Structure of the YajR Transporter Suggests a Transport Mechanism Based on the Conserved Motif A

Daohua Jiang[1,2], Yan Zhao[1,3], Xianping Wang[1], Junping Fan[1], Jie Heng[1], Xuehui Liu[1], Wei Feng[1], Xusheng Kang[1], Bo Huang[1], Jianfeng Liu[2], Xuejun C. Zhang[1]

[1] National Laboratory of Macromolecules, National Center for Protein Science- Beijing, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing, China 100101 [2] School of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei, China 430074 [3] School of Life Sciences, University of Science and Technology of China, Hefei, Anhui, China 230027

Structural description

Overall structure

Within the TM core, the compact N-domain of YajR is formed by TMs 1–6 and possesses an internal pseudo two-fold symmetry (with the axis roughly parallel to the membrane plane) connecting TMs 1–3 to 4–6. Furthermore, the N-terminal domain is related to the six C-terminal TM helices (*i.e.* TMs 7–12) through another pseudo two-fold symmetry (with the axis perpendicular to the membrane plane). Thus, similar to known MFS structures, YajR consists of four symmetrical, three-helix repeats. The first helix in each repeat (*i.e.* TMs 1, 4, 7, and 10) contributes to formation of the cavity, and is thus referred to as the central helix. These central helices appear to play two major roles. Firstly, they determine substrate specificity and thus usually contain more polar residues facing the central cavity than other helices, allowing more specific interactions with a substrate than hydrophobic residues could offer. Compared with other MFS structures, TM4 in YajR losses two a.a. residues in the sequence region $G^{114}AIAA^{118}$. Of the known 3D structures of MFS transporters, YajR is currently the only one that contains a break in the middle of a TM helix. Such a broken TM helix provides more cavity space thus likely contributing to substrate specificity. Secondly, the central helices contribute to the interactions in both the inward and outward states, for example, by dynamically participating in distinct inter-domain helical bundles. The second helix in each repeat (*i.e.* TMs 2, 5, 8, and 11) is usually long and curved, assuming a banana-like shape. Together, they form the side walls of the central cavity and are referred to here as rocker helices to indicate their direct involvement in inter-domain conformational changes. The third group of helices (*i.e.* TMs 3, 6, 9, and 12) are located at both ends of the longest dimension of the TM core (Fig. 1B). They do not directly contribute to the formation of the cavity, however, they do contact the lipid bilayer directly and are referred to as supporting helices. The fact that they are more hydrophobic than the central and rocker helices and are in general less conserved among orthologs is in agreement with their assumed role (Fig. S1B). Together, these 12, symmetry related, TM helices form the structural basis of the rocker-switch mechanism of the YajR transporter.

In general, the helical type of integral membrane proteins contain more a.a. residues of short side chains (*e.g.* Gly, Ala, Ser, or Thr) in the helical region than soluble proteins. Such residues favor more compact helix-helix packing (1). Intuitively, such residues would allow smoother dynamic movement between TM helices should such movement be required by a function of the membrane protein. By searching for SxxSxxxS or SxxxSxxS patterns in the a.a. sequence (where *S* stands for short side chain a.a. residues and *x* stands for any residues), we identified twelve such sequence regions in the TM core domain of YajR (Table S2). Four of them are observed to be involved in the interface between the N- and C-domain in the outward conformation, and other two are potentially involved in the domain interface in the inward conformation. This observation further supports the notion that MFS proteins undertake rocker switch movement during substrate transporting.

Moreover, charge-dipole interactions are commonly used to stabilize soluble proteins (2). They are also frequently observed in the helix-type of integral membrane proteins. In our YajR crystal structure, all helices in the TM core domain have at least one charge-helix dipole interaction at either the N- or C-terminal (Table S5). Six of them are strictly or highly conserved among YajR from different species. For motif A, such an inter-domain, charge-dipole interaction is used to stabilize one of the two low energy states during the transporting process.

Central cavity

The path of substrate transport is believed to be through the central cavity between the N- and C- domains of the TM core (3, 4). In the outward conformation of YajR, the cavity is *ca.* 20 Å deep and 10 Å wide. Inside the cavity, the N-domain side is positively charged, whereas the C-domain side is negatively charged (Fig. S4). The open side of the cavity has two clefts between the N- and C-domains, one on

each side-wall. Most hydrophilic residues in the transmembrane region of the TM core point into the cavity. In particular, in the outward conformation, Arg24, Tyr57, Gln61, and Gln65 from the N-terminal domain face the cavity. Of these, Arg24 serves as a C-cap for the N-terminal portion of the broken TM4, and Gln65 serves as an N-cap for the C-terminal portion of TM4. From the C-terminal domain, His225, Asn317, Glu320, Tyr343, and Gln347 face the central cavity. In particular, His225 and Glu320 are candidates for protonation. While these residues are absolutely conserved in YajR and AraEP proteins from a variety of Gram-negative bacteria (Fig. S1B), most of them are not conserved in the homologous multi-drug efflux exporter EmrD (Fig. S1A), suggesting that YajR transports a different class of substrates from the hydrophobic substrates transported by EmrD. Importantly, all residues forming the surface of the central cavity, both hydrophobic and hydrophilic, are highly conserved among YajR and AraEP proteins, indicating that these transporters share conserved functions in their corresponding host organisms.

Characterization of the YAM domain

The 65-residue small domain at the C-terminal of YajR possesses a modified ferredoxin-like fold and has dimensions of 22 × 25 × 37 Å^3 (Fig. 1B). While ferredoxin-like folds usually have a βαββαβ secondary structural pattern, the last β-strand is missing in the small domain at the C-terminal of YajR. The remaining three β-strands form one antiparallel β-sheet with both α-helices located on one side of the sheet and the other side facing the TM core. This folding topology is similar to that of the metal-binding domain (MBD) of P1-type ATPase (PDB ID: 3DXS) (5), which is located at the N-terminal of the ATPase and is postulated to play a regulatory role. Thus, we termed the small domain at the C-terminal of YajR proteins as well as their homologous AraEP proteins as the YajR/AraEP/MBD or YAM domain. The modified ferredoxin-like fold of the YAM domain is predicted to be conserved in all YajR transporters (Fig. S1B).

To study the function of the YAM domain, we expressed and purified the YAM domain of *E. coli* YajR and measured its thermo-denaturation using thermofluor assays (Fig. S6). Our results showed that the YAM domain has a Tm above 75°C, and indicated that the thermal stability increased at increasing concentrations of halogen ions, but not Na^+/K^+ cations. In addition, the stability increased with increasing pH over an experimental range of 5.5–6.2. Therefore, the YAM domain of *E. coli* YajR exhibits high thermal stability which is independent of the TM core, and this stability may change in response to pH and halogen ion concentration.

In the outward conformation of YajR, the YAM domain is located outside of the TM core, and has only minor interactions with the TM core (with a total of *ca.* 840 Å^2 buried solvent-accessible surface from both the TM core and the YAM domain). In addition, pull-down experiments were performed with YajR peptides consisting of the isolated TM core (YajR-ΔYAM) and the YAM domain, and no interactions were observed between them. In a thermofluor assay, the recombinant protein of YajR-ΔYAM showed only slightly lower thermal stability than the full length protein (Fig. 1A). These results suggest that YAM domain may have a regulatory instead of a structural role.

