Recursive use of evolutionary conservation data in molecular modeling of membrane proteins

A model of the multidrug H\(^+\) antiporter EmrE

Jaume Torres and Isaiah T. Arkin

Cambridge Centre for Molecular Recognition, Department of Biochemistry, University of Cambridge, UK

Membrane proteins are currently the most biomedically important family of proteins, serving as targets for the majority of pharmaceutical agents. It is also clear that they are invariably abundant in all of the genomes sequence so far, representing up to a third of all open reading frames. Finally, and regrettably, it is clear that they are highly resistant to structural elucidation, representing less than 0.2% of the Protein Data Bank. Recent accomplishments in genome sequencing efforts, however, may help offset this imbalance through the availability of evolutionary conservation data.

Herein, we develop a novel approach, utilizing a combination of evolutionary conservation data and global searching molecular dynamics simulations to model membrane proteins, deriving a model for the multidrug H\(^+\) antiporter EmrE, a transmembrane four-helix bundle. Structures resulting from an extensive, rotational molecular dynamics search, were evaluated by comparing the residue specific interaction energy and the evolutionary conservation data. Subsequent rounds of molecular dynamics, in which confinement of the search space was undertaken in order to achieve a self consistent result, point to a structure that best satisfies the evolutionary conservation data. As the conservation patterns calculated for each of the helices suggested that the different conservation pattern for helix 3 (as well as being the most conserved) might be due to the oligomeric nature of EmrE, a dodecamer of helices was constructed based on the result of a search of helix 3 as a trimer. The resulting interaction energy per residue in the final model is in reasonable agreement with the evolutionary data and consistent with recent site directed mutagenesis experiments, pointing to the strength of this method as a general tool.

Keywords: EmrE; oligomerization; molecular dynamics; transporters; antibiotic resistance.

The dramatic achievements of structural biology provide chemical insights into a multitude of biological processes. However, despite numerous efforts, understanding processes taking place within lipid bilayers has been met with little success due to the fact that membrane proteins have remained resistant subjects to structural elucidation. This paucity of structural data is further exacerbated by the fact that membrane proteins are by far the most biomedically important family of proteins, in that they serve as the targets for the majority of pharmaceuticals agents.

As is usually the case, where experimental methods fail, theoretical approaches come into their own. This is particularly true for membrane proteins due to the fact the simple hydrophobicity algorithms [1] (followed by experimental verification) can outline the topology of most α-helical membrane proteins. Furthermore, one can assign sequence regions within the protein to each of the topological elements, classifying the protein (or domain) as a helical bundle. In contrast, defining the topology of a soluble protein is only possible through a complete structure determination with the prediction methods in use, unless structural data from homologous sequences are known. Therefore, what remains to be done is to determine the correct packing of the helices with respect to one another, a process which is clearly more difficult the more helices the transmembrane bundle possesses. However, even in simple cases it is virtually impossible to empirically determine what is the correct structure.

One possibility of overcoming such difficulties is making use of evolutionary conservation data, with the underlying assumption that conserved residues are more likely to reside in the protein–protein interface, than elsewhere [2,3]. Herein, we develop a method making use of such data, in an unbiased way, to obtain a model for a well studied membrane protein, EmrE.

The multidrug transporter EmrE is a member of a family of small (about 100 amino acids) multidrug-transporters from bacteria [4]. This family, termed Smr [4] or MiniTEXANs [5] has to date four established members: EmrE and SugE in Escherichia coli, Smr in Staphylococcus aureus and QacE from Klebsiella aerogenes. The best characterized member of the family, EmrE, has been purified and reconstituted in a functional form [5] and has been shown to catalyse H\(^+\)/cation antiport in proteoliposomes reconstituted with purified transporter, recognizing a wide range of inhibitors and substrates. EmrE can also be solubilized in a chloroform/methanol mixture and subsequently reconstituted in a functional mode [5], a feature that has been used in its purification.
Hydropathy analysis predicts that these proteins contain four transmembrane segments. Fourier transform IR (FTIR) spectroscopic studies [6] have shown that EmrE adopts a highly helical secondary structure (about 80%), when reconstituted in liposomes (made from either using dimyristoylphospho- choline or E. coli lipids) as well as in a chloroform/methanol mixture. The same authors have shown using polarized ATR-FTIR that the minimum average tilt angle of the helices is 27°.

