doi:10.1016/j.jmb.2006.08.072 J. Mol. Biol. (2006) 364, 54–67

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Quasi-symmetry in the Cryo-EM Structure of EmrE Provides the Key to Modeling its Transmembrane Domain

Sarel J. Fleishman[1], Susan E. Harrington[1], Angela Enosh[2] Dan Halperin[2], Christopher G. Tate[3] and Nir Ben-Tal[1]*

[1]*Department of Biochemistry George S. Wise Faculty of Life Sciences, Tel-Aviv University Ramat Aviv 69978, Israel*

[2]*School of Computer Sciences Tel-Aviv University, Ramat Aviv 69978, Israel*

[3]*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK*

**Corresponding author*

Small multidrug resistance (SMR) transporters contribute to bacterial resistance by coupling the efflux of a wide range of toxic aromatic cations, some of which are commonly used as antibiotics and antiseptics, to proton influx. EmrE is a prototypical small multidrug resistance transporter comprising four transmembrane segments (M1–M4) that forms dimers. It was suggested recently that EmrE molecules in the dimer have different topologies, i.e. monomers have opposite orientations with respect to the membrane plane. A 3-D structure of EmrE acquired by electron cryo-microscopy (cryo-EM) at 7.5 Å resolution in the membrane plane showed that parts of the structure are related by quasi-symmetry. We used this symmetry relationship, combined with sequence conservation data, to assign the transmembrane segments in EmrE to the densities seen in the cryo-EM structure. A C^{α} model of the transmembrane region was constructed by considering the evolutionary conservation pattern of each helix. The model is validated by much of the biochemical data on EmrE with most of the positions that were identified as affecting substrate translocation being located around the substrate-binding cavity. A suggested mechanism for proton-coupled substrate translocation in small multidrug resistance antiporters provides a mechanistic rationale to the experimentally observed inverted topology.

Keywords: dual topology; protein structure prediction; structural bioinformatics; cryo-EM; mechanism of action

Introduction

Bacterial multidrug resistance is a growing challenge to medical treatment, with previously harmless bacteria inducing life-threatening infections.[1] One of the mechanisms for the acquirement of multidrug resistance is the active extrusion of toxic compounds from the bacterial cell through membrane transporters. Efflux of toxic compounds is driven either by ATP hydrolysis, as in the ABC transporter superfamily,[2] or by coupling the extrusion of toxic compounds to the inward movement of protons down their electrochemical gradient, as in the small multidrug resistance (SMR) family of antiporters. Of the SMRs, EmrE is a representative from *Escherichia coli*, which has been extensively characterized structurally, phylogenetically, and biochemically[3,4] These analyses have provided evidence that EmrE contains four transmembrane (TM) segments that form α-helices.[5,6]

A recent electron cryo-electron microscopy (cryo-EM) analysis of 2D crystals of EmrE bound to one of its substrates, tetraphenylphosphonium (TPP^+), clearly resolved the eight α-helices comprising the EmrE dimer at an in-plane resolution of 7.5 Å and 16 Å perpendicular to the membrane plane.[7] However, at this resolution, the individual amino acid residues were not observed, and the TM segments could not be assigned unambiguously to

Abbreviations used: SMR, small multidrug resistance; TM, transmembrane; cryo-EM, cryo-electron microscopy; TPP^+, tetraphenylphosphonium.

E-mail address of the corresponding author: nirb@tauex.tau.ac.il

the densities representing the α-helices. The 2D crystals of EmrE bind TPP^+ with the same high affinity as detergent-solubilized EmrE, and EmrE in the native *E. coli* membrane,[4,8] so it is thought that the cryo-EM structure of EmrE is a faithful representation of the protein's physiological conformation. Quasi-symmetry between six helices was detected around an axis lying within the plane of the membrane, suggesting that the EmrE monomers might assume dual topology in the membrane, with the monomers arranged in an inverted or upside-down manner with respect to one another.[7] In contrast, no obvious symmetry relationship was observed around axes perpendicular to the membrane plane in either the 3D structure or a previous 2D projection map.[9]

Two atomic-resolution X-ray structures of EmrE have been solved in recent years. The first structure at 3.8 Å resolution appears to have trapped the molecule in an unphysiological state,[10] and is incompatible with much of the biochemical data on this protein.[11] Recently, another X-ray structure of EmrE was solved at 3.7 Å resolution,[12] which included one molecule of bound substrate TPP^+ per dimer. However, it has been argued that this structure too may not be physiologically relevant,[13] for three main reasons. (i) The X-ray structure is very different from the cryo-EM structure of EmrE.[12] (ii) Several key residues that were shown to be critical for substrate binding are not in a position to bind substrate in the structure. For instance, it was demonstrated by different experimental approaches that Glu14 residues from both monomers are crucial for translocation, participate in substrate and proton binding,[14–19] and are in proximity to one another.[20] By contrast, the X-ray structure shows that Glu14 from only one monomer forms partial contact with substrate and the two glutamate residues are over 20 Å apart. (iii) Evolutionary conservation has been shown to be a powerful predictor of helix orientations in integral membrane proteins, with conserved amino acid positions usually occupying locations that are buried in the protein core, whereas lipid-facing positions are evolutionarily variable;[21–28] the X-ray structure of EmrE orients many conserved positions (Figure 1(a)) towards lipid, and conversely, variable amino acids are placed at helix–helix interfaces.[29]

The difficulties that have arisen in determining a high-resolution structure of EmrE that accounts for the body of experimental evidence and recent data supporting the dual topology of EmrE and other members of the SMR family[30,31] gave us the impetus to try to understand the cryo-EM structure through modeling strategies. The proposal of the dual topology architecture of EmrE contradicts previous experimental data that suggested EmrE had one unambiguous topology,[32] but could obviously have crucial implications for structural modeling. Here, we show that the most straightforward structural interpretation of dual topology, i.e. that EmrE is arranged as an anti-parallel homodimer, provides the key for determining a model of EmrE based on the cryo-EM structure.

Results

Quasi-symmetry and helix assignment

The assignment of the two sets of four hydrophobic segments seen in the sequence of EmrE to the eight helices observed in the cryo-EM structure is potentially the most significant hurdle in the structural modeling (theoretically having $4\times8! = 161{,}280$ different permutations).[7] However, if a symmetry relationship existed between two parts of the structure, this problem could be greatly simplified (to $2\times4! = 48$ permutations). The previous analysis of the cryo-EM structure of EmrE identified symmetry between two parts of the structure around an axis of symmetry within the plane of the membrane, but there were no symmetry relationships around axes perpendicular to the membrane plane.[7] Recent data suggesting dual topology in EmrE molecules provide additional support for the in-plane 2-fold symmetry axis.[30,31] Indeed, several integral membrane proteins contain two structurally related domains that are related by a rotational axis of quasi-symmetry within the membrane plane (e.g. GlpF,[33] ClC,[34] and SecYEβ[35]).

To derive the most likely helix arrangement for EmrE, four pieces of experimental data were used. (1) Positions of α-helices were based on the cryo-EM structure.[7] (2) The continuous density between the ends of helices F and H suggested that they were adjacent in the amino acid sequence (Figure 1(b)).[7] (3) The two monomers in the EmrE dimer are represented by A-D and E-H, based upon the symmetrical relationship between A-B-C and H-G-F, correspondingly (Figure 1(b)). (4) Densities A-B-C-F-G-H that form the substrate-binding chamber are composed of helices M1, M2, and M3, because amino acid residues that are involved in substrate binding and translocation are found only in these three helices (Table 1). These data alone were insufficient to give a conclusive model, so evolutionary conservation was used to guide the assignment of sequence segments to helices. The rationale behind the use of evolutionary conservation for helix assignment is that residues that are packed against other helices are conserved during evolution, since even minor substitutions in such positions often weaken interhelix contacts and adversely affect protein function.[21,22,25–28,36] Conversely, lipid-exposed positions are expected to be generally accommodating to sequence variability. Hence, we correlated the conservation of sequences with the extent of burial of each of the helices observed in the cryo-EM structure against other helices to constrain the possible assignments.

