

Direct Phasing of One-Wavelength Anomalous-Scattering Data of the Protein Core Streptavidin

BY SHA BING-DONG, LIU SHEN-PING, GU YUAN-XIN AND FAN HAI-FU

Institute of Physics, Chinese Academy of Sciences, Beijing 100080, People's Republic of China

HENGMING KE

Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC 27599, USA

AND YAO JIA-XING AND M. M. WOOLFSON

Department of Physics, University of York, Heslington, York YO1 5DD, England

(Received 31 August 1994; accepted 28 October 1994)

Abstract

The direct method [Fan, Hao, Gu, Qian, Zheng & Ke (1990). *Acta Cryst.* **A46**, 935–939] was used to break the phase ambiguity intrinsic to one-wavelength anomalous scattering data from a known protein of moderate size, core streptavidin, which was solved originally with three-wavelength anomalous diffraction data [Hendrickson, Pähler, Smith, Satow, Merritt & Phizackerley (1989). *Proc. Natl Acad. Sci. USA*, **86**, 2190–2194]. Unlike that in the previous test with a small protein, the Fourier map calculated with the direct-method phases could not clearly reveal the moderate-sized protein structure. However, the phases can be improved step by step using Wang's solvent-flattening method, non-crystallographic symmetry averaging and the skeletonization method. The final electron-density map clearly shows most C_α positions and some side chains and it is traceable without prior knowledge of the structure. It is concluded that the direct method is capable of breaking the OAS phase ambiguity of a moderate-sized protein at moderate resolution such as 3 Å, while the combination of direct methods with macromolecular techniques may produce phases good enough for unknown protein structure to be traced.

Introduction

Great efforts have been made to resolve the phase ambiguity arising from one-wavelength anomalous scattering (OAS) without using additional diffraction data. This is of importance in protein crystallography since most protein crystals are sensitive to X-ray irradiation and isomorphous derivatives are not always easy to prepare. 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

reflexion 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 distribution 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, McRee, Williamson, Collett, Xuong, Furey, Wang & Stout, 1991). Apart from the above, procedures based on the Wilson statistics (Ralph & Woolfson, 1991) and P_5 -function-related techniques (Okaya, Saito & Pepinsky, 1955; Hao & Woolfson, 1989; Fan, Hao & Woolfson, 1990) have also been proposed to break the OAS phase ambiguity. Tests with experimental OAS 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; Hazell, 1970; Sikka, 1973; Heinerman, Krabbendam, Kroon & Spek, 1978; Hauptman, 1982; Giacovazzo, 1983; Fan, Han & Qian, 1984; 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 is the only direct-methods procedure which has been successfully tested with experimental OAS data of a small protein (Fan, Hao, Gu, Qian, Zheng & Ke, 1990). Here we present a test with a much more complicated protein, core streptavidin (Hendrickson, Pähler, Smith, Satow, Merritt & Phizackerley, 1989) using an improved phasing procedure which combines the *ab initio* direct-method phasing and the step-by-step phase improvement by solvent flattening, non-crystallographic symmetry averaging and the method of skeletonization. An interpretable electron-density map was obtained at the end. To demonstrate the advantage of

using the direct method, a comparison was made between the *ab initio* phasing power of the direct method, the method of solvent flattening and the combination of bimodal OAS distribution with Sim distribution.

Data

The crystal structure of the selenobiotin-binding core of streptavidin was determined at 3.1 Å resolution by Hendrickson, Pähler, Smith, Satow, Merritt & Phizackerley (1989) using three-wavelength anomalous-diffraction data. The crystals are in space group *I222* with unit-cell parameters $a = 95.27$, $b = 105.41$, $c = 47.56$ Å. There are four tetramers in the unit cell. In the following test, the one-wavelength ($\lambda = 0.9795$ Å, $\Delta f'_{\text{Se}} = -6.203$, $\Delta f''_{\text{Se}} = 3.663$) anomalous-scattering data collected at 3 Å were used. The total number of reflexions is 4578. In comparison to the previous test example aPP (Blundell, Pitts, Tickle, Wood & Wu, 1981) the protein streptavidin in this test is about nine times bigger, while the anomalous scatterer is even weaker.

