Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure determination using poorly diffracting membrane-protein crystals: the H^+-ATPase and Na^+,K^+-ATPase case history

Bjørn P. Pedersen, J. Preben Morth and Poul Nissen*

Centre for Membrane Pumps in Cells and Disease, Danish National Research Foundation, Aarhus University, Department of Molecular Biology, Gustav Wieds Vej 10C, DK 8000 Aarhus, Denmark

Correspondence e-mail: pn@mb.au.dk

Received 28 September 2009
Accepted 10 December 2009

An approach is presented for the structure determination of membrane proteins on the basis of poorly diffracting crystals which exploits molecular replacement for heavy-atom site identification at 6–9 Å maximum resolution and improvement of the heavy-atom-derived phases by multi-crystal averaging using quasi-isomorphous data sets. The multi-crystal averaging procedure allows real-space density averaging followed by phase combination between non-isomorphous native data sets to exploit crystal-to-crystal nonisomorphism despite the crystals belonging to the same space group. This approach has been used in the structure determination of H^+-ATPase and Na^+,K^+-ATPase using Ca^{2+}-ATPase models and its successful application to the Mhp1 symporter using LeuT as a search model is demonstrated.

1. Introduction

Membrane-protein crystals often display weak and anisotropic diffraction properties owing to inherent flexibility and disordered detergent and lipid components. New generations of synchrotron sources allow difficult targets such as membrane proteins and large complexes to be approached by low-resolution crystallography based on poorly diffracting crystals. When high-accuracy atomic models are not readliy available, low-resolution X-ray crystallography still offers significant information on the three-dimensional structure (Brunger *et al.*, 2009). Experimental electron-density maps should also be appreciated as important end-results, as demonstrated by the fatty-acid synthase complexes (Jenni *et al.*, 2006; Maier *et al.*, 2006). In these cases model-unbiased heavy-atom-derived phasing therefore becomes critical. Rationales for the derivatization of membrane-protein crystals are available (Morth *et al.*, 2006), but heavy-atom site identification and phasing is obstructed when only weak diffraction data of poor quality are available and is further complicated by, for example, radiation damage and severe anisotropy. Scaling and combined use of data sets may also be significantly hampered by non-isomorphism between individual crystals. Finally, phase refinement by density modification is nontrivial at low resolution where solvent-mask definition and histogram matching are inaccurate, while reciprocal-space refinement of atomic models is complicated by poor diffraction data and systematic variations in the solvent region introduced by detergents and lipids that impede bulk-solvent correction.

Based on our experience with the structure determination of two ion pumps (representing two very different approaches in purification and crystallization), we have established a general low-resolution methodology for heavy-atom site identification and phase extension that yields high-quality electron-density maps for molecular analysis. These structures were (i) *Arabidopsis thaliana* auto-inhibited H^+-ATPase isoform 2 (AHA2), which was heterologously overexpressed in yeast and crystallized in detergent micelles (Pedersen *et al.*, 2007), and (ii) the Na^+,K^+-ATPase enzyme (NKA) isolated from pig kidney and crystallized in stacked bilayers (Morth *et al.*, 2007). Further details are provided as supplementary information.[1] For both NKA

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: KW5018). Services for accessing this material are described at the back of the journal.

and AHA2 the low-resolution diffraction intensity was strong; the diffraction intensity then displayed a minimum at 6–7 Å resolution as generally observed for proteins (Svergun *et al.*, 2001; Popov & Bourenkov, 2003), followed by weak diffraction intensity at higher resolution, reaching a maximum attainable resolution of 3.5 Å by data collection at a third-generation synchrotron source (Supplementary Fig. 1). AHA2 also displayed severe anisotropy, only yielding usable diffraction data at 3.6 *versus* 5.5 Å maximum resolution on extensive exposure (Supplementary Fig. 2). These are frequently observed characteristics of membrane-protein crystals and are likely to be the result of semi-ordered packing owing to detergent/lipid-solubilized regions hindering crystal contacts.

2. Determination of the heavy-atom substructure

The AHA2 and NKA crystals were both highly fragile and sensitive to manipulation, resulting in only low-resolution heavy-atom derivative data sets being available for phasing (6–9 Å maximum resolution). The few data sets of acceptable quality that were obtained from screening hundreds of crystals were practically non-isomorphous, with merging *R* factors on intensities of ∼20–30% even at low resolution (Supplementary Tables 1–3). As a result of the weak diffraction properties (which prompted high-dose data collection), most data sets also displayed signs of radiation damage. Despite indications of heavy-atom signal in the derivative data sets, exhaustive efforts to identify heavy-atom sites failed without prior phase information.

