

## Direct-Method Structure Determination of the Native Azurin II Protein Using One-Wavelength Anomalous Scattering Data

ZHENG XIAO-FENG,<sup>a</sup> FAN HAI-FU,<sup>a</sup> Q. HAO,<sup>a,b,\*</sup> F. E. DODD<sup>c</sup> AND S. S. HASNAIN<sup>c</sup>

<sup>a</sup>*Institute of Physics, Chinese Academy of Sciences, Beijing 100080, China, and* <sup>b</sup>*Department of Applied Physics, De Montfort University, Leicester LE1 9BH, England, <sup>c</sup>CCLRC Daresbury Laboratory, Warrington WA4 4AD, England, and School of Applied Sciences, De Montfort University, Leicester LE1 9BH, England. E-mail: qhao@dmu.ac.uk*

(Received 11 January 1996; accepted 19 March 1996)

### Abstract

The one-wavelength anomalous scattering (OAS) X-ray diffraction data of azurin II, a copper-containing protein from *Alcaligenes xylosoxidans* were collected at the Photon Factory, Japan at a 'routine' wavelength of 0.97 Å. The structure had been originally solved by the molecular-replacement method [Dodd, Hasnain, Abraham, Eady & Smith (1995). *Acta Cryst.* D51, 1052–1064]. As a technique of *ab initio* structure determination, the direct method [Fan, Hao, Gu, Qian, Zheng & Ke (1990). *Acta Cryst.* A46, 935–939] was attempted to break the phase ambiguity intrinsic to OAS data. The phases were then improved using the solvent-flattening method. The final electron-density map clearly shows most  $C\alpha$  positions and many side chains and it is traceable without prior knowledge of the structure. It is concluded that the direct method is capable of phasing anomalous scattering data collected at one wavelength from moderate-sized native proteins ( $M_w \approx 20$  kDa) which contain copper or atoms with a similar scattering power.

### 1. Introduction

The multi-wavelength anomalous dispersion (MAD) method (generally data collected at three wavelengths) is capable of solving protein structures (Hendrickson *et al.*, 1989). However, the MAD experiments and data processing require high degree of precision and sometimes protein crystals are sensitive to X-ray irradiation limiting the ultimate quality of data. Moreover, the requirement of MAD to have the wavelengths coincide with the  $f'$  dip means long wavelengths for Fe (1.74 Å at  $K$  edge) and Cu (1.38 Å at  $K$  edge) where absorption is increased, which sometimes result in excessive radiation damage making the data collection from one crystal impractical. In principle  $F_{\lambda_1}^+$ ,  $F_{\lambda_1}^-$  and  $F_{\lambda_2}^+$  is adequate to uniquely solve the phase and involves two wavelengths. However, MAD experiments have generally used three or more wavelengths to reach a least-squares determination of the phase angle, which puts increased demand on the highly scarce synchrotron radiation beamtime. In a recent effort to minimize this demand, several strategies

for a MAD experiment have been explored and it is likely that at least in some cases the requirement could be reduced to data collection at two wavelengths (Peterson *et al.*, 1996). Therefore, there have been great efforts to resolve the phase ambiguity arising from a one-wavelength anomalous scattering (OAS) technique without using additional diffraction data. This is of particular interest when many native proteins contain iron or copper or atoms of a similar scattering power. There are successful procedures to break the phase ambiguity of OAS for solving unknown protein structures. Ramachandran & Raman (1956) proposed that for the two possible phases of each reflection one can always make that choice which has a phase closer to that of the heavy-atom contribution. Hendrickson & Teeter (1981) used a similar but improved method in the structure determination of the hydrophobic protein crambin. Their method combines the bimodal OAS phase distribution with the Sim-weight calculated from the known positions of anomalous scatterers. Wang's solvent-flattening technique (Wang, 1985) has been used to break the phase ambiguity of OAS in the structure determination of Cd, Zn metallothionein (Robbins *et al.*, 1991). Apart from the above, procedures based on the Wilson statistics (Ralph & Woolfson, 1991) and  $P_r$ -function-related techniques (Hao & Woolfson, 1989; Fan, Hao & Woolfson, 1990) have also been proposed to break the OAS phase ambiguity. Tests with data from known proteins showed that these methods are of use in practice. In a different context, direct methods have continuously been trying to break the OAS phase ambiguity (Fan, 1965; Karle, 1966; Hauptman, 1982; Giacovazzo, 1983; Fan & Gu, 1985). A common feature of this kind of method is to use three-phase structure invariants. So far the procedure of Fan & Gu (1985) is the only direct-method procedure which has been successfully tested with experimental OAS data (Fan, Hao, Gu *et al.*, 1990; Sha *et al.*, 1995). In Sha *et al.* (1995), the anomalous scattering data from core streptavidin (Hendrickson *et al.*, 1989) at the wavelength tuned very close to the Se absorption edge ( $\lambda = 0.9795$  Å,  $f'' = 3.66$ ) was used. An interpretable electron-density map was obtained after the *ab initio* direct-method phasing and the phase improvement

