Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Severe diffraction anisotropy, rotational pseudosymmetry and twinning complicate the refinement of a pentameric coiled-coil structure of NSP4 of rotavirus

Anita R. Chacko,[a] Peter H. Zwart,[b] Randy J. Read,[c] Eleanor J. Dodson,[d] C. D. Rao[e] and Kaza Suguna[a]*

[a]Molecular Biophysics Unit, Indian Institute of Science, C. V. Raman Avenue, Bangalore, Karnataka 560 012, India, [b]Lawrence Berkeley Laboratory, One Synchrotron Road, Berkeley, CA 94720, USA, [c]Department of Haematology, University of Cambridge, Cambridge Institute for Medical Research, Wellcome Trust MRC Building, Hills Road, Cambridge CB2 2XY, England, [d]Department of Chemistry, York Structural Biology Laboratory, University of York, Heslington, York YO10 5YW, England, and [e]Department of Microbiology and Cell Biology, Indian Institute of Science, C. V. Raman Avenue, Bangalore, Karnataka 560 012, India

Correspondence e-mail: suguna@mbu.iisc.ernet.in

Received 23 June 2012
Accepted 7 September 2012

The crystal structure of the region spanning residues 95–146 of the rotavirus nonstructural protein NSP4 from the asymptomatic human strain ST3 was determined at a resolution of 2.5 Å. Severe diffraction anisotropy, rotational pseudosymmetry and twinning complicated the refinement of this structure. A systematic explanation confirming the crystal pathologies and describing how the structure was successfully refined is given in this report.

1. Introduction

The first virus-encoded enterotoxin (Ball *et al.*, 1996; Tian *et al.*, 1996), nonstructural protein 4 (NSP4) of rotavirus, is associated with diverse functions such as membrane destabilization (Newton *et al.*, 1997; Tian *et al.*, 1996), diarrhoea induction (Ball *et al.*, 1996; Horie *et al.*, 1999), double-layered particle binding (Au *et al.*, 1993; Chan *et al.*, 1988; O'Brien *et al.*, 2000; Taylor *et al.*, 1993), calcium binding (Estes *et al.*, 2001), intracellular calcium mobilization (Tian *et al.*, 1994; Dong *et al.*, 1997), VP4 binding (Estes *et al.*, 2001), inhibition of the microtubule-mediated secretory pathway (Xu *et al.*, 2000) and interaction with extracellular proteins such as laminin $\beta3$ and fibronectin (Boshuizen *et al.*, 2004), integrins $\alpha1\beta1$ and $\alpha2\beta1$ (Seo *et al.*, 2008), calnexin (Mirazimi *et al.*, 1998), tubulin (Xu *et al.*, 2000) and caveolin (Parr *et al.*, 2006; Mir *et al.*, 2007).

The pleotropic functional nature of this protein is reflected in the variety of structures adopted by the fragment consisting of residues 95–146 of this protein. It was reported to be a tetrameric coiled coil in the case of the simian strain SA11 of rotavirus (PDB entries 1g1j and 2o1k; Deepa *et al.*, 2007; Bowman *et al.*, 2000) and the human asymptomatic strain I321 of rotavirus (PDB entry 2o1j; Deepa *et al.*, 2007). Recently, this fragment of NSP4 from the human asymptomatic strain ST3 of rotavirus was reported as a pentameric coiled coil for the first time (PDB entry 3miw; Chacko *et al.*, 2011).

After overcoming the initial challenge with molecular replacement (Chacko *et al.*, 2012), we encountered several pathologies associated with the data set that complicated the refinement process in the case of the pentameric structure. The manner in which the data pathologies were identified and overcome is described in this paper.

2. Materials and methods

2.1. Data collection

Data were collected at 100 K to a resolution of 3.2 Å at the home source on a MAR300 imaging-plate system using Cu $K\alpha$ radiation of 1.54 Å wavelength generated by a Rigaku RU-200

Table 1
Crystal data and data-collection statistics obtained at the synchrotron and the home source.

Values in parentheses are for the highest resolution shell.

	Synchrotron		Home source	
Temperature (K)	100	100	100	100
Point group	422	4	422	4
Unit-cell parameters (Å)				
a	63.33	63.46	63.66	64.17
b	63.33	63.46	63.66	64.17
c	122.86	123.21	122.90	123.02
Matthews coefficient (V_M; Å^3 Da^{-1})	1.90	1.90	1.92	1.95
Solvent content (%)	35.2	35.5	35.8	37.0
No. of molecules in asymmetric unit	1 pentamer	2 pentamers	1 pentamer	2 pentamers
Resolution range (Å)	50–2.5 (2.6–2.5)	50–2.5 (2.6–2.5)	63.7–3.5 (3.7–3.5)	64.2–3.2 (3.4–3.2)
Observed reflections	100818	106383	46164	52206
No. of unique reflections	8158 (785)	16133 (1254)	3557 (496)	8224 (1202)
Completeness (%)	97.7 (96.4)	95.1 (74.6)	100 (100)	99.7 (99.3)
Multiplicity	12.4 (9.4)	6.6 (4.7)	13.0 (13.3)	6.3 (6.3)
$\langle I/\sigma(I)\rangle$	31.0 (4.1)	22.2 (2.2)	10.5 (6.1)	13.7 (4.1)
R_{merge}† (%)	5.5 (20.0)	5.5 (18.7)	20.1 (45.0)	9.9 (48.7)
Unique reflections after anisotropic truncation	6926	12629		
Completeness after anisotropic truncation (%)	84	75		
Completeness in the ellipsoid (%)	100	100		

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is its mean intensity.

X-ray generator operating at 44 kV and 70 mA and focused with an Osmic mirror system. A total of 180 frames with an oscillation range of 1° were collected with a crystal-to-detector distance of 250 mm. Subsequently, 180 frames with an oscillation range of 1° were collected using a 1 Å wavelength at 100 K at SPring-8, Japan, yielding a data set to a resolution of 2.5 Å. The crystal-to-detector distance was 184 mm. The data were indexed, integrated and scaled using *DENZO* and *SCALEPACK* from the *HKL*-2000 suite (Otwinowski & Minor, 1997; Table 1).

