Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 29 January 2003
Accepted 26 February 2003

© 2003 International Union of Crystallography
Printed in Great Britain – all rights reserved

Ab initio protein phasing at 1.4 Å resolution

Maria C. Burla,[a] Benedetta Carrozzini,[b] Giovanni L. Cascarano,[b] Liberato De Caro,[b] Carmelo Giacovazzo[c,b]* and Giampiero Polidori[a]

[a]Dipartimento di Scienze della Terra, Piazza Università, 06100 Perugia, Italy, [b]Istituto di Cristallografia, CNR, c/o Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy, and [c]Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy. Correspondence e-mail: carmelo.giacovazzo@ic.cnr.it

All the techniques today available for the *ab initio* crystal structure solution of proteins require that the atomicity condition is satisfied. Accordingly, diffraction data at resolution equal or better than 1.2 Å are necessary. This condition reduces the role of the *ab initio* techniques in macromolecular crystallography. The computer program *SIR2002* has been modified in such a way that it may succeed also with 1.4 Å resolution diffraction data. The modifications concern all the modules of the program: the modified program also benefits by the efficiency of a figure of merit.

1. Introduction

SIR–MIR (single–multiple isomorphous replacement), SAD–MAD (single–multiple wavelength anomalous dispersion) and SIRAS–MIRAS (the combination of isomorphous replacement and anomalous dispersion) techniques are rather insensitive to the data resolution: *i.e.* data at 3 Å or lower often provide an amount of information sufficient for the crystal structure solution. *Ab initio* methods for phasing proteins rely on only one set of diffraction data and require that the atomicity condition is satisfied: *i.e.* diffraction data at resolution equal to or better than 1.2 Å are usually necessary. Few cases are reported in the literature in which this limit has been overcome (see, for example, Mukherjee & Woolfson, 1995; Mukherjee *et al.*, 1999, 2000). Four computer programs for the *ab initio* phasing of proteins have well documented reports: *Shake-and-Bake* (Weeks *et al.*, 1994; Rappleye *et al.*, 2002), *SHELX-D* (Sheldrick, 1998), *ACORN* (Foadi *et al.*, 2000) and *SIR2002* [Burla *et al.* (2002); this is the heir of *SIR2000-N* (Burla *et al.*, 2001)]. The four programs use different phasing techniques:

(*a*) *Shake-and-Bake* constrains the phases by using the minimum principle (De Titta *et al.*, 1994) and refines them by repeatedly cycling from direct to reciprocal space.

(*b*) *SHELX-D* develops the molecular model by alternating tangent refinement (*i.e.* the reciprocal-space search) and direct-space techniques [*i.e. via* the use of the *iterative peaklist optimization*, see Sheldrick & Gould (1995)].

(*c*) *ACORN* locates a small fragment of the molecule (possibly by molecular replacement techniques) to obtain a useful non-random starting set of phases. Then solvent-flattening techniques are used to refine them.

(*d*) *SIR20002* combines the random phase approach with the tangent procedure. Then real-space techniques are used to refine the phases, without alternating them with reciprocal-space methods.

All four programs require, for the success of the *ab initio* phasing, that the data resolution is better than 1.2 Å. Such a limit is generally considered critical because the atomicity condition (so basic for direct methods) is violated at lower data resolutions. On the other hand, the resolution limit of 1.2 Å strongly reduces the impact of direct-methods programs on macromolecular crystallography: indeed, in spite of the experimental advances produced by the use of synchrotron radiation and by the cryocrystallographic techniques, only a small percentage of protein structures are able to produce reliable diffraction data at 1.2 Å.

This paper breaks down the limit of 1.2 Å: it shows that *SIR2002* can be suitably modified so as to succeed with a resolution limit of 1.4 Å. In the following the new program will be called *SIR2003*.

