Supporting Information

Hartmann et al. 10.1073/pnas.1211872110

SI Materials and Methods

Sequence Analysis. The protein used in this study is SadA of *Salmonella enterica* subsp. *enterica serovar Typhimurium* strain *LT2* (GenBank accession no. NP_462591). To determine its genomic localization in enterobacteria, homologs were detected by BLAST against the nonreduntant database at National Center for Biotechnology Information. For selected hits, the genomic context was extracted and visualized by using GCView (1). The domain composition of the trimeric autotransporters from different *Enterobacteria* was evaluated by using daTAA (2).

Cloning. The full-length *SadA* gene was amplified by PCR from genomic DNA of *S. enterica* subsp. *enterica serovar Typhimurium* strain *LT2* by using primers SP1 (5′-GGAACCTTTCTAGATAACGAGGGCAAAAAATGAATAGAATATTTAAAGTCCTCTGGAATGCC) and SP2 (5′-CCAAGGTTAAGCTTATTACCACTGGAA GCCCGCGCC). The obtained 4.4-kbp fragment was digested with HindIII and XbaI and cloned in pASK-IBA2 (IBA BioTAGnology). The resulting clone pSadA served as a template for amplification of shorter SadA constructs by PCR using primers SK5fw (5′-GACCATGGTCTCCGATTTATGAAACCAACCAGAAGGTGGATC) and SK5rev (5′-GACCATGGTCTCCTCATTCAGCCGTTACCCGTTGCGTATGCATC) for SadAK5, SK9fw (5′-GACCATGGTCTCCGATTCA AAATGCCATTGGTGCGG TCAC) and SK9rev (5′-GACCATGGTCTCCTCATTTGCGCCACATTAACCGCGTC AGTG) for SadAK9, SK12fw (5′-GACCATGGTCTCCGATTTATTCTTTAAGTCAAT CCGTCGCCGACCGACTCGGCGG) and SK12rev (5′-GACCATGGTCTCCT CATCTGAGAGCCGTTAACGGCATCGGTGCTGTCCGCAGCCAGG) for SadAK12, and SK14fw (5′-GACCATGGTCTCCGATTAAAGTAACGGACGCGCAGGTTTCC) and SK14rev (5′-GACCATGGTCTCCTCATCTTGTTTTCTACGCCTTTGATTTTGC) for SadAK14. The yielded fragments of expected size were digested with BsaI and ligated into vector pIBA-GCN4tri-His (3). Correctness of the clones was verified by DNA sequencing.

Protein Expression and Purification. The purification of SadAK1 and SadAK3 was described (3). SadAK5, which contains amino acids 823–947 of SadA, was expressed as a fusion with only an N-terminal GCN4pII-adaptor. This was achieved by insertion of a stop codon in primer SK5rev immediately before the C-terminal GCN4pII adaptor and the following $(His)_6$ linker. SadAK5 was overexpressed in *Escherichia coli* TOP10 as described above. Pelleted cells were ruptured by using a French press in 20 mM Tris·HCl, 40 mM NaCl, 4 mM $MgCl_2$ at pH 7.4 containing a protease inhibitor mix (Roche), PMSF, and Dnase I. After a centrifugation step (140,000 × g, 40 min at 4 °C), the supernatant was diluted 1:5 with 0.5 M Mes at pH 5.5 and applied onto a cation exchange column (SP Sepharose; GE Healthcare) equilibrated with 20 mM Mes and 40 mM NaCl at pH 5.5. SadAK5 bound to the column and was eluted with a gradient of 0–1 M KCl. The final step was a size exclusion chromatography step (S75; GE Healthcare) in 20 mM Mops and 150 mM NaCl at pH 7.2. The other SadA constructs comprise SadA residues 1049–1304 in SadAK9, 255–358 in SadAK12, and 1185–1386 in SadAK14, each fused to GCN4pII adaptors on both the N- and C terminus. The C-terminal adaptor is followed by a $(His)_6$ linker to simplify purification. The proteins were expressed and purified under denaturing conditions by using a NiNTA column (GE Healthcare) as described for His-tagged SadAK3 (3). Refolding was performed at 4 °C by dialysis by using the following buffers: 20 mM Mops, 450 mM NaCl, and 10% glycerol at pH 7.2 for SadAK9; 20 mM Mops and 150 mM NaCl at pH 7.2 for SadAK12, and 20 mM Tris·HCl and 150 mM NaCl at pH 7.4 for SadAK14. For SadAK9, an additional size exclusion chromatography step on a Superdex 200 column (GE Healthcare) equilibrated with the appropriate refolding buffer was necessary to obtain pure protein.

