BIOLOGICAL CRYSTALLOGRAPHY
ISSN 1399-0047

Crystallographic study of a MATE transporter presents a difficult case in structure determination with low-resolution, anisotropic data and crystal twinning

Jindrich Symersky,[a] Yi Guo,[a] Jimin Wang[b] and Min Lu[a]*

[a]Department of Biochemistry and Molecular Biology, Rosalind Franklin University of Medicine and Science, North Chicago, Illinois, USA, and [b]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, USA. *Correspondence e-mail: min.lu@rosalindfranklin.edu

Received 9 July 2015
Accepted 10 September 2015

Edited by Z. Dauter, Argonne National Laboratory, USA

Keywords: multidrug resistance; MATE transporter; data anisotropy; twinning; difference Fourier analysis.

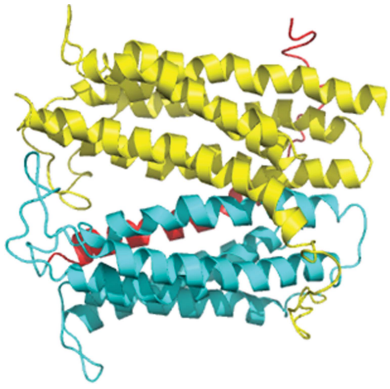
NorM from *Neisseria gonorrhoeae* (NorM-NG) belongs to the multidrug and toxic compound extrusion (MATE) family of membrane-transport proteins, which can extrude cytotoxic chemicals across cell membranes and confer multidrug resistance. Here, the structure determination of NorM-NG is described, which had been hampered by low resolution (~4 Å), data anisotropy and pseudo-merohedral twinning. The crystal structure was solved using molecular replacement and was corroborated by conducting a difference Fourier analysis. The NorM-NG structure displays an extracellular-facing conformation, similar to that of NorM-NG bound to a crystallization chaperone. The approaches taken to determine the NorM-NG structure and the lessons learned from this study are discussed, which may be useful for analyzing X-ray diffraction data with similar shortcomings.

1. Introduction

Multidrug transporters belonging to the multidrug and toxic compound extrusion (MATE) protein family are integral membrane proteins that can move cytotoxic chemicals across cell membranes (Kuroda & Tsuchiya, 2009). Known MATE transporters utilize either the Na^+ or H^+ electrochemical gradient, and fall into the NorM, DinF and eukaryotic subfamilies (Brown *et al.*, 1999). Since MATE proteins can render cells resistant to a wide variety of drugs, they hold great promise as potential therapeutic targets for overcoming multidrug resistance (Fischbach & Walsh, 2009). Despite its medical relevance, the molecular mechanism underlying MATE-mediated multidrug transport is not fully understood.

Molecular structures of MATE transporters are essential for elucidating their transport mechanisms. To date, the X-ray structures of four NorM and DinF transporters have been reported, providing insights into how the MATE transporters interact with drug substrates and how they couple the influx of cations (Na^+ or H^+) to the efflux of structurally and chemically distinct drugs (He *et al.*, 2010; Lu, Symersky *et al.*, 2013; Tanaka *et al.*, 2013; Lu, Radchenko *et al.*, 2013; Radchenko *et al.*, 2015). Among these reported structures is that of a NorM transporter from *Neisseria gonorrhoeae* (NorM-NG), which portrays the MATE transporter in an extracellular-facing, substrate-bound state (Lu, Symersky *et al.*, 2013).

The structure determination of NorM-NG had been hindered by severe defects in the protein crystals, which diffracted X-rays both weakly and anisotropically. To overcome these defects, we identified a crystallization chaperone



called a monobody that bound to NorM-NG with nanomolar affinity and improved its crystallization behavior (Lu, Symersky *et al.*, 2013). We subsequently determined the structure of NorM-NG in complex with the monobody by combining molecular replacement and MIRAS phasing. Although the overall structure of NorM-NG bound to the monobody resembles those of other MATE transporters which were determined in the absence of any crystallization chaperone (He *et al.*, 2010; Tanaka *et al.*, 2013), it remains unclear as to whether the binding of the monobody has altered the structure of NorM-NG.

To address this question, we examined the X-ray diffraction data collected from NorM-NG crystals which were obtained in the absence of any added substrate. Our analysis revealed that the NorM-NG crystals were pseudomerohedrally twinned, with an estimated twin fraction of 0.4–0.5. Despite the challenges posed by the low resolution, data anisotropy and crystal twinning, we determined the NorM-NG structure using molecular replacement and verified the correctness of the structure solution by conducting a difference Fourier analysis. We found that the NorM-NG structure is largely identical to that of NorM-NG in complex with the monobody. Over the course of this work, we have learned a number of lessons in analyzing weak, anisotropic and twinned X-ray diffraction data, which may be valuable for structural biologists who face similar technical challenges when examining crystals with analogous issues.

2. Methods

2.1. Protein expression and purification

NorM-NG was expressed and purified as follows. Briefly, the pET-15b-derived expression vector containing the gene encoding NorM-NG (Lu, Symersky *et al.*, 2013) was introduced into *Escherichia coli* BL21 (DE3) cells, which were grown in Luria–Bertani medium to an optical density of 0.6 at 600 nm and induced with 1 m*M* isopropyl β-D-1-thiogalactopyranoside at 37°C for 3 h. Cells were disrupted using a microfluidizer at 6°C. Cell membranes were collected by ultracentrifugation and were extracted with 1% *n*-dodecyl-β-maltoside (DDM) in 20 m*M* HEPES pH 7.5, 100 m*M* NaCl, 20% glycerol, 1 m*M* tris(2-carboxyethyl)phosphine (TCEP) at 4°C. The soluble fraction was passed onto an Ni–NTA column pre-equilibrated in 20 m*M* HEPES pH 7.5, 100 m*M* NaCl, 0.05% DDM, 20% glycerol, 1 m*M* TCEP at 4°C. The protein was eluted using the same buffer supplemented with 450 m*M* imidazole and was further purified by gel-filtration chromatography at 4°C.

2.2. Protein crystallization and derivatization

NorM-NG was concentrated to ∼10 mg ml^{-1} and dialyzed extensively against 20 m*M* Tris–HCl pH 8.0, 0.05% DDM, 20% glycerol, 1 m*M* TCEP at 4°C. Crystallization of NorM-NG was performed at 22°C using the hanging-drop vapor-diffusion method. The protein sample was mixed with an equal volume (*e.g.* 5 µl + 5 µl) of a crystallization solution consisting of 100 m*M* Tris–HCl pH 9.0, 0.05% DDM, 30% PEG 400. For heavy-atom derivatization, the NorM-NG crystals were incubated with 5–10 m*M* heavy-metal compounds for >2 h at 22°C.