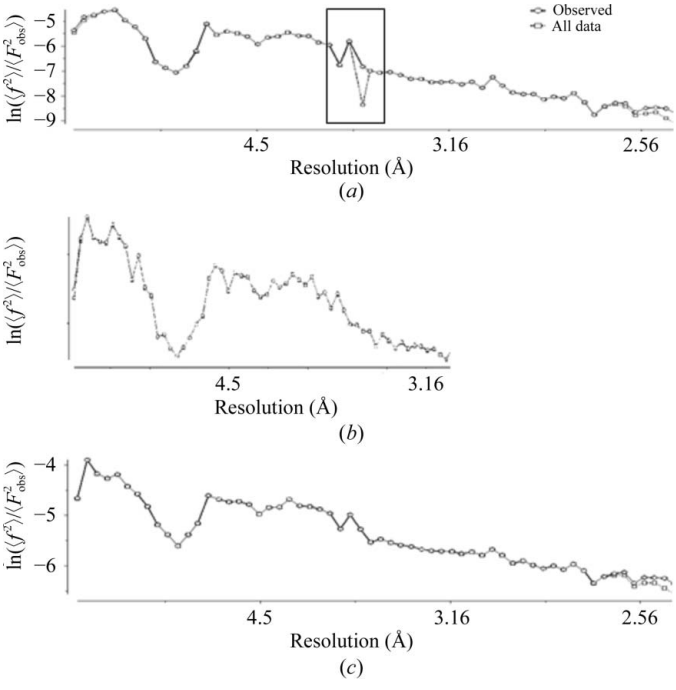

Figure 1
Wilson plots of (*a*) the synchrotron data (showing the sharp dip) before merging extra reflections from the home source, (*b*) the data collected at the home source and (*c*) the synchrotron data after merging extra reflections from the home source.

2.2. Computations

Model building was carried out with *Coot* (Emsley & Cowtan, 2004) and refinement was carried out with *CNS* v.1.2 (Brünger *et al.*, 1998) and *PHENIX* (Adams *et al.*, 2002). Superposition of the structures was carried out by aligning the main-chain atoms using *ALIGN* (Cohen, 1997) and *SSM* superposition (Krissinel & Henrick, 2004) in *Coot*.

3. Results and discussion

The molecular-replacement solution was not straightforward and was obtained from data in space group $P4_22_12$ (Chacko *et al.*, 2012) using *Phaser* (McCoy *et al.*, 2007). Discussed below are the steps in identifying the inherent crystal pathology, which led to the determination of the true space group ($P4_2$) and ultimately resulted in successful refinement of the structure.

3.1. Improving data completeness

A close examination of the synchrotron data revealed serious overloads and a missing shell of reflections corresponding to the ice ring. This was reflected as a sharp dip in the Wilson plot (Fig. 1). The missing reflections were taken from the home-source data and appended to the synchrotron data set after shell-wise scaling. These merged data (Fig. 1) with 277 further reflections (the completeness increased from 97.7% to 98.7%) were used in refinement.

3.2. Diffraction data analysis

In space group $P4_22_12$, the Matthews coefficient (Matthews, 1968) indicated the possibility of one pentamer in the

asymmetric unit and a solvent content of 35%. The self-rotation peaks were strongly masked by the crystallographic fourfold and twofold peaks and hence no information could be gained about the oligomeric state and the symmetry of the molecule. Since a clear molecular-replacement solution was obtained in space group $P4_22_12$, with a rotational Z-score of 5.6, a translational Z-score of 26.6 and an LLG of 1121 (Chacko *et al.*, 2012), model building and refinement were carried out in this space group. However, the R factors could not be improved beyond an R_{work} of 31.2 and an R_{free} of 39.0%. The map looked clean, with no significant unaccounted density and little scope for further building. TLS refinement was not stable and resulted in abnormally high R factors. Twinning is not a possibility in this space group ($P4_22_12$) and therefore there is no need for this correction to be applied. All of these indications suggested that there was probably some other anomaly in the data which had to be taken care of before the structure could be refined any further.

3.3. Anisotropy

Analysis of the data using the anisotropy-correction server (http://www.doe-mbi.ucla.edu/~sawaya/anisoscale/) revealed that the data had severe diffraction anisotropy. The server, which identified the data to be stronger along the c^* axis than in the a^*/b^* directions, detected an anisotropic δB of 52.09 Å^2. The pseudoprecession images clearly reflect the anisotropic nature of the data, with more data observed along the c^* direction in the $h0l$ or $0kl$ planes, whereas the data appeared to be isotropic in the $hk0$ plane (Fig. 2). The output of *CTRUNCATE* (French & Wilson, 1978) in *CCP*4 (Winn *et al.*, 2011) implied serious anisotropy, with eigenvalue ratios of 0.3, 0.3, 1, which clearly deviate from the isotropic value of 1.

Because of the presence of severe anisotropy, a three-dimensional ellipsoid was defined in order to use as many significant data measurements as possible, rather than the usual sphere used for scaling and refinement. The total number of reflections reduced from 16 316 to 12 629 (Table 1) and the resolution limits were reduced to 2.9 and 2.5 Å along the a^* and b^*/c^* directions, which is where $F/\sigma(F)$ fell below 3.0, defining the resolution limits for ellipsoidal truncation of data. The correction was applied to exclude very weak data so that better electron-density maps could be generated for further model building and refinement at this stage. In order to make the data isotropic, B-factor corrections of 17.36, 17.36 and −34.73 Å^2 along the a^*, b^* and c^* directions, respectively, were applied to the observed structure factors, resulting in improved eigenvalue ratios of 0.79, 0.79, 1 and implying that the data were now more isotropic. This correction of the data in space group $P4_22_12$ lowered the R