2. The *SIR2002* approach and its applications to 1.4 Å resolution data

SIR2002 is a program for the solution of small size (say less than 80 non-hydrogen atoms in the asymmetric unit), medium size (from 81 to 200 non-hydrogen atoms in the asymmetric unit) and large size (more than 200 non-hydrogen atoms in the asymmetric unit) crystal structures. To reduce computing time, *SIR2002* uses different phasing procedures for different structure sizes. Since we are interested only in the crystal structure solution of proteins, we will avoid any reference to the procedural aspects concerning small and medium size molecules.

Table 1
Code name and crystallochemical data for protein test structures.

PDB is the file code in the Protein Data Bank, when available; R (Å) is the experimental data resolution in Å; NASYM is the number of non-hydrogen atoms in the asymmetric unit, H_2O is the number of water molecules. In the last column, the species and number of heavy atoms in the asymmetric unit are specified.

Structure code	PDB	R (Å)	Space group	NASYM – H_2O	Heavy atoms
App[(a)]	–	0.99	$C2$	302	Zn
Collagen[(b)]	2knt	1.20	$P2_1$	465 – 50	S_6 P
Conotoxin[(c)]	1a0m	0.90	$I4$	255 – 42	S_{10}
Crambin[(d)]	–	0.83	$P2_1$	329 – 75	S_6
H42q[(e)]	1b0y	0.93	$P2_12_12_1$	594 – 206	S_9 Fe_4
Hipip[(e)]	1cku	1.20	$P2_12_12_1$	1229 – 334	S_{18} Fe_8
Hirustasin[(f)]	1bx7	1.20	$P4_32_12$	365 – 52	S_{11}
Lactalbumin[(g)]	1b9o	1.15	$P2_12_12_1$	935 – 164	S_{10} Ca
Oxidoreductase[(h)]	1mfm	1.02	$P2_12_12_1$	1106 – 283	S_2 Cl_2 Cu Zn Cd_9
Vancomycin[(i)]	1sho	1.09	$P4_32_12$	207 – 108	Cl_6

References: (*a*) Glover *et al.* (1983); (*b*) Merigeau *et al.* (1998); (*c*) Hu *et al.* (1998); (*d*) Weeks *et al.* (1995); (*e*) Parisini *et al.* (1999); (*f*) Uson *et al.* (1999); (*g*) Harata *et al.* (1999); (*h*) Ferraroni *et al.* (1999); (*i*) Schafer *et al.* (1996).

The modules of *SIR2002* and their use in a standard procedure for protein crystal structure solution may be shortly described as follows:

(i) *TT*: the triple tangent module, which aims at producing useful sets of phases starting from random phase sets.

(ii) *EDM*: an electron-density modification procedure. It performs 3 supercycles, each constituted by 7 microcycles $\rho \rightarrow \varphi \rightarrow \rho$.

(iii) *HAFR*: associates the heaviest atomic species present in the crystal to a number of peaks selected from the electron-density map. The occupancy factor is proportional to the peak height and to its chemical connectivity. *HAFR* is constituted by 12 cycles $\rho \rightarrow \varphi \rightarrow \rho$.

(iv) *LSQH*: refines the isotropic displacement parameters of the heavy atoms *via* a least-squares (4 cycles) procedure.

(v) *FR*: the isotropic displacement parameters of all the atoms are modified (6 cycles) in such a way that those corresponding to the strongest peaks will be given the smallest displacement parameters.

(vi) *DLSQ*: the completion and refinement of the structural model is performed *via* a procedure that alternates least squares and $(2F_{obs} - F_{cal})$ Fourier maps.

The goodness of a trial solution is judged by the RAT figure of merit:

$$\mathrm{RAT} = CC/\langle R_{cal}^2 \rangle, \qquad (1)$$

where

$$CC = \frac{(\langle R_{obs}^2 w^2 \rangle - \langle R_{obs}^2 \rangle \langle w^2 \rangle)}{(\langle R_{obs}^4 \rangle - \langle R_{obs}^2 \rangle^2)^{1/2} (\langle w^4 \rangle - \langle w^2 \rangle^2)^{1/2}} \qquad (2)$$

is the correlation coefficient between the R_{obs}'s comprised in the interval (0.3, 1.2) and the corresponding Sim-like coefficients (Sim, 1959)