We found that the most informative helices in the cryo-EM structure were C and F, which are related to one another by the in-plane symmetry axis

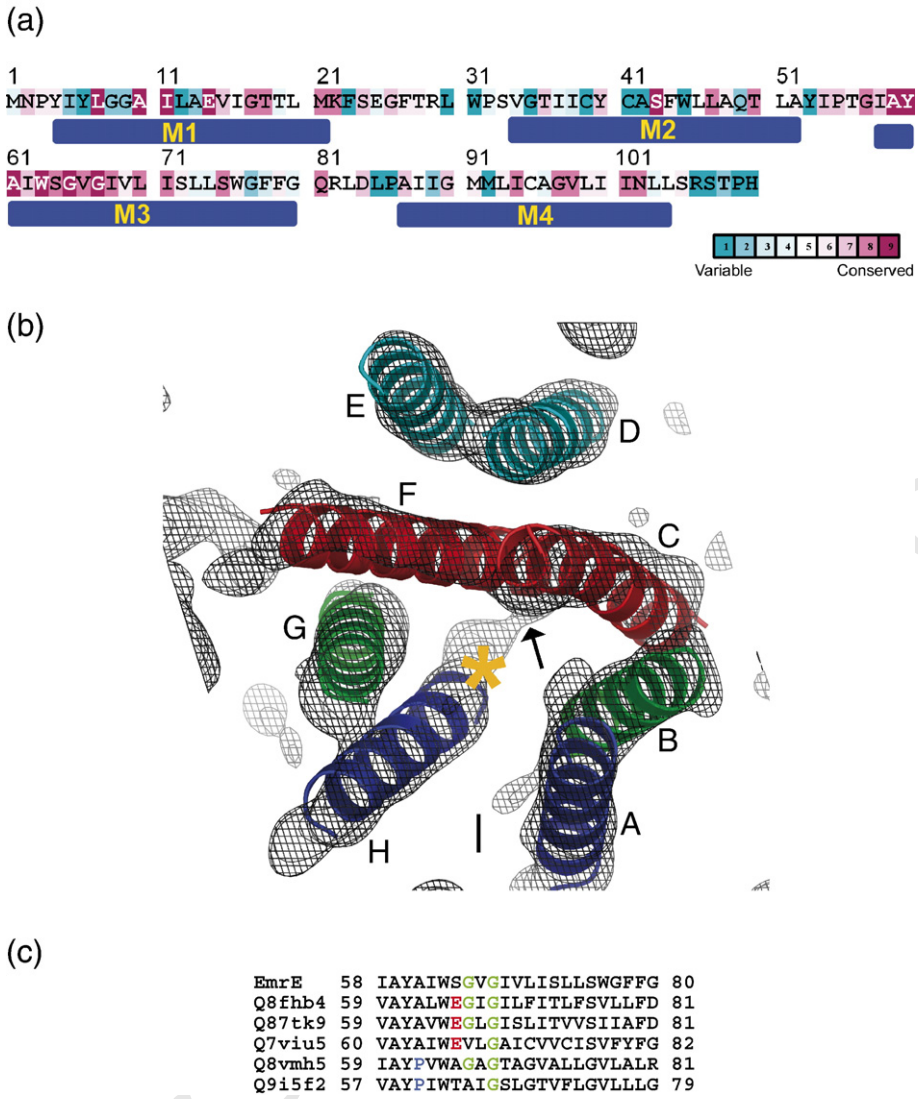

Figure 1. (a) Evolutionary conservation of amino acid residues in EmrE. Sequence conservation was color-coded using the *ConSeq* webserver,[55] and the predicted hydrophobic segments are marked M1–M4. Note that segment M3 is completely conserved in its N terminus with helically periodic variability emerging only in its C terminus. (b) Positions and tilt angles of α-helices inferred from the cryo-EM structure of EmrE[7] viewed perpendicular to the membrane plane. The helices are marked A–H following the notation used by Ubarretxena-Belandia *et al.*[7] The gray mesh indicates electron density at 1.1σ. The arrow marks the position where helices F and H are connected *via* what could be a rigid loop. The orange star marks the approximate in-plane position of the center of the TPP^+ molecule. Notice that helices A-B-C are related to helices H-G-F, respectively, by an approximate 2-fold rotation around the in-plane axis marked by a continuous line. Symmetry-related helices are denoted by the use of the same color. (c) Multiple-sequence alignment of selected SMR sequences in the M3 region. The N terminus of M3 contains the sequence signatures of backbone flexibility, such as fully and highly conserved glycine residues in positions 67 and 64, respectively (green). Some sequences have proline (blue) in positions aligned with Ala61 from EmrE, and others have glutamate aligned with Ser64 of EmrE (red). These polar, small, and helix-deforming residues could elicit flexibility in the M3 segment, correlating with a kink observed in helix C (Figure 1(a)). The complete alignment of SMR homologues is available at http://www.ashtoret.tau.ac.il/~sarel/EmrE.html. Figures 1(b), 3, and 4(b) were generated with PyMol [http://pymol.sourceforge.net/].

(Figure 1(b)). These helices are unique in EmrE, because one half of each helix is buried on all sides by other helices and the other half is exposed to lipid on only one of its faces; of the four hydrophobic segments, only M3 contains conserved amino acid residues in this identical pattern. The N terminus of M3 is highly conserved, implying it is packed on all sides by other helices, but its C terminus shows a helical periodic pattern of variable residues, suggesting that one face is lipid-exposed (Figure 1(a)). In Figure 1(b), the most lipid-exposed, C-terminal portion of M3 is represented by the right-hand (distant) end of C and the left-hand (near) end of F.

The assignment of M3 to C and F is supported partly by the observation that M3 is predicted by our analysis to be the longest hydrophobic stretch (23 residues compared to 18 or 19 for the other TM segments, Figure 1(a)) paralleling its assignment to the most tilted helices in the structure. Further

Table 1. Summary of experimental data gathered on residues of helices M1 to M4 of EmrE, and locations of those residues in the model structure

Residue	Environment predicted from model	Activity data ++=wt	Ref	Environment of label	Ref
M1					
Tyr4	Substrate chamber	−	2,4,5	Accessible	4
Ile5	Lipid facing	++	2,4	Accessible	4
				Lipid-facing	2
Tyr6	Lipid facing	++	2,3,4,5	Partly accessible	4
Leu7	Chamber	−	2,3,4	Accessible	4
				Water-exposed	2
Gly8	Lipid facing	++	2,4	Inaccessible	4
				Lipid-facing	2
Gly9	Lipid facing	++	2,4	Inaccessible	4
Ala10	Interhelix contact	−	2,3,4	Accessible	4
				Water-exposed	2
Ile11	Chamber	++	2	Partly accessible	4
		−	1,3,4	Water-exposed	2
Leu12	Lipid facing	++	1,2,3,4	Partly accessible	4
				lipid-facing	2
Ala13	Lipid facing	++	1,2,3,4	Inaccessible	4
Glu14	Chamber	−	2,3,4,6	Accessible*	4
				Proximal to E14	2
Val15	Interhelix contact	++	2,3,4	Inaccessible	4
Ile16	Lipid facing	++	2,3,4	Inaccessible	4
				Lipid-facing	2
Gly17	Interhelix contact	+	1,2	Partly accessible	4
		−	3,4	Water-exposed	2
Thr18	Chamber	−	1,2,3,4	Accessible	4
				Proximal to T18	2
Thr19	Interhelix contact	++	2,4	Inaccessible	4
				Constrained	2
Leu20	Interhelix contact	++	2,4	Inaccessible	4
				Constrained	2
Met21	Interhelix contact	++	2,3,4	Inaccessible	4
				Constrained	2
M2					
Val34	Substrate chamber	+	1		
Gly35	Lipid facing	++	1		
Thr36	Interhelix contact	+	1		
Ile37	Chamber	++	1		
Ile38	Lipid facing	++	1		
Cys39	Interhelix contact	++	1		
Tyr40	Chamber	−	1,5	Proximal to substrate	4
Cys41	Chamber	+	1		
Ala42	Lipid facing	+	1		
Ser43	Interhelix contact	++	1		
Phe44	Chamber	−	1		
Trp45	Chamber	++	1,7		
Leu46	Interhelix contact	++	1		
Leu47	Chamber	+	1		
Ala48	Chamber	−	1		
Gln49	Lipid facing	+	1		
Thr50	Interhelix contact	++	1		
Leu51	Chamber	++	1		
Ala52	Chamber	−	1		
M3					
Ile58	Interhelix contact				
Ala59	Interhelix contact				
Tyr60	Binding chamber	+	5	Proximal to substrate	5
Ala61	Interhelix contact				
Ile62	Lipid facing				
Trp63	Chamber	−	7	Proximal to substrate	7
Ser64	Interhelix contact				
Gly65	Interhelix contact				
Val66	Interhelix contact				
Gly67	Interhelix contact				
Ile68	Interhelix contact				
Val69	Lipid facing				
Leu70	Interhelix contact				
Ile71	Chamber				
Ser72	Interhelix contact	+	1		
Leu73	Lipid facing	++	1		
Leu74	Interhelix contact	+	1		
Ser75	Interhelix contact	+	1		
Trp76	Lipid facing	++	7		
Gly77	Lipid facing				
Phe78	Chamber				
Phe79	Lipid facing				
Gly80	Lipid facing				
M4					
Ala87	Interhelix contact				
Ile88	Lipid facing				
Ile89	Lipid facing				
Gly90	Interhelix contact				
Met91	Interhelix contact				
Met92	Lipid facing				
Leu93	Interhelix contact	−	1		
Ile94	Interhelix contact	++	1		
Cys95	Lipid facing	++	1		
Ala96	Lipid facing	++	1		
Gly97	Interhelix contact	++	1		
Val98	Interhelix contact				
Leu99	Lipid facing				
Ile100	Interhelix contact				

