Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Properties of an electron-density map derived from a limited number of experimentally determined triplet phases

Kerstin Hölzer,† Edgar Weckert* and Klaus Schroer†

Institut für Kristallographie, Universität Karlsruhe (TH), D-76128 Karlsruhe, Germany

† Present address: Biology Department, Brookhaven National Laboratory, Upton, NY 11973, USA.

Correspondence e-mail: edgar.weckert@physik.uni-karlsruhe.de

Received 26 June 1999
Accepted 22 December 1999

In a previous communication [Weckert *et al.* (1999). *Acta Cryst.* D**55**, 1320–1328], the feasibility of the measurement of a large set of triplet phases by three-beam interference was demonstrated. This paper reports the methodology for the calculation of an electron-density map from this limited amount of experimental phase information and the map's properties with respect to model building and refinement. The tetragonal form of hen egg-white lysozyme (HEWL) was chosen as a test structure for the development of this method. The quality of the electron-density map obtained from all measured triplet phases allows a straightforward and nearly complete interpretation. The starting model was refined to a final R value of 17.4%. In a second step, the minimum number of phased reflections needed for the interpretation of an electron-density map was investigated, applying criteria based on $|F|$ and resolution.

1. Introduction

The feasibility of the direct experimental determination of triplet phases has been shown in various contributions (Hümmer *et al.*, 1991; Chang *et al.*, 1991; Weckert *et al.*, 1993; Weckert & Hümmer, 1997; Mo *et al.*, 1998). Very recently, a set of 847 triplet phases from the tetragonal form of lysozyme has been published (Weckert *et al.*, 1999). The ∼750 single phases derived from these triplet phases extend to slightly higher than 3 Å resolution and belong to those reflections with the largest structure-factor moduli. Crystals suitable for three-beam interference experiments must only have a small mosaic spread and must be of reasonable size. These experiments require synchrotron radiation. At an ESRF bending magnet, about 100 triplet phases can be measured per day using good-quality crystals.

This paper will give a short outline of how to obtain the single phases necessary for the calculation of an electron-density map from experimental triplet phases. Subsequently, the quality of this map will be discussed with respect to model building and refinement. Since the measurement of triplet phases is slow compared with an intensity-data collection, we will discuss the minimum number of triplet phases necessary to solve a protein by this approach.

2. Deriving single phases from triplet phases

The three-beam interference experiment provides invariant phase differences $\Phi_3 = \varphi(\mathbf{g}) + \varphi(\mathbf{h} - \mathbf{g}) - \varphi(\mathbf{h})$, which resemble triplet phases $\Phi_T = \varphi(\mathbf{g}) + \varphi(\mathbf{h} - \mathbf{g}) + \varphi(-\mathbf{h})$ in the

case of negligible anomalous dispersion effects. Therefore, one must devise a method to derive the single phases that are necessary for any further analysis. In general, the number of experimentally determined triplet phases will be small compared with the total number of reflections. For this reason, an efficient use of the measured phase information is required.

The approach we have developed is quite similar to the one applied in direct methods (*e.g.* Ladd & Palmer, 1980; Giacovazzo, 1980), with the difference that direct methods rely mainly on reflections with large $|E|$ and use statistical estimates in order to obtain phase information, whereas three-beam diffraction provides directly experimental triplet phases only for reflections with large $|F|$ for experimental reasons (Weckert & Hümmer, 1997). Depending on the space group, the phases of up to three reflections can be fixed arbitrarily, taking into account possible phase restrictions owing to symmetry, in order to obtain a standard origin. This is equivalent to the selection of an origin in direct space. Since the three-beam interference experiment provides not only the modulus of the triplet phases but also their signs, no additional enantiomorph-fixing reflection need be used. The experimental determination of triplet phases showed that reflections with the largest structure-factor moduli participate in a large number of measurable three-beam cases. Therefore, taking into account the properties of the experiment, these reflections are preferable to fix the origin. Another source of directly determined single phases are semi-invariant reflections *via* Σ_1 relations (Giacovazzo, 1980). If a three-beam case in which the reciprocal-lattice vectors resemble a Σ_1 triplet [that means with the reflections

