International Journal of Medical Microbiology 305 (2015) 265–275

Contents lists available at ScienceDirect

International Journal of Medical Microbiology

journal homepage: www.elsevier.com/locate/ijmm

A domain dictionary of trimeric autotransporter adhesins

Jens Bassler, Birte Hernandez Alvarez, Marcus D. Hartmann, Andrei N. Lupas*

Department of Protein Evolution, Max-Planck-Institute for Developmental Biology, Spemannstr. 35, 72076 Tuebingen, Germany

ARTICLE INFO

Keywords:
Trimeric autotransporter adhesins
Gram-negative bacteria
Domain dictionary approach
Domain structure

ABSTRACT

Trimeric autotransporter adhesins (TAAs) are modular, highly repetitive outer membrane proteins that mediate adhesion to external surfaces in many Gram-negative bacteria. In recent years, several TAAs have been investigated in considerable detail, also at the structural level. However, in their vast majority, putative TAAs in prokaryotic genomes remain poorly annotated, due to their sequence diversity and changeable domain architecture. In order to achieve an automated annotation of these proteins that is both detailed and accurate we have taken a domain dictionary approach, in which we identify recurrent domains by sequence comparisons, produce bioinformatic descriptors for each domain type, and connect these to structural information where available. We implemented this approach in a web-based platform, daTAA, in 2008 and demonstrated its applicability by reconstructing the complete fiber structure of a TAA conserved in enterobacteria. Here we review current knowledge on the domain structure of TAAs.

Introduction

Trimeric autotransporter adhesins (TAAs) are a widespread family of outer membrane proteins (OMPs) in Gram-negative bacteria (reviewed in Linke et al., 2006; Łyskowski et al., 2011). Most TAAs in the non-redundant protein database at NCBI are from proteobacteria and here almost exclusively from the α, β and γ branches, but this is clearly due to a sampling bias, as TAAs are also well-represented in bacterial clades whose detailed exploration has only begun in recent years. Thus, Verrucomicrobia, Bacteroidetes, Synergistetes, Fusobacteria, and the Gram-negative branch of Firmicutes (the Negativicutes) all show a broad representation of TAAs in their genomes. In addition, TAAs are also found some species of Chlorobia and unicellular Cyanobacteria.

As mediators of adhesion, TAAs play a crucial role in attaching bacteria to surfaces in their environment, both of a biotic and an abiotic nature. Due to the strong focus of Microbiology on pathogenesis, it is unsurprising that the TAAs explored in most detail are determinants of infection and host colonization. These include YadA of *Yersinia enterocolitica* (Bölin et al., 1982), Hia of *Haemophilus influenzae* (St Geme and Cutter, 2000), UspA1 and A2 of *Moraxella catarrhalis* (Lafontaine et al., 2000), BadA of *Bartonella henselae* (Riess et al., 2004), SadA of *Salmonella enterica* (Raghunathan et al., 2011), and the Eib proteins of *Escherichia coli* (Sandt and Hill, 2000). A more detailed list of experimentally explored TAAs is provided in Supplementary Table 1.

TAAs are trimeric fibers ranging in length over almost two orders of magnitude (Fig. 1). They share a basic architecture, consisting of variable arrays of typically more slender stalks and bulkier head regions, and ending in a C-terminal membrane anchor. In simple TAAs, such as YadA, the fiber consists of only one head and one stalk, while in more complex TAAs, head and stalk regions alternate multiply along the fiber (Fig. 1). Heads and stalks are assembled from a set of analogous, structurally conserved building blocks, whose rearrangement in evolution has resulted in considerable diversity. The trimeric anchor, in contrast, is homologous in all TAAs and represents their defining element.

A domain dictionary approach to TAA annotation

The initial description of TAAs as a new class of adhesins (Hoiczyk et al., 2000) had already recognized their basic head-stalk-anchor architecture, the universal presence of the anchor, and shared sequence motifs specific to each part of the fiber. As crystal structures for TAA fragments were determined (Nummelin et al., 2004; Yeo et al., 2004; Meng et al., 2006), it became possible to relate the sequence motifs to specific domains, allowing their recognition in other TAAs by sequence comparisons. The comprehensive annotation of such TAAs was however hindered by the high sequence diversity and mosaic-like arrangement of their constituent domains, fuzzy domain boundaries, limited coverage of the sequences by domains of known structure, and the frequent presence of extended regions of low sequence complexity, some

* Corresponding author. Tel.: +49 7071 601 340; fax: +49 7071 601 349.
E-mail address: andrei.lupas@tuebingen.mpg.de (A.N. Lupas).

http://dx.doi.org/10.1016/j.ijmm.2014.12.010

Table 1
Domains and motifs of TAAs, with a brief description of their structural properties. Where they move the path of the chain by topological crossover around the trimer axis, this is indicated as viewed from the N-terminus of the fiber. For domains of known structure, an example is listed as a Protein Data Bank ID. Entries are listed in the order in which they appear in the text.

Name	Description	Topological crossover	PDB
Anchor	Trimeric, 12-stranded outer membrane β-barrel	–	2gr7
Stalk	Slender segment of predominantly coiled-coil structure	–	
N@*d*	Asparagine in coiled-coil position *d*. This is the most common polar residue in the core of TAA stalks.	–	
FGG	Insertion of a 3-stranded β-meander into a coiled-coil segment	120° ccw	2yo2
Eib saddle	Non-helical insertion into a coiled-coil segment, exclusive to Eib proteins	100° ccw	2xqh
Connectors	Domains mediating the transition between heads and stalks		
Neck	β to α connector		
Short neck	Neck variant of 19 residues length	120° cw	3d9x
Long neck	Neck variant of 22 residues length	120° cw	3laa
Insert necks	Neck variants containing an extended insertion	120° cw	1s7m
KG	Insert neck variant lacking the first β-strand	–	3emi
DALL	α to β connector. Three conserved, taxonomically widespread variants (DALL1-3) are described so far. Always followed by a neck domain	–	2yo3, 2ynz
HANS	Short α to β connector exclusively found before Ylheads	120° ccw	2yo3
Heads	Domains of predominantly β secondary structure		
Transversal heads	Analogous group of head domains with β-strands perpendicular to the trimer axis. Consist of a variable number of repetitive motifs		
Ylhead	Most common transversal head domain	–	1p9h
GIN	Transversal head, appears exclusively after interleaved head domains	120° cw	3d9x
Interleaved heads	Homologous group of head domains with β-strands parallel to the trimer axis. Unlike transversal heads, they are not repetitive and have a fixed size		
TrpRing	Most common interleaved head domain	120° ccw	3d9x
FxG	Variant of TrpRing	120° ccw	–
GANG	Deletion variant of TrpRing	?	–

of which have since been recognized as compositionally unusual coiled coils (Hartmann et al., 2009).

In order to obtain automated annotations that are both detailed and accurate, we took a 'domain dictionary' approach to the description of TAAs. We performed a comprehensive analysis of TAA sequences, deriving manually curated alignments for conserved regions, using the modular nature of these regions to define domain boundaries, constructing profile Hidden Markov Models (HMMs) as higher-level descriptors of the domains, and extracting knowledge-based rules of domain order to support assignments made by sequence similarity. We then built a server, daTAA (domain annotation of TAAs; http://toolkit.tuebingen.mpg.de/dataa; Szczesny and Lupas, 2008), which uses the profile HMMs, some of the knowledge-based rules, and a modified recognition protocol for coiled coils in order to annotate TAA sequences. daTAA has been found to be very sensitive, correctly predicting even highly divergent exemplars of domains (e.g. Meng et al., 2008).

In going from sequence to structure, two observations were critical: (I) The domains of TAAs are strung up consecutively along the fiber with generally minimal interaction (Yeo et al., 2004; Szczesny et al., 2008; Meng et al., 2008; Edwards et al., 2010; Leo et al., 2011; Hartmann et al., 2012). This suggested that context would exert little influence on domain structure. (II) Even after extensive sequence divergence, domains retain their structure closely, presumably due to the constraints imposed by their trimeric symmetry and frequently interleaved fold (Szczesny et al., 2008). This suggested that once identified, a domain could be modeled accurately from known prototypes. As a test case, we reconstructed the complete fiber structures of a family of complex TAAs conserved in enterobacteria: SadA, UpaG, and EhaG (Hernandez Alvarez et al., 2008; Hartmann et al., 2012). We annotated the proteins using daTAA and built models from known structures. In parallel, we solved crystal structures of segments of SadA, covering both the domains whose structure was still unknown and several domains of known structure that we had modeled computationally. In this way, we showed that TAAs could be annotated with high accuracy, both in sequence and in structure, using the domain dictionary approach.