Antibody Purification. For antibody production, full-length SadA was expressed in *E. coli* BL21 Omp8 DE3 cells (4) at 25 °C in LB medium overnight, supplemented with 100 μg/mL Ampicillin. Outer membrane fractions were isolated essentially as described, using the method of differential membrane solubilization with Sarkosyl (5). The outer membranes were then solubilized in 3% Octylpolyoxyethylene (C_8POE), 50 mM EDTA, 150 mM NaCl, and 20 mM Tris·HCl at pH 8.0 and were diluted after 2 h to a final detergent concentration of 1%. The solution was cleared of debris by centrifugation and was subjected to phase separation by using ice cold 20% $(NH_4)_2SO_4$ (6). The detergent-rich phase was collected, dialyzed against 1% C8POE and 20 mM Tris·HCl at pH 8.0, and subjected to anion exchange chromatography with a MonoQ column (GE Healthcare) by using a 0–1 M NaCl gradient to remove residual lipopolysaccharides. Finally, the fractions containing SadA were pooled, precipitated with 90% ice-cold Acetone, and resuspended as a slurry in PBS to a concentration of approximately 0.5 mg/mL The purified protein was used to raise rabbit anti-SadA antibodies. After testing the specificities of several bleeds against SadA, the polyclonal antiserum was purified by using SadAK9 as bait. SadAK9 was coupled to a 1-mL HiTrap NHS-activated HP column (GE Healthcare) according to the manufacturer instruction. One milliliter of anti-SadA antiserum was diluted 1:20 with 50 mM Mops and 150 mM NaCl at pH 7.2 and loaded on the column equilibrated with the same buffer. Bound antibodies were eluted with a gradient of 0–4 M $MgCl_2$ and tested for specificity in Western blotting.

Electron Microscopy. *E. coli* Top10 cells carrying either an empty pASK-IBA3 or pASK-IBA2-SadA were grown at 37 °C until OD_{600} = 0.6–0.8. Overproduction was induced by the addition of anhydrotetracycline, and cells were grown for another 4 h, spun down gently to concentrate cells, and immobilized on poly-L-Lysine–coated coverslips.

For immunolabeling, immobilized cells were blocked with 0.5% (wt/vol) BSA and 0.2% (wt/vol) gelatin in PBS and labeled with affinity-purified SadA specific rabbit IgG. Bound antibodies were detected with Nanogold coupled to goat IgG anti-rabbit IgG (no. 2003; Nanoprobes). After washing (2× blocking buffer, 3× PBS; 5 min each) the antigen–antibody–marker complexes were stabilized with 0.5% GA in PBS for 5 min and washed six times with H_2O (total 25 min). Gold markers were enhanced for 35 min with silver lactate, hydrochinone, and gum arabic in citrate buffer at pH 3.8 according to Danscher (7).

For scanning electron microscopy, colonies were postfixed with 1% osmium tetroxide in PBS for 1 h on ice, dehydrated in ethanol and critical-point-dried from CO_2. The samples were sputter-coated with 7 nm gold-palladium and examined at 20 kV accelerating voltage in a Hitachi S-800 field emission scanning electron microscope.

X-Ray Crystallography. Crystallization trials were set up by using the Honey bee 961 robot from Genomic Solutions, mixing 400 nL

of protein solution with 400 nL of reservoir solution in 96-well sitting-drop Corning 3550 plates. The reservoir volume was 75 μL. The crystallization conditions for all crystals used in the diffraction experiments are listed in Table S1, together with soaking conditions for cryoprotection where applicable. All crystals were loop mounted, flash frozen in liquid nitrogen, and all data were collected at beamline X10SA (PXII) at the SLS (Paul Scherrer Institute, Villigen, Switzerland) under cryo conditions at 100 K by using a mar225 CCD detector (Marresearch). Diffraction images were processed and scaled by using the *XDS* program suite (8). All structures were solved by molecular replacement using MOLREP (9), in a hierarchical fashion: The shortest construct SadAK1, comprising an FGG domain, was solved as described for SadAK3 (10), using the trimeric GCN4 structure 1GCM as a search model. For K12, comprising an FGG domain and a DALL/Neck tandem, was solved by using the K1 structure as a search model, the resulting structure of the DALL/Neck tandem from K12 was subsequently used to solve K5. K14 was solved by using K5, and K9 by using K14.

