Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Combination of methods used in the structure solution of pyruvate:ferredoxin oxidoreductase from two crystal forms

E. Chabrière, A. Volbeda,*† J. C. Fontecilla-Camps, M. Roth and M.-H. Charon†

Laboratoire de Cristallographie et de Cristallogenèse des Protéines, Institut de Biologie Structurale J.-P. Ebel CEA-CNRS, 41 avenue Jules Horowitz, 38027 Grenoble CEDEX 01, France

† Both these authors contributed equally to this project.

Correspondence e-mail: volbeda@lccp.ibs.fr

Received 12 April 1999
Accepted 22 June 1999

The structure of the homodimeric 267 kDa pyruvate:ferredoxin oxidoreductase (PFOR) of *Desulfovibrio africanus* was solved with data from two crystals forms, both containing two monomers per asymmetric unit. Phases were obtained from multiwavelength anomalous dispersion (MAD), solvent flattening (SF), molecular replacement (MR) using a 5 Å resolution electron-density search model, multiple isomorphous replacement (MIR) and, finally, electron-density averaging (DA) procedures. It is shown how the combination of all these techniques was used to overcome problems arising from incompleteness of MAD data and weak phasing power of MIR data. A real-space refinement (RSR) procedure is described to improve MR solutions and obtain very accurate protein envelopes and non-crystallographic symmetry (NCS) transformations from 5 Å resolution phase information. These were crucial for the phase extension to high resolution by DA methods.

1. Abbreviations

DA, density averaging; MAD, multiwavelength anomalous dispersion; MIR, multiple isomorphous replacement; NCS, non-crystallographic symmetry; PFOR, pyruvate:ferredoxin oxidoreductase; RSR, real-space refinement; SF, solvent flattening.

2. Introduction

In spite of great progress made during the last decade, solving large protein structures is still not a trivial problem. One reason is that, when the classical MIR method is used, the relative heavy-atom contribution to the X-ray diffraction decreases as a function of the protein weight (Blundell & Johnson, 1976). Hence, for large proteins, the location of heavy-atom sites in derivatives is often difficult. Furthermore, the signal provided by the heavy atoms is perturbed if crystals are not isomorphous. Unit-cell changes are quite frequent when X-ray data collection is performed at low temperature after flash-freezing of the protein crystal, which is often a prerequisite for high-resolution work. For a protein with anomalous scatterers, the latter problem can be overcome by using the tuneable wavelength of a synchrotron source. Once anomalous scatterers are located from MAD data, phases can be obtained. However, if atoms like iron are used for MAD phasing, there is significant absorption at the used wavelengths, and this complicates precise phase information at high resolution. Fortunately, in many cases, the existence of NCS or other crystal forms allows for further improvement of the quality of experimental phases and their extension to higher resolution by DA methods.

Table 1
Statistics of used MAD and native data sets.

Values in parentheses refer to the highest resolution shell. The native 2 data are more than 50% complete up to 2.5 Å resolution. MAD data are ∼64% complete up to 3.6 Å resolution.

Data set	MAD2A	MAD2B	MAD2C	MAD2D	Native 1	Native 2
Crystal form	2	2	2	2	1	2
X-ray wavelength (Å)	1.772	1.744	1.739	1.709	1.04	1.074
Resolution (Å)	2.98 (3.06–2.98)	2.98 (3.06–2.98)	2.98 (3.06–2.98)	2.98 (3.06–2.98)	2.85 (2.92–2.85)	2.31 (2.49–2.31)
No. of unique reflections	28819	29471	29721	30240	61632	75568
Data redundancy	3.0	3.1	3.1	3.1	3.9	3.2
Completion (%)	55.6 (28.7)	56.9 (35.6)	57.3 (36.6)	58.3 (40.7)	97.2 (78.4)	68.1 (28.2)
$\langle I/\sigma \rangle$	6.8	7.8	6.8	7.2	15.3	7.4
$R_{\rm sym}$ (%)†	8.1	7.2	8.1	7.9	4.0	8.2
$R_{\rm anom}$ (%)‡	5.3	4.9	5.8	6.0		
$R_{\rm scale}$ (%)§	5.2		4.2	6.1		

† $R_{\rm sym} = \sum_{hkl} \sum_j |I_{hkl,j} - \langle I_{hkl} \rangle| / \sum_{hkl,j} \langle I_{kl} \rangle$, using Friedel mates as independent observations. ‡ $R_{\rm anom} = \sum_{hkl} |I_{hkl}{}^+ - \langle I_{hkl} \rangle^-| / \sum_{hkl} \langle I_{hkl} \rangle$. § $R_{\rm scale} = \sum_{hkl} |F_i - F_{\rm MAD2B}| / \sum_{hkl} \langle F_{\rm MAD2B} \rangle$.

In this paper we discuss how the structure of pyruvate: ferredoxin oxidoreductase (PFOR) of *Desulfovibrio africanus* was solved (Chabrière, Charon, Volbeda *et al.*, 1999). PFORs occur in most anaerobes where they catalyze the oxidative decarboxylation of pyruvate to form acetyl-coenzyme A. The *D. africanus* enzyme is active as a homodimer of 267 kDa, containing a total of six [4Fe–4S] clusters (Pieulle *et al.*, 1997). The polypeptide chain contains 1 231 residues. The crystallization protocol used (Pieulle *et al.*, 1999) gives two orthorhombic crystal forms (space group $P2_12_12_1$) with cell dimensions $a = 86.0$, $b = 146.3$, $c = 211.9$ Å for form 1 and $a = 84.8$, $b = 144.9$, $c = 203.0$ Å for form 2. Each form contains one PFOR dimer per asymmetric unit. For solving the structure a combination of standard crystallogaphic methods was used: MAD, MIR, SF, MR and DA. Specific methods were developed to overcome problems caused by the incompleteness of the MAD data, and to refine protein masks and NCS transformations. We hope that the strategy employed by us will be useful to other crystallographic structure determinations.