$$\mathbf{h}_S, \quad \mathbf{g}, \quad -\mathbf{g}\,\mathbf{R}, \tag{1}$$

where $\mathbf{R}$ represents the matrix of the rotational part of a space-group symmetry operation] gives rise to an interpretable interference effect, the phases of semi-invariant reflections $\mathbf{h}_S$ can be determined directly, as the phases of the two other three-beam reflections $\mathbf{g}$, $-\mathbf{g}\mathbf{R}$ cancel except for a symmetry phase shift. In the case of tetragonal lysozyme, 26 Σ_1 triplets could be measured.

Starting with these phases, further single phases can be derived from a search for triplets with two phases already known, thus obtaining the phase of the third reflection. A limit was imposed for the maximum number of triplet phases on which a single derived phase depends, to keep phase-error propagation small. In our case, up to 7–8 dependencies can be tolerated without a significant increase of the mean phase error. Reflections important for this 'phasing tree' can be phased *via* different paths or branches in order to increase the reliability of their phase value. At a certain stage, it will not be possible to find further triplets with two known phases of sufficient reliability. The procedure chosen is again similar to direct methods. In order to continue, the phase of a reflection with large $|F|$, being part of many unused experimental triplets, has to be assigned a symbolic phase value. This will give rise to a number of new connections and derived single phases. The procedure is repeated until all experimental triplets have been used or a maximum number of symbolic phases has been reached. During the course of the experiment, we tried to replace such symbolic phases as far as possible by newly determined experimental values. Details of the algorithm used for the optimal selection of symbolic phases and for the proposal of new triplets that should be measured will be published elsewhere (Weckert, in preparation).

The remaining unknown symbolic phases have to be determined by other means. For the case study presented here, in the final data set four symbolic phases giving rise to 40 permutations were necessary to derive 752 single phases from 847 triplet phases. In total, only six reflections depend on symbolic phases. This is a consequence of the fact that during the experiment the selection procedure for new three-beam cases tries first to select cases with any reflection for which the phase has been assigned a symbol and two reflections of known phase. The application of this procedure enables us to keep the number of symbolic phases small, as well as the number of reflections that depend on symbolic phases. To fix the origin, two (tetragonal space group) very low resolution reflections with large $|F|$ values (210 and 521) were chosen. Their phase values were assigned to the phases calculated from the known model in order to be able to compare the determined electron densities with the model coordinates. The phase values of the symbolic phases were permuted by the 'magic integer' method (White & Woolfson, 1975; Main, 1977, 1978), which probes the phase space for the symbols with a mean phase error of about 45°. It is expected that the figures of merit available from 'direct methods' might not work very well for the determination of the most probable phase combination for the unknown symbols. Therefore, a different approach was adopted. Each permuted phase set was treated as a base set of known phases and exposed to an entropy-maximization algorithm. A maximum-entropy electron density based on normalized structure factors $U(\mathbf{h})$ using the derived and permuted phases was thus calculated for each phase permutation. Applying the extrapolating possibilities of maximum-entropy density, one can obtain a measure of the moduli $|U(\mathbf{k})|$ of the so far unused and unphased reflections with known measured moduli of the non-base set in order to probe the likelihood of a particular phase permutation. The procedure follows very closely the ideas of Bricogne (Bricogne, 1984, 1991; Bricogne & Gilmore, 1990), except for details in the algorithm (Weckert, in preparation). If the number of reflections within the base set is beyond some critical limit (which is still under investigation), there is a statistical correlation between the relative entropy, the relative likelihood and the mean phase error (compared with PDB entry 193l) of the different permuted sets (Fig. 1). This enables the determination of the most likely phase combinations for the symbolic phases that lead to a consistent set of single phases, which can be used to calculate an electron-density map. This method, however, only works reliably in cases where a minimum number of phases depend on a symbolic phase. Entropy and likelihood show a small sensitivity to the change of the phase of only one reflection out of 700.

