Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The use of low-resolution phasing followed by phase extension from 7.6 to 2.5 Å resolution with noncrystallographic symmetry to solve the structure of a bacteriophage capsid protein

Nicola G. A. Abrescia,[a,b,c] Jonathan M. Grimes,[b,d] Hanna M. Oksanen,[e] Jaana K. H. Bamford,[f] Dennis H. Bamford[e] and David I. Stuart[b,d]*

[a]Structural Biology Unit, CIC bioGUNE, CIBERehd, Bizkaia Technology Park, 48160 Derio, Spain, [b]Division of Structural Biology, The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Headington, Oxford OX3 7BN, England, [c]IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain, [d]Diamond Light Source Ltd, Diamond House, Harwell Science and Innovation Campus, Didcot, England, [e]Institute of Biotechnology and Department of Biosciences, PO Box 56 (Viikinkaari 5), 00014 University of Helsinki, Finland, and [f]Department of Biological and Environmental Science and Nanoscience Centre, University of Jyväskylä, PO Box 35 (Survontie 9), 40014 Jyväskylä, Finland

Correspondence e-mail: dave@strubi.ox.ac.uk

Received 28 July 2010
Accepted 14 January 2011

P2, the major capsid protein of bacteriophage PM2, adopts the double β-barrel fold characteristic of the PRD1–adenoviral lineage. The 2.5 Å resolution X-ray data obtained by analysis of the two major lattices of a multiple crystal of P2 were phased by molecular replacement, using as a search model structure factors to 7.6 Å resolution obtained from electron density cut from the map of the entire PM2 virion. Phase extension to 2.5 Å resolution used solely sixfold cycling averaging and solvent flattening. This represents an atypical example of an oligomeric protein for which the structure has been determined at high resolution by bootstrapping from low-resolution initial phases.

1. Introduction

Structural information on capsid proteins and entire viruses can be used to group virus families (Bamford *et al.*, 2002, 2005; Abrescia *et al.*, 2004, 2008, 2010). Such analyses rely on the availability of atomic resolution three-dimensional structures (Abrescia *et al.*, 2008). Molecular-replacement (MR; Rossmann & Blow, 1962) and multi-wavelength anomalous dispersion (Hendrickson, 1991) methods are commonly used to solve the phase problem in X-ray crystallography (Villeret *et al.*, 1995) and advances in electron microscopy (EM) have allowed the use of low-resolution three-dimensional reconstructions as MR models to phase X-ray data (Rossmann, 1995; Lee & Johnson, 2003; Fry *et al.*, 2007; Qian *et al.*, 2007; Trapani *et al.*, 2006, 2010).

Recently, we elucidated the structure of PM2, an ∼45 MDa membrane-containing bacteriophage, by X-ray crystallography at 7 Å resolution using 60-fold averaging and solvent flattening (Abrescia *et al.*, 2008). The interpretation was aided by an SeMet difference Fourier map (Kivelä *et al.*, 2007), but the modest resolution prevented the unequivocal determination of the fold of the major capsid protein P2. Thus, we initiated structural analysis of isolated P2 (Abrescia *et al.*, 2005).

Here, we describe the structure determination of P2 using the low-resolution electron density derived from the averaged map of bacteriophage PM2 as the search model. This provides a successful phasing protocol that is applicable to other cases in which low-resolution information is available for the target molecule and noncrystallographic symmetry (NCS) is present within the protein crystal.

2. Crystal growth

P2 produced using recombinant technology aggregated, forcing us to purify protein from virus particles. P2 was dissociated from the virus and purified in Helsinki and shipped diluted (0.2 mg ml^{-1}) to Oxford (Abrescia *et al.*, 2005; Kivelä *et al.*, 2002). P2 purified from the selenomethionated virus tendened to aggregate and failed to produce diffraction-quality crystals (Kivelä *et al.*, 2007).

The temperature during the shipping of the protein from Helsinki to Oxford was monitored by an EL-USB-1 temperature-recording device (Lascar Electronics Ltd, UK) packed with the vial. Once in

Oxford the protein solution was concentrated to ∼3 mg ml^{-1} (in 150 m*M* NaCl and 20 m*M* Tris pH 7.5) and crystallization was carried out using in-house protocols (Abrescia *et al.*, 2005). Thin plate-like crystals grew in several conditions and considerable effort was made to increase their thickness. Eventually, a crystal suitable for high-resolution (2.5 Å) data collection was obtained at room temperature from an optimization screen (Walter *et al.*, 2005) of PACT Premier (Molecular Dimensions Ltd, USA) condition No. 64 [20%(*w*/*v*) PEG 3350, 100 m*M* bis-tris propane pH 6.5 and 200 m*M* potassium thiocyanate] with 6% sucrose added. Owing to the fragile nature of the thin crystal, we pipetted 0.5 µl cryoprotectant solution (20% glycerol in reservoir buffer) directly onto the nanodrop (0.3 µl volume; 2:1 protein:reservoir ratio). Then, using a cryoloop (0.2 mm diameter), the crystal was gently removed and flash-frozen in a cryostream jet (Oxford Instruments).