In both crystal forms obtained for K9, K9cfI, and K9cfII, the N-terminal 41 residues that should form a continuous coiled coil are not traceable, albeit overall electron density resembling the expected shape of the coiled coil is visible. As described in the main text, we explain this phenomenon with the intrinsic flexibility of the HANS connector following the coiled coil, giving rise to the elevated R factors of the two structures. In K5, the C-terminal 43 residues are not traceable and supposedly unstructured—the C terminus ending directly after a HANS domain was chosen unfortunate, which we learned after obtaining the first structure of a HANS domain in K9.

K12 shows an interesting crystallographic peculiarity: In the crystal packing in $P6_322$, cavities of 32-symmetry are formed along the c-direction, which can accommodate full K12 trimers. In the later stages of refinement, weak electron density characteristic for the rod-shaped K12 trimer emerged in these cavities. This density can be explained by an average of two K12 trimers in opposite orientations, related by the twofold axes of the crystal system. Refinement of an additional protein chain in these cavities was tried in space group $P6_322$ with an occupancy of 50% and also after expansion to lower symmetry, including P1. In all cases, the refinement resulted in only faint electron density in the resulting $2F_O$-F_C maps for this chain, which was always, also in lower symmetry, interpreted best as an average over both orientations with low occupancy. Because none of these trials led to improved R factors (but to more complicated models), we decided to refine and deposit the final model in $P6_322$ without interpreting the density in the cavities. We assume that, throughout the crystal, approximately 50% of these cavities are stochastically filled with trimers, randomly in one of the two orientations, which might be of importance for crystal integrity.

For K1, K5, K12, and K14, ARP/WARP (11) was used for automated rebuilding. All structures were completed in cyclic manual modeling with Coot (12) and refinement with REFMAC5 (13). For K9cfII and K14, the refinement was carried out by using NCS restraints. In all structures, the nature of structural solvent molecules, such as the central water molecules in β-layers and chloride ions in N@*d* layers, was identified on the basis of their coordination distances to the protein, refined B-factors, and fit to the electron density maps. Analysis with Procheck (14) showed good geometries for all structures. Data collection and refinement statistics are summarized in Table S2 together with PDB accession codes. All molecular depictions were prepared by using MolScript (15) and Raster3D (16).

Modeling of Full-Length Fibers. The models of full-length SadA, UpaG, and EhaG proteins were computed in Modeler by using the structure of the YadA head (PDB ID code 1P9H), the reported SadA structures, and the structure of the Hia membrane anchor domain (PDB ID code 2GR8) as templates. The modeling process was conducted with symmetry constraints.

1. Grin I, Linke D (2011) GCView: The genomic context viewer for protein homology searches. *Nucleic Acids Res* 39(Web Server issue):W353–W356.
2. Szczesny P, Lupas A (2008) Domain annotation of trimeric autotransporter adhesins—daTAA. *Bioinformatics* 24(10):1251–1256.
3. 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.
4. Prilipov A, Phale PS, Van Gelder P, Rosenbusch JP, Koebnik R (1998) Coupling site-directed mutagenesis with high-level expression: large scale production of mutant porins from E. coli. *FEMS Microbiol Lett* 163(1):65–72.
5. Arnold T, Linke D (2008) The use of detergents to purify membrane proteins. *Curr Protoc Protein Sci*, Chapter 4:Unit 4.8.1–4.8.30.
6. Arnold T, Linke D (2007) Phase separation in the isolation and purification of membrane proteins. *Biotechniques* 43(4):427–430, 432, 434 passim.
7. Danscher G (1981) Histochemical demonstration of heavy metals. A revised version of the sulphide silver method suitable for both light and electronmicroscopy. *Histochemistry* 71(1):1–16.
8. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* 26:795–800.
9. Vagin A, Teplyakov A (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr* 56(Pt 12):1622–1624.
10. 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.
11. Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6(5):458–463.
12. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
13. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr D Biol Crystallogr* 55(Pt 1):247–255.
14. Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) Procheck - A program to check the stereochemical quality of protein structures. *J Appl Cryst* 26:283–291.
15. Kraulis PJ (1991) Molscript - A program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* 24:946–950.
16. Merritt EA, Bacon DJ (1997) Raster3D: Photorealistic molecular graphics. *Methods Enzymol* 277:505–524.