Studies using thiol reactive reagents [7] have shown that the cysteines of EmrE are not reactive with small hydrophilic molecules. Reactivity was observed, however, in two of the cysteines (Cys41 and Cys95), with the hydrophobic cation 4-(chloromercuri)benzoic acid and its derivative 4-(chloromercuri)benzenesulfonic acid. In contrast, Cys39 was found to be nonreactive even with these reagents. Recently, a study in which roughly 50% of the residues of EmrE have been substituted by cysteine [8] has shown that none of the residues located in a transmembrane region reacts with N-ethylmaleimide, suggesting that the lumen of the channel is not hydrophilic. This is in agreement with previous H+/D+ exchange experiments [6] that show that the transmembrane segments are very resistant to H+/D+ exchange.

Fig. 1. EmrE and its homologous sequences after the alignment (A) and evolutionary conservation per residue in EmrE (B). (A) The reference of each sequence is indicated on the left, where: YKKC_BACSU, Hypothetical 11.9-kDa protein in HMP 3' region (Bacillus subtilis); CFU217271, Citrobacter freundii lipocalcin precursor (blic); Suge homolog (sueE), enterocin R (ecrnR), enterocin B precursor (ecnB), and enterocin A precursor (ecnA) genes, complete cds ecrR (C. freundii); SUGE_ECOLI, SUGE protein (E. coli); SUGE_PROVU, SUGE protein homolog (Proteus vulgaris); E35207, fltD homolog secA2 (Mycobacterium santhii); A70035, chaperonin homolog yvR (B. subtilis); B69857, chaperonin homolog yk kDa (B. subtilis); S1Y19441, Staphylococcus sp. plasmid pSt94 qacG and rep94 genes (Staphylococcus sp.); SAU319801, S. aureus plasmid phK4 replication protein Rep (rep) and quaternary ammonium compounds resistance protein Qac genes (S. aureus); SYY169451, S. saprophyticus plasmid pST2H6 qacH and rep2H6 genes (S. saprophyticus). EBR_STAAU, Ethidium bromide resistance protein (multidrug resistance protein) (S. aureus); D69619, multidrug resistance protein ebrB (B. subtilis); E69619, multidrug resistance protein ebrB (B. subtilis); MTCY22D16, Mycobacterium tuberculosis cosmid SCY22D7 (Mycobacterium tuberculosis); EMRE_ECOLI, EMRE protein (methyl viologen resistance protein C) (ethidium resistance protein) (E. coli); C70027, multidrug-efflux transporter homolog yvoE (B. subtilis); EBR_ECOLI, putative ethidium bromide resistance protein (E1 protein) (E. coli; Salmonella typhimurium and Pseudomonas aeruginosa); B69196, hypothetical protein b1600 (E. coli, strain K-12); A64916, hypothetical protein b1599 (E. coli, strain K-12). The more conserved residues in a 3–4 residue periodicity are indicated in boxes and black circles underneath. (B) The more conserved residues in helices 1, 2 and 4 are indicated by black circles and their number is indicated above. These residues were used to represent the face of the helix oriented towards the centre of the monomer bundle.
An intriguing feature of EmrE is that it is much smaller than the classical 12 transmembrane segment consensus for multi-drug transporters in cells [9], suggesting that EmrE may function as an oligomer. This assumption has been supported by a recent study [10], in which the activity of wild-type EmrE reconstituted in liposomes was decreased upon co-reconstitution with mutant inactive EmrE. It was shown that this pattern could also be observed by co-expressing wild-type and mutant EmrE in the same cells. These results were explained by assuming that inactive EmrE associates with wild-type monomers, which inactivates totally or partially the latter. Assuming that the inactivation caused by a mutant monomer is total, the results are consistent with a trimeric arrangement, although the evidence for this was not conclusive. In attempting to model results are consistent with a trimeric arrangement, although the evidence for this was not conclusive. In attempting to model proteins ab initio of the size of EmrE, several simplifications and assumptions have been made in order to cope with CPU time limitation. Even with these assumptions (discussed and justified as detailed below) we have spent 1 month CPU time using eight processors (DEC Alpha 433 MHz CPU units) in parallel, which is equivalent to more than 36 000 h (4 years) of CPU time using a standard SGI R8000 O2 workstation. We believe that each of the necessary assumptions made were justified based on this CPU time limitation. In the present study, models obtained from molecular dynamics simulations have been selected with respect to their correspondence with the evolutionary conservation data. The results not only suggest a model that is compatible with evolutionary and site-directed mutagenesis data, but also reveal a model for the inter-monomeric interaction.