We therefore turned to molecular replacement to obtain initial phases and facilitate heavy-atom (HA) site identification by difference Fourier maps.

Molecular replacement (MR) is typically considered for structure determination when a search model with >25% sequence identity (Taylor, 2003) combined with reasonable-quality resolution data (4–5 Å or better) for the unknown structure are available. One caveat of MR is that as the resolution becomes lower the MR model bias of the map becomes higher, to the point where new structural features cannot be identified (DeLaBarre & Brunger, 2006). MR can be successful for α-helical membrane proteins using low-homology search models and very low resolution (Strop *et al.*, 2007) because the large proportion of mass arranged as nearly parallel transmembrane helices provides strong and fold-specific signals at low resolution. While being extremely model-biased, phases from such a low-resolution molecular-replacement solution could be sufficient to identify HA scatterers in derivative crystals, in particular using anomalous difference Fourier maps.

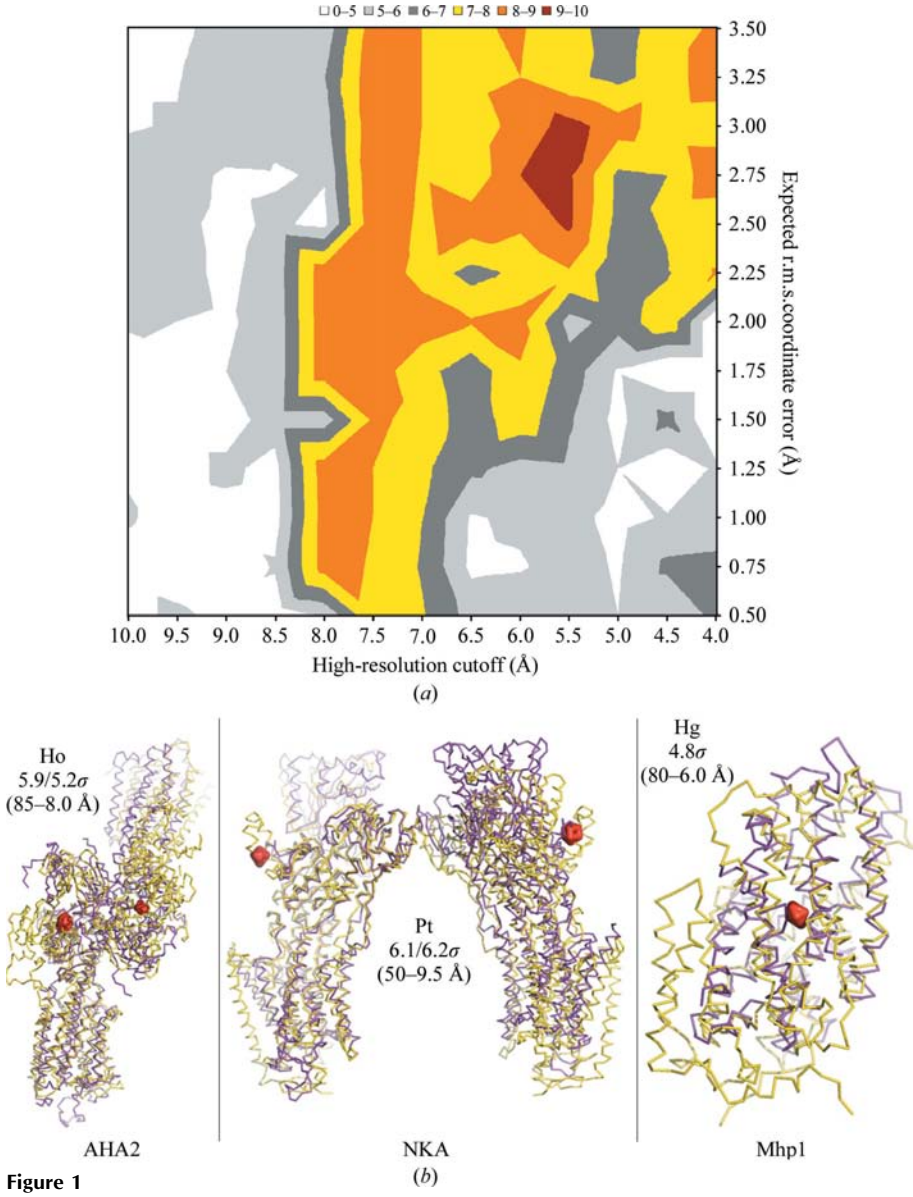

Figure 1
Z score of molecular-replacement solutions of AHA2 and the identification of HA sites. (*a*) The final top-scoring translational *Z* score of molecular replacement of AHA2 is shown as a function of estimated r.m.s. coordinate error and high-resolution cutoff. The coloured regions (yellow, orange and red) indicate the borders within which a correct MR solution was found. As expected, high values of the *Z* score correlate with the identification of a correct solution. The clearest solution was found using a cutoff of 5.5 and an r.m.s. of 3.0 (*Z* score of 9.4). Note that the use of all data (to 4 Å) and default values of the r.m.s. coordinate error from 0.8 to 1.5 Å would not result in a correct MR solution. (*b*) The MR solutions described in Table 1 are shown in ribbon representation, with the search model used for MR in magenta and the published model in yellow (all models were prepared using *PyMOL*; DeLano, 2002). The red mesh shows the position of the HA peaks in each of the three cases (the HA type, peak height and anomalous map cutoff are noted). Peaks are related by local symmetry in AHA2 and NKA, serving as a further validation.