(continued on next page)

Table 1 (*continued*)

Residue	Environment predicted from model	Activity data ++=wt	Ref	Environment of label	Ref
M4					
Ile101	Interhelix contact				
Asn102	Interhelix contact				
Leu103	Lipid facing				
Leu104	Interhelix contact				

Notice that 22 positions have been probed, but so far, have not been implicated in protein function (++ in the Activity data column); 11 of these are lipid-facing in the model structure. Moreover, 23 positions have been implicated in protein function (− and + in the Activity data column); 21 of these have straightforward structural explanations, with the positions either lining the translocation chamber or situated at helix interaction sites.

Reference 1 (Mordoch *et al.*[45]): Cys-scanning mutagenesis was performed on an active Cys-less mutant. Activity was assessed by performing transport assays for three different substrates. Reference 2 (Koteiche *et al.*[20]): The environment of a spin-label on the Cys mutant is described as either water-exposed, lipid facing or proximal to the corresponding residue in the dimer. Although all residues in M1 were tested, assignments are given only to residues that are unambiguous, with other residues presumably at environmental boundaries. Reference 3 (Gutman *et al.*[16]): Cys mutants were assayed for binding of TPP^+. Reference 4 (Sharoni *et al.*[18]): Accessibility refers to the ability of alkylating agents to react with a Cys mutation at the position indicated and an asterisk (*) shows that the experiment was performed on a heterodimer to ensure proper folding of EmrE. Reference 5: Rotem *et al*. tested the effects of mutations of tyrosine residues to cysteine on function and changes in fluorescence quenching in response to ligand binding. Reference 6: Several studies showed that Glu14 is a critical residue for substrate binding and translocation.[14–19] Reference 7: Elbaz *et al*. tested the effects of mutations of tryptophan residues to cysteine on function and changes in fluorescence quenching in response to ligand binding.[59]

support is provided by the observation that the M3 N terminus contains several sequence signatures that would favor flexibility of the helix backbone, correlating with a kink in helix C observed in the cryo-EM structure (Figure 1(b)). These sequence signatures include (Figure 1(c)): (a) the presence of two highly conserved glycine residues in positions 65 and 67; (b) the observed substitution of position Ser64 with glutamate residues in other SMR members; and (c) the fact that position Ala61 is substituted by proline in several homologues. Notably, proline residues in multiple-sequence alignments of TM domains have been shown to be indicators for kinks, even in cases where the sequence of the protein, for which a structure is available, does not exhibit a proline.[37] Although these sequence features would favor flexibility of the helix backbone, the segment does not necessarily exhibit a kink and, thus, helix F is seen to be mostly straight in the cryo-EM structure (Figure 1(b)).[7]

Given the assignment of M3 to helices C and F, and the experimental constraints listed above, there is only one solution for the assignment of the remaining helices. As the termini of helices F and H are apparently connected by density on the side of the structure away from the viewer in Figure 1(b) (indicated by an arrow), and based on the assignment of the portion of F near the connection to H to be the N terminus of M3, then helix H must be M2. Since the M2–M3 interconnecting loop is predicted to contain only five amino acid residues (Figure 1(a)), it might be rigid and could, therefore, be visible in the cryo-EM structure. If M2 is helix H, then, by symmetry, helix A is also M2. Multiple sources of biochemical data have implicated residues on M1 as crucial for substrate binding and translocation;[14–19] given the assignment of M2 and M3 above, M1 must occupy the symmetry-related B and G helices around the translocation chamber. In contrast, amino acid residues in M4 are not involved directly in substrate binding or translocation (Table 1). The lack of data implicating residues on M4 in substrate binding is in agreement with the location of helices D and E, separated from the substrate-binding chamber by helices C and F. Finally, the helix assignment suggested here (M1=B,G; M2=A,H, M3=C,F, and M4=D,E) is consistent with constraints imposed by the short interconnecting loops observed in the EmrE sequence (Figure 1(a)) on the distances between the helix ends seen in the cryo-EM structure.[38] In addition to this most likely helix assignment, we tested each of the 47 other permutations against the known functional data on EmrE, the interconnecting loop lengths, and SMR evolutionary conservation. None of these other permutations fit the aggregate data on EmrE nearly as well as the suggested assignment (data not shown).

We note that domain swapping, where helices from one monomer interpenetrate between helices in the other monomer, could confound the proposed helix assignment. However, this possibility would connect helices that are distant from one another, and is therefore made unlikely by the very short lengths of the interconnecting loops (Figure 1(a)).[7] The only loop that would allow domain swapping is between M1 and M2. However, the swapping of these domains involves conformations in which the loop blocks substrate entry to the binding chamber.

Structural modeling

Canonical α-helices were constructed to fit the helix axes[39] extracted from the cryo-EM structure.[7] For each helix, all the rotations around its principal axis were sampled in 5° increments; each conformation was scored according to a rule that favors situations in which evolutionarily conserved amino acid positions were packed inside the protein core, with variable positions facing the lipid.[23] Following the orientation of each of the helices, we introduced a kink into helix C to account for the deviation from α-helical regularity observed for this helix in the cryo-EM structure (Figure 1(b)).[7] At the vertical resolution of the cryo-EM structure (16 Å), the position of the kink cannot be determined unambiguously. We therefore estimated this position on the basis of the direction of the kink observed in the cryo-EM structure and features observed in SMR

sequences of M3 that imply backbone flexibility at the N terminus of this helix (Figure 1(c)), and placed the kink so that it affects mainly the backbone hydrogen bond between positions Ser64 and Ile68. We note that this approximate location for the kink within helix C does not affect the conclusions we draw below on the support that biochemical and biophysical data provide to the model structure.

The computed conformation fits the conservation profile of each helix quite closely (Figure 2) with all of the variable residues facing the lipid, and the conserved residues facing the protein core. It is important to note that no experimentally derived information was used to constrain the orientations of the helices. As shown below, these orientations nevertheless provide a structural framework for understanding most of the biochemical data on EmrE.

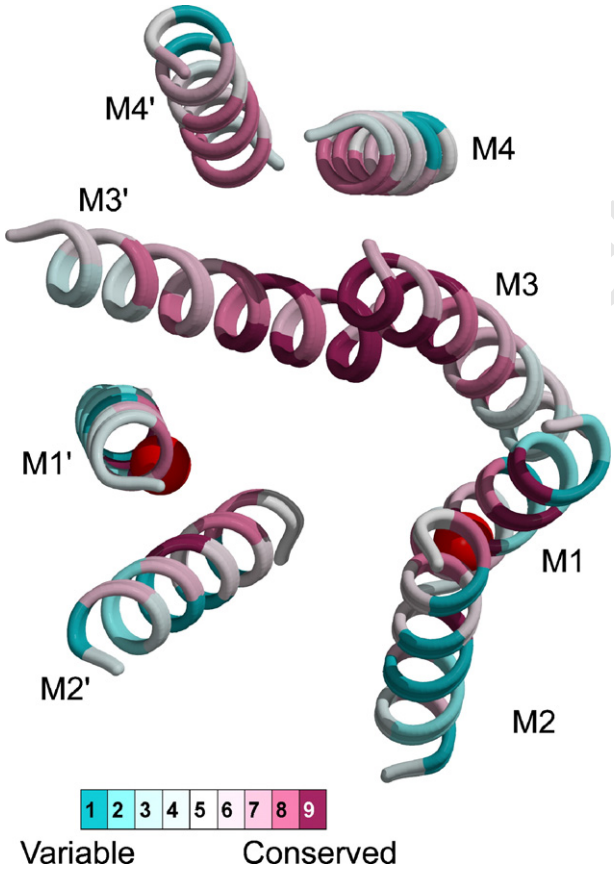

Figure 2. A view of the EmrE model perpendicular to the membrane bilayer color-coded according to evolutionary conservation. On all helices, the conservation signal closely matches the pattern of exposure of residues to lipid, with conserved residues buried at interhelix contact regions, and variable residues placed in membrane-exposed positions. The Glu14 residues on both monomers are shown as red spheres. The two monomers are distinguished by the presence or the absence of an apostrophe. The Figure was generated with MOLSCRIPT[60] and rendered with Raster3D.[61]

Comparison of the EmrE model with data from biochemical and biophysical experiments

It is difficult to interpret the pertinent biochemical and biophysical data on EmrE on the basis of the model structure, because the model does not contain side-chains. Furthermore, we estimate that the orientations of the individual α-helices around their principal axes might vary by up to 20°, and that the positions of C^{α} atoms on the terminal turns of each of the helices might diverge from the positions specified in our model.[23] Even with this level of uncertainty, however, it is possible to provide a rough account of the majority of the experimental data on the basis of the model structure.

