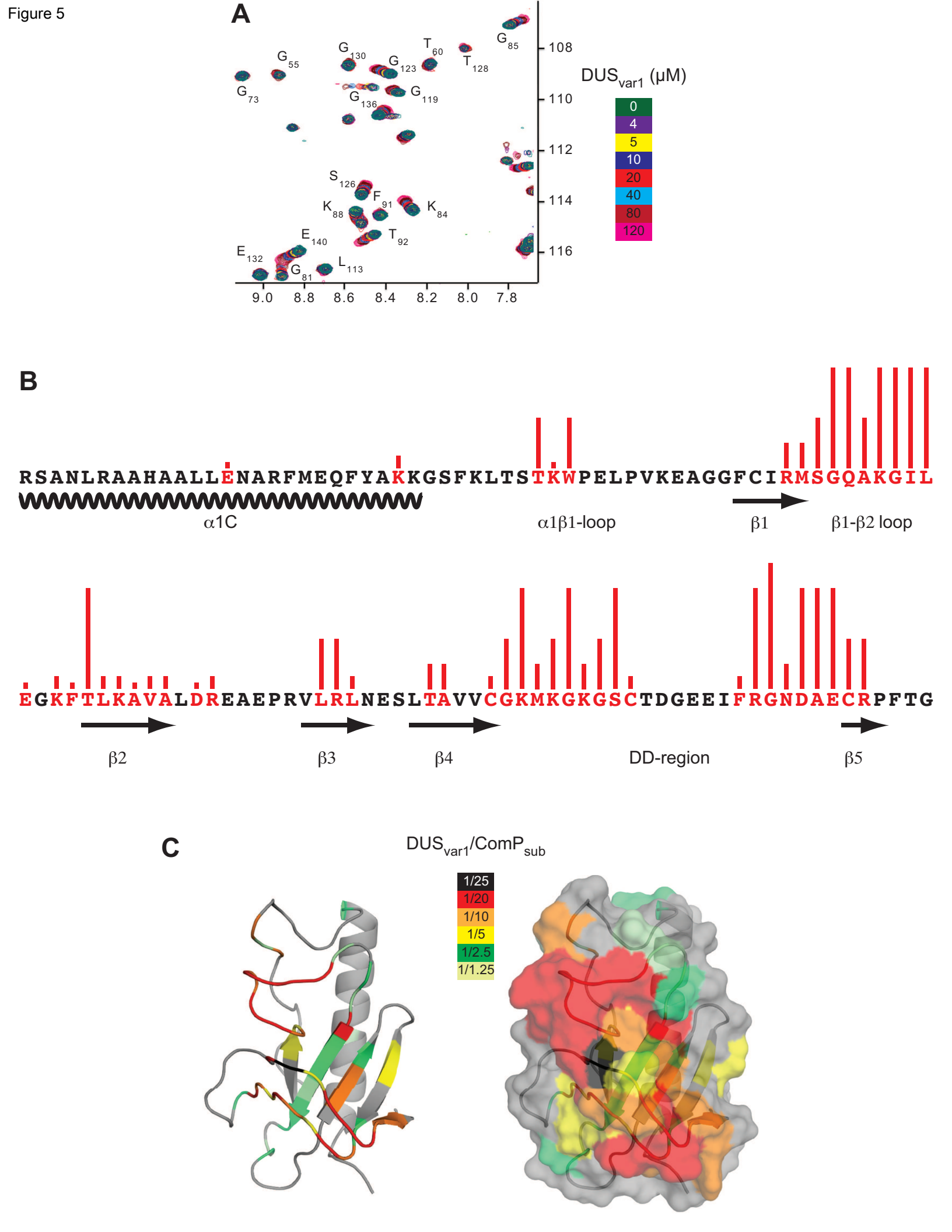