by solvent flattening, non-crystallographic symmetry averaging and skeletonization.

Here we present a test with the anomalous scattering data from a copper-containing protein, azurin II (Dodd, Hasnain, Abraham, Eady & Smith, 1995) collected at a routine wavelength 0.97 Å at the Photon Factory, Japan. The azurin II structure was originally solved by the molecular-replacement method using the structure of azurin from *Alcaligenes denitrificans* NCTC 8582 (Baker, 1988) with which azurin II shows a close homology. In the present study, an *ab initio* solution of the azurin II structure has been attempted; a procedure which combines direct-method phasing and solvent-flattening phase improvement has produced a traceable electron-density map.

## 2. Experimental data

In a denitrification bacterium *Alcaligenes xylosoxidans*, we have identified two azurins, azurin I being the same as previously identified (Suzuki & Iwasaki, 1962) while azurin II is a new azurin (Dodd, Hasnain, Hunter *et al.*, 1995). Both of these azurins have been found to donate electrons to the copper-containing nitrite reductase of the organism, a key enzyme in the denitrification pathway. Azurin II forms blue elongated rectangular prisms, in the tetragonal space group  $P4_122$  with unit-cell parameters  $a = b = 52.65$ ,  $c = 100.63$  Å. There is one molecule (129 residues) in the asymmetric unit with molecular weight 13 759 Da. The one-wavelength ( $\lambda = 0.97$  Å,  $f_{Cu}^{\prime} = -0.083$ ,  $f_{Cu}^{\prime\prime} = 2.206$ ) anomalous-scattering diffraction data extending to 1.9 Å were collected using the Weissenberg camera at the Photon Factory. The crystal was mounted with the spindle aligned along its  $c$  axis so that Friedel pairs were recorded on the same image plate. Rotation angles of 6° were used with a 0.5° overlap. The high speed of data collection offered by the Weissenberg method coupled with an off-line image-plate scanner system made it possible to obtain the whole data set from one crystal. The 43 645 measurements to 1.9 Å were reduced using *WEIS* program (Higashi, 1989) then scaled and merged using the *CCP4* suite (Collaborative Computational Project, Number 4, 1994), giving a merging  $R_{merge}$  of 4.0% for 6217 independent Friedel pairs (or unique reflections for centric ones). These data have a completeness of 62% to 1.9 Å and 79% to 2.5 Å. Further details are given in Dodd, Hasnain, Abraham *et al.* (1995).

## 3. Locating the copper site

The copper anomalous scatterer was located by the conventional direct-method program *SAPI* (Fan, Yao, Zheng, Gu, Qian, 1991). Magnitudes of anomalous differences,

$$|\Delta F(\mathbf{H})| = ||F(\mathbf{H})| - |F(-\mathbf{H})||,$$

for reflections within 3.0 Å resolution were used as per the strategy suggested by Mukherjee, Helliwell & Main (1989). The phase set number 10 was selected by a default run of the program as the solution which had figures of merit: ABSFOM = 0.6266, PSIZERO = 2.272, RESIDUAL = 22.35. The resultant electron-density map produced a highest peak which was three times higher than the next highest one. The highest peak was chosen as the copper site (fractional coordinates  $x = 0.1849$ ,  $y = 0.1404$ ,  $z = 0.2479$ ) which is very close to the copper site ( $x = 0.1827$ ,  $y = 0.1395$ ,  $z = 0.2497$ ) found by the original molecular replacement method (Dodd, Hasnain, Abraham *et al.*, 1995).

## 4. Absolute configuration

The space group  $P4_122$  is one of the exceptional space groups where the correct analysis of anomalous scattering effects leads neither simply to an absolute configuration nor to polar-axis direction (Jones, 1984). However, once the space group itself is determined the absolute configuration is determined simultaneously because the space-group specific symmetry information has been encoded into the algebraic representation of the triplet phase-invariant relationships in direct-method programs (Langs, Guo & Hauptman, 1995).

