

# Complete fiber structures of complex trimeric autotransporter adhesins conserved in enterobacteria

Marcus D. Hartmann, Iwan Grin, Stanislaw Dunin-Horkawicz, Silvia Deiss, Dirk Linke, Andrei N. Lupas<sup>1</sup>, and Birte Hernandez Alvarez

Department of Protein Evolution, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany

Edited by William I. Weis, Stanford University School of Medicine, Stanford, CA, and accepted by the Editorial Board November 8, 2012 (received for review July 15, 2012)

Trimeric autotransporter adhesins (TAAs) are modular, highly repetitive surface proteins that mediate adhesion to host cells in a broad range of Gram-negative pathogens. Although their sizes may differ by more than one order of magnitude, they all follow the same basic head-stalk-anchor architecture, where the head mediates adhesion and autoagglutination, the stalk projects the head from the bacterial surface, and the anchor provides the export function and attaches the adhesin to the bacterial outer membrane after export is complete. In complex adhesins, head and stalk domains may alternate several times before the anchor is reached. Despite extensive sequence divergence, the structures of TAA domains are highly constrained, due to the tight interleaving of their constituent polypeptide chains. We have therefore taken a “domain dictionary” approach to characterize representatives for each domain type by X-ray crystallography and use these structures to reconstruct complete TAA fibers. With SadA from *Salmonella enterica*, EhaG from enteropathogenic *Escherichia coli* (EHEC), and UpaG from uropathogenic *E. coli* (UPEC), we present three representative structures of a complex adhesin that occur in a conserved genomic context in Enterobacteria and is essential in the infection process of uropathogenic *E. coli*. Our work proves the applicability of the dictionary approach to understanding the structure of a class of proteins that are otherwise poorly tractable by high-resolution methods and provides a basis for the rapid and detailed annotation of newly identified TAAs.

coiled coil |  $\beta$ -layer

Gram-negative pathogenic bacteria express numerous different virulence factors to overcome host defenses, to mobilize and take up nutrients within the host, to invade cells or tissues, and to adhere to host cells (1). Adhesion to host cells is a key event during the onset of infection; the mediators of this process, called adhesins, are a heterogeneous group of bacterial surface proteins, which vary in architecture, domain content, and mode of binding. One distinct class of adhesins are the trimeric autotransporter adhesins (TAAs) (2, 3), also referred to as type Vc secretion systems (4). TAAs are important virulence factors of many well-studied pathogens: Examples include YadA of *Yersinia enterocolitica*, a species causing enteritis, mesenteric lymphadenitis, and reactive arthritis (5); NadA of *Neisseria meningitidis* (6), an agent of meningitis and sepsis; BadA of *Bartonella henselae* (7), which is the agent of cat scratch disease; UspA1 and A2 of *Moraxella catarrhalis* (8), a prominent species in respiratory tract infections, and Hia of *Haemophilus influenzae* (9), an organism causing meningitis and respiratory tract infections. Despite their role in the context of unrelated diseases, these TAAs always fulfill similar functions—adhesion to host cells, autoagglutination, and biofilm formation (3).

All TAAs display the same basic architecture: The N-terminal head typically mediates molecular interactions such as autoagglutination or binding to extracellular matrix proteins. It is followed by an extended and typically coiled-coil rich stalk, which projects the head from the bacterium and often provides binding sites for host serum factors (10, 11). The protein ends in a membrane anchor (2). In architecturally complex adhesins, head and stalk segments may alternate several times before the anchor

is reached (12). Whereas head and stalk are assembled from an array of analogous domains (13), the anchor is homologous in all TAAs and represents the defining element of this protein family (2). It trimerizes in the outer membrane to form a 12-stranded  $\beta$ -barrel pore (14), through which the head and the stalk exit the periplasm, thus giving rise to the name “autotransporter.” The C-terminal end of the folded stalk occludes the pore after export is completed.

A number of partial TAA structures were solved recently. Several head structures, from YadA (15), Hia (16), BadA (13), and BpaA (17), revealed different trimeric complexes with novel folds. Partial stalk structures from UspA1 (18), SadA (19), and YadA (20) substantiated earlier predictions that coiled coils are the dominant structural motif of TAA stalks, albeit sometimes with noncanonical properties such as unusual periodicities or ion binding sites in their core. Finally, one structure of a TAA membrane anchor could also be determined, from Hia (14), showing a size and architecture similar to that of single-chain autotransporters, albeit built of three chains rather than a single one.

Despite their strong sequence divergence, structures of homologous TAA domains are so closely conserved that one structure can be used to solve the next one by molecular replacement (13, 21); this characteristic, and the fact that the domains can be predicted from sequence using state-of-the-art homology detection methods (12), prompted us to suggest a dictionary approach to understand the structure of TAAs, given that their flexibility and extreme length otherwise precludes their analysis by high-resolution methods. We proposed to solve representatives for all TAA domains defined from sequence analysis, which could then be used to model full TAA fibers from fragments (13, 22). Strictly speaking, the term “domain,” which has been developed on globular proteins to denote independently folding units, does not fully describe the structural elements of TAA or oligomeric fibers in general. We therefore use the term here for a unique and complete TAA building block, defined evolutionarily as a segment with a specific structure that can be shuffled in TAAs with few constraints from adjacent segments. By this definition, not all TAA domains are independently folding units.

During our work on a web-based annotation platform for TAAs (12), we identified a chromosomally encoded TAA from *Enterobacteriaceae* as an excellent model system, due to its domain complexity and the genetic tractability of its parent organisms. This protein is called SadA in *Salmonella enterica* (22), EhaG in enteropathogenic *Escherichia coli* (EHEC), and UpaG in uropathogenic *E. coli* (UPEC). UpaG has been found to be

Author contributions: M.D.H., D.L., A.N.L., and B.H.A. designed research; M.D.H., I.G., S.D.-H., S.D., and B.H.A. performed research; M.D.H., A.N.L., and B.H.A. analyzed data; and M.D.H., D.L., A.N.L., and B.H.A. wrote the paper.