3. Maximum-entropy electron-density maps

After the determination of a consistent set of phases, an electron-density map can be calculated using only the moduli and the phases of the reflections phased by three-beam diffraction. The number of known phases is comparatively small. They belong, however, to those reflections which have the largest structure-factor moduli (Weckert *et al.*, 1999). Under these circumstances, an ordinary Fourier synthesis is very likely to show significant truncation effects, especially in regions of low density. For this reason, a maximum-entropy (ME) Fourier synthesis based on $F(\mathbf{h})$ was calculated and used for further interpretations and in general shows less artificial features.

At certain stages of the 'triplet-phase' data collection, such ME electron-density maps were calculated based on the experimentally determined structure-factor phases and moduli that were available, applying the principles described in the previous paragraph. These maps were compared with the known structure model from PDB entry 193l (Vaney *et al.*, 1996) in order to monitor the improvement of the electron-density map by the increasing number of experimentally determined triplet phases.

The results of four different stages will be summarized briefly.

(i) Using 375 triplet phases (corresponding to 322 structure-factor phases) the protein molecule could already be recognized in the density map distinct from the surrounding solvent.

(ii) After the collection of 600 triplet phases (corresponding to 535 structure-factor phases), some of the secondary structure (*e.g.* α-helices) can already be identified.

(iii) In the electron-density map based on 700 triplet phases (corresponding to 625 structure-factor phases) it is possible to fit fragments of approximately ten amino-acid residues in length to the electron-density map; some characteristic side chains can be identified.

(iv) With 847 triplet phases (corresponding to 752 structure-factor phases) two large fragments can be fitted straightforwardly to almost the entire electron-density map, except in the regions around Trp63, Pro70, Glu71 and Asn74–Asn77, where the electron density is not well defined. This agrees with the reported conformational changes for the residues Pro70–Leu75 (Vaney *et al.*, 1996). The obtained fragments can be linked and lead to a starting model. The program *XFIT* (McRee, 1993) was used to display the ME electron-density maps and to fit the starting model.

In Figs. 2 and 3, the ME electron-density based on 752 structure factors and the structure model from PDB entry 193l are presented. These two figures show different regions of the protein molecule in order to demonstrate the quality of the density map. Fig. 2 presents the region near the centre of the protein molecule and Fig. 3 shows part of an α-helix. Many details can be recognized and the electron density is already well defined around the side chains.

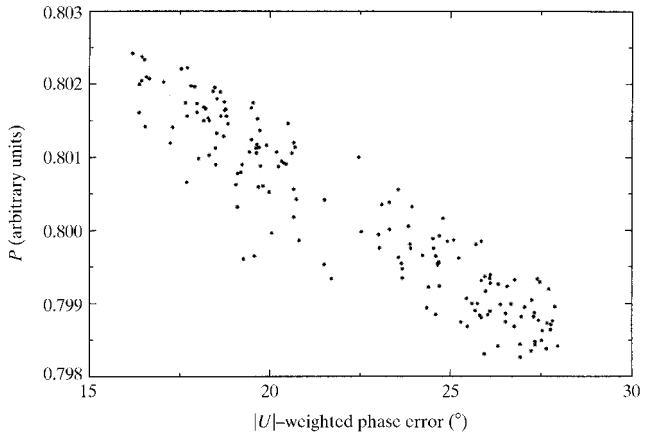

Figure 1
The probability $P \propto \exp(S + L)$ (S, entropy; L, log likelihood) of a set of phase permutations as function of the U_{obs}-weighted phase error (compared with PDB entry 193l; Vaney *et al.*, 1996). The example is calculated for 600 triplet phases involving 544 reflection phases including 18 single-phase semi-invariants. In this case, seven symbolic phases are necessary, which give rise to 164 permutations. Of 544 phases, 495 are independent from symbolic phases. The correlation between P and the phase error is 93%.