Locating heavy-atom sites

The heavy-atom sites were located by conventional direct methods with the OAS data. For details of the technique readers are referred to Mukherjee, Helliwell & Main (1989) and Fan, Hao, Gu, Qian, Zheng & Ke (1990). Magnitudes of anomalous differences for reflexions at 5 Å resolution were used. A default run of the program *SAPI* (Fan, Yao, Zheng, Gu & Qian, 1991) resulted in an *E*-map containing three large peaks in addition to a number of smaller ones. The first and the third peaks correspond to the two independent Se atoms while the second is a ghost peak which was eliminated easily after Karle recycling. Coordinates of the two Se atoms are comparable with those found with three-wavelength anomalous-diffraction data by the original authors (Table 1).

Selection of absolute configuration

While the heavy-atom sites are derived from the value of

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

it is impossible to determine the absolute configuration of the heavy-atom arrangement without additional information. In protein crystallography a way round this difficulty is to try both alternatives and then the correct one may be obvious by the appearance of the resultant map. Recently a procedure for solving this problem was proposed by Woolfson & Yao (1994). The procedure is based on the properties of the P_s function (Okaya, Saito & Pepinsky, 1955). Hao & Woolfson (1989) showed that a sum function $Q(\mathbf{r})$ of the P_s function based on the known positions of the anomalous scatterers could reveal

Table 1. *Coordinates of Se atoms*

	Results from the direct method			Results of Hendrickson <i>et al.</i> (1989)		
	<i>x</i>	<i>y</i>	<i>z</i>	<i>x</i>	<i>y</i>	<i>z</i>
Se1	0.3394	0.1114	-0.0946	0.33861	0.11314	-0.09436
Se2	0.2058	0.0471	0.2468	0.20736	0.05111	0.24254

the protein structure in some cases. The function $Q(\mathbf{r})$ is defined as,

$$Q(\mathbf{r}) = (1/V) \sum_{\mathbf{H}} \text{Sum}(\mathbf{H}) [|\mathbf{F}(\mathbf{H})| - |\mathbf{F}(-\mathbf{H})|] \sin(2\pi\mathbf{H} \cdot \mathbf{r}),$$

$$\text{Sum}(\mathbf{H}) = \sum_{j=1}^N \exp(i2\pi\mathbf{H} \cdot \mathbf{R}_j),$$

where \mathbf{R}_j is the positional vector of the *j*th anomalous scatterer. If the sum function $Q(\mathbf{r})$ has been calculated for the correct configuration then it will contain a positive image of the structure with the weight N (the number of times the superposition was performed to generate the sum function) plus other sundry positive and negative images of lesser weight. If the opposite configuration was used, then the sum function will contain a negative image. Now the value of,

$$C = \int Q(\mathbf{r})^3 d\mathbf{v},$$

taken over the whole unit cell can be used to predict the correct configuration of the anomalous scatterers. A positive value of C means that the configuration used to calculate the sum function $Q(\mathbf{r})$ is correct, otherwise the configuration used should be changed to its opposite. In the present example, the C value calculated with the heavy-atom sites found by *SAPI* (listed in Table 1) was 2.05×10^6 , while that from the alternative was -2.05×10^6 . Hence the coordinates listed in Table 1 correspond to the correct configuration. As a cross check, direct-method phasing of the OAS data (see following sections) based on the correct heavy-atom configuration resulted in an average phase error of 70° while that based on the alternative resulted in an error of 89° .

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 φ'_H is the phase of,

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

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

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

In the present test, a large number of reflexions have a value of $\cos \Delta\varphi_H$ outside the range of -1 to $+1$. This is due to the error in data measurement and scaling. In order to reduce the error, the whole set of $\cos \Delta\varphi_H$ values were not used directly in further calculations, instead they were arranged in descending order and then modified to fit into a uniform distribution between -1 and $+1$. We found that such a fitting makes the result even worse, if the accuracy of experimental data is good enough, as was the case in testing with the OAS data of the protein aPP (Fan, Hao, Gu, Qian, Zheng & Ke, 1990). However when the error is large, as in the test with the single isomorphous replacement data of Rhe (Fan & Wang, 1988) and in the present example, considerably better results can be obtained.

