

Using electron-microscopy images as a model for molecular replacement

Eleanor J. DodsonYork Structural Biology Laboratory, Chemistry
Department, University of York,
York YO10 5DD, EnglandCorrespondence e-mail:
e.dodson@ysbl.york.ac.uk

This review addresses the technical problems encountered while using models based on electron microscopy to generate initial phases for crystallographic studies. The test cases used were the gp6 portal protein with 13-fold rotational symmetry and the human hepatitis virus HepB, a viral assembly with $T = 4$ icosahedral symmetry.

Received 16 May 2001
Accepted 8 August 2001

1. Introduction

The electron microscope provides a way of visualizing the structures of large macromolecular assemblies. To complete a three-dimensional reconstruction of a single particle, images of the specimen viewed from different angles are aligned and the model built up. If the same assembly can be crystallized, the electron-microscopy (EM) model and phases can be used as a starting point for the X-ray structure determination. Before this can be performed, the EM model must be correctly positioned in the unit cell; in other words, the molecular-replacement solution must be obtained.

Molecular replacement is a technique for matching observed crystal diffraction intensities to those predicted by a suitable model correctly orientated and positioned within the unit cell. Usually, the model is described using coordinates and the predicted intensity pattern can be generated either within the molecular-replacement program (*AMoRe*, *MOLREP*) or by using standard software for calculating structure factors, which then provide the input to an MR package such as *ALMN*. It is convenient to generate these assuming that the model lies at the origin of a *P1* cell with 'crystal' axes at least twice the model diameter. Many packages such as *AMoRe* use model structure factors which are more finely sampled in reciprocal space; *i.e.* the 'crystal' into which the model is placed has axes three or four times the model diameter (Lattman, 1985; Navaza, 2001). This means that the structure factors corresponding to different orientations of the model can be obtained by interpolation from the initial set. This saves a great deal of computer time without a serious loss of accuracy and allows many orientations to be explored rapidly. The structure factors generated by the model at a given position in the unit cell are the sum of symmetry-equivalent reflections with phase corrections appropriate for the space group.

The usual procedure for generating structure factors for large molecules first generates an 'atom map' where the contribution from all atoms is summed onto a regular grid.

Once such a map has been constructed, the structure factors can be generated very quickly using the inverse fast Fourier transform. Since electron microscopists conventionally construct 'maps' from their phase estimations, these too can generate a structure-factor list in a suitable 'cell' for input to the existing MR packages.

In addition to the technical issues to be resolved, in any interdisciplinary research we face the serious problem of properly understanding the difficulties and the conventions of our partners. The CCP4 Study Weekend on Low Resolution Phasing in 2000 had brought together electron microscopists and crystallographers and the Proceedings [*Acta Cryst.* (2000), D56, 1205–1357] give an excellent starting point with good descriptions of the techniques. The comprehensive review by Baker *et al.* (1999) is also very helpful.

Another difficulty is the paucity of test data sets. There is not yet a mechanism for the deposition of the experimental results of electron microscopy. I am very indebted to E. Orlova (Orlova *et al.*, 1999), R. A. Crowther (Bottcher *et al.*, 1997) and A. Leslie (Wynne *et al.*, 1999) for providing images and X-ray data for these studies.

2. Problems

Technical problems to consider are as follows.

(i) Electron microscopy is not able to determine the precise scale of the particle. Dr Orlova warned us that her images may have an error of up to 5%. Fortunately, the correlation coefficient used by the *AMoRe* rotation function is sensitive to any change of scale of the image and clearly indicated the best scale (1.01 in this case). The image scale factor can be easily modified by changing the 'cell' dimensions in the header records.

(ii) Although the image is stored as a three-dimensional array, the particle representation is only valid within a masked volume. Within this volume, the base level for the density is somewhat arbitrary and it may be necessary to add some constant value to all these grid points to bring the density to the true base level. Unfortunately, the mask limits are not recorded and the crystallographer must deduce them from the image. Standard density-modification techniques can be used and Dr Leslie (Wynne *et al.*, 1999) reports obtaining better results for the phase-extension step after experimenting with a variety of masks and base levels. Errors in the base level have little effect on the MR results; the model is already truncated within a spherical volume and the correlation coefficients are fairly insensitive to this correction.