3. Methods and results

3.1. Strategy

All data sets (Table 1, see also Chabrière, Charon, Volbeda *et al.*, 1999) were recorded at the BM02 beamline of the European Synchrotron Radiation Facility and further processed with the program *XDS* (Kabsch, 1993) and the *CCP*4 program package (Collaborative Computational Project, Number 4, 1994). Because it was not possible to locate heavy-atom sites in derivatives of crystal form 1 by Patterson map interpretation, it was decided to try to take advantage of the presence of 24 Fe atoms in the PFOR dimer by performing a MAD experiment on crystal form 2. The strategy that was followed is summarized in Fig. 1.

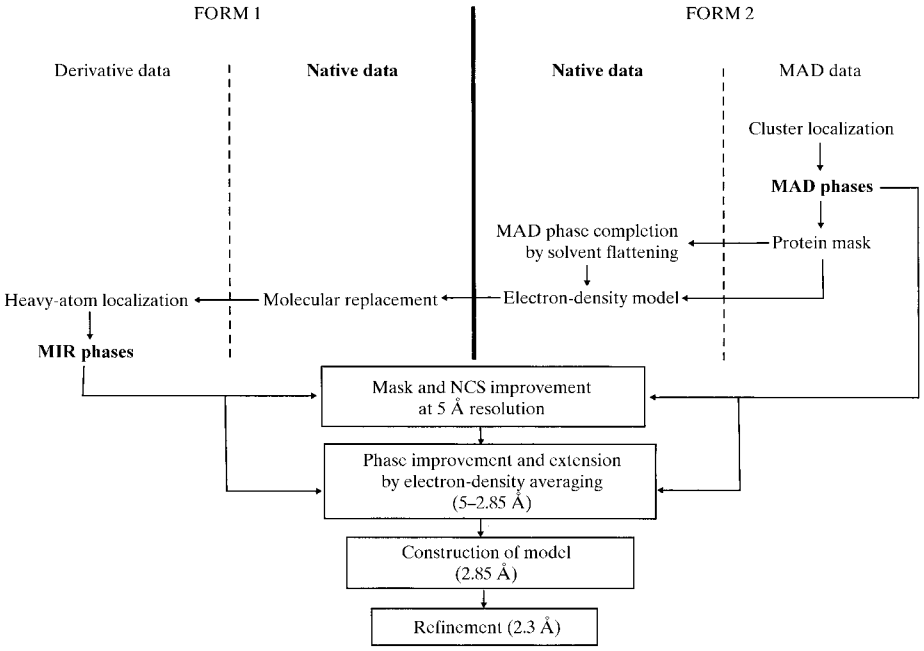

Figure 1
Strategy for the PFOR structure determination.

The first step was the location of the six [4Fe–4S] clusters for MAD phasing of crystal form 2. Next, an approximate protein mask was constructed which served to improve and complete the obtained phases by SF. Subsequently, the MAD phases were transformed to crystal form 1, using the method of MR. This allowed us to locate the heavy-atom sites in four derivatives of this form for MIR phasing. Next, by using the independent data and experimental phases coming from the two crystal forms, the protein masks and all internal and cross NCS transformations were optimized. All these steps were carried out at 5 Å resolution. Finally, phases were extended to 2.85 Å resolution by intra- and inter-crystal DA. This allowed the construction of an atomic model which

Table 2
MR solutions for crystal form 1.

MR solutions are expressed in polar rotation angles κ, φ and ψ (Rossmann & Blow, 1962) and translation components along x, y and z. Electron-density models $ed2_{MAD}$ and $ed2_{MAD\_SF}$ are explained in the text (§3.4). All solutions were found with the program *AMoRe* (Navaza, 1994).

Model	Function	d (Å)†	κ (°)	φ (°)	ψ (°)	T_x (Å)	T_y (Å)	T_z (Å)	c.c.f. (%)‡
None	Self rotation	15.0–2.85	180.0	87.1	68.2	—	—	—	21.3 (7.8)
$ed2_{MAD}$	Cross rotation	8.0–5.0	49.2	181.1	83.6	—	—	—	15.9 (9.8)
	Cross rotation	8.0–5.0	5.5	195.2	102.6	—	—	—	14.3 (9.8)
$ed2_{MAD\_SF}$	Cross rotation	8.0–5.0	48.6	180.2	83.2	—	—	—	25.1 (14.1)
	Cross rotation	8.0–5.0	5.0	193.8	91.9	—	—	—	24.7 (14.1)
	Translation	8.0–5.0	5.0	193.8	91.9	−0.95	−1.59	−3.56	30.0 (19.1)
	Fiting	8.0–5.0	5.4	183.2	94.2	−0.86	−1.62	−3.41	33.5

† Resolution range of used data. ‡ Correlation coefficient, with highest noise peak in parentheses.

was refined with data to a maximal resolution of 2.3 Å. Details of the different steps are described below.

3.2. MAD phasing

An X-ray fluorescence wavelength scan of a PFOR form 2 crystal recorded around the iron absorption edge showed the presence of a significant anomalous signal (Fig. 2). Four wavelengths were chosen for data collection and MAD data sets were collected up to 2.9 Å resolution (Table 1). Because of limited available beam time and a suboptimal data-collection strategy, the MAD data were only ∼64% complete up to 3.6 Å resolution (Figs. 3*a*−3*c*). A later attempt to complete the missing data with another MAD experiment was not successful because of insufficient crystal quality. Therefore, it was decided to continue the structure determination with the incomplete MAD data. Solution of Hendrickson's equations (Hendrickson, 1991) with the program *nYn* (M. Roth, unpublished work) provided the F_A values, which correspond to the wavelength-independent part of the structure factors of the 24 anomalously scattering Fe atoms.

In order to find the position of the anomalous scatterers the Patterson method was used. Because a number of F_A data contained significant errors, it was crucial to select the most reliable ones. At low resolution, the Fe atoms of each [4Fe–4S] cluster coalesce into a single atom, and thus six independent peaks (one per cluster) were expected in each Harker section. Visual inspection of Harker sections of the F_A^2 Patterson was used for choosing resolution, the maximum amplitude of F_A and the maximum value for the standard deviation associated to each F_A. The maximum amplitude was the most important criterion: several strong F_A outliers which introduced a severe bias in the Patterson map had to be removed. In spite of the presence of ghost peaks and shape distortions arising from data incompleteness (Samudzi & Rosenberg, 1992), the Patterson map of the selected F_A coefficients, corresponding to 92% of the collected data and ∼60% of all possible data up to 5 Å resolution, could be solved with the program *SHELXS* (Sheldrick, 1997). A clear solution was obtained, constituted by two groups of three sites.