The structure of EmrE has been probed using a number of biophysical and biochemical techniques. The model presented here does not seriously conflict with any of these data and, in fact, can be used to rationalize and simplify a number of observations. The experiments discussed in this section used spin labels to probe the environment of helix M1,[20] and site-directed mutagenesis to define amino acid residues important for folding and transport activity.

The TM region M1 of EmrE contains Glu14, which is essential for substrate binding and translocation.[14–16,18,19] Therefore, this region has been studied intensively using a number of biophysical and biochemical approaches. Site-directed spin-labeling experiments were applied to all the residues of M1 to infer which of them are packed against other helices, exposed to lipid, or are in the vicinity of M1 residues of the neighboring monomer.[20] All of the residues that were identified as lipid-exposed by the spin-labeling experiments are predicted to be lipid facing in our EmrE model (dark blue spheres in Figure 3(a) and Table 1). Interestingly, lipid-exposed positions were identified by spin-labeling to be restricted mainly to the N-terminal part of M1; side-chains in its C terminus were found to be motionally more restricted (light blue spheres in Figure 3(a)). These results are in close agreement with the model structure, in which the N-terminal part is more exposed to lipid and the C terminus is packed against other helices from almost all directions. Thus, the spin-labeling data[20] verify the assignment of M1 to helices B and G as well as the helices' orientations around their principal axes. The spin-labeling experiments also identified only two residues on M1 (Glu14 and Thr18) that are vicinal to their counterparts on the other monomer. Indeed, these two residues face one another according to the model structure, and these two pairs have the closest C^{α}–C^{α} distances of all residues on M1 in the model (~16 Å). The model proposed by Koteiche *et al.* for the relative orientations of the two M1s considered only parallel helix packing;[20] however, their results fit equally well with our antiparallel model shown here.

Hsmr is a homologue of EmrE from the archaeon *Halobacterium salinarium*, which is unique among SMR members, in that approximately 40% of its sequence is comprised of Ala and Val residues.[40]

Presumably, this composition reflects an evolutionary pressure to increase the G+C content of the genome, while maintaining the relatively high hydrophobicity necessary for a TM protein. Positions that are not Ala or Val in Hsmr are, therefore, considered important for structure or function; conversely, positions that are Ala or Val in Hsmr, but not Ala or Val in EmrE, can be presumed to be unimportant.[40,41] As expected, the vast majority of these positions are lipid exposed according to the model structure (Figure 3(b)). This observation suggests that, despite the low level of sequence identity between SMR proteins, the overall fold of the homologous proteins is conserved.

The cryo-EM structure of EmrE was derived from crystals of the transporter bound to TPP^+. The position of TPP^+ in the plane of the membrane is clear from the 3D structure and from comparisons of projection maps of EmrE with and without TPP^+.[8] However, the position of TPP^+ along the axis perpendicular to the membrane plane is less certain due to the low resolution along this axis.[7] To provide

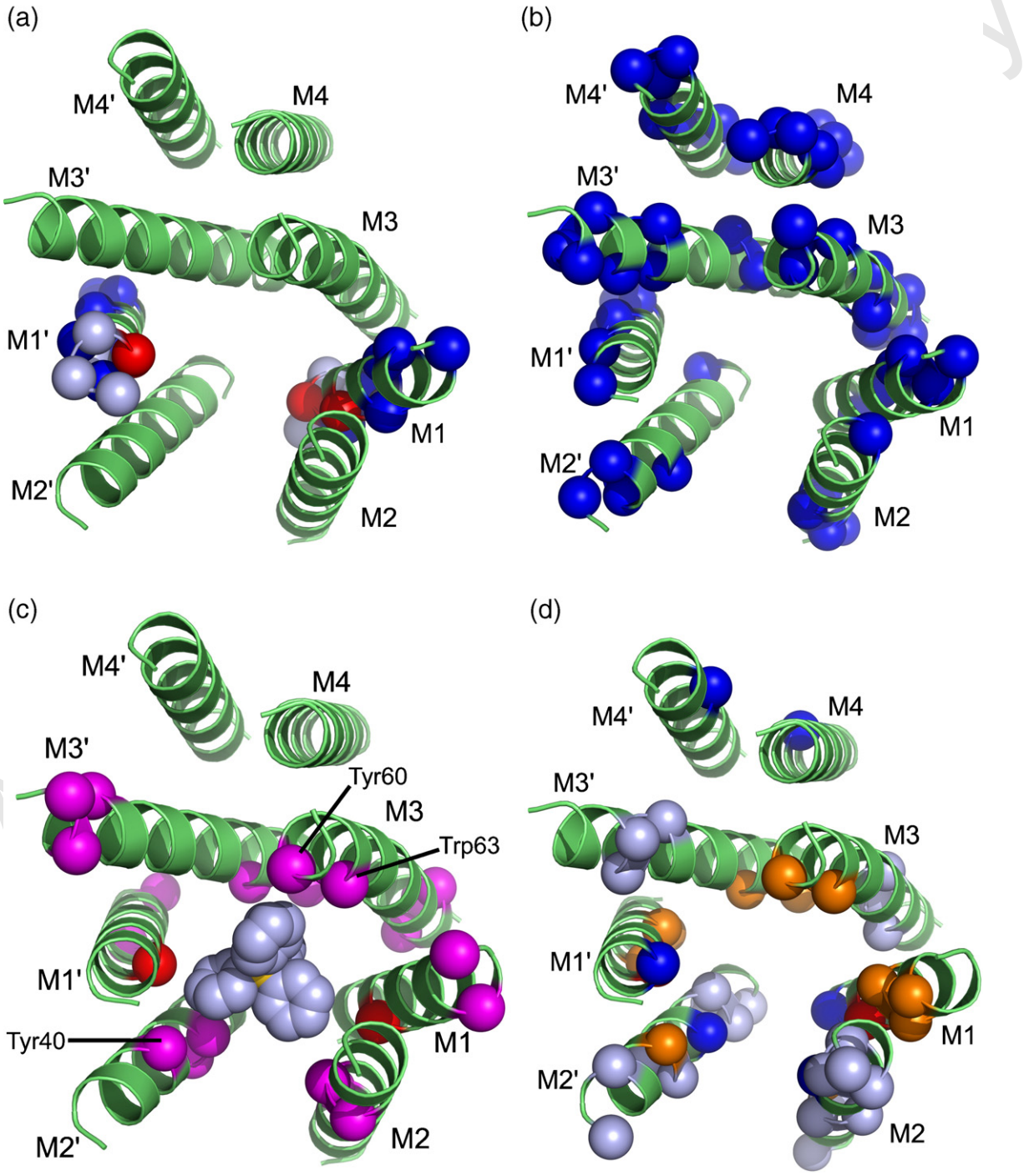

Figure 3 (*legend on opposite page*)

rough constraints for which residues are located around the substrate, we docked TPP^+ manually (Figure 3(c)) based on the constraint that the position of TPP^+ with respect to Glu14 should roughly match that seen in the atomic-resolution structure of the water-soluble multidrug receptor BmrR,[42] which was also crystallized in a TPP^+-bound form. In harmony with various experimental assays, the Glu14 residues from both monomers are in position to form contact with the substrate.[17,19] It is also notable that several aromatic residues are found in the vicinity of the modeled TPP^+, providing partners for aromatic interactions with substrate. In particular, positions Tyr40, Tyr60, and Trp63 were shown experimentally to bind substrate,[18] and their C^α atoms are located within 6 Å from carbon atoms of the modeled TPP^+ (Figure 3(c)). Other aromatic residues are located between α-helices, where they might increase structural stability. Table 1 lists 52 amino acid residues that line the translocation chamber or mediate interhelix contacts in our model; 21 of these residues have been mutated and implicated experimentally in substrate binding and translocation. It should be noted, however, that it is likely that the specific residues mediating EmrE binding to substrates other than TPP^+ might vary from those specified here, in analogy to the differences observed in the binding modes of different substrates to bacterial multidrug gene regulators,[43] and bacterial multidrug resistance transporters.[44]