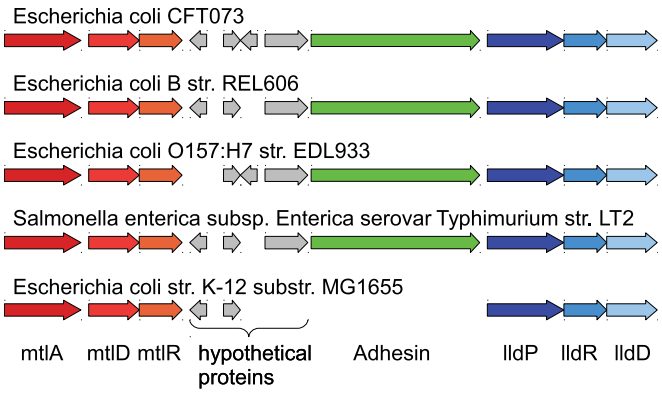

Fig. S1. Genomic context of the enterobacterial adhesins. Genomic region of *E. coli* and *S. typhimurium* strains containing SadA or its homologs (UpaG in *E. coli* CFT 073, EhaG in *E. coli* O157:H7). The adhesin is located between the *mtl*-operon (mannitol metabolism) and the *lld*-operon (L-lactate metabolism). This genomic structure is conserved in all strains of *E. coli* and *Salmonella spp* except for *E. coli* K-12 and derivatives that appear to have a 5-kb deletion at this location.

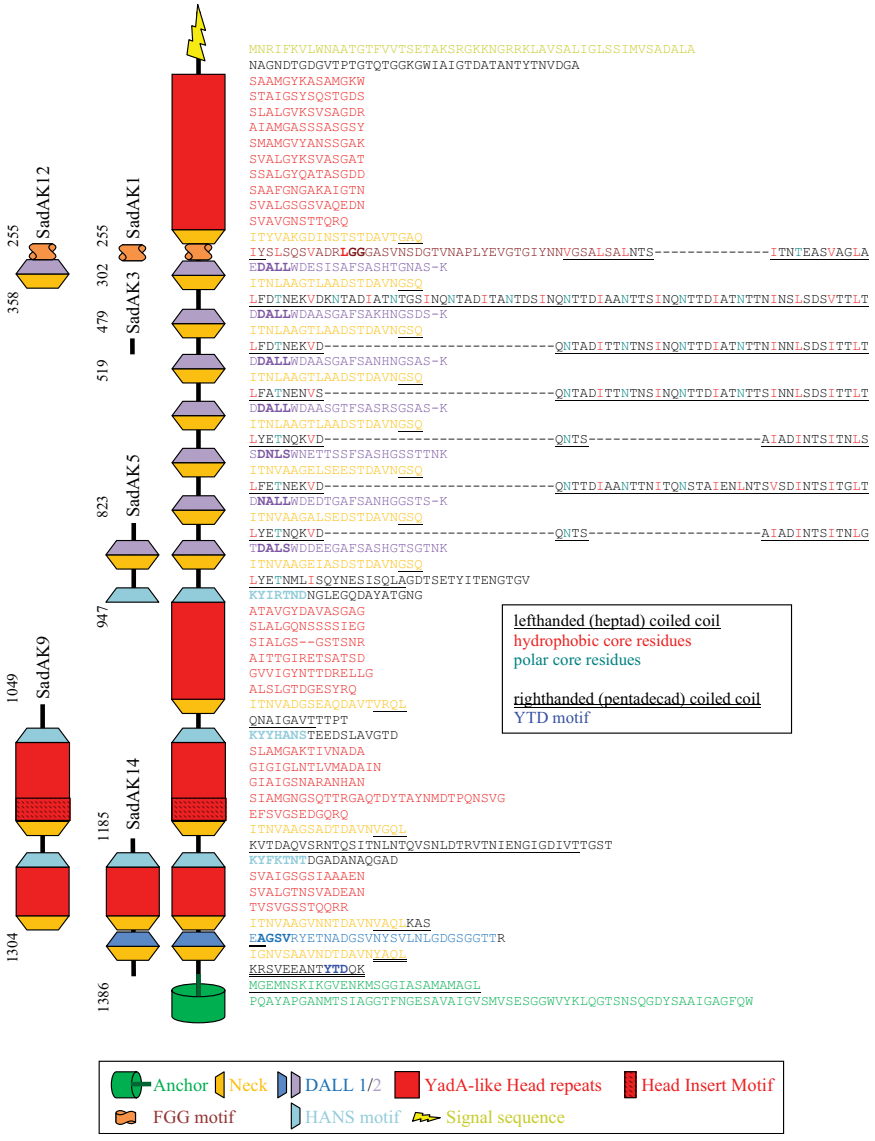

Fig. S2. Sequence of SadA and constructs used in this study. Domains and constructs are marked individually and colored as indicated.