(iii) Mancina & Fuller (2000) described the difficulties electron microscopists sometimes have in determining a reliable phase-contrast function. Errors in this can result in some images being inverted and thus generating phases in reciprocal space which are 180° in error. As EM techniques improve this is less likely to occur, but crystallographers need to be aware of the problem.

(iv) It is difficult to decide what resolution range of X-ray and EM data should be used for the MR search and how best to scale them together. An EM image shows the shape and

three-dimensional boundary of the particle clearly (Fig. 1), but has little atomic detail. It generates diffraction data which is complete and good at very low resolutions (50 Å say) but falls off rapidly with resolution and rarely extends beyond 10 Å. On the other hand, there are technical difficulties in collecting X-ray data at these very low resolutions; the structure factors will contain a significant scattering component from the solvent volume in the unit cell. Translation searches raise slightly different problems to rotation ones. Rotation functions are essentially Patterson search functions which are independent of phase, whereas the translation function matches the observed amplitude with that of a vector sum of the calculated structure factors. The best resolution range may well be different for the two functions.

(v) Many structures solved by EM have a high degree of internal symmetry. This can be helpful for the MR search within the crystal, although it leads to many equivalent solutions, which can be confusing. The high symmetry can also disguise the hand of the EM model; at very low resolution it can appear to be a centrosymmetric arrangement of spherical blobs.

3. Test cases

3.1. Bacteriophage SPP1 portal protein (gp6)

This portal protein, located at one single vertex of the icosahedral viral capsid, is a central component acting at different steps of tailed bacteriophage's assembly. It is a circular multi-subunit assembly composed of 13 identical subunits through which DNA movements occur and where the

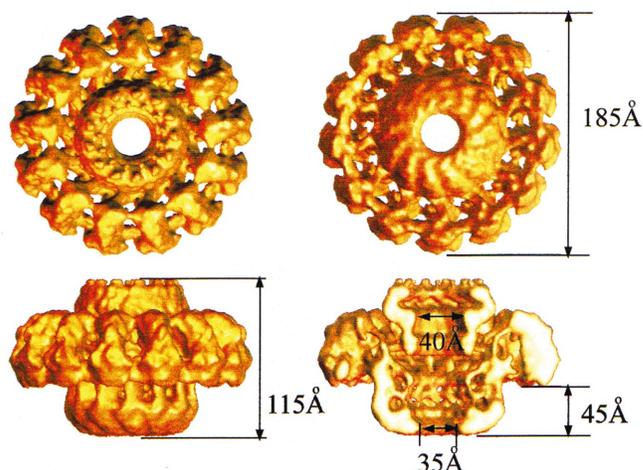


Figure 1

Bacteriophage SPP1 assembly pathway. A spherical procapsid is assembled by copolymerization of gp11 (scaffolding protein) and gp13 (major head protein) initiated at a specialized vertex characterized by the presence of gp6 (portal protein) and gp7 (accessory protein) (Lurz *et al.*, 1997; Dröge *et al.*, 2000). The symmetry and EM structures of the gp6 cyclical oligomer are shown for two morphogenetic steps of viral assembly (Orlova *et al.*, 1999; Orlova *et al.*, unpublished results).

phage tail is attached. This protein has been a subject of electron-microscopy single-particle reconstruction studies which have led to the determination of the structure of native gp6 and of the mutant gp6 SizA (Fig. 1) (Tavares *et al.*, 1992; Orlova *et al.*, 1999 and references therein).

gp6 has also been crystallized in space group $C222_1$, with unit-cell parameters 173.42, 222.61, 420.12 Å (Jekow *et al.*, 1998), and X-ray data has now been collected to 3.4 Å. With the current technology at the ESRF it was possible to record all the low-resolution data to 80 Å. The unit-cell volume

