CellPress

Type IV Pilus: One Architectural Problem, Many Structural Solutions

Julien R.C. Bergeron[1] and Nikolaos G. Sgourakis[2,*]
[1]Department of Biochemistry and Molecular Biology, and Centre for Blood Research, University of British Columbia, 2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada
[2]National Institute of Diabetes & Digestive & Kidney Diseases, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA
*Correspondence: nikolaos.sgourakis@nih.gov
http://dx.doi.org/10.1016/j.str.2015.01.004

Type IV pili are long appendages found at the surface of many bacteria, composed of an oligomerized pilin protein and involved in processes such as adherence, motility and DNA transfer. In this issue of *Structure*, Piepenbrink and colleagues report the first structure a major pilin from a Gram-positive bacterium, revealing an unprecedented stabilization mechanism that may have implications for pilus evolution.

Most bacteria are decorated with filamentous appendages on their surface, such as flagella, pili, secretion systems, and fimbriae. These appendages perform a number of functions including motility, cell-cell communication, surface adherence, biofilm formation, and eukaryotic cell invasion. One of these appendages, the type IV pilus (T4P), is found in many Gram-negative pathogenic bacteria, and its primary function is to allow twitching motility. In addition, T4Ps have been linked to surface adherence, biofilm formation, and DNA transfer. Interestingly, the T4P is evolutionarily related to other appendages; the Type II secretion system (T2SS), whose role is to promote the secretion of proteins (such as toxins) in the extracellular environment in some Gram-negative bacteria (Ayers et al., 2010); and the archaellum, which allows motility in archaea (Albers and Pohlschröder, 2009).

The T4P is constituted of pilin proteins, which oligomerizes into long fiber structures (Figure 1A). Typically, a pilus contains many copies of a single protein, the major pilin, as well as a few copies of several different minor pilins. Sequence variation between pilin proteins, as well as the exact composition of major and minor pilins, and posttranslational modifications, are thought to permit different functions reported for the various T4Ps, as well as to generate diversity needed to evade immune detection. In the case of the T2SS, it has been proposed that a structure resembling the pilus, labeled pseudopilus, is generated to "push" secreted proteins through the peptidoglycan layer and outer membrane. Similarly to the T4P, the T2SS pseudopilus is composed of a major pseudopilin and several minor pseudopilins. Lastly, the archaellum is formed by up to five archaellins, although the precise composition is not known.

The structures of pilins and pseudopilins from a number of bacteria have been reported (Giltner et al., 2012). Although these proteins vary in size and in sequence, they all share a common lollipop-like architecture consisting of a long, hydrophobic helix at the N terminus followed by a globular domain containing a central β sheet (Figure 1A). The more conserved N-terminal helix is embedded in the inner membrane at the proximal end of the pilus and promotes pilus assembly by forming the hydrophobic core