In the following, we will now review individually the component entries of the TAA domain dictionary (Table 1), progressing from the membrane anchor via the stalk to the head domains. More detailed descriptions of the structures are available in the Supplementary material, where we will also address the question of domain recombinations between TAAs and the fibrous surface proteins of bacteriophage. Given that the two classes of proteins have architectural and functional similarities (Fig. S1), one might have expected such events, but at present only a few instances can be observed (Fig. S2).

The anchor

The anchor domain consists of three subunits, each composed of one long, amphipathic helix followed by a four-stranded β-meander (Hoiczyk et al., 2000; Koretke et al., 2006; Meng et al., 2006). After export of the TAA through the inner membrane via the Sec pathway, triggered either by a canonical or an extended N-terminal signal peptide (Sijbrandi et al., 2003), the β-strands insert into the outer membrane and assemble into a 12-stranded β-barrel in a process dependent on the Bam complex (Lehr et al., 2010). The barrel is thought to provide the pore through which the upstream head and stalk domains exit the periplasm (Leo et al.,

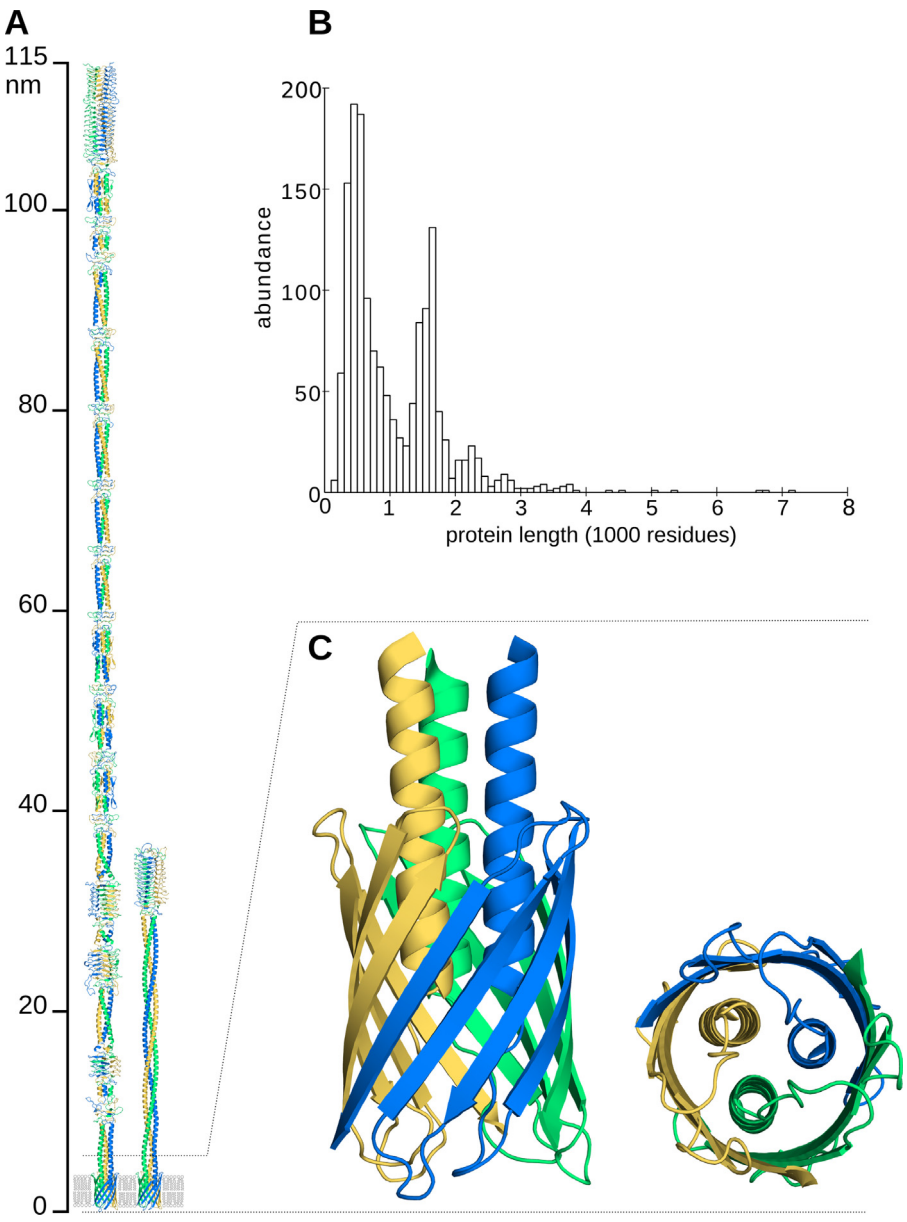


Fig. 1. TAA fibers. (A) Full fiber reconstructions of UpaG (left) and YadA (right) as representatives of the two main size ranges encountered in TAAs (panel B). The structures are colored by subunit chain. (B) Size distribution of 1517 complete TAAs in the non-redundant protein database at NCBI (i.e. containing both an N-terminal signal peptide and a C-terminal membrane anchor). The shortest is BN80_143 from the Yersinia phage phiR1_RT (122 res.) and the longest heg44 from *Veilonella atypica* (7187 res.). (C) Side and bottom view of the TAA membrane anchor.

2012), a view supported by the observation that in YadA, a nearly invariant glycine on the inner face of the barrel is essential for the export of the polypeptide chains (Grosskinsky et al., 2007). After export is complete, the helices of the anchor trimerize to form a coiled coil at the center of the barrel, obstructing the pore and stabilizing the structure (Fig. 1, Meng et al., 2006). The shortest TAA we have encountered, BN80_143 of the Yersinia phage phiR1_RT, consists entirely of the anchor domain, with the central coiled-coil stalk barely extended by 15 N-terminal residues.

The anchor domain is most similar to the translocation domain of canonical single-chain autotransporters (Oomen et al., 2004), and to that of intimins and invasins (Fairman et al., 2012), both of which are 12-stranded β-barrels as well; in the latter, however, located at the N-, not the C-terminus. A recent study, which analyzed the similarity of outer membrane proteins in sequence and structure, concluded on the common origin of all OMPs by amplification and divergence of an ancestral ββ-hairpin (Remmert et al., 2010). In this study, the three types of autotransporter mentioned above, often referred to as secretion systems of type Va (canonical single-chain), type Vc (trimeric), and type Ve (invasin), were recovered as nearest neighbors in a cluster map based on sequence similarity. This suggests that they may have diverged from a common ancestor that was already able to translocate polypeptide chains.

Stalks

All stalk segments of TAAs represent variants of the basic parallel, three-helical coiled coil and as such adhere to the rules of coiled-coil structure (reviewed in Lupas and Gruber, 2005). Coiled coils are bundles of α-helices that interact through seams of residues that run along the length of the helices. In their simplest form, these seams are formed by hydrophobic residues (h) arranged in a repeating pattern of seven residues, separated mainly by polar

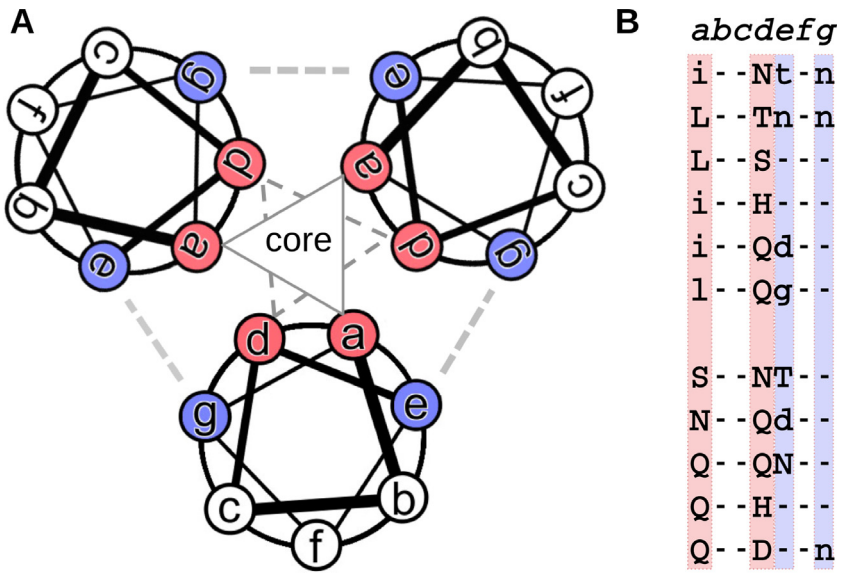


Fig. 2. Coiled coils with polar cores. (A) Schematic representation of a parallel, trimeric, heptad coiled coil viewed from the N-terminus. Heptad positions are denoted by the letters *a*-*g*; *a* and *d* point to the center of the bundle and form separate layers of interaction (*a*-*a*-*a* and *d*-*d*-*d*). Polar interactions between e and g frequently further stabilize the structure. (B) Common motifs with polar cores in TAAs. The upper block shows motifs with polar residues in *d*, the lower block motifs with polar residues in *a* and *d*. Consensus sequences show residues present in more than 50% of instances in upper case and in 25–50% of instances in lower case.

residues (p): hpphppp. This pattern is referred to as the heptad repeat and its positions are denoted *abcdefg*, with the hydrophobic residues in positions *a* and *d* (the 'core', Fig. 2). The sequence periodicity of such a coiled coil is 3.5 residues per turn, as the 7 residues of each repeat are arranged over 2 turns of the α-helix. Since the periodicity of an undistorted α-helix is slightly larger, on average 3.63 residues per turn, and the α-helix is right-handed, it follows that coiled coils with heptad periodicity have to wind into a left-handed supercoil to compensate for the drift in their seams.