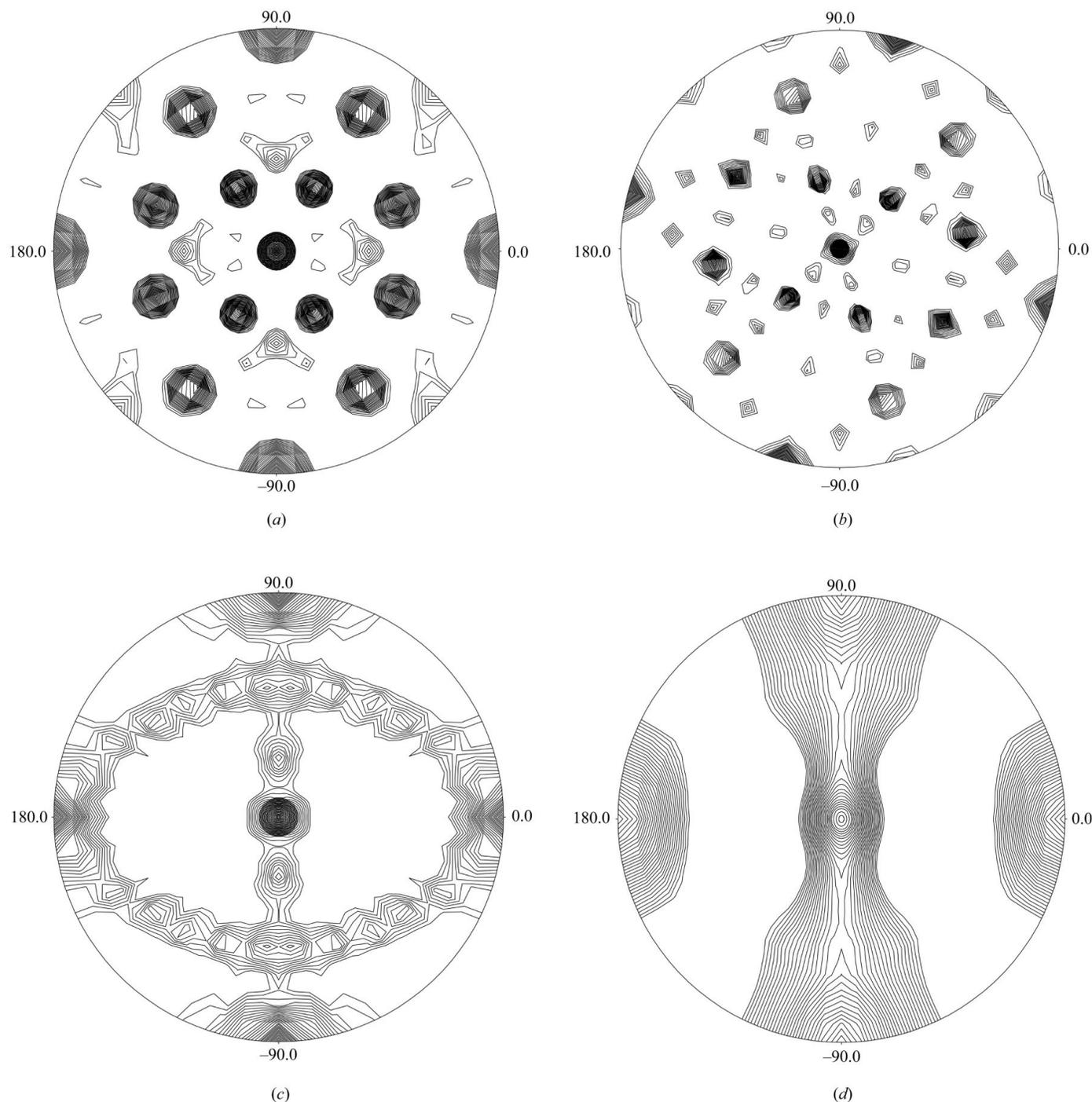


Figure 2

$\kappa = 180^\circ$ sections of self-rotation functions showing the directions of the twofold axes of symmetry. All were calculated using *POLARRFN* (Collaborative Computational Project, Number 4, 1994). (a) Using data generated from the EM image of hepatitis B showing the viral $T = 4$ symmetry. Resolution range 50–10 Å and search sphere radius 60 Å. (b) Using X-ray data for HepB. The map is viewed down the crystallographic twofold axis. Resolution range 20–8 Å and search sphere radius 60 Å. (c) Using X-ray data for gp6. Crystal symmetry requires there be 13 twofold axes. Resolution range 30–12 Å and search sphere radius 50 Å. (d) Using X-ray data for gp6. Resolution range 80–12 Å. The 13 twofold axes have been obscured by the pattern generated from the solvent content of the crystal.

requires that there be only one assembly of gp6 in the asymmetric unit, sited at a general position in the unit cell.