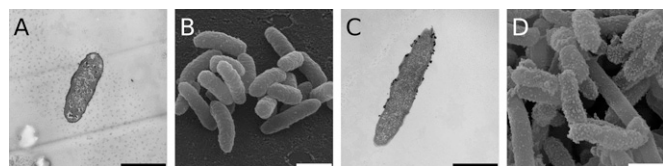
The authors declare no conflict of interest.

This article is a PNAS Direct Submission. W.I.W. is a guest editor invited by the Editorial Board.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, [www.pdb.org](http://www.pdb.org) [PDB ID codes 2YNY (SadAK1), 2YNZ (SadAK5), 2YO0 (SadAK9cfl), 2YO1 (SadAK9cfl), 2YO2 (SadAK12), and 2YO3 (SadAK14)].

<sup>1</sup>To whom correspondence should be addressed. E-mail: [andrei.lupas@tuebingen.mpg.de](mailto:andrei.lupas@tuebingen.mpg.de).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1211872110/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1211872110/-DCSupplemental).



**Fig. 1.** Electron microscopy of SadA. Transmission (A and C) and scanning electron micrographs (B and D) of *E. coli* Top10 carrying an overexpression vector for *Salmonella typhimurium* SadA (C and D) or an empty vector as control (A and B). Cells were grown in liquid culture and immobilized on polylysine-coated coverslips. In A and C, cells were labeled after immobilization with an antibody raised against SadA and affinity purified against SadAK9 and chemically fixed by using glutaraldehyde. (Scale bars: 1  $\mu$ m.)

essential for the colonization of the urinary tract by UPEC (23), whereas EhaG mediates binding of EHEC specifically to colorectal epithelium (24); SadA promotes biofilm formation and host cell adherence in *Salmonella* (25). Here, we reconstruct the full SadA, UpaG, and EhaG fibers from the structures of representative SadA fragments, describing in the process a number of unusual structural motifs with functional implications. These motifs include a coiled coil elaborated by a “collar” of three-stranded  $\beta$ -meanders, which may provide stiffness to the stalk; a  $\beta$ -layer motif, which acts as a universal adaptor for transitions between  $\alpha$ -helical and  $\beta$ -stranded domains; and a connector with an intrinsic rotational flexibility of up to 25°, which offers an attractive mechanism for fine-tuning the relative orientation of consecutive binding sites along the fiber. Our work proves the domain dictionary approach and provides a basis for the rapid and detailed annotation of newly identified TAAs.

## Results and Discussion

**Chromosomally Encoded Enterobacterial Adhesin.** Enterobacteria of the genera *Escherichia*, *Salmonella*, and *Shigella* contain a chromosomal TAA located between the *mil*-operon, which encodes genes responsible for mannitol metabolism, and the *lld*-operon, which contains genes for the uptake and metabolism of L-lactate (Fig. S1). The TAAs of pathogenic organisms are characterized by a high genetic turnover (2), probably due to selective pressure by the host immune system; correspondingly, in some strains, this adhesin may be disrupted by frame shifts, internal stop codons, or insertion elements. This observation is made only rarely in *Escherichia* and *Salmonella* species, but frequently in *Shigella*. In the nonpathogenic laboratory strain *E. coli* K12, this genomic region is lost entirely and the L-lactate operon is located directly downstream of the mannitol operon.

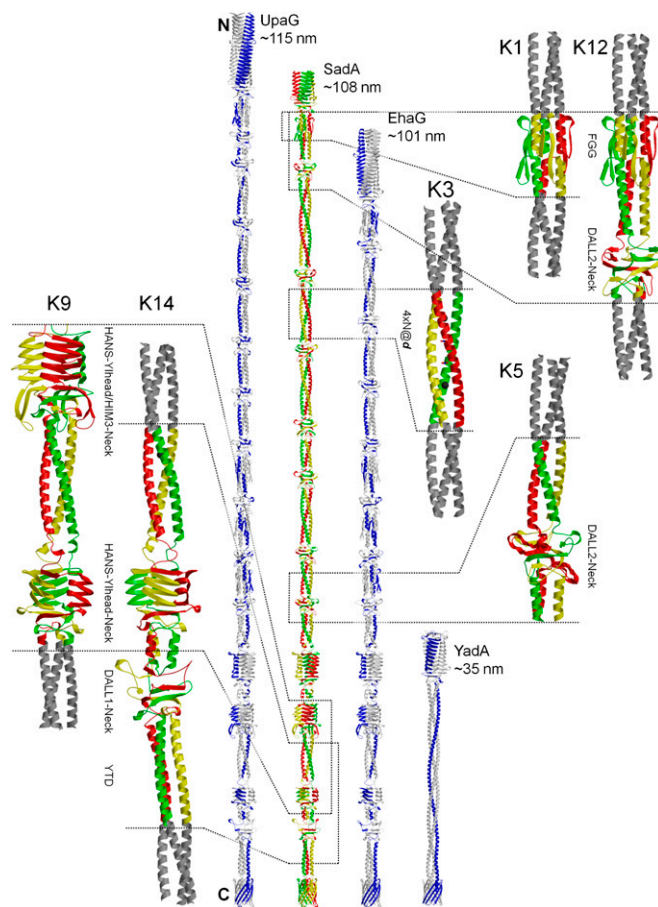
Early on in the project, we settled on the *Salmonella* ortholog, SadA, as a model system, because the laboratory strain LT2 contains this protein whereas the *Escherichia* laboratory strain K12 does not. We resorted to recombinant expression of SadA in *E. coli*. In scanning electron micrographs, SadA-expressing cells displayed a rough cell surface densely covered with knobs, in contrast to uninduced control cells. Transmission electron micrographs of sections labeled with anti-SadA antibodies confirmed the insertion of SadA into the outer membrane and its exposure to the cell surface (Fig. 1).