## 5. Evaluation of phase doublets

The phase doublets inherent in the OAS method are expressed as,

$$\varphi_H = \varphi'_H \pm |\Delta\varphi_H|. \quad (1)$$

Where  $\varphi'_H$  is the phase of,

$$F''_{ano} = \sum_{j=1}^N i f_j'' \exp(i2\pi \mathbf{H} \cdot \mathbf{r}_j), \quad (2)$$

which can be calculated from the known positions of the anomalous scatterers and the known value of  $f''$ ,  $|\Delta\varphi_H|$  is obtained from (see, Blundell & Johnson, 1976),

$$\cos\Delta\varphi_H = (F_H^+ - F_H^-) / 2|F''_{ano}|. \quad (3)$$

The phase problem in the OAS case is in fact a sign problem according to (1). The probability for  $\Delta\varphi_H$  positive is given by Fan & Gu (1985),

$$P_+(\Delta\varphi_H) = \frac{1}{2} + \frac{1}{2} \tanh\left\{ \sin|\Delta\varphi_H| \left[ \sum_{H'} m_{H'} m_{H-H'} \kappa_{H, H'} \right. \right. \\ \left. \left. \times \sin(\Phi'_3 + \Delta\varphi_{H'best} + \Delta\varphi_{H-H'best}) \right. \right. \\ \left. \left. + \chi \sin\delta_H \right] \right\}. \quad (4)$$

The procedure to use (4) for *ab initio* phasing of the OAS data of azurin is the same as that described by Fan, Hao, Gu *et al.* (1990). The program used, *TSAPI* was written in FORTRAN 77. All 6217 Friedel pairs

Table 1. Initial phase errors from different methods

Reflections were sorted in descending order of  $F_{\text{obs}}$  and then cumulated into seven groups with number of reflections shown in the first column. Phase errors were calculated against the refined structure (Dodd, Hasnain, Abraham, Eady & Smith, 1995). Weighted average phase errors were calculated using  $F_{\text{obs}}$  as weights. I, initial phase errors for Wang's (1985) real-space filter method. II, phase errors of direct-method results.

Number of reflections	Phase error (°)			
	Mean	I Weighted	Mean	II Weighted
1000	67.0	67.8	60.1	60.9
2000	66.2	67.0	61.5	61.7
3000	68.3	68.0	63.9	63.1
4000	70.5	69.2	65.9	64.2
5000	72.6	70.2	68.8	65.6
6000	75.5	71.3	72.5	67.0
6217	76.1	71.4	73.5	67.3

at 1.9 Å were used as input. The direct-method phases were compared with phases calculated from the refined model (Dodd, Hasnain, Abraham *et al.*, 1995) and the results are arranged in descending order of  $F_{\text{obs}}$  and then cumulated into seven groups as listed in Table 1. For a comparison, initial phasing results obtained by the real-space-filter method (Wang, 1985) are also listed. It is evident in the present test that the direct method gave better initial phases.

### 6. Phase improvement by solvent-flattening technique

Initial phases obtained by the direct method in the present test are not sufficiently accurate for producing an interpretable Fourier map if no prior knowledge of the structure is known. The solvent-flattening technique (Wang, 1985) available in CCP4 suite (Collaborative Computational Project, Number 4, 1994) has been used to improve the direct-method phases. The solvent content was calculated to be 45%. The initial phase error of 73.5° from the direct method reduced and converged to 64.0° after four cycles (Table 2). The map correlation coefficient between the electron-density map calculated from the improved phases and that from phases of Dodd, Hasnain, Abraham *et al.* (1995) was 0.51. The electron-density map calculated from solvent-flattening-improved direct-method phases was then put in the modelling program *O* (Jones, Zou, Cowan & Kjeldgaard, 1991). Using the automatic bone search option in *O*, the majority of the  $C\alpha$  trace was recognizable in addition to many side chains. The map is traceable without prior knowledge of the structure. Three sections of the map superimposed on the refined model of Dodd, Hasnain, Abraham *et al.* (1995) are shown in Fig. 1 as representative examples. The electron density for the  $\beta$ -strand (residues 4–10) and around the Cu site is good, while some discontinuity occurs in the loop region (residues 86–92).