3. Data collection and processing

Data were collected on the MRC UK-operated BM14 beamline at the ESRF synchrotron (Grenoble, France). The crystal diffracted to 2.5 Å resolution but showed the presence of two predominant crystal lattices (Fig. 1) from two stacked crystals with the same lattice constants. *HKL*-2000 (Otwinowski & Minor, 1997) was used to process the images, confirming that the crystals belonged to space group $P2_12_12$ with unit-cell parameters $a = 164.0$, $b = 77.6$, $c = 127.8$ Å, similar to previously collected data (Abrescia *et al.*, 2005). 672 images were recorded, 312 with 1° oscillation and 30 s exposure and the remainder with 0.5° oscillation and 35 s exposure, to produce a high-redundancy data set and to minimize the overlap of spots from the multiple lattices. The two major lattices per image were processed independently and then scaled together with *HKL*-2000 as if they were derived from two different sets of images. Since a small fraction of the reflections from each lattice overlapped reflections from the other and not all of these data were rejected at the stage of data integration, an iterative process was used to scale the data (four rounds of *SCALEPACK*). During this process the error model was kept fixed at a set of values appropriate to the detector and experimental conditions (taking into account experience from other projects using the same beamline and detector). The initial values were never readjusted to reduce the χ^2, knowing that the presence of overlapping lattices introduces a subset of reflections which, when compared with non-overlapped symmetry equivalents, will show large discrepancies. As a result, the χ^2 was initially relatively high (overall $\chi^2 = 2.4$); however, after the four rounds of weeding out reflections the process converged (overall $\chi^2 = 1.4$) and the number of rejected reflections stabilized at ∼102 400 (less than 5% of the total number of observations). The resulting merged data set was of reasonable quality [2.5 Å resolution, $\langle I/\sigma(I)\rangle = 16.3$, completeness = 94.6%, $R_{\rm merge} = 25.5\%$, 30-fold multiplicity; for more details, see Table 1 in Abrescia *et al.* (2008) and see below for the quality of the electron-density map].

4. Structure determination

Analysis of the self-rotation function, native Patterson and solvent-content calculations suggested that the crystallographic asymmetric unit contained two trimers, each with the molecular threefold aligned with the *c* axis (Abrescia *et al.*, 2005). The atomic models of the trimers of the major capsid proteins of PRD1 and STIV (Benson *et al.*, 2002; Khayat *et al.*, 2005) were fitted by hand into the averaged electron density of the virus corresponding to the capsid protein P2 and were used as MR templates in programs such as *X-PLOR* (Brünger, 1992), *MOLREP* and *Phaser* (Collaborative Computational Project, Number 4, 1994) run at different resolutions. Despite finding solutions for which the packing was plausible, none of the candidates led to interpretable electron density after twofold averaging and phase extension. We therefore decided to use as our search model the electron density of the P2 trimer cut out from the averaged virus electron density at 7 Å resolution (600 × 600 × 600 grid points with cell dimensions $a = 900$, $b = 680$, $c = 1070$ Å, $\beta = 102°$). To do so, we first averaged the electron densities of the four trimers within the icosahedral asymmetric unit using the *General Averaging Program* (*GAP*; D. I. Stuart & J. M. Grimes, unpublished work; software for computers running Unix is available on request from the authors). The molecular envelope was based on a pseudo-atomic model of the P2 trimer created by manually adjusting and trimming another trimeric protein, the P3 major capsid protein of PRD1, to produce a model that filled the low-resolution P2 density (a 5 Å radius was included around each atom). The averaged electron density for the trimer was then placed into a *P*1 cell (126 × 126 × 126 grid points with cell dimensions $a = b = c = 150$ Å, $\alpha = \beta = \gamma = 90°$; Fig. 2*a*), back-