SadA is a protein of 1,461 residues with a complex domain composition. It includes a signal sequence of 54 residues, four YadaA-like head domains and extended stalk regions, multiply segmented by FGG, HANS, DALL, and neck motifs (for domain and motif definitions; ref. 12). Whereas the domain architecture of its N-terminal half is quite variable between orthologs in different species, its C-terminal half, encompassing three of the four YadaA-like heads and the membrane anchor, is highly conserved. Overall, SadA illustrates very well the modularity and the combination of conserved and variable regions characteristic for TAAs.

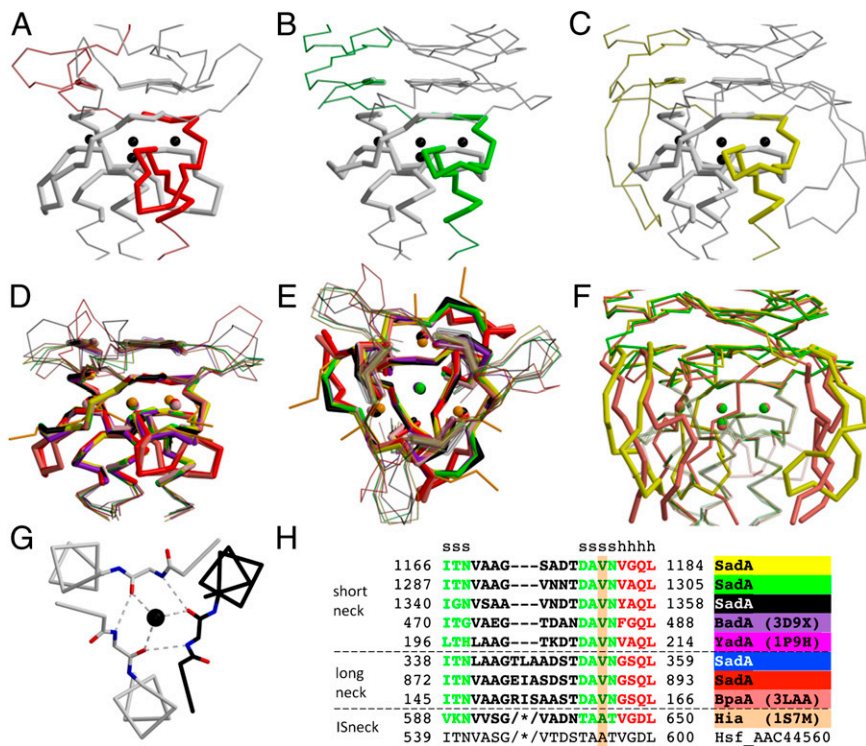
**Structure Determination of SadA Fragments and Reconstruction of Full Fibers.** To gain sufficient structural information for reconstruction of the full fiber, we designed a redundant set of SadA

fragments, SadAK1–16, each covering at least one copy of the domains and motifs found in the protein. With one exception (SadAK5), the constructs were designed to start and end with coiled-coil segments, so that they could be extended by adaptors derived from the trimeric form of the GCN4 leucine zipper, GCN4pII, to increase their stability (22). The constructs SadAK1, K3, K5, K9, K12, and K14, which together cover all desired domain types (Fig. 2 and Fig. S2), yielded crystals that diffracted to high resolution. We obtained their structures by molecular replacement in a hierarchical fashion: The shortest constructs, K1 and K3, were solved by using the trimeric GCN4 structure 1GCM as a search model and each further construct was solved with parts of the previous ones: K12 with K1, K5 with K12, K14 with K5, and K9 with K14. Whereas K1, K12, K3, and K14 were solved in their entirety, the C-terminal 43 residues of K5 were not visible and the N-terminal 41 residues of K9 were recognizable, but not traceable, in the electron density. We obtained structures for K9 from two different crystal forms, K9cfI and K9cfII, neither of which allowed us to trace the N-terminal part. All constructs were solved from crystals containing the full trimers in the asymmetric unit and display more or less pronounced asymmetries; only K12 and one K9 structure (K9cfI) were built by crystallographic symmetry and are therefore fully symmetrical.

On the basis of these structures, we reconstructed the whole fibers of SadA, UpaG, and EhaG according to the domain annotation in Figs. S2–S4. To yield straight fibers, the modeling



**Fig. 2.** Crystal structures of fragments and reconstructed full fibers of SadA, UpaG, and EhaG. The three chains of the SadA trimers are colored individually; the coiled-coil adaptors fused to the fragments are shown in gray. Only one of the two structures obtained for SadAK9 (K9cfI) is shown. The reconstructed fibers are compared with a prototypical simple trimeric adhesin, YadaA. Indicated lengths are measured on the atomic coordinates of the models and include the membrane anchor.