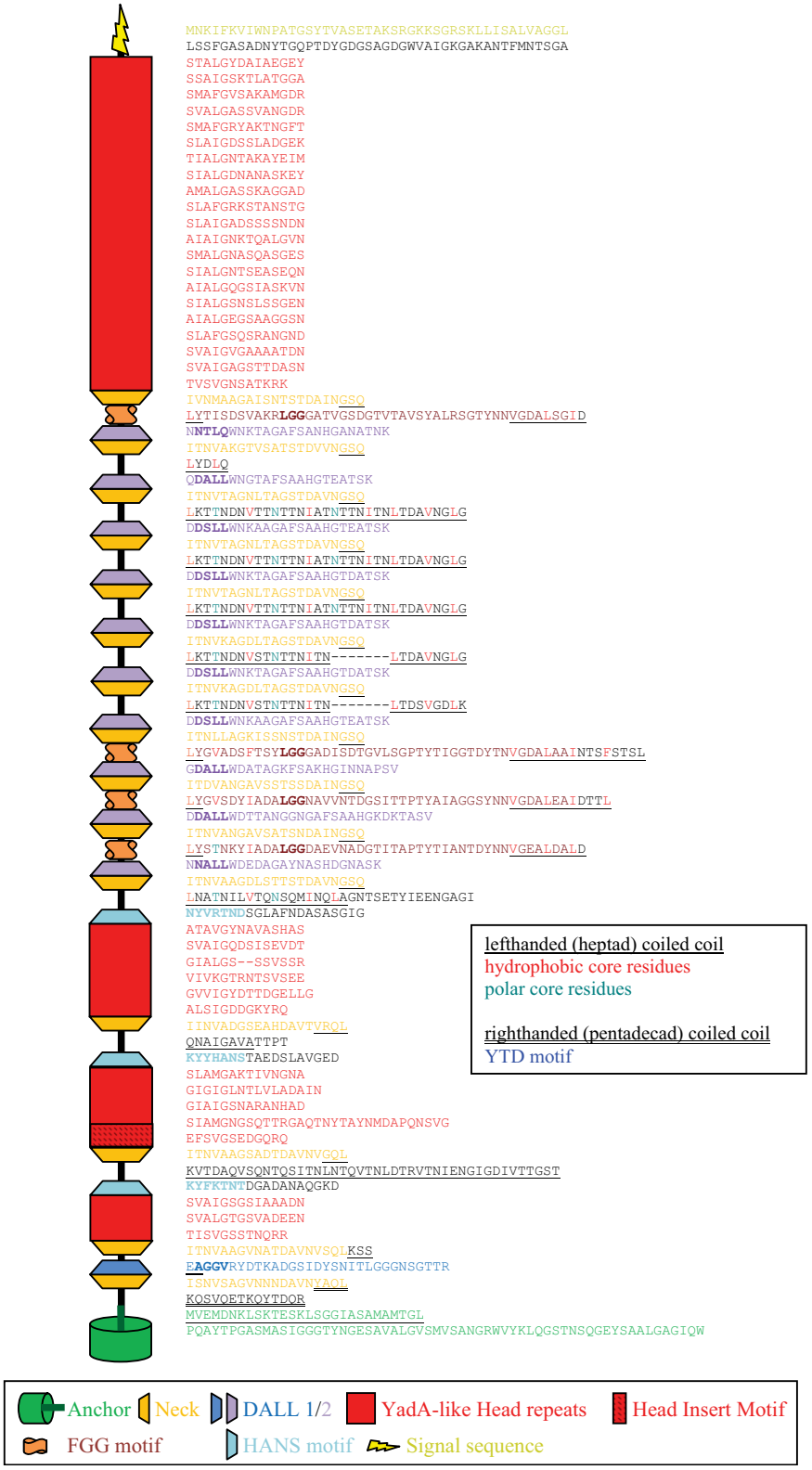

Fig. S3. Sequence of UpaG and symbolic domain arrangement. Domains are marked individually and colored as indicated.