MR solutions were found for both AHA2 and NKA using Ca^{2+}-ATPase as a low-homology search model (Table 1), as identified by *Z* score and proper crystal packing. It was of critical importance to use native data sets collected with particular attention to the redundancy and completeness of low-resolution diffraction data, as has been observed previously (*e.g.* Meyer *et al.*, 2006) and as expected from the formulation of the maximum-likelihood target in molecular replacement (Bricogne, 1992; McCoy, 2004). An example of the MR parameters used can be found in the supplementary material. The quality of the obtained molecular-replacement phases was low. For instance, the map correlation coefficient to the final model in the case

Table 1
Summary of MR solutions.

The table summarizes the input data and shows the final optimized values for the molecular replacement and derived *Z* scores (see also Fig. 1).

Protein	AHA2 (3b8c)	NKA (3kdp)	Mhp1 (2jlo)
Input			
Search model	Ca^{2+}-ATPase (1t5s)	Ca^{2+}-ATPase (1xp5)	LeuT (3gjd)
Modification	Loops removed, no A domain	No modification	Loops removed, polyalanine model
R.m.s. deviation† (Å)	4.4 (885 C^{α} atoms)	2.4 (1344 C^{α} atoms)	3.9 (254 C^{α} atoms)
Sequence identity before/after‡ (%)	25/27	30/30	18/(22)§
Coverage in MW¶	78.9/97.4 (81%)	109.5/148.0 (74%)	23.7/50.6 (47%)
Molecules in ASU	2	2	1
MR			
R.m.s. of MR search††	3.0	2.0	3.0
High-resolution cutoff	5.5	6.0	7.0
Z score of solution	9.4	10.9	5.4
LLG score of solution	110	209	16
HA peaks (σ)	5.95/5.23	6.11/6.15	4.8

† R.m.s. deviation gives the r.m.s. deviation between the search model and the final refined model. ‡ The sequence identity before and after modification of the search model (No. of identical residues/No. of residues in shortest sequence, given as a percentage). § The sequence identity of the search model was calculated before conversion to polyalanine. ¶ Coverage of the target by the search model (in kDa). †† R.m.s. of MR search lists the estimated r.m.s. coordinate error used as input to *Phaser*.

of AHA2 was 0.36 for main-chain atoms and 0.13 for side-chain atoms. However, the reason for not using these phases directly was not their poor quality but rather that the resultant map would exhibit model bias obstructing the identification of new structural features.