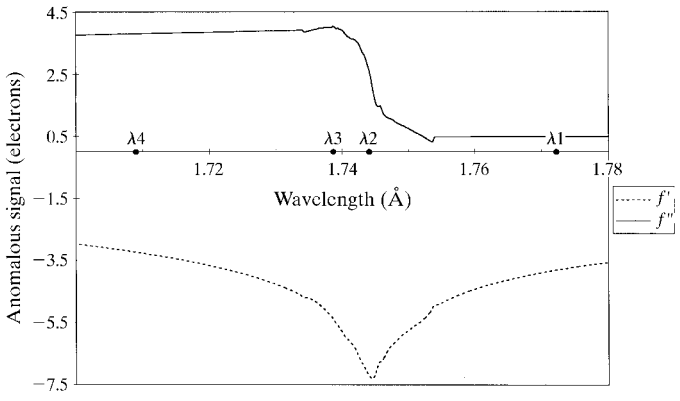

Figure 2
Anomalous scattering factors in the Fe absorption edge region for a PFOR form 2 crystal, with f' computed from f'' with the Kramers–Kronig relations. The indicated wavelengths were selected for MAD data collection (see Table 1).

Next, the MAD data were treated as a special case of MIR (Ramakrishnan & Biou, 1997). The MAD2B data set recorded at the λ2 wavelength (1.744 Å, see Fig. 2), where f' is the lowest, was used as native. The [4Fe–4S] clusters were modelled as single atoms, with both f' and f'' set equal to one electron, and with a fixed temperature factor of 80 Å^2 in order to take partially account for the cluster size. Occupancies (real and anomalous) and positions were refined with the program *MLPHARE* (Otwinowski, 1991). The refinement was carried out at 5 Å resolution, where the one-atom approximation for the [4Fe–4S] clusters is reasonable. At the end an incomplete MAD phase set was obtained with an overall figure of merit of 0.60 between 20.0 and 5.0 Å resolution, and 0.27 in the 20–3.0 Å resolution range (Fig. 4). The absolute configuration of the cluster coordinates was easily found, as the correct enantiomorph gave a significantly better contrast between solvent and protein in the 5 Å resolution MAD-phased map than the incorrect one (Fig. 5). The lack of data and phase completeness resulted in a deformed electron-density map. This effect is clearly seen in the anomalous difference Fourier (Fig 3*d*): peaks are elongated along the z axis, which corresponds to the direction in reciprocal space were most of the data are missing. Nevertheless, the MAD map obtained provided a good starting point for the determination of the protein envelope in the next step.

3.3. Protein mask and NCS

Taking into account that up to 5 Å resolution the phasing power was greater than 1.0 and that the signal-to-noise ratio of the cluster peaks remained high in the electron-density map, we concluded that the MAD phases were reliable up to this resolution. For further phase improvement we decided to use the non-crystallographic symmetry, which required the determination of an accurate molecular mask and NCS operator. These are respectively defined here as $NCS2_{int}$ and mask2 in crystal form 2, where the subscript 'int' specifies an internal transformation. The self-rotation function suggested the presence of local twofold symmetry axes in both crystal forms. Their orientations and positions were refined as described below.

A starting $NCS2_{int}$ operator was obtained by superimposing the three clusters of one monomer over those of the other. Next, $NCS2_{int}$ was improved with density-correlation methods incorporated in the program *IMP* from the *RAVE* package (Kleywegt & Jones, 1994), initially selecting only electron density within 5 Å of each of the clusters. Next, in order to obtain the molecular mask, the program *ENVNCS* (A. Volbeda, unpublished work) was used. This program uses the local correlation between the experimental and the NCS-transformed map. and multiplies this with the blurred experimental map in order to discriminate between protein and solvent regions (Wang, 1985). The result is a 'probability map' for the protein region. The mask is built to the desired size, starting from a high probability level, by adding shells of grid points with a value above a gradually decreasing level, and taking account of the crystallographic packing in order to prevent overlap with neighbouring molecules.

Because the NCS operator and mask depend on each other, their joint refinement is tricky. In our case, the difficulty was increased by the lack of completeness, which caused bias. For these reasons, mask2 and $NCS2_{int}$ were improved iteratively using, respectively, *ENVNCS* and *IMP* on the 5 Å resolution MAD map. During the refinement procedure $NCS2_{int}$ turned out to correspond to a proper twofold symmetry axis, as already predicted by the self-rotation function. Therefore, only one mask covering the complete PFOR homodimer needed to be determined. Refinement of $NCS2_{int}$ was obtained by maximizing the density correlation between the PFOR dimer and its NCS-transformation within the gradually growing mask. The V_m (Matthews, 1968) of the asymmetric