Ab initio phasing of OAS data

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

$$P_+(\Delta\varphi_H) = \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right) \tanh \left\{ \sin |\Delta\varphi_H| \times \left[\sum_{H'} m_{H'} m_{H-H'} \kappa_{H,H'} \times \sin(\Phi_3 + \Delta\varphi_{H',\text{best}} + \Delta\varphi_{H-H',\text{best}}) + \chi \sin \delta_H \right] \right\} \quad (4)$$

The procedure to use (4) for *ab initio* phasing of the OAS data of streptavidin is the same as that described by Fan, Hao, Gu, Zheng & Ke (1990). The program used was originally written by Hao Quan and modified by two of the authors, FHF and GYX. 4578 reflexions at 3.1 Å resolution were used as input. The κ_{\min} value for accepting \sum_2 relationships in the direct-method procedure was set to 0.03. The results are arranged in descending order of F_{obs} and then cumulated into five groups as listed in Table 2. For a comparison, *ab initio* phasing results obtained by the solvent-flattening technique (Wang, 1985) and by the method combining bimodal OAS phase distribution and Sim distribution are also listed. It is evident in the present test that the direct method gave the best initial phases.

Improvement of direct-method phases

Initial phases obtained by the direct method in the present test are encouraging but they are not accurate enough for producing an interpretable Fourier map, if nothing is known about the structure in advance. A number of techniques have been tried to improve the direct-method phases. The solvent-flattening technique was found to be most powerful. A program for iterative solvent flattening was written by one of the authors

Table 2. *Initial phase errors from different methods*

Reflexions were sorted in descending order of F_{obs} and then cumulated into five groups with number of reflexions shown in the first column. Phase errors were calculated against the refined structure. I, OAS phase ambiguity broken by the solvent-flattening method. II, OAS phase ambiguity broken by combining the bimodal phase distribution with Sim distribution of the anomalous scatterers. III, OAS phase ambiguity broken by the direct method.

Number of reflexions	Average phase error (°)		
	I	II	III
1000	64.4	61.8	56.7
2000	65.8	64.2	60.8
3000	68.1	65.9	63.5
4000	71.1	69.2	67.9
4578	73.2	71.3	69.8

(SBD) according to the principle described by Wang (1985). The initial average phase error of 69.8° from the direct method reduced and converged to 63.9° after a few cycles of solvent flattening. The solvent content was assumed to be 50% at the beginning, it was changed only slightly after solvent flattening. At this stage, a twofold axis inside the asymmetric unit was clearly recognized on the electron-density map, the orientation and position of which was accurately determined. Based on this, non-crystallographic symmetry averaging was then carried out resulting in an average phase error of 57.6° . Finally the skeletonization method (Baker, Bystroff, Fletterick & Agard, 1993) was used to further improve the phases bringing the error down to 54.1° . The program for performing the skeletonization was written by one of the authors (LSP). Table 3 shows the cumulative results of phase errors in various stages. Fig. 1 shows a portion of the corresponding electron-density maps containing three β -strands. The result from *ab initio* direct-method phasing of the OAS data (Fig. 1a) reveals good electron density for both backbone and side chains of the left strand and some scattering density for the middle and right strands. This indicates that the direct method is capable of breaking the phase ambiguity of OAS data, but not sufficient to give a traceable map. When the solvent-flattening method is incorporated, the overall quality of the electron density is improved although it is still not traceable. For example, the density around Trp92 shows up in Fig. 1(b), in contrast to the noise in Fig. 1(a). The incorporation of local symmetry averaging dramatically improved the electron density. Many side chains, such as Trp75 in Fig. 1(c), are visible and density is connected to several fragments. Finally, the combination of the skeletonization with the above methods resulted in an interpretable map (Fig. 1d) which is of similar quality to that calculated from the refined structure (Fig. 1e).

Discussion

The final Fourier map at 3.1 Å resolution shows that the electron density is connected into seven fragments. The electron density for the eight-stranded β -barrels, the hydrophobic core of the structure and for the biotin is