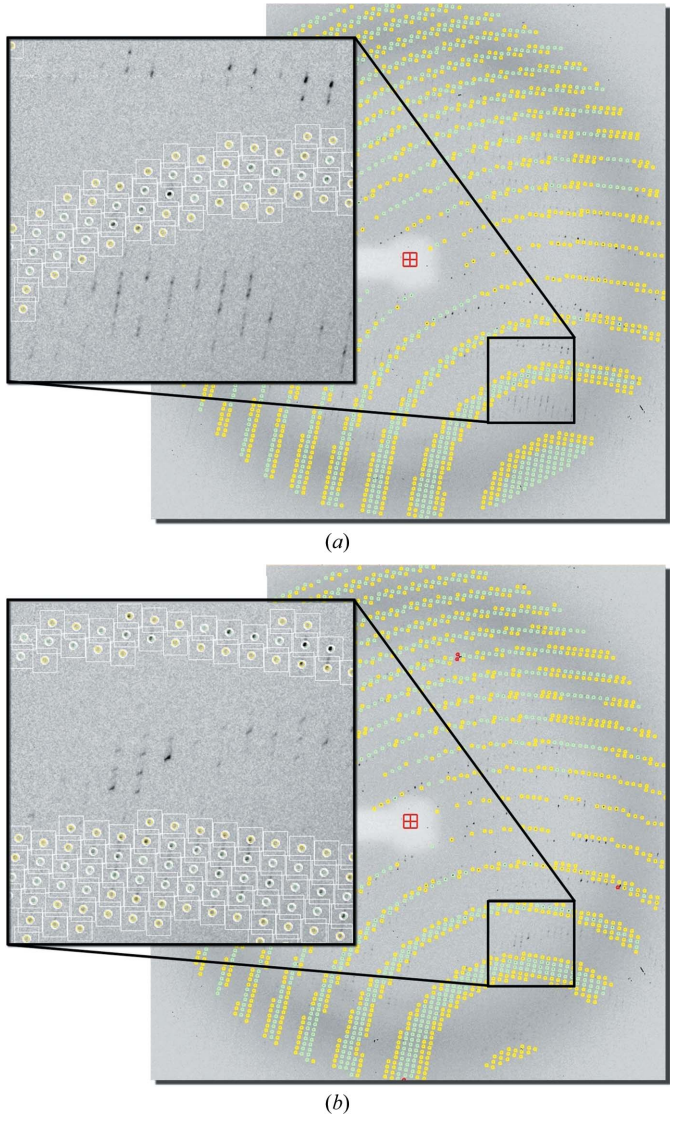

Figure 1
A typical diffraction image showing partially overlapped reciprocal lattices. (*a*) Predicted spots belonging to the first lattice. (*b*) Predicted spots for the second lattice.

