

From electron microscopy to X-ray crystallography: molecular-replacement case studies

Yong XiongDepartment of Molecular Biophysics and
Biochemistry, Yale University, New Haven,
CT 06511, USA

Correspondence e-mail: yong.xiong@yale.edu

Received 21 February 2007

Accepted 29 October 2007

Multi-component molecular complexes are increasingly being tackled by structural biology, bringing X-ray crystallography into the purview of electron-microscopy (EM) studies. X-ray crystallography can utilize a low-resolution EM map for structure determination followed by phase extension to high resolution. Test studies have been conducted on five crystal structures of large molecular assemblies, in which EM maps are used as models for structure solution by molecular replacement (MR) using various standard MR packages such as *AMoRe*, *MOLREP* and *Phaser*. The results demonstrate that EM maps are viable models for molecular replacement. Possible difficulties in data analysis, such as the effects of the EM magnification error, and the effect of MR positional/rotational errors on phase extension are discussed.

1. Introduction

Developments in technology and methodology have increasingly narrowed the gap between the two major structural biology techniques: X-ray crystallography and single-particle cryo-electron microscopy (cryo-EM). X-ray crystallography, the most effective method for obtaining the atomic model of a molecule, is being used more and more often in tackling the structures of large molecular complexes. Structures of macromolecules that are megadaltons in molecular weight, such as the ribosome and fatty-acid synthase, have been solved by X-ray crystallography (Jenni *et al.*, 2006; Korostelev *et al.*, 2006; Nissen *et al.*, 2000; Selmer *et al.*, 2006; Lomakin *et al.*, 2007). These large molecular complexes are traditionally studied by cryo-EM, in which two-dimensional projection images of the molecule are used to reconstruct a three-dimensional map at low resolution (Baker *et al.*, 1999). Because of recent improvements in EM instruments and reconstruction techniques, it is now common to achieve a cryo-EM reconstruction at a resolution better than 10 Å (Rossmann *et al.*, 2005). Structures have even reached 4 Å resolution (van Heel *et al.*, 2000), which is approaching the atomic resolution obtained by X-ray crystallography.

These advances facilitate the interplay of the two methods, either by providing an EM image for X-ray studies or by providing an X-ray structure for EM studies. In the former case, an EM image is used as a search model for X-ray crystal structure determination by molecular replacement (MR). Initial MR phases at low resolution can then be extended to high resolution by density-modification techniques such as noncrystallographic symmetry (NCS) averaging. Initial MR phases can also be used to locate heavy-atom positions using difference Fourier calculations, followed by subsequent heavy-

Table 1

X-ray and EM data used in the MR case studies.

Molecule	X-ray data			Reference	EM data			Data for MR
	Space group	Molecules per AU	Resolution (Å)		Resolution (Å)	EMDB code	Reference	Resolution range (Å)/completeness (%)
50S	<i>C222₁</i>	1	10	Nissen <i>et al.</i> (2000)	7.5	1019	Matadeen <i>et al.</i> (1999)	86–10/99
70S	<i>P2₁2₁2₁</i>	2	2.8	Selmer <i>et al.</i> (2006)	11	1008	Rawat <i>et al.</i> (2003)	50–12/96
70S	<i>I422</i>	1	3.7	Korostelev <i>et al.</i> (2006)				73–12/99
FAS	<i>P21</i>	1	5	Lomakin <i>et al.</i> (2007)	24		Kolodziej <i>et al.</i> (1997)	82–24/90
FAS	<i>P4₃2₁2</i>	1/2	4					60–24/83

atom phasing in isomorphous replacement or anomalous dispersion methods. In the latter case of using X-ray crystal structures to interpret EM data, parts of a large molecular complex whose structures have been solved by X-ray crystallography can be docked into the EM reconstruction map. This frequently results in a more detailed structural description of a complex than revealed by the EM map or the individual X-ray structures.

The exchange of information between the two techniques has been facilitated by a concerted effort to make EM images available with a standard convention (Heymann *et al.*, 2005). A significant development is the establishment of a public database, the Electron Microscopy Data Bank (EMDB), for three-dimensional EM data deposition (<http://www.ebi.ac.uk/msd>). The EM images are deposited in the same CCP4 map format as used in X-ray crystallography. This standard format will significantly contribute to the increased use of EM data by the X-ray crystallographic community.