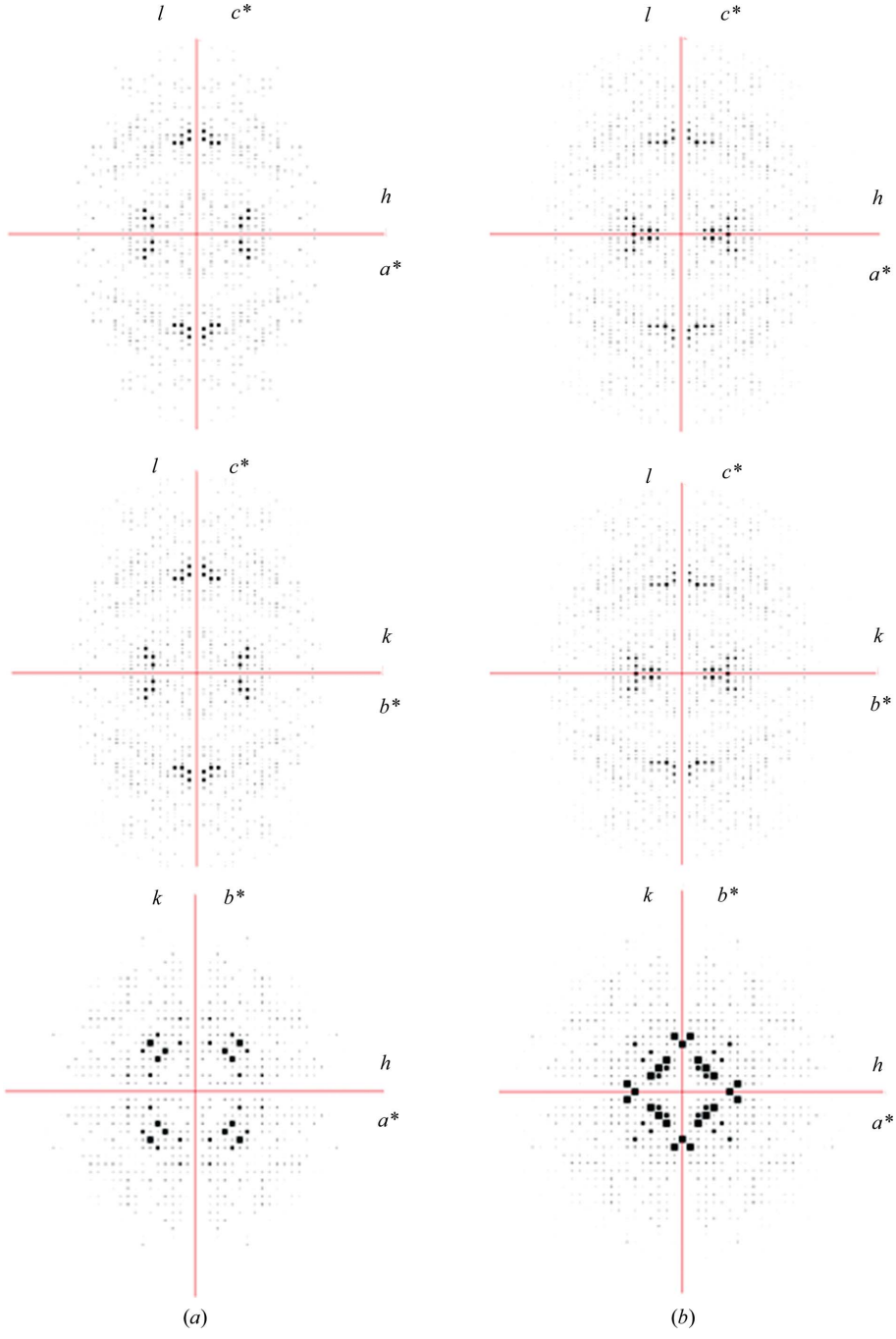

Figure 2
Pseudo-precession images of the data in the $h0l$, $0kl$ and $hk0$ planes showing the anisotropic nature of the data (*a*) before and (*b*) after adding the extra reflections corresponding to the ice ring and overloads.

Table 2
Intensity statistic tests for the detection of twinning.

		Theoretical	
	Observed	Untwinned	Perfect twin
Acentric reflections			
$\langle I^2\rangle/\langle I\rangle^2$	2.532	2.000	1.500
$\langle F\rangle^2/\langle F^2\rangle$	0.714	0.785	0.885
$\langle\lvert E^2-1\rvert\rangle$	0.869	0.736	0.541
Centric reflections			
$\langle I^2\rangle/\langle I\rangle^2$	3.943	3.000	2.000
$\langle F\rangle^2/\langle F^2\rangle$	0.564	0.637	0.785
$\langle\lvert E^2-1\rvert\rangle$	1.204	0.968	0.736
$\langle\lvert L\rvert\rangle$	0.541	0.500	0.375
$\langle\lvert L\rvert^2\rangle$	0.386	0.333	0.200
H-test	0.48 (twin fraction)		
Britton analysis	0.45 (twin fraction)		

factor to 30.9% from 31.2%, but $R_{\rm free}$ remained the same at about 39%.

Anisotropic scaling corrections have been used in many cases to account for possible disorder or movement within the protein that generates high *B* factors and leads to data anisotropy, as observed in the DsbC protein from *Haemophilus influenzae* (Zhang *et al.*, 2004), the PE–PPE protein complex from *Mycobacterium tuberculosis* (Strong *et al.*, 2006) and the complex between G-protein-coupled receptor kinase 2 and G$\beta\gamma$ (Lodowski *et al.*, 2003). When the number of lattice contacts is lower in one cell direction than another, diffraction anisotropy results (Sheriff *et al.*, 1987). This is also observed in the present case, where there are more lattice contacts along the *c* axis compared with the *a* or *b* axes (Fig. 3).

3.4. Rotational pseudosymmetry

Little support could be gained from comparing the $R_{\rm merge}$ values of the synchrotron data in point groups 422 and 4, since the $R_{\rm merge}$ was 5.5% in both cases. 422, being the higher point group, was chosen for molecular replacement and initial refinement runs. The *R* factors remained high even after anisotropic correction, implying that there was some other inherent problem. Suspecting rotational pseudosymmetry (RPS), we refined the data in space groups $P4_2$ and $P2_12_12$ (subgroups of $P4_22_12$), where clear solutions were obtained in molecular replacement. RPS occurs when the point-group symmetry of the lattice is higher than that of the crystal (Zwart *et al.*, 2008), as observed in the structure of a cyclic nucleotide-modulated (HCN) pacemaker channel (PDB entry 1q43; Zagotta *et al.*, 2003), similar to the present case. Zwart *et al.* (2008) suggested that in such cases refinement should be carried out in both lower and higher symmetry space groups and the resulting *R* factors and model-quality indicators should be compared. A significant drop in $R_{\rm work}$ and $R_{\rm free}$ was observed in the case of the structure of superoxide dismutase from *Pyrobaculum aerophilum*, where the noncrystallographic twofold axis deviates from the crystallographic twofold axis by a rotation of 2° (Lee *et al.*, 2003) when refined in the true space group. In calcium-depleted C-reactive protein (Ramadan *et al.*, 2002) this angle is only 0.6°. A similar situation of RPS occurred in the DUF55 domain of human thymocyte nuclear protein 1 (Yu *et al.*, 2009). In the present case, changing the space group to $P4_2$ reduced the $R_{\rm work}$ and $R_{\rm free}$ values to 29.9% and 35.1%, respectively. The two molecules in the asymmetric unit appear to be related by rotational pseudosymmetry, which should be probed further. However, refinement in space group $P2_12_12$ yielded similar statistics as refinement in space group $P4_22_12$ and could not be improved.