Table 2. Average phase errors and map correlation coefficients

Reflections were sorted in descending order of  $F_{\text{obs}}$  and then cumulated into seven groups with number of reflections shown in the first column. Phase errors were calculated against the refined structure (Dodd, Hasnain, Abraham *et al.*, 1995). Weighted average phase errors were calculated using  $F_{\text{obs}}$  as weights. Correlation coefficients were calculated with respect to the Fourier maps phased by the refined structure. I, results from solvent flattening alone (initial phases based on column I, Table 1). II, results from direct-method phases (column II, Table 1) and solvent-flattening improvement.

Number of reflections	Phase error (°)			
	Mean	I Weighted	Mean	II Weighted
1000	58.1	54.6	53.4	50.0
2000	58.0	54.5	55.0	51.0
3000	61.2	56.3	57.8	52.8
4000	63.5	57.9	59.8	54.1
5000	64.8	58.8	61.7	55.2
6000	66.4	59.7	63.7	56.3
6217	66.6	59.9	64.0	56.5
Correlation coefficient		0.48		0.51

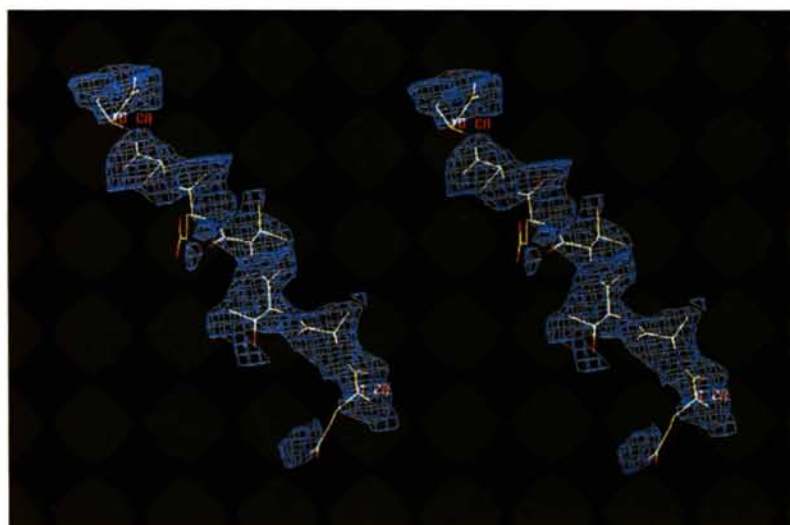
Table 3. Statistics of four data sets used for examining the importance of data quality

Data sets *A* and *D* provided the successful solution. Data set *B* was constructed from the first 45° of data *A*. Data set *D* was obtained by merging three data sets, one each for oxidized, reduced and mixed form.

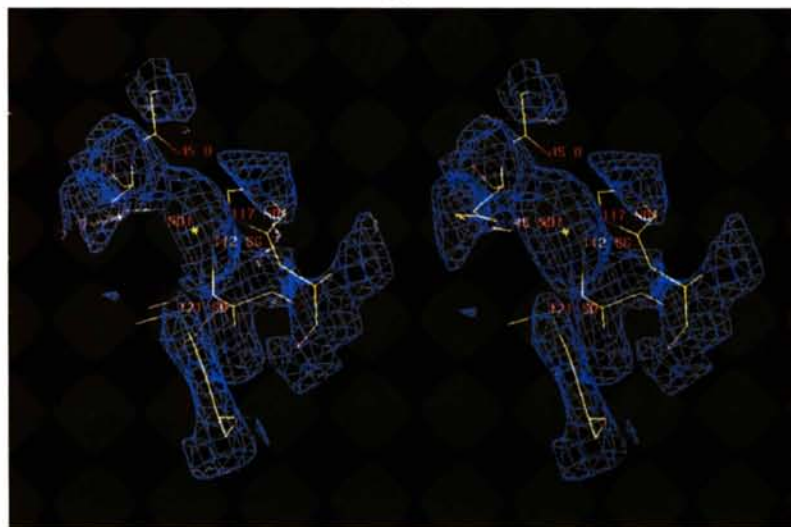
	Data (A)	Data (B)	Data (C)	Data (D)
Data resolution (Å)	1.9	1.75	1.9	1.75
1.9 Å				
$R_{\text{merge}}$ (%)	4.0	6.0	3.5	8.0
Completeness (%)	61.8	88.5	52.3	95.4
Multiplicity	6.0	3.0	3.5	8.7
3.0 Å				
$R_{\text{merge}}$ (%)	3.4	4.5	2.9	5.9
Completeness (%)	88.8	95.3	87.4	98.4
Multiplicity	7.9	3.4	4.2	10.0