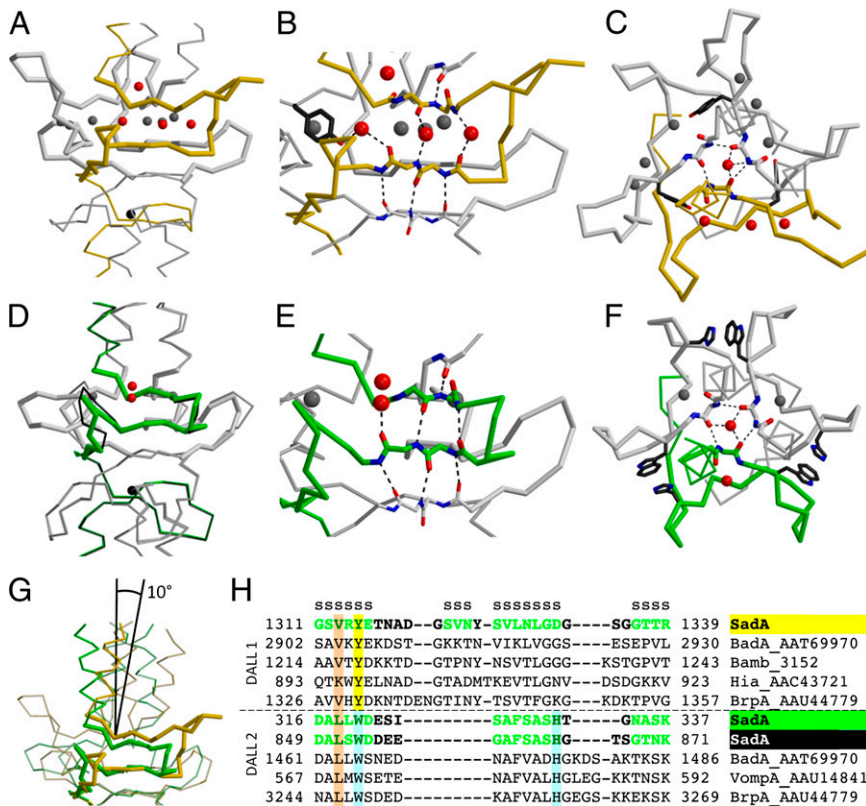


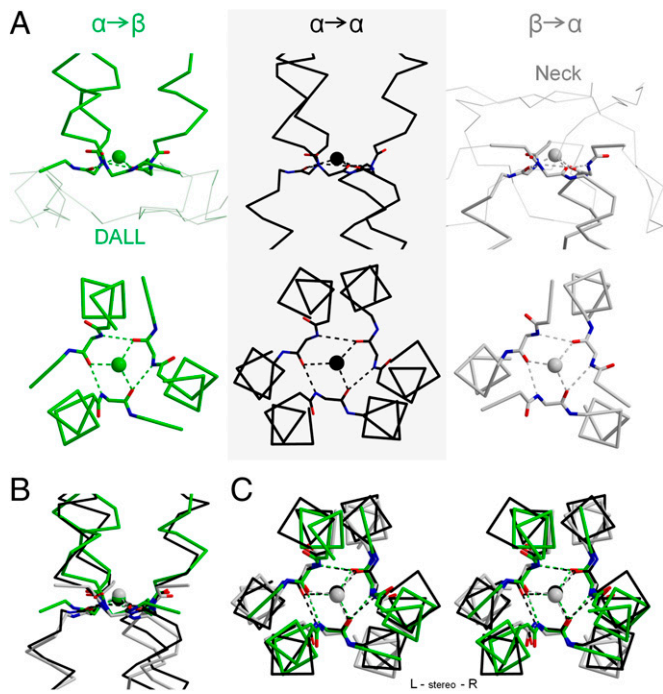
**Fig. 3.** Neck domains in context with different upstream domains. (A–C) Neck domains in SadA in context with different upstream domains. One chain of each trimer is colored individually. The neck (thick lines) forms a continuous  $\beta$ -sheet with the last  $\beta$ -strand of the respective upstream domain (transparently thick). Structurally invariant water molecules are in black. (A) Long neck following a DALL2 domain in SadAK5. (B) Short neck following a Yhead domain in SadAK14. (C) Short neck following a HIM3 domain in SadAK9. (D) Superposition of nine known neck structures as indicated in *H*, solved in context with their native upstream domain, colored by trimer. The insertion sequence in the neck of Hia (1S7M) was cut for clarity. (E) Top view of *D*. (F) Head insert motifs HIM2 and HIM3 fold tightly around their respective neck: Superimposition of *B* and *C* (HIM3) with a HIM2-neck domain, highlighting the head insert motif (thick lines). (G) Close up of the  $\beta$ -layer in the neck of SadAK5, highlighting the hydrogen bonding network that involves the central water molecule. (H) Sequence alignment of neck sequences, highlighting the central  $\beta$ -layer residue (mostly valine) from the DAVN consensus sequence.

process was conducted with symmetry constraints; the membrane anchor was modeled by using the Hia membrane anchor domain (PDB ID code 2GR8) as a template. With a length of  $\sim$ 115

nm, the model of UpaG is slightly longer than the  $\sim$ 108-nm-long SadA and the  $\sim$ 101-nm-long EhaG model. In the following, we will describe the structure of the individual domains and motifs.

**Fig. 4.** The DALL domain. (A) Structure of the DALL1 domain in SadAK14 together with the downstream neck domain. One chain is colored yellow; the DALL1 domain is drawn in thick lines. The central water molecule of the DALL1 and the water molecules interacting mostly with the yellow chain are drawn in red, the central water of the neck in black, and others in gray. Notably, the N-terminal coiled coil is kinked with respect to the trimer axis of the DALL1-neck tandem. (B) Close-up of the  $\beta$ -sheet on one face of the DALL1-neck tandem. Compared with DALL2 in *E*, the  $\beta$ -sheet of the DALL1 is invaded by bridging water molecules. (C) Top view of the DALL1 domain, highlighting the  $\beta$ -layer interactions. Although the backbone interactions of the central  $\beta$ -layer residues are almost undisturbed, the coordination of the central water molecule is asymmetric because of the kink of the coiled coil. (D) Structure of the DALL2 domain in SadAK12 together with the downstream neck domain. One chain is colored green; the DALL2 domain is drawn in thick. The central water molecule and the water interacting with the green chain are shown in red, the central water molecule of the neck in black, and others in gray. Superposed in thin black is the structure of the DALL2-neck tandem in SadAK5. (E) Close-up of the continuous  $\beta$ -sheet formed between the DALL2 and neck domain. (F) Top view of the DALL2 domain in SadAK12, highlighting the interchain interaction between the conserved Tryptophan and Histidine and the  $\beta$ -layer interactions. (G) Superposition of the DALL1 structure of SadAK14 and the DALL2 structure of SadAK12 on the C-terminal halves. Because of the waters invading the  $\beta$ -sheet in DALL1, the spacing of the  $\beta$ -strands differs between the two structures. The N-terminal coiled-coil of DALL1 in SadAK14 is  $10^\circ$  kinked with respect to the symmetric DALL2 in SadAK12. (H) Sequence alignment of DALL1 and DALL2 domains, highlighting the central  $\beta$ -layer residues and the conserved residues in interchain interactions.