Methods

Protein expression and purification

The *YajR* gene (GenBank ID: 251783932) was cloned from the genome of *E. coli* BL21(DE3) and was found to express well as a recombinant protein in high-throughput expression screening (6). The DNA encoding the full length YajR was subcloned into the pET-28a vector (Novagen, US) with an incleavable His_6 tag fused at the C-terminus, and the nucleotide sequence was confirmed using DNA sequencing. The plasmid was transformed into *E. coli* C43 (DE3) strain (7). Cells were grown at 37°C in TB medium until the cell density reached an OD_{600} of 1.3. The temperature was then reduced to 16°C before induction with 0.5 mM isopropyl β-d-thiogalactoside (IPTG). After growth at 16°C for 18 h, cells were harvested by using centrifugation at 4,200 *g*. Cells were resuspended in ice-cold buffer A (20 mM Tris-HCl (pH 8.0), 300 mM NaCl, and 10% (v/v) glycerol) and subjected to two runs of homogenization at 10,000–15,000 psi using an EmulsiFlex-C3 homogenizer (AVESTIN, Canada). The homogenate was centrifuged at 17,000 *g* for 15 min at 4°C, and then the supernatant was ultra-centrifuged at 100,000 *g* for 90 min. The membrane fraction was resuspended in buffer A supplemented with 1% (w/v) DDM (Anatrace, US) and was slowly stirred for 2 h at 4°C. After another ultracentrifugation at 100,000 *g* for 30 min, the supernatant was collected and loaded on 2 ml of Ni^{2+}-nitrilotriacetate affinity resin (Ni-NTA from Qiagen, Germany) pre-equilibrated with buffer A supplemented with 5 mM imidazole and 0.03% (w/v) DDM. After incubating for 1 h, the resin was washed with 50 ml buffer A supplemented with 20 mM imidazole, and 0.03% (w/v) DDM. The protein sample was eluted with 15 ml elution buffer containing buffer A, 200 mM imidazole, and 0.03% (w/v) DDM, and was concentrated to 500 µl. The concentrated protein sample was then loaded onto a Superdex-200 column (10/30, GE Healthcare, US) pre-equilibrated with 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, and a detergent (*e.g.* β-NG*)* at 2x CMC. Peak fractions were collected, and the pooled protein sample was concentrated to 20 mg/ml before carrying out crystallization trials. The yield was typically 6 mg protein per liter of cell culture.

To obtain a seleno-methionine derivative protein, the transformed *E. coli* C43 (DE3) strain was grown in minimal medium with seleno-L-methionine replacing methionine, and purified in the same manner as the native protein.

Crystallization

Initial crystallization trials were carried out at 16°C using commercial crystallization screening kits from Molecular Dimension (UK) and Hampton Research (US) and the hanging-drop vapor-diffusion method. Protein crystals

appeared in the presence of detergents DDM, decyl-β-d-maltopyranoside (DM), nonyl-β-d-maltopyranoside (NM), nonyl-β-d-thiomaltopyranoside (NTM), 6-cyclohexyl-1-hexyl-β-d-maltoside (Cymal-6), octaethylene glycol monododecyl ether ($C_{12}E_8$), and β-NG (Anatrace). While most crystals showed diffractions up to only 7–8 Å resolutions, crystals grown with β-NG diffracted significantly better on a synchrotron beamline. After optimization, hexagonal rod shaped crystals appeared in 1–2 d and reached full size (typically 200 x 80 x 80 μm^3) in 10 d with a reservoir solution of 0.1 M Na-acetate (pH 5.0–5.3), 25–29% (v/v) PEG300. Crystals were directly flash-cooled in liquid nitrogen for storage and data collection.

For heavy atom derivatives, native crystals were soaked in mother liquor containing 0.1 M Na-acetate (pH 5.1), 30% (v/v) PEG300, 1% (w/v) β-NG, and 2–10 mM $Hg(CN)_2$ overnight. Se-Met derivative crystals were grown under the same condition as the native crystals.

Data collection and structure determination

Diffraction data for the native YajR crystal were collected up to 3.15 Å resolution at the 17U beamline of the Shanghai Synchrotron Radiation Facility (SSRF). All anomalous data sets, including $Hg(CN)_2$-soaked derivatives and seleno-L-methionine derivatives, were collected at the 41XU beamline of SPring-8 facility (Japan). All data were processed with *HKL2000* (8).

Two mercury atoms in the Hg-derivative crystal of the wild type protein were located using *SheIx-C/D/E* (9) and initial phases were calculated using *Phenix.autosol* (10) with a two-wavelength MAD data set. A crude YajR model containing 12 transmembrane helices was built manually using *Coot* (11). Building of amino-acid side chains was facilitated with selenium and mercury sites as markers to ensure correct fitting of the a.a. sequence to the peptide model. A significant improvement in the map quality was achieved by multi-crystal averaging using the native, mercury-, and selenium-derivative data sets. Refinement of the model was carried out with *Phenix* (the phased maximum likelihood method) using the highest resolution data set available supplemented with the experimental phase information. Strong secondary structure restraint and reference-model restraint were used throughout the entire refinement. Model validation was carried out using the *Molprobity* server (12). In the final refined model, only one residue, Thr7, was associated with disallowed regions of the Ramachandran plot and was located in the N-terminal flexible region with weak electron density.

Thermofluor stability assay

Thermofluor analysis (13) was performed with a qPCR instrument, Rotor-Gene 6600 (Corbett Research, Australia) equipped with both a blue and a green fluorescence channel. Thiol-specific fluorochrome N-[4-7- (diethylamino-4-methyl-3-coumarinyl) phenyl] maleimide (CPM, Invitrogen, US) was used as the fluorescence probe for full length YajR and YajR-ΔYAM. Using the blue light channel, CPM fluorescence was measured with a 387 nm excitation and a 463 nm emission wavelength (14). In addition, Sypro orange dye (Invitrogen) was used together with the green fluorescence channel (*i.e.* 492 nm excitation and 516 nm emission) for the isolated YAM domain protein sample which was soluble but contained no Cys residue. The protein samples (*ca.* 4 μg/μl) were typically diluted from one buffer (*i.e.* 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, and 0.03% (w/v) DDM for YajR and YajR-ΔYAM; and 20 mM MES (pH 6.0) and 100 mM NaCl for YAM) into the buffer to be analyzed at a 1:20 ratio. The volume of each sample was 20 μl, and the final concentration of the protein was 0.2 μg/μl. Buffers tested included sodium acetate, MES, HEPES, Tris, and bicine. The pH range varied from 5 to 10, and the salt concentration varied from 0.1 to 2.0 M. The qPCR instrument was programmed to increase temperature from 25 to 99°C at a rate of 1°C/min. Melting temperature, Tm, was estimated to be the temperature corresponding to the minimum of the first derivative of the denaturation curve.