MATERIALS AND METHODS

Sequence analysis

An estimation of the evolutionary conservation of the sequence of EmrE was determined using the programs available in seqlab, the wisconsin package v10.0, Genetics Computer Group (GGC), Madison, WI, USA. Homologous sequences were found from owl (v31.3) using FASTA, and these sequences were aligned using pileup. A consensus sequence and a numerical value for similarity were obtained using the scoring matrix blossom62 [11] that assigns positive or negative scores depending on the similarity.

Hydropathicity and surface probability plots were obtained using plotstructure (seqlab). Hydropathicity was obtained using the Kyte-Doolittle scale [12], using a window of 13 residues. The surface probability was obtained as described in [13], slightly modified for the end values of the protein chains.

Molecular modelling

A global search, in which the helices were rotated about their helical axis, was carried out as described elsewhere [14] assuming an asymmetric four-helix bundle. Several modifications were needed in order to take into account the large number of structures, mostly in the structure rmsd calculations (see below). Two models were used, A and B, which were created simply by exchanging helices 2 and 4. All calculations were performed with pcns, the parallel-processing version of the Crystallography and NMR System (CNS Version 0.3) [15]. The OPLS parameter set with united atom topology was used, representing explicitly polar hydrogens and aromatic side-chain atoms [16]. All calculations were carried out in vacuo with the initial coordinates of a canonical α helix (3.6 residues per turn). The searches used an initial crossing angle of −25°, introduced by rotating the long helix axis with respect to the long bundle axis. Each one of the helices was rotated independently every 20°. For each starting conformation, three trials were carried out using different initial random velocities in each trial. For a search over 180°, this procedure results in a total of 12 288 different trials, which in turn produce 12 288 final structures. Each structure was subjected to a simulated annealing and an energy minimization protocol, as described elsewhere [14]. The resulting structures were grouped in clusters, defined by having more than 10 structures, and where every structure pertaining to a cluster was within 1.0 Å rmsd from all the other structures in the same cluster. Note that this is a modification of the chl program [14] in which a cluster is composed of structures in which every structure is similar to at least one other in the cluster. The structures pertaining to the same cluster were averaged and this averaged structure was subjected to a simulated annealing protocol identical to that used in the systematic search and taken as representative of each cluster.

First guess orientation

For helices 1, 2 and 4, the use of helical wheel representations allowed us to restrict the conformational space searched to 180°, with the center of this arc pointing towards the center of the bundle and being defined by the face of the helix containing the most conserved residues. These helical wheel representations showed that the most conserved residues clustered preferentially in left handed configurations.

For helix 3, the first guess for the orientation of the helix in the bundle was obtained by assuming that this helix connects different protomers. A systematic search (both left and right handed) was performed using a symmetric trimer of only helices 3 searched every 10°. By assuming that helix 3 is also a left handed coiled-coil, only the left handed coiled-coil clusters were selected. As only two clusters were left handed, with a tilt angle of 11° and only separated by a ϕ angle of approximately 30° (not shown), the average orientation of these clusters was taken as a first guess for the orientation of helix 3 in the dodecamer.