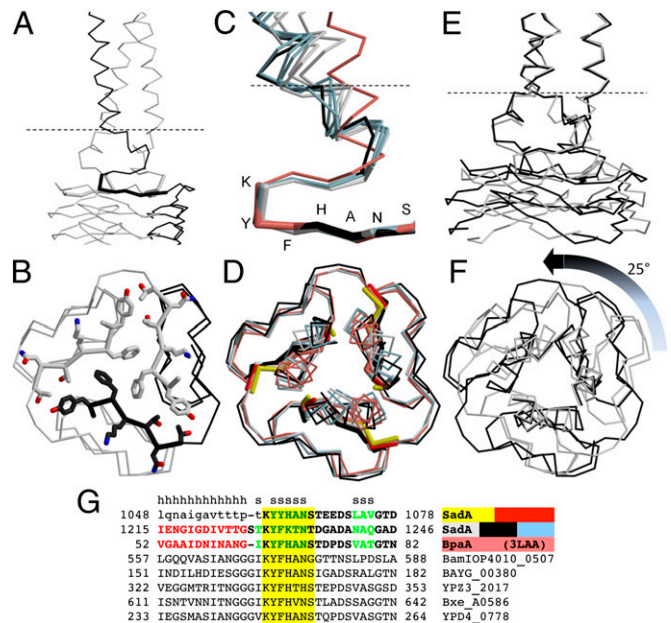
**Fig. 5.**  $\beta$ -Layers as universal adaptors. The structure of  $\beta$ -layers in three different contexts, highlighting the  $\beta$ -interactions of the three chains around the central water molecule. (A) Side and top views are shown for the  $\beta$ -layer in a DALL domain (Left; DALL2 from SadAK12) leading from alpha to beta, a  $\beta$ -layer in a neck domain (Right; from SadAK12), leading from beta to alpha, and a  $\beta$ -layer separating two coiled-coil segments, leading from alpha to alpha, from structure 2BA2 (Center). B and C show a superposition of the three different  $\beta$ -layers in side and top view, respectively. The superposition is based on the central  $\beta$ -layer residues and the central water molecule. With a total of 8 DALL and 12 neck domains, the whole SadA fiber comprises 20  $\beta$ -layers.

**YadA-Like Head Domains.** SadA contains four YadA-like head domains (Ylheads), of which we solved the third in construct K9 and the fourth in K9 and K14. Ylheads are trimers of single-stranded, left-handed  $\beta$ -helices with a repetitive substructure. The general repeat length is 14 residues, and each repeat consists of an outer and an inner  $\beta$ -strand, running perpendicularly to the fiber axis. The repeats stack to form continuous  $\beta$ -sheets running along the inner and outer faces of the  $\beta$ -helices (15). Like the heptads of coiled coils, individual repeats have a structure that is independent of their sequence, but short insertions and deletions of one or two residues are common and, in contrast to coiled coils, do not affect the overall structure. In the four head domains of SadA, 20 of 24 repeats are 14 residues long and three others have 12, 15, and 16 residues, respectively. Longer insertions occur on occasion and are almost invariably found in the last repeat of a head segment. These head insert motifs (HIMs) can be assigned to individual subtypes on the basis of their sequence properties (12). We identified two HIM subtypes previously, HIM1 and HIM2, and here observe a third type, HIM3, in the structure of the third Ylhead of SadA. C-terminally, Ylheads always merge into neck domains, continuing the inner  $\beta$ -sheet into the  $\beta$ -sheet of the latter. In the case of the third Ylhead of SadA (construct K9; Fig. 3C), a HIM3 forms additional interactions, folding tightly around the succeeding neck domain. Comparable interaction patterns can be observed for a HIM2 in the partial *Burkholderia pseudomallei* TAA head structure 3LAA (17) (Fig. 3F).

**Connectors from Beta to Alpha: Neck Domain.** Neck domains mediate the transition from a variety of  $\beta$ -stranded domains into downstream coiled-coil regions in TAAs (Fig. 3). They are rep-

resented by multiple structures in the Protein Data Bank in context with different upstream domains, to which we add another five exemplars from SadA. Necks can be subdivided into short necks, long necks, and necks with an insertion sequence (ISneck). They only differ in the length of the loop that connects their two  $\beta$ -strands. A superposition shows an invariant structure for all three subclasses, mediating a perfectly straight and symmetric transition into the downstream coiled coil (Fig. 3D and E). Moreover, the interactions with the upstream domains are conserved and independent of the nature of the latter. The last  $\beta$ -strand of the first chain of the upstream domain forms a continuous  $\beta$ -sheet with the first  $\beta$ -strand of the second chain and the second  $\beta$ -strand of the third chain of the neck, yielding a highly intertwined structure. This  $\beta$ -sheet is, apart from a common hydrophobic core, the only interaction the neck forms with its upstream domain, with the exception of Ylhead/HIM domains described above. Notably, HIM3 and HIM2 are always associated with short and long necks, respectively.

The crux of the neck is, however, located directly at the transition to the coiled coil where the three chains of the trimer cross each other and form a tight symmetric network of hydrogen bonds. Here, the valine of the consensus sequence DAVN from one chain forms  $\beta$ -sheet interactions with the valines of the other chains, and these three valines together coordinate a central water molecule on the trimer axis (Fig. 3G). This structural motif constitutes the core of the transition from the  $\beta$ -stranded structure of the neck to the  $\alpha$ -helical structure of the downstream coiled coil, forming a plane of  $\beta$ -sheet like interactions perpendicular to the coiled-coil axis. We refer to this motif as a  $\beta$ -layer and call the valine the central  $\beta$ -layer residue (see Fig. 5).