Correct MR solutions appeared from a combination of large values for the estimated root-mean-square coordinate error (r.m.s.) of the search model as implemented in *Phaser* (McCoy *et al.*, 2007) and an appropriate high-resolution cutoff. High r.m.s. values make the MR likelihood function relatively insensitive to high-resolution data, relaxing the need for a high-resolution cutoff (Fig. 1), while underestimation of the r.m.s. results in negative log-likelihood gain scores even for correct solutions (Supplementary Fig. 3). Conversely, a negative LLG score can be circumvented by a proper setting of the r.m.s. value, facilitating the identification of a reasonable starting value of the r.m.s. (compare Fig. 1*a* and Supplementary Fig. 3).

For AHA2 trimming of the search model by the removal of non-conserved loop regions was necessary and several Ca^{2+}-ATPase models representing different conformational states were tested for both AHA2 and NKA. Search models formed by ensembles of superpositioned Ca^{2+}-ATPase structures (representing all available states) were not tested as it seemed clear that a particular functional state would form the optimal match, as was indeed observed. Heavy-atom sites were identified in both AHA2 and NKA by anomalous difference Fourier maps using phases derived from these MR solutions (Fig. 1). As a further validation of the identification of the MR-phased site, peaks in the anomalous difference maps were confirmed by their relation through noncrystallographic symmetry (NCS) operators (Fig. 1).

To generalize the low-resolution MR approach we have tested it on an unrelated α-helical membrane-protein structure, the Mhp1 symporter (Weyand *et al.*, 2008), which is part of a large group of evolutionarily related membrane transporters that all display a LeuT fold (Abramson & Wright, 2009). Representative structures show different conformations alternating between occluded, inward-facing and outward-facing conformations of the substrate-binding site. Trying various combinations of models to find a molecular-replacement solution for the Mhp1 data, we succeeded using a trimmed model of LeuT which, like Mhp1, adopts an outward-facing conformation (Yamashita *et al.*, 2005). Using the LeuT-based MR phases (*i.e.* from a model containing less than 1600 of the 3570 atoms found in Mhp1), we were able to identify the Hg site in the ethylmercury thiosalicylate derivative used by Weyand and coworkers for phasing in combination with a Pt derivative (Weyand *et al.*, 2008).

In all of the three cases described here a solution was found using a high-resolution cutoff in the range 5.5–7 Å and an r.m.s. in the range 2.0–3.0 Å (Fig. 1). The requirement for a high-resolution cutoff in the 5.5–7 Å range is in part caused by poor data quality, with the intensities displaying a minimum in this range (*cf.* Supplementary Table 4 and Supplementary Fig. 1).

3. Density modification

Following HA-derived phasing, density modification is of critical importance. Multi-crystal averaging has been recognized as a powerful tool in the determination of low-resolution structures when several crystal forms are available (Ban *et al.*, 1999; Faham *et al.*, 2008; Chen *et al.*, 2005). A few cases have reported the use of same-form (*i.e.* quasi-isomorphous) crystals, where all crystals have identical space group and unit-cell parameters (Maeda *et al.*, 2009; Weyand *et al.*, 2008; Morth *et al.*, 2007; Pedersen *et al.*, 2007; Lescar *et al.*, 2001; Jeruzalmi *et al.*, 2001). Low-resolution data sets derived from crystals with a large solvent content often display poor crystal-to-crystal isomorphism that impedes scaling and phase combination, but multi-crystal averaging may then be useful, as was indeed the case for AHA2 and NKA and as was recently reported for the gap junction channel (Suga *et al.*, 2009).

After experimental phasing in *SHARP* (de La Fortelle & Bricogne, 1997; Supplementary Fig. 4), the *DMMULTI* (Cowtan, 1994) procedure was initiated with MR-derived transformation matrices (later on from updated models) and applied to unmodified amplitudes and HA-derived phases. Initial masks were based on the MR-derived models where excluded regions had been re-inserted to complete the mask. Later masks were recalculated from updated models. Details of how to obtain transformation matrices and generate masks are given in the supplementary material. The phases were extended to the 3.5–3.6 Å high-resolution limit (Fig. 2 and Supplementary Figs. 5 and 6). Phase extension worked best when initiated from a conservative starting point where phases display an average figure of merit of >0.25 (Supplementary Fig. 7 and Supplementary Table 4), followed by slow gradual steps of increased resolution with more than 100 cycles starting from 8–12 Å maximum resolution (Keller *et al.*, 2006). An example of the *DMMULTI* parameters used can be seen in the supplementary material. Multi-crystal averaging from same-form crystals indeed improves phase refinement, as indicated by the continuous features of the main-chain fold and its map correlation coefficient, CC_{main}, which increased from 0.58 to 0.71 for AHA2 (Fig. 2 and Supplementary Fig. 5). Application of real-space density averaging between phased non-isomorphous data sets also provides a small improvement over reciprocal-space phase combination, as reflected by a further increased CC_{main} (0.74 *versus* 0.71) and by the electron density showing more complete features in loop regions.