Disulfide bond formation assay

Membrane fractions of *E. coli* cells expressing the wild type YajR or mutation variants were collected each from 5 ml of freshly prepared cell culture. After resuspended in 400 μl of buffer A (20 mM Tris-HCl (pH 8.0), 300 mM NaCl, and 10% (v/v) glycerol), each membrane sample was divided into 20-μl aliquots. Each sample was incubated with 2 mM (final) copper phenanthroline (Sigma, US) for 1 h at 30°C. For reducing controls, samples were incubated with additional 50 mM (final) dithiothreitol (DTT) for 10 min at 30°C. All samples were subjected to sodium dodecyl sulfate-10% polycrylamide gel electrophoresis (SDS-10% PAGE) with 5 min pre-heating at 37°C. Immuno-blotting was performed with anti-His antibodies (Tagmouse monoclonal antibody) and HRP-goat anti-mouse IgG antibodies (both from EarthOx, US). Experiments were repeated multiple times, and the results were reproducible. Representative results are shown in Fig. 2B.

References

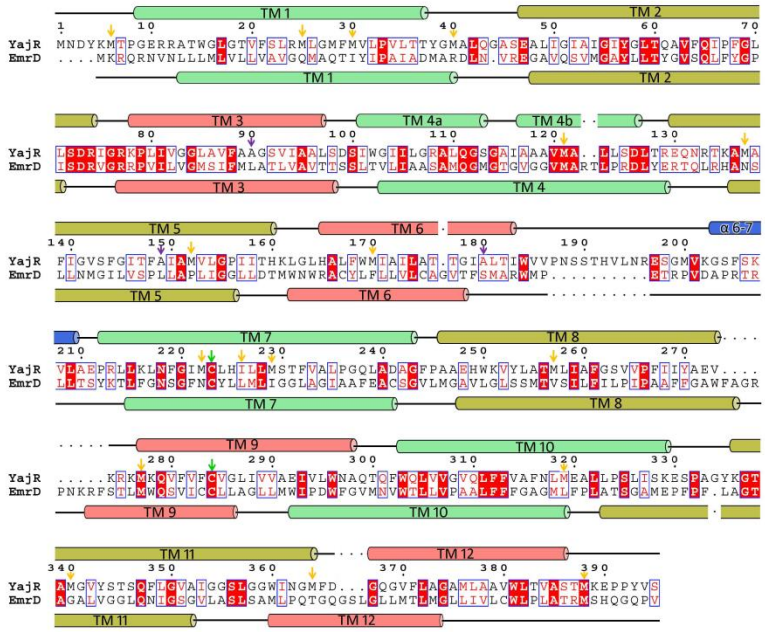
1. Eilers M, Patel AB, Liu W, & Smith SO (2002) Comparison of helix interactions in membrane and soluble alpha-bundle proteins. *Biophys J* 82(5):2720-2736.
2. Nicholson H, Becktel WJ, & Matthews BW (1988) Enhanced protein thermostability from designed mutations that interact with alpha-helix dipoles. *Nature* 336(6200):651-656.
3. Huang Y, Lemieux MJ, Song J, Auer M, & Wang DN (2003) Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. *Science* 301(5633):616-620.
4. Abramson J*, et al.* (2003) Structure and mechanism of the lactose permease of Escherichia coli. *Science* 301(5633):610-615.
5. Zimmermann M*, et al.* (2009) Metal binding affinities of Arabidopsis zinc and copper transporters: selectivities match the relative, but not

the absolute, affinities of their amino-terminal domains. *Biochemistry* 48(49):11640-11654.

6. Fan J*, et al.* (2011) An efficient strategy for high throughput screening of recombinant integral membrane protein expression and stability. *Protein Expr Purif*.
7. Miroux B & Walker JE (1996) Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* 260(3):289-298.
8. Otwinowski Z & Minor W (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* 276(Macromolecular Crystallography, part A):307-326.
9. Schneider TR & Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* 58(Pt 10 Pt 2):1772-1779.
10. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
11. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.
12. Davis IW, Murray LW, Richardson JS, & Richardson DC (2004) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* 32(Web Server issue):W615-619.
13. Cummings MD, Farnum MA, & Nelen MI (2006) Universal screening methods and applications of ThermoFluor. *J Biomol Screen* 11(7):854-863.
14. Alexandrov AI, Mileni M, Chien EY, Hanson MA, & Stevens RC (2008) Microscale fluorescent thermal stability assay for membrane proteins. *Structure* 16(3):351-359.
15. Thompson JD, Gibson TJ, & Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* Chapter 2:Unit 2 3.
16. Gouet P, Robert X, & Courcelle E (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* 31(13):3320-3323.
17. McGuffin LJ, Bryson K, & Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16(4):404-405.
18. Crooks GE, Hon G, Chandonia JM, & Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6):1188-1190.
19. Sonnhammer EL, Eddy SR, Birney E, Bateman A, & Durbin R (1998) Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res* 26(1):320-322.
20. Finn RD, Clements J, & Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39(Web Server issue):W29-37.
21. Jones DT, Taylor WR, & Thornton JM (1994) A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry* 33(10):3038-3049.