Search protocol

In a first approximation, the search was constrained to a range of 180° for each helix. The interaction energy per residue in
each of the clusters (typically around 200–300 clusters) was compared to the evolutionary conservation per residue. The clusters having a $\chi^2$ smaller than an arbitrary value were selected. For these clusters, the average and standard deviation ($\sigma$) of the $\phi$ angle in each of the four helices was calculated.

In a second step, another search was performed over a range of $2\sigma$ around the average. This process was repeated until $\sigma$ no longer decreased. The cluster having the lowest $\chi^2$ (correlation between the evolutionary conservation data and the interaction energy per residue) in this last search was taken as the model for EmrE.

Finally, a second search for a trimer of helices 3 was performed by restricting the helix tilt (i.e. the angle between the vectors connecting every $C_\alpha$ of residue $i$ and $C_\alpha$ of residue $i+7$ and the z-axis) to $5^\circ$, as described previously [17]. The three protomers of EmrE were fitted to each one of the helices of the helix 3 trimer, to generate a dodecamer of helices, i.e. a trimer of EmrE protomers. The molecular graphics were generated with MOLSCRIPT [18].

Transmembrane segments

The sequence assigned to each transmembrane segment was chosen based on hydrophilicity/surface probability plots. Essentially, these segments are similar to the reported helix-forming segments of EmrE in chloroform/methanol 4–26, 32–53, 58–76 and 85–106 [19] for helices 1, 2, 3 and 4, respectively. Whereas we have used almost identical segments for helices 2 (31–53) and 4 (85–105), we have shortened helix 1 (4–21) based on the hydrophobicity and surface probability plots. Helix 3 (58–81) has been extended from residue 76–81, as these four residues (77–81) have a low surface probability.

Correlation between evolutionary conservation data and residue interaction energy

The correlation between evolutionary conservation data and residue interaction energy was calculated using $\chi^2$ analysis given by:

\[ \chi^2 = \sum \frac{(O - E)^2}{E} \]

Fig. 3. Helical wheel diagrams for the four transmembrane helices of EmrE at 3.5 (top), 3.6 (middle) and 3.9 (bottom) residues per turn. The conservation of the residues is indicated by the color of the circles, according to the quantitative values plotted in Fig. 1. The scale is indicated on the bar at the bottom, with blue being the more conserved during evolution. The residues indicated in Fig. 1 are indicated here by numbers of bigger font size. The face pointing towards the center of the bundle (first guess) is indicated by an arrow.
whereby \( o_i \) is the normalized interaction energy of residue \( i \), and \( e_i \) is the normalized evolutionary conservation of residue \( i \) calculated using the blosum62 scoring matrix [11].

**RESULTS AND DISCUSSION**

**Derivation of evolutionary constraints**

Sequence comparison of EmrE and all existing databases resulted in 19 highly homologous proteins, all of similar lengths. Multiple alignment of these sequences enabled determination of a consensus sequences and a numerical value for similarity, using the blosum62 scoring matrix. As shown in Fig. 1, there is a clear periodicity in the conservation pattern, which is consistent with residues on one side of the helix being preferentially conserved. No distinction can obviously be made *a priori* between residues that are conserved for structural purposes to those conserved for functional purposes. However, in terms of molecular modelling, neither set of residues is likely to reside in the protein–lipid interface. Thus, the most conserved face of the helix has been oriented towards the center of the bundle as a first guess. The possibility that this face intervenes in interhelical contacts is accounted for by allowing the search to proceed at ±90° from this first guess orientation.