Such left-handed, heptad coiled coils are the predominant form of stalk segments in TAAs, but they often show highly unusual residue compositions, making their prediction very challenging. Stalk segments of TAAs are therefore often not recognized by coiled-coil prediction programs and in fact frequently considered unstructured by disorder prediction programs. As a rule, segments that show a sequence periodicity compatible with coiled-coil structure and are flanked at the N-terminal end by a β-to-α connector and at the C-terminal end by an α-to-β connector (see the next section for "Connectors") should be treated as coiled coils.

The left-handed coiled-coil stalks of TAAs show a second substantial departure from canonical trimeric coiled coils: their core residues are often polar, particularly in position *d*, but occasionally in both *a* and *d*. Several heptad patterns with polar core residues are seen recurrently in TAAs (Fig. 2B), the most frequent by far containing asparagine in *d* (which we refer to as N@*d*; Hartmann et al., 2009), often repetitively in consecutive heptads. A structural and biophysical analysis of such repeats from a stalk segment of SadA showed that the asparagines form a highly ordered network of polar interactions with an anion at its center (Fig. 3A). Although they decrease the stability of the coiled coils in which they occur, they increase their solubility and preserve their ability to fold at elevated concentrations. Considering the need of TAAs to maintain an unfolded but soluble state of the polypeptide chain during export, these properties suggest a functional role of N@*d* residues in maintaining an export-competent state (Hartmann et al., 2009). Additionally, coiled-coil segments with polar cores have been proposed to provide bending sites for the TAA fiber (Leo et al., 2011).

Although generally of heptad periodicity, TAAs also contain coiled coils with other periodicities, primarily hendecad (repeats of 11 residues over 3 helical turns) and pentadecad (repeats of 15 residues over 4 helical turns). These differ from the heptad periodicity by progressive insertions of 4 residues, which are slightly more than one turn of an undistorted α-helix (3.63). Such insertions alter the path of the seams of interaction and therefore the supercoiling angle of the resulting coiled coil, progressively shifting it from a left-handed to a right-handed geometry (Gruber and Lupas, 2003). Hendecads have a periodicity of 3.67 (11/3) and are thus fairly straight, 3.67 being close to the undistorted 3.63, while pentadecads have a periodicity of 3.75 (15/4) and are thus supercoiled to the right as strongly as canonical coiled coils are to the left (3.75 − 3.63 = 0.12; 3.50 − 3.63 = −0.13). As mentioned above, most TAA stalk segments are left-handed, built of heptads or of mixtures of heptads and hendecads. The last stalk segment, however, which extends into the pore of the anchor, frequently starts out right-handed and makes a transition to left-handed supercoiling prior to entering the pore, as seen for example in the stalk of YadA (Hernandez Alvarez et al., 2010;Fig. 3B and D). The reasons for this feature remain unclear.

A common sequence motif of right-handed coiled coils in TAAs is YxD, which can be used to identify them by their sequence. In the structure of the YadA stalk (Hernandez Alvarez et al., 2010;Fig. 3B and D), the residues of this motif are seen to form a ring of polar interactions, providing stability and structural specificity to the stalk. This is analogous to the RxD motifs frequently encountered in left-handed trimeric coiled coils (Kammerer et al., 2005), which are also seen in the left-handed part of the YadA stalk (Fig. 3B and D).

A striking elaboration of TAA stalks is a non-helical motif inserted into the coiled coil, which we named FGG (Fig. 3C and E) for its most conspicuous sequence pattern at the time when we first observed it (Riess et al., 2004); today, the most frequent occurrence of this pattern is LGG (Fig. 3E). It interrupts the N-terminal part of the coiled coil and forms a three-stranded β-meander, which ends near the point of origin of the insertion in the next chain of the trimer. FGG motifs are widespread in TAAs, but few other such non-helical insertions have been identified so far. The main other example is the 'saddle' of Eib proteins, which in EibD connects the upper, hendecad part of the stalk to the lower, heptad part (Leo

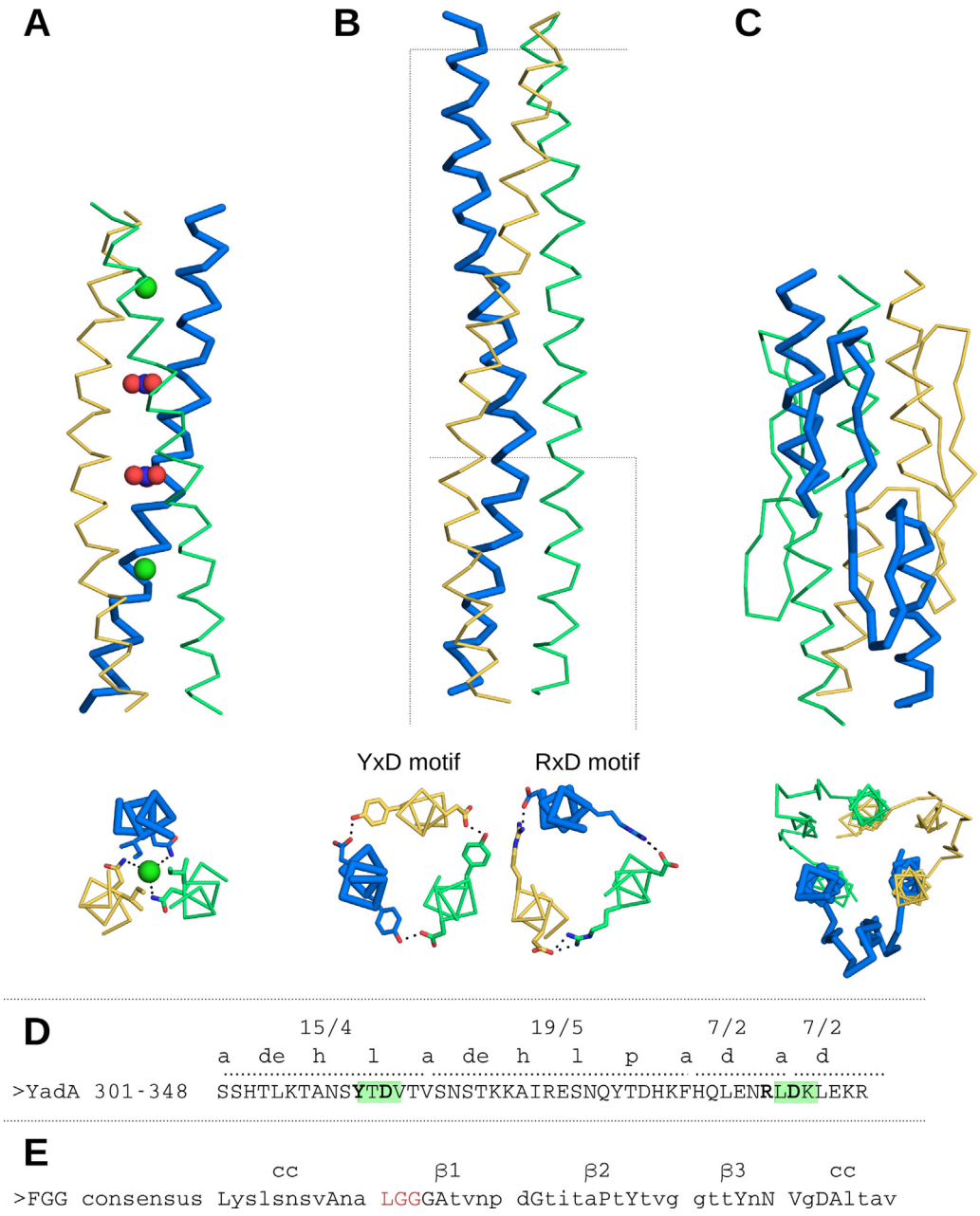


Fig. 3. Stalk motifs of TAAs in side and top view. (A) Left-handed coiled coil with chloride and nitrate ions coordinated at the center by N@*d* motifs, from the coiled-coil stalk of SadA. (B and D) Transition from a right-handed to a left-handed coiled-coil in the YadA stalk, including the indicative YxD and RxD motifs. (C and E) FGG stalk motif inserted into the coiled coil in the stalk of SadA, causing a 120° shift of the subunit chains around the trimer axis. The consensus sequence shows the relatively most frequent residues in lower-case, or in upper case if present in more than 50% of instances.