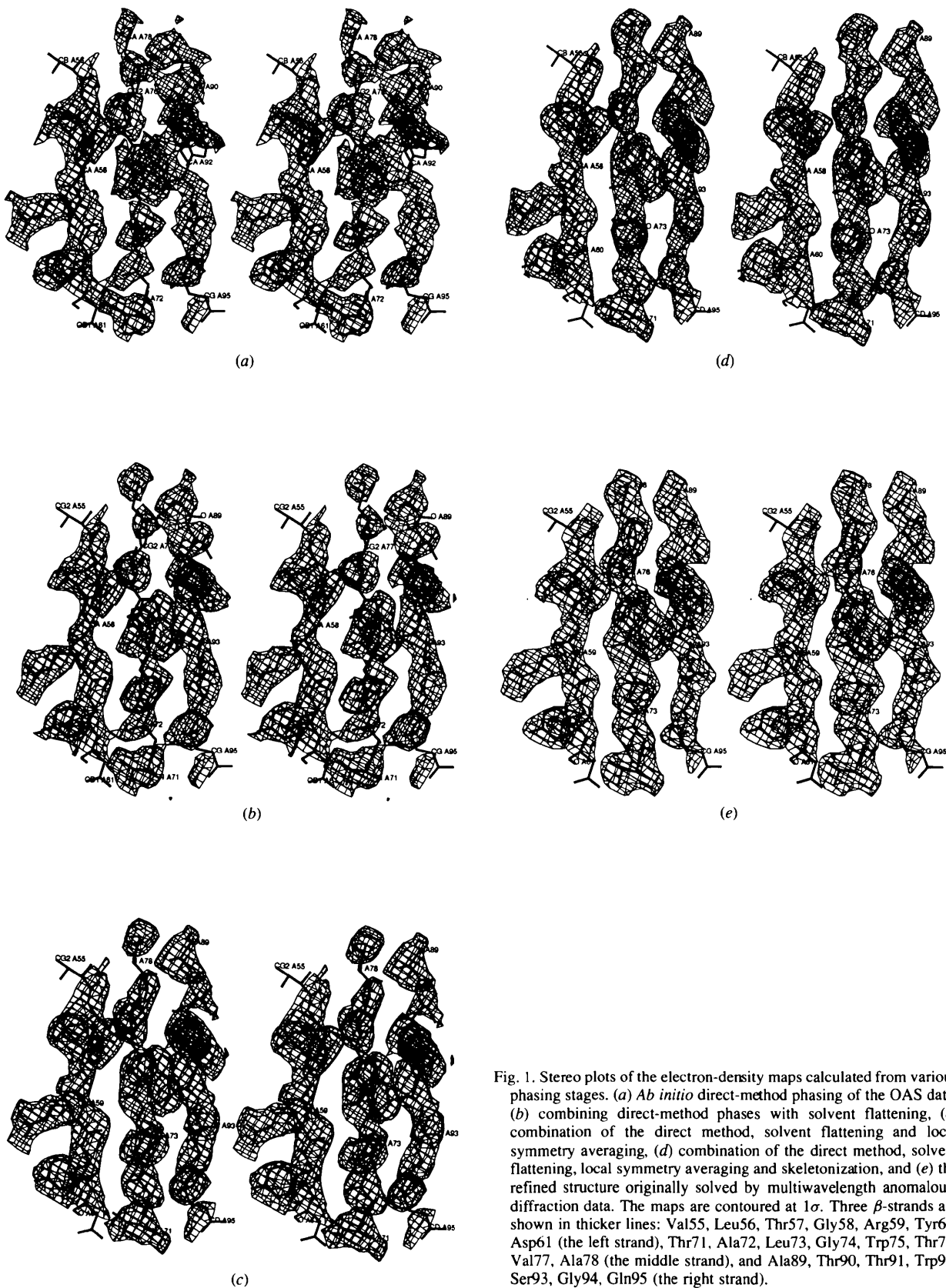


Fig. 1. Stereo plots of the electron-density maps calculated from various phasing stages. (a) *Ab initio* direct-method phasing of the OAS data, (b) combining direct-method phases with solvent flattening, (c) combination of the direct method, solvent flattening and local symmetry averaging, (d) combination of the direct method, solvent flattening, local symmetry averaging and skeletonization, and (e) the refined structure originally solved by multiwavelength anomalous-diffraction data. The maps are contoured at 1σ . Three β -strands are shown in thicker lines: Val55, Leu56, Thr57, Gly58, Arg59, Tyr60, Asp61 (the left strand), Thr71, Ala72, Leu73, Gly74, Trp75, Thr76, Val77, Ala78 (the middle strand), and Ala89, Thr90, Thr91, Trp92, Ser93, Gly94, Gln95 (the right strand).

