Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallography of the integral membrane protein EmrE from *Escherichia coli*

Che Ma‡ and Geoffrey Chang*

Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, CB-105, La Jolla, CA 92037, USA

‡ Current address: Genomics Research Center, Academia Sinica, Taipei 115, Taiwan.

Correspondence e-mail: gchang@scripps.edu

Received 28 July 2004
Accepted 8 October 2004

Crystals of the EmrE membrane-protein imposed several technical challenges for X-ray crystallography, including high mosaicity, poor diffraction and a relatively large number of heavy atoms. Consequently, the heavy-atom substructure solution was difficult to obtain. By removing the histidine tag for protein purification, the mosaicity and the diffraction quality were greatly improved. The direct-methods *Shake-and-Bake* program *SnB* was successful in locating the heavy-atom sites from a mutant of EmrE which lacks a cysteine and therefore has a reduction in the number of heavy-atom sites. The substructure solution was solved from data with anomalous difference at a resolution of 5.5 Å and the structure was determined to 3.8 Å.

1. Introduction

Membrane proteins are challenging targets for structural biologists and their structures are underrepresented. Of more than 26 000 structures deposited in the Protein Data Bank (http://www.rcsb.org/pdb/), about 80 are from integral membrane proteins. The difficulties lie in all aspects of membrane-protein structure determination by X-ray crystallography. Firstly, sufficient quantities of suitable material are not always easy to obtain, since the expression levels of membrane proteins are often low. Secondly, the requirement of using appropriate detergent(s) complicates the protein-purification procedures. Thirdly, the crystallization of membrane proteins can be a difficult task. Often, membrane-protein crystals give poor diffraction owing to structural heterogeneity and disorder. Finally, the structures can be a challenge to solve at this modest resolution, which is frequently also the case for other protein crystals. The structure determination of the multidrug transporter EmrE from *Escherichia coli* is an example of all the above (Ma & Chang, 2004).

EmrE from *Escherichia coli* belongs to the small multidrug-resistance (SMR) family (Paulsen *et al.*, 1996). SMR pumps utilize the proton electrochemical gradient as an energy source to pump out various cationic hydrophobic compounds such as ethidium bromide, methyl viologen and tetracycline, as well as other antiseptics and intercalating dyes. EmrE is a small 12 kDa protein with 110 amino-acid residues and four predicted α-helices. The wild-type apo EmrE crystallizes in space group *F*222 and diffracts to a maximum resolution of about 5.5 Å. Upon derivatization by soaking with mercury chloride, the diffraction resolution limit increased to 3.5 Å. Despite three Harker sections, we were unable to determine the heavy-atom substructure from the wild-type crystals owing to the complexity of the Patterson maps and the low resolution limit for the diffractions. In an effort to reduce the number of the heavy-atom sites, we attempted to crystallize 25 cysteine mutants of EmrE. Only one mutant, C41S, crystallized and led to a mercury substructure with a reduced number of sites and was solvable. The direct-methods *Shake-and-Bake* program *SnB* was used to locate the heavy-atom positions for the structure determination using anomalous difference data to 5.5 Å.

2. Methods

2.1. Expression, purification and crystallization

The expression, purification and crystallization procedures have been described previously (Ma & Chang, 2004). In brief, the *emrE* gene was PCR-amplified from the genomic DNA from *E. coli*, cloned into pET15b and pET19b expression vectors (Novagen), which added an N-terminal histidine tag and a thrombin- or an enterokinase-cleavage site, respectively, and overexpressed in *E. coli* host BL21(DE3). During protein expression, tetracycline was added to the medium to select for cells expressing functional EmrE drug transporters. Extraction of EmrE was performed with 2% *N*-nonyl-β-glucopyranoside (NG). 0.3% NG was present during the purification process by nickel-chelation, ion-exchange and gel-filtration chromatography. The histidine tag was removed by thrombin digestion and the protein was

Table 1
Diffraction data statistics as a function of resolution shell used to determine heavy-atom sites using mutant C41S (16-site search).

R_{meas} is the multiplicity-weighted R_{merge} relative to I^+ or I^-. R_{meas0} is relative to the overall mean.