2.3. X-ray diffraction data collection and analysis

Prior to data collection, the NorM-NG crystals were plunged into liquid nitrogen. X-ray diffraction data were collected from the flash-cooled crystals on beamlines 22-ID and 23-ID at Argonne National Laboratory using MAR300 CCD detectors with a sample-to-detector distance of 400 mm. For heavy-metal-treated crystals, fluorescence scans around the metal absorption edges were carried out on the cooled crystals using the automated procedures implemented at the beamlines. X-ray diffraction data were then collected at the wavelengths of the absorption peaks as suggested by the fluorescence scans to maximize the anomalous signals. More than 4000 NorM-NG crystals were screened using robotic crystal handlers and the resulting X-ray diffraction images were examined using *HKL*-2000 (Otwinowski & Minor, 1997) or *MOSFLM* (Leslie *et al.*, 2002). Further X-ray data analysis was performed using the *CCP*4 (Winn *et al.*, 2011) and *PHENIX* (Adams *et al.*, 2010) program suites, unless otherwise specified. Model building and electron-density inspection were conducted using *O* (Jones *et al.*, 1991) or *Coot* (Emsley & Cowtan, 2004). Structure refinement was carried out using *REFMAC* (Murshudov *et al.*, 2011). Figures were prepared with *PyMOL* (DeLano, 2002).

3. Results and discussion

3.1. Data collection and radiation damage

The NorM-NG crystals grew as thin plates with typical dimensions of 200 × 50 × 20 µm. Both streak-seeding and post-crystallization dehydration were attempted but failed to improve the morphology or the diffraction properties of the crystals. Furthermore, substantial differences in the diffraction properties were found among the NorM-NG crystals, even among those from the same batch of the protein sample or the same crystallization drops. As such, many NorM-NG crystals had to be examined at synchrotron beamlines before a data-quality crystal could be identified.

Moreover, a high-brilliance synchrotron-radiation source equipped with undulator devices was essential for obtaining the best possible diffraction data from the NorM-NG crystals, which diffracted X-rays only weakly. The best NorM-NG crystals diffracted to beyond 3.7 Å resolutions, but such 'high-resolution' diffraction features were visible only with sufficient X-ray exposure, typically with a twofold attenuated beam for 1 s. With such a high X-ray dose, the radiation damage suffered by the NorM-NG crystals became an issue (O'Neill *et al.*, 2002). This problem was manifested by the observation that reflections beyond 4.0 Å resolution visually vanished after ∼30 diffraction images (1° oscillation per image) had been collected. Moreover, the later collected reflections often have *B* factors that exceed 20 Å^2 during data reduction, which

is consistent with significant detrimental effects of radiation damage.

To mitigate radiation damage and yet to obtain X-ray diffraction data sets that are as complete as possible, we reduced the X-ray exposure to such a degree that reflections at ∼4.0 Å, rather than 3.7 Å, resolution were visible (with the beam typically attenuated threefold). We mounted the NorM-NG crystals with their longest dimensions aligned perpendicular to the beam and we collected ∼30° wedges of data from each spot on the crystals before severe radiation damage set in. For a typical NorM-NG crystal and a 50 × 50 µm collimated X-ray beam, after each wedge of data was collected we shifted the crystal by ∼80 µm along the oscillation axis to expose fresh regions to X-rays. We subsequently selected the most isomorphous wedges of diffraction data collected from the same or different crystals, and then merged and scaled the diffraction data together using the `.x` files for individual images in *HKL*-2000. On average, each data set examined in this study was derived from 15–20 wedges of data, corresponding to an overall completeness of >80% and a multiplicity of ∼10.

Table 1
Data-collection statistics for NorM-NG.

Data	Native	Mercury derivative	Native
Space group	*P*1	*P*1	*C*2
Unit-cell parameters			
a (Å)	73.9	73.3	74.4
b (Å)	89.1	88.4	163.6
c (Å)	107.9	107.6	108.7
α (°)	84.1	83.2	90.0
β (°)	75.3	74.9	104.9
γ (°)	65.6	65.5	90.0
Resolution (Å)	50–3.8 (3.87–3.80)	50–4.0 (4.07–4.00)	50–3.8 (3.87–3.80)
R_{sym}	0.097 (0.487)	0.119 (0.751)	0.100 (0.448)
$R_{p.i.m.}$	0.032 (0.171)	0.032 (0.153)	0.024 (0.133)
R_{meas}	0.118 (0.459)	0.110 (0.454)	0.118 (0.471)
$\langle I/\sigma(I)\rangle$	18.0 (1.2)	18.8 (1.8)	21.4 (1.7)
Completeness (%)	91.8 (65.8)	86.9 (51.3)	94.8 (66.1)
Multiplicity	11.4 (5.3)	13.1 (9.3)	19.8 (8.4)
No. of unique reflections	22113	17827	11782
Total No. of reflections	250996	1232588	232744
Wilson *B* factor (Å^2)	180	140	140
Twinning operator	$-h, -h + k, -l$	$-h, -h + k, -l$	N/A
Twin fraction from Britton plot	0.35	0.39	N/A

3.2. Data reduction and anisotropy

Diffraction-pattern indexing indicated that the data could accommodate both *P*1 and *C*2 space groups, and subsequent merging and scaling yielded similar statistics for the *P*1 and *C*2 space groups (Table 1), suggesting that the space group was likely to be *C*2. By contrast, processing the diffraction data in any higher-symmetry space group, including an orthorhombic space group, led to an unacceptable level of rejected reflections (>90%) and an exceedingly high R_{sym} (>40%). Assuming a solvent/detergent content of 50% (Matthews, 1968), there would be four and two NorM-NG molecules per asymmetric unit for the *P*1 and *C*2 space groups, respectively. Furthermore, we observed that the NorM-NG crystals diffracted rather anisotropically, *i.e.* they diffracted X-rays to markedly different Bragg spacings along distinct directions.