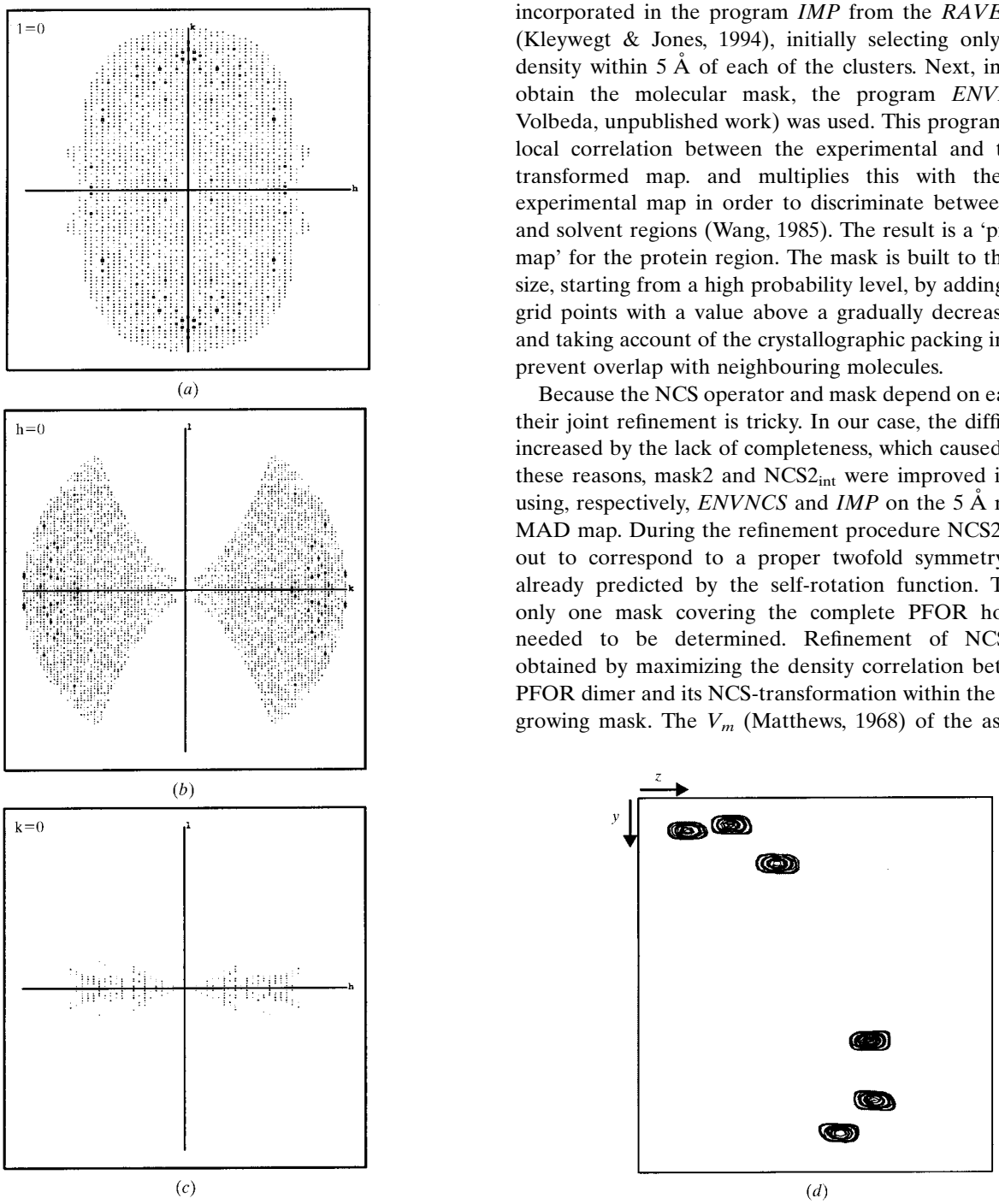

Figure 3
Analysis of the incompleteness of the MAD data with pseudo-precession images of observed diffraction intensities (no σ cutoff used), calculated with the program *HKLVIEW* (Collaborative Computational Project, Number 4, 1994); (*a*) the *hk*0 plane, (*b*) the 0*kl* plane and (*c*) the *h*0*l* plane; (*d*) effect of data incompleteness on the shape of the peaks in a close-up of a projection of the 20.0–5.0 Å resolution MAD-phased anomalous difference map.

unit of crystal form 2 is 2.34 Å^3 Da^{-1}, which corresponds to a protein content of 53%. In order not to remove parts of the protein and keeping in mind the severe bias in the map arising from data incompleteness, a large error margin was added and the volume of the final mask was chosen at 74%(*v*/*v*) of the asymmetric unit. Using this mask, a real-space correlation of 0.218 between the MAD map and its NCS-transformed duplicate was obtained.

3.4. MR phasing

The next problem to be solved was the transformation of the MAD phases from crystal form 2 to form 1 in order to use the data of this form and to locate the heavy-atom sites in its derivatives. A model called ed2$_{MAD}$ was calculated from the 64% complete 5 Å resolution MAD phases by applying mask 2 to the MAD map, thereby isolating the electron density of one PFOR dimer. When this was used as a search model for MR with the program *AMoRe* (Navaza, 1994), two relatively low signal-to-noise peaks for the rotation function were obtained (Table 2). However, the translation function did not provide any convincing solution, presumably owing to the map deformation arising from data incompleteness. Therefore, using a more complete data set obtained from another form 2 crystal (native 2 in Table 1), a phase-completion step was performed by SF at 5 Å resolution with the program package *DEMON/ANGEL* (Vellieux *et al.*, 1995), using the already described mask2 as the protein envelope. No DA was used at this stage in order to prevent bias arising from an imprecise NCS2$_{int}$. The electron-density model called ed2$_{MAD_SF}$ that was generated with the completed amplitude and phase set gave two well contrasted rotation-function solutions, with which it was easy to solve the translation function (Table 2). A change of origin was introduced in order to select the NCS transformation closest to the identity operation. The final refined solution is defined as NCS$_{ext}$, where the subscript 'ext' is used in order to indicated an external transformation, *i.e.* to the other crystal form. In this way a 5 Å resolution electron-density model for crystal form 1 was obtained with an *R* factor of 0.50 between the observed structure factors and those calculated by Fourier inversion of the transformed MAD map.

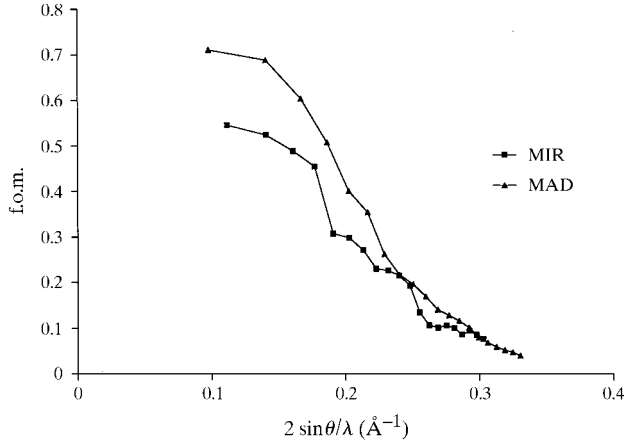

Figure 4
Resolution dependence of the figure of merit (f.o.m.) of the experimental phases obtained for crystals forms 1 (MIR) and 2 (MAD).