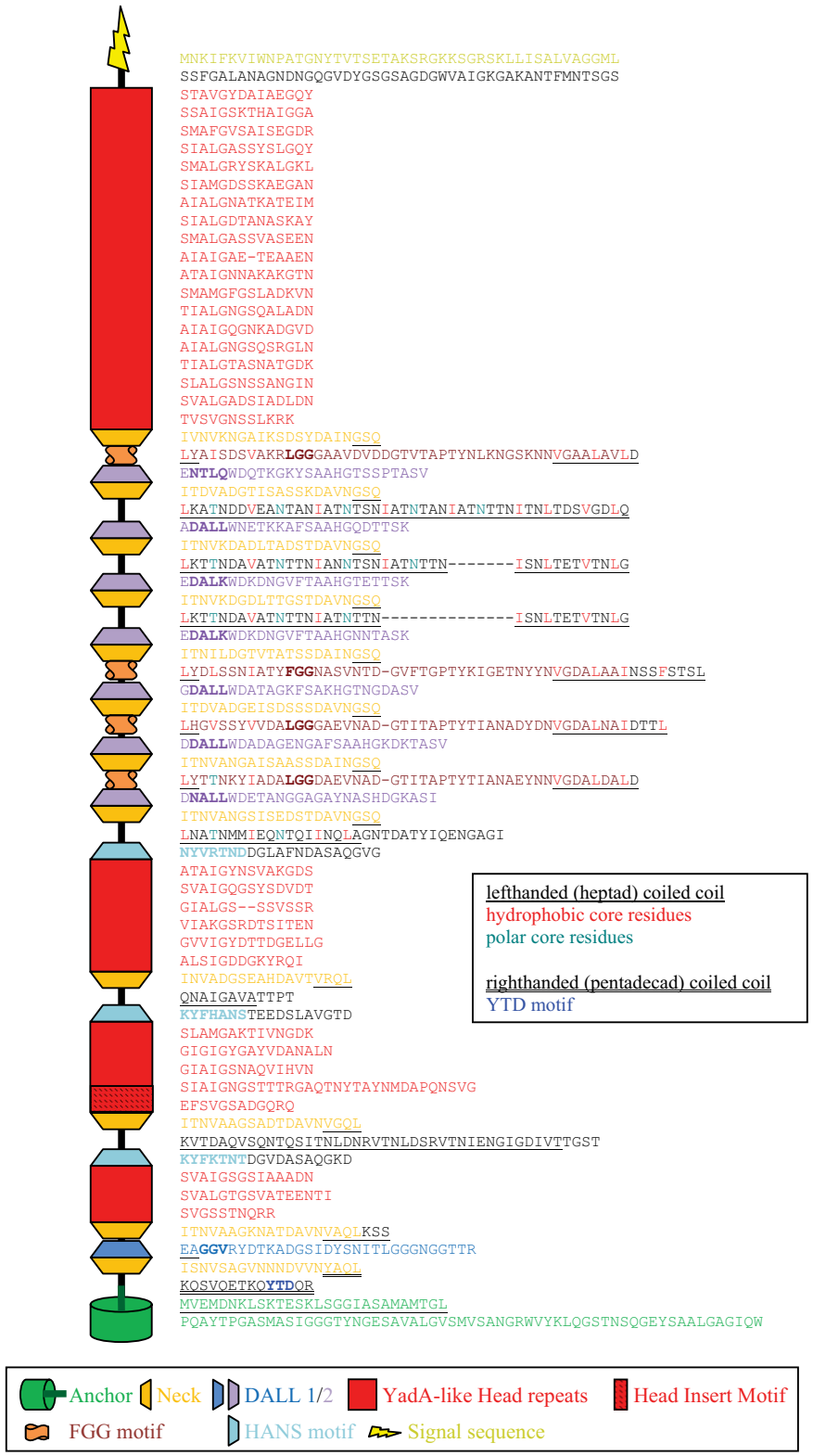

Fig. S4. Sequence of EhaG and symbolic domain arrangement. Domains are marked individually and colored as indicated.