**Delineation of putative transmembrane α helices**

In order to analyse the conservation pattern, it was first necessary to delineate the putative transmembrane helices in the protein using hydrophobicity analysis [1]. The results of this analysis combined with surface probability calculations, which is also a tool to assess the hydrophobicity of a residue, are shown in Fig. 2. This clearly indicates the presence of four putative transmembrane α helices, consistent with previous FTIR data [6]. The helix-forming segments in chloroform/methanol have also been considered, as shown in this figure. It is recognized that precise delineation of the ends of the transmembrane α helices would be difficult. However, during the molecular dynamics, simulated annealing protocol, some adjustment could be made in terms of lateral movements of the helices with respect to one another in order to maximize their interactions.

**Use of evolutionary conservation data helical wheel representation**

With the sequence of the putative helices it was possible to determine the correlation between the evolutionary conservation pattern and the helical periodicity. Three helical wheels were analysed with varying pitches: 3.6, 3.5 and 3.9 amino acids per turn representing a canonical helix, a left handed coiled-coil and a right handed coiled-coil, respectively (see Fig. 3). Out of the four α helices in EmrE, only in helix 3 was it impossible to define the one face of the helix that was preferentially conserved (see below). The results shown in Fig. 3 clearly indicate that a far better agreement between helical periodicity and the evolutionary conservation pattern exists in a right handed coiled-coil, and does not necessitate any sort of Fourier power spectrum analysis.

In justification of the approach described above, we note that it has been shown that evolutionary conserved residues preferentially reside in the protein–protein interface [2,3]. As such, a left handed bundle would readily accommodate such a requirement. Furthermore, it is noted that this approach has been successfully applied to the structure of human glycophorin A [20], whereby lining up the residues with equal sensitivity to mutations, resulted in a determination of the bundle topology, which was later confirmed by NMR spectroscopy (K. R. MacKenzie, J. Prestegard & D. M. Engelman, unpublished results). Further comparison with a representative set of water soluble four-helix bundles from the SCOP database [21] (ID codes: 1eci, 1fha, 1lpe, 1rop and 256b) revealed that they all are left handed bundles. Based on the above we have decided to model EmrE as a left handed helical bundle in order to reduce by half the CPU requirement.

**Global search protocol: initial conditions**

As the handedness of the helix bundle was based upon the correlation between the helical periodicity and the evolutionary conservation data, as well as the topology of known four-helix bundles, what remained to be determined was the bundle configuration (Fig. 4). These two distinct configurations are termed throughout the text as A and B. The search was limited to these two models based on the reasonable assumption that the loops connecting the transmembrane segments in EmrE are too small (around four residues) to allow two contiguous helices that are connected in diagonal across the tetramer. Any other configuration would require the interhelix loops to traverse diagonally across the bundle. Furthermore, inspection
of several water soluble four-helix bundle proteins from the protein data bank (ID numbers: 1eci, 1lpe, 1rop, and 256b), showed that these proteins all share the same topology, even though their loop sizes were not as short as found in EmrE. Once again, CPU limitation necessitated modelling solely plausible configurations, which in this case maintained concentrating on helices connected through adjacent loops.

Once both bundle configuration and handedness were determined, the stage was set for a global search molecular dynamics protocol, aimed at determining local clusters of structures which represent energy minima [14]. However, a complete global search, at 20° rotation increments would require to perform molecular dynamics and energy minimization on \( \left(\frac{360°}{20°}\right)^4 \times 10^4 \) structures (not taking into account repetition with different initial random velocities). Even on a fast computer each calculation takes about 10 min per structure; it is clear that the rotational search space needed to be restricted.

Utilizing the correlation between the helical periodicity and evolutionary conservation pattern, it was possible to limit the helix rotational search space to ±90°, centered at the helix conservation vector depicted in Fig. 3. In contrast to helices 1, 2 and 4, the pattern of conservation in helix 3 is not as clear (Figs 1 and 3) because almost all residues are well conserved (being the most conserved helix in EmrE) and a periodicity is not evident. This problem was overcome by considering that previous reports [10] have shown that EmrE is an oligomer, probably a trimer, i.e. a dodecamer of helices. Indeed, the fact that helix 3 is the only helix with an anomalous conservation pattern, as well as being the most conserved, suggests that helix 3 connects the three monomers in the trimer, and this is what generates this unique pattern of conservation, shown schematically in Fig. 5.