3.5. MIR phasing

With the MR phases obtained from the transformed electron-density map of crystal form 2, and using difference Fourier methods, it was possible to locate 30 heavy-atom sites, including several common ones, distributed over four derivatives of crystal form 1 (Chabrière, Charon, Volbeda *et al.*, 1999). The heavy-atom parameters were refined with *MLPHARE* (Otwinowski, 1991), using the native 1 data set as reference. Using a procedure for phase-bias reduction called PBR (Chabrière, Charon & Vellieux, 1999), developed for improving phases when two derivatives contain common sites, phases were obtained with an overall figure of merit of 0.28 between 20 and 3.3 Å resolution (Fig. 4). The MIR

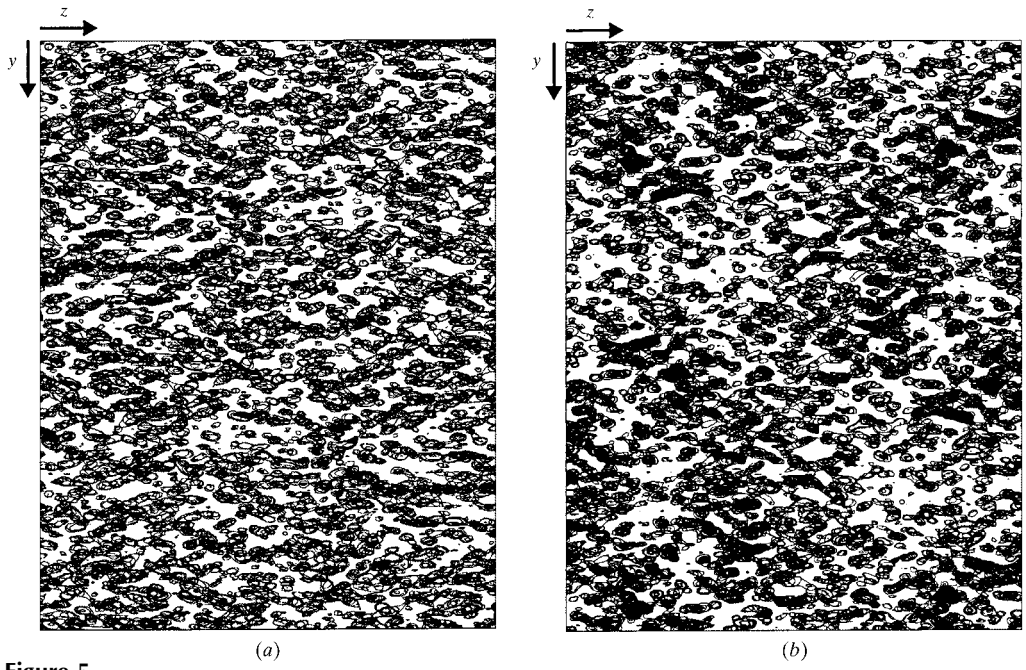

Figure 5
MAD-phased electron-density maps of crystal form 2, calculated with data between 20.0 and 5 Å resolution: (*a*) with the incorrect absolute configuration of the cluster locations, and (*b*) with the correct enantiomorph. The complete unit cell is shown, using a contour level of 2σ.

phase set was instrumental in improving masks and NCS operators in the next step.

3.6. Mask and NCS refinement

Initial attempts to improve the experimental phases by DA led to unacceptably high *R* factors between observed and calculated structure factors, the latter ones being obtained by Fourier back-transformation of averaged electron-density maps. Consequently, the masks and NCS operators, which were first defined as described above in §3.3, were improved in the following iterative procedure. In step 1, the experimental density map of each crystal form was averaged using the internal NCS operators and masks. The averaged maps were used to refine NCS_{ext} (step 2), using map correlation techniques incorporated in the program *SUPERMAP* (A. Volbeda, unpublished work). Next, MAD phases could be transformed from crystal form 2 to form 1, and MIR phases from form 1 to form 2, using NCS_{ext} (or its inverse) and the respective masks, by transformation of the unaveraged map (step 3) followed by Fourier inversion (step 4). After application of Sim weights (Sim, 1960) the transformed phases were combined with the experimental ones (MIR and MAD) in step 5. The electron-density maps called $ed1_{comb}$ and $ed2_{comb}$ that were computed for each crystal form from the combined phases (step 6) were of much better quality and allowed to re-define the internal NCS transformations and masks (step 7), following the methods described in §3. Next, the whole sequence of step 1 to step 7 was repeated until convergence was obtained, as indicated by the absence of further shifts.

Each step of the procedure was used in parallel for both crystal forms. As the internal NCS transformations are independent from the external ones, NCS_{int} was used as additional information to refine NCS_{ext}, and *vice versa*. In form 1 the same native 1 structure-factor amplitudes were used for all map calculations, but in form 2 the incomplete reference MAD2B data were used for the experimental MAD map and the much more complete native 2 data set for Sim weighting and the calculation of the phase-combined map. In this way the bias in the mask and NCS transformation due to data incompleteness was effectively minimized. Phases calculated from averaged maps were not used by any point in the refinement procedure, in order to prevent the development of bias in the refined NCS transformations and masks.