Figures

A


TM 1
TM 2
TM 3
TM 4a
TM 4b
TM 4
TM 5
TM 6
α6-7
TM 7
TM 8
TM 9
TM 10
TM 11
TM 12
YajR
EmrD

B

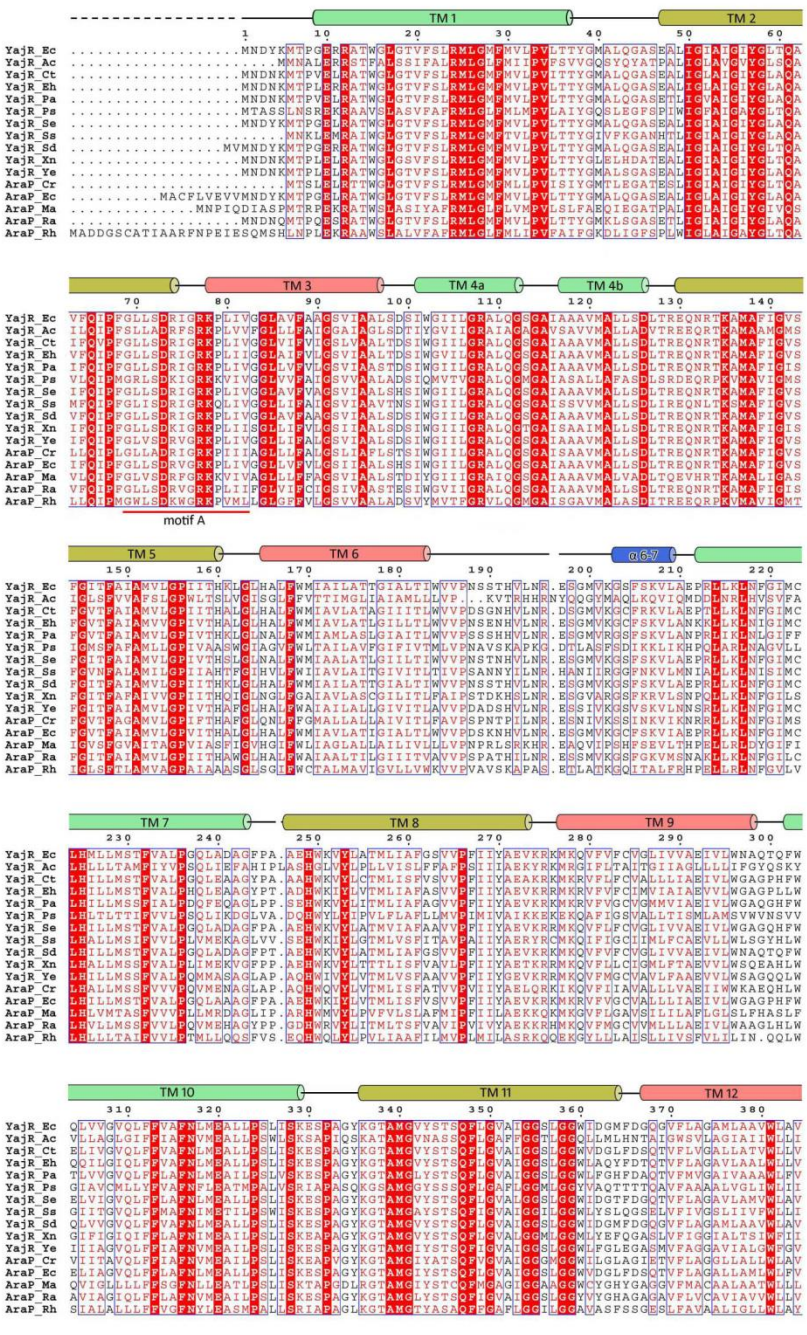

TM 1
TM 2
TM 3
TM 4a
TM 4b
motif A
TM 5
TM 6
α6-7
TM 7
TM 8
TM 9
TM 10
TM 11
TM 12
β 1
α 1
β 2
β 3
α 2

Figure S1. Amino acid sequence analysis of *E. coli* YajR. *A*. Alignment of primary sequences of YajR and EmrD. Corresponding secondary structure elements are marked above and below the sequences, respectively. Positions of observed Se-Met residues in the crystal are indicated with orange arrows, and native and engineered Cys residues binding Hg-atoms are indicated with green and purple arrows, respectively. *B*. Alignment of sequences of YajR and AraP from a variety of Gram negative bacteria. YajR_Ec: *E. coli*; YajR_Sd: *Shigella dysent*eriae; YajR_Se: *Salmonella enteria*; YajR_Ct: *Cronobacter turicensis*; YajR_Eh: *Escherichia hermannii*; YajR_Pa: *Pantoea anantis*; YajR_Ye: *Yersinia enterocolitica*; YajR_Xn: *Xenorhabdus nematophila*; YajR_Ss: *Serratia symbiotica*; YajR_Ps: *Pseudoalteromonas*; YajR_Ac: *Acinetobacter calcoaceticus*; AraP_Ec: *Enterobacter cloacae*; AraP_Ra: *Rahnella aquatilis*; AraP_Cr: *Candidatus Regiella insecticola*; AraP_Rh: *Rheinheimera*; and AraP_Ma: *Methylomicrobium album*. Identical and similar residues are colored in white on red background and in red on white background, respectively. The secondary structural elements of *E. coli* YajR are indicated above the sequence alignment and color coded as in Fig. 1B, along with selected residue position numbers. Conserved motifs A and B of MFS transporters are marked below the sequences. Sequences were aligned with the program ClustalX (15) and formatted with ESPript (16) .

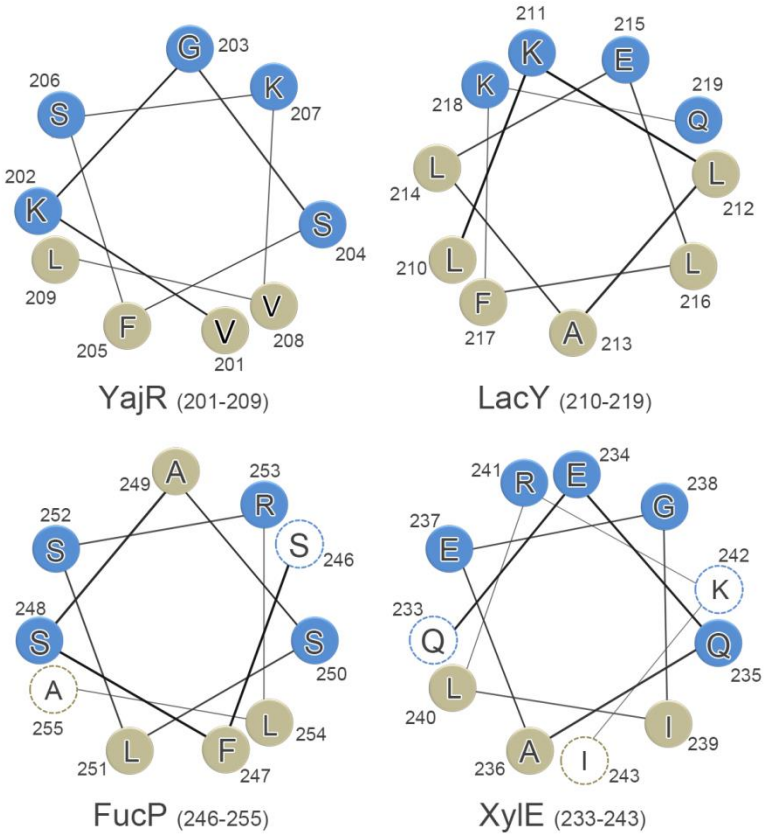


Figure S2. Helix wheels of the amphipathic helices (α6-7) in the inter-domain linkers of known structures of MFS proteins. Polar residues are shown in blue color circle, and hydrophobic residues are in wheat color. Potential terminal residues that fit the pattern are included in open circles. Note that the structures of $PepT_{St}$ and $PepT_{So}$ contain an insertion of a TM helix pair instead of an amphipathic α-helix; and the EmrD/2GFP structure does not show an amphipathic α-helix (albeit a 3-turn amphipathic helix can be predicted in the linker region using the PSIPRED server (17)). Thus, these three structures are not included.

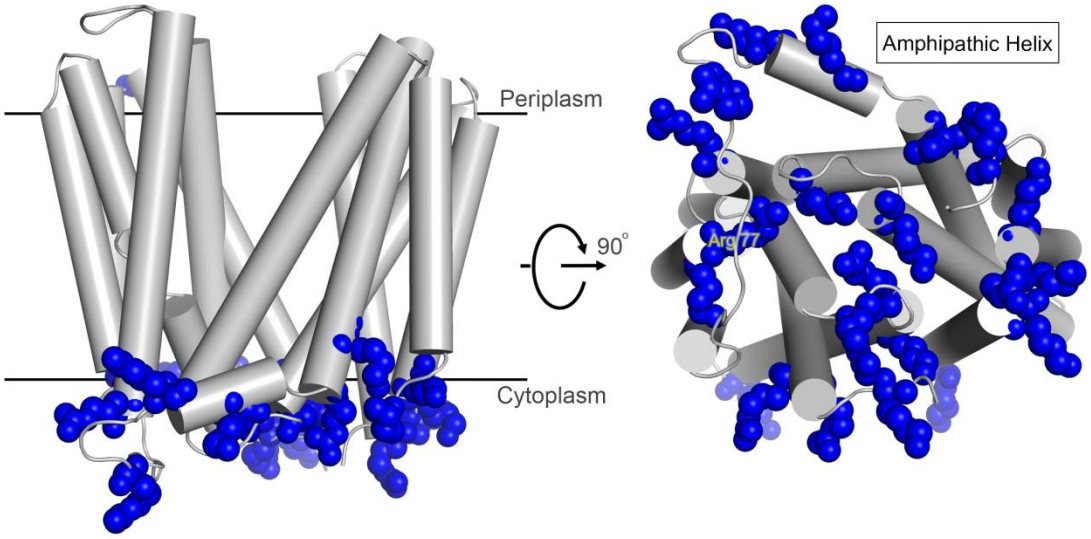


Figure S3. Distribution of basic amino acid residues in the TM core of YajR. The transmembrane core is shown in a cartoon presentation, and basic residues in space filling models. The distribution clearly follows the positive-inside rule.