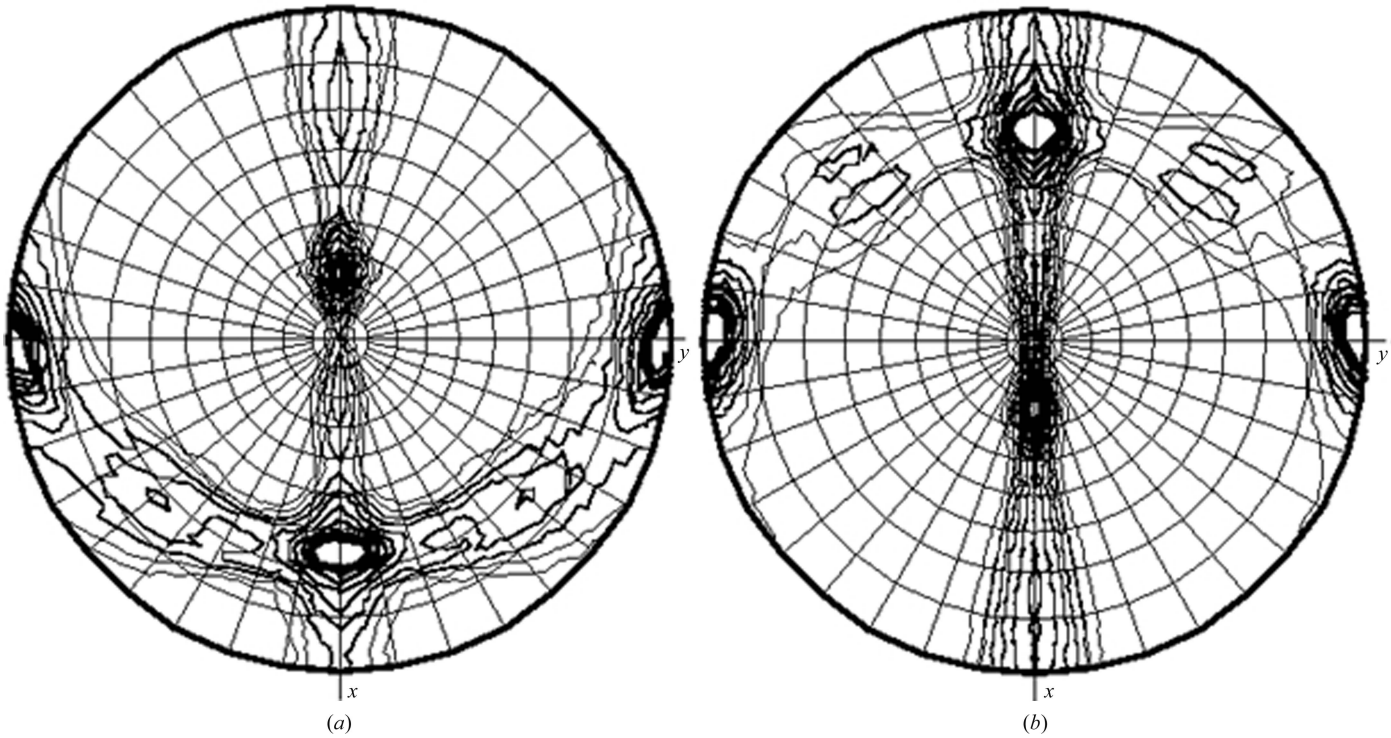


Figure 1
Self-rotation function for NorM-NG. The self-rotation function was calculated using data from 50 to 4.5 Å resolution and an integration radius of 37 Å. Data processed in space group *P*1 were used in (*a*) and those in space group *C*2 in (*b*). Peaks in this section (κ = 180°) suggest the positions of three twofold symmetry axes in both the *P*1 and *C*2 space groups.

The data processed in space group $C2$ extended to ∼3.8 Å resolution along the l direction, but to lower than 5 Å resolution along the h or k directions. By using *Phaser* in *CCP*4, we found that the B factors for data along the best and worst directions differed by >110 Å^2. Such a high level of data anisotropy, although undesirable, is not uncommon for crystals of detergent-purified membrane proteins, including NorM-NG. It is generally thought that the detergent-covered transmembrane domains in integral membrane proteins are incapable of forming strong crystal-packing interactions between neighboring protein molecules, as opposed to extracellular and/or intracellular protein loops, which can make stronger and more specific crystal-packing interactions (Michel, 1991).

Although not unexpected, we reasoned that the data anisotropy seen in the X-ray diffraction would pose several problems for further data analysis. Specifically, it has been demonstrated that data anisotropy can obscure the detection of pathological defects in protein crystals, including twinning (Padilla & Yeates, 2003). Moreover, data anisotropy may exert negative effects in the calculation and inspection of experimentally phased electron-density maps, since the diffraction data are rather noisy and of low resolution along the poorly diffracting directions. Based on such considerations, we initiated our data analysis with molecular replacement, since the structure of NorM-NG bound to a monobody has been determined and since the commonly used software for molecular replacement can account for data anisotropy.

3.3. Twinning detection and analysis

We used the structure of monobody-bound NorM-NG as the search model for molecular replacement using *MOLREP* or *Phaser*, and found only a one-molecule solution for data processed in space group $C2$. This was unexpected since the self-rotation function (Fig. 1) and packing considerations both suggested that more than one NorM-NG molecule is present in the asymmetric unit. In addition, structure refinement in space group $C2$ stalled at a free R factor of >43%. Furthermore, we failed to verify the correctness of this structure solution by conducting a difference Fourier analysis on heavy-atom derivatives (see below). We also noticed that the intensity statistics for X-ray data processed in space group $C2$ suggested that the data are substantially different from those expected for untwinned data, although no twin law could be found in this space group.

To examine the possibility of crystal twinning, we analyzed the data processed in space group $P1$ by using the L-test (Padilla & Yeates, 2003), which has been developed to overcome the difficulties in detecting crystal twinning in anisotropic data. As shown in Fig. 2, the results of the L-test suggested that the NorM-NG data are substantially twinned, and analysis based on the Britton plot indicated that the diffraction data has a twin fraction of 0.35 (Britton, 1972). Furthermore, we identified a twinning operator $(-h, -h + k, -l)$ using *phenix.xtriage* or using *CTRUNCATE* in *CCP*4. As noted above, the L-test for data processed in space group $C2$ also suggested that the data are twinned, implying that the detected twinning is unlikely to be caused by treating crystallographic symmetry as noncrystallographic symmetry (NCS), *i.e.* processing data conforming to the $C2$ symmetry in the reduced symmetry space group $P1$.

Therefore, the NorM-NG crystals seemed to suffer from pseudomerohedral twinning (Yeates, 1997), probably arising from a fortuitous unit-cell geometry (Fig. 3). The $P1$ space group (Table 1) appears to represent the 'true' crystal symmetry, whereas the $C2$ space group seems to correspond to the higher symmetry caused by crystal twinning. It is apparent that the axis of the twinning operator $(-h, -h + k, -l)$ coincides with the twofold symmetry axis in the $C2$ unit cell (Fig. 3), thereby explaining the monoclinic symmetry found during data merging and scaling (Table 1). Furthermore, as previous studies have shown (Yeates, 1997), the twin fraction can vary considerably among crystals and even within the same crystal. However, we had not found any substantial differences in estimated twin fractions in data collected from different NorM-NG crystals, provided that they were analyzed using the same software and resolution cutoffs *etc*.