Table 3. Average phase errors and map correlation coefficients in various phasing stages

Reflexions were sorted in descending order of F_{obs} and then cumulated into five groups with number of reflexions shown in the first column. Phase errors were calculated against the refined structure. Weighted average phase errors were calculated using F_{obs} as weights. Correlation coefficients were calculated with respect to the Fourier map phased by the refined structure. I, results from the direct method. II, results from solvent flattening based on direct-method phases. III, results from non-crystallographic symmetry averaging based on II. IV, results from skeletonization based on III.

Number of reflexions	Average phase error (°)			
	I	II	III	IV
1000	56.7	45.4	32.9	27.0
2000	60.8	52.3	41.4	35.3
3000	63.5	55.6	47.6	42.1
4000	67.9	60.7	54.2	49.4
4578	69.8	63.9	57.6	54.1
Weighted average phase error	63.9	55.9	47.3	42.3
Correlation coefficient	0.46	0.56	0.68	0.71

excellent. Discontinued density occurred at the surface of the molecule. Some of the unobserved loops are even disordered in the refined structure. Positions of most C_{α} atoms are recognizable in addition to many side chains such as Trp, Tyr and Phe which are markers for tracing of an unknown protein structure. Worth mentioning especially is the fact that four out of six tryptophan residues are directly recognizable. This will greatly help with tracing. In short, the map is traceable without prior knowledge of the structure.

The present test showed that the direct method is a powerful technique of breaking the phase ambiguity inherent in one-wavelength anomalous-diffraction experiments, while a combination of the direct method with solvent flattening and other macromolecular methods is important for obtaining an interpretable Fourier map. This provides a useful alternative in solving a *de novo* protein structure without either preparing isomorphous crystals or collecting multiwavelength diffraction data.

We wish to thank Dr Wayne Hendrickson, whose data of streptavidin have enabled us to carry out this study. We are indebted to Professors Liang Dong-chai and

Chang Wen-zhui for their helpful discussion. Our gratitude is also due to the British Royal Society and the Chinese Academy of Sciences for supporting collaboration between our laboratories in Beijing, China, and York, England.

References

- BAKER, D., BYSTROFF, C., FLETTERICK, R. J. & AGARD, D. A. (1993). *Acta Cryst.* **D49**, 429–439.
- BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography*, p. 338. New York: Academic Press.
- BLUNDELL, T. L., PITTS, J. E., TICKLE, I. L., WOOD, S. P. & WU, C. W. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 4175–4179.
- 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., HAN, F. S. & QIAN, J. Z. (1984). *Acta Cryst.* **A40**, 495–498.
- 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. & WANG, B. C. (1988). Unpublished work, partial results of which are quoted in *Physical and Non-physical Methods of Solving Crystal Structures*, Chapter V, by WOOLFSON, M. M. & FAN, H. F. (1994). London: Cambridge Univ. Press.
- FAN, H. F., YAO, J. X., ZHENG, C. D., GU, Y. X. & QIAN, J. Z. (1991). *SAPI-91, 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. (1993). *Acta Cryst.* **A39**, 585–592.
- HAO, Q. & WOOLFSON, M. M. (1989). *Acta Cryst.* **A45**, 794–797.
- HAUPTMAN, H. (1982). *Acta Cryst.* **A38**, 632–641.
- HAZELL, A. C. (1970). *Nature (London)*, **227**, 269.
- HEINERMAN, J. J. L., KRABBENDAM, H., KROON, J. & SPEK, A. L. (1978). *Acta Cryst.* **A34**, 447–450.
- HENDRICKSON, W. A., PÄHLER, 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.
- KARLE, J. (1966). *Acta Cryst.* **21**, 273–276.
- MUKHERJEE, A. K., HELLIWELL, J. R. & MAIN, P. (1989). *Acta Cryst.* **A45**, 715–718.
- OKAYA, Y., SAITO, Y. & PEPINSKY, R. (1955). *Phys. Rev.* **98**, 1857–1858.
- 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., MCRREE, 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.
- SIKKA, S. K. (1973). *Acta Cryst.* **A29**, 211–212.
- WANG, B. C. (1985). *Methods Enzymol.* **115**, 90–111.
- WOOLFSON, M. M. & YAO, J. X. (1994). *Acta Cryst.* **D50**, 7–10.