### 7. Importance of data quality

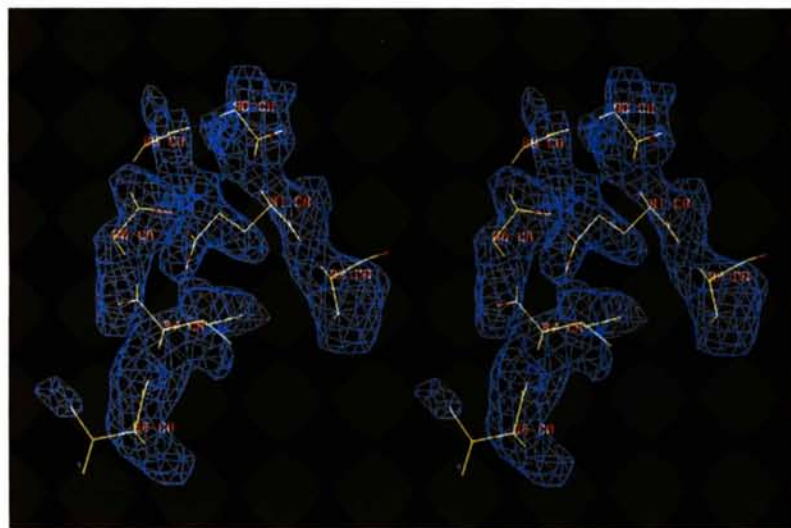
A second data set collected at the Photon Factory in a later session has also been processed (Dodd & Hasnain, unpublished results). This data set was also collected from a single azurin II crystal but using X-ray wavelength of 1.0 Å and two-image-plate camera, 2B. Thus, a higher completeness of 82% to 1.75 Å resolution (see Table 3) was achieved but Friedel pairs were recorded on different image plates as a two-plate camera was used for data collection. The plates were divided into two groups corresponding to the upper and lower position in the cassette. Each group was scanned on the same scanner system. The 34 309 measurements were reduced using the *WEIS* program and then scaled and merged using the *CCP4* suite, giving a merging  $R_{\text{merge}}$  of 6.0% for 7672 Friedel pairs. However, with this data, the copper site could not be located using either the Patterson method or the direct-method program *SAPI*



(a)



(b)



(c)

Fig. 1. Regions of the electron-density map calculated from direct-method plus solvent-flattening phases: (a) a piece of  $\beta$ -strand (residues 4–10); (b) near the copper site; (c) a loop region (residues 86–92).

even after the number of phasing trials was increased to 500.

A comparison of the 1.75 Å data (set *B*) with the 1.9 Å data (set *A*) reveals several differences;  $R_{\text{merge}}$  for data *B* is worse than data *A* (4.5 versus 3.4% at 3 Å). Also, data *B* has a lower multiplicity compared with data *A* (3.4 versus 7.9 at 3 Å). Even though data *B* has a better completeness at 1.9 Å (89 versus 62%), the completeness of two data sets is very similar at 3 Å (95 versus 89%). We note that data *A* was collected with single-plate camera while data *B* was obtained with a two-plate cassette.

We have examined the effect of multiplicity and  $R_{\text{merge}}$  by attempting the structure determination using two additional data sets, details of which are included in Table 3. Data *C* was constructed from data *A* by only merging data for the first 45°, while data *D* was obtained by merging data *B* with two additional data sets, one of reduced protein and the other of partially reduced protein crystal (Dodd & Hasnain, unpublished data). No copper site was found when data *C* was used which has a multiplicity of 4.2, similar to data *B*, and  $R_{\text{merge}}$  of 2.9% (at 3 Å). In contrast, data *D* provided a clear solution, even though this data set was obtained by merging data sets from three different crystals, each of which was obtained using a two-plate cassette. Data *D* has a multiplicity of 10.0 at 3 Å and  $R_{\text{merge}}$  of 5.9%.