3.2. Human hepatitis B virus capsid (HepB)

This viral assembly consists of 240 subunits arranged with $T = 4$ icosahedral symmetry. Its structure was determined first by single-particle electron cryomicroscopy at 7.4 Å resolution (Bottcher *et al.*, 1997) and then crystallized and solved using X-ray techniques (Wynne *et al.*, 1999). The space group is $C2$, with unit-cell parameters 538.4, 354.8, 370.1 Å, $\beta = 132.3^\circ$. The unit-cell volume required that the crystal asymmetric unit contained only one half of the icosahedra and that the complete unit was generated by the crystallographic twofold axis. This meant that the particle be positioned on the twofold

rotation axis and limited the MR problem to orientating the particle relative to the crystal axes.

4. Methodology

4.1. Self-rotation searches

Since the expected results for the self rotations were known (a 13-fold rotation axis for gp6 and $T = 4$ viral symmetry for HepB), we were able to use self-rotation functions to test various parameters, such as the sphere radii, and the best resolution limits. The outer range was restricted to the limit available from the EM model (about 12 Å for gp6 and 7.4 Å for HepB). The inner resolution limit was varied to give the strongest signal from the protein symmetry, *i.e.* where the

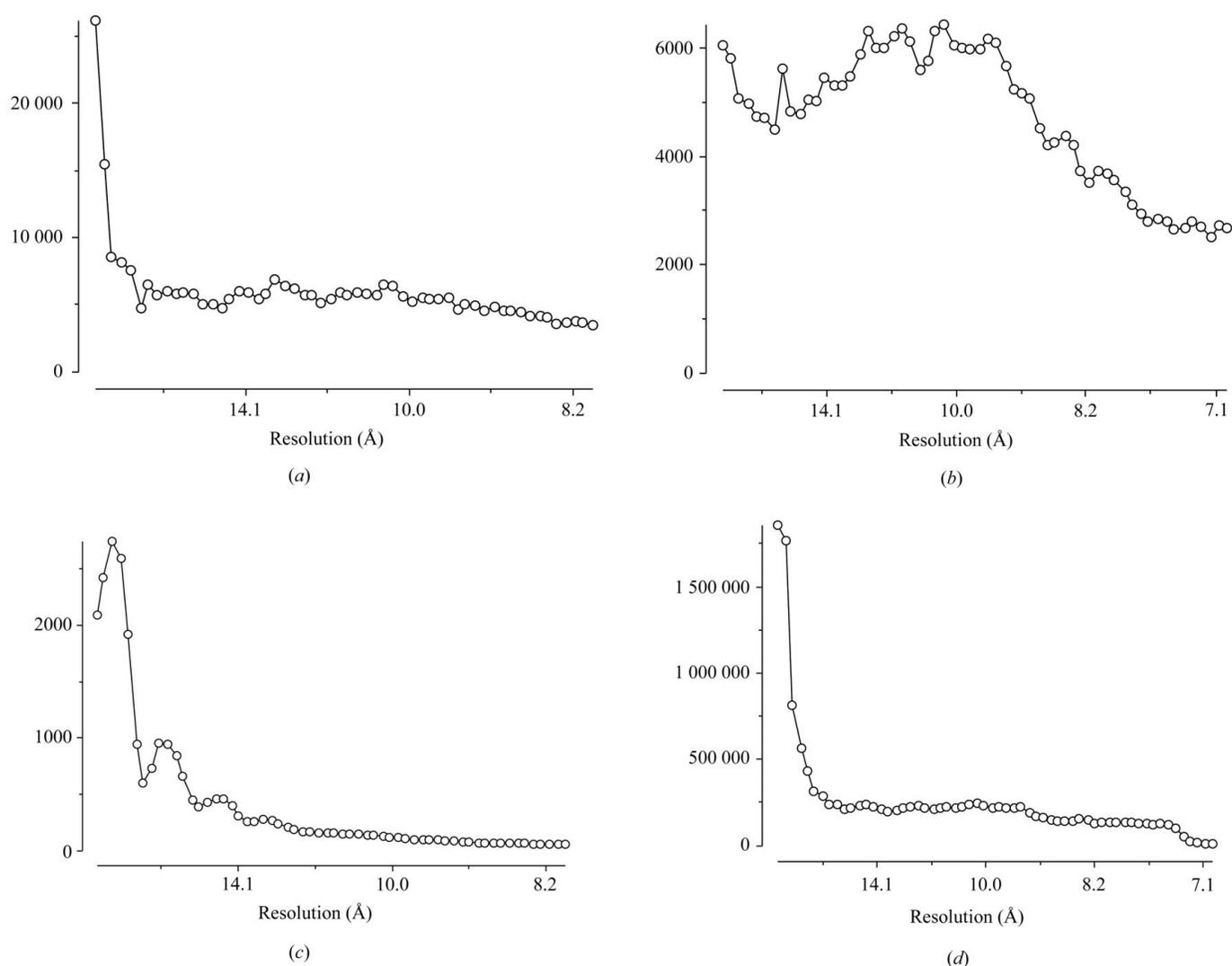


Figure 3

The $\langle F \rangle$ distributions plotted against resolution for the observed X-ray data and the calculated structure factors generated from the EM images. (a) The distribution for gp6 X-ray data in the range 80–8.2 Å. The distribution is quite smooth, but increases rapidly at very low resolution. (b) The distribution for X-ray data from HepB in the range 30–7 Å. (c) The distribution of $\langle F \rangle$ obtained from the gp6 EM images. This was normalized then ‘blunted’ by the application of an artificial temperature factor of 400 Å² before being used as model data for molecular-replacement searches. The information from 12–8.2 Å did not improve the molecular-replacement signal. (d) The distribution of $\langle F \rangle$ obtained from the HepB EM images. These images showed more detail than those for gp6 and the distribution from 20 to 7.4 Å fell off at much the same rate as that of the X-ray data. Including them in the MR calculations also improved the signal.