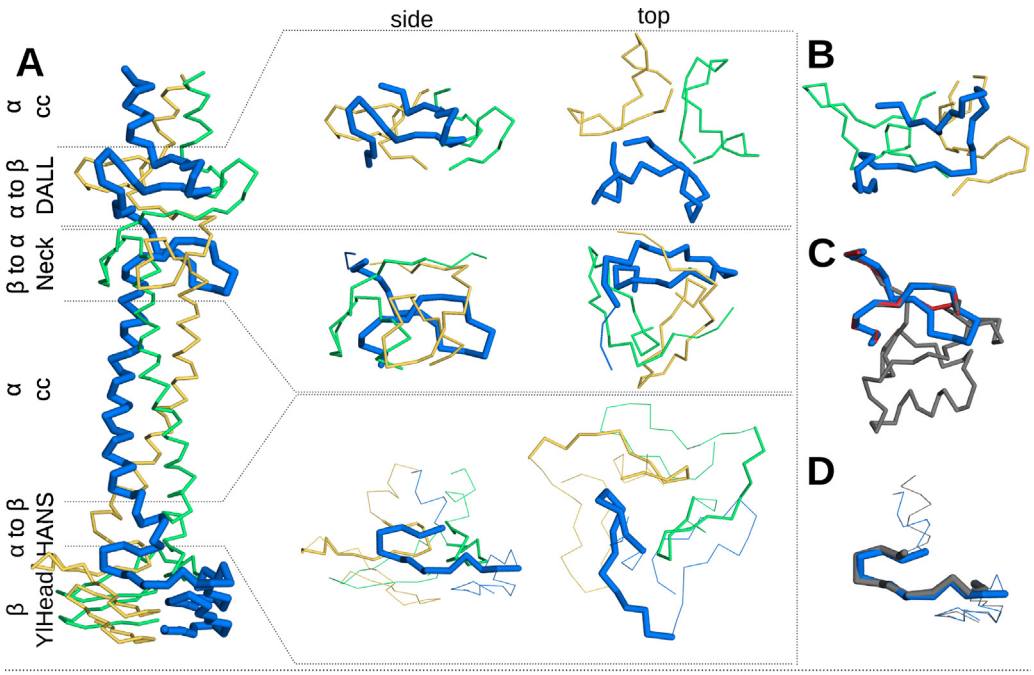
et al., 2011). The 'saddle' forms fewer and less tight interactions with the coiled coil than FGG and occurs only in Eib proteins.

Connectors

In TAAs, the transitions between the slimmer stalks, which are mainly α-helical, and the bulkier heads, which are primarily composed of β-strands, are mediated by a set of conserved connectors (Fig. 4). Basically, these can be classified into β-to-α connectors, going from heads to stalks, and α-to-β connectors, going from stalks to heads.

The main connector from β to α is the neck domain (Fig. 4A, C and F). Neck domains represent the most highly conserved part of TAAs (Hoiczyk et al., 2000) and are often identifiable in unannotated sequences by simple visual scanning, providing a rapid first parsing. The overwhelming majority of necks occurs in two length variants, short necks (19 res.) and long necks (22 res.), which differ by the insertion of three residues in the loop connecting their two β-strands (Fig. 4C and F). In addition, several necks with longer insertions at the same location have been observed (Yeo et al., 2004; Meng et al., 2008), which we refer to as IS necks.

Necks share a common architecture: A first β-strand connects the neck to the preceeding domain and anchors it to the core of the fiber by two hydrophobic, inward-projecting residues (this strand is however deleted in the KG variant (Fig. 4F); see also the Supplementary material). A second β-strand then leads to a triangular nexus at the valine of the consensus motif DAVN (Fig. 4F), which mediates the transition from α to β structure. Here, the chains are tightly connected by a symmetric network of hydrogen bonds,



E

```
                h  ssss                    ssssss      s
>DALL1          aAVkYDkndGvny------------nsVTLGGgn-gtp
>DALL2          DALLWdeta----------------gAFsAsHGgtnSK
>DALL3          DgLVqQd------gtglITIGketgGteIniAnkdGarR
```

F

```
                sbb                                                    ss  hhhh
>short          ItNVAaG-------------------------------------------tadTDAVNVgQL
>YV             ITNVAdP-------------------------------------------tddtDAvNKkYV
>long           ItNVAaGdis----------------------------------------atSTDAVNGSQL
>insert 1       VTNVtSGLkkYGnaNfdalnnsatdltkqvDdAYkGLlNLnEKgagktplVADNTAATVGDL
>insert 2       ITNVASGLkgagvls-dAtg-------------------------dvLnNAANVGDL
>KG             -------------------------kTsvikpknGKlvtgktannattdtNkGsGlVTGNtV
```

G

```
                                  cc          HANS          YlHead
                                hhhhhhh    sssssss        sss    sss
>HANS           consensus              GsTKYFHANSTeaD
>SadA           1226-1239    IGDIVTT GSTKYFKTNTDGAD ANAQGADSVAIGSG
>BpaA PDB 3laa    84- 117    IDNINAN G-IKYFHANSTDPD SVATGTNSVAIGPN
```

Fig. 4. Connector domains. (A) A segment of the SadA fiber containing DALL2 (α to β), long neck (β to α) and HANS (α to β) connectors. The connectors are shown individually as well, in side and top view. (B) DALL1 variant. Compared to DALL2, the β-strands are further apart and contain intervening water molecules (not shown; Hartmann et al., 2012). (C) Superposition of representative structures of a short neck (red), long neck (blue), and insert neck 1 (gray). (D) Superposition of two HANS motifs from BpaA (gray) and SadA (blue). Deletion of the second residue in the BpaA HANS motif alters the angle of the preceding coiled coil. (E–G) Alignment of the consensus sequences of (E) DALL domain variants, (F) neck variants, and (G) the HANS motif. The latter is shown together with individual HANS motifs of SadA and BpaA from panel D. (E–G) Red highlighting indicates eponymous motifs, bold print indicates β-layers. Secondary structure annotation: b: β-bulge; h: helix; s: strand.

which is anchored by a central water molecule (Fig. 5). We refer to this motif as the β-layer.

Necks generally end in the sequence QL (DL in IS necks), with the leucine forming the first core position (*a*) of the following coiled coil. In cases where this coiled coil is right-handed, the transition may contain a YxD motif (see above), in which case the neck ends in the sequence YV, not QL. We refer to this variant as the YV neck (Fig. 4F). Further, rarer versions of the neck appear to exist, which we will gradually add to the classification as data become available. Beyond this we note that a number of β-to-α transition regions are still unannotated. These regions do not have any of the sequence properties characteristic of necks and probably represent topologically novel connectors.

Currently, two types of α-to-β connectors are classified in daTAA, HANS and DALL, named for prominent motifs in their consensus sequences (Fig. 4E and G). HANS motifs mediate an abrupt transition from stalk segments to YadA-like heads (see the next section) by the formation of a short β-hairpin, which interacts with the YadA-like head. The segment between helix and strand in the HANS motif appears to allow for some conformational flexibility and may provide a bending site for the TAA fiber (Hartmann et al., 2012). Although the characteristic residues of the KYFHANS consensus motif are usually easy to recognize, they may on occasion show considerable divergence. In some proteins, only the length, location and general hydrophobic pattern of a connector between coiled coil and YadA-like head let us suspect the presence of a particularly divergent HANS motif, but these may well turn out to be novel α-to-β connectors.