3.7. Phase extension, model building and refinement

The two experimentally phased electron-density maps obtained for each crystal form were used as the starting point for phase improvement and extension to high resolution by DA (Fig. 6) with *DEMON/ANGEL* (Vellieux *et al.*, 1995). Phases were first improved and completed at 5 Å resolution. Twofold DA and SF within each crystal form, using the respective protein envelopes and NCS_{int}, was followed by inter-crystal DA, using NCS_{ext}, Fourier inversion of the fourfold averaged maps and phase combination between Sim-weighted calculated map phases and available experimental (MIR or MAD) phases. This was finally followed by the calculation of an improved phase-combined map, which was subsequently subjected to the same averaging procedure. This sequence of steps was repeated ten times, after which the resolution was extended by approximately one reciprocal lattice point (1/210 Å), which is within the width of the *G* function describing the existence of phase correlations between nearby reflections in reciprocal space (Rossmann & Blow, 1962; Vellieux & Read, 1997). Next, ten iterations were performed at the extended resolution, but now experimental phases were combined only in the added resolution shell, and NCS was used as the only phasing source for the rest of the reflections, including those without MAD or MIR phases. In this way phases were gradually extended in 32 steps, each consisting of ten iterations, to 2.85 Å, the resolution limit of the native data set of crystal form 1. The final *R* factor for the calculated

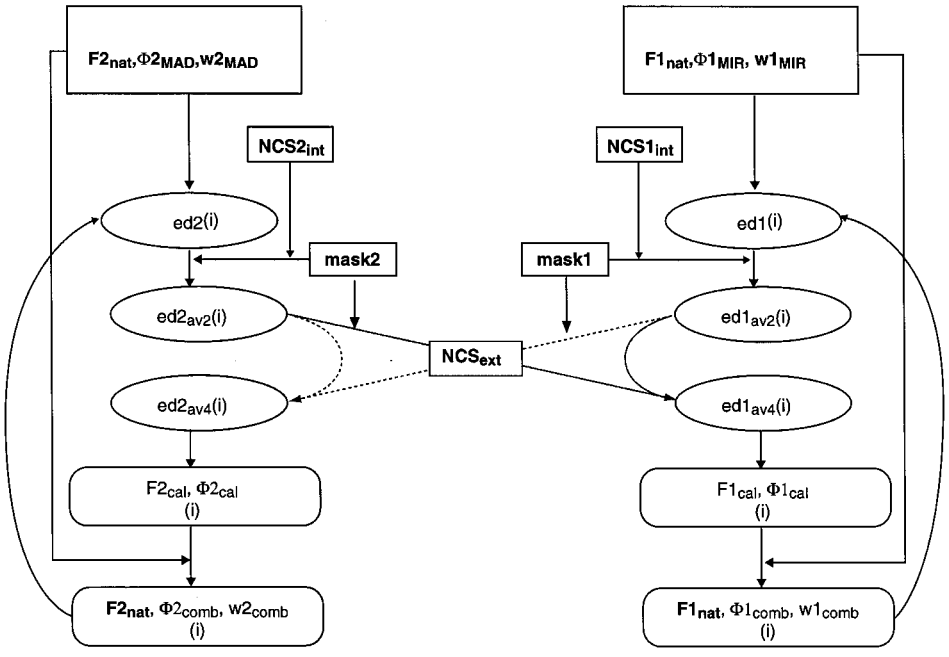

Figure 6
Procedure of phase improvement and extension by DA. Symbols are: F = amplitude, Φ = phase, *w* = figure of merit or Sim weight, ed = electron density, 1 or 2 = crystal form, av2 and av4 = averaged over 2 and 4 PFOR monomers, cal = calculated from Fourier inversion of fourfold averaged maps, comb = phase-combined, (*i*) = iteration number. Bold characters indicate non-changing data. Further explanations are given in the text.

Table 3
Refinement of internal and external NCS-transformations for the PFOR dimer in crystal form 2.

A comparison of the results obtained from MR, RSR and the final atomic models is given. Electron-density (ed) models are explained in the text (§§3.4 and 3.6). Transformations are defined as in Table 2. The final column contains root-mean-square deviations (Δr.m.s.) calculated between NCS-transformed monomers (for internal transformations) or dimers (external transformations) of atomic models of form 2, using the refined atomic model subjected to the respective final transformations as the reference.

Transformation	Origin:	κ (°)	φ (°)	ψ (°)	T_x (Å)	T_y (Å)	T_z (Å)	Δr.m.s. (Å)
Internal	Self-rotation function	180.0	87.3	62.7	—	—	—	–
	$ed2_{MAD}$ (RSR)	180.0	87.3	63.4	−39.7	−69.5	−37.0	0.87
	$ed2_{comb}$ (RSR)	180.0	87.0	63.0	−39.1	−68.4	−37.0	0.30
Final NSC_{int}	Atomic model	180.0	86.9	62.9	−39.0	−68.1	−37.0	0.27
External	$ed2_{MAD\_SF}$ (MR)	5.4	183.2	94.2	−0.86	−1.62	−3.41	0.78
	$ed1_{comb}$, $ed2_{comb}$ (RSR)	5.8	186.4	95.2	−0.78	−1.36	−3.42	0.26
Final NCS_{ext}	Atomic models	6.0	189.9	94.8	−1.12	−1.39	−3.58	0.00

structure factors obtained from Fourier inversion of fourfold averaged maps was 21.8% for crystal form 1, using 60 198 observations, and 19.4% for crystal form 2, using 50 020 observations between 20.0 and 2.85 Å resolution.

No updates of protein masks or NCS transformations were performed during the phase-extension procedure, as this was found to be unnecessary. Most of the polypeptide chain could be easily traced in the final averaged 2.85 Å resolution electron-density map, using the program *O* (Jones *et al.*, 1991). Part of the remaining chain tracing problems could be solved after phases from a partially refined model were combined with the ones obtained from the DA procedure. In addition, secondary-structure predictions were helpful for the modelling of some regions. At the end of the refinement, for which crystal form 2 was used, an almost complete model was obtained containing 19 411 atoms (see PDB entry 1B0P) with good geometry and an R factor of 19.9% (for 69 688 reflections) and an R_{free} of 27.7% between 6.0 and 2.3 Å resolution (Chabrière, Charon, Volbeda *et al.*, 1999).

4. Discussion

One of the largest proteins of which the structure was solved with a combination of DA and MAD phasing from Fe atoms in FeS clusters is *B. subtilis* glutamine 5-phosphoribosyl-1-pyrophosphate amidotransferase, which is a homotetramer of 200 kDa containing one [4Fe–4S] cluster per subunit (Smith *et al.*, 1994). Based on this success, it appeared to be feasible to obtain phase information from a similar approach for *D. africanus* PFOR, which is a 267 kDa homodimer with a total of six [4Fe–4S] clusters. As the phasing power that can be obtained from twofold DA is relatively low, a strategy was developed for making use of the two crystal forms obtained, which are slightly different and each contain one PFOR dimer in the asymmetric unit. It was considered that this should permit the use of fourfold DA, which is much more efficient. The first MAD phases after MR allowed us to obtain supplementary phase information from MIR data of the other crystal form.