protein contribution to the X-ray amplitudes dominated that from the solvent (Fig. 2). For gp6, the maps generated using data from 80–50 Å showed no 13-fold symmetry at all, indicating that this shell of data is dominated by the solvent diffraction. This low-resolution information helped to determine the translation parameters and was valuable for phase-extension procedure, but was not useful for the rotation search.

4.2. Cross-rotation searches

As described above, structure factors were generated from both the EM maps. The $\langle F \rangle$ distribution for these are shown in Fig. 3 for comparison with the $\langle F \rangle$ distribution of the X-ray amplitudes. The cross rotation requires that these distributions are similar for the measured and calculated amplitudes; for more conventional studies, this is usually controlled by modifying the relative temperature-factor corrections. It is often advantageous to sharpen both data sets to increase the signal from the outer resolution shells, but matching the EM and X-ray distributions is tricky. In many cases, the strongest X-ray observations at low resolution have been lost; the Lorentz polarization correction increases sharply, increasing the risk that the reflections will overload the detector and be rejected during data processing. However, we found the best results were obtained using data in the range 30–15 Å, after applying a relative temperature factor of 400 Å to sharpen the calculated EM amplitudes. Although EM amplitudes had been generated to 10 Å, the 15–10 Å data did not improve the signal, presumably because the reliability of the EM results falls off. The best sphere radius was 50 Å, which covered a reasonable volume of the assembly without including too much of the central cavity. We knew the direction of the 13-fold axis in the crystal from the self-rotation results; it lay in the bc^* plane at 25° to c^* . This meant that the 13 cross-rotation peaks should satisfy $\omega = 25^\circ$, $\Phi = 90^\circ$ and occur at regular intervals of 27.7° round κ . *AMoRe*, *MOLREP* and *ALMN* all gave clear answers, with the best results obtained after scaling the EM images by a factor of 1.01.