Here, I will discuss issues related to the use of EM images as MR models for X-ray crystal structure determination. This study complements an excellent review by Dodson (2001). I will present three case studies, in which EM maps of the 50S ribosomal subunit, 70S ribosome and fatty-acid synthase (FAS) were used to solve their X-ray crystal structures. Information about the EM images and the crystal data is summarized in Table 1. A brief procedure of EM map preparation for MR program packages such as *AMoRe* (Navaza, 2001), *MOLREP* (Vagin & Teplyakov, 1997) and *Phaser* (Read, 2001) will be described. The effect of errors, including those in EM data and MR solutions, on structure determination and phase extension will be discussed.

2. EM map preparation for molecular replacement

The use of EM maps for molecular replacement exactly parallels the use of atomic coordinates. Before an MR run, a model-preparation step is carried out, which involves placing the model in an artificial unit cell with dimensions three or four times as large as those of the model. This ensures fine sampling of reciprocal space when the model is Fourier transformed to structure factors. These finely sampled model structure factors can then be used to compute 'crystal' structure factors by interpolation whenever they are needed within the MR procedure, such as preparing inputs to the rotation and translation programs and computing structure factors for the moving models during rigid-body refinement. The proce-

dures of preparing the EM map are described below with the required programs mentioned, which are either from the *CCP4* suite (Collaborative Computational Project, Number 4, 1994) or from the *RAVE* package of the Uppsala Software Factory (Kleywegt & Jones, 1999). A more EM-friendly version of *AMoRe* distributed by the program's author does not require items (ii) and (iii) described below.

(i) Scale the EM map to the correct magnification. The EM map is frequently deposited in the database on a three-dimensional grid of arbitrary step/pixel size, resulting in an arbitrary size of the molecule. The true grid step size should be obtained from the EM experimentalists and then applied to the EM map to yield an image of the correct size. Alternatively, the size of the EM map can be adjusted by comparing it with a separate atomic model when it is available. Adjusting the size of the EM map can be conveniently performed by using *MAPMAN* to convert the *CCP4*-format map to an ASCII/text-format map such as the *X-PLOR/CNS* format and then editing the unit-cell parameters in the map header. *MAPMAN* also produces a *NEWZD*-type ASCII map whose unit-cell parameters may be changed as described above.

(ii) Place the EM map into a large *P1* cell. This can be achieved using the program *MAPROT* with desired unit-cell parameters and grid steps specified. The program requires a mask that covers the molecule. The mask can be converted from the EM map itself using programs such as *MAMA* or *MAPMASK*. The resulting EM map in the large unit cell can then be directly supplied to the MR program *MOLREP* as the search model.

(iii) Calculate structure factors from the placed model. When using the current *CCP4* distribution of the MR programs *AMoRe* or *Phaser*, the EM map needs to be Fourier transformed to structure factors using a program such as *SFALL*. The structure factors can then be input into either *AMoRe* to generate the model structure-factor table, using the sorting function (*SORTFUN*), which takes the place of *TABLING* in an ordinary *AMoRe* run or *Phaser* to define the search ensemble.

(iv) Compare the average intensity/amplitude distribution against resolution (power spectrum) between the data generated from the EM map and the X-ray experimental data. It is advantageous to have similar distributions in the range used in the molecular replacement so that the scaling of F_c to F_{obs} is more accurate. This can be used as a guide to decide what resolution range should be included in the MR search and the

proper *B* factor by which to either sharpen or blur the EM image.