Table S1. Crystallization conditions and cryo protection

Structure	Protein solution	Concentration, mg/mL	Reservoir solution (RS)	Cryo solution
K1	20 mM Mops at pH 7.2, 150 mM NaCl	4	16% (wt/vol) PEG 4000, 80 mM sodium acetate, 100 mM Hepes at pH 7.5	RS+15% (vol/vol) PEG 400
K5	20 mM Mops at pH 7.2, 150 mM NaCl	17	10% (wt/vol) PEG 10000, 200 mM magnesium nitrate	RS+20% (vol/vol) PEG 400
K9cfI	20 mM Tris at pH 7.4, 150 mM NaCl, 10% (vol/vol) glycerol	8.5	20% (vol/vol) butanediol, 100 mM sodium acetate at pH 4.5	—
K9cfII	20 mM Mops at pH 7.2, 150 mM NaCl, 10% (vol/vol) glycerol	4	15% (vol/vol) butanediol, 100 mM sodium acetate at pH 4.2	—
K12	20 mM Mops at pH 7.2, 150 mM NaCl	10	500 mM ammonium tartrate, 100 mM sodium acetate at pH 5.0	RS+25% (vol/vol) EG
K14	20 mM Tris at pH 7.4, 150 mM NaCl	12	12% (wt/vol) PEG 8000, 100 mM magnesium acetate, 100 mM Tris at pH 8.5	RS+10% (vol/vol) PEG 400

Crystallization conditions for SadAK3 have been published (10).

Table S2. Data collection and refinement statistics

Structure	SadAK1	SadAK5	SadAK9cfI	SadAK9cfII	SadAK12	SadAK14	SadAK3
PDB ID code	2YNY	2YNZ	2YO0	2YO1	2YO2	2YO3	2WPQ
Monomers/AU	3	3	1	3	1	3	3
Space group	$P2_1$	$P2_12_12_1$	P 6_3 (*)	$P2_1$	P 6_322	C2	$P2_1$
a, Å	34.8	44.6	54.0	81.9	48.6	189.3	26.0
b, Å	40.4	60.3	54.0	48.7	48.6	46.2	37.0
c, Å	98.6	135.6	306.7	135.6	366.0	103.8	178.4
β, °	93.5	90	90	105.1	90	98.0	92.7
Resolution range, Å	38.0–1.35 (1.43–1.35)	19.9–1.40 (1.49–1.40)	37.2–2.80 (2.97–2.80)	39.1–3.10 (3.28–3.10)	38.3–2.00 (2.12–2.00)	37.3–2.00 (2.12–2.00)	34.1–1.85 (1.96–1.85)
Completeness, %	96.0 (89.6)	98.8 (97.9)	99.5 (99.4)	96.9 (94.6)	99.5 (98.6)	98.3 (96.8)	98.3 (94.0)
Redundancy	4.30 (3.22)	4.38 (4.14)	3.25 (3.22)	2.31 (2.35)	7.30 (6.01)	3.91 (3.60)	3.97 (3.82)
I/σ(I)	13.9 (1.98)	12.3 (1.97)	8.59 (2.07)	8.93 (2.35)	13.8 (2.29)	12.4 (2.32)	10.1 (1.91)
R_{merge}, %	5.5 (62.8)	6.9 (67.9)	14.0 (59.8)	10.0 (42.3)	8.7 (63.7)	7.1 (61.0)	7.7 (65.0)
R_{cryst}/R_{free}, %	13.3/19.8	17.6/23.2	22.1/29.7	26.5/32.0	24.2/28.4	20.0/25.3	22.1/28.8
Bond length/angle rmsd, Å/°	0.021/1.68	0.011/1.10	0.010/1.14	0.014/1.32	0.015/1.29	0.011/1.10	0.007/0.84
Ramachandran plot statistics, %	98.9/1.1/0/0	97.7/2.3/0/0	72.6/26.2/1.2/0	77.3/20.7/1.0/1.1	97.3/2.7/0/0	93.6/6.1/0.3/0	100/0/0/0

Values in parentheses refer to the highest resolution shell. Statistics for SadAK3, which have been published (10), are shown for the sake of completeness. The ramachandran plot statistics show the percentage of residues in the most favored/additionally allowed/generously allowed/disallowed regions, respectively, as defined and determined by using the program Procheck (14).

*The crystals of SadAK9cfI were hemihedrally twinned with apparent 622 symmetry, twinning operator -H-K,K,-L and a twin fraction of 43%.