$$w = D_1(R_{obs} R_{cal}/2); \qquad (3)$$

$R_{obs} = |E_{obs}|$ and $R_{cal} = |E_{cal}|$ are the moduli of the normalized structure factors; w is the reliability factor of the phase estimate, available after the last cycle of *EDM*, $D_1(x) = I_1(x)/I_0(x)$, I_i is the modified Bessel function of order i. The average in the denominator of (1) is calculated over the reflections with the smallest values of F_{obs} (about 30% of the total). Obviously, RAT is expected to be maximum for the correct structure. RAT operates as follows: if RAT < RATM, where RATM is the maximum value of RAT obtained in the previous trials, then the procedures *FR* and *DLSQ* are skipped (to save computing time) and a new trial is launched. For the trials with favourable values of RAT, the crystallographic residual

$$\mathrm{RES} = \frac{\sum ||F_{obs}| - |F_{cal}||}{\sum |F_{obs}|}$$

is calculated at the end of the least-squares module *DLSQ*. The program stops when RES < 0.25.

We have selected for our applications ten protein structures (see Table 1), chosen so as to provide a variety of space groups and subsets of structures with and without atoms heavier than sulfur. The standard *SIR2002* program succeeds in solving them (Burla *et al.*, 2002), provided the full experimental data resolution is used: this is quoted as R (Å) in column 3 of Table 1.

We then cut the data to 1.4 Å resolution and explored, by *SIR2002*, 500 trials for each test structure (*SIR2002*, at the experimental data resolution, was always able to find the correct solution among the first 87 trials). Success was attained only for three structures: H42q, Hipip and Oxidoreductase. In Table 2, we show the correlation-factor values (COR2002) between the electron-density map calculated with the published phases at 1.4 Å resolution and that obtained at the end of a successful run of *SIR2002*. In the three successful cases, COR2002 > 0.60, otherwise COR2002 was always less than 0.3 (when a dash is given in the fourth column of Table 2).

We may then conclude that *SIR2002* is inefficient at data resolutions for which the atomicity condition is violated.

3. The *SIR2003* approach and its applications to 1.4 Å resolution data

Cutting the data at 1.4 Å resolution produces several severe effects:

(*a*) The atomicity condition is violated (*i.e.* two electron-density peaks may partially overlap if they correspond to bonded atoms).

(*b*) The number of measurable reflection intensities markedly decreases. For the structures listed in Table 1, the ratio between the reflections measured up to a resolution of 1.4 Å and those measured up to the full experimental data resolution ranges from 0.17 to 0.46, indicating the loss of a great amount of experimental information. As a consequence, the tangent procedures will be less efficient because a smaller number of reliable triplet invariants will be available at 1.4 Å resolution.

Table 2
For each test structure, we show: NREF, the number of unique reflections experimentally measured; NO/NP, the ratio of the number of observations to the number of structural parameters at the experimental resolution (we excluded water molecules from the calculation) (in parentheses the value NO/NP when the data resolution is truncated at 1.4 Å); the correlation factor between the best map obtained by the phasing procedure and the true (published) map for both *SIR2002* and *SIR2003*; the CPU time needed by *SIR2003* for solving each structure, expressed in hours.

			CORMAP		
Structure code	NREF	NO/NP	COR2002	COR2003	CPU time
App	15653	7.6 (2.5)	–	0.71	0.4
Collagen	16000	5.2 (3.3)	–	0.61	9.7
Conotoxin	9243	5.6 (2.7)	–	0.64	0.3
Crambin	28727	13.0 (3.0)	_	0.60	1.6
H42q	41951	10.9 (3.3)	0.64	0.73	3.2
Hipip	51872	6.4 (4.1)	0.72	0.81	6.6
Hirustasin	15894	6.6 (4.1)	–	0.85	25.6
Lactalbumin	37648	6.7 (3.4)	–	0.73	10.5
Oxidoreductase	69678	9.7 (3.7)	0.71	0.80	1.1
Vancomycin	11867	8.5 (4.1)	–	0.76	2.7

(*c*) The electron-density modification process will be less effective in driving badly approximated phases to the correct values.