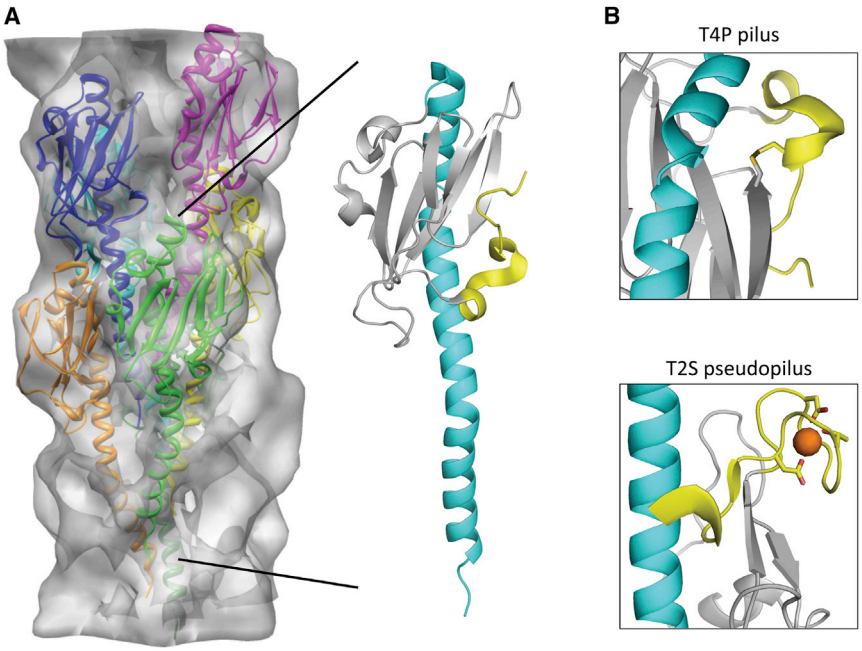

Figure 1. Structure of the Type IV Pilus
(A) EM map of the prototypical *Nisseria gonorrhoeae* T4P (EMDB ID 1236), with the fitted crystal structures of the corresponding pilin protein (PDB ID 2HIL). The structure of a pilin monomer is shown on the right, with the hydrophobic N-terminal helix in cyan, the globular domain in gray, and the D region in yellow. (B) Close-up view of the D region from a T4P pilin (PDB ID 2HI2, top) or a T2SS pseudopilin (PDB ID 3G20, bottom), showing the canonical disulfide bond or calcium coordination site, respectively.

of the growing filament. The globular domain of pilin proteins is located on the outside of the pilus and varies in sequence, size, and fold between species and/or pilus type. At its C-terminal end, a hypervariable region called the D region performs an essential role in surface adherence for many pili. Despite the sequence variation of this D region, it consistently possesses a disulfide bond; this disulfide bond is replaced by a calcium coordination motif in most T2SS major pseudopilins (Figure 1B; Craig et al., 2006).

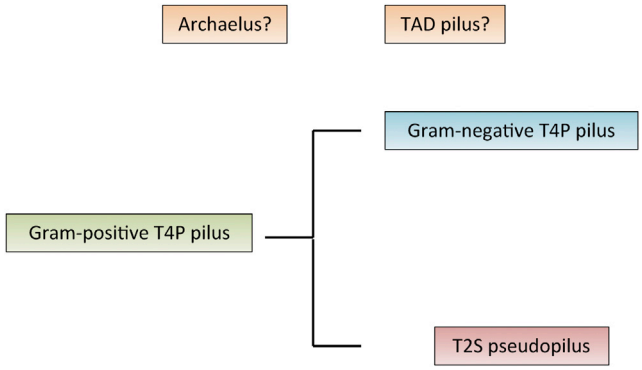

Figure 2. Putative Phylogenetic Tree for T4S-like Appendages
The structure of a Gram-positive pilus reported by Piepenbrink and colleagues (Piepenbrink et al., 2015) suggests that it may be an archetypal appendage which further evolved to specialized T4P and T2SS in Gram-positive bacteria. Further structural work would be required to localize the position of other appendages such as TAD pilus and archaellus in this evolutionary tree.

Recently, the increase in available genome sequences for bacterial strains have led to the identification of gene loci encoding for T4Ps in a number of Gram-positive bacteria, including many *Streptococcus* and *Clostrodium* species (Melville and Craig, 2013). In a few cases, T4P-like appendages have been observed, confirming that these encode functional T4Ps. However the role of these pili is currently poorly understood, and no evidence to date suggests that they can provide twitching motility (Melville and Craig, 2013).

In this issue of *Structure*, Piepenbrink and coworkers demonstrate that the proteins PilA1 and PilJ are the major and minor pilins, respectively, for the T4P found in the Gram-positive pathogen *Chlostrodium difficile* (Piepenbrink et al., 2015). They build on their previous structural work that focused on PilJ (Piepenbrink et al., 2014) and now describe a crystal structure of PilA1. The overall architecture of PilA1 is similar to other pilins, and the variable globular domain mostly resembles a subset of Gram-negative major pilins involved in cellular adherence. Notably, a major difference with other T4P pilins lies in the D region of PilA1, which does not possess the canonical disulfide bond (nor the calcium binding site found in T2SS pseudopilins). Instead, in the PilA1 structure, the globular domain is stabilized with two β strands inserted in loop regions, which form a second β sheet (called B2). Even more surprisingly, Piepenbrink et al. observe that the PilA1 sequence conservation is quite low between *C. difficile* strains, particularly so for the B2 sheet region. Furthermore, they go on to solve the structures of PilA1 in two additional strains, one of which did not include a B2 sheet; instead, water molecules mediate a network of hydrogen bonds that stabilize the globular domain.