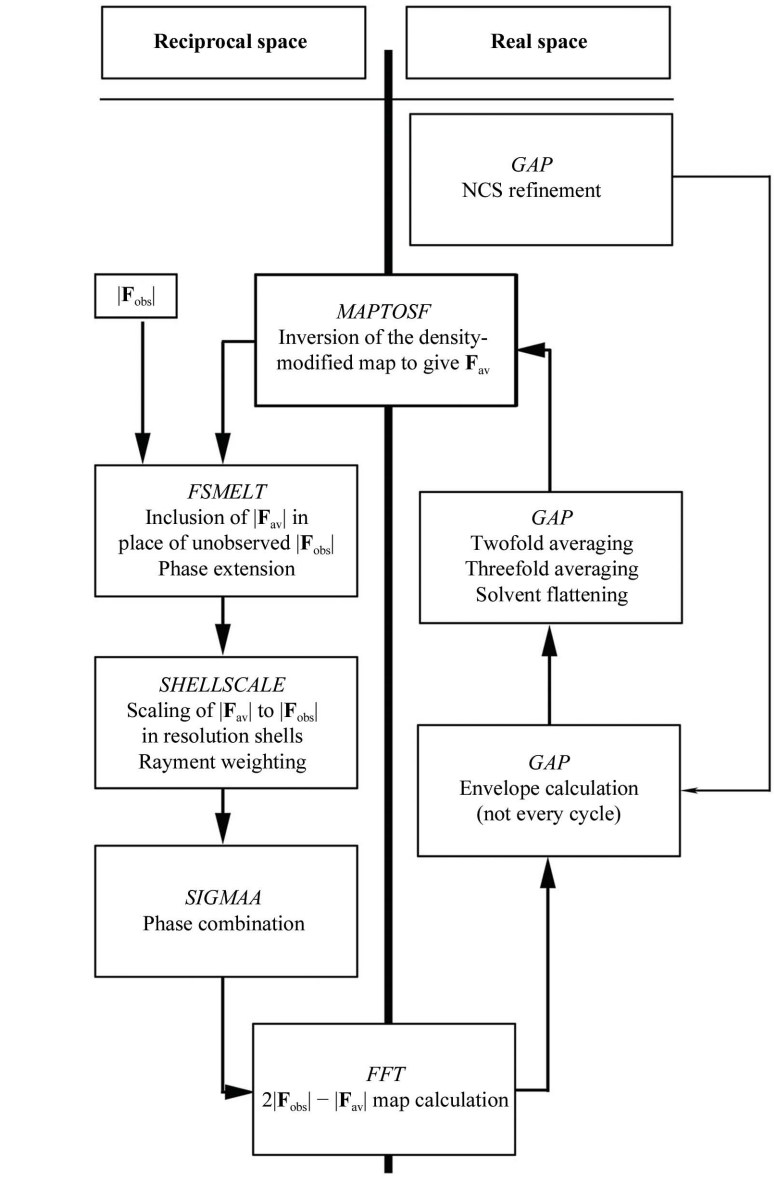

Figure 2
Structure-determination protocol and map quality. (*a*) Flowchart of the structure-determination and averaging protocols (adapted from Diprose, 2000). (*b*) Stereoview of the low-resolution map of a P2 trimer whose structure factors and phases were used as a search model in *Phaser* viewed from the top and along the molecular threefold axis.

transformed (20 961 reflections between 150 and 7 Å) using *SFALL* (Collaborative Computational Project, Number 4, 1994) and the resultant structure-factor amplitudes and phases were used in *Phaser* (Collaborative Computational Project, Number 4, 1994; see Fig. 2*b*). MR carried out using reflections in the 30–7.6 Å resolution range (the resolution limit was set automatically by *Phaser*) produced a sensible top solution with two trimers in the asymmetric unit. The conclusive proof of correctness was the ability to obtain a traceable map at 2.5 Å by a combination of cycling averaging, solvent flattening and phase extension using NCS (see below).

5. Phase extension

Cyclic averaging, solvent flattening, phase extension and NCS refinement were carried out using *GAP*; the procedure is summarized in Fig. 2(*a*). Initial iterative sixfold NCS averaging, solvent flattening and phase extension in resolution steps of 1/2000 Å of the $(2F_o - F_c)\exp(i\psi_c)$ map obtained in *Phaser* did not produce interpretable electron density even when the initial phases at 7.6 Å were recombined with the new set of averaged phases in *SIGMAA* (Collaborative Computational Project, Number 4, 1994) and the NCS operators were refined during the process.

Successful phase determination was achieved when the NCS operators derived from *Phaser* were refined against the observed structure amplitudes at 7.6 Å resolution (using a steepest-descent search procedure in real space with the correlation coefficient as the target function in *GAP*) prior to averaging and phase extension. Refinement of the translation vector between the two trimers produced little improvement, whilst refinement of the relative orientations and rotational register of the two trimers led to a significant decrease in the crystallographic *R* factor and an increase in the real-space correlation coefficient (cycle 0 in Table 1). In this step the proper threefold symmetry of the trimer was unlocked: this was fundamental, providing a proper description of the NCS. The importance of determining the orientation accurately can be appreciated if one models the P2 trimer as a disc of radius 36 Å; a 1° error in the rotation of this disc then leads to a 90° phase error in the contribution of atoms on the periphery of the disc to reflections with a Bragg spacing of 2.5 Å.

With the newly refined NCS operators, the cycling averaging was restarted and the phases were gradually extended to 2.5 Å resolution. The NCS operators were repeatedly re-refined during averaging (Table 1) and molecular envelopes were recalculated every fifth cycle. This process converged after 450 cycles, producing an excellent electron-density map (Figs. 3 and 4). The atomic model was built manually (*Coot*; Emsley & Cowtan, 2004), refined initially in *REFMAC*

Table 1
NCS refinement statistics.

R and CC represent the real-space *R* factor and correlation coefficient between the NCS (i = initial, f = final; the superscripts represent two of the threefold operators).