The HepB case was simpler. The X-ray self rotation showed the $T = 4$ viral symmetry beautifully with its twofold, fivefold and sixfold rotation axes and after a rotation of 14.9° around the crystallographic twofold it aligned perfectly with the pattern from the EM images. As expected, *ALMN* gave a clear solution with a sphere radius of 60 Å using sharpened data in the resolution range 18–7.4 Å.

4.3. Translation searches

Crystal symmetry and asymmetric volume required that the HepB particle be positioned on the crystallographic twofold axis, so there was no need to perform a translation search for this case. For GP6, one copy of the assembly had to be positioned in the asymmetric unit. The best signal was obtained from *AMoRe* using rescaled data in the range 50–15 Å. The correlation coefficient for the centre-overlap function was 39% increasing to 61% after fitting. The next ranking corre-

lation coefficient was 59%, only 2% lower than the correct solution but not consistent with the self-rotation results. We repeated the search using a variety of resolution ranges and temperature-factor corrections and found that this result was robust.

It must be remembered that even when a particle is correctly positioned in the cell, the phase-extension step is challenging. This has been discussed in the review by Rossmann (1995) and by Mancini & Fuller (2000).

5. Conclusions

The joint exploitation of electron microscopy and crystallography will doubtless be used more extensively as X-ray data from larger and larger macromolecular assemblies becomes available. At least in the test cases reported here, which have a high degree of internal symmetry, optimal parameters for resolution limits and integration sphere radius could be found which led to successful molecular replacement.

The gp6 study depended on the beautiful EM images provide by Elena Orlova and her collaborators. The data collection by Fred Antson and Margaret Krause was an epic struggle, especially to obtain excellent very low resolution terms. The molecular-replacement analysis was performed in collaboration with Dr Julie Wilson. I am very grateful to Andrew Leslie and R. A. Crowther for allowing me the opportunity to use their superb HepB images and X-ray data. Further details of their analysis will be published elsewhere. Funding is provided by Wellcome Trust (grant 062788).

References

- Baker, T. S., Olson, N. H. & Fuller, S. D. (1999). *Microbiol. Mol. Biol. Rev.* **63**, 862–922.
- Bottcher, B., Wynne, S. A. & Crowther, R. A. (1997). *Nature (London)*, **386**, 88–91.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Dröge, A., Santos, M. A., Stiege, A., Alonso, J. C., Lurz, R., Trautner, T. A. & Tavares, P. (2000). *J. Mol. Biol.* **296**, 117–132.
- Dube, P., Tavares, P., Lurz, R. & van Heel, M. (1993). *EMBO J.* **12**, 1303–1309.
- Jekow, P., Schaper, S., Günther, D., Tavares, P. & Hinrichs, W. (1998). *Acta Cryst.* **D54**, 1008–1011.
- Lattman, E. (1985). *Methods Enzymol.* **115**, 55–77.
- Lurz, R., Trautner, T. A. & Alonso, J. (1997). *J. Mol. Biol.* **268**, 822–839.
- Mancini, E. J. & Fuller, S. D. (2000). *Acta Cryst.* **D56**, 1278–1287.
- Navaza, J. (2001). *Acta Cryst.* **D57**, 1367–1372.
- Orlova, E., Dube, P., Beckmann, E., Zemlin, F., Lurz, R., Trautner, T. A., Tavares, P. & van Heel, M. (1999). *Nature Struct. Biol.* **6**, 842–846.
- Rossmann, M. G. (1995). *Curr. Opin. Struct. Biol.* **5**, 650–655.
- Tavares, P., Santos, M. A., Lurz, R., Morelli, G., Lencastre, H. & Trautner, T. A. (1992). *J. Mol. Biol.* **225**, 81–92.
- Wynne, S. A., Crowther, R. A. & Leslie, A. G. W. (1999). *Mol. Cell*, **3**, 771–780.