3. Case study 1: the effects of EM magnification or scaling errors

Because electron microscopy is unable to directly determine the absolute size of the molecule and the magnification error can be as large as 5% (Dodson, 2001), it is important to investigate the effect of this error on structure determination by molecular replacement. Test studies were conducted using an EM map of the *Escherichia coli* 50S ribosomal subunit (Matadeen *et al.*, 1999; EMDB accession code 1019) as the search model for molecular replacement with the X-ray data from the *Haloarcula marismortui* 50S crystals (Nissen *et al.*, 2000). The initial EM map was adjusted to approximately the correct size by comparing it with the known atomic model of the 50S ribosomal subunit. Various magnification errors from -15% to 15% were then introduced by altering the map header as described above. MR tests using the altered EM maps as search models were carried out using *MOLREP* at resolutions of 10 and 20 Å. In a second series of tests, structure factors were calculated from these EM maps and served as test data for an MR search by *Phaser* using an atomic model of *H. marismortui* 50S.

The test results indicate that MR solutions are not very sensitive to EM magnification errors and are a function of the highest resolution used in the search. For example, up to $\pm 10\%$ error can be tolerated at 20 Å resolution and up to $\pm 5\%$ error can be tolerated at 10 Å resolution (Fig. 1*a*). Similarly, when using the atomic model to search against ‘crystal’ data converted from EM maps of altered sizes, a solution could be found with up to $\pm 10\%$ size errors (Fig. 1*b*).

The correct magnification is important for later steps such as phase extension because it provides better starting phases, even though it may not be crucial for molecular replacement.

Table 2

Phaser Z-score table for the MR test with the 70S ribosome.

	Z scores	
	Rotation function	Translation function
<i>I422</i> crystal	5.6	25.3
<i>P2₁2₁2₁</i> crystal		
Run 1		
Molecule 1	6.4	17.3
Molecule 2	5.2	12.3
Molecule 2	3.6	5.9
Run 2†		
Molecule 2		25.5

† The orientation and position of molecule 1 and the orientation of molecule 2 are fixed.

Therefore, it is advantageous to find out the correct magnification by testing the correlation of the X-ray data with data calculated from EM maps at different sizes. Alternatively, this can also be achieved by choosing the optimal size that gives the best MR solution. In this test case the correct EM map size is $\sim 2\%$ smaller than the starting map (Fig. 1).

4. Case study 2: *E. coli* 70S ribosome in two crystal forms

The two recently published crystal structures of the *E. coli* 70S ribosome (Korostelev *et al.*, 2006; Selmer *et al.*, 2006) provide excellent MR test cases to study the effects of different space groups and different numbers of independent molecules in the asymmetric unit (AU) of the crystal. There is one 70S molecule in the AU of the *I422* crystal and there are two molecules in the AU of the *P2₁2₁2₁* crystal. A cryo-EM map of the 70S particle at 11 Å resolution (Rawat *et al.*, 2003; EMDB accession code 1008) was used as the search model. The test runs were carried out at 12 Å resolution using the program *Phaser*.

The MR search in the *I422* crystal form was straightforward, with outstanding peaks for the correct solution in both the