The second widespread type of α-to-β connector is DALL, which encompasses three main variants (Fig. 4A, B, and E). Of these, DALL2

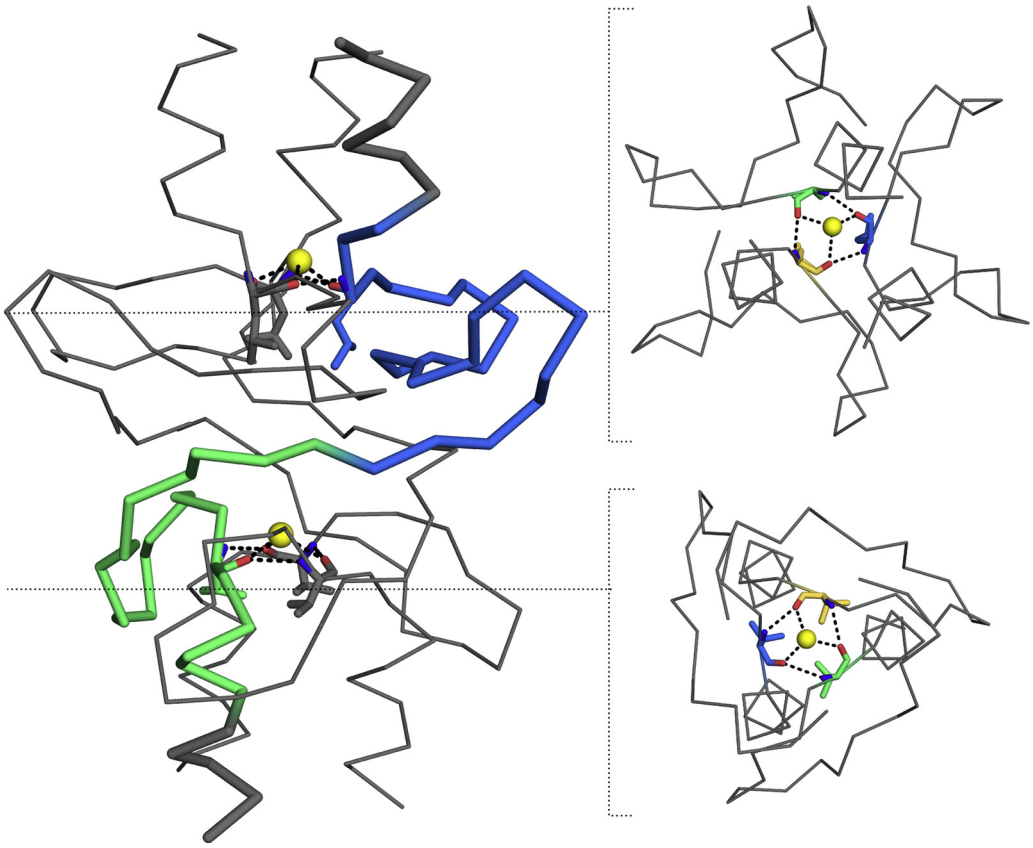


Fig. 5. The β-layers of DALL (blue) and neck (green) in side and top view from the SadA fiber. β-Layers mediate secondary structure transitions from α to β (in DALLs) and from β to α (in necks). The backbone atoms of the third residue in the respective β-layer motif (Fig. 4E and F) coordinate a single central water molecule (yellow).

was the first one discovered and it is the only one that carries the signature DALL motif in its consensus sequence (Fig. 4E). DALLs are peculiar in that they do not mediate a transition from stalk to head. Rather, they are invariably followed by a neck and the DALL-neck tandem can be considered a minimal-size α-to-β-to-α unit, topologically akin to a head domain (Figs. 4A and 5). DALLs contain two β-strands, which form a hairpin perpendicular to the fiber axis and connect to the first β-strand of the following neck. In DALL1, the β-strands of the hairpin are separated by bridging water molecules (Hartmann et al., 2012). As in HANS motifs, the α-to-β transition in DALLs is conformationally flexible and may also allow for fiber bending. Conceptually, DALL domains can be viewed as upside-down necks, with the transition between α and β mediated by a β-layer (Fig. 5). In DALL, the β-layer occurs at the beginning of the domain, at an α-to-β junction, rather than as in the neck at the end of the domain, at a β-to-α junction. In contrast, HANS does not contain a β-layer.

Heads

Heads are composed of β-strands and two basic topologies can be distinguished: transversal and interleaved. In transversal heads, the β-strands are oriented perpendicularly to the fiber axis whereas in interleaved heads, they run in parallel to it (Fig. 6). The domains with transversal topology currently classified in daTAA are YadA-like heads (Ylhead) and GIN domains; both transition into necks. The domains with interleaved topology are tryptophan ring domains (TrpRing), FxG domains, and GANG domains; all three generally transition into GIN domains.

The most widespread head domains are Ylheads, trimers of left-handed β-helices in which each turn of the helix formed by a single Ylhead motif (Fig. 6A). Ylheads are not only repetitive in structure, but also in sequence. In its canonical form, each Ylhead repeat consists of 14 residues, with 7 residues forming the inner face (inner strand and turn to the outer strand), and 7 the outer face (outer strand and turn to the inner stand). The residue patterns of the inner and outer faces are therefore similar (Fig. 6C), with two main differences. First, the β-strand residues pointing away from the β-helix are hydrophobic in the case of the inner strands, as they form the trimer core (Fig. 6), and hydrophilic in the case of the outer strands, as they are oriented toward the solvent. Second, the turns connecting the inner to the outer strands are structurally constrained, as they point toward another subunit, whereas the turns connecting the outer to the inner strands are not, as they point away from the trimer. For this reason, the inner turns almost never carry insertions, whereas the outer turns do so on occasion, and the glycine in the inner turns is conserved nearly invariantly, whereas it is often absent in the outer turns.

The most common form of insertion into Ylheads occurs in the outer turn leading to the last inner strand of the domain, before the neck is reached. Insertions at this location carry conserved sequence motifs, three of which we have classified as head insert motifs 1–3 (HIM1-3; Fig. 7). Structures of HIM2 from BpaA (Edwards et al., 2010) and UspA1 (Agnew et al., 2011), and for HIM3 from SadA (Hartmann et al., 2012) show that HIMs follow a common principle, extending downward over the neck to make contact with the stalk (Fig. 7A), allowing the head to grip the following segment and stiffen the fiber. Although the structures do not suggest steric conflicts with particular neck variants, HIM1 and HIM3 always occur with short necks and HIM2 with long necks.

The other head domain with transversal topology is GIN, named for a common motif in its last strand before the following neck