Shell	Resolution (Å)	$I/\sigma(I)$	Redundancy	Completeness (%)	R_{meas}	R_{meas0}	R_{merge}
1	12.65	13.2	3.2	87.7	0.056	0.067	0.041
2	8.94	9.2	3.2	87.0	0.076	0.084	0.055
3	7.30	6.7	3.0	91.9	0.090	0.094	0.066
4	6.32	5.2	3.1	93.5	0.144	0.151	0.106
5	5.66	2.2	3.2	93.5	0.239	0.240	0.175
6	5.16	3.0	3.2	94.2	0.255	0.300	0.191
7	4.78	2.5	3.2	94.4	0.280	0.320	0.250
8	4.47	2.4	3.3	94.7	0.321	0.331	0.281
9	4.22	2.2	3.3	94.7	0.360	0.363	0.340
10	3.80	2.1	3.2	94.4	0.398	0.397	0.385

concentrated to ∼15 mg ml^{-1}. Crystals of *E. coli* EmrE were obtained using the vapor-diffusion hanging-drop method at 277 K by combining the protein solution with precipitant in a ratio of 2–3:1. The precipitant solution contained 15–30% polyethylene glycol 200, 20 m*M* sodium chloride, 200–600 m*M* ammonium sulfate, 20 m*M* sodium acetate pH 4 and 0.3–0.6% NG. Crystals appeared within 5 d and continued to grow for two weeks to final dimensions of 0.3 × 0.4 × 0.4 mm.

2.2. Data collection

Anomalous difference data were collected at Advanced Light Source BL 5.0.2 near the Hg L_{III} edge (λ = 1.0072 Å) and inflection point (λ = 1.0094 Å) from EmrE C41S mutant crystals derivatized with 10 m*M* mercury chloride. The anomalous data were suitable to 5.5 Å (Table 1). Data were not collected at the remote wavelength owing to severe radiation decay. The redundancy and completeness of the data were 3.5 and 90.1%, respectively.

2.3. Substructure determination by *Shake-and-Bake*

The mercury substructure of the crystals of EmrE C41S mutant was determined using the program *SnB* v.2.2 (Miller *et al.*, 1994; Weeks & Miller, 1999) with the following parameters: number of expected sites, 12; data resolution, 40–5.5 Å; both Bayseian estimates and locally normalized $|E|$ were used to generate the normalized structural factor E; number of reflections, 360; minimum $E/\sigma(E)$, 2.0; maximum $|E|$, 5.00; number of triplet invariants, 3600. For reciprocal-space phase refinement (shaking), parameter shift with default setting (90° phase shift, two shifts and three passes through phase set) was chosen. For real-space Fourier refinement (baking), 12 peaks were selected with grid size 1.83 and minimum inter-peak distance 3.0 Å. 1000 trials were performed starting with a randomized atom structure.

3. Results and discussion

The approach using natural sequence variations among the EmrE homologs increased the probability of discovering better protein targets for expression, purification and crystallization. EmrE from *E. coli* was targeted for structure determination from a panel of nine bacterial orthologs based on its high protein-expression levels and suitability for three-dimensional crystallization. The protein-expression level of *E. coli* EmrE is about 0.1 mg per litre of culture, which is modest by the standards of most structural biologists. In order to obtain sufficient quantities of EmrE protein, the cell culture was scaled up using fermentors with 100 l operational volume. During expression, tetracycline was added to the culture to select for cells expressing functional EmrE drug transporters. This step not only selected cells that expressed EmrE, but also improved the purity, homogeneity and yield of the final protein product. In addition, our EmrE construct has demonstrated functional activity (Zhang *et al.*, 2004). EmrE can be solubilized with several different detergents. This includes the maltosides (decyl, undecyl, dodecyl and tridecyl), glucosides (octyl, nonyl and decyl) and fos-choline (carbon-chain length 10–14). Nonyl glucoside was selected because it maintained the integrity of the protein (EmrE appears as tetramer, the proposed oligomerization state, using gel-filtration chromatography) and was suitable for crystallization. Moreover, the sample remained active and bound one of its substrates, tetraphenylphosphonium, with high affinity. The apo EmrE crystals were obtained from a crystallization screen of detergents, precipitants, salts, pHs, additives and temperatures.