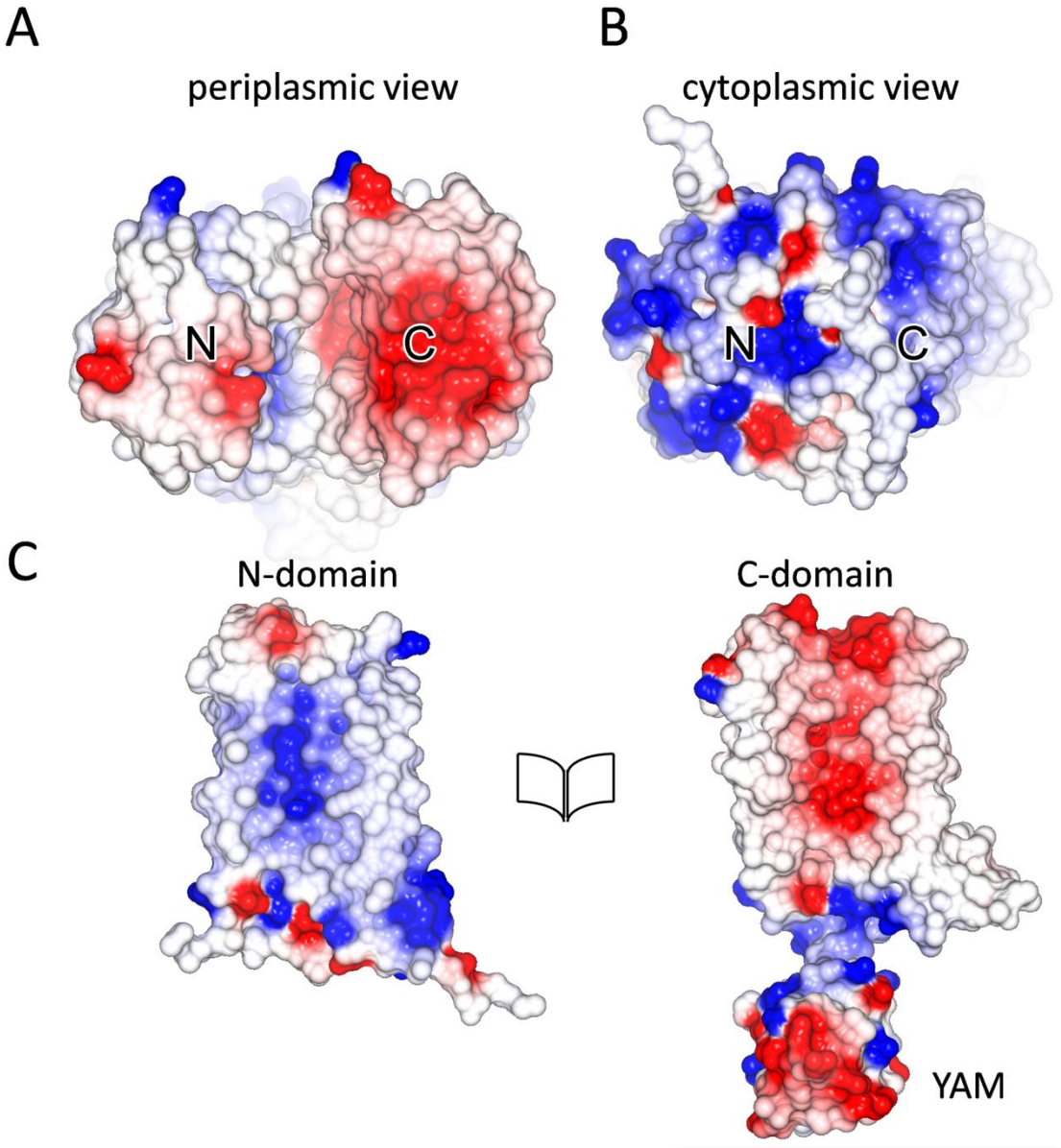


Figure S4. Molecular surface of YajR color-coded for electrostatic potential. *A*. View from the periplasmic side. *B*. View from the cytoplasmic side (without the YAM domain). *C.* An open-book view of the central cavity. All surface figures were generated using the default settings in *Coot*. The positively charged region is in blue and the negatively charged region in red.

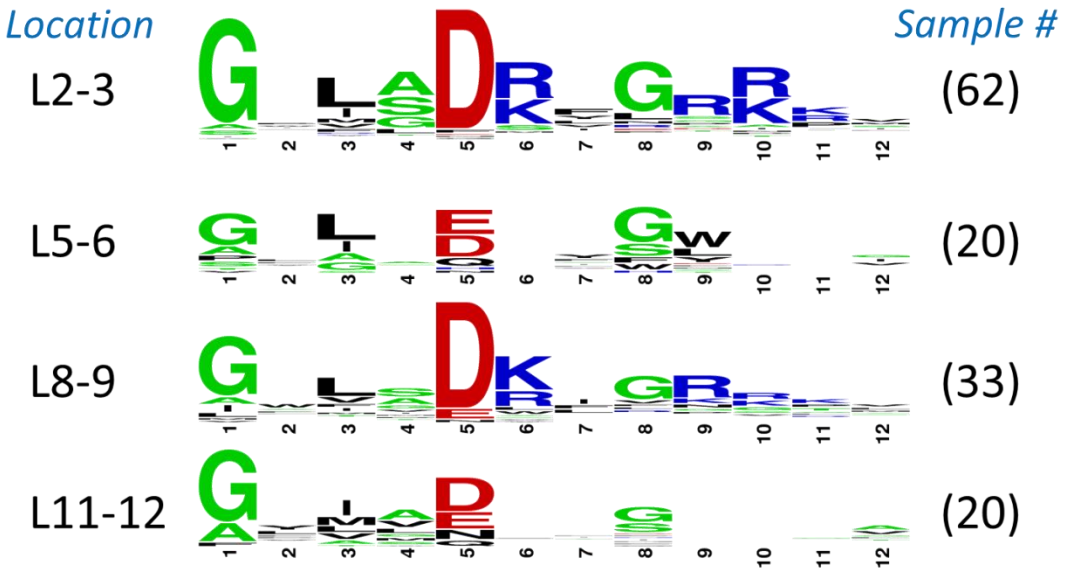


Figure S5. Conservativeness of A-like motifs in *E. coli* MFS transporters. Data were from Table S4. The higher the frequency of a residue occurs at a given position, the larger the letter size is. Short side chain residues (*i.e.* Gly, Ala, Ser, and Thr) are in green; acidic residues (Asp and Glu) in red; basic residues (Arg, Lys, and His) in blue; and the rest in black. The figure was generated with the web program WebLogo (18).