**Fig. 6.** The HANS domain. (A) Side view of the last HANS domain in SadA from structure K9cfl, second in the alignment in G. The dashed line indicates where the N-terminal coiled coil breaks into a second, distorted segment. (B) Top view of the same HANS domain, showing the side chains of the second  $\beta$ -strand. (C and D) Superposition of all structures of the last HANS domain of SadA (K9cfl black, K9cflII gray, K14 light blue) with the HANS from 3LAA (salmon) on the KYFHANS consensus motif, illustrating the conformational flexibility of the N-terminal helical segment. C shows a side view of a superposition of all individual chains, and D shows the top view of the full trimers. (E and F) Superposition of two different conformations of the last HANS domain in SadA (from K9cfl and from K9cflII) onto the coiled coil above the distorted segment (indicated by the dashed line). Between the two conformations, the C-terminal Ylhead is rotated by 25° with respect to the N-terminal coiled-coil. (G) Sequence alignment of HANS domains.

**Connectors from Alpha to Beta.** *DALL domain.* SadA comprises eight DALL domains, one of type 1 and seven of type 2, which always occur in tandems with neck domains; K5 and K12 cover two different DALL2-neck tandems, K14 the DALL1-neck tandem. DALL domains mediate the transition from preceding coiled-coil segments to the  $\beta$ -stranded structure of their respective neck domains and form continuous  $\beta$ -sheets with them.

Architecturally, DALL2 is simpler than DALL1 (Fig. 4). DALL2 forms a  $\beta$ -sheet between the first and second  $\beta$ -strand of the same chain that continues into the neck domain. A conserved tryptophan in the first and a conserved histidine in the second strand form a  $\pi$ -stacking interaction between the chains. In DALL1, which carries an insertion between the first and second strand, the  $\beta$ -sheet is invaded by bridging water molecules (Fig. 4B). Consequently, the two strands have a wider spacing so that DALL1 and DALL2 can be structurally superimposed only on either their N- or C-terminal halves (Fig. 4G). Compared with the neck domain, the transition mediated by DALL domains has a considerable degree of conformational flexibility; the coiled-coil axis preceding the DALL1 domain in the structure of K14 is kinked  $\sim 10^\circ$  with respect to the trimer axis of the DALL1-neck tandem.

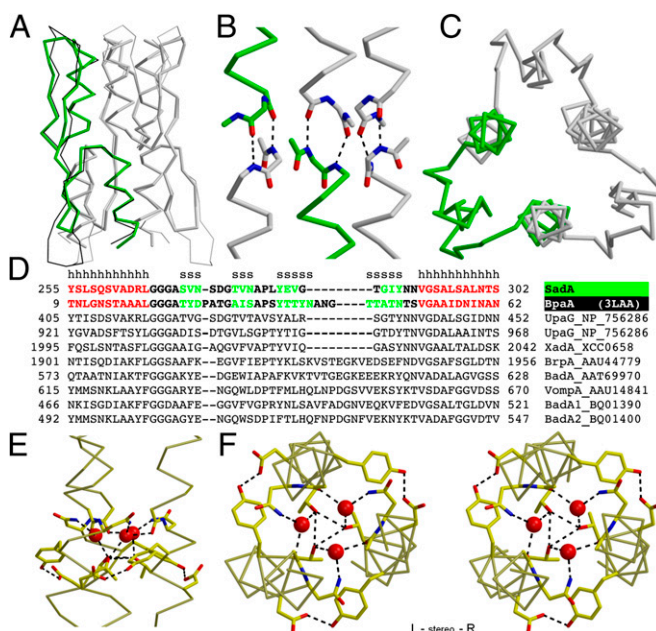
DALL domains can be viewed as upside-down neck domains. The transition from the coiled-coil to the  $\beta$ -stranded structure of the domain is mediated by a  $\beta$ -layer in the same way as in the neck, just that the transition leads from an  $\alpha$ -helical to a  $\beta$ -stranded structure; the central  $\beta$ -layer residue is the first leucine of the DALL consensus motif. Moreover,  $\beta$ -layers are not limited to transitions from alpha to beta and beta to alpha, and not to TAAs, underscoring the universality of  $\beta$ -layers: In the structure 2BA2 of a DUF16 domain from *Mycoplasma pneumoniae*, a  $\beta$ -layer is found that directly connects two trimeric coiled-coil segments, thus forming a transition from alpha to alpha. In a superposition of  $\beta$ -layers of all three types (from a DALL domain, a neck domain, and the one from 2BA2), both the central  $\beta$ -layer residues and the central water molecule match perfectly (Fig. 5).

**HANS domain.** Three transitions from  $\alpha$ -helical to  $\beta$ -stranded domains in SadA are mediated without a  $\beta$ -layer by a structurally simpler connector, the HANS domain. It precedes the last three heads of the enterobacterial TAAs. Although our construct K9 comprises the last two HANS domains of SadA, in both crystal forms, K9cfl and K9cflII, only the last HANS is fully ordered. This last HANS is also captured in the structure of K14, and in all three structures it assumes different conformations.

In the HANS connector, each helix of the coiled coil passes seamlessly into a  $\beta$ -sheet via a divergent  $\beta$ -turn (Fig. 6). This turn is formed by the residue preceding the KYFHANS consensus sequence and the phenylalanine therein. Starting with the tyrosine of the motif, the sheet passes into the inner  $\beta$ -sheet of the neighboring chain of the downstream Yhead domain.