Since no significant off-origin peak (9.66% of the height of the origin peak) and a *p*-value (height) of >0.1955 (Zwart & Adams, 2005) was observed in the self-Patterson map, the possibility of translational pseudosymmetry was ruled out and hence it seemed to be a case of rotational pseudosymmetry. Anisotropy correction of the data in the true space group, $P4_2$, reduced the $R_{\rm work}$ and $R_{\rm free}$ values to 28.4% and 29.9%, respectively. Since the MR solutions in both $P4_22_12$ and $P4_2$ had reasonable *R* values and good packing, it appears that the crystal suffers from pseudosymmetry.

3.5. Twinning

Unlike space group $P4_22_12$, twin laws are available in space group $P4_2$. Data analysis by *phenix.xtriage* showed the presence of merohedral twinning along the $h, -k, -l$ direction. Although twinning could not be detected using conventional tests owing to the presence of anisotropy and pseudosymmetry, it was confirmed by other tests as described below.

3.5.1. *L*-test and twin detection. The presence of rotational pseudosymmetry (especially when parallel to the twin law) or translational pseudosymmetry can countervail the effects of twinning on the intensity statistics, thus making it more difficult to detect the presence of twinning. Merohedral twinning and twin estimation can be complicated by pseudosymmetry and/or anisotropy, as in the cases of superoxide dismutase from *P. aerophilum* (Lee *et al.*, 2003), human semicarbazide-sensitive amine oxidase (Jakobsson *et al.*, 2005) and Rio1 kinase (Dauter *et al.*, 2005). The *L*-test to detect twinning

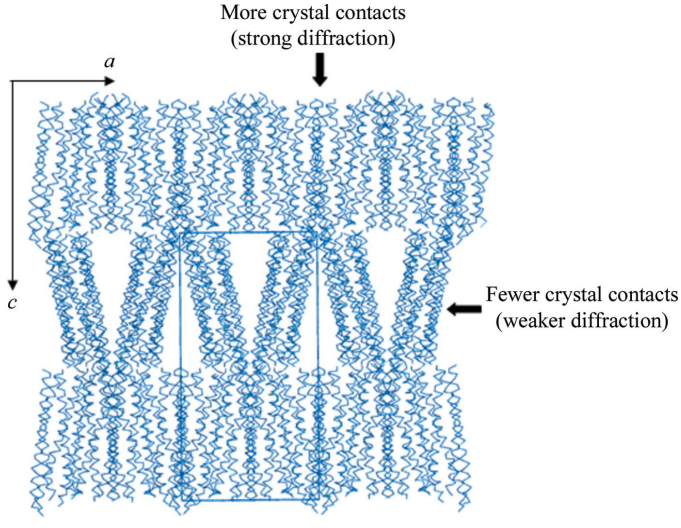

Figure 3
More crystal contacts are observed along the *c* axis, giving rise to data anisotropy.

Table 3
RvR test.

RvR test	Present case		Theoretical	
	R_{twin} (observed)	R_{twin} (calculated)	R_{twin} (observed) when data are twinned	R_{twin} (calculated) when data are twinned
RPS + perfect twin	0.036	0.34	∼0	<0.5
Misspecified crystal symmetry			0	0
No twinning			R_{twin} (observed) ≃ R_{twin} (calculated)	
Partial twinning			R_{twin} (observed) < R_{twin} (calculated)	
Perfect twinning			R_{twin} (observed) ≃ 0	
No RPS			R_{twin} (calculated) ≃ 1/2	
RPS			R_{twin} (calculated) < 1/2	

(Padilla & Yeates, 2003) is supposedly insensitive to crystal pathologies such as anisotropic diffraction and pseudo-centring. However, in the present case the *L*-test proved negative (Table 2). The $\langle|L|\rangle$ and $\langle|L^2|\rangle$ values were 0.541 and 0.386, respectively (the theoretical untwinned values are 0.500 and 0.333, respectively, and the twinned values are 0.375 and 0.200, respectively), ruling out the possibility of twinning. The traditional cumulative intensity plots (Fig. 4*a*) and the conventional intensity statistics also failed to detect twinning (Table 2). However, a high value of 5.63 for the multivariate *Z*-score or the Mahalanobis distance (Mahalanobis, 1936), which is the multivariate mean in units of standard deviation, indicated that the ($\langle|L|\rangle$, $\langle|L^2|\rangle$) pair is outside the range expected for experimental data sets and hence showed the presence of twinning. *Z*-score values larger than 3 indicate a twin fraction greater than 6% (Zwart & Adams, 2005). The *L*-test thus did not directly detect the presence of any twinning, but the high *Z*-scores implied the presence of twinning.

3.5.2. *H*-test and Britton plots to estimate the twin fraction. The twin fraction can be estimated by the *H*-test (Yeates, 1988, 1997) and the Britton analysis (Fisher & Sweet, 1980) in addition to the *L*-test. The *H*-test and Britton plot have proved to be more reliable than the *L*-test (Zwart & Adams, 2005) in estimating larger twin fractions. Twinning fractions of 0.48 and 0.45 were detected by the *H*-test and Britton analysis, respectively (Figs. 4*b* and 4*c*). The ML α test, which is a likelihood-based twin-estimation method, also detected a twin fraction of 0.48.