Thus, in order to obtain a first guess for the orientation of helix 3 in the tetramer, a search was performed using only a trimer of helix 3, the rationale behind this attempt being that the orientation of the helix 3 in the trimer would be similar to that present in the dodecamer (trimer of EmrE molecules). Fig. 6 shows the result of this search from inside the lumen of an EmrE monomer, i.e. the indicated residues are the ones that face the inside of the EmrE monomer. The face of helix 3 facing the lumen of the monomer was used as the anchor point for the rotational variation in a similar manner to the conservation vectors of helix 1, 2 and 4.

**Global search results**

A global search was conducted to cover a range of 180° of each helix, using three different initial random velocities resulting in the formation of 12 288 structures. Lateral movement of the helices was accounted for during the molecular dynamics, simulated annealing protocol. Both configurations A and B (Fig. 4) of a left handed coiled-coil were simulated and the \( \chi^2 \) comparison between the evolutionary conservation data of the resulting clusters is shown in Fig. 7.

Clearly, the \( \chi^2 \) values for the clusters obtained searching configurations A are lower (Fig. 7), suggesting that this is the configuration adopted by EmrE. Hence, subsequent searches were performed only assuming configuration A. After selection of the best clusters, the average and standard deviation for the rotational angle difference between the original position and the final state, \( \phi \), were calculated. The results for the four helices were then used as inputs for the following recursive searches aimed at achieving a self-consistent agreement with the evolutionary conservation data. This iterative approach is schematically illustrated in Fig. 8.

The recursive process was repeated three times until the SD was unchanged (Table 1). As seen in Fig. 7, the second step in the recursive process indeed did result in a better fit with the evolutionary conservation data, as evident from an average reduction in the \( \chi^2 \) values. Finally, the cluster having the smallest \( \chi^2 \) value in this last search was chosen as the model for EmrE. A comparison of the interaction energy per residue and the conservation data is represented in Fig. 9. The conservation per residue in this figure does not correlate extremely well with the interaction energy. This would be worrying if the residues had a high interaction energy according to the model but were not conserved through evolution. However, in almost all residues where a discrepancy exists, residues are more conserved than expected from their interaction energies,
suggesting that these residues have a functional significance rather than a structural one, because in that case their interaction energy does not need to be high.

**EmrE trimer generation**

In each one of the searches it was observed that the tilt angle for helix 3 was consistently smaller than for the other helices (≈ 5°). As the tilt angle for helix 3 in the helix 3 trimer was around 11° (see above), a second search involving only a trimer of helices 3, restricting the tilt angle to 5° (see Materials and methods) was performed. The final clusters were selected only according to their energy (Fig. 10). In this figure, the best clusters are represented in black and share around 95% of the structures, which means that their structure is essentially the same.

Helix 3 in these clusters (Fig. 10) is oriented in a way that is entirely consistent with the orientation observed for the best cluster in the search (Fig. 11). In this figure, the cytoplasmic and periplasmic sides have been suggested on the basis of experiments [7] that show that 4-(chloromercu)benzenesulfonic acid, a reagent that in contrast to 4-(chloromercu)benzoic acid does not permeate through the protein, reacts with Cys41 only when allowed to permeate to the cell interior and with Cys95 only when the reagent is present in the outside face of the membrane.

Finally, superposition by the fitting of helix 3 in three EmrE monomers to the helix 3 trimer in the model represented in Fig. 10, produces a model for the putative dodecamer, i.e. a trimer of EmrE monomers (Fig. 12).

**Structural and functional implications**

In analysing the final structure, we note that the three cysteines are located in the interhelical region, and not oriented toward the lumen of the EmrE monomer. This is consistent with the fact that none of them seems to be necessary to confer multiple drug resistance to *E. coli* [7]. According to our model, Cys39 would be the least accessible to the lumen and this is also in agreement with experimental data [7] that show that a mutant with only Cys39 is the least reactive with the mercurial reagent 4-(chloromercu)benzoic acid. The most sensitive was C41, and this residue was found to be also the most accessible from the lumen.