The absence of the canonical T4P disulfide bond in major pilins is not entirely unprecedented: it is also lacking from the *Dichelobacter nodosus* major pilin FimA (Hartung et al., 2011). However, in that case the disulfide bond is replaced by a network of hydrogen bonds, forming a similar architecture to that observed in disulfide bond-containing pilins. In contrast, the structures of PilA (and PilJ) demonstrate that Gram-positive pili can employ a number of different strategies to stabilize the globular domain. Interestingly, no disulfide-linked cysteines are found in the D region of any Gram-positive predicted pilin proteins, indicating that this is a conserved feature of this branch of the T4P family.

The variety of strategies employed by the D region becomes relevant in the light of T4P evolution. In particular, it has been suggested that Gram-positive T4Ps might be akin to an archaetypal pilus, which lacks both the retraction apparatus of Gram-negative T4Ps and secreted proteins found in T2SSs (Figure 2). It can therefore be envisioned that ancestor pilins used a diverse range of mechanisms for stabilization of the D2 region, which allowed the emergence of a more specialized disulfide bond for T4P pilins or calcium-binding loops for T2SS pseudopilins. Further structural characterization of Gram-positive pilins, as well as a more detailed functional characterization of these appendages (which is currently largely lacking), will help validate this hypothesis.

It should also be pointed that there is currently no structural data for pilin proteins in two other T4P-like appendages: the archaeolus as well as the Tad pilus (a specialized T4P found in a number of Gram-negative pathogens involved in adherence; Tomich et al., 2007). Their highly divergent sequences suggest that they may utilize entirely new mechanisms for stabilization of the globular region. Additional structural characterization will allow positioning these appendages on the T4P evolutionary tree (Figure 2). The availability of new structures for carefully chosen targets, as exemplified by Piepenbrink and coworkers (Piepenbrink et al., 2015), continues to shed light on the remarkable diversity of emerged structural solutions that support a seemingly simple architectural arrangement in the type 4 pilus assembly, confirming that nature, like Wayne Gretzky, "does not skate to where the puck is, but to where the puck is going to be."

ACKNOWLEDGMENTS

J.R.C.B. is a MSFHR postdoctoral fellow and supported in part by a UBC CBR transition grant. N.G.S. acknowledges funding by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health.

REFERENCES

Albers, S.V., and Pohlschröder, M. (2009). Extremophiles *13*, 403–410.

Ayers, M., Howell, P.L., and Burrows, L.L. (2010). Future Microbiol. *5*, 1203–1218.

Craig, L., Volkmann, N., Arvai, A.S., Pique, M.E., Yeager, M., Egelman, E.H., and Tainer, J.A. (2006). Mol. Cell *23*, 651–662.

Giltner, C.L., Nguyen, Y., and Burrows, L.L. (2012). Microbiol. Mol. Biol. Rev. *76*, 740–772.

Hartung, S., Arvai, A.S., Wood, T., Kolappan, S., Shin, D.S., Craig, L., and Tainer, J.A. (2011). J. Biol. Chem. *286*, 44254–44265.

Melville, S., and Craig, L. (2013). Microbiol. Mol. Biol. Rev. *77*, 323–341.

Piepenbrink, K.H., Maldarelli, G.A., de la Peña, C.F., Mulvey, G.L., Snyder, G.A., De Masi, L., von Rosenvinge, E.C., Günther, S., Armstrong, G.D., Donnenberg, M.S., and Sundberg, E.J. (2014). J. Biol. Chem. *289*, 4334–4345.