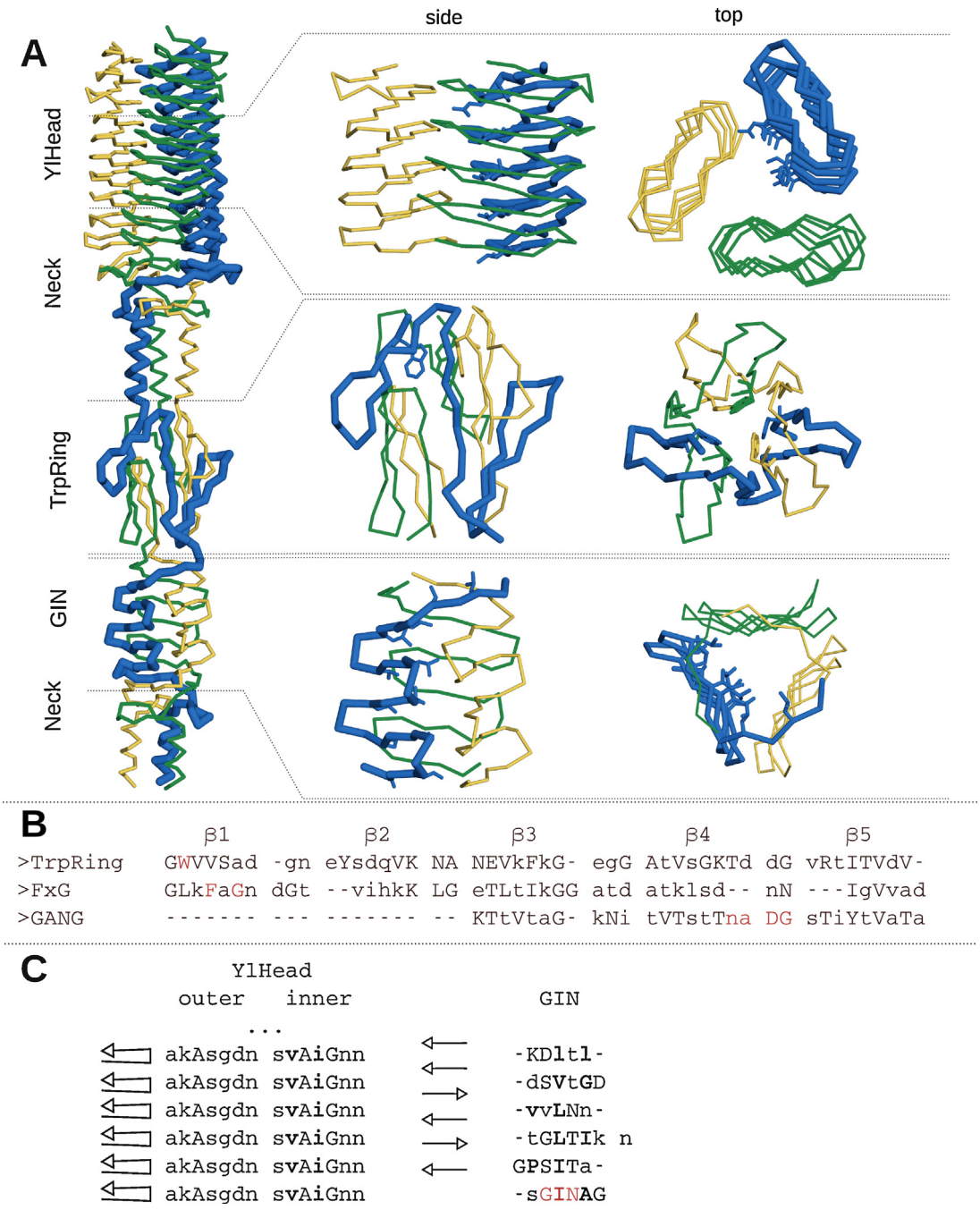


Fig. 6. Head domains. (A) A reconstructed segment of the BadA fiber containing a YadA-like head (Ylhead), a Tryptophan Ring domain (TrpRing), and a GIN domain. The head domains are shown individually as well, in side and top view. The eponymous tryptophan residues of the TrpRing domain and the core hydrophobic residues in the β-sheets of Ylhead and GIN are shown in stick representation. (B) Alignment of the consensus sequences of the TrpRing, FxG and GANG domains. (C) Comparison of Ylhead and GIN consensus sequences. The Ylhead is formed by repetition of the Ylhead motif, each forming an outer and an inner strand of the β-helix spiral. GIN is less obviously repetitive and consists of a single β-sheet, whose first strand represents a topological crossover from an adjacent subunit. Both GIN and the inner strands of Ylhead form a β-prism, whose core is formed by two stripes of hydrophobic residues from each subunit (bold print). Red highlighting indicates eponymous motifs.

(Fig. 6A and C). GIN domains consist of a single, mainly antiparallel β-sheet, which trimerizes to form a triangular prism. The first strand is followed by a topological crossover (Fig. 6C), after which the remaining strands meander downward in antiparallel arrangement. Each β-strand projects a core hydrophobic residue toward the center of the prism (Fig. 6), flanked by a second hydrophobic position, which, depending on the direction of the strand, is two residues before or after. GIN domains are structurally repetitive, but this repetition is not recognizable from their sequences.

Unlike GIN domains and Ylheads, which are analogous, the three interleaved heads are homologous and, though divergent, their sequences can be related to each other by profile Hidden Markov Model comparisons. The most frequent interleaved head is the tryptophan ring domain (TrpRing), named for a highly conserved tryptophan residue at the beginning of the first strand, which together with the equivalent tryptophans of the other two subunits forms a ring with a geometry unique in proteins of known structure (Fig. 6). TrpRing domains consist of five β-strands with

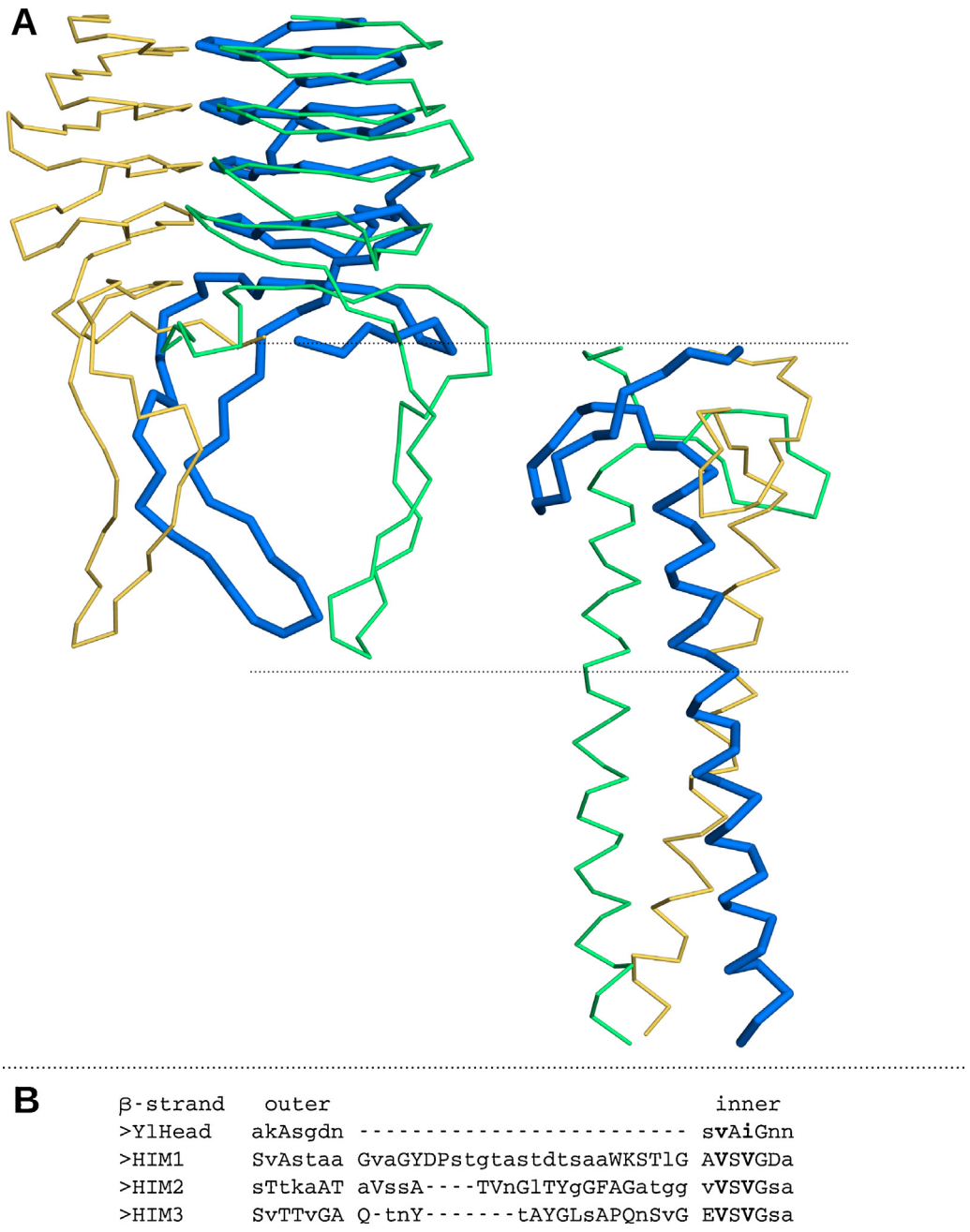


Fig. 7. Head insert motifs (HIM). (A) Four Ylhead repeats followed by a head insert motif (HIM2 of UspA1), a neck and a stalk segment. Head-HIM and neck-stalk are separated, in order to illustrate how the HIM insertion allows the head to grip the following stalk. (B) Alignment of the consensus sequences for HIM variants 1–3, compared to the consensus sequence of a regular Ylhead. Bold letters indicate hydrophobic residues of the inner β-sheet.

up-and-down topology, which interleave to form three β-sheets, each containing strands from all three subunits.

The N- and C-termini of TrpRings are fairly close to the trimer axis, so they do not need connectors to make the transition from stalk segments or into them. Indeed, at their N-terminus, TrpRings usually connect to the coiled coil of a neck. At their C-terminus, however, they either connect to GIN or, occasionally, to another interleaved head domain. Very rarely, they connect directly to a truncated neck variant, KG. We have never observed a coiled-coil segment at the C-terminus of TrpRings, even though this should be topologically unproblematic. The reason for this is unclear, but we note that in a number of TAAs from diverse bacteria, TrpRing domains are followed by a region with repeated Gly-X-Y patterns, which may represent a novel, collagen-like stalk.

Although only the crystal structure of the TrpRing is currently known, from Hia (Yeo et al., 2004; Meng et al., 2008) and from BadA (Szczesny et al., 2008), the robust sequence profile matches to FxG and GANG domains allow the expectation that these will show substantially the same interleaved topology. Because GANG domains are shorter and only match the C-terminal three strands of TrpRing and FxG, we anticipate that they represent a truncated form of interleaved head. In terms of domain architecture, they are found in the same context as TrpRings, generally connected at their N-terminal end to a neck and at their C-terminal end to GIN, but able to connect at either end to another interleaved head domain.

Role of structural domains in adhesion

The functional role of TAAs is adhesion, both to other cells and to surfaces in the environment. For the pathogenic bacteria of animals, which are the ones that have been studied in by far the greatest detail, this means autoagglutination (for the formation of biofilms) and adherence to the host tissue. The host molecules bound most frequently are components of the extracellular matrix, such as collagen, fibronectin, and laminin (e.g. Nummelin et al., 2004; Kaiser et al., 2012; Leo et al., 2010; Valle et al., 2008); and of the immune response, like immunoglobulins, carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1), and factor H (e.g. Leo and Goldman, 2009; Conners et al., 2008; Biedzka-Sarek et al., 2008). Other TAAs, however, also bind abiotic surfaces like plastics, glass, or steel (Ishikawa et al., 2012).