The conformational flexibility of the most C-terminal HANS, observed when comparing the three crystal structures, resides in the coiled-coil before the KYFHANS motif: Approximately 10 residues before reaching the  $\beta$ -turn, the coiled coil breaks into a second, distorted segment with a steeper crossing angle of the helices. This distorted segment assumes different conformations in all three structures and entirely accounts for the different HANS conformations (Fig. 6). Most striking is the difference between the two structures of K9: When superimposed on the upstream coiled coil, the downstream Yhead is rotated by  $25^\circ$  around the trimer axis between the two HANS conformations (Fig. 6E and F). Additionally, although the transition at this HANS is straight or almost straight in K9cfl or K14, it is kinked by  $\sim 6^\circ$  in K9cflII. This conformational freedom is further underlined by the N-terminal HANS domain in the two structures of the K9 construct. In both, the whole N-terminal coiled coil down to the KYFHANS consensus motif is poorly ordered. Because faint overall electron density for the N-terminal coiled-coil is visible, albeit not interpretable, we can conclude that it is in fact present and folded.

**Special Coiled-Coil Motifs of the Stalk.** SadA contains a single copy of a FG domain after the N-terminal head, which we covered in the structures of K1 and K12. This domain is a stalk variant that connects two coiled-coil segments (Fig. 7A–D). At the FG sequence motif, the individual chains break the N-terminal coiled coil to form a  $\beta$ -hairpin that packs against the N-terminal coiled-coil segment, continue into a second  $\beta$ -hairpin that packs against the C-terminal coiled-coil segment, and finally return to where they departed from the coiled coil to fold into the C-terminal coiled-coil segment. In doing so, the three chains undergo a perfect  $120^\circ$  rotation around the trimer axis so that the helices of the N-terminal segment appear to pass seamlessly into the helices of the neighboring chain of the C-terminal segment: At the junction of the N- and C-terminal helices, the backbone hydrogen bonding pattern of  $\alpha$ -helices is continued between the N-terminal helix of one and the C-terminal helix of the other chain (Fig. 7B). In a superposition with the only other available FG structure in 3LAA (ref. 17; Fig. 7A), the structures only deviate in the loops of the  $\beta$ -hairpins where they differ in length. By forming a tight collar around them, the FG domain presumably provides rigidity to the adjacent coiled-coil segments. For instance, the  $\beta$ -hairpins of the FG domain in 3LAA fold



**Fig. 7.** FGG domain and NxYTD motif. Side (A) and top (C) view of the FGG domain. The  $\beta$ -hairpins inserted between the coiled-coil segments fold around the bundle like a collar and cause a  $120^\circ$  rotation of the helical bundle. (B) Directly at the junction of the two coiled-coil segments, the N-terminal helices appear to pass seamlessly into the C-terminal helices. (D) Sequence alignment of FGG domains. (E and F) Side and top view of the NxYTD motif in SadAK14, highlighting hydrogen bonding interactions and the invariant water network. Before reaching the anchor, many TAAs contain a right-handed coiled-coil segment in the stalk that contains a YTD sequence motif at the transition to the final left-handed segment; as described recently for YadA (20), this motif forms an elaborate network of hydrogen bonds: The threonines of the three chains form hydrogen bonds with each other in the core via their hydroxyl groups, whereas the tyrosines and aspartates form interchain hydrogen bonds around the bundle. Careful analysis however shows that the extent of the network is even larger. In more than half of all occurrences, YTD is preceded by an asparagine two residues before the tyrosine, yielding the motif NxYTD. These asparagines, together with the backbone oxygen of the residue preceding them and the threonines, coordinate three water molecules in the core of the bundle. Inspection of the YadA structures and their experimental data reveals that the three water molecules are also present in the two NxYTD motifs in YadA, albeit not modeled consistently.

around the whole  $\alpha$ -helical half of an adjacent HANS domain, so that the latter is perfectly well ordered and forms a straight transition. It may therefore play a role in the fine tuning of the flexibility of the fiber.

The other two stalk variants found in SadA are polar core motifs within the coiled coil. One is the NxYTD motif, with one occurrence in SadA, which is described in detail in Fig. 7. The other motif, N@d, is found throughout the whole SadA fiber. As we reported (19), this motif sequesters ions to the core of the coiled coil and has functional implications for the auto-transport process.

**Implications for the Adhesion Process.** In the light of the adhesion process, the DALL and HANS connector domains are of special functional importance: Both confer well-defined degrees of flexibility to the fiber. DALL domains can act as hinges that allow local bending of the fiber, whereas the hinge point is localized to its  $\beta$ -layer; the maximum bending angle we observed in crystalline state was  $10^\circ$  for a single DALL domain. Because some complex adhesins often comprise a large number of DALL domains, the total overall possible bending angle can sum up considerably. The HANS domain, also allowing for local bending, additionally provides rotational freedom around the trimer axis by partial unwinding of its helices. We observed rotation angles of up to  $25^\circ$  between different crystal structures of the same HANS domain—if this angle was the maximum deflection achievable by one HANS domain, only the three HANS domains in SadA would already allow for overall rotations of  $75^\circ$  around the trimer axis. Both these rotational degrees of freedom, perpendicular to and around the trimer axis, are conceivably facilitating the adjustment of the relative and overall orientation of the substrate binding domains of the fiber. For further fine tuning, the flexibility conferred by HANS may be delimited by FGG domains. The importance of flexibility for function has been pinpointed in detail for different TAAs, including the *Moraxella catarrhalis* adhesin UspA1 (26), *Haemophilus influenzae* Hia (21), and *E. coli* EibD (11). All of these adhesins bind to large receptors or matrix proteins on the surface of host cells, which disallows a rigid, straight fiber.