3.5.3. The RvR test. When twinning occurs in association with RPS, the RvR test can be used to distinguish between true twinning and misspecified crystal symmetry (Lebedev *et al.*, 2006). In this method, R_{twin} (observed) and R_{twin} (calculated) values are compared. In cases where both RPS and perfect twinning occur, R_{twin} (observed) will be approximately zero and R_{twin} (calculated) will be less than 0.5 (Lebedev *et al.*, 2006). In the present case, the values of R_{twin} (observed) and R_{twin} (calculated) were 0.036 and 0.34, respectively, clearly indicating the presence of both twinning and RPS (Table 3). Misspecified crystal symmetry would give R_{twin} (observed) and R_{twin} (calculated) values close to zero. The presence of twinning and RPS was also identified using this test in the case of homologous antibody Fv fragments (Brooks *et al.*, 2008).

The accuracy of deconvoluting twinned data suffers when the twin fraction is close to 0.5. All of the tests described above estimated the twin fraction to be less than 0.5, but the refined value of 0.499 from *phenix.refine* was very close to 0.5. However, twin correction using this value by the program *PHENIX* resulted in significant improvement of the refinement statistics and the quality of the electron-density maps.

3.6. Refinement

Although the MR solution was very clear, refinement always resulted in high R_{work} and R_{free} values with no apparent

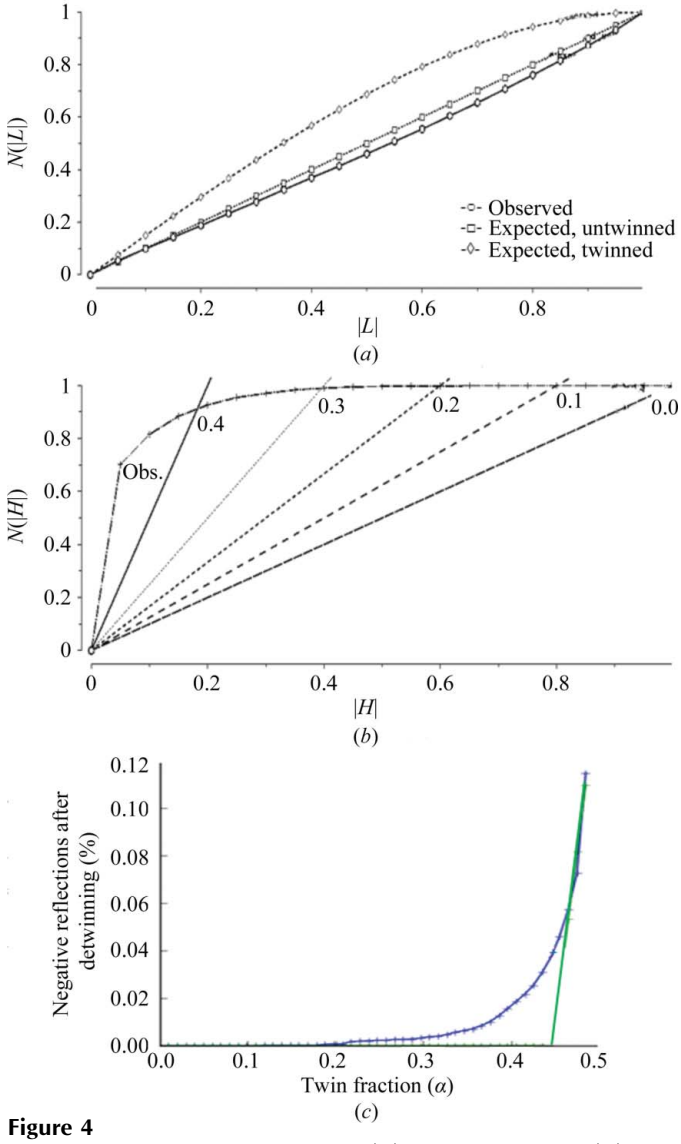

Figure 4
(*a*) Twinning analysis using the *L*-test: $|L|$ is plotted against $N(|L|)$. *L* is given by $(I_1 - I_2)/(I_1 + I_2)$, where I_1 and I_2 are the intensities of unrelated reflections and $N(|L|)$ is the cumulative probability distribution of *L*. (*b*) The *H*-test: $|H|$ is plotted against $N(|H|)$, where $H = |I_1 - I_2|/(I_1 + I_2)$, I_1 and I_2 are the intensities of twin-related reflections and $N(|H|)$ is the cumulative distribution of *H*. From the observed values, the twin fraction is estimated to be 0.48. (*c*) The Britton plot: percentage of twin fractions. Extrapolation of the curve (green line) yields a twin fraction of 0.45.

way to improve the coordinate set or the R factor until the presence of anisotropy, pseudosymmetry and twinning were detected. The refinement steps are described in Fig. 5. The model used for refinement was initially optimized for perfect fivefold symmetry with the interhelical angles close to 72°. NCS was applied to all ten chains during refinement to improve the observation-to-parameter ratio since the overall data completeness had fallen to 78% owing to the ellipsoidal truncation. This was released completely towards the end of the refinement. About ten residues at the C-terminus of each chain could not be traced in the electron-density maps owing to disorder, as observed in the previous structures of NSP4 (Deepa *et al.*, 2007). The geometry was initially kept relaxed and was gradually tightened in each cycle. In the final stages, each chain was treated as one TLS group and refinement was carried out using *PHENIX* (Adams *et al.*, 2002). Only after the above corrections had been applied did the TLS refinement became stable. Rotamer correction and weight optimization finally helped to improve the model. The R_{work} and R_{free} were 23.5% and 27.3%, respectively, using the anisotropically truncated data. Towards the final cycles, reflections excluded by ellipsoid truncation were included in the refinement and the final R_{work} and R_{free} were 26.1% and 29.7%, respectively (Table 4). In comparison, the R_{work} and R_{free} values without applying any corrections were 29.9% and 35.1%, respectively (Fig. 5).

Table 4
Refinement statistics for the synchrotron data.