It has been the goal of many studies to narrow down these binding activities to specific sites on the TAA fibers, but this appears to have been successful in only two cases: for the IgA and IgG binding sites on the EibD stalk (Leo et al., 2011) and the CEACAM1 binding site on the UspA1 stalk (Conners et al., 2008). In other cases, such as the binding of collagen to the YadA head (Nummelin et al., 2004; Leo et al., 2010) or of factor H to the YadA stalk (Biedzka-Sarek et al., 2008), the studies showed the existence of multiple, promiscuous, low-affinity binding sites along extended parts of the fiber. By and large, it would appear that compact, globular ligands, such as immunoglobulin variable domains and CEACAM1 are bound with high affinity via specific sites, whereas extended molecules, such as collagen and factor H, are bound with high avidity via multible non-specific sites. Correspondingly, the hope to correlate specific binding activities with individual TAA domains, which we had when we set up daTAA, does not appear realistic.

Conclusion

As outlined here, the structure of TAAs lends itself to a description domain by domain. However, as with the species concept in organismic biology, to which Charles Darwin remarked in the *Origin of Species* that "I was much struck how entirely vague and arbitrary is the distinction between species and varieties", we are faced with a number of issues regarding the domain concept. Keeping in mind the biophysical definition of domains as autonomously folding units of the polypeptide chain, it is not clear how many of the TAA "domains" would actually satisfy this criterion. In our nomenclature, we have referred to segments as domains if they include a recognizable, geometrically conserved hydrophobic core, and have otherwise referred to them as motifs.

A second area of uncertainty concerns domain (and motif) boundaries. In some cases, these are fairly obvious. For example, the boundaries of Ylheads or TrpRings are not in doubt, within a few residues leeway. In other cases, however, things are less clear. For example, should FGG be seen as a motif inserted into a coiled-coil stalk domain (such as for example HIMs would be seen as motifs inserted into Ylhead domains), or should the formation of a hydrophobic core make FGG a domain, consisting jointly of the helices and strands? We have ourselves not been consistent on this issue, calling FGG a domain previously (Hartmann et al., 2012), but opting for a motif today. Similar considerations apply to the question of whether the first strand of GIN domains, prior to the topological crossover, is actually part of GIN or should be considered separately as a connector (e.g. as in Szczesny et al., 2008), akin to the HANS motif.

A third area of uncertainty concerns our categories of heads, stalks and connectors. While these are very clean in simple cases such as YadA, where one head is connected to one stalk by one connector, without any elaborations or insertions, they become far less clean in complex TAAs. For example, should DALL-neck tandems, which have similar dimensions to interleaved heads, be considered head structures? What about FGGs? These considerations show that, as useful as it is, a "domain dictionary" always remains an approximation of reality.

This said, our dictionary approach, as implemented in daTAA, allows the automated annotation of, on average, about two-thirds of residues in a newly sequenced TAA. With improved domain identification methods and progress in sequence and structure databases, we anticipate to reach three-quarters in the near future.

Acknowledgements

We thank our many collaborators for fruitful discussions, particularly the groups of Dirk Linke (University of Oslo), Ingo Autenrieth (Tübingen University), Volkhard Kempf (University Hospital Frankfurt), Adrian Goldman (University of Leeds) and Katsutoshi Hori (University of Nagoya). Our work on trimeric autotransporters was supported by institutional funds of the Max Planck Society and by the Deutsche Forschungsgemeinschaft (SFB 766, TP B4).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijmm.2014.12.010.

References

Agnew, C., Borodina, E., Zaccai, N.R., Conners, R., Burton, N.M., Vicary, J.A., Cole, D.K., Antognozzi, M., Virji, M., Brady, R.L., 2011. Correlation of in situ mechanosensitive responses of the *Moraxella catarrhalis* adhesin UspA1 with fibronectin and receptor CEACAM1 binding. Proc. Natl. Acad. Sci. U.S.A. 108, 15174–15178, http://dx.doi.org/10.1073/pnas.1106341108.

Biedzka-Sarek, M., Salmenlinna, S., Gruber, M., Lupas, A.N., Meri, S., Skurnik, M., 2008. Functional mapping of YadA- and Ail-mediated binding of human factor H to *Yersinia enterocolitica* serotype O:3. Infect. Immun. 76, 5016–5027, http://dx.doi.org/10.1128/IAI 00314-08.

Bölin, I., Norlander, L., Wolf-Watz, H., 1982. Temperature-inducible outer membrane protein of *Yersinia pseudotuberculosis* and *Yersinia enterocolitica* is associated with the virulence plasmid. Infect. Immun. 37, 506–512.

Conners, R., Hill, D.J., Borodina, E., Agnew, C., Daniell, S.J., Burton, N.M., Sessions, R.B., Clarke, A.R., Catto, L.E., Lammie, D., Wess, T., Brady, R.L., Virji, M., 2008. The Moraxella adhesin UspA1 binds to its human CEACAM1 receptor by a deformable trimeric coiled-coil. EMBO J. 27, 1779–1789, http://dx.doi.org/10.1038/emboj.2008.101.

Edwards, T.E., Phan, I., Abendroth, J., Dieterich, S.H., Masoudi, A., Guo, W., Hewitt, S.N., Kelley, A., Leibly, D., Brittnacher, M.J., Staker, B.L., Miller, S.I., van Voorhis, W.C., Myler, P.J., Stewart, L.J., 2010. Structure of a *Burkholderia pseudomallei* Trimeric Autotransporter Adhesin head. PLoS ONE 5, http://dx.doi.org/10.1371/journal.pone.0012803.

Fairman, J.W., Dautin, N., Wojtowicz, D., Liu, W., Noinaj, N., Barnard, T.J., Udho, E., Przytycka, T.M., Cherezov, V., Buchanan, S.K., 2012. Crystal structures of the outer membrane domain of intimin and invasin from enterohemorrhagic *E. coli* and enteropathogenic *Y. pseudotuberculosis*. Structure 20, 1233–1243, http://dx.doi.org/10.1016/j.str.2012.04.011.

Grosskinsky, U., Schütz, M., Fritz, M., Schmid, Y., Lamparter, M.C., Szczesny, P., Lupas, A.N., Autenrieth, I.B., Linke, D., 2007. A conserved glycine residue of trimeric autotransporter domains plays a key role in *Yersinia* adhesin A autotransport. J. Bacteriol. 189, 9011–9019, http://dx.doi.org/10.1128/JB 00985-07.

Gruber, M., Lupas, A.N., 2003. Historical review: another 50th anniversary—new periodicities in coiled coils. Trends Biochem. Sci. 28, 679–685, http://dx.doi.org/10.1016/j.tibs.2003.10.008.

Hartmann, M.D., Grin, I., Dunin-Horkawicz, S., Deiss, S., Linke, D., Lupas, A.N., Hernandez Alvarez, B., 2012. Complete fiber structures of complex trimeric autotransporter adhesins conserved in enterobacteria. Proc. Natl. Acad. Sci. U.S.A. 109, 20907–20912, http://dx.doi.org/10.1073/pnas.1211872110.

Hartmann, M.D., Ridderbusch, O., Zeth, K., Albrecht, R., Testa, O., Woolfson, D.N., Sauer, G., Dunin-Horkawicz, S., Lupas, A.N., Alvarez, B.H., 2009. A coiled-coil motif that sequesters ions to the hydrophobic core. Proc. Natl. Acad. Sci. U.S.A. 106, 16950–16955, http://dx.doi.org/10.1073/pnas.0907256106.

Hernandez Alvarez, B., Gruber, M., Ursinus, A., Dunin-Horkawicz, S., Lupas, A.N., Zeth, K., 2010. A transition from strong right-handed to canonical left-handed supercoiling in a conserved coiled-coil segment of trimeric autotransporter adhesins. J. Struct. Biol. 170, 236–245, http://dx.doi.org/10.1016/j.jsb.2010.02.009.

Hernandez Alvarez, B., Hartmann, M.D., Albrecht, R., Lupas, A.N., Zeth, K., Linke, D., 2008. A new expression system for protein crystallization using trimeric coiled-coil adaptors. Protein Eng. Des. Sel. 21, 11–18, http://dx.doi.org/10.1093/protein/gzm071.

Hoiczyk, E., Roggenkamp, A., Reichenbecher, M., Lupas, A., Heesemann, J., 2000. Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins. EMBO J. 19, 5989–5999, http://dx.doi.org/10.1093/emboj/19.22.5989.

Ishikawa, M., Nakatani, H., Hori, K., 2012. AtaA, a new member of the trimeric autotransporter adhesins from *Acinetobacter* sp, Tol 5 mediating high adhesiveness to various abiotic surfaces. PLoS ONE 7 (11), http://dx.doi.org/10.1371/journal.pone.0048830.