This result implied that we could not overcome the twinning problem simply by selecting data collected from untwinned crystals or crystals with negligible twinning (*i.e.* with twin fractions of <0.2). In addition, to test whether the detected twinning was introduced by merging data from

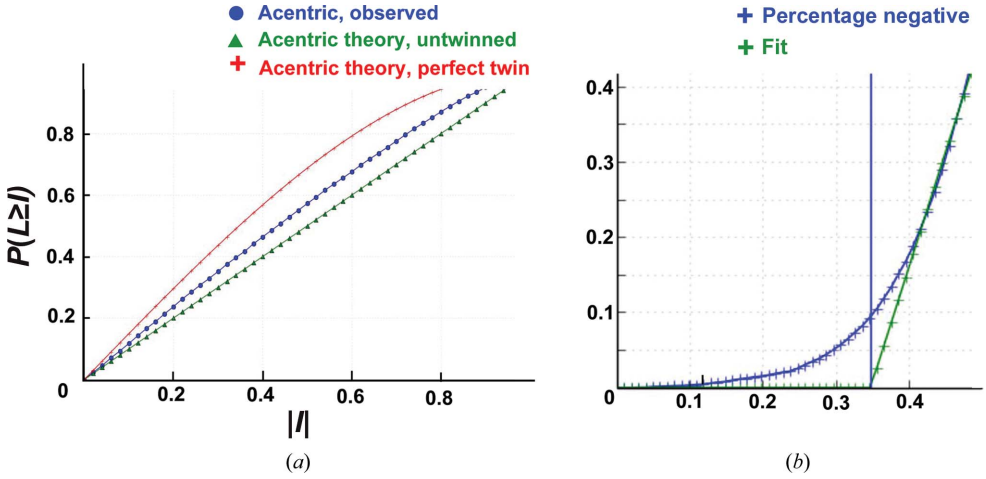


Figure 2
L-test of diffraction data and twin-fraction estimation. (*a*) An *L*-test for twinning was conducted using the native data processed in space group *P*1. The theoretical distribution of acentric data is shown by a green curve; that for perfectly twinned (*i.e.* twin fraction = 0.5) data is shown by a red curve. (*b*) Britton plot showing the fraction of negative intensities after detwinning as a function of the assumed twin fraction. The blue line highlights the estimated twin fraction.

different crystals, we further attenuated the beam (typically by ninefold instead of threefold) and collected 'large' wedges of diffraction data (180° instead of 30°) from NorM-NG crystals. These large wedges were individually processed in space group $P1$ to ∼4 Å resolution and >80% completeness, and all behaved the same way in the twinning analysis (*e.g.* the L-test) as those data sets derived from multiple 'small' wedges. This finding strongly suggested that the detected twinning is intrinsic to the NorM-NG crystals, thus ruling out the possibility that the twinning was inadvertently caused by the merging of multiple wedges of diffraction data.

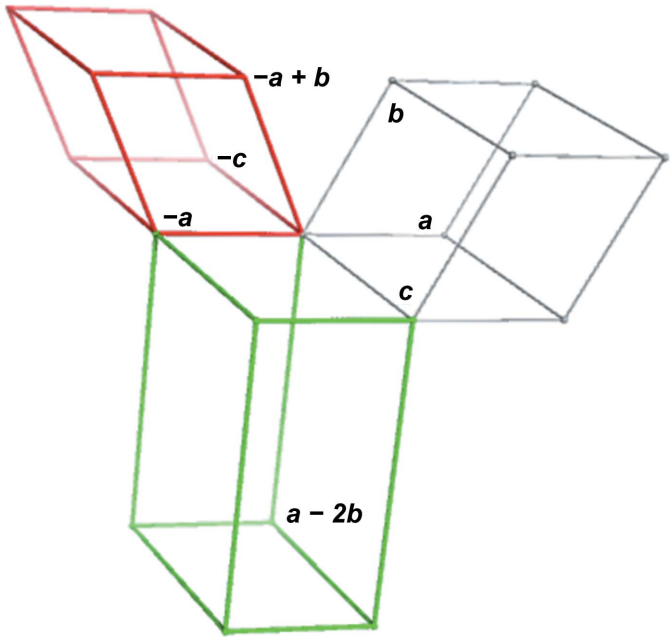


Figure 3
Unit-cell relationship. The $P1$ cell is drawn in gray; the unit-cell vectors are a, b and c, respectively. The twin-related $P1$ cell is in red; its unit-cell vectors are $-a$, $-a + b$ and $-c$, respectively. The $C2$ cell is in green; its unit-cell vectors are $-a$, $a - 2b$ and c, respectively.

3.4. Molecular replacement and structure solution

Taking the crystal twinning into consideration, we pursued molecular replacement further by using the data processed in space group $P1$. We expected to find a correct molecular-replacement solution in the true unit cell; if so, subsequent structure refinement will further verify the crystal twinning. We examined both the uncorrected and twin-corrected data, and using the uncorrected data we found a two-molecule solution in space group $P1$ with an estimated solvent/detergent content of ∼80% and a reasonable crystal-packing arrangement (see below). Those two NorM-NG molecules are related by a twofold screw NCS axis in the asymmetric unit (Fig. 4), which is perpendicular to the axis corresponding to the twinning operator $(-h, -h + k, -l)$. Significantly, this packing arrangement is consistent with the self-rotation function (Fig. 1), in contrast to the one-molecule solution found in space group $C2$.

Structure refinement of the two-molecule solution to 3.8 Å resolution yielded a free R factor of ∼38% and a structure model with reasonable stereochemistry. Further phase improvement *via* density modification, including solvent flattening and twofold NCS averaging, was subsequently performed. Notably, despite the high solvent content of the NorM-NG crystals, solvent flattening only improved the electron-density map slightly. This seemingly unexpected result may be owing to the facts that our data are weak and of modest resolution and the unit cell is relatively small, which all limit the effectiveness of solvent flattening (Wang, 1985; Abrahams & Ban, 2003). The density-modified electron-density map was calculated to 3.8 Å resolution and used for subsequent model rebuilding. The final results of structure refinement with or without using the twinning operator are summarized in Table 2. In both cases, twofold NCS restraints were used throughout structure refinement in order to improve the observables-to-parameters ratio, although the NCS averaging failed to yield a significant improvement of the electron-density map.