The precision of transformation matrices is critical to averaging and they are conveniently derived from MR models. Combinations of

multi-crystal and NCS averaging using both whole-monomer and domain masks should be tested. The AHA2 monomer was covered by a single mask, while the NKA monomer was divided into four separate domain masks.

Weak data are favourably included when a maximum-likelihood target is used in structural refinement (McCoy, 2004). The *R* factors of merged reflections are poor-quality measures of weak structure-factor amplitudes obtained by redundant data collection, especially in

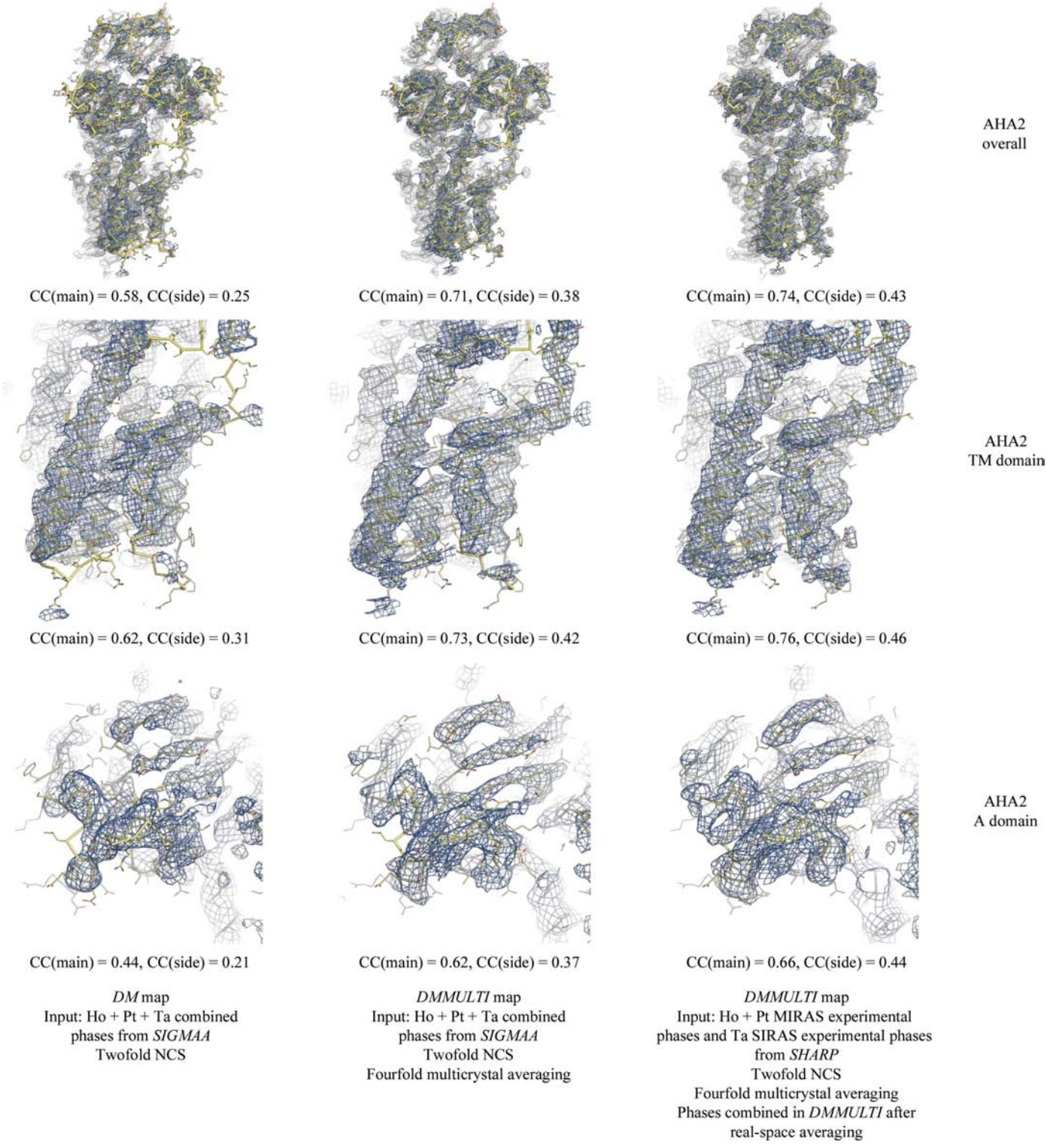

Figure 2
Improvements in the electron density of AHA2. A sequential look at the improvement in the experimental electron-density maps (1σ) of the H^+-ATPase monomer, transmembrane domain (TM) and actuator domain (A) as phases are improved by multi-crystal averaging (from column 1 to 2) and by phase combination after real-space density averaging (from column 2 to 3). The correlation coefficient is listed for main-chain atom density and for side-chain atom density compared with the published model (PDB entry 3b8c).

the case of anisotropic data, where the spherical resolution shells used to calculate statistics will invariably include both weak and strong reflections along the bad and good axes, respectively. In most cases this is nicely corrected by the anisotropic diffraction correction step employed in *Phaser* and all refinement programs (Murshudov *et al.*, 1998), but in extreme cases ellipsoidal truncation of the data might be beneficial [and is possible in *MOSFLM* (Leslie, 1992) and the *Anisotropic Diffraction Server* (Strong *et al.*, 2006)]. That the merging R factors are poor-quality measures is clear in the highly anisotropic AHA2 data sets, where merging R factors far exceed 100% in the higher resolution bins owing to the inclusion of a large number of very weak diffraction features along with few stronger reflections. Here, the data were nevertheless included in refinement and map calculations because this helped to reveal new features in the experimental maps. With a reliable model available, a proper data cutoff can be based on electron-density map correlations and model-based $R_{\rm free}$ values in thin resolution shells can be used if reciprocal-space refinement is feasible. Furthermore, sharpening of B factors has been observed to improve side-chain density on several occasions and should be considered for map calculations (DeLaBarre & Brunger, 2006). Sharpening is conveniently performed using *PHENIX* (Adams *et al.*, 2002), *CNS* (Brünger *et al.*, 1998) or the abovementioned anisotropy server.