Cycle	Resolution (Å)	Twofold	Threefold (unlocked)
0	7.60	R_i = 30.6%, CC_i = 91.1%	R_i^2 = 43.5%, CC_i^2 = 81.6%
		R_f = 29.3%, CC_f = 91.8%	R_f^2 = 37.0%, CC_f^2 = 86.9%
			R_i^3 = 46.6%, CC_i^3 = 78.7%
			R_f^3 = 37.3%, CC_f^3 = 86.7%
117	5.25	R_i = 29.5%, CC_i = 91.9%	R_i^2 = 32.5%, CC_i^2 = 90.2%
		R_f = 28.5%, CC_f = 92.5%	R_f^2 = 31.7%, CC_f^2 = 90.6%
			R_i^3 = 32.0%, CC_i^3 = 90.4%
			R_f^3 = 31.7%, CC_f^3 = 90.6%
152	4.79	R_i = 31.6%, CC_i = 90.7%	R_i^2 = 35.2%, CC_i^2 = 88.5%
		R_f = 31.6%, CC_f = 90.7%	R_f^2 = 35.0%, CC_f^2 = 88.6%
			R_i^3 = 35.2%, CC_i^3 = 88.5%
			R_f^3 = 35.0%, CC_f^3 = 88.6%

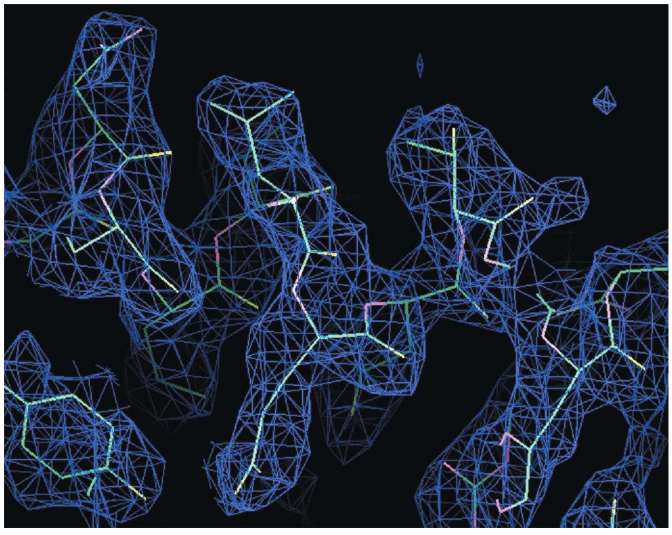

Figure 3
Detail of the experimentally phased map (blue) to 2.5 Å resolution contoured at ∼1.2σ. The manually built atomic model is shown in stick representation.

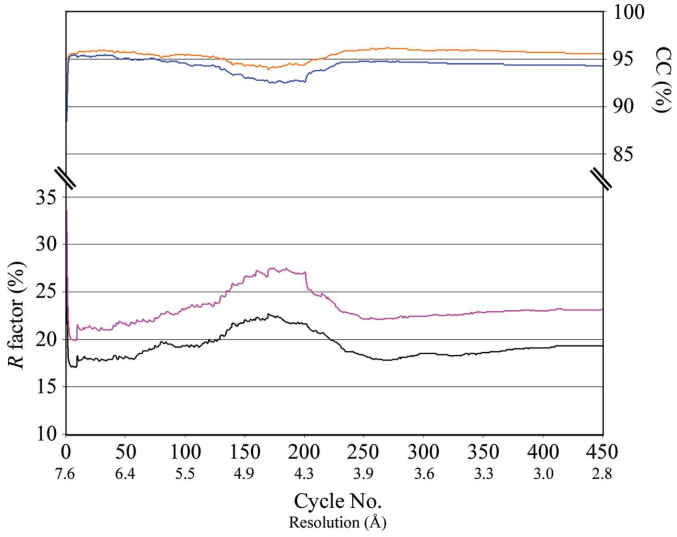

Figure 4
Black and magenta lines illustrate the evolution of the *R* factor corresponding to the twofold and threefold averaging during cycling, whilst blue and orange lines describe the real-space correlation coefficient for the same cycles. The improvement as the phase extension proceeds beyond about 4.5 Å resolution may correlate with the point at which β-strands become resolved (in this predominantly β protein).

(Collaborative Computational Project, Number 4, 1994) and finalized in *BUSTER* (Roversi *et al.*, 2000). The coordinates and structure factors for the P2 protein have been deposited in the Protein Data Bank with accession code 2vvf (Abrescia *et al.*, 2008).

We may estimate an upper bound on the errors in the phases determined with this density-modification procedure by comparing these phases with those derived from the final refined model (Fig. 4). The observed agreement corresponds to a figure of merit (Drenth, 1999) of 0.74 (calculated as the cosine of the mean phase difference of 42° between the two sets of phases; for perfect phases the figure of merit is 1).