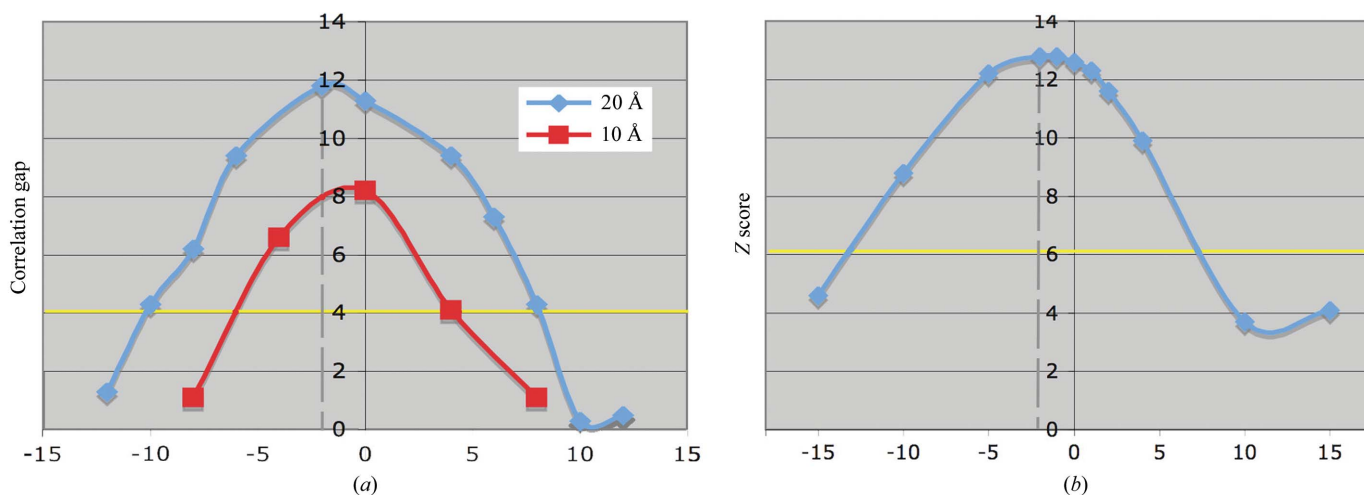


Figure 1

Effect of EM map magnification errors on the MR solution. (a) The vertical axis shows the gap in Patterson correlation separating the correct solution and the next highest noise peak using the program *MOLREP*. A minimum of 4 (yellow line) is the cutoff for a correct solution. The broken vertical line shows the optimal magnification of the EM map for the best MR solution. (b) Tests repeated using the program *Phaser* with the Z score shown on the vertical axis.

rotational and translational searches (Table 2). In the $P2_12_12_1$ crystal form two orientations are clearly found in the rotational search and two corresponding translational solutions were also found when looking for the first molecule alone (Table 2). The two translational solutions obtained from the automated search routine represent correct solutions individually but with arbitrary choices of origins that may be different. However, a solution for the second molecule was not found using the automated routine. The structure was solved by running a second translational search in *Phaser* with molecule 1 fixed manually and the rotational angles for

molecule 2 also fixed. The angles for molecule 2 were taken from the first *Phaser* run searching for molecule 1.

5. Test case 3: an unknown structure of yeast fatty-acid synthase (FAS)

Yeast FAS, a 2.6 MDa $\alpha_6\beta_6$ protein complex with point-group symmetry 32, was crystallized in two space groups, with one FAS particle in the AU of a $P2_1$ crystal and half an FAS particle in the AU of a $P4_32_12$ crystal (Lomakin *et al.*, 2007). A 24 Å resolution cryo-EM map of yeast FAS (courtesy of James