(*d*) The ratio NO/NP = number of observations/number of structural parameters is strongly reduced (see column 3 of Table 2). The average value of NO/NP (calculated over all the test structures) is equal to 8.0 when the resolution is the experimental one, it is equal to 3.4 when the data have been cut to 1.4 Å resolution. In these conditions, any least-squares procedure (trying to refine atomic positions) would become unreliable: it would diminish the residual factor between observed and calculated structure-factor moduli without providing a sound structural model. This behaviour is confirmed in all our tests by *SIR2002* at 1.4 Å resolution.

In order to overcome the limits of *SIR2002*, we introduced some new features in *SIR2003*, which are described below:

(i) The grid of the electron-density maps in *SIR2002* was always 0.33 Å. In *SIR2003*, it is fixed to 1/3 of the data resolution (*i.e.* for the tests at 1.4 Å resolution the grid is fixed at 0.47 Å).

(ii) The module *TT* uses the P_{10} formula (Cascarano *et al.*, 1984) as a default tool for estimating the triplet cosine invariants. The default threshold for their reliability factor was 0.30 for *SIR2002*. Owing to the lack of the triplet invariants at 1.4 Å resolution, the threshold has been decreased to 0.25 in *SIR2003*.

(iii) The weight w defined by (3) has been used in *SIR2002* as a criterion for estimating the reliability of the phases calculated *via* the map inversions. In *SIR2003*, the argument of the D_1 function has been increased (according to the *EDM* cycle) from $0.5R_{\rm obs}R_{\rm cal}$ to $2R_{\rm obs}R_{\rm cal}$.

(iv) In *SIR2002*, the atomicity concept was introduced in the module *HAFR*. The reader could observe that such a concept may hardly be applied to data with 1.4 Å resolution. However, *HAFR* is successfully applied also by *SIR2003*: in some way, its results are equivalent to those produced by an *EDM* inversion using spherical peaks (in practice, *HAFR* calculates the structure factors by using the locations of the electron-density peaks as atomic positions). In *SIR2003*, *HAFR* is applied to the largest NASYM/6 peaks (NASYM is the number of non-hydrogen atoms in the asymmetric unit), and the number of cycles $\rho \rightarrow \varphi \rightarrow \rho$ has been reduced from 12 to 5.

(v) In accordance with the point (*d*) of this section, the modules *FR* and *DLSQ* have been eliminated.

(vi) *Use of the negative electron density map*. In the *EDM* techniques, it is usual to fix a threshold value for the electron density, say TRρ, such that $\rho_{\rm mod} = \rho$ when $\rho >$ TRρ, $\rho_{\rm mod} = 0$ when $\rho <$ TRρ. Usually TR$\rho \gg 0$. Accordingly, *SIR2002* fixes TRρ in such a way that only a percentage of the pixels, variable from 2.5 to 3.5%, is not zeroed in the electron-density modification process. Owing to the Babinet principle, inverting the full positive region of the electron-density map provides phases that are correlated with those arising from the full negative region of the map. This suggests that even the negative regions of the map contain structural information available for active use. In *SIR2003*, we leave unvaried the *SIR2002* criterion for the positive region of the map, and we introduce a supplementary threshold TRρn for the negative regions. TRρn is chosen in such a way that no more than 1.8% of the total number of pixels (those with the most negative values of the electron density) are actively used in the map inversion. In this way, the correlation between the phase indications provided by the selected positive and by the selected negative regions of the map is not unitary, owing to the quite small percentage of pixels involved in the inversion.