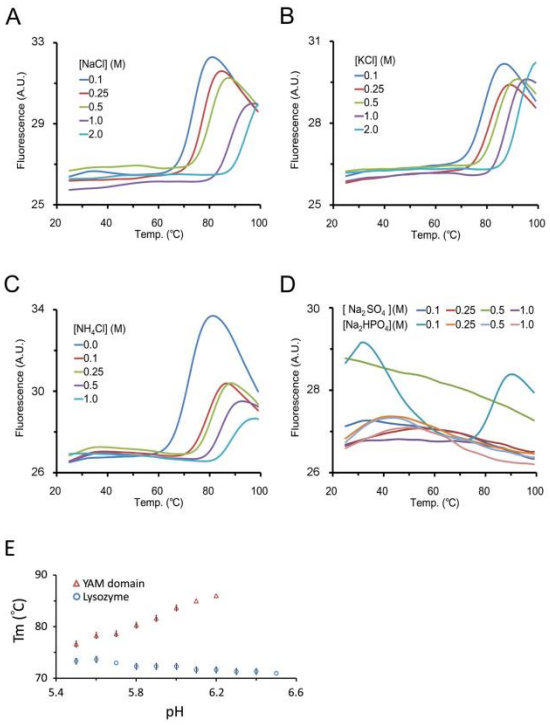


Figure S6. Thermal stability of YAM domain of YajR. Thermal denaturation curves of YAM domain were measured using the green fluorescence of Sypro Orange dye (Invitrogen). *A*. Effects of NaCl concentration on YAM thermal stability. *B*. Effects of KCl concentration. *C*. Effects of NH_4Cl concentration. *D*. Effects of sodium salts. In panel *A*–*D*, the buffers contained 50 mM MES (pH 6.0) and specified salt. Multiple measurements were performed, and the results were reproducible. Raw data from a representative measurement are shown. *E*. Effects of pH on YajR YAM thermal stability. Buffers contained 50 mM MES at specified pH and 100 mM NaCl. At pH higher than 6.2, the fluorescence signal diminished. Chicken egg white lysozyme was used as a control under identical conditions as the YAM sample. Error bars were calculated based on two or more independent measurements.

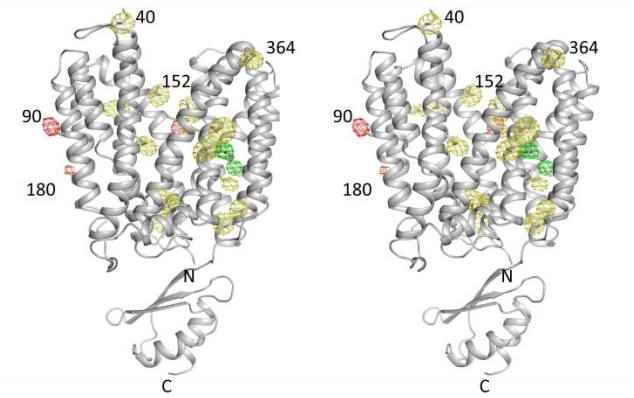


Figure S7. Stereo view of heavy atom positions. Yellow chicken wires represent electron densities from a seleno anomalous difference *Fourier* map contoured at 3.5 σ level, and they indicate positions of Se atoms in the Se-Met derivative of YajR. Similarly, green and magenta chicken wires represent electron densities from mercury anomalous difference *Fourier* map contoured at 5.0 σ, indicating positions of mercury atoms bound to cysteine residues of the native and Ala-to-Cys point mutation crystals, respectively.

Tables

Table S1. *Statistics from data collection and refinement*

	Native	$Hg(CN)_2$		Se	A90C $Hg(CN)_2$	A180C $Hg(CN)_2$
		peak	inflection	peak	peak	peak
Data processing						
Beamline	BL17U at SSRF	BL41XU at SPring-8		BL41XU at SPring-8	BL41XU at SPring-8	BL41XU at SPring-8
Wavelength (Å)	0.9200	1.0063	1.0089	0.9791	1.0063	1.0063
Space group	$P6_122$	$P6_122$		$P6_122$	$P6_122$	$P6_122$
Cell dimensions *a = b, c* (Å)	128.7, 165.1	128.9, 166.7		128.7, 163.9	128.1, 164.4	128.8, 164.2
Resolution (Å)	50–3.15 (3.26–3.15)[a]	50–4.20 (4.35–4.20)	50–4.50 (4.66–4.50)	50–3.70 (3.83–3.70)	50–3.75 (3.88–3.75)	50–3.70 (3.83–3.70)
Completeness (%)	99.6 (100)	99.5 (100)	99.5 (100)	98.8 (96.0)	98.7 (100)	98.9 (99.8)
R_{merge} (%)[b]	8.9 (99.5)	20.3 (91.0)	17.8 (66.5)	13.0 (>100)	12.7 (>100)	12.0 (79.1)
I /σ(I)	22.9 (2.2)	11.0 (2.1)	12.8 (3.7)	10.7 (1.5)	14.7 (2.5)	20.1 (3.3)
Unique reflections	14,541 (1,419)	6,308 (608)	5,275 (510)	9,017 (847)	8,636 (838)	9,049 (876)
Redundancy	7.9 (7.4)	12.1 (11.1)	12.1 (11.0)	5.5 (4.6)	13.2 (11.0)	13.9 (13.9)
Refinement						
Resolution (Å)	39–3.15					
No. of reflections (test)	14,472 (707)					
R_{work} / R_{free} (%)[c]	27.1 / 29.0					
Average *B*–factor ($Å^2$)	66					
Rmsd						
bond lengths (Å)	0.011					
bond angles (°)	1.800					
Ramachandran plot (%)[d]						
favored region	92.5					
allowed region	7.3					
disallowed region	0.2					

[a] Values in parentheses are for the highest resolution shell. [b] $R_{merge} = \sum_{hkl}\sum_{i}|I_i - (I)|/|(I)|$ where I_i is the intensity for the *i* th measurement of an equivalent reflection with indices h, k, l.

[c] $R = \sum|F_o - F_c| / \sum|F_o|$ where F_o and F_c are the observed and calculated structure factors, respectively.

[d] Calculated using *MolProbity*.

Table S2. *Patterns of short side chain distribution in TM helices of YajR.*

No.	Pattern	Position	Helix	Contact
1	**A**LI**G**IAI**G**IY**G**LTQ**A**	48–62	TM2	TM11 (potentially inter-domain w/ #11)
2	**G**GL**A**VF**AA**G**S**VIA**A**	83–96	TM3	TM4 (#3) and TM6
3	**S**IW**G**IIL**G**	100–107	TM4a	TM3 (#2)
4	**A**AVM**A**LL**S**	117–125	TM4b	TM2 (motif A) & TM11 (inter-domain w/ #10)
5	**T**KAM**A**FI**G**VSF**G**	134–145	TM5	TM10 (inter-domain w/ #9)
6	**A**TT**G**IAL**T**	175–182	TM6	TM1
7	**S**TFV**A**LP**G**	230–237	TM7	TM11 (#11) & TM12
8	**T**MLI**A**FG**S**	256–263	TM8	TM10
9	**A**LLP**S**LI**S**	321–328	TM10	TM5 (inter-domain w/ #5)
10	**GT**AM**G**VY**ST**	337–344	TM11	TM2 (inter-domain w/ motif A) & TM4 (#4)
11	**A**IG**G**SLG**G**	352–359	TM11	TM2 (potentially inter-domain w/ #1) & TM7 (#7)
12	**A**VWL**A**VA**S**	379–386	TM12	