Owing to the significant incompleteness of the MAD data, a number of problems were encountered that needed to be solved by the development of specific methods. The 5 Å resolution MAD-phased electron-density map was severely biased because about one third of the MAD data were missing. A reasonable looking protein mask could be generated by combining the density correlation between the two PFOR monomers with solvent contrast information, but the electron density within this mask was not good enough to solve the other crystal form by MR. As it is well known that the success of the MR method is quite dependent on both model accuracy and data completeness, we decided to try to complete the MAD phases by SF, using the independent but much more complete native 2 data set. As only 26% of the unit cell was attributed to solvent, not much phasing power could be obtained from SF. Nevertheless the electron density within the mask improved sufficiently to obtain two clear MR solutions for crystal form 1. We conclude that, even though the new phases were quite poor, the introduction of the extra observed structure factors was instrumental for finding these solutions. Our results may be compared with those of Vellieux *et al.* (1993), who were able to solve the structure of *T. brucei* glyceraldehyde-3-phosphate dehydrogenase with MR by applying amplitude and phase extension to only 37% complete Laue data through sixfold DA.

The next problem requiring a careful strategy was the quality of protein masks and NCS transformations, which need to be accurate for the successful application of DA methods. Using the MAD and MIR phases from the two crystal forms a mask and NCS real-space refinement (RSR) procedure was developed that obeyed the rule that at all stages the refined parameters were practically independent from the used data. The quality of the obtained transformations may be quantified by the structural superposition of different NCS-transformed atomic models to the final refined model, using the operations obtained before and after the RSR procedure (Table 3). The unrefined transformations contained significant errors, which explains why initial attempts of DA did not give good results. The NCS transformations obtained after RSR at 5 Å resolution were quite accurate, since they gave r.m.s. deviations similar both to the mean coordinate error of the final refined model and to the r.m.s. deviation obtained from a superposition of its two subunits (Table 3). Therefore, no updates were necessary during phase extension to 2.85 Å resolution. The protein mask obtained from the RSR procedure was

smoother and more continuous than the one obtained from the incomplete MAD data. The mask quality is further illustrated in Fig. 7, which shows very few surface loops of the refined protein model protruding from it.

The quality of the MIR and MAD phases is analyzed in Fig. 8 by comparing them to phases calculated from the final refined atomic models of the two crystal forms. The MIR phases are less accurate, but they have the advantage of forming an almost complete set at 5 Å resolution. They provided a very useful extra source of phase information for refining the NCS transformations and protein masks, as discussed above. It is also evident that beyond about 5 Å resolution the quality of the incomplete MAD phases drops rapidly. This is partially a result of the modelling of the clusters as single atoms, which becomes a progressively poorer approximation at higher resolutions. In addition the refined cluster positions were significantly inaccurate because of the incompleteness of the MAD data, especially in the z direction. Both the MIR and MAD phases show hardly any correlation to the model phases at resolutions higher than 3.9 Å. This was not an insurmountable problem, however, because the presence of NCS provided another independent phasing source. Most phasing power was obtained from fourfold DA, as shown by the resulting phase improvement after extension to high resolution (Fig. 8). An important factor for the success obtained is the fact that the two crystal forms are almost structurally independent, in spite of the relatively small differences in cell dimensions and crystal packing. Only a weak correlation of 0.114 was found in the 8.0–2.85 Å resolution range between the diffraction data of the two forms, which is possibly due to the contribution of the similar solvent regions. No significant correlation was found between the phase sets of the refined models, which differ by a rotation of 6.0° and a translation parallel to the rotation axis of 0.61 Å.

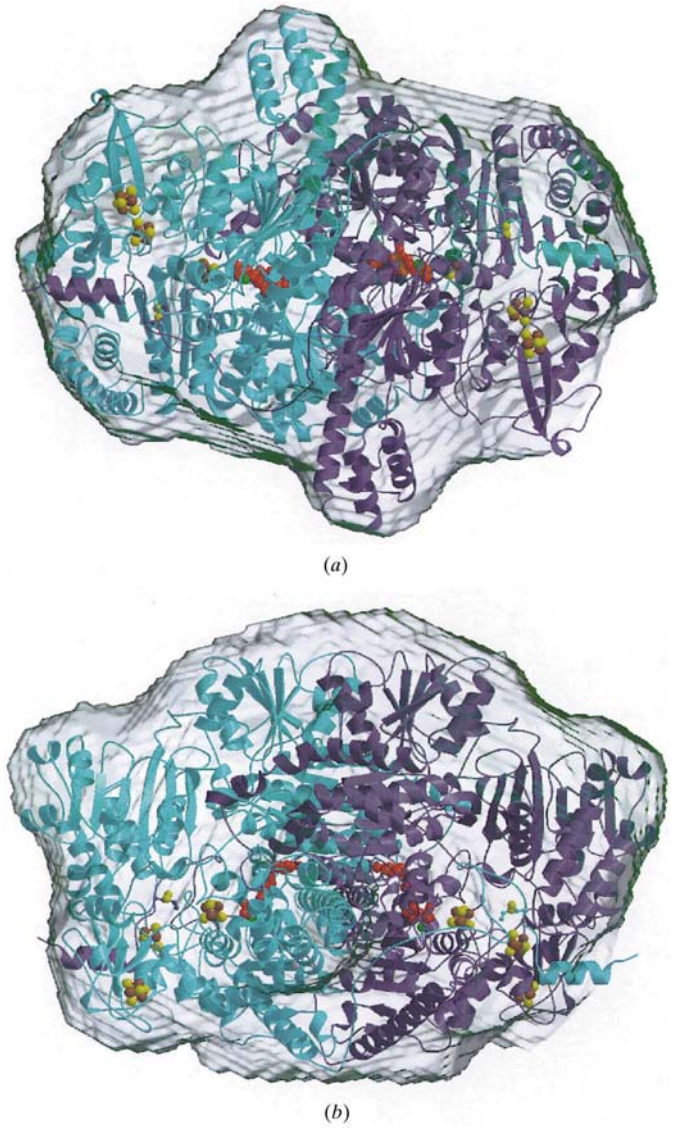

Figure 7
Fit of the PFOR protein mask, determined by real-space refinement at 5 Å resolution to the polypeptide chain of the final refined atomic model: (*a*) with the local twofold symmetry axis pointing towards the reader, and (*b*) with the same axis vertical in the plane of the picture. Arrows and spirals depict β-strands and α-helices, respectively. This figure was produced with *BOBSCRIPT* (Esnouf, 1999) and *RASTER3D* (Merrit & Bacon, 1997).