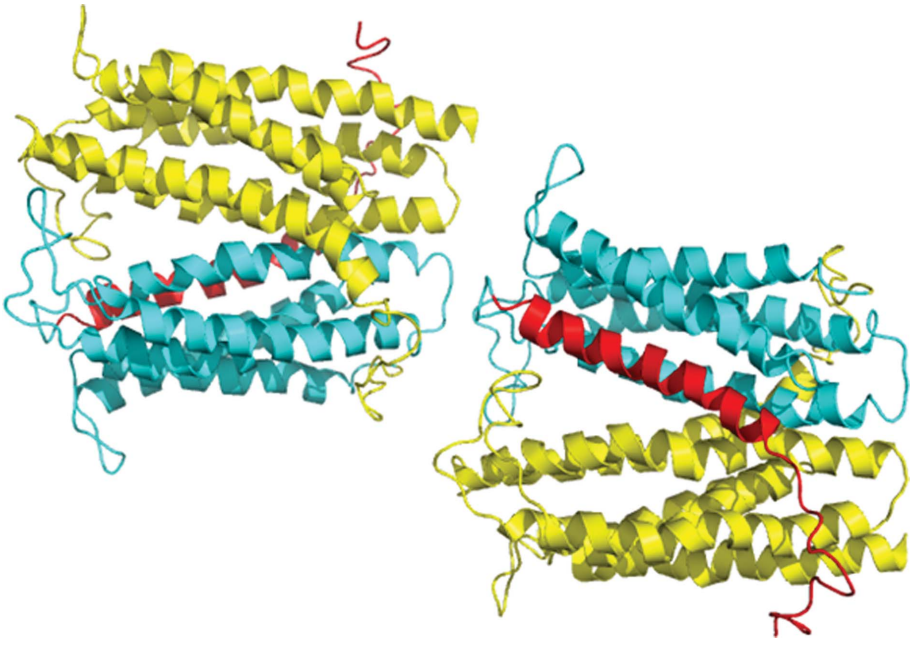


Figure 4
The two-molecule solution found in space group $P1$. The N- and C-terminal domains of NorM-NG are colored cyan and yellow, respectively, except for the first transmembrane helices (TM1), which are colored red.

As shown in Table 2, structure refinement with the twinning operator yielded a free R factor that was >11% lower and a protein model with better stereochemistry compared with that without the twinning operator,

Table 2
Refinement statistics for NorM-NG.

Data	Native	Native
Space group	*P*1	*P*1
Twinning operator	$-h, -h + k, -l$	None
Twin fraction from *REFMAC*	0.50	N/A
Resolution (Å)	20–3.8	20–3.8
No. of reflections	20899	20899
R/R_{free}	0.328 (0.395)/ 0.364 (0.440)	0.435 (0.527)/ 0.482 (0.548)
No. of protein atoms	6852	6852
Average temperature factor (Å^2)	199	175
R.m.s.d., bond lengths (Å)	0.01	0.02
R.m.s.d., bond angles (°)	2.0	2.4
Ramachandran plot allowed (%)	100.0	98.9
Ramachandran plot disallowed (%)	0	1.1

supporting substantial twinning of the NorM-NG crystals. The twin fraction was refined to 0.5, which is consistent with the low R_{sym} obtained for the data processed in space group *C*2 (Table 1). This result also implied that detwinning of the NorM-NG data would be impossible without introducing a substantial amount of additional noise into the data. For instance, detwinning with a twin fraction of 0.48 would increase the estimated intensity error by ∼25-fold (Dauter, 2003). Our results also indicated that estimation of the twin fraction based on the intensity statistics, such as the Britton plot, underestimated the level of twinning of the NorM-NG crystals (Fig. 2). This underestimation may result from the presence of NCS coupled with twinning, measurement errors and/or data anisotropy (Wittmann & Rudolph, 2007).

Furthermore, structure refinement with anisotropically corrected diffraction data using *Phaser* gave rise to structural models with markedly higher free *R* factors and poorer stereochemistry compared with the uncorrected data. Moreover, the electron-density maps calculated with the 'corrected' data were no better than those with the uncorrected data. This result implied that the rescaling of weak, anisotropic and twinned diffraction data could do more damage than good, probably by introducing additional errors and/or improperly treating the twin-related reflections, as reflected in the markedly increased R_{twin} after the anisotropic 'correction' (Lebedev *et al.*, 2006). In parallel, we also attempted to ellipsoidally truncate the anisotropic data using the UCLA twinning server (Strong *et al.*, 2006), but the truncated data only worsened the structure refinement. Given the modest data-to-parameter ratio in this case, it seems likely that the ellipsoidal truncation of the diffraction data reduced the data completeness to such a degree that it offset the benefit of removing weak and non-existent reflections.

Based on the refined model, we conducted an RvR analysis using *phenix.xtriage* to gain further insights into crystal twinning and space-group assignment (Lebedev *et al.*, 2006). This analysis revealed that with respect to the twinning operator $(-h, -h + k, -l)$, R_{twin}^{obs} (0.08–0.27) is consistently lower than R_{twin}^{calc} (0.46–0.52) throughout the resolution range (50–3.8 Å), indicating that severe twinning is indeed associated with this twinning operator and is unlikely to be owing to rotational pseudosymmetry. Moreover, the *P*1 space group is unlikely to be mis-specified (Barends *et al.*, 2005), since in the case of a mis-specified crystal symmetry both R_{twin}^{obs} and R_{twin}^{calc} would be close to 0 (Lebedev *et al.*, 2006). Taken together, our RvR analysis supported the *P*1 space-group assignment and verified the twinning operator $(-h, -h + k, -l)$.

3.5. Difference Fourier analysis

Given the modest quality of our X-ray data, we strove to verify the correctness of the structure solution by conducting an analysis that is independent of structure refinement. To this end, we incubated the NorM-NG crystals with >20 different heavy-metal compounds. We then employed the inverse-beam method to collect X-ray diffraction data near the heavy-metal absorption edges from these crystals, aiming to measure the diffraction intensities of Friedel pairs as accurately as possible. We subsequently used the model-derived or density-modified phases to carry out a difference anomalous Fourier analysis. However, we were unable to identify any strong difference anomalous Fourier peaks that are located near reactive amino acids, for instance Cys residues for mercury compounds.

In parallel, we conducted a difference isomorphous Fourier analysis. For a mercury (thimerosal) derivative, we found four distinct peaks per asymmetric unit at contour levels higher than 7σ in the difference Fourier maps calculated using the model-derived phases (Fig. 5). The same maps did not reveal any other peaks at a >4σ contour level. Among the four strong peaks, two were located in close proximity to the Cys381 residues from the two NorM-NG molecules, whereas the other two peaks were near residues Ala264, Ile287 and Ser290, making little chemical sense. However, after the fractional coordinates of the two twin-related NorM-NG molecules had been generated by applying the transpose of the inverse of the matrix corresponding to the twinning operator, the two 'inexplicable' difference Fourier peaks were placed into close proximity to the two Cys381 residues within the twin-related NorM-NG molecules (Fig. 5).