High mosaicity and poor diffraction are often observed with membrane-protein crystals. This disorder is commonly attributed to the positional disorder of the target protein construct. EmrE fused with a histidine tag and an enterokinase cleavage site (pET19b) crystallized in space group $I4_122$ and only diffracted to ∼7 Å with a relatively high mosaicity of 3.5°. We speculated that a possible source of disorder was caused by the purification tag and attempted to remove it using enterokinase. However, these efforts failed and we subcloned EmrE into the vector pET15b containing a thrombin-cleavage site. Upon incubation with thrombin, we obtained wild-type protein that was active. This untagged EmrE construct yielded the *F*222 crystal form with unit-cell parameters a = 178, b = 238, c = 289 Å, $\alpha = \beta = \gamma = 90°$ (Fig. 1*a*), which diffracted to about 5.5 Å. Upon derivatization with mercury chloride, methyl mercury chloride, dimethyl mercury or ethyl mercury chloride, the mosaicity fell to 0.7° and the resolution limit of the diffraction greatly improved to 3.5 Å.

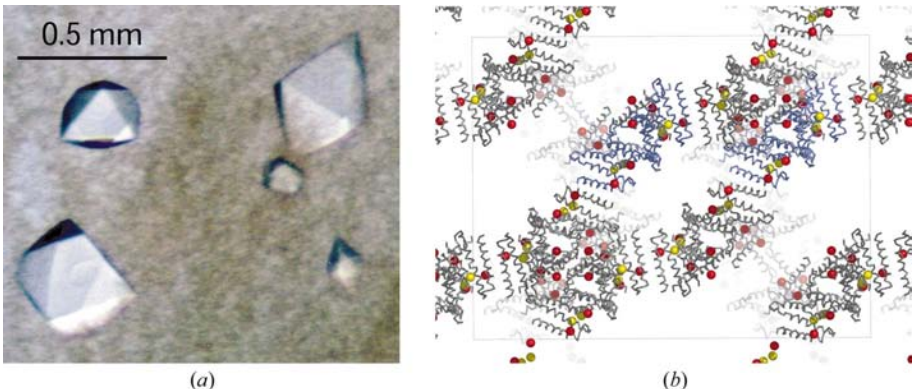

Figure 1
Crystals and the crystal-packing diagram of *E. coli* EmrE. (*a*) A picture of the crystals from *E. coli* EmrE. (*b*) Crystal packing of EmrE transporters looking down the *F*222 *c* axis. EmrE is represented by black lines. One of the asymmetric units is colored blue. Red spheres represent the mercury sites identified from the mutant C41S crystals and yellow spheres represent the additional mercury sites from the wild-type crystals. All molecular graphics were generated with *PyMOL* (http://pymol.sourceforge.net/).