Undoubtedly, the structure determination of PFOR would have been easier if the MAD data had been more complete. However, the location of the individual Fe atoms in the [4Fe–4S] clusters, a prerequisite for determining high-resolution MAD phases, requires accurate high-resolution MAD data, which may not always be obtainable. MAD data collection is very time-consuming when the unit cell is large and when long wavelengths have to be used. In such cases it may be easier to follow the strategy chosen here, consisting of collecting reliable (and better complete) medium-resolution MAD data, modelling of clusters as single atoms, obtaining starting MAD phases and, finally, improving and extending these to high resolution by density-modification methods, using independent native data collected at short wavelengths (see also Nicolet *et al.*, 1999).

This work has confirmed the general notion that in order to obtain good phases for large structures, a large number of

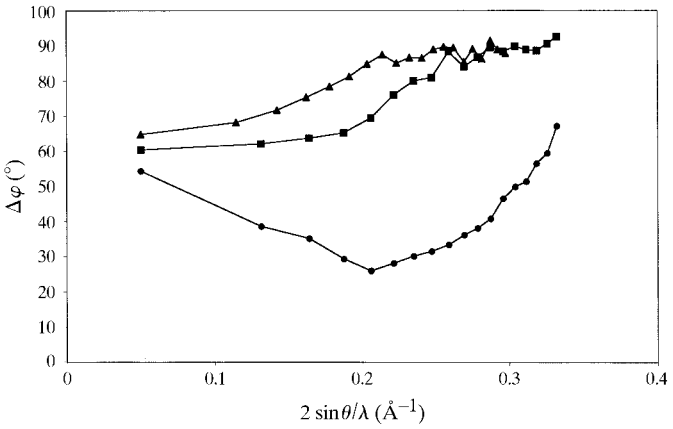

Figure 8
Average difference, as a function of resolution, between the experimental and the refined model phases, using filled triangles for MIR phases in crystal form 1, filled squares for MAD phases in form 2, and filled circles for phases obtained after the final cycle of fourfold DA in form 2.

independent data should be combined, especially if some of these provide only weak phasing power. One of the most powerful phasing methods is electron-density averaging, which may be exploited by taking advantage of several crystal forms. We have shown that even a severely biased experimental electron-density map obtained for one form may allow to solve another form by MR. In addition we found that the NCS transformations and masks required for DA can be very accurately obtained at medium resolution (5 Å for PFOR) by an RSR procedure. In conclusion, it may be anticipated that for solving larger and larger structures the availability of several well diffracting crystal forms will become increasingly useful.

We thank Fred Vellieux for his advice concerning density-averaging methods and Pierre Legrand for helpful discussions. The Commissariat à l'Energie Atomique, the Centre National de la Recherche Scientifique and the Ministère de l'Education Nationale, de la Recherche et de la Technologie are gratefully acknowledged for financial support.

References

Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*, ch. 6. London: Academic Press.
Chabrière, E., Charon, M.-H. & Vellieux, F. M. D. (1999). *Acta Cryst.* D**55**, 469–472.
Chabrière, E., Charon, M.-H., Volbeda, A., Pieulle, L., Hatchikan, E. C. & Fontecilla-Camps, J. C. (1999). *Nature Struct. Biol.* **6**, 182–190.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Esnouf, R. M. (1999). *Acta Cryst.* D**55**, 938–940.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Kleywegt, G. J. & Jones, T. A. (1994). *From First Map to Final Model*, edited by S. Bailey, R. Hubard & D. Waller, pp. 59–66. Warrington: Daresbury Laboratory.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merrit, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Nicolet, Y., Piras, C., Legrand, P., Hatchikian, C. E. & Fontecilla-Camps, J. C. (1999). *Structure*, **15**, 13–23.
Otwinowski, Z. (1991). *Isomorphous replacement and anomalous scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 60–68. Warrington: Daresbury Laboratory.
Pieulle, L., Chabrière, E., Hatchikian, E. C., Fontecilla-Camps, J. C. & Charon, M.-H. (1999). *Acta Cryst.* D**55**, 329–331.
Pieulle, L., Magro, V. & Hatchikian, E. C. (1997). *J. Bacteriol.* **179**, 5684–5692.
Ramakrishnan, V. & Biou, V. (1997). *Methods Enzymol.* **276**, 538–557.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Samudzi, C. T. & Rosenberg, J. M. (1992). *J. Appl. Cryst.* **25**, 65–68.
Sheldrick, G. M. (1997). *Methods Enzymol.* **276**, 628–641.
Sim, G. A. (1960). *Acta Cryst.* **13**, 511–512.
Smith, J. L., Zaluzec, E. J., Wery, J.-P., Niu, L., Switzer, R. l., Zalkin, H. & Satow, Y. (1994). *Science*, **264**, 1427–1433.
Vellieux, F. M. D., Hajdu, J., Verlinde, C. L., Groendijk, H., Read, R. J., Greenhough, T. J., Campbell, J. W., Kalk, K. H., Littlechild, J. A., Watson, H. C. & Hol, W. G. J. (1993). *Proc. Natl Acad Sci. USA*, **90**, 2355–2359.
Vellieux, F. M. D. A. P., Hunt, J. F., Roy, S. & Read, R. J. (1995). *J. Appl. Cryst.* **28**, 347–351.
Vellieux, F. M. D. A. P. & Read, R. J. (1997). *Methods Enzymol.* **277**, 18–53.
Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–112.