A plausible explanation for the twin-related difference Fourier peaks is that the size of the twin domains in the NorM-NG crystals is similar to the X-ray beam coherence length, which enabled the X-rays waves scattered from the two domains to interfere (Yeates, 1997). As such, the observed intensities are derived from the complex average of the scattering factors from each twin domain, and it is thus conceivable that the difference isomorphous Fourier maps showed features for mercury binding in both twin domains. In accord with this scenario, we observed several, but not all, cross-peaks between the mercury-binding sites within the two twin domains in the difference isomorphous Patterson maps (see below), implying that the size of the twin domains in the NorM-NG crystals indeed approaches the beam coherence length (Yeates, 1997).

Notably, similar observations in difference isomorphous Fourier analysis have been made by others, although the presence of additional difference Fourier peaks was attributed to similar molecular packing between the twin domains (Yuan *et al.*, 2003). Furthermore, besides crystal twinning, static

disorder can also give rise to additional, 'inexplicable' difference Fourier peaks (Shah & Brunger, 1999). Static disorder can in principle be distinguished from crystal twinning based on the diffraction intensity statistics, because these two defects affect crystal diffraction rather differently. As described above, since the results from the *L*-test and RvR plot are both suggestive of crystal twinning in our case, we argue that the NorM-NG crystals suffered from pseudomerohedral twinning rather than static disorder.

Significantly, the binding of the mercuric compound to Cys381 in NorM-NG, as observed in the monobody-bound crystal form (Lu, Symersky *et al.*, 2013), can also explain the difference isomorphous Patterson maps calculated for the mercury derivative. In these maps, we found not only the peaks corresponding to the mercury-binding sites within the same twin domains (Fig. 6), but also some of the cross-peaks for the binding sites between the twin domains, which is consistent with our difference isomorphous Fourier maps (Fig. 5). Taken together, the identification of the mercury-binding sites in NorM-NG using model-derived phases verified the correctness of our structure solution and further confirmed the crystal twinning in space group *P*1. By contrast, we were unable to identify any heavy-metal binding site by using difference Fourier analysis in space group *C*2, suggesting that the correct space group is *P*1, not *C*2.

3.6. Experimental phasing attempts

Because unbiased experimental phases would be useful for uncovering potential structural differences between NorM-NG and its monobody-bound form, we set out to calculate the SIR phases for the mercury derivative using either the uncorrected or detwinned diffraction data. We detwinned both the native and derivative data by using *DETWIN* in

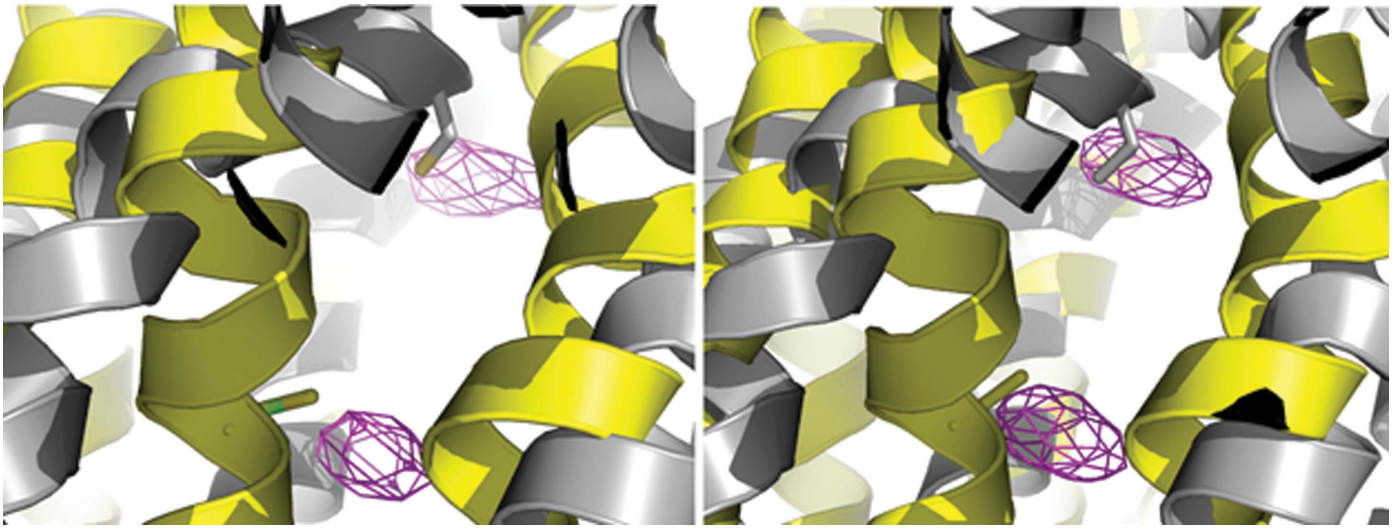


Figure 5
Difference isomorphous Fourier maps. The twin-related mercury-binding sites in the two NorM-NG molecules (yellow) in the asymmetric unit are highlighted by the difference Fourier peaks (magenta mesh; 7σ contour level). The twin-related NorM-NG molecules are colored gray. Cys381 residues are drawn as stick models. Data from 15 to 3.8 Å resolution were used for map calculation.