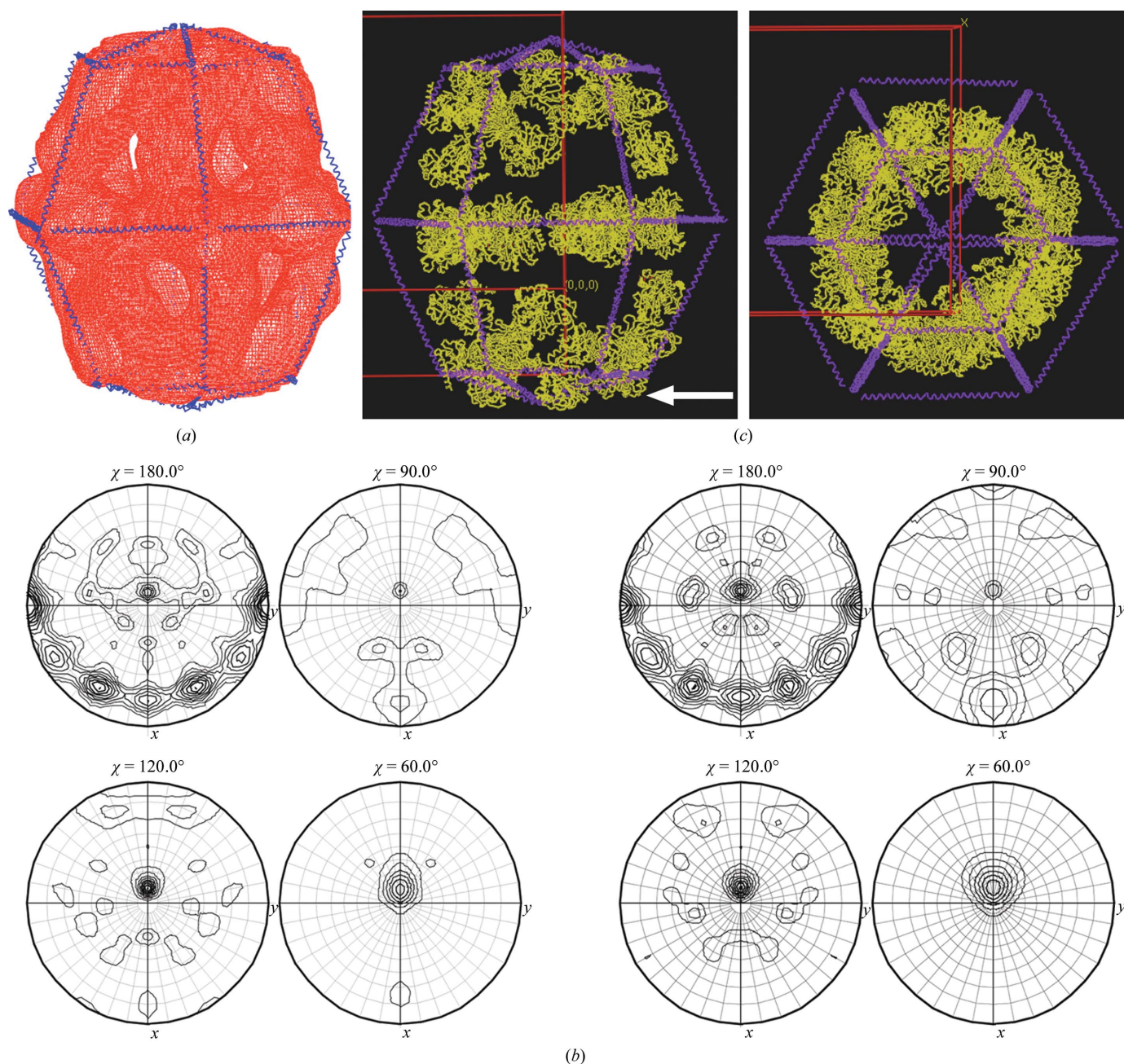


Figure 2

MR solution in the $P2_1$ crystal form. (a) The EM map model. The blue skeleton outlines the EM image. (b) The self-rotation plots from the X-ray data (left) and from the MR solution (right). (c) Comparison of the MR solutions from the EM map model (blue skeleton) and from the atomic model (yellow). The arrow points to the positional errors in the solution with the EM model.

Stoop, University of Texas; Kolodziej *et al.*, 1997) was used as the search model (Fig. 2a).

5.1. MR solution in the $P2_1$ crystal

An outstanding solution was found at 24 Å resolution using either *AMoRe*, *MOLREP* or *Phaser*. This solution was corroborated by packing analysis and by comparing the calculated self-rotation peaks from the positioned model with those of the observed X-ray data (Fig. 2b). However, phase extension to high resolution by sixfold NCS-averaging was unsuccessful, presumably owing to inaccuracy in the MR solution at low resolution, which results in errors in the NCS operators. When compared with the final solution from an atomic model that later became available, the solution obtained from the EM map model had a positional error of a few angstroms in the xz (ac) plane (Fig. 2c). This error is presumably too large to be corrected by NCS-operator refinement during the averaging process.

5.2. MR solution in the $P4_32_12$ crystal

An MR solution could not be found in the $P4_32_12$ crystal form with standard searches using the full or half FAS particle as the model after extensive trials with various resolution cutoffs, magnification and intensity corrections. However, analysis of crystal content and the X-ray data greatly simplifies molecular replacement. Because there is only half an FAS particle in the AU, the other half of the molecule must be generated by the crystallographic twofold symmetry. Therefore, the particle is located on the twofold crystallographic axis. The orientation of the FAS around the twofold axis can be defined by the threefold axis of the particle, which can be found from the peaks in the threefold section ($\chi = 120^\circ$) of the self-rotation plot (Fig. 3a). This analysis reduced the search

from six-dimensional to just one-dimensional along the xy diagonal where the twofold crystallographic axis is located (Fig. 3b).