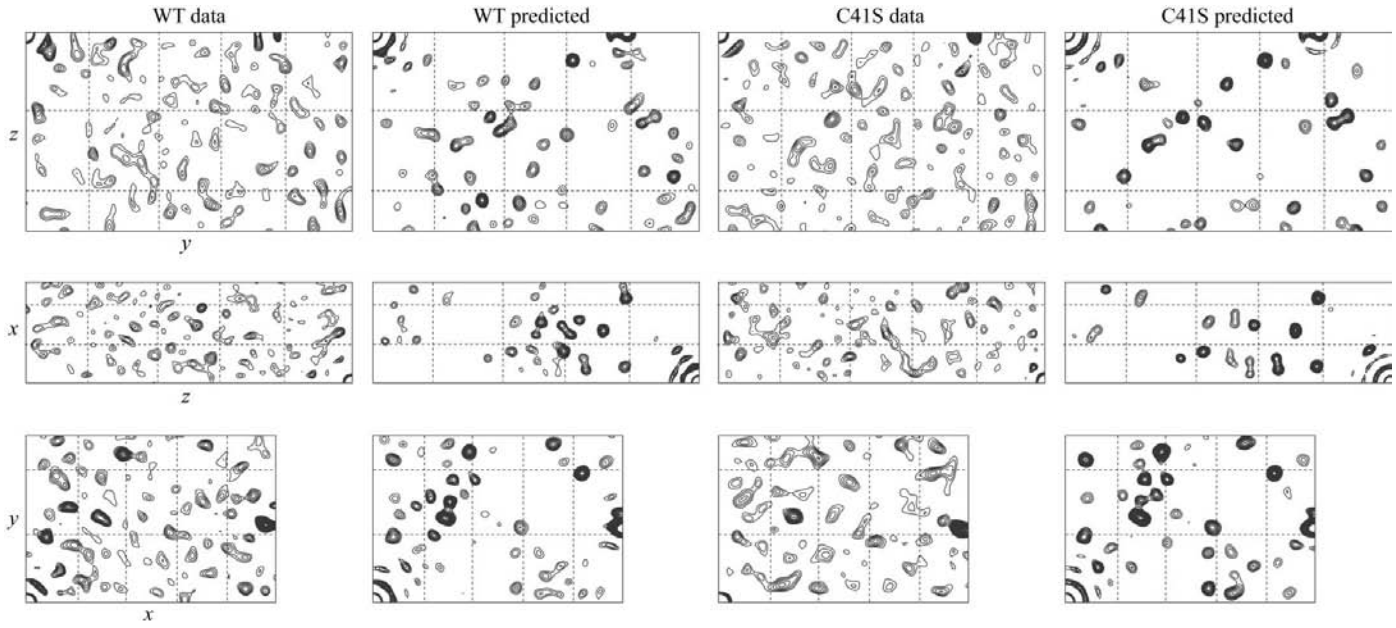

Figure 2
Anomalous difference Patterson maps of the Harker sections from both mercury-derivatized wild-type and C41S mutant crystals and their corresponding predicted maps. Only minimal Patterson asymmetric units are shown. The axis units for the maps are $y = 0$ to 0.5 and $z = -0.25$ to 0 for the yz plane, $x = 0$ to 0.25 and $z = -0.5$ to 0 for the zx plane, and $x = 0$ to 0.5 and $y = 0$ to 0.25 for the xy plane. All maps were generated with *CNS* using the minimum contour level of 1σ with 0.5σ contour increments.

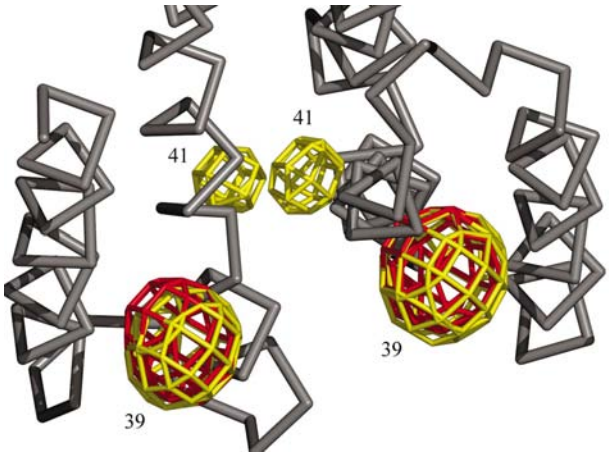

Figure 3
Close-up view of the EmrE structure (PDB code 1s7b) with anomalous difference density obtained using the model phases for mercury sites from C41S data (red) and wild-type data (yellow). The locations of cysteines 39 and 41 are indicated.

The anomalous difference Patterson maps from the wild-type crystals appeared to be complicated (Fig. 2) and we were unable to solve the heavy-atom substructure using Patterson search programs such as *SOLVE* (Terwilliger & Berendzen, 1999), *CNS* (Brünger *et al.*, 1998), *CCP*4 (Collaborative Computational Project, Number 4, 1994), genetic algorithms (Chang & Lewis, 1994) or by inspection of Harker sections. Indeed, in retrospect, a significant number of the cross-vectors and self-vectors in the anomalous Patterson matrix had negative densities (data not shown) because of the relatively large number of sites. We attribute this failure to the lower resolution limit of the anomalous difference data (only good to 5.5 Å), the large number of heavy-atom sites and the fairly high-symmetry space group *F*222. Attempts using direct-methods programs that utilize the *Shake-and-Bake* algorithm such as *SHELXD* (Schneider & Sheldrick, 2002) and *SnB* (Miller *et al.*, 1994; Weeks & Miller, 1999) were also unsuccessful in identifying the mercury locations in the wild-type crystals.