Piepenbrink, K.H., Maldarelli, G.A., Martinez de la Peña, C.F., Dingle, T.C., Mulve, G.L., Lee, A., von Rosenvinge, E., Armstrong, G.D., Donnenberg, M.S., and Sundberg, E.J. (2015). Structure *23*, this issue, 385–396.

Tomich, M., Planet, P.J., and Figurski, D.H. (2007). Nat. Rev. Microbiol. *5*, 363–375.

A Consensus on Protein Structure Accuracy in NMR?

Martin Billeter[1,*]
[1]Department of Chemistry and Molecular Biology, University of Gothenburg, Box 462, 40530 Gothenburg, Sweden
*Correspondence: martin.billeter@chem.gu.se
http://dx.doi.org/10.1016/j.str.2015.01.007

The precision of an NMR structure may be manipulated by calculation parameters such as calibration factors. Its accuracy is, however, a different issue. In this issue of *Structure*, Buchner and Güntert present "consensus structure bundles," where precision analysis allows estimation of accuracy.

Biomolecular NMR spectroscopy is arguably among the most versatile experimental methods to characterize proteins. NMR can be used to investigate 3D structures at atomic resolution, dynamic behavior on time scales from nanoseconds to hours and intermolecular interactions covering a wide range of affinities. Despite this versatility, NMR does have some limitations. Not unlike other structural biology methods, the coordinates resulting from an NMR structure determination procedure contain a certain degree of "imprecision" that is to some extent due to intrinsic protein dynamics. The question of how well the experimental structure reflects the "true" one is a significant one for those interested in using the structural information for follow-up work such as computational structure-based drug design. Estimating this accuracy is never trivial (true structures are not known, methods may be prone to inherent biases, etc.). Historically, the field of NMR-based structural biology has had difficulties developing an accepted accuracy measure. The result of an NMR structure determination is typically presented as a bundle of structures. The width of the bundle varies along the protein sequence. One may be tempted to interpret the extent of these variations both as an indication of local dynamics as well as the accuracy of the structure. The first assumption is often qualitatively correct; however, local dynamics can be determined in an independent and better manner by NMR relaxation studies.

The second assumption is critical! Even while in infancy, a qualitative correlation between the variation in an NMR structure bundle and the difference between the bundle's mean structure and a corresponding crystal structure was observed for NMR structure determinations (Billeter, 1992). Assuming that the latter represents the true structure, this difference becomes an estimate of accuracy. However, the study very consistently showed that optimal superpositions still place the crystal structures largely outside of the NMR bundles; this was confirmed by numerous later NMR structure determinations. Although it is arguable whether crystal structures are the true structures, the persistent difference is nonetheless disturbing.

Simple case difficulties in estimating the accuracy of an NMR structure are illustrated in Figure 1. A lysine side chain, for example, has two potential partners (aspartic acids) for salt bridge formation; both corresponding conformations may be populated, leading each to the observation of a Lys-Asp NOE (Figures 1A and 1B). Automatic peak assignment during structure determination may assign only one NOE (the other one may be ambiguous due to overlap) or both. The conformations of Figures 1A and 1B are obtained with one NOE; use of both NOEs results in the (chemically unlikely) conformation of Figure 1C. In all three cases, high precision will result, whereas a correct accuracy should encompass all conformations—this may result when considering both NOEs as ambiguous throughout the structure determination. Similar problems may affect also the backbone fold. Structure validation procedures are efficient in identifying errors in structures but have difficulties in detecting overestimation of accuracy (Spronk et al., 2004).

In this issue of *Structure*, Buchner & Güntert (2015) present a novel idea to improve accuracy estimations. The experimental input is unchanged from the conventional approach and consists of unassigned peak lists from NOESY spectra augmented by the protein sequence and a list with chemical shifts. With this input, the standard CYANA procedure with automated NOE assignment yields both a structure bundle and a set of distance restraints derived from the NOESY peaks (Herrmann et al., 2002). Importantly, the result also depends on random numbers used to construct starting structures for the CYANA optimization algorithm. Thus, repetitions of the CYANA procedure, using different random numbers, will yield different structure bundles, each with a different set of distance