Mordoch *et al.* conducted an extensive substituted cysteine-accessibility method analysis of Cys-less EmrE[45] (Table 1). They replaced 48 positions throughout the protein with cysteine, and tested the mutant transport properties. Only five positions were absolutely sensitive to replacement, in that the mutants were incapable of conferring resistance to known EmrE substrates. Of these five, two mutations (Ile11Cys and Thr18Cys) led to good expression of EmrE, but transport was reduced severely; these positions are on the same face of M1 as the essential Glu14 and are probably involved in substrate transport (Figure 1(d)). The other three mutations (Tyr40Cys, Phe44Cys, and Leu93Cys) resulted in no expression of EmrE, so it may be that these residues are important in the folding or stability of EmrE in the membrane. Subsequent mutational analysis of Tyr40 showed that this residue is also important in substrate recognition.[18] Both Phe44 and Leu93 are predicted to be at the interfaces between helices (Figure 3(d)), which may explain their effect on protein folding and/or stability. This substituted cysteine-accessibility method study also identified ten residues in the TM regions that showed decreased resistance to only one of the antiporter cognate substrates.[45] Eight out of these ten residues are positioned at or around the substrate-binding chamber (Table 1 and light blue spheres in Figure 3(d)). Presumably, substituting these positions alters the properties of the protein surface that lines the chamber, hence reducing substrate affinity. As Mordoch *et al.* note, the residues on M2 are distributed on two helical faces, and indeed the two sensitive positions (Ala42 and Gln49), which the model does not place at or around the substrate-binding chamber, are located on this helix. Notably, mutants of these two residues are sensitive only to acrylflavine among the cognate substrates that were tested,[45] implying that these residues might be involved in the binding of only certain substrates.[43,44]

There are two reports of experimental data that conflict with the model we present here, because they both suggest that the monomers within EmrE have an identical orientation in the membrane with the N and C termini in the cytoplasm.[32,46] A third study reported single topology for the QacC homologue of EmrE, although the data were inconclusive regarding the localization of the C terminus.[47] These results are, however, contrary to

Figure 3. Structural interpretation of biochemical and phylogenetic data on EmrE and its homologues. (a) Spin-labeling experiments identified lipid-exposed residues (dark blue), and motionally restricted (light blue) positions.[20] Red spheres identify Glu14 and Thr18, which were shown to be close to their counterparts on the other monomer, as indeed they are in the model. Notice that in both monomers, the motionally restricted residues on the N-terminal turn of M1 are surrounded by other helices from almost all sides, whereas positions that were identified experimentally as lipid-exposed are indeed located in lipid-facing parts of the protein. (b) Green spheres mark positions on EmrE that are aligned with Ala or Val in the Hsmr homologue from *H. salinarium*, but are not Ala and Val in EmrE. Such positions are thought to have little structural or functional importance,[40] and indeed the majority face the lipid environment. (c) Docking of a molecule of TPP^+ in the EmrE model structure. TPP^+ was docked manually such that it approximately fits the orientation seen in a crystal structure of the cytoplasmic receptor for TPP^+ BmrR.[42] The two Glu14 residues (red spheres) are in proximity to aromatic rings from the substrate TPP+. Aromatic residues in the TM domain of EmrE are marked by purple spheres. Some of these residues surround TPP^+, thus providing possible interaction partners for the substrate. Positions Tyr40, Tyr60, and Trp63, which are marked on the Figure, have been implicated directly in substrate binding.[18] Others are placed in spacious regions of the structure, where they might serve to enhance the interactions between helices (e.g. the M2/M2′, M1/M3, and M1′/M3′ interfaces). The substrate TPP+ molecule is shown in space-filling spheres, with light blue corresponding to carbon and yellow to phosphate atoms. (d) Blue spheres indicate four positions where mutations to Cys abolish functionality, and green spheres indicate positions that change resistance to only some of the transporter's cognate substrates.[45] Orange spheres mark positions that are involved in substrate binding.[18,19] All of the blue spheres map to positions at the interfaces between the helices, where mutations might disrupt protein folding or oligomerization, or around the binding chamber. Most of the light blue spheres map to positions around the translocation chamber at least in one of the monomers, where changes to the surface of the protein might modify substrate recognition. The orange spheres are all located around the Glu14 residue. A listing of all residues in the TM domain and their experimentally determined structural or functional roles is provided in Table 1.

the topology analysis reported by Daley *et al.*,[30] who showed that the predominant orientation of EmrE has the N and C termini in the periplasm. The cross-linking data identifying helix–helix interactions[46] are difficult to reconcile with our model, and will require an atomic-resolution structure to be determined before the conflict can be resolved, as was the case for the cross-linking data for the lactose permease.[48] The most internally consistent cross-linking data showing that helix M4 lies parallel with and adjacent to M4 from the neighboring monomer could be explained easily by suggesting that EmrE is a tetramer in the membrane, related by a 2-fold perpendicular to the membrane,[4] a proposal that is supported by recent data from studying the interaction between peptides representing individual TM regions of Hsmr, the archaeal homologue of EmrE.[49]

An alternate-access mechanism for substrate translocation

Transport of drug substrates from the cytoplasm or cytoplasmic leaflet of the lipid bilayer out of the bacterium is thought to occur in essentially two steps[3,4] (Figure 4(a)). First, the drug substrate binds