In their recent paper, Schuldiner and colleagues [8] have substituted around 50% of the residues of EmrE with cysteine, identifying four different groups of residues. The first group are

Table 1. Averages and standard deviations for the $\phi$ angles corresponding to the clusters selected after the first, second and third searches for conformation A.

<table>
<thead>
<tr>
<th>Search</th>
<th>Helix 1</th>
<th>Helix 2</th>
<th>Helix 3</th>
<th>Helix 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>188 ± 44</td>
<td>89 ± 51</td>
<td>57 ± 64</td>
<td>100 ± 50</td>
</tr>
<tr>
<td>2</td>
<td>148 ± 29</td>
<td>90 ± 26</td>
<td>34 ± 20</td>
<td>67 ± 22</td>
</tr>
<tr>
<td>3</td>
<td>136 ± 24</td>
<td>103 ± 19</td>
<td>36 ± 19</td>
<td>60 ± 17</td>
</tr>
</tbody>
</table>

Fig. 7. Plots corresponding to the chi-squared value of the difference between the evolutionary data and the clusters obtained from the search. Top: values for conformation A (middle) and B (top) obtained in the first trial. Bottom: results obtained from the second trial in conformation A.
residues that lead to lack of expression (Tyr40 and Phe44 in TM2 and Leu93 in TM4). The second group consists of residues that, when exchanged for a cysteine, lead to expression but are not functional. These are Thr18, Glu14 and Ile11, all in TM1. The residues pertaining to these two first groups are labelled with red numbers in Fig. 11. The third group, Val34, Thr36, Ala42, Cys41, Leu47, Gln49 and Ala48, all of them in TM2, labelled in green in Fig. 11, conferred reduced resistance, mainly to methyl viologen. The remaining residues in TM2 (labelled in black in Fig. 11) could be substituted without loss of function.

As is shown in Fig. 11, the residues labelled with red numbers are in close spatial proximity. Residues Phe44, Ile11 and Leu47 on one hand, and Thr18 and Thr36 on the other, are located, respectively, immediately above and below Glu14, an essential residue that cannot be substituted even by aspartate without loss of function. In turn, Glu14 is facing Tyr40 and Leu93.

As reported by Schuldiner and colleagues [8] the sensitive residues in TM2 (coloured in green in Fig. 11) are clustered on opposite sides of the helix. Whereas some of these are facing the lumen (Thr36, Tyr40, Phe44 and Leu47), others (Val34, Ala42, Cys41, Ala48 and Gln49) are oriented preferentially away from the lumen. According to our model, this face would
be interacting with residues in TM4 of another protomer of EmrE (see Fig. 12), perhaps conferring structural stability. The fact that these residues oriented away from the lumen are important for function is consistent with experiments that show the oligomeric nature of EmrE [10], in which function was impaired when active and inactive molecules were reconstituted at the same time. Consequently, we would expect that mutants containing cysteine in residues of TM4 that, according to our model interact with the sensitive residues in TM2 would also show a decreased resistance to methyl viologen. Unfortunately, there is no data available for these residues. We cannot speculate which one of the residues in TM4 would be sensitive to mutation because the precise relative orientation of helices 2 and 4 in different protomers depends in our model upon the correct orientation of the trimer of helix 3. As we have not performed any energy minimization on the EmrE trimer structure such speculation should be deferred to a later stage. Nevertheless, the model describes a pocket around Glu14 lined with essential residues and predicts for the first time a configuration for the EmrE trimer.

**ACKNOWLEDGEMENTS**

We thank Prof. Shimon Schuldiner for a valuable discussion of the manuscript. This work was supported by a grant from the BBSRC and the Wellcome trust to I. T. A.

**REFERENCES**