4. Search strategy for triplet phases to measure

One of the problems that arises during the course of such an experiment is the following question: given a certain set of phased reflections, which triplet phases are most favourable to measure? As mentioned already, three-beam cases are searched which connect the phases of reflections with the largest $|F|$ values. Among these should be the origin-defining phases. It has been mentioned that ME electron-density maps were

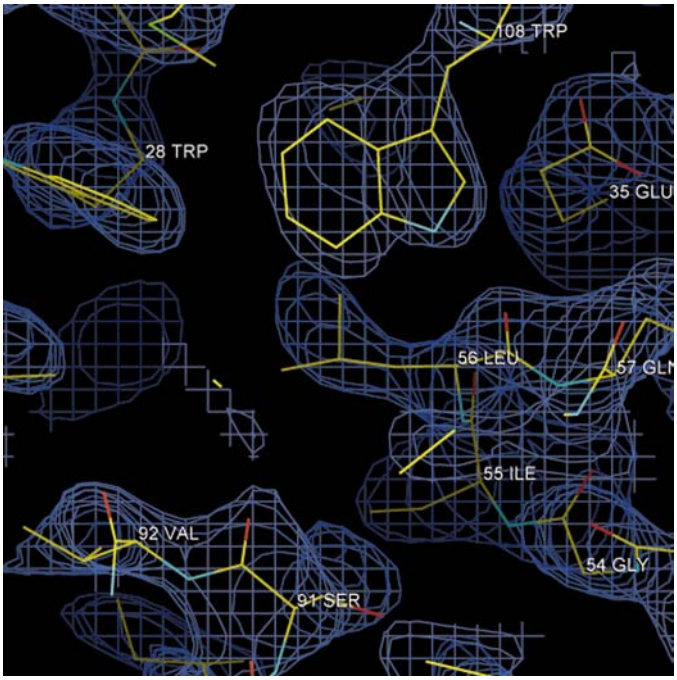

Figure 2
ME electron-density map based on 752 structure-factor phases and the structure model from PDB entry 193l. Region *A*: the center of the protein molecule.

calculated during the triplet-phase data collection in order to study the influence of newly phased reflections. During the development of the method, it was realised that reflections that can be combined with the existing phasing tree and that have a large difference between the measured structure-factor modulus and the one extrapolated from the maximum-entropy density have the most significant influence on the electron-density map. This correlates well with similar results found in applying purely statistical means for phase assignments (Bricogne & Gilmore, 1990).

After the collection of 847 triplet phases, analysis of the phased reflections showed that it is possible to reduce the number of experimentally determined triplet phases and nevertheless calculate an interpretable ME electron-density map. Thus, one can contemplate two different cutoff criteria to be applied: the structure-factor moduli and the resolution limit. Both have been applied to the phased reflections. In each case, the obtained ME electron-density map was examined. As a result, an ME electron-density map based on only 550 phased reflections with structure-factor moduli larger than 360, which corresponds to about one-fifth of the modulus of the largest structure factor, and with resolution lower than 3.5 Å is still interpretable, even though the number of chain breaks increased to 12 and some details were lost. Hence, the strategy of selecting the reflections to be phased can be summarized as follows: (i) they have to connect to the existing phasing tree, (ii) the structure-factor modulus has to show a large difference between the measured one and that extrapolated from the existing maximum-entropy density, (iii) in general, large structure-factor moduli are of advantage and a resolution ($1/|\mathbf{h}|$) lower than 3.5 Å seems to be sufficient. In most cases, the three-beam interference measurements benefit from reflections obeying criterion (iii), because of the large interference effects expected, hence the shorter time required for sufficient counting statistics.

Figs. 4 and 5 show portions of the ME electron-density map based on the 548 phased structure factors with structure-factor moduli larger than 360 and a resolution lower than 3.5 Å and the structure model from PDB entry 193l. Compared with Figs. 2 and 3, some of the details are lost and chain breaks (Fig. 4; side chain Trp108) or additional connections (Fig. 5; Ser91–Ala95) occur. However, the quality of the ME electron-density map still allowed the fitting of the majority of the starting model.

5. Refinement of the starting model

The starting model obtained from the model building was refined using the program *SHELXL* (Sheldrick, 1997). The observed structure-factor file r193lsf (PDB entry 193l; Vaney *et al.*, 1996) was kindly provided by M. C. Vaney. The refinement was carried out against 95% of the measured data. The