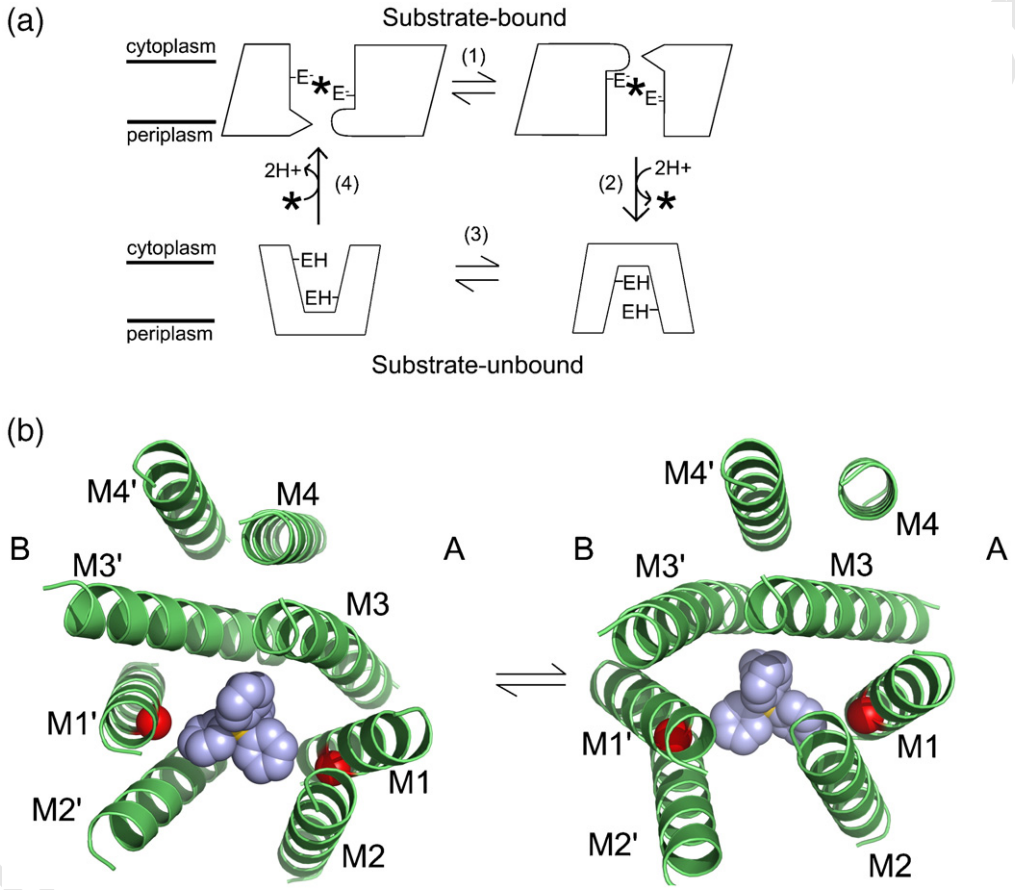

Figure 4. (A) A mechanism for proton-coupled translocation of substrates by the SMR family of proteins. (1) Two substrate-bound forms of the protein interconvert between conformations, in which the substrate, marked by an asterisk (*) faces the cytoplasm or the periplasm due to conformational changes. (2) In the periplasmic-facing conformation, the substrate is supplanted by the binding of two protons to the Glu14 positions (marked by E^-) on both monomers, thus driving the equilibrium towards substrate translocation. (3) A conformational change reorients the binding site towards the cytoplasm. (4) Substrate binding on the cytoplasmic side forces the protons out of the translocation chamber into the cytoplasm. (b) A suggestion for the conformational change represented by Step 1 in Figure 4(a). Periplasmic-facing and cytoplasmic-facing conformations of the EmrE dimer based on the cryo-EM structure. The transition between the two conformations involves a reorientation of the M1–M3 helices in both monomers by approximately 20° with respect to the in-plane axis of symmetry; a kinking and straightening of M3; and a small translation of M1–M3 in both monomers with respect to the M4 helices. As these changes occur in the protein dimer, the TPP^+ substrate, which is accessible from the near end in the conformation on the left, moves downwards and becomes accessible from the far end of the EmrE dimer in the right-hand conformation. Thus, interconversion between these two conformations could alter the accessibility of substrate from cytoplasmic facing to periplasmic facing. The conformation on the right was obtained by rotating the conformation on the left by 180° with respect to the in-plane axis of quasi-symmetry, so that the two conformations are completely superimposable. Thus, inverted topology would reproduce a single substrate-binding mode as two conformations, one of which is accessible to the periplasm and the other to the cytoplasm. The two monomers are arbitrarily marked A and B.

to EmrE, which induces a conformational change, so that the inward-facing binding pocket is opened to the periplasm and closed to the cytoplasm. The high concentration of protons in the periplasm competes directly with the drug for binding at the two Glu14 residues, so protonation results in release of the drug. A further conformation change then re-orients the protein to face the cytoplasm, where it may bind another drug molecule.

The nature of these conformation changes is uncertain, but the cryo-EM structure and our model suggest its basic features. To explain the substrate-translocation process, we propose that M1-M4 from monomer A adopt the conformation of M1-M4 observed in monomer B in our model and *vice versa* (Figure 4(b)) during the step marked as (1) in Figure 4(a). Due to the in-plane symmetry, this transition results in a structure identical with the original model rotated by 180° with respect to the in-plane axis of symmetry; the two symmetry-related structures are shown on both sides of the chemical equilibrium in Figure 4(b). To analyze the details of this transition, it is useful to divide the model structure into three subunits: (1) M1, M2 and M3 (monomer A); (2) M1′, M2′ and M3′ (monomer B); and (3) M4 and M4′.

Helices M1–M3 from monomer A are virtually superimposable on M1′–M3′ of monomer B, except for the kink in M3, suggesting that M1–M3 move as one unit during the transition described in Figure 4(b). The two M4 helices are seen to make minimal movements with respect to one another during the transition, suggesting that they are stable as a helix pair. Indeed, the cryo-EM and model structures show this pair to be closely packed with glycine residues (positions 90 and 97) lining the interhelix interface, which can stabilize helix packing.[50] By contrast, the interfaces between helices M3 and M4 are small in both monomers in comparison to any of the other pairs of interacting helices in the structure (Figure 4(b)). It therefore comes as no surprise that the most significant conformational change that occurs during the transition can be localized to the contact region between M3 and M4, with the crossing angles between these helices changing by approximately 20° around the in-plane axis of symmetry in order to switch the M3-M4 packing from that observed in monomer A to that observed in monomer B, and *vice versa*. The kinking and straightening of the two M3 helices and a small translation of the M1-M3 helices in both monomers with respect to the M4 helices, coupled to the movement of the TPP^+ molecule perpendicular to the membrane plane would then complete the transition. Thus, although residues on M4 have so far not been recognized as important for substrate binding and translocation (Table 1), this putative mechanism suggests a crucial role for M4 in stabilizing the dimer interface during the translocation process. The short M3-M4 loop, consisting of six residues in the SMR family, would hold the two parts of the structure together in the face of these relative motions. It is interesting to note in this connection that recent results have suggested a role for the M4 helix in mediating the formation of SMR tetramers.[49]

The sum of these conformational changes would alternately open the substrate-translocation chamber to the cytoplasmic and periplasmic media, allowing substrate to bind in the cytoplasmic-facing conformation, and then to be replaced by protons when the protein faces the periplasm (Figure 4(a)). Interestingly, this mechanism suggests that the periplasmic- and cytoplasmic-facing conformations of substrate-bound EmrE are essentially identical, and would thus require a single substrate-binding mode to be optimized structurally, which would then by symmetry be reproduced in both cytoplasmic-facing and periplasmic-facing conformations. Currently, there are only two conformations of EmrE (Figure 4(a), the upper panels) for which we have structural information (Figure 4(b)),[7,8] and the structure of further transport intermediates (i.e. Figure 4(a), lower panel) will be essential to identify conformational changes that occur during the transport cycle. However, the availability of our model will now allow the design of specific experiments, such as using site-specific spin labels to monitor movements in EmrE during the transport cycle.

Discussion

The suggestion of dual topology of EmrE,[7] and the recent support for this from global-topology analyses,[30,31] were the key for the successful modeling of EmrE presented here. The presence of 2-fold quasi-symmetry between the monomers of the EmrE dimer within the plane of the membrane (Figure 1(b)) implied an antiparallel orientation of the EmrE monomers. Our previous modeling attempts (not described), which were not guided by the in-plane pseudo 2-fold axis, were unsuccessful in providing explanations for the biochemical and biophysical observations on EmrE. In contrast, our model with the monomers in an antiparallel orientation explains virtually all the biochemical and biophysical data. The model makes many predictions about the structure of EmrE that will provide a platform for further experimental work, such as the identification of other residues in the translocation pathway that have not yet been studied (Table 1), and residues that may be important in mediating helix packing, and therefore could be involved in the conformational changes.

The suggestion of oppositely oriented monomers in EmrE has been made only recently,[7] and is reinforced by global topology analysis of bacterial proteins,[30,31] but, so far, mechanistic advantages of dual topology have not been proposed. The mechanism of translocation that we invoked above suggests a potential advantage. That is, if the cytoplasmic-facing and periplasmic-facing conformations of substrate-bound EmrE are essentially identical, then only one mode of substrate binding should be

devised by evolution, which would be replicated as two conformations, one facing the cytoplasm and another facing the periplasm. This might also provide partial solution to a long-standing puzzle in SMR research; namely, how these small proteins consisting of roughly 100 amino acid residues can catalyze the coupled translocation of substrate and protons,[3] a feat that is accomplished in other antiporter families, such as the major-facilitator family, by much larger proteins.[51] Thus, inverted topology might be a parsimonious evolutionary solution to the problem of vectorial transport. In this connection, it is interesting to note that two of the five proteins identified as having the dual-topology architecture are from the SMR family (the others have not been fully characterized mechanistically), and a sixth case of dual topology was identified involving two homologous proteins (*YdgE* and *YdgF*), which are also SMR members that are likely to have arisen from a gene-duplication event.[30]

Although much of the biochemical and biophysical data gathered on EmrE are in harmony with the model structure, there are one or two pieces of data that are not in agreement. The topology of the protein is clearly the most important point of disagreement, because dual topology provided the basis for the model structure reported here and for the suggested mechanism of substrate translocation; ultimately, if inverted topology for EmrE is incorrect, then so is the model. We have found that dual topology provides the most satisfactory model for EmrE, but Ninio *et al*.[32] predict, on the basis of labeling data, that the monomers have identical topology, conflicting with other lines of experimental data that suggest inverted topology for EmrE;[30,31] the reasons for this discrepancy among different lines of experimental data are unclear. The possible conflicts of our EmrE model with the cross-linking data[46] have been discussed above. The difficulties inherent in the structural interpretations of cross-linking data on dynamic structures are well known, because even rarely sampled conformations might elicit crosslinks, as was underscored recently in the case of lactose permease.[48]

Despite many years of structural studies of the SMR transporter EmrE, an atomic-resolution structure of this representative protein that can explain much of the biochemical and structural data has not emerged. Here, we have used phylogenetic analysis combined with constraints obtained from a cryo-EM structure of EmrE and some biochemical experiments in order to produce a model structure specifying the approximate positions of individual amino acid residues for EmrE and its homologues. Although this model was constrained only by some biophysical data on EmrE, it is encouraging that the model is capable of accounting for so much of the biochemistry. By revealing the locations of individual amino acid residues in the membrane-spanning regions, the model can be used in order to plan and interpret experiments aimed at deciphering the molecular details of the substrate-translocation mechanism in EmrE and its homologues.