(vii) *The powering of the electron-density map.* The Sayre (1952) equation, and therefore the tangent methods, have their counterpart in direct space in the squared structure. Other authors (Hoppe & Gassmann, 1968; Gassmann & Zechmeister, 1972) proposed to modify the electron density according to

$$\rho_{\rm mod} = a\rho + b\rho^2 + c\rho^3. \qquad (4)$$

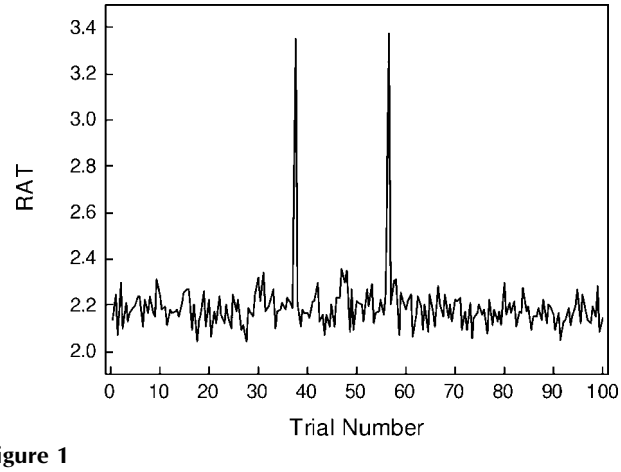

Figure 1
The final values of RAT for the first 100 trials for the Hipip structure. Only two of them are solutions of the structure. In these trials, RAT assumes values (>3) much higher than those assumed for unsuccessful trials (near 2).

We found it useful to modify the electron density according to the following criterion (Refaat & Woolfson, 1993):

$$\rho_{\rm mod} = \rho^{1.3} \quad \text{if} \quad \rho > 0, \quad \rho_{\rm mod} = 0 \quad \text{otherwise.} \qquad (5)$$

We apply the above modification once per *EDM* supercycle.

(viii) *Identification of the correct solution*. We found that in *SIR2002* the figure of merit RES was effective for identifying the correct solution only after the use of *DLSQ*, which has been eliminated from *SIR2003*. Luckily, RAT proved to have a reserve of power: *SIR2003* stops when RAT is larger than a suitable threshold (on the basis of our experimental results, RAT > 3 is a sensible choice). In order to show the usefulness of this figure of merit, in Fig. 1 we plot, for Hipip, the final values of RAT given by the program *SIR2003* for the first 100 trials. Only two of them correspond to solutions of the structure. For these trials, RAT assumes values much higher than those relative to the unsuccessful trials.

(ix) *The trial iteration*. The *SIR2003* phasing process is just a tool for driving random phases to the correct values. Only a few trials are successful: quite often the process is trapped in false minima and any effort to escape them is in vain. Sometimes the driving process is incomplete: one run of *SIR2003* is not enough to perform the complete driving process. We realized that the phases available at the end of an *SIR2003* trial can, on some occasions, be used as input for a supplementary run of *SIR2003* and lead to better phases. How to distinguish trials suitable for iteration from trials unsuitable? The numerator of RAT (*i.e. CC*) proved to be quite efficient in identifying the favourable trials: the condition for the iteration we use is

$$CC > \langle CC \rangle + \sigma_{CC},$$

where $\langle CC \rangle$ is the average value of *CC* calculated over the previous *SIR2003* trials, and σ_{CC} is the corresponding standard deviation. The iteration stops when RAT does not increase any more. In the iterated trials, the maximum RAT value is sometimes smaller than 3. In these cases, the program stops when

$$\text{RAT} > \langle \text{RAT} \rangle + 5\sigma_{\rm RAT},$$

where $\langle \text{RAT} \rangle$ is the average value of RAT calculated over the previous *SIR2003* trials, and $\sigma_{\rm RAT}$ is the corresponding standard deviation.

The application of *SIR2003* to the test structures listed in Table 1 (at 1.4 Å resolution) gives the results shown in the fifth column of Table 2. COR2003 is the correlation factor between the electron-density map calculated with the published phases at 1.4 Å resolution and that corresponding to the highest RAT value. In all the cases, COR2003 > 0.60: all the maps are therefore interpretable. As an example, in Fig. 2 we show some details of the Lactalbumin electron-density map as automatically provided by *SIR2003*.