Space group	$P4_2$
Resolution (Å)	28.4–2.5
No. of reflections	16272
R_{work}/R_{free}	0.261/0.297
No. of atoms	
Protein	3328
Ligand/ion	108
Water	162
B factors (Å^2)	
Protein	72.88
Ligand	72.73
Water	65.00
R.m.s. deviations	
Bond lengths (Å)	0.013
Bond angles (°)	1.508

3.7. *Post hoc* analysis

Pseudosymmetry, twinning and anisotropy are known to complicate structure solution and refinement of many structures (Dauter *et al.*, 2005; Jakobsson *et al.*, 2005; Lebedev *et al.*, 2006; Guelker *et al.*, 2009). It is interesting to observe that in the present case twinning, anisotropy and pseudosymmetry did not complicate the process of molecular replacement. The presence of a pseudosymmetric twofold operator close to a twinning operator made it difficult to detect twinning by conventional means; therefore, molecular replacement was carried out in the pseudo (apparent) space group $P4_22_12$. Indicators such as crystal morphology did not suggest the possibility of twinning, nor was there an abnormally low V_M value as detected in perfect twins (Yeates, 1997); hence, all phasing attempts by molecular replacement were carried out in space group $P4_22_12$ rather than $P4_2$.

Although the mode of packing seems to be the same for space group $P4_2$ with two pentamers in the asymmetric unit and for space group $P4_22_12$ with only one pentamer in the asymmetric unit, *post hoc* analysis revealed that the two pentamers in $P4_2$ were indeed different. Alignment of the first pentamer in space group $P4_2$ with the pentamer in space group $P4_22_12$ gives an r.m.s.d. of 0.44 Å and a rotation angle of 0.16°. Alignment of the second pentamer in space group $P4_2$ with the symmetry-related molecule of the pentamer in space group $P4_22_12$ gives an r.m.s.d. of 0.5 Å and a rotation angle of 1.8°. However, alignment of the

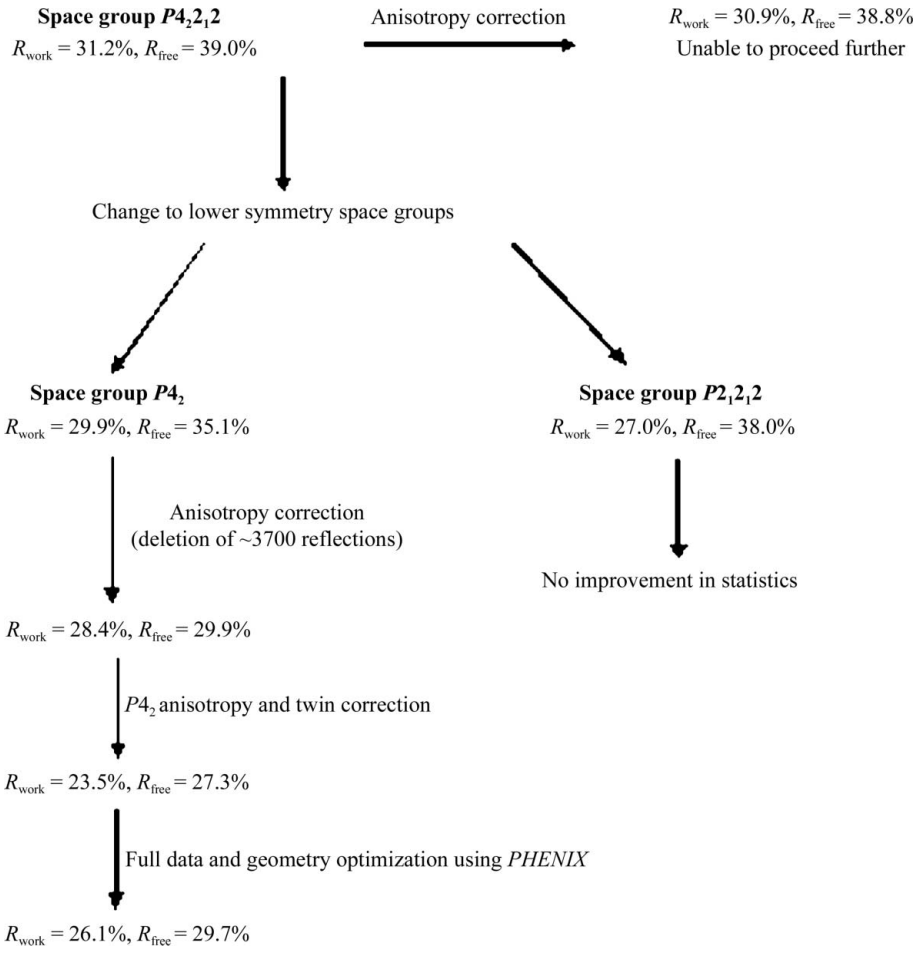

Figure 5
Various steps in the refinement process.

Table 5
Alignment parameters of the two pentamers in the asymmetric unit of space group $P4_2$.

Molecule aligned	Main-chain r.m.s.d. (Å)	Rotation angle (°)
Pentamer II on pentamer I	0.70	179.6
Pentamer I on $P4_22_12$ pentamer	0.44	1.6
Pentamer II on symmetry-related $P4_22_12$ pentamer	0.50	1.8

Table 6
Alignment parameters of the individual helices of two pentamers in the asymmetric unit of space group $P4_2$.

Alignment	Direction cosines of the fivefold axis	Angle w.r.t. the x, y, z axes (°)	Main-chain r.m.s.d. (Å)	Rotation angle (°)
A on *B*	0.9997 −0.0245 −0.0002	1.40, 91.40, 90.01	0.55	179.2
C on *D*	0.9997 −0.0241 0.0021	1.40, 91.38, 89.87	0.69	179.8
E on *F*	0.9998 −0.0167 0.0111	1.14, 90.95, 89.36	0.52	179.8
G on *H*	0.9983 −0.0574 −0.0055	3.34, 93.29, 90.31	0.74	179.1
I on *J*	0.9999 0.0027 −0.0020	0.81, 89.84, 90.11	0.68	179.1

Table 7
Alignment of successive helices of pentamer I and pentamer II.