A one-dimensional search was carried out using the line search option in *Phaser*. The EM map of half an FAS particle was placed on the xy diagonal with its threefold axis oriented according to the self-rotation peak. A solution was found but was initially unrecognized for two reasons. Firstly, the Z score of the correct solution was only 3.7, well below the value (Z score 5) above which a solution is considered to be correct. Secondly, this solution scores below a false solution of Z score 4.0. The solution was identified *a posteriori* when an MR solution was determined from an atomic model consisting of C^α atoms of homologous proteins of several domains of the molecule (Jenni *et al.*, 2006). Although the atomic model consists of less than 5% of the diffracting mass of the FAS particle, the MR solution in *Phaser* was readily obtained using data in the 60–10 Å resolution range (rotational Z score of 8.7 and translational Z score of 17.3). This is because these C^α atoms are from domains that are relatively evenly distributed within the FAS particle; they define the shape of the molecule sufficiently well and correlate well with the FAS electron density at low resolution. However, the C^α model worked only at resolutions higher than 12 Å in both crystal forms. Neither the cryo-EM model nor the C^α model yielded a solution for the tetragonal crystal at a resolution lower than 12 Å, presumably owing to the strong effect of the solvent contribution to the diffraction at low resolution.

5.3. EM model phase extension

Phase extension by cross-crystal averaging between the two crystal forms was carried out using the program *DMMULTI* starting from phases calculated from an EM map correctly

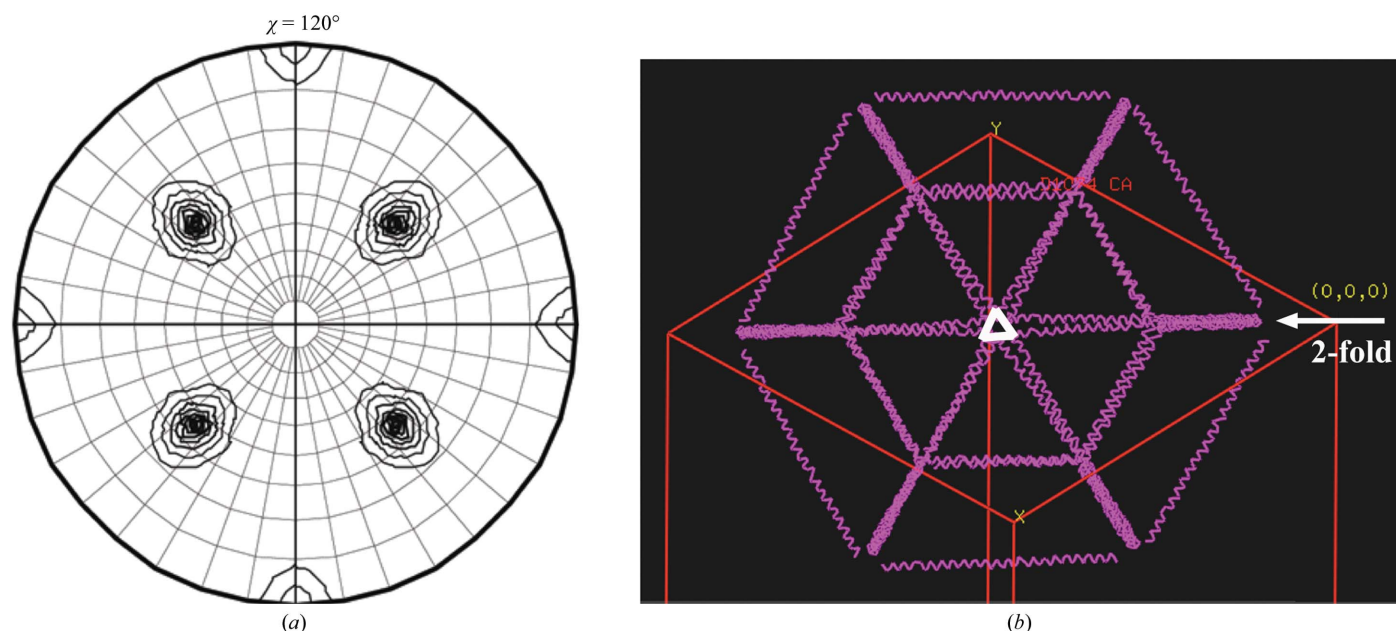


Figure 3 Self-rotation (a) and packing analysis (b) of the $P4_32_12$ FAS crystal. The threefold axis of the FAS particle is indicated by the white triangle in (b).