It is worthwhile noting that the solution of six test structures required the use of the iterative process: App (2 cycles), Collagen (6 cycles), Conotoxin (3 cycles), Crambin (5 cycles), Lactalbumin (3 cycles), Vancomycin (2 cycles). As an example, in Fig. 3, we show the results obtained for Collagen: we give for each iteration the values of RAT, of CORMAP (the correlation factor between the electron-density map available at the end of the run and the true map at 1.4 Å resolution), and of ERPHA (the phase error). RAT attains its maximum in correspondence with the highest value of the CORMAP and the minimum value of ERPHA: it regularly decreases with decreasing correlation values and increasing phase errors.

A further detail of Table 2 deserves to be noticed: for the three structures solvable also by *SIR2002*, COR2003 values are significantly larger than COR2002: this suggests better convergence properties of the new program. Finally, the last column of Table 2 gives the CPU time necessary for solving each structure, expressed in hours (all the numerical tests have been performed using a Xeon 1.7 GHz processor, Linux operating system).

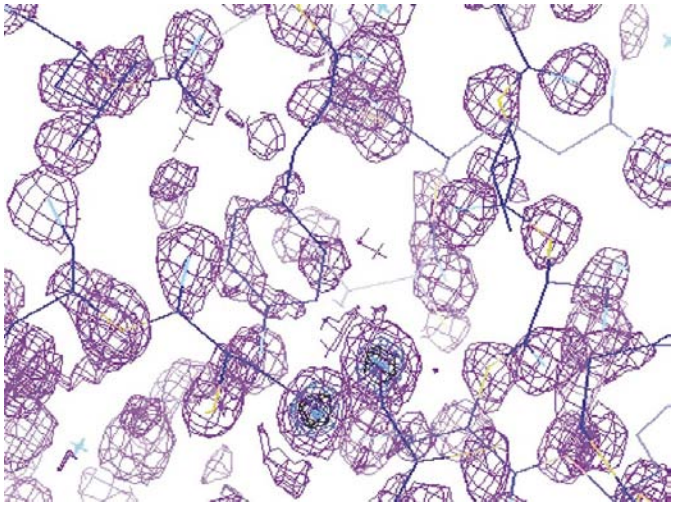

Figure 2
Region of the electron-density map of Lactalbumin (as provided by *SIR2003*) close to sulfur atoms Cys61 A SG and Cys77 A SO (in black). The image was obtained using *XtalView* (v. 4.1; McRee, 1999).

4. Discussion

None of the documented computing programs aiming at solving *ab initio* protein structures is supplied with tools for

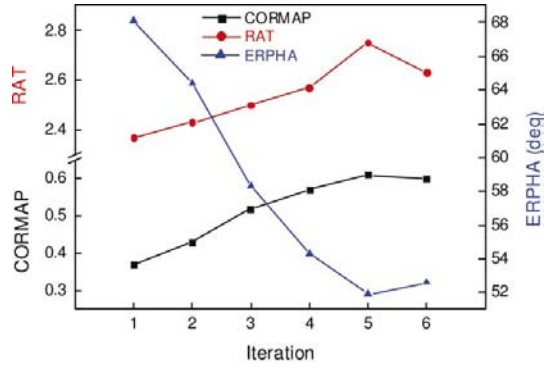

Figure 3
Collagen: the values of RAT (red curve), CORMAP (black curve) and ERPHA (blue curve) for each iterated run.

succeeding with 1.4 Å resolution data. The common belief that atomic resolution (1–1.2 Å) is a mandatory condition for success is denied by the clear success of the approach implemented in *SIR2003*.

It may be useful to note that:

(*a*) Truncating the experimental data at 1.4 Å resolution (while the crystals diffract to higher resolution) is not equivalent to using data from crystals that only diffract at 1.4 Å: such truncated data are expected to be of much higher quality.

(*b*) Some of the test structures contain atoms heavier than sulfur and this considerably improves the performances of direct methods at lower resolution.