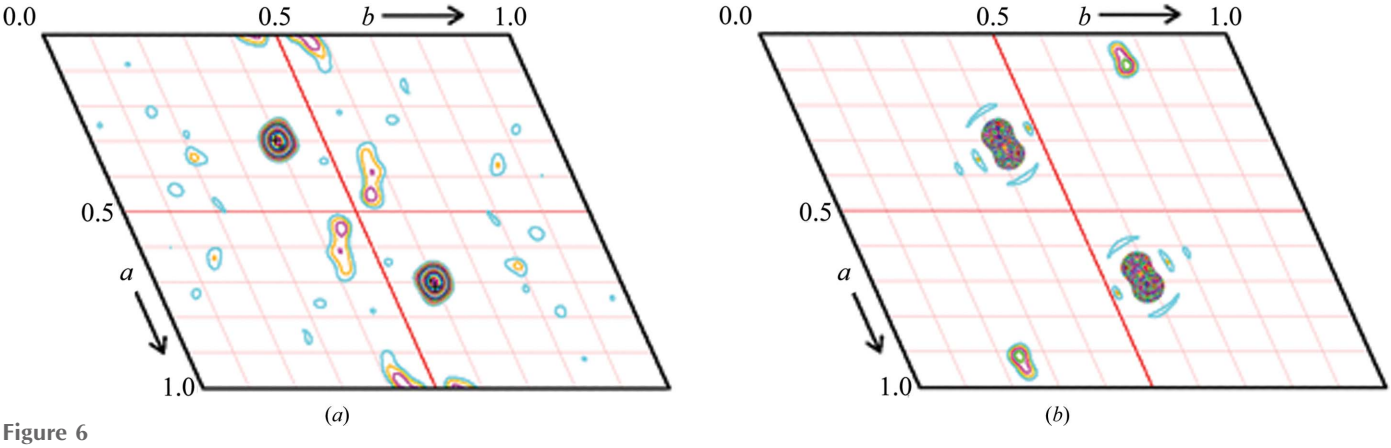


Figure 6
Difference isomorphous Patterson maps. (*a*) Observed difference Patterson maps for the mercury derivative at w = 0.5; data were processed in space group *P*1. The peaks (16.6σ) are the superposition of the cross-peaks between the two Hg atoms within the two twin domains. Cross-peaks in one twin domain are highlighted by '+' and those in the second domain are marked by '*'. (*b*) Predicted difference Patterson based on the coordinates of the mercury-binding sites. This figure was prepared using *XPREP* (Bruker); data from 15 to 4 Å resolution were used for this analysis.

*CCP*4 and used various twin fractions ranging from 0.1 to 0.4 with an increment of 0.1. Although the peak heights for the twin-related peaks gradually decreased in the difference isomorphous Fourier maps as we increased the twin fractions to 0.4, structure refinement against the 'detwinned' data always yielded twin fractions greater than 0.4, probably reflecting the difficulty in correcting the twinned data computationally.

Furthermore, the refinement of heavy-atom positions and occupancies using *SHARP* (de La Fortelle & Bricogne, 1997) fared better using the uncorrected diffraction data than the computationally 'detwinned' data, which are likely to be more noisy than the uncorrected data. Notably, the solvent-flattened SIR maps calculated at 6 Å resolution without twofold NCS averaging revealed a clear protein–solvent boundary and no additional protein molecules in the asymmetric unit other than the two known molecules (Fig. 4). The quality of the experimentally phased electron-density maps, however, was insufficient for model building even after extensive density-modification procedures including solvent flattening and NCS averaging. Inclusion of the experimental phases also failed to improve the structure refinement. As we only found one heavy-metal derivative, we did not pursue experimental phasing any further, since it has been suggested that at least four different heavy-metal derivatives are necessary for MIR phasing using highly twinned X-ray data (Yeates & Rees, 1987).

3.7. Comparison of the NorM-NG structures

The final structure model includes residues 8–455. Within the crystals, the neighboring NorM-NG molecules are packed against each other in a 'head-to-tail' fashion *via* interactions between the intracellular and extracellular loops, particularly along the *c* direction (Fig. 7). By contrast, the largely hydrophobic surface of the transmembrane domain in NorM-NG, which is presumably covered by detergents and/or lipids, plays little role in forming crystal-packing interactions.

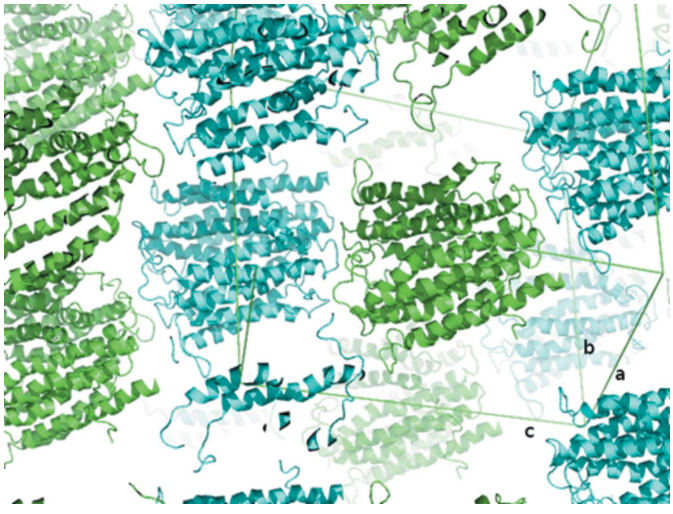


Figure 7
Crystal packing of NorM-NG. One molecule in the asymmetric unit is colored green and the other cyan. The unit-cell axes in space group *P*1 are highlighted by '*a*', '*b*' and '*c*'.

This crystal-packing arrangement helps to explain why the diffraction of the NorM-NG crystals extended to higher resolution along the *l* direction, as opposed to the *h* or *k* directions, since the neighboring NorM-NG molecules form more extensive packing interactions along the *c* direction, which approximately corresponds to the *l* direction in reciprocal space. Furthermore, a similar crystal-packing arrangement was seen for another NorM transporter from *V. cholerae* (NorM-VC), which was also crystallized in a detergent-purified form (He *et al.*, 2010). However, the NorM-VC molecules were packed in a 'tail-to-tail' or 'head-to-head' manner in the crystals, rather than a 'head-to-tail' fashion as seen in NorM-NG.

As observed in the monobody-bound crystal form (Lu, Symersky *et al.*, 2013), NorM-NG adopts an overall V-shaped, extracellular-facing conformation. The C-terminal residues 456–463 in NorM-NG, which were found to make close contacts with the monobody, were unresolved in the current structure and are likely to be disordered. The two NorM-NG molecules in the asymmetric unit have largely the same structure, with an r.m.s. deviation of ∼0.3 Å for 448 C^{α} positions. The structure of NorM-NG is also similar to that of the monobody-bound form, with an r.m.s. deviation of ∼0.6 Å for 448 C^{α} positions. Therefore, although the NorM-NG structure has only been refined to 3.8 Å resolution, we argue that the monobody has not altered the transporter structure to any significant extent. Indeed, composite OMIT maps revealed no major discrepancy between the refined NorM-NG model and electron density.

3.8. Ligand-binding site

In the structure of monobody-bound NorM-NG, we observed a conspicuous electron-density feature in the substrate-binding site, which was attributed to an as yet unidentified ligand (Lu, Symersky *et al.*, 2013). However, no such density feature was found in the electron-density maps calculated for NorM-NG. Despite this difference, the two extracellular protein loops that cap the substrate-binding cavity have not altered their conformations in NorM-NG compared with their counterparts in the monobody-bound structure. Our data thus raised the question of why NorM-NG remains in its substrate-bound state in the absence of any bound ligand.