Alignment	Direction cosines of the fivefold axis	Angle w.r.t the x, y, z axes (°)	Main-chain r.m.s.d. (Å)	Rotation angle (°)
Pentamer I				
B on *C*	−0.1092 0.1665 0.9799	83.73, 80.41, 11.50	0.58	69.1
C on *E*	−0.0403 0.1667 −0.9805	92.30, 80.40, 11.33	0.78	76.6
E on *G*	−0.0927 0.1851 −0.9783	95.31, 79.33, 11.95	0.74	74.6
G on *I*	−0.1260 0.1769 −0.9761	97.23, 79.81, 12.55	0.69	67.7
I on *B*	−0.1104 0.1747 −0.9784	96.33, 79.93, 11.93	0.60	70.5
Pentamer II				
A on *D*	−0.1161 −0.1533 0.9813	96.99, 98.81, 11.09	0.76	70.5
D on *F*	−0.1163 −0.1690 0.9787	96.67, 99.72, 11.84	0.67	76.4
F on *H*	−0.1296 −0.1592 0.9786	97.44, 99.16, 11.87	0.59	70.7
H on *J*	−0.1197 −0.1494 0.9815	96.87, 98.59, 11.03	0.55	73.9
A on *J*	−0.1183 −0.1706 0.9782	96.79, 99.82, 11.98	0.51	66.9

two pentamers within the asymmetric unit of space group $P4_2$ results in an r.m.s.d. of 0.7 Å, which is slightly larger than the previously mentioned r.m.s.d. values (Table 5). The relationship between the two pentamers is very close to $\frac{1}{2} + x$, $\frac{1}{2} - y$, $\frac{1}{2} - z$, with the pseudo-rotation relating the two pentamers being offset by a small angle (<0.5°) from 180° and the translations deviating by about 0.1–0.2 Å from $\frac{1}{2}\,\frac{1}{2}\,\frac{1}{2}$.

The rotation matrix relating the two pentamers is

$$\begin{pmatrix} 0.999962 & 0.008415 & -0.002067 \\ 0.008400 & -0.999942 & 0.006797 \\ -0.002124 & -0.006779 & -0.999975 \end{pmatrix}.$$

The translation is 31.85, −31.94, −60.69 Å about a vector with direction cosines 0.99999, 0.00420, −0.00105.

The rotation between corresponding chains of the two pentamers averaged at 179.4° (179.2, 179.8, 179.8, 179.1, 179.1°) and the average r.m.s.d. between the chains was 0.7 Å (Table 6). The mean main-chain r.m.s.d. when each chain of the pentamers in space group $P4_2$ was superposed on the neighbouring chain ranged from 0.5 to 0.8 Å and the rotation angles varied from 66.9 to 76.6°. The rotation angles between successive helices in each of the two pentamers were 69.1, 76.6, 74.6, 67.7, 70.5° and 70.5, 76.4, 70.7, 73.9, 66.9°, respectively

Table 8
Superposition of the *A* chain of pentamer II on the *B* chain of pentamer I in space group $P4_2$.

Chains under study	R.m.s.d. (Å)
A and *B*	0.47
C and *D*	1.11
E and *F*	0.59
G and *H*	1.32
I and *J*	0.73

Table 9
Superposition of *B* chain of the pentamer in space group $P4_22_12$ and the *B* chain of the pentamer in space group $P4_2$.

Pentamer 1, *BCEGI*, space group $P4_2$; pentamer 2, *ADFHJ*, space group $P4_2$; $P4_22_12$ pentamer, $B_1C_1E_1G_1I_1$; symmetry-related $P4_22_12$ pentamer, $B_2C_2E_2G_2I_2$.

Chains under study	R.m.s.d. (Å)
B and B_1	0.24
C and C_1	0.71
E and E_1	0.97
G and G_1	1.44
I and I_1	0.53
A and B_2	1.34
D and C_2	1.51
F and E_2	1.58
H and G_2	1.15
J and I_2	1.87

(Table 7). This clearly shows the differences between the two pentamers in the asymmetric unit. All of the alignments carried out were in the residue range 95–135.

Only one chain of the first pentamer in space group $P4_2$ was superposed on the corresponding chain of the second pentamer in space group $P4_2$ and the r.m.s.d.s of all chains were calculated (Table 8). Similarly, we superposed only one chain of the first pentamer in space group $P4_2$ on its corresponding chain in the pentamer in space group $P4_22_12$ and calculated the r.m.s.d.s between all corresponding chains (including the r.m.s.d. between the second pentamer in space group $P4_2$ and the symmetry-related pentamer in space group $P4_22_12$; Table 9). The individual chains in the two pentamers are not related by an exact fivefold relation and the two pentamers in the asymmetric unit in space group $P4_2$ are also different (Table 7). In space group $P4_22_12$, the single pentamer in the asymmetric unit is most likely to be the average of the two pentamers in space group $P4_2$ (Table 6). The relation between the pentamers is very close to the crystallographic symmetry, but the difference between the two pentamers in the asymmetric unit is more pronounced than the deviation of the pseudo-symmetric axis from the perfect symmetry axis, making this a special case of rotational pseudosymmetry.

The diffraction data (Fig. 2) are a combination of a few very strong reflections and a very large number of weak reflections and thus have an intensity distribution that differs from that of a typical globular protein. The removal of the weakest of the reflections by an ellipsoidal truncation helped in the initial stages of model building as it resulted in better electron-density maps. The addition of missing reflections to improve the completeness of the synchrotron data and the detection

and correction of the data pathologies resulted in successful refinement of the structure. From this study, it is clear that the diffraction data should be recorded in the best possible way and also that all possible factors affecting the intensities of the reflections such as anisotropy, twinning and pseudosymmetry should be considered and evaluated critically if problems with structure solution and/or refinement are encountered.

This work was supported by the Indian Council of Medical Research and the Department of Biotechnology (DBT) Infrastructure Program to the Indian Institute of Science. Preliminary data were collected at the X-ray Facility for Structural Biology in the Molecular Biophysics Unit of the Institute funded by the DBT. We thank J. Jeyakanthan, K. Sekar and G. Ueno for synchrotron data collection and processing. ARC acknowledges the University Grants Commission (UGC), Government of India for the award of a fellowship.

References

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.

Au, K. S., Mattion, N. M. & Estes, M. K. (1993). *Virology*, **194**, 665–673.

Ball, J. M., Tian, P., Zeng, C. Q.-Y., Morris, A. P. & Estes, M. K. (1996). *Science*, **272**, 101–104.

Boshuizen, J. A., Rossen, J. W., Sitaram, C. K., Kimenai, F. F., Simons-Oosterhuis, Y., Laffeber, C., Büller, H. A. & Einerhand, A. W. (2004). *J. Virol.* **78**, 10045–10053.