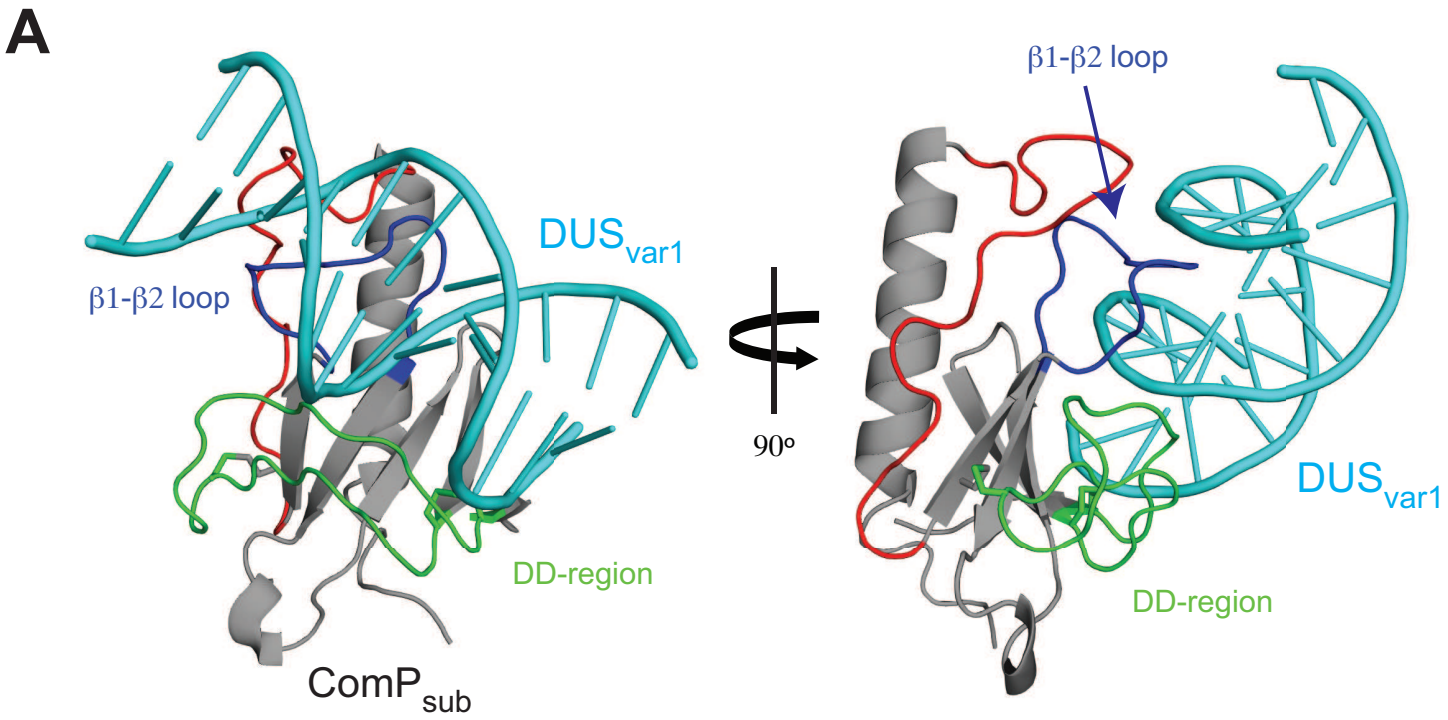