The answer to this question may be found in at least three mutually non-exclusive factors. Firstly, the crystallization conditions for NorM-NG and its monobody-bound complex are different. It is plausible that under the conditions in which NorM-NG was crystallized the binding of the unidentified ligand was suppressed to such a level that it was no longer crystallographically detectable owing to low occupancy. Secondly, the diffraction data for NorM-NG are much weaker, more anisotropic and of lower resolution compared with those for the monobody-bound complex. Furthermore, well determined experimental phases were used to calculate the

electron-density maps for the monobody-bound NorM-NG (Lu, Symersky *et al.*, 2013). As such, the electron density for the unidentified ligand in NorM-NG might be 'invisible' simply because of the limited quality of the diffraction data, *e.g.* modest resolution and/or phase bias from the protein-only model. Thirdly, since the extracellular loops that cap the substrate-binding site are intimately involved in the crystal-packing interactions (Fig. 7), it seems likely that the ligand-bound conformation of NorM-NG was stabilized by crystal-packing forces.

4. Concluding remarks

The structure determination of NorM-NG presents a particularly difficult case because the crystals diffracted X-rays weakly and anisotropically and suffered from severe twinning defects. Despite these issues, we were able to obtain a structure solution thanks to the availability of a very similar structure model. During the course of this study, we also attempted to use other protein models for molecular replacement, including the published NorM-VC structure (He *et al.*, 2010), but without success. This failure may in part be owing to the amino-acid sequence and structural differences between NorM-NG and NorM-VC, which were reflected in ∼33% sequence identity and an r.m.s. deviation of >5 Å for 440 common C^{α} positions, or >3 Å for 300 C^{α} positions if only the membrane-spanning helices were considered. Indeed, previous studies have suggested that molecular replacement would become increasingly difficult once the r.m.s. deviation between the search model and target structure exceeds 2 Å (Terwilliger *et al.*, 2012).

Another challenge in this structure determination stemmed from the fact that the diffraction data were available only to moderate resolution. At ∼4 Å resolution the ratio between the observables and the parameters is so low that structure refinement becomes less powerful in improving the protein model and it is easy to fall victim to model bias during map interpretation. Such problems can be mitigated or rectified if high-number NCS restraints and/or accurate experimental phases are available, as they can be included during structure refinement to enhance the observables to parameters ratio (Brunger *et al.*, 2009). Furthermore, unbiased experimental maps are extremely useful for revealing the subtle differences between the search models and target structures at moderate resolutions, especially if molecular replacement is initially used to solve the phase problem, *i.e.* to overcome phase bias (Lu, Radchenko *et al.*, 2013).

In the case of NorM-NG we only had twofold NCS and poor SIR phases, which made a modest and no contribution to improving the structure refinement, respectively. As such, it was essential for us to validate the structure solution by means of difference Fourier analysis. Despite the fact that the NorM-NG crystals were highly twinned, we were able to use model-derived phases to identify a mercury derivative and locate the heavy-atom binding sites. Although the quality of the resulting SIR phases was insufficient for model building or improving structure refinement, we could explain both the difference isomorphous Fourier and Patterson maps. Therefore, in situations where only low-resolution, anisotropic and twinned X-ray diffraction data are attainable, we argue that combining molecular replacement with difference Fourier analysis may still lead to a correct structure solution, provided that a similar structure model and a heavy-metal derivative are both available.

Acknowledgements

We thank the beamline staff at GM/CA-CAT (23-ID) and SER-CAT (22-ID) of Argonne National Laboratory for assistance during X-ray data collection. We also thank T. Yeates (UCLA) and A. Leslie (MRC) for discussions, and the anonymous reviewers for their insightful suggestions. This work was supported in part by the National Institutes of Health (R01-GM094195 to ML).

References

Abrahams, J. P. & Ban, N. (2003). *Methods Enzymol.* **374**, 163–188.
Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Barends, T. R. M., de Jong, R. M., van Straaten, K. E., Thunnissen, A.-M. W. H. & Dijkstra, B. W. (2005). *Acta Cryst.* D**61**, 613–621.
Britton, D. (1972). *Acta Cryst.* A**28**, 296–297.
Brown, M. H., Paulsen, I. T. & Skurray, R. A. (1999). *Mol. Microbiol.* **31**, 394–395.
Brunger, A. T., DeLaBarre, B., Davies, J. M. & Weis, W. I. (2009). *Acta Cryst.* D**65**, 128–133.
Dauter, Z. (2003). *Acta Cryst.* D**59**, 2004–2016.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Fischbach, M. A. & Walsh, C. T. (2009). *Science*, **325**, 1089–1093.
He, X., Szewczyk, P., Karyakin, A., Evin, M., Hong, W.-X., Zhang, Q. & Chang, G. (2010). *Nature (London)*, **467**, 991–994.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kuroda, T. & Tsuchiya, T. (2009). *Biochim. Biophys. Acta*, **1794**, 763–768.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Lebedev, A. A., Vagin, A. A. & Murshudov, G. N. (2006). *Acta Cryst.* D**62**, 83–95.
Leslie, A. G. W., Powell, H. R., Winter, G., Svensson, O., Spruce, D., McSweeney, S., Love, D., Kinder, S., Duke, E. & Nave, C. (2002). *Acta Cryst.* D**58**, 1924–1928.
Lu, M., Radchenko, M., Symersky, J., Nie, R. & Guo, Y. (2013). *Nature Struct. Mol. Biol.* **20**, 1310–1317.
Lu, M., Symersky, J., Radchenko, M., Koide, A., Guo, Y., Nie, R. & Koide, S. (2013). *Proc. Natl Acad. Sci. USA*, **110**, 2099–2104.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Michel, H. (1991). Editor. *Crystallization of Membrane Proteins*, pp. 73–88. Boca Raton: CRC Press.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
O'Neill, P., Stevens, D. L. & Garman, E. (2002). *J. Synchrotron Rad.* **9**, 329–332.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* D**59**, 1124–1130.
Radchenko, M., Symersky, J., Nie, R. & Lu, M. (2015). *Nature Commun.* **6**, 7995.
Shah, S. A. & Brunger, A. T. (1999). *J. Mol. Biol.* **285**, 1577–1588.
Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D. & Eisenberg, D. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 8060–8065.

Tanaka, Y., Hipolito, C. J., Maturana, A. D., Ito, K., Kuroda, T., Higuchi, T., Katoh, T., Kato, H. E., Hattori, M., Kumazaki, K., Tsukazaki, T., Ishitani, R., Suga, H. & Nureki, O. (2013). *Nature (London)*, **496**, 247–251.
Terwilliger, T. C., Read, R. J., Adams, P. D., Brunger, A. T., Afonine, P. V., Grosse-Kunstleve, R. W. & Hung, L.-W. (2012). *Acta Cryst.* D**68**, 861–870.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wittmann, J. G. & Rudolph, M. G. (2007). *Acta Cryst.* D**63**, 744–749.
Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.
Yeates, T. O. & Rees, D. C. (1987). *Acta Cryst.* A**43**, 30–36.
Yuan, Y.-R., Martsinkevich, O. & Hunt, J. F. (2003). *Acta Cryst.* D**59**, 225–238.