6. Conclusions

We have solved the structure of a trimeric protein composed of 32 kDa protein subunits using a low-resolution (7.6 Å) electron-density map as a template for molecular replacement and phase extension to 2.5 Å by means of noncrystallographic symmetry, an approach that is likely to have special application where EM reconstructions are available.

We thank M. Bahar for help with data collection and the staff of BM14 (ESRF, Grenoble, France) for beamline support. This work was supported by Academy of Finland Centre of Excellence Program in Virus Research Grant 1129684 (2006–2011 to JKHB and DHB) and Academy of Finland Grant 1127665 (to HMO), the Medical Research Council UK (for support of DIS and NGAA), a grant from the Human Frontiers Science Program (RGP0320/2001-M), the Spanish Ministerio de Ciencia y Tecnología (BFU2009-08123 to NGAA) and the Basque Government (PI2010-20 to NGAA). JMG was supported by FP7 SPINE2-Complexes (LSHGCT-2006-031220).

References

Abrescia, N. G., Cockburn, J. J., Grimes, J. M., Sutton, G. C., Diprose, J. M., Butcher, S. J., Fuller, S. D., San Martín, C., Burnett, R. M., Stuart, D. I., Bamford, D. H. & Bamford, J. K. (2004). *Nature (London)*, **432**, 68–74.
Abrescia, N. G., Grimes, J. M., Fry, E. E., Ravantti, J. J., Bamford, D. H. & Stuart, D. I. (2010). *Emerging Topics in Physical Virology*, edited by P. Stockley & R. Twarock, pp. 35–58. London: Imperial College Press.
Abrescia, N. G., Grimes, J. M., Kivelä, H. M., Assenberg, R., Sutton, G. C., Butcher, S. J., Bamford, J. K., Bamford, D. H. & Stuart, D. I. (2008). *Mol. Cell*, **31**, 749–761.
Abrescia, N. G. A., Kivelä, H. M., Grimes, J. M., Bamford, J. K. H., Bamford, D. H. & Stuart, D. I. (2005). *Acta Cryst.* F**61**, 762–765.
Bamford, D. H., Burnett, R. M. & Stuart, D. I. (2002). *Theor. Popul. Biol.* **61**, 461–470.
Bamford, D. H., Grimes, J. M. & Stuart, D. I. (2005). *Curr. Opin. Struct. Biol.* **15**, 655–663.
Benson, S. D., Bamford, J. K. H., Bamford, D. H. & Burnett, R. M. (2002). *Acta Cryst.* D**58**, 39–59.
Brünger, A. T. (1992). *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR.* Yale University, Connecticut, USA.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Diprose, J. M. (2000). *Structural Studies on Orbiviruses.* DPhil thesis, University of Oxford.
Drenth, J. (1999). *Principles of Protein X-ray Crystallography.* Heidelberg: Springer-Verlag.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Fry, E., Abrescia, N. G. & Stuart, D. I. (2007). *Macromolecular Crystallography*, edited by M. R. Sanderson & J. V. Skelly, pp. 245–263. Oxford University Press.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Khayat, R., Tang, L., Larson, E. T., Lawrence, C. M., Young, M. & Johnson, J. E. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 18944–18949.
Kivelä, H. M., Abrescia, N. G., Bamford, J. K., Grimes, J. M., Stuart, D. I. & Bamford, D. H. (2007). *J. Struct. Biol.* **161**, 204–210.

Kivelä, H. M., Kalkkinen, N. & Bamford, D. H. (2002). *J. Virol.* **76**, 8169–8178.
Lee, K. K. & Johnson, J. E. (2003). *Curr. Opin. Struct. Biol.* **13**, 558–569.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J. & Baker, D. (2007). *Nature (London)*, **450**, 259–264.
Rossmann, M. G. (1995). *Curr. Opin. Struct. Biol.* **5**, 650–655.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Roversi, P., Blanc, E., Vonrhein, C., Evans, G. & Bricogne, G. (2000). *Acta Cryst.* D**56**, 1316–1323.
Trapani, S., Abergel, C., Gutsche, I., Horcajada, C., Fita, I. & Navaza, J. (2006). *Acta Cryst.* D**62**, 467–475.
Trapani, S., Schoehn, G., Navaza, J. & Abergel, C. (2010). *Acta Cryst.* D**66**, 514–521.
Villeret, V., Tricot, C., Stalon, V. & Dideberg, O. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 10762–10766.
Walter, T. S. *et al.* (2005). *Acta Cryst.* D**61**, 651–657.