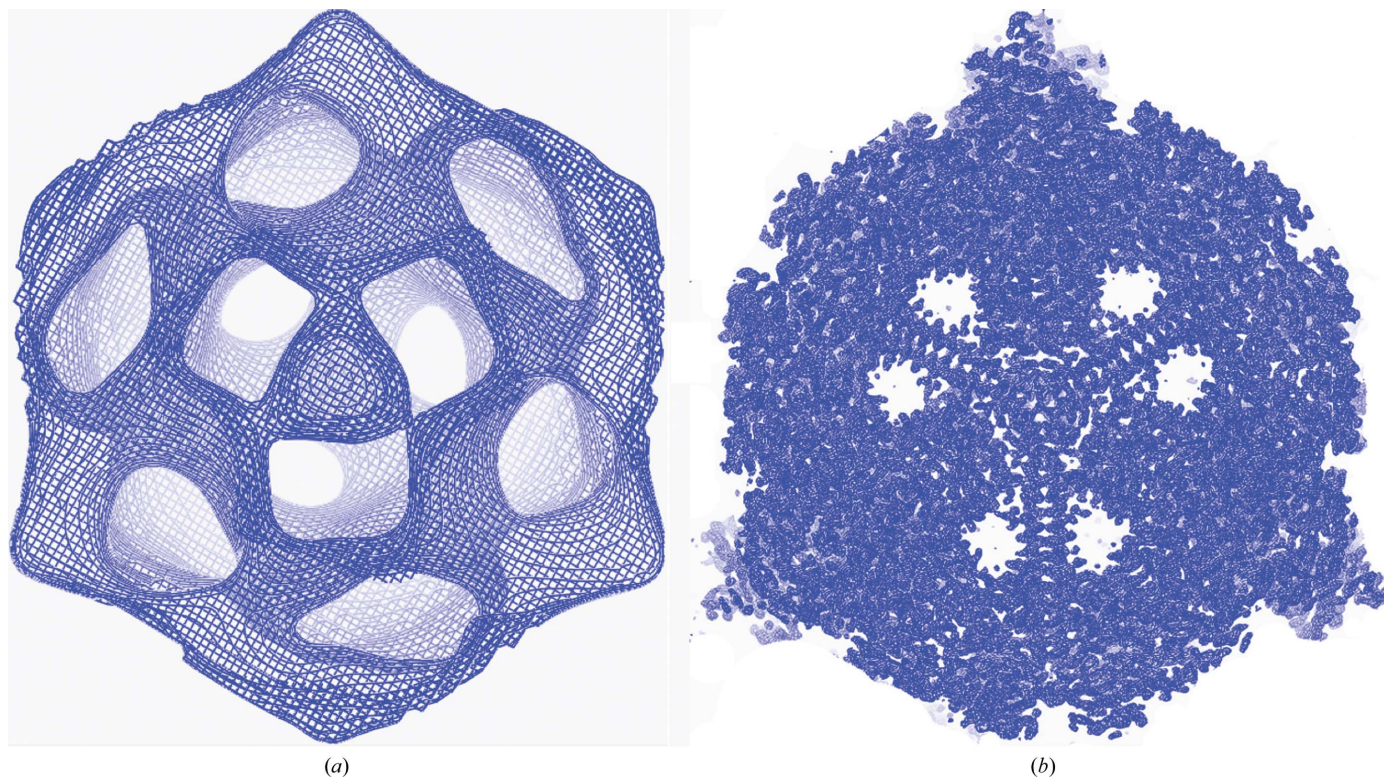


Figure 4
Phase extension from the starting FAS EM map (a) to the final map at 4 Å (b).