Methods

Sequence data

An initial alignment of a few tens of EmrE homologues was constructed using CLUSTAL W.[52] On the basis of this alignment, we then constructed a hidden Markov model (HMM),[53] which was then calibrated and used to search SWISSPROT and TrEMBL[54] for additional sequence homologues. Sequences showing over 90% identity with other sequences in the set were removed to obtain 98 sequences, which were then aligned†.[52] Conservation scores were then computed for each amino acid position using the ConSeq server and the Rate4Site algorithm.[55,56] The sequence alignment was inspected to identify hydrophobic stretches that correspond to the hydrophobic cores of the helices in forming the TM domain. Starting from the secondary structure assignment derived from NMR,[6] we manually modified the N and C termini of each hydrophobic domain so that the longest stretches of hydrophobic residues would be aligned. The following segments of EmrE were used as the hydrophobic stretches: TM1, 4–21; TM2, 34–52; TM3, 58–80; TM4, 87–104. The conservation scores and the hydrophobic segments are shown in Figure 1(b).

Conformation scoring function

The method for conformational search was as described.[23] In brief, this scoring function favors the burial of evolutionarily conserved amino acid positions in the protein core and the exposure of variable positions to the lipid, without biasing helix orientations according to experimentally derived data. Conformations that expose charged amino acids to the lipid milieu are penalized (in EmrE, this applies only to M1 due to position Glu14). The following scoring function is used to score each conformation:

$$Score = \sum_i (2(B^i - 1/2)(H^i - C^i)) \quad (1)$$

where B^i quantifies the extent of burial of amino acid residue i in the protein core.[39,57] It assumes values of 0 to 1, with 1 signifying complete burial against another helix, and 0 signifying complete exposure to the lipid or the pore lumen. The function is computed by iterating over all of the helices in the structure other than the one on which i is located, and taking into account distance from, and orientation of i with respect to each of these helices. B^i is then taken as the maximum of the values calculated for each of the helices.[23,39] Thus, high values of B^i imply that i is in close contact with another helix, whereas low values indicate that it is not interacting with any of the helices.

The C^i values are the normalized evolutionary-rate scores assigned by *Rate4Site*.[55,56] High-through-low values of C^i are assigned to variable-through-conserved positions, respectively. H^i is the free energy of transfer from water to lipid of amino acid i according to the Kessel and Ben-Tal scale.[58] H^i values are taken into account only if they are greater than 7 kcal/mol, and only for residues i that are exposed to the membrane, i.e. for which the burial scores B^i are less than 0.5. Thus, the hydrophobicity scale

† The multiple-sequence alignment of SMR proteins can be downloaded from http://ashtoret.tau.ac.il/~sarel/EmrE.html

serves as a significant penalty on the exposure of the most polar residues to the membrane environment.

Conformational search

Canonical C^{α}-trace models of eight α-helices were constructed according to the helix axes parameters derived from helical models that were made to fit the cryo-EM structure, and their geometric centers were placed at the hypothetical membrane midplane. The amino acid identities of positions in the hydrophobic segments M1–M4 were assigned to the relevant positions on these helices.

Each helix was rotated around its principal axis independently, in 5° steps, and its optimal orientation was derived. Then, the optimal orientations of all helices were superimposed to yield the optimal conformation of the entire complex.

Data Base accession number

The cryo-EM structure is available from the EM data bank with accession code 1087‡. The coordinates of the model structure of a dimer of EmrE containing backbone atoms has been deposited in the PDB with accession number 2i68.

Acknowledgements

The authors gratefully acknowledge many comments and discussions with S. Schuldiner regarding biochemical and biophysical data, and their implications, for providing unpublished data. This study was supported by a grant 222/04 from the Israel Science Foundation to N.B.-T. S.J.F. was supported by a doctoral fellowship from the Clore Israel Foundation. Work by A.E. and D.H. was supported in part by the IST Programme of the EU as shared-cost RTD (FET Open) Project under Contract No IST-006413 (ACS - Algorithms for Complex Shapes), and by the Hermann Minkowski-Minerva Center for Geometry at Tel-Aviv University.

References

1. Nikaido, H. (1994). Prevention of drug access to bacterial targets: permeability barriers and active efflux. *Science*, **264**, 382–388.
2. van Veen, H. W., Higgins, C. F. & Konings, W. N. (2001). Multidrug transport by ATP binding cassette transporters: a proposed two-cylinder engine mechanism. *Res. Microbiol.* **152**, 365–374.
3. Schuldiner, S., Granot, D., Mordoch, S. S., Ninio, S., Rotem, D., Soskin, M. *et al.* (2001). Small is mighty: EmrE, a multidrug transporter as an experimental paradigm. *News Physiol. Sci.* **16**, 13013–13014.
4. Ubarretxena-Belandia, I. & Tate, C. G. (2004). New insights into the structure and oligomeric state of the bacterial multidrug transporter EmrE: an unusual asymmetric homo-dimer. *FEBS Letters*, **564**, 234–238.
5. Arkin, I. T., Russ, W. P., Lebendiker, M. & Schuldiner, S. (1996). Determining the secondary structure and orientation of EmrE, a multi-drug transporter, indicates a transmembrane four-helix bundle. *Biochemistry*, **35**, 7233–7238.
6. Schwaiger, M., Lebendiker, M., Yerushalmi, H., Coles, M., Groger, A., Schwarz, C. *et al.* (1998). NMR investigation of the multidrug transporter EmrE, an integral membrane protein. *Eur. J. Biochem.* **254**, 610–619.
7. Ubarretxena-Belandia, I., Baldwin, J. M., Schuldiner, S. & Tate, C. G. (2003). Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer. *EMBO J.* **22**, 6175–6181.
8. Tate, C. G., Ubarretxena-Belandia, I. & Baldwin, J. M. (2003). Conformational changes in the multidrug transporter EmrE associated with substrate binding. *J. Mol. Biol.* **332**, 229–242.
9. Tate, C. G., Kunji, E. R., Lebendiker, M., Schuldiner, S. & Yerushalmi, H. (2001). The projection structure of EmrE, a proton-linked multidrug transporter from *Escherichia coli*, at 7 Å resolution. *EMBO J.* **20**, 77–81.
10. Ma, C. & Chang, G. (2004). Structure of the multidrug resistance efflux transporter EmrE from *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **101**, 2852–2857.
11. Butler, P. J., Ubarretxena-Belandia, I., Warne, T. & Tate, C. G. (2004). The *Escherichia coli* multidrug transporter EmrE is a dimer in the detergent-solubilised state. *J. Mol. Biol.* **340**, 797–808.
12. Pornillos, O., Chen, Y. J., Chen, A. P. & Chang, G. (2005). X-ray structure of the EmrE multidrug transporter in complex with a substrate. *Science*, **310**, 1950–1953.
13. Tate, C. G. (2006). Comparison of three structures of the multidrug transporter EmrE. *Curr. Opin. Struct. Biol.* **16**, 457–464.
14. Muth, T. R. & Schuldiner, S. (2000). A membrane-embedded glutamate is required for ligand binding to the multidrug transporter EmrE. *EMBO J.* **19**, 234–240.
15. Grinius, L. L. & Goldberg, E. B. (1994). Bacterial multidrug resistance is due to a single membrane protein which functions as a drug pump. *J. Biol. Chem.* **269**, 29998–30004.
16. Gutman, N., Steiner-Mordoch, S. & Schuldiner, S. (2003). An amino acid cluster around the essential Glu-14 is part of the substrate- and proton-binding domain of EmrE, a multidrug transporter from *Escherichia coli*. *J. Biol. Chem.* **278**, 16082–16087.
17. Soskine, M., Adam, Y. & Schuldiner, S. (2004). Direct evidence for substrate-induced proton release in detergent-solubilized EmrE, a multidrug transporter. *J. Biol. Chem.* **279**, 9951–9955.
18. Sharoni, M., Steiner-Mordoch, S. & Schuldiner, S. (2005). Exploring the binding domain of EmrE, the smallest multidrug transporter. *J. Biol. Chem.* **280**, 32849–32855.
19. Weinglass, A. B., Soskine, M., Vazquez-Ibar, J. L., Whitelegge, J. P., Faull, K. F., Kaback, H. R. & Schuldiner, S. (2005). Exploring the role of a unique carboxyl residue in EmrE by mass spectrometry. *J. Biol. Chem.* **280**, 7487–7492.
20. Koteiche, H. A., Reeves, M. D. & McHaourab, H. S. (2003). Structure of the substrate binding pocket of the multidrug transporter EmrE: site-directed spin