A
G 85
T 60
G 130
T 128
G 123
G 55
G 73
G 119
G 136
S 126
F 91
K 88
K 84
E 140
T 92
E 132
L 113
G 81
108
110
112
114
116
9.0
8.8
8.6
8.4
8.2
8.0
7.8
DUS var1 (µM)
0
4
5
10
20
40
80
120
B
RSANLRAAHAALLENARFMEQFYAKKGSFKLTSTKWPELPVKEAGGFCIRMSGQAKGIL
α1C
α1β1-loop
β1
β1-β2 loop
EGKFTLKAVALDREAEPRVLRLNESLTAVVCGKMKGKGSCTDGEEIFRGNDAECRPFTG
β2
β3
β4
DD-region
β5
C
DUS var1/ComP sub
1/25
1/20
1/10
1/5
1/2.5
1/1.25

Figure 6


A
β1-β2 loop
DUS var1
DD-region
ComP sub
90°
β1-β2 loop
DUS var1
DD-region

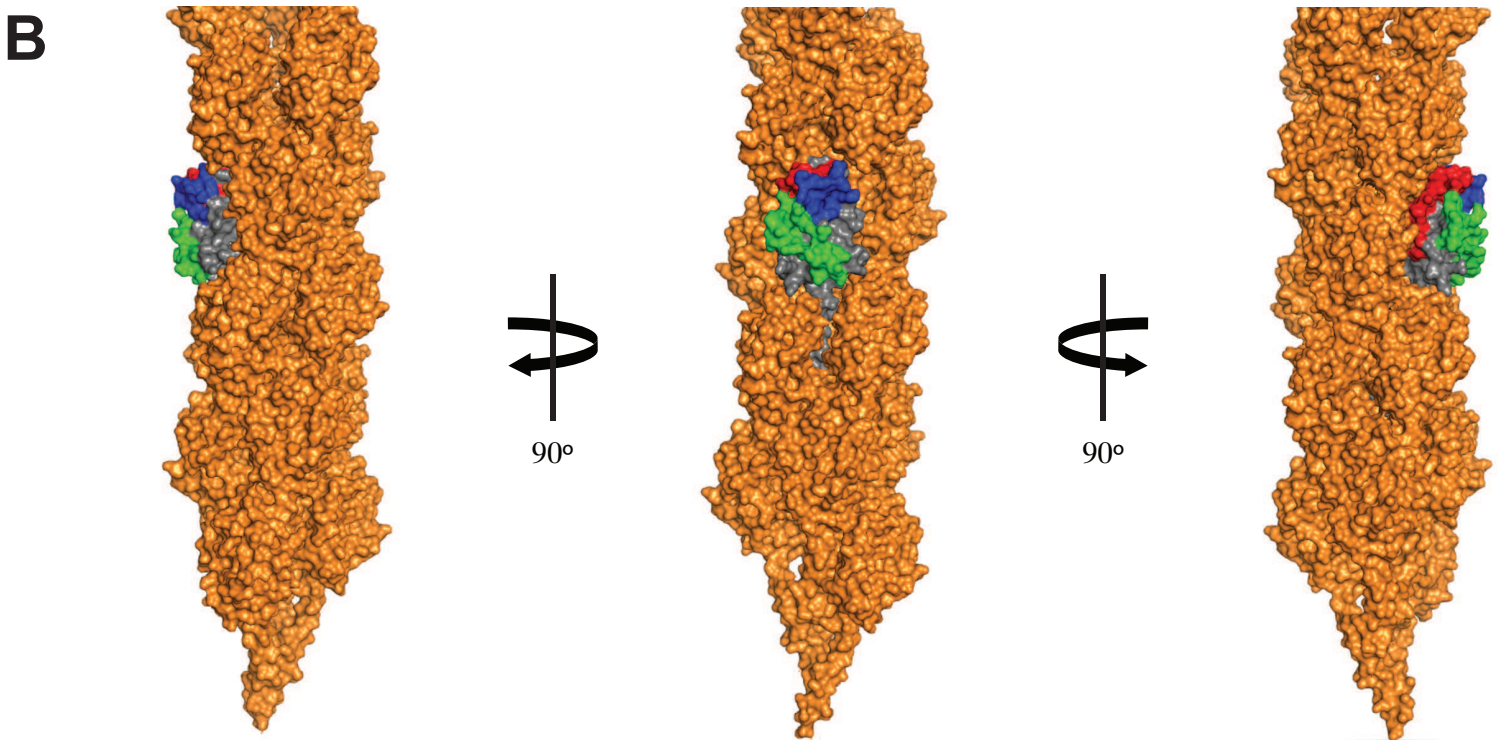

B
90°
90°

Figure 7