Location of the patterns:

MNDYKMTPGE RRATWGLGTV FSLRMLGMFM VLPVLTTYGM **ALQGASEALI GIAIGIYGLT**
QAVFQIPFGL LSDRIGRKPL IV**GGLAVFAA GSVIAA**LSD**S IWGIILG**RAL QGSGAI**AAAV**
MALLSDLTRE QNR**TKAMAFI GVSFG**ITFAI AMVLGPIITH KLGLHALFWM IAIL**ATTGIA**
LTIWVVPNSS THVLNRESGM VKGSFSKVLA EPRLLKLNFG IMCLHMLLM**S TFVALPG**QLA
DAGFPAAEHW KVYLA**TMLIA FGS**VVPFIIY AEVKRKMKQV FVFCVGLIVV AEIVLWNAQT
QFWQLVVGVQ LFFVAFNLME **ALLPSLIS**KE SPAGYK**GTAM GVYS**TSQFLG V**AIGGSLGG**W
IDGMFDGQGV FLAGAMLA**AV WLAVAS**TMKE PPYVSSLRIE IPADIAANEA LKVRLLETEG
VKEVLIAEEE HSAYVKIDSK VTNRFEVEQA IRQA

Table S3. A-like motifs in known structures of MFS proteins

Protein	PDB ID	Location	A-like motif*	Putative charge-relay system*
YajR (antiporter)	4XYZ (outward)	L2-3; obs.	GLLSD^{73} RIGRKP	Asp73-Arg77-Asp126
		L8-9; obs.	IIFAE^{272}VKRKMK	Glu272-Lys329-Glu130
LacY (symporter)	1PV7 (inward)	L2-3	GLLSD^{68} KLGLRK	Asp68-Lys131-Glu130
		L11-12; obs. periplasmic	GNMYE^{374}SIGFQG	none
GlpT (Pi^- antip.)	1PW4 (inward)	L8-9	GWMSD^{314}KVFRGNRG	Asp314-Arg321-Glu374
		L11-12; obs. periplasmic	GYTVD^{409}FFGWDG	none
EmrD (antiporter)	2GFP (intermediate)	L2-3	GPISD^{68} RVGRRP	Asp68-Arg72-Asp123
$PepTs_o$ (symporter)	2XUT (inward)	L2-3	GWIAD^{79} RFFGKYN	Asp79-Lys84-Asp136
FucP (symporter)	3O7Q (outward)	L11-12 periplasmic	GFVSD^{406}AAGNIP	none
$PepT_{St}$ (symporter)	4APS (inward)	L2-3	GFVAD^{79} RIIGARP	Asp79-Arg80-Asp210
		L5-6; obs. periplasmic	GAANE^{171}AAGYHV	none
XylE (symporter)	4GC0 (outward)	L2-3; obs.	GYCSN^{80} RFGRRD	Asn80-Arg84-Glu153
		L8-9; obs.	IMTVD^{337}KFGRKP	Asp337-Arg341-Glu397
PiPT (symporter)	4J05 (intermediate)	L2-3	GVLGD^{97} SFGRK	none
		L8-9; obs.	LFLID^{381}IVGRKK	Asp381-Arg385-Glu440
NarK (exchanger)	4JRE (intermediate)	L8-9; obs.	GALSD^{313}RLGGTR	none
NarU (exchanger)	4IU9 (intermediate)	L8-9; obs.	GAISD^{311}KFGGVR	none

*These structural features were identified based on visual investigation of 3D structures, complemented by manual analysis of the primary sequences.

***Table S4**. Distribution of potential A-like motifs in MFS proteins from E. coli.*

protein	ID	# res.	# TM	L2-3	#	TM4 -C	L5-6	#	L8-9	#	TM10 -C	L11-12	#
AMPG	P0AE16	491	14	SPLMDRYTPPFF	66	FDAW			LSITDKHLYSMG	308	LEYT		
ARAE	P0AE24	472	12	GWLSFRLGRKYS	80	SEMA	AFLSDTAFSYSG	164	VFTVDKAGRKPA	315	SEIQ		
ARAJ	P23910	394	12										
BCR	P28246	396	12	GPMADSFGRKPV	66	RDYI							
CYNX	P17583	384	12										
DGOT	P0AA76	430	12	GWFLDRVGSRVT	72	QEML			GWVADLLVRKGF	297			
DTPA	P77304	500	14	GWLGDKVLGTKR	78				AAIYNKMGDTLP	337			
DTPB	P36837	489	14	GYVGDHLLGTKR	71		PVIADRFGYSVT	164	AFKLDKTGRNKM	257			
DTPD	P75742	493	14	GFLADKVLGNRM	66	GELY	GYAQEEYSWAMG	159				GVIADQTSQASF	431
EMRB	P0AEJ0	512	14				GYISDNYHWGWI	159					
EMRD	P31442	394	12	GPISDRVGRRPV	64	RDLY	GGLLDTMWNWRA	152					
EMRY	P52600	512	14	GRLAQRIGELRL	74	LDVL	GYICDNFSWGWI	163					
ENTS	P24077	416	12	GVLADRYERKKV	73	RENL							
EXUT	P0AA78	472	12									GALADTIGFSPL	416
FSR	P52067	406	12	GYWTDKYPMPWS	81				GPVGDKIGRKYV	284	QELL	GLIADHTSIELV	371
FUCP	P11551	438	12									GFVSDAAGNIPT	397
GALP	P0AEP1	464	12	GWLSFKLGRKKS	73	SEIA			IGLVDRWGRKPT	308	SEIQ		
GARP	P0AA80	444	12	GWLLDKFGSKKV	74		AWGWEHVFTVMG	174	GVFSDYLIKRGL	307	SDTA		
GLPT	P08194	452	12	GSVSDRSNPRVF	84				GWMSDKVFRGNR	310	LELA	GYTVDFFGWDGG	400
HSRA	P31474	475	12	GWLADRFGTRRI	66		SSGIELFGEKIV	212					
KGTP	P0AEX3	432	12	GRIADKHGRKKS	84	GEYG			GALSDKIGRRTS	300	AEMF		
LACY	P02920	417	12	GLLSDKLGLRKY	64	IEKV						GNMYESIGFQGA	365