‡ http://www.ebi.ac.uk/msd/index.html

labeling of transmembrane segment 1. *Biochemistry*, **42**, 6099–6105.
21. Fleishman, S. J., Unger, V. M. & Ben-Tal, N. (2006). Transmembrane protein structures without X-rays. *Trends Biochem. Sci.* **31**, 106–113.
22. Fleishman, S. J., Unger, V. M., Yeager, M. & Ben-Tal, N. (2004). A C-alpha model for the transmembrane alpha-helices of gap-junction intercellular channels. *Mol. Cell*, **15**, 879–888.
23. Fleishman, S. J., Harrington, S., Friesner, R. A., Honig, B. & Ben-Tal, N. (2004). An automatic method for predicting the structures of transmembrane proteins using cryo-EM and evolutionary data. *Biophys. J.* **87**, 3448–3459.
24. Baldwin, J. M., Schertler, G. F. & Unger, V. M. (1997). An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *J. Mol. Biol.* **272**, 144–164.
25. Beuming, T. & Weinstein, H. (2005). Modeling membrane proteins based on low-resolution electron microscopy maps: a template for the TM domains of the oxalate transporter OxlT. *Protein Eng. Des. Sel.* **18**, 119–125.
26. Briggs, J. A., Torres, J. & Arkin, I. T. (2001). A new method to model membrane protein structure based on silent amino acid substitutions. *Proteins: Struct. Funct. Genet.* **44**, 370–375.
27. Adamian, L. & Liang, J. (2006). Prediction of buried helices in multispan alpha helical membrane proteins. *Proteins: Struct. Funct. Genet.* **63**, 1–5.
28. Hurwitz, N., Pellegrini-Calace, M. & Jones, D. T. (2006). Towards genome-scale structure prediction for transmembrane proteins. *Phil. Trans. Roy. Soc. ser. B*, **361**, 465–475.
29. Fleishman, S. J. & Ben-Tal, N. (2006). Progress in structure prediction of alpha-helical membrane proteins. *Curr. Opin. Struct. Biol.* **16**, 496–504.
30. Daley, D. O., Rapp, M., Granseth, E., Melen, K., Drew, D. & von Heijne, G. (2005). Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science*, **308**, 1321–1323.
31. Rapp, M., Granseth, E., Seppala, S., von Heijne, G., Daley, D. O., Melen, K. & Drew, D. (2006). Identification and evolution of dual-topology membrane proteins. *Nature Struct. Mol. Biol.* **13**, 112–116.
32. Ninio, S., Elbaz, Y. & Schuldiner, S. (2004). The membrane topology of EmrE—a small multidrug transporter from *Escherichia coli*. *FEBS Letters*, **562**, 193–196.
33. Fu, D., Libson, A., Miercke, L. J., Weitzman, C., Nollert, P., Krucinski, J. & Stroud, R. M. (2000). Structure of a glycerol-conducting channel and the basis for its selectivity. *Science*, **290**, 481–486.
34. Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. (2002). X-ray structure of a ClC chloride channel at 3.0 A reveals the molecular basis of anion selectivity. *Nature*, **415**, 287–294.
35. Van den Berg, B., Clemons, W. M., Jr, Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C. & Rapoport, T. A. (2004). X-ray structure of a protein-conducting channel. *Nature*, **427**, 36–44.
36. Baldwin, J. M. (1993). The probable arrangement of the helices in G protein-coupled receptors. *EMBO J.* **12**, 1693–1703.
37. Yohannan, S., Faham, S., Yang, D., Whitelegge, J. P. & Bowie, J. U. (2004). The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. *Proc. Natl Acad. Sci. USA*, **101**, 959–963.
38. Creighton, T. E. (1993). *Proteins, Structures and Molecular Properties.* Freeman, New York.
39. Fleishman, S. J. & Ben-Tal, N. (2002). A novel scoring function for predicting the conformations of tightly packed pairs of transmembrane alpha-helices. *J. Mol. Biol.* **321**, 363–378.
40. Ninio, S. & Schuldiner, S. (2003). Characterization of an archaeal multidrug transporter with a unique amino acid composition. *J. Biol. Chem.* **278**, 12000–12005.
41. Gottschalk, K. E., Soskine, M., Schuldiner, S. & Kessler, H. (2004). A structural model of EmrE, a multi-drug transporter from *Escherichia coli*. *Biophys. J.* **86**, 3335–3348.
42. Zheleznova, E. E., Markham, P. N., Neyfakh, A. A. & Brennan, R. G. (1999). Structural basis of multidrug recognition by BmrR, a transcription activator of a multidrug transporter. *Cell*, **96**, 353–362.
43. Godsey, M. H., Zheleznova Heldwein, E. E. & Brennan, R. G. (2002). Structural biology of bacterial multidrug resistance gene regulators. *J. Biol. Chem.* **277**, 40169–40172.
44. Lewinson, O. & Bibi, E. (2001). Evidence for simultaneous binding of dissimilar substrates by the *Escherichia coli* multidrug transporter MdfA. *Biochemistry*, **40**, 12612–12618.
45. Mordoch, S. S., Granot, D., Lebendiker, M. & Schuldiner, S. (1999). Scanning cysteine accessibility of EmrE, an H+-coupled multidrug transporter from *Escherichia coli*, reveals a hydrophobic pathway for solutes. *J. Biol. Chem.* **274**, 19480–19486.
46. Soskine, M., Steiner-Mordoch, S. & Schuldiner, S. (2002). Crosslinking of membrane-embedded cysteines reveals contact points in the EmrE oligomer. *Proc. Natl Acad. Sci. USA*, **99**, 12043–12048.
47. Paulsen, I. T., Brown, M. H., Dunstan, S. J. & Skurray, R. A. (1995). Molecular characterization of the staphylococcal multidrug resistance export protein QacC. *J. Bacteriol.* **177**, 2827–2833.
48. Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R. & Iwata, S. (2003). Structure and mechanism of the lactose permease of *Escherichia coli*. *Science*, **301**, 610–615.
49. Rath, A., Melnyk, R. A. & Deber, C. M. (2006). Evidence for assembly of small multidrug resistance proteins by a "two-faced" transmembrane helix. *J. Biol. Chem*, **281**, 15546–15553.
50. Curran, A. R. & Engelman, D. M. (2003). Sequence motifs, polar interactions and conformational changes in helical membrane proteins. *Curr. Opin. Struct. Biol.* **13**, 412–417.
51. Kaback, H. R., Sahin-Toth, M. & Weinglass, A. B. (2001). The kamikaze approach to membrane transport. *Nature Rev. Mol. Cell. Biol.* **2**, 610–620.
52. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
53. Eddy, S. R. (1996). Hidden Markov models. *Curr. Opin. Struct. Biol.* **6**, 361–365.
54. Bairoch, A. & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucl. Acids Res.* **28**, 45–48.
55. Berezin, C., Glaser, F., Rosenberg, J., Paz, I., Pupko, T., Fariselli, R. *et al.* (2004). ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics*, **20**, 1322–1324.

56. Pupko, T., Bell, R. E., Mayrose, I., Glaser, F. & Ben-Tal, N. (2002). Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*, **18**, S71–S77.
57. Fleishman, S. J., Schlessinger, J. & Ben-Tal, N. (2002). A putative activation switch in the transmembrane domain of erbB2. *Proc. Natl Acad. Sci. USA*, **99**, 15937–15940.
58. Kessel, A. & Ben-Tal, N. (2002). Free energy determinants of peptide association with lipid bilayers. In *Current Topics in Membranes* (Simon, S. & McIntosh, T., eds), vol. 52, pp. 205–253, Academic Press, San Diego.
59. Elbaz, Y., Tayer, N., Steinfels, E., Steiner-Mordoch, S. & Schuldiner, S. (2005). Substrate-induced tryptophan fluorescence changes in EmrE, the smallest ion-coupled multidrug transporter. *Biochemistry*, **44**, 7369–7377.
60. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
61. Merritt, E. A. & Bacon, D. J. (1997). Raster3D photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524.

Edited by J. Bowie

(Received 19 July 2006; received in revised form 25 August 2006; accepted 25 August 2006)
Available online 30 August 2006