## Conclusions

In our quest to obtain the complete molecular structures of the chromosomally encoded TAAs of *Enterobacteriaceae*, we have characterized various hitherto undescribed domains and elements. We describe a universal structural motif for the transition from  $\alpha$ -helical to  $\beta$ -stranded structures that we name  $\beta$ -layer. This recurrent motif facilitates the domain shuffling that is the hallmark of these highly repetitive, fibrous proteins, but also occurs outside of the TAA family. In addition, we describe the structure and dynamics of connector domains with well-defined intrinsic rotational degrees of freedom that confer structural flexibility to the fiber that is highly relevant for the adhesion process. This work completes the list of structural templates for frequently occurring TAA domains and proves both the robustness and the applicability of the dictionary approach to understanding TAA structure. Despite low sequence identity, domains of the same type are structurally nearly identical—crystal structures of new fragments could be solved by molecular replacement by using search models of homologous domains. With the structures reported here, we were able to reconstruct the full fibers of SadA, UpaG, and EhaG; likewise, we can now reconstruct other TAA fibers either completely or to a large extent, including many important pathogenicity factors, using standard molecular modeling techniques.

## Materials and Methods

The protein used for structure determination is SadA of *Salmonella enterica* subsp. *enterica* serovar *Typhimurium* strain LT2 (GenBank accession no. NP\_462591). The other two proteins modeled are UpaG (NP\_756286) and EhaG (NP\_290185). Methodological details about sequence analysis, cloning, expression, protein and antibody purification, electron microscopy, crystallization, crystal structure solution, and molecular modeling are described in *SI Materials and Methods*. Crystallization conditions are listed in *Table S1*. Data collection and refinement statistics are summarized in *Table S2*.

**ACKNOWLEDGMENTS.** We thank Ines Wanke, Reinhard Albrecht, and Kerstin Baer for setting up the crystallization experiments and the staff of beamline PXII/Swiss Light Source for excellent technical support. This work was supported by institutional funds from the Max Planck Society and by German Science Foundation Grants FOR449/LU1165 and SFB766/B4.

- Gerlach RG, Hensel M (2007) Protein secretion systems and adhesins: The molecular armory of Gram-negative pathogens. *Int J Med Microbiol* 297(6):401–415.
- Hoiczky 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(22):5989–5999.
- Linke D, Riess T, Autenrieth IB, Lupas A, Kempf VA (2006) Trimeric autotransporter adhesins: Variable structure, common function. *Trends Microbiol* 14(6):264–270.
- Leo JC, Grin I, Linke D (2012) Type V secretion: Mechanism(s) of autotransport through the bacterial outer membrane. *Philos Trans R Soc Lond B Biol Sci* 367(1592):1088–1101.
- 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(2):506–512.
- Comanducci M, et al. (2002) NadA, a novel vaccine candidate of Neisseria meningitidis. *J Exp Med* 195(11):1445–1454.
- Riess T, et al. (2004) Bartonella adhesin A mediates a proangiogenic host cell response. *J Exp Med* 200(10):1267–1278.
- Lafontaine ER, et al. (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(5):1364–1373.
- St Geme JW, 3rd, 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(21):6005–6013.
- Biedzka-Sarek M, et al. (2008) Functional mapping of YadA- and Ail-mediated binding of human factor H to Yersinia enterocolitica serotype O:3. *Infect Immun* 76(11):5016–5027.
- Leo JC, et al. (2011) The structure of E. coli IgG-binding protein D suggests a general model for bending and binding in trimeric autotransporter adhesins. *Structure* 19(7):1021–1030.
- Szczesny P, Lupas A (2008) Domain annotation of trimeric autotransporter adhesins—daTAA. *Bioinformatics* 24(10):1251–1256.
- Szczesny P, et al. (2008) Structure of the head of the Bartonella adhesin BadA. *PLoS Pathog* 4(8):e1000119.
- Meng G, Surana NK, St Geme JW, 3rd, Waksman G (2006) Structure of the outer membrane translocator domain of the Haemophilus influenzae Hia trimeric autotransporter. *EMBO J* 25(11):2297–2304.
- Nummelin H, et al. (2004) The Yersinia adhesin YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. *EMBO J* 23(4):701–711.
- Yeo HJ, et al. (2004) Structural basis for host recognition by the Haemophilus influenzae Hia autotransporter. *EMBO J* 23(6):1245–1256.
- Edwards TE, et al. (2010) Structure of a Burkholderia pseudomallei trimeric autotransporter adhesin head. *PLoS ONE* 5(9):e12803.
- Conners R, et al. (2008) The Moraxella adhesin UspA1 binds to its human CEACAM1 receptor by a deformable trimeric coiled-coil. *EMBO J* 27(12):1779–1789.
- Hartmann MD, et al. (2009) A coiled-coil motif that sequesters ions to the hydrophobic core. *Proc Natl Acad Sci USA* 106(40):16950–16955.
- Alvarez BH, et al. (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(2):236–245.
- Meng G, St Geme JW, 3rd, Waksman G (2008) Repetitive architecture of the Haemophilus influenzae Hia trimeric autotransporter. *J Mol Biol* 384(4):824–836.
- Hernandez Alvarez B, et al. (2008) A new expression system for protein crystallization using trimeric coiled-coil adaptors. *Protein Eng Des Sel* 21(1):11–18.
- Valle J, et al. (2008) UpaG, a new member of the trimeric autotransporter family of adhesins in uropathogenic Escherichia coli. *J Bacteriol* 190(12):4147–4161.
- Totsika M, et al. (2012) Molecular characterization of the EhaG and UpaG trimeric autotransporter proteins from pathogenic Escherichia coli. *Appl Environ Microbiol* 78(7):2179–2189.
- Raghunathan D, et al. (2011) SadA, a trimeric autotransporter from Salmonella enterica serovar Typhimurium, can promote biofilm formation and provides limited protection against infection. *Infect Immun* 79(11):4342–4352.
- Agnew C, et al. (2011) Correlation of in situ mechanosensitive responses of the Moraxella catarrhalis adhesin UspA1 with fibronectin and receptor CEACAM1 binding. *Proc Natl Acad Sci USA* 108(37):15174–15178.