## 8. Discussions

The present test showed that the direct method is capable of phasing one-wavelength anomalous scattering (OAS) diffraction data from a native protein which contains a weak anomalous scatterer like Cu ( $f'' \approx 2.2$ ). However, the combination of the direct method with solvent flattening is important for obtaining an interpretable Fourier map. This provides a useful alternative in solving a *de novo* protein structure without either preparing isomorphous heavy-atom derivative crystals or collecting multi-wavelength diffraction data. The success of the method with two of the four data sets clearly suggests that a data of high multiplicity is essential for a successful solution. It is uncertain whether such multiplicity is of importance in the MAD method. This method does not require the wavelength to be tuned near the absorption edge, thus it may be of particular interest for phasing data collected at conventional X-ray sources. However, data collected near the absorption edge should have stronger anomalous scattering signals. For copper proteins, data collected at an optimized wavelength 1.380 Å ( $f'' = 3.90$ ) may provide an interpretable map for a 20 kDa molecule.

Drs Z. H. L. Abraham and R. R. Eady are thanked for providing azurin proteins and for their continuous interest in the project. We are indebted to Professor N. Sakabe for providing experimental facilities at the

Photon Factory. We acknowledge the help of Drs I. Harvey and R. Strange during the data collection. We are also grateful to the Royal Society and Chinese Academy of Science for supporting collaboration between our groups at De Montfort, UK, and Beijing, China.

## References

- Baker, E. N. (1988). *J. Mol. Biol.* **203**, 1071–1095.
- Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*, p. 338. New York: Academic Press.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Dodd, F., Hasnain, S. S., Abraham, Z. H., Eady, R. R. & Smith, B. E. (1995). *Acta Cryst.* **D51**, 1052–1064.
- Dodd, F., Hasnain, S. S., Hunter, W. N., Abraham, Z. H., Debenham, M., Kanzler, H., Eldridge, M., Eady, R. R., Ambler, R. P. & Smith, B. E. (1995). *Biochemistry*, **34**, 10180–10186.
- Fan, H. F. (1965). *Acta Phys. Sin.* **21**, 1114–1118.
- Fan, H. F. & Gu, Y. X. (1985). *Acta Cryst.* **A41**, 280–284.
- Fan, H. F., Hao, Q., Gu, Y. X., Qian, J. Z., Zheng, C. D. & Ke, H. (1990). *Acta Cryst.* **A46**, 935–939.
- Fan, H. F., Hao, Q. & Woolfson, M. M. (1990). *Acta Cryst.* **A46**, 659–664.
- Fan, H. F., Yao, J. X., Zheng, C. D., Gu, Y. X. & Qian, J. Z. (1991). *SAPI91, A Computer Program for Automatic Solution of Crystal Structures from X-ray Diffraction Data*, Institute of Physics, Chinese Academy of Sciences, Beijing, People's Republic of China.
- Giacovazzo, C. (1983). *Acta Cryst.* **A39**, 585–592.
- Hao, Q. & Woolfson, M. M. (1989). *Acta Cryst.* **A45**, 794–797.
- Hauptman, H. (1982). *Acta Cryst.* **A38**, 632–641.
- Hendrickson, W. A., Pahler, A., Smith, J. L., Satow, Y., Merritt, E. A. & Phizackerley, R. P. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 2190–2194.
- Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
- Higashi, T. (1989). *J. Appl. Cryst.* **22**, 9–18.
- Jones, P. G. (1984). *Acta Cryst.* **A40**, 660–662.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.
- Karle, J. (1966). *Acta Cryst.* **21**, 273–276.
- Langs, D. A., Guo, D. & Hauptman, H. A. (1995). *Acta Cryst.* **D51**, 1020–1024.
- Mukherjee, A. K., Helliwell, J. R. & Main, P. (1989). *Acta Cryst.* **A45**, 715–718.
- Peterson, M. R., Harrop, S. J., McSweeney, S. M., Leonard, G. A., Thompson, A. W., Hunter, W. N. & Helliwell, J. R. (1996). *J. Synchrotron Rad.* **3**, 24–34.
- Ralph, A. C. & Woolfson, M. M. (1991). *Acta Cryst.* **A47**, 533–537.
- Ramachandran, J. N. & Raman, S. (1956). *Curr. Sci.* **25**, 348–351.
- Robbins, A. H., McRee, D. E., Williamson, M., Collett, S. A., Xuong, N.-H., Furey, W. F., Wang, B. C. & Stout, C. D. (1991). *J. Mol. Biol.* **221**, 1269–1293.
- Sha, B. D., Liu, S. P., Gu, Y. X., Fan, H. F., Ke, H., Yao, J. X. & Woolfson, M. M. (1995). *Acta Cryst.* **D51**, 342–346.
- Suzuki, H. & Iwasaki, H. (1962). *J. Biochem.* **52**, 193–199.
- Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–111.