Mercurial compounds are expected to react with cysteines, of which EmrE contains three: Cys39, Cys41 and Cys95. In an attempt to reduce the number of mercury sites in the derivatized crystals, we constructed single and double mutants with either one or two of these cysteines replaced by the conserved residues found in the sequence alignment from EmrE homologs, such as serine, glycine, alanine, valine, leucine or isoleucine. Surprisingly, we were only able to crystallize one mutant out of 25: C41S. The C41S mutant crystal also crystallized in space group *F*222 and is isomorphous with the wild-type crystal. Although the anomalous difference Patterson maps of the mutant C41S also appeared to be complicated (Fig. 2), the reduction in the number of heavy-atom sites produced sufficient simplification for *SnB* to find the substructure solution.

The direct-methods program *SnB* is based on the dual-space *Shake-and-Bake* algorithm (Miller *et al.*, 1994; Weeks & Miller, 1999). With high-resolution data, *SnB* was successful in determining the *ab initio* phasing of a complete crystal structure of a triclinic lysozyme containing 1001 non-H atoms and ~200 bound water molecules (Deacon *et al.*, 1998). *SnB* can also be effective in solving the heavy-atom substructures of proteins derivatized with a large number of anomalous scatterers (Deacon *et al.*, 2000; von Delft *et al.*, 2003). However, the success rate is dependent on several factors such as the space group, the number of residues in the asymmetric unit, the number of anomalous scatterers in the asymmetric unit, the resolution of the data, the size of the unit cell and the data quality (Wang & Ealick, 2003). The asymmetric unit of EmrE crystal contains eight EmrE

monomers arranged as two non-crystallographic tetramers. Therefore, the crystal of wild-type EmrE protein contains 24 (8×3) mercury sites in the asymmetric unit. *SnB* was unable to find the solution from the wild-type crystals at this resolution. Nevertheless, because of the reduction in the number of anomalous scattering sites from 24 to 16, *SnB* was successful in finding the solution using single anomalous difference (Hg $L_{\rm III}$ edge) data from C41S mutant crystals using the resolution range 40–5.5 Å. Initially, we set the number of search sites in *SnB* to 12 as we were not sure of the number of molecules in the asymmetric unit. *SnB*, however, was able to consistently identify the positions of 14 (out of 16) heavy atoms even when searching for 16 possible sites. The position of the two remaining sites was confirmed first by difference Fourier using protein phases and also by non-crystallographic symmetry.

16 mercury sites from C41S EmrE were used to generate initial protein phases using the program *PHASES* (Furey & Swaminathan, 1997). A cross-phased Fourier map was calculated using phases from the C41S mutant in combination with the anomalous difference data from the wild-type crystals. Eight additional mercury sites corresponding to the Cys41 position from the wild-type crystal were then identified (Fig. 3). The final electron-density map for model building was generated by the combined protein phases, solvent flattening, fourfold non-crystallographic averaging and the resolution was then extended from 4.5 to 3.8 Å using in-house programs (G. Chang, unpublished work). The crystal-packing diagram (Fig. 1*b*) reveals the complexity of the mercury substructure with EmrE molecules in the unit cell.

This work is supported by grants from NIH GM067744, NASA NAG8-1834 and the Beckman Young Investigator Award.

References

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Chang, G. & Lewis, M. (1994). *Acta Cryst.* D**50**, 667–674.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Deacon, A. M., Ni, Y. S., Coleman, W. G. & Ealick, S. E. (2000). *Structure*, **8**, 453–462.

Deacon, A. M., Weeks, C. M., Miller, R. & Ealick, S. E. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 9284–9289.

Delft, F. von, Inoue, T., Saldanha, S. A., Ottenhof, H. H., Schmitzberger, F., Birch, L. M., Dhanaraj, V., Witty, M., Smith, A. G., Blundell, T. L. & Abell, C. (2003). *Structure*, **11**, 985–996.

Furey, W. & Swaminathan, S. (1997). *Methods Enzymol.* **277**, 590–620.

Ma, C. & Chang, G. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 2852–2857.

Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.

Paulsen, I. T., Skurray, R. A., Tam, R., Saier, M. H. Jr, Turner, R. J., Weiner, J. H., Goldberg, E. B. & Grinius, L. L. (1996). *Mol. Microbiol.* **19**, 1167–1175.

Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.

Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.

Wang, J. & Ealick, S. E. (2003). *J. Appl. Cryst.* **36**, 1397–1401.

Weeks, C. M. & Miller, R. (1999). *Acta Cryst.* D**55**, 492–500.

Zhang, Z., Ma, C., Pornillos, O., Chang, G. & Saier, M. H. Jr (2004). In the press.