Bowman, G. D., Nodelman, I. M., Levy, O., Lin, S. L., Tian, P., Zamb, T. J., Udem, S. A., Venkataraghavan, B. & Schutt, C. E. (2000). *J. Mol. Biol.* **304**, 861–871.

Brooks, C. L., Blackler, R. J., Gerstenbruch, S., Kosma, P., Müller-Loennies, S., Brade, H. & Evans, S. V. (2008). *Acta Cryst.* D**64**, 1250–1258.

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.

Chacko, A. R., Arifullah, M., Sastri, N. P., Jeyakanthan, J., Ueno, G., Sekar, K., Read, R. J., Dodson, E. J., Rao, D. C. & Suguna, K. (2011). *J. Virol.* **85**, 12721–12732.

Chacko, A. R., Jeyakanthan, J., Ueno, G., Sekar, K., Rao, C. D., Dodson, E. J., Suguna, K. & Read, R. J. (2012). *Acta Cryst.* D**68**, 57–61.

Chan, W. K., Au, K. S. & Estes, M. K. (1988). *Virology*, **164**, 435–442.

Cohen, G. H. (1997). *J. Appl. Cryst.* **30**, 1160–1161.

Dauter, Z., Botos, I., LaRonde-LeBlanc, N. & Wlodawer, A. (2005). *Acta Cryst.* D**61**, 967–975.

Deepa, R., Durga Rao, C. & Suguna, K. (2007). *Arch. Virol.* **152**, 847–859.

Dong, Y., Zeng, C. Q.-Y., Ball, J. M., Estes, M. K. & Morris, A. P. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 3960–3965.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Estes, M. K., Kang, G., Zeng, C. Q.-Y., Crawford, S. E. & Ciarlet, M. (2001). *Novartis Found. Symp.* **238**, 82–96.

Fisher, R. G. & Sweet, R. M. (1980). *Acta Cryst.* A**36**, 755–760.

French, S. & Wilson, K. (1978). *Acta Cryst.* A**34**, 517–525.

Guelker, M., Stagg, L., Wittung-Stafshede, P. & Shamoo, Y. (2009). *Acta Cryst.* D**65**, 523–534.

Horie, Y., Nakagomi, O., Koshimura, Y., Nakagomi, T., Suzuki, Y., Oka, T., Sasaki, S., Matsuda, Y. & Watanabe, S. (1999). *Virology*, **262**, 398–407.

Jakobsson, E., Nilsson, J., Källström, U., Ogg, D. & Kleywegt, G. J. (2005). *Acta Cryst.* F**61**, 274–278.

Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.

Lebedev, A. A., Vagin, A. A. & Murshudov, G. N. (2006). *Acta Cryst.* D**62**, 83–95.

Lee, S., Sawaya, M. R. & Eisenberg, D. (2003). *Acta Cryst.* D**59**, 2191–2199.

Lodowski, D. T., Pitcher, J. A., Capel, W. D., Lefkowitz, R. J. & Tesmer, J. J. (2003). *Science*, **300**, 1256–1262.

Mahalanobis, P. C. (1936). *Proc. Natl Acad. Sci. USA*, **1**, 49–55.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

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.

Mir, K. D., Parr, R. D., Schroeder, F. & Ball, J. M. (2007). *Virus Res.* **126**, 106–115.

Mirazimi, A., Nilsson, M. & Svensson, L. (1998). *J. Virol.* **72**, 8705–8709.

Newton, K., Meyer, J. C., Bellamy, A. R. & Taylor, J. A. (1997). *J. Virol.* **71**, 9458–9465.

O'Brien, G. J., Bryant, C. J., Voogd, C., Greenberg, H. B., Gardner, R. C. & Bellamy, A. R. (2000). *Virology*, **270**, 444–453.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* D**59**, 1124–1130.

Parr, R. D., Storey, S. M., Mitchell, D. M., McIntosh, A. L., Zhou, M., Mir, K. D. & Ball, J. M. (2006). *J. Virol.* **80**, 2842–2854.

Ramadan, M. A. M., Shrive, A. K., Holden, D., Myles, D. A. A., Volanakis, J. E., DeLucas, L. & Greenhough, T. J. (2002). *Acta Cryst.* D**58**, 992–1001.

Seo, N.-S., Zeng, C. Q.-Y., Hyser, J. M., Utama, B., Crawford, S. E., Kim, K. J., Höök, M. & Estes, M. K. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 8811–8818.

Sheriff, S., Hendrickson, W. A. & Smith, J. L. (1987). *J. Mol. Biol.* **197**, 273–296.

Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D. & Eisenberg, D. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 8060–8065.

Taylor, J. A., O'Brien, J. A., Lord, V. J., Meyer, J. C. & Bellamy, A. R. (1993). *Virology*, **194**, 807–814.

Tian, P., Ball, J. M., Zeng, C. Q.-Y. & Estes, M. K. (1996). *J. Virol.* **70**, 6973–6981.

Tian, P., Hu, Y., Schilling, W. P., Lindsay, D. A., Eiden, J. & Estes, M. K. (1994). *J. Virol.* **68**, 251–257.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Xu, A., Bellamy, A. R. & Taylor, J. A. (2000). *EMBO J.* **19**, 6465–6474.

Yeates, T. O. (1988). *Acta Cryst.* A**44**, 142–144.

Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.

Yu, F., Song, A., Xu, C., Sun, L., Li, J., Tang, L., Yu, M., Yeates, T. O., Hu, H. & He, J. (2009). *Acta Cryst.* D**65**, 212–219.

Zagotta, W. N., Olivier, N. B., Black, K. D., Young, E. C., Olson, R. & Gouaux, J. E. (2003). *Nature (London)*, **425**, 200–205.

Zhang, M., Monzingo, A. F., Segatori, L., Georgiou, G. & Robertus, J. D. (2004). *Acta Cryst.* D**60**, 1512–1518.

Zwart, P. H., Grosse-Kunsteleve, R. W. & Adams, P. D. (2005). *CCP4 Newsl.* **42**, contribution 10.

Zwart, P. H., Grosse-Kunstleve, R. W., Lebedev, A. A., Murshudov, G. N. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 99–107.