MDFA	P0AEY8	410	12	GPLSDRIGRRPV	72	QESF							
MDTD	P36554	471	14	GWLADKVGVRNI	66	REQY	GLLVEYASWHWI	166					
MDTG	P25744	408	12	GGLADRKGRKLM	75		GLLADSYGLRPV	163	GKLGDRIGPEKI	279			
MDTH	P69367	402	12	GAIADRFGAKPM	67	RDTL							
MDTL	P31462	391	12	GKVADRSGRKPV	59	RDTL							
MDTM	P39386	410	12	GPLSDRIGRRPV	69	QEAF							
MELB	P02921	473	12										
MHPT	P77589	403	12	GMLADRYGRKRI	71	SEAA			GALMDKLRPVTM	271			
NANT	P41036	496	14	GAMGDRYGRRLA	76	IESW	VVAAQVYSLVVP	158	GFLGDWLGTRKA	332		ALIAQRLDLGTA	423
NEPI	P0ADL1	396	12				SFLGELIGWRNV	165				GYALDNIGLTSP	356
NUPG	P0AFF4	418	12	GIVADKWLSAKW	59								
PROP	P0C0L7	500	12	GMLGDKYGRQKI	87	GEYT			GLLSDRFGRRPF	312			
RHMT	P76470	429	12				GALLEMHGFMGH	160					
SETA	P31675	392	12	AKRSDSQGDRRK	70	REYA							
SETB	P33026	393	12	AGRSDKRGDRKS	72	REHA						GIVAEIWNYHAV	361
SETC	P31436	394	12	ARHSDKQGDRKL	70	REHA							
SHIA	P76350	438	12	GHFGDRLGRKRM	85	VESA			AWLADRFGRRRV	307	TEMF		
SOTB	P31122	396	12									ALVGNQVSLHWS	350
TSGA	P60778	393	12									GPIVEHSGPQAA	349
UHPC	P09836	439	12	GIVSDRSNARYF	86				GWGSDKLFNGNR	307	AECS		
UHPT	P0AGC0	463	12	SYYADGKNTKQF	86				GWLSDLANGRRG	316	SEIF		
UIDB	P0CE44	457	12	VRVFDAFADVFA	53							GYIANQVQTPEV	390
XAPB	P45562	418	12	GIIADKWLRAER	59	LDPV						GMAVDYFSVDGV	366
XYLE	P0AGF4	491	12	GYCSNRFGRRDS	76	AELA	ARSGDASWLNTD	186	IMTVDKFGRKPL	333	SEIF	FPMMDKNSWLVA	426
YAAU	P31679	443	12	GYISDKVGRRKM	74								

YAGG	P75683	460	12	GLLVDRTRTRHG	64		GKGDEQVGYFGM	171						
YAJR	P77726	454	12	GLLSDRIGRKPL	69	SDLT	PIITHKLGLHAL	156	IIYAEVKRKMKQ	268	KESP			
YBJJ	P75810	402	12						GWFIDRYSRVAV	272	SDTG	GYLGEHYGLRSA	359	
YCAD	P21503	382	12						GRLADKFGRLLV	256	CEKV			
YDEE	P31126	395	12	GILADKFDKKRY	64	ADNL								
YDFJ	P77228	427	12	GKMGDRIGRKKV	51	AEYA			GWLSDKIGRRIP	267	LENI			
YDHC	P37597	403	12	GPLSDRYGRKPV	62	TDYY								
YDHP	P77389	389	12				TWLGETIGWRMS	150	GKLADRSVNGTL	258				
YDIM	P76197	404	12	GLLSDRFGRRPF	63	LDAG						AHLSQRSIADIM	355	
YDIN	P76198	421	12	GVISDKFGRRAV	69									
YDJE	P38055	452	12	GFIGDYFGRRRA	78									
YDJK	P76230	459	12	GIIGDKTGRRNA	79	TEYM			MLVMDKIPRKTM	320	PEIW			
YEAN	P76242	393	12											
YEBQ	P76269	457	14	SFLGDMFGYRRI	72				GYLIERVHAGLL	315		ALMLNQFGDNGT	417	
YEGT	P76417	425	12	GSITDRFFSAQK	58		GFLPQILGYADI	146				GVMMEKMFAYQE	358	
YFCJ	P77549	392	12	GRLADQYGAKRS	72				GWMPDRFGGVKV	267	VEVV			
YGAY	P76628	394	12	VPLGDMFERRRL	67		GLLANLGGWRTV	155	GGFADKGKSHHT	268	LDLT			
YHHS	P37621	405	12	GRYADSLGPKKI	74									
YHJE	P37643	440	12	GHFGDRVGRKAT	85	TENA			GLLADAFGRRKS	305	PELF			
YHJX	P37662	402	12	GKLQERFGVKRV	67				GILSDKIARIRV	271	SEFF			
YICJ	P31435	460	12	GLLADRTRSRWG	66	NDPT								
YIHN	P32135	421	12	IYFADKLPRRYT	64				GMISDKILKSPS	276		GYILDLNPGIIG	372	
YIHO	P32136	467	12	GFLLDSRKNIGP	71									
YIHP	P32137	461	12	GIMLDSRRKIGP	74									
YJDL	P39276	485	14	GWLADRLLGNRT	65	GELY	GLAAQWYGWHVG	158	LASPESRGNSTL	328				

YJHB	P39352	405	12	GAMADKYGRKPM	71	GEYA			GFVGDKIGVKKA	269	YDYF		
YJIJ	P39381	453	12										
YJJL	P39398	453	12	GPLLDRKGPRLM	98				GYVTDWLVKGGM	314		GFIVDTTHSFRL	412
YNFM	P43531	417	12	GPLSDAIGRKPV	92	SEEI							
YQCE	P77031	425	12	GVIADKFSHRKM	66								

Table methods: Open reading frames of *E. coli* (K12) were down-loaded from the Swiss-Prot database. HMM models of MFS family were down-loaded from the Pfam database (19) and were used with the program HMMER-3.0 (20) to identify MFS proteins from *E. coli* (K12) genome. TM helices were additionally predicted using the program MEMSAT (21). A-like motifs were identified at proper loop regions using pattern search supplemented with visual inspection. For motifs A^{L2-3} and A^{L8-9}, short side chain residues for helix packing (*i.e.* positions +1 and +4), the central acidic residue for the charge-helix dipole interaction (*i.e.* +5), short side chain residues at position +8, and basic residue(s) for the charge-relay system (*e.g.* position +9) were included in the pattern search. For motifs A^{L5-6} an A^{L11-12}, only the first three factors were considered. Main results: Seventy-seven unique sequences were identified to belong to the MFS group. Of them, 67 were predicted to fold into 12-TM MFS proteins; and the rest ten were predicted to be 14-TM MFS proteins. Of the 77 sequences, only four do not show any detectable A-like motif. Sixty-two sequences contain potential motif A^{L2-3}; 21 sequences contain potential motif A^{L5-6}; 33 sequences contain potential motif A^{L8-9}; and 20 sequences contain potential motif A^{L11-12}. Of the 62 motifs A^{L2-3}, 36 are accompanied by an acidic residue from the C-terminus of the predicted TM4; and of the 33 motifs A^{L8-9}, 21 are accompanied by an acidic residue from the C-terminus of the predicted TM10. In general, motifs A^{L2-3} and A^{L8-9} are more clearly defined.

***Table S5**. Charge-helix dipole interactions in the YajR crystal structure.*

Helix dipole	Position	Charged residue	Comments
TM1-N	11	Glu10	strictly conserved in YajRs
TM2-N	48	Glu47	conserved in YajRs
TM3-C	97	His165	
TM4a-C	111	Arg24	strictly conserved in YajRs
TM4b-C	125	Arg133	conserved in YajRs
TM5-N	134	Glu272	inter-domain, conserved in YajRs
TM6-C	185	Lys78	strictly conserved in YajRs
α6-7-N	204	Glu197	conserved in YajRs
TM7-N	213	Glu211	
TM8-C	274	Arg275	
TM9-N	277	Glu390	from the TM-YAM domain linker
TM10-C	330	Lys336	highly conserved (*i.e.* Lys-Arg)
TM11-N	337	Asp73	inter-domain, strictly conserved in YajRs
TM12-N	367	Asp366	
TM12-C	386	Lys278	highly conserved