placed by *Phaser*. An initial averaging mask was defined by the EM map. The NCS matrices within each crystal and those relating the two crystals were defined by the MR solutions from the atomic model. Because of the ninefold NCS relationship (sixfold in the $P2_1$ crystal and threefold in the $P4_32_12$ crystal) and the accuracy of the NCS axes, the phase extension was successful and produced an electron-density map of extraordinary quality at 4 Å resolution (Fig. 4). The X-ray structure factors were sharpened with a B factor of -120 \AA^2 to calculate the map. The B -factor sharpening dramatically enhanced the higher resolution details of the electron-density map. The sharpening magnitude was chosen to maximize the enhancement while not overinflating the noise at higher resolution. It is possible that the NCS matrices defined by the MR solution from the EM map model would also work in the case of cross-crystal averaging because the high degree of NCS averaging might produce a radius of convergence large enough to refine the NCS matrices and phases to the correct values.

6. Conclusions

Test cases demonstrate that cryo-EM maps serve as viable models for MR solution of X-ray crystal structures. Of the five X-ray diffraction data sets tested, only one could not be solved by molecular replacement with an EM model using standard procedures. A cryo-EM model worked at very low resolution (82–24 Å) in a case when an incomplete atomic model could not produce a solution at the same resolution range. When a

high degree of NCS is present, phases generated from the low-resolution EM map can be extended to high resolution by averaging. NCS occurs frequently as a large fraction of macromolecules fortuitously crystallize with more than one molecule in the asymmetric unit of the crystal. Furthermore, the use of an EM map as a search model is likely to be applied to large molecular complexes, which are frequently highly symmetrical and therefore may crystallize with a high degree of NCS.

I thank Ivan Lomakin for assistance, Jimin Wang, Joao Cabral and Jesse Cochrane for comments on the manuscript, and James Stoops for providing EM maps of yeast FAS.

References

- Baker, T. S., Olson, N. H. & Fuller, S. D. (1999). *Microbiol. Mol. Biol. Rev.* **63**, 862–922.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Dodson, E. J. (2001). *Acta Cryst.* **D57**, 1405–1409.
- Heel, M. van, Gowen, B., Matadeen, R., Orlova, E. V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M. & Patwardhan, A. (2000). *Q. Rev. Biophys.* **33**, 307–369.
- Heymann, J. B., Chagoyen, M. & Belnap, D. M. (2005). *J. Struct. Biol.* **151**, 196–207.
- Jenni, S., Leibundgut, M., Maier, T. & Ban, N. (2006). *Science*, **311**, 1263–1267.
- Kleywegt, G. J. & Jones, T. A. (1999). *Acta Cryst.* **D55**, 941–944.
- Kolodziej, S. J., Penczek, P. A. & Stoops, J. K. (1997). *J. Struct. Biol.* **120**, 158–167.

- Korostelev, A., Trakhanov, S., Laurberg, M. & Noller, H. F. (2006). *Cell*, **126**, 1065–1077.
- Lomakin, I. B., Xiong, Y. & Steitz, T. A. (2007). *Cell*, **129**, 319–332.
- Matadeen, R., Patwardhan, A., Gowen, B., Orlova, E. V., Pape, T., Cuff, M., Mueller, F., Brimacombe, R. & van Heel, M. (1999). *Structure*, **7**, 1575–1583.
- Navaza, J. (2001). *Acta Cryst. D* **57**, 1367–1372.
- Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. (2000). *Science*, **289**, 920–930.
- Rawat, U. B., Zavialov, A. V., Sengupta, J., Valle, M., Grassucci, R. A., Linde, J., Vestergaard, B., Ehrenberg, M. & Frank, J. (2003). *Nature (London)*, **421**, 87–90.
- Read, R. J. (2001). *Acta Cryst. D* **57**, 1373–1382.
- Rossmann, M. G., Morais, M. C., Leiman, P. G. & Zhang, W. (2005). *Structure*, **13**, 355–362.
- Selmer, M., Dunham, C. M., Murphy, F. V. IV, Weixlbaumer, A., Petry, S., Kelley, A. C., Weir, J. R. & Ramakrishnan, V. (2006). *Science*, **313**, 1935–1942.
- Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.