4. Conclusion

In summary, we point to low-resolution molecular replacement as a way of initiating heavy-atom-derived phasing of poorly diffracting membrane-protein crystals. Surprisingly, a number of unrelated sequences of membrane proteins display similar tertiary structures (Abramson & Wright, 2009; Wang *et al.*, 2009), suggesting that the number of unique folds of membrane proteins may in fact be lower than expected and that the Protein Data Bank (Berman *et al.*, 2000) will be an increasingly powerful resource for low-resolution MR phasing of new membrane-protein crystals. All-atom refinement of the search models by molecular dynamics (Qian *et al.*, 2007) or the use of generic model representations (Strop *et al.*, 2007) and automated procedures (Keegan & Winn, 2007) will be useful features. Non-isomorphous data sets can be obtained by variations of crystal conditions and dehydration, enabling multi-crystal averaging to be used as a powerful and generally applicable tool for low-resolution density modification and phase extension, complementing NCS averaging. The MR procedure yields transformation matrices and mask definitions and provides a starting model for map validation and model building. Although low-resolution membrane-protein crystallography may not necessarily yield fully refined crystal structures, a high-quality and unbiased electron-density map derived from crystal diffraction at 3.5–7 Å maximum resolution still allows imaging and molecular analysis at a favourable resolution compared with other techniques such as small-angle scattering, atomic force microscopy and electron microscopy as shown here using well established procedures and resources for X-ray crystallography.

We thank the SLS staff at X06SA for maintaining an excellent beamline and Drs Clemens Schulze-Briese, Jimin Wang and Thomas Schneider for valuable discussions on low-resolution imaging. We are grateful to Peter Henderson, So Iwata and Alexander Cameron for access to derivative data from the Mhp1 crystals. BPP is supported by a postdoctoral fellowship from the Carlsberg Foundation, PN is supported by a Hallas–Møller stipend from the Novo Nordisk Foundation and JPM was funded by a postdoctoral fellowship from the DANSYNC programme of the Danish Research Council during part of this study. All authors contributed equally.