Kaiser, P.O., Linke, D., Schwarz, H., Leo, J.C., Kempf, V.A.J., 2012. Analysis of the BadA stalk from *Bartonella henselae* reveals domain-specific and domain-overlapping functions in the host cell infection process. Cell. Microbiol. 14, 198–209, http://dx.doi.org/10.1111/j.1462-5822.2011.01711.x.

Kammerer, R.A., Kostrewa, D., Progias, P., Honnappa, S., Avila, D., Lustig, A., Winkler, F.K., Pieters, J., Steinmetz, M.O., 2005. A conserved trimerization motif controls the topology of short coiled coils. Proc. Natl. Acad. Sci. U.S.A. 102, 13891–13896, http://dx.doi.org/10.1073/pnas.0502390102.

Koretke, K.K., Szczesny, P., Gruber, M., Lupas, A.N., 2006. Model structure of the prototypical non-fimbrial adhesin YadA of *Yersinia enterocolitica*. J. Struct. Biol. 155, 154–161, http://dx.doi.org/10.1016/j.jsb.2006.03.012.

Lafontaine, E.R., Cope, L.D., Aebi, C., Latimer, J.L., McCracken, G.H., Hansen, E.J., 2000. The UspA1 protein and a second type of UspA2 protein mediate adherence of *Moraxella catarrhalis* to human epithelial cells in vitro. J. Bacteriol. 182, 1364–1373, http://dx.doi.org/10.1128/JB.182.5.1364-1373.2000.

Lehr, U., Schütz, M., Oberhettinger, P., Ruiz-Perez, F., Donald, J.W., Palmer, T., Linke, D., Henderson, I.R., Autenrieth, I.B., 2010. C-terminal amino acid residues of the trimeric autotransporter adhesin YadA of *Yersinia enterocolitica* are decisive for its recognition and assembly by BamA. Mol. Microbiol. 78, 932–946, http://dx.doi.org/10.1111/j.1365-2958.2010.07377.x.

Leo, J.C., Elovaara, H., Bihan, D., Pugh, N., Kilpinen, S.K., Raynal, N., Skurnik, M., Farndale, R.W., Goldman, A., 2010. First analysis of a bacterial collagen-binding protein with collagen Toolkits: promiscuous binding of YadA to collagens may explain how YadA interferes with host processes. Infect. Immun. 78, 3226–3236, http://dx.doi.org/10.1128/IAI 01057-09.

Leo, J.C., Goldman, A., 2009. The immunoglobulin-binding Eib proteins from *Escherichia coli* are receptors for IgG Fc. Mol. Immunol. 46, 1860–1866, http://dx.doi.org/10.1016/j.molimm.2009.02.024.

Leo, J.C., Grin, I., Linke, D., 2012. Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. Philos. Trans. R. Soc. B: Biol. Sci. 1592, 1088–1101, http://dx.doi.org/10.1098/rstb.2011.0208.

Leo, J.C., Lyskowski, A., Hattula, K., Hartmann, M.D., Schwarz, H., Butcher, S.J., Linke, D., Lupas, A.N., Goldman, A., 2011. The structure of *E. coli* IgG-binding protein D suggests a general model for bending and binding in trimeric autotransporter adhesins. Structure 19, 1021–1030, http://dx.doi.org/10.1016/j.str.2011.03.021.

Linke, D., Riess, T., Autenrieth, I.B., Lupas, A., Kempf, V.A.J., 2006. Trimeric autotransporter adhesins: variable structure, common function. Trends Microbiol. 14, 264–270, http://dx.doi.org/10.1016/j.tim.2006.04.005.

Lupas, A.N., Gruber, M., 2005. The structure of alpha-helical coiled coils. Adv. Protein Chem. 70, 37–78, http://dx.doi.org/10.1016/S0065-3233(05)70003-6.

Łyskowski, A., Leo, J.C., Goldman, A., 2011. Structure and biology of trimeric autotransporter adhesins. Adv. Exp. Med. Biol. 715, 143–158, http://dx.doi.org/10.1007/978-94-007-0940-9_9.

Meng, G., Geme, St., Waksman, J.W.G., 2008. Repetitive architecture of the *Haemophilus influenzae* Hia trimeric autotransporter. J. Mol. Biol. 384, 824–836, http://dx.doi.org/10.1016/j.jmb.2008.09.085.

Meng, G., Surana, N.K., St Geme, J.W., Waksman, G., 2006. Structure of the outer membrane translocator domain of the *Haemophilus influenzae* Hia trimeric autotransporter. EMBO J. 25, 2297–2304, http://dx.doi.org/10.1038/sj.emboj.7601132.

Nummelin, H., Merckel, M.C., Leo, J.C., Lankinen, H., Skurnik, M., Goldman, A., 2004. The *Yersinia* adhesin YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. EMBO J. 23, 701–711, http://dx.doi.org/10.1038/sj.emboj.7600100.

Oomen, C.J., van Ulsen, P., van Gelder, P., Feijen, M., Tommassen, J., Gros, P., 2004. Structure of the translocator domain of a bacterial autotransporter. EMBO J. 23, 1257–1266, http://dx.doi.org/10.1038/sj.emboj.7600148.

Raghunathan, D., Wells, T.J., Morris, F.C., Shaw, R.K., Bobat, S., Peters, S.E., Paterson, G.K., Jensen, K.T., Leyton, D.L., Blair, J.M.A., Browning, D.F., Pravin, J., Flores-Langarica, A., Hitchcock, J.R., Moraes, C.T.P., Piazza, R.M.F., Maskell, D.J., Webber, M.A., May, R.C., MacLennan, C.A., Piddock, L.J., Cunningham, A.F., Henderson, I.R., 2011. SadA, a trimeric autotransporter from *Salmonella enterica* serovar typhimurium, can promote biofilm formation and provides limited protection against infection. Infect. Immun., http://dx.doi.org/10.1128/IAI 05592-11.

Remmert, M., Biegert, A., Linke, D., Lupas, A.N., Söding, J., 2010. Evolution of outer membrane beta-barrels from an ancestral beta beta hairpin. Mol. Biol. Evol. 27, 1348–1358, http://dx.doi.org/10.1093/molbev/msq017.

Riess, T., Andersson, S.G.E., Lupas, A., Schaller, M., Schäfer, A., Kyme, P., Martin, J., Wälzlein, J.-H., Ehehalt, U., Lindroos, H., Schirle, M., Nordheim, A., Autenrieth, I.B., Kempf, V.A.J., 2004. *Bartonella* adhesin a mediates a proangiogenic host cell response. J. Exp. Med. 200, 1267–1278, http://dx.doi.org/10.1084/jem.20040500.

Sandt, C.H., Hill, C.W., 2000. Four different genes responsible for nonimmune immunoglobulin-binding activities within a single strain of *Escherichia coli*. Infect. Immun. 68, 2205–2214, http://dx.doi.org/10.1128/IAI.68.4.2205-2214.2000.

Sijbrandi, R., Urbanus, M.L., ten Hagen-Jongman, C.M., Bernstein, H.D., Oudega, B., Otto, B.R., Luirink, J., 2003. Signal recognition particle (SRP)-mediated targeting and Sec-dependent translocation of an extracellular *Escherichia coli* protein. J. Biol. Chem. 278, 4654–4659, http://dx.doi.org/10.1074/jbc.M211630200.

St Geme, J.W., Cutter, D., 2000. The *Haemophilus influenzae* Hia adhesin is an autotransporter protein that remains uncleaved at the C terminus and fully cell associated. J. Bacteriol. 182, 6005–6013, http://dx.doi.org/10.1128/JB.182.21.6005-6013.2000.

Szczesny, P., Linke, D., Ursinus, A., Bär, K., Schwarz, H., Riess, T.M., Kempf, V.A.J., Lupas, A.N., Martin, J., Zeth, K., 2008. Structure of the head of the *Bartonella* adhesin BadA. PLoS Pathog. 4, http://dx.doi.org/10.1371/journal.ppat.1000119.

Szczesny, P., Lupas, A., 2008. Domain annotation of trimeric autotransporter adhesins—daTAA. Bioinformatics 24, 1251–1256, http://dx.doi.org/10.1093/bioinformatics/btn118.

Valle, J., Mabbett, A.N., Ulett, G.C., Toledo-Arana, A., Wecker, K., Totsika, M., Schembri, M.A., Ghigo, J.-M., Beloin, C., 2008. UpaG, a new member of the trimeric autotransporter family of adhesins in uropathogenic *Escherichia coli*. J. Bacteriol. 190, 4147–4161, http://dx.doi.org/10.1128/JB 00122-0.

Yeo, H.-J., Cotter, S.E., Laarmann, S., Juehne, T., St Geme, J.W., Waksman, G., 2004. Structural basis for host recognition by the *Haemophilus influenzae* Hia autotransporter. EMBO J. 23, 1245–1256, http://dx.doi.org/10.1038/sj.emboj.7600142.