However, the relatively modest computing times necessary for attaining most of the crystal structure solutions seem to indicate that *SIR2003* has a reserve of power and that the program may succeed also for crystals with poorer resolution limit and not containing atoms heavier than sulfur. *SIR2003* may therefore be considered a tool for future efforts aiming at: (*a*) reducing the computing time by the use of early (*i.e.* operating at an early stage of the program) figures of merit; (*b*) increasing the range of structural complexity affordable by the procedure; (*c*) bringing the resolution limit to 1.5–1.6 Å, by using experimental data from crystals diffracting at this resolution.

References

Burla, M. C., Camalli, M., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Polidori, G. & Spagna, R. (2001). *J. Appl. Cryst.* **34**, 523–526.

Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C. & Polidori, G. (2002). *Z. Kristallogr.* **217**, 629–635.

Cascarano, G. L., Giacovazzo, C., Camalli, M., Spagna, R., Burla, M. C., Nunzi, A. & Polidori, G. (1984). *Acta Cryst.* A**40**, 278–283.

De Titta, G. T., Weeks, C. M., Thuman, P., Miller, R. & Hauptman, H. A. (1994). *Acta Cryst.* A**50**, 203–210.

Ferraroni, M., Rypniewski, W., Wilson, K. S., Viezzoli, M. S., Banci, L., Bertini, I. & Mangani, M. (1999). *J. Mol. Biol.* **288**, 413–426.

Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Jia-xing, Y. & Chao-de, Z. (2000). *Acta Cryst.* D**56**, 1137–1147.

Gassmann, J. & Zechmeister, K. (1972). *Acta Cryst.* A**28**, 270–280.

Glover, I., Haneef, I., Pitts, J. E., Wood, S. P., Moss, D., Tickle, I. J. & Blundell, T. L. (1983). *Biopolymers*, **22**, 293–304.

Harata, K., Abe, Y. & Muraki, M. (1999). *J. Mol. Biol.* **287**, 347–358.

Hoppe, W. & Gassmann, J. (1968). *Acta Cryst.* B**24**, 97–107.

Hu, S. H., Loughnan, M., Miller, R., Weeks, C. M., Blessing, R. H., Alewood, P. F., Lewis, R. J. & Martin, J. L. (1998). *Biochemistry*, **37**, 11425–11433.

McRee, D. E. (1999). *J. Struc. Biol.*, **125**, 156–165.

Merigeau, K., Arnoux, B., Norris, K., Norris, F. & Ducruix, A. (1998). *Acta Cryst.* D**54**, 306–312.

Mukherjee, M., Ghosh, S. & Woolfson, M. M. (1999). *Acta Cryst.* D**55**, 168–172.

Mukherjee, M., Maiti, S. & Woolfson, M. M. (2000). *Acta Cryst.* D**56**, 1132–1136.

Mukherjee, M. & Woolfson, M. M. (1995). *Acta Cryst.* D**51**, 626–628.

Parisini, E., Capozzi, F., Lubini, P., Lamzin, V., Luchinat, C. & Sheldrick, G. M. (1999). *Acta Cryst.* D**55**, 1773–1784.

Rappleye, J., Innus, M., Weeks, C. M. & Miller, R. (2002). *J. Appl. Cryst.* **35**, 374–376.

Refaat, L. S. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 367–371.

Sayre, D. (1952). *Acta Cryst.* **5**, 60–65.

Schafer, M., Schneider, T. R. & Sheldrick, G. M. (1996). *Structure*, **4**, 1509–1515.

Sheldrick, G. M. (1998). In *Direct Methods for Solving Macromolecular Structures*, edited by by S. Fortier. Dordrecht: Kluwer Academic Publishers.

Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* B**51**, 423–431.

Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.

Uson, I., Sheldrick, G. M., de La Fortelle, E., Bricogne, G., Di Marco, S., Priestle, J. P., Grutter, M. G. & Mittl, P. R. (1999). *Structure Fold. Des.* **7**, 55–63.

Weeks, C. M., De Titta, G. T., Hauptman, H. A., Thuman, P. & Miller, R. (1994). *Acta Cryst.* A**50**, 210–220.

Weeks, C. M., Hauptman, H. A., Smith, G. D., Blessing, R. H., Teeter, M. M. & Miller, R. (1995). *Acta Cryst.* D**51**, 33–38.