References

Abramson, J. & Wright, E. M. (2009). *Curr. Opin. Struct. Biol.* **19**, 425–432.
Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.
Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P. B. & Steitz, T. A. (1999). *Nature (London)*, **400**, 841–847.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Bricogne, G. (1992). *Proceedings of the CCP4 Study Weekend. Molecular Replacement*, edited by E. J. Dodson, S. Gover & W. Wolf, pp. 62–75. Warrington: Daresbury Laboratory.
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.
Brunger, A. T., DeLaBarre, B., Davies, J. M. & Weis, W. I. (2009). *Acta Cryst.* D**65**, 128–133.
Chen, B., Vogan, E. M., Gong, H., Skehel, J. J., Wiley, D. C. & Harrison, S. C. (2005). *Structure*, **13**, 197–211.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
DeLaBarre, B. & Brunger, A. T. (2006). *Acta Cryst.* D**62**, 923–932.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* http://www.pymol.org.
Faham, S., Watanabe, A., Besserer, G. M., Cascio, D., Specht, A., Hirayama, B. A., Wright, E. M. & Abramson, J. (2008). *Science*, **321**, 810–814.
Jenni, S., Leibundgut, M., Maier, T. & Ban, N. (2006). *Science*, **311**, 1263–1267.
Jeruzalmi, D., O'Donnell, M. & Kuriyan, J. (2001). *Cell*, **106**, 429–441.
Keegan, R. M. & Winn, M. D. (2007). *Acta Cryst.* D**63**, 447–457.
Keller, S., Pojer, F., Heide, L. & Lawson, D. M. (2006). *Acta Cryst.* D**62**, 1564–1570.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Lescar, J., Roussel, A., Wien, M. W., Navaza, J., Fuller, S. D., Wengler, G., Wengler, G. & Rey, F. A. (2001). *Cell*, **105**, 137–148.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Maeda, S., Nakagawa, S., Suga, M., Yamashita, E., Oshima, A., Fujiyoshi, Y. & Tsukihara, T. (2009). *Nature (London)*, **458**, 597–602.
Maier, T., Jenni, S. & Ban, N. (2006). *Science*, **311**, 1258–1262.
McCoy, A. J. (2004). *Acta Cryst.* D**60**, 2169–2183.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Meyer, P. A., Ye, P., Zhang, M., Suh, M. & Fu, J. (2006). *Structure*, **14**, 973–982.
Morth, J. P., Pedersen, B. P., Toustrup-Jensen, M. S., Sørensen, T. L., Petersen, J., Andersen, J. P., Vilsen, B. & Nissen, P. (2007). *Nature (London)*, **450**, 1043–1049.
Morth, J. P., Sørensen, T. L. & Nissen, P. (2006). *Acta Cryst.* D**62**, 877–882.
Murshudov, G. N., Davies, G. J., Isupov, M., Krzywda, S. & Dodson, E. J. (1998). *CCP4 Newsl. Protein Crystallogr.* **35**, 37–43.
Pedersen, B. P., Buch-Pedersen, M. J., Morth, J. P., Palmgren, M. G. & Nissen, P. (2007). *Nature (London)*, **450**, 1111–1114.
Popov, A. N. & Bourenkov, G. P. (2003). *Acta Cryst.* D**59**, 1145–1153.
Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J. & Baker, D. (2007). *Nature (London)*, **450**, 259–264.
Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D. & Eisenberg, D. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 8060–8065.
Strop, P., Brzustowicz, M. R. & Brunger, A. T. (2007). *Acta Cryst.* D**63**, 188–196.
Suga, M., Maeda, S., Nakagawa, S., Yamashita, E. & Tsukihara, T. (2009). *Acta Cryst.* D**65**, 758–766.
Svergun, D., Petoukhov, M. V. & Koch, M. H. J. (2001). *Biophys. J.* **80**, 2946–2953.
Taylor, G. (2003). *Acta Cryst.* D**59**, 1881–1890.
Wang, Y., Huang, Y., Wang, J., Cheng, C., Huang, W., Lu, P., Xu, Y. N., Wang, P., Yan, N. & Shi, Y. (2009). *Nature (London)*, **462**, 467–472.
Weyand, S. *et al.* (2008). *Science*, **322**, 709–713.
Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. (2005). *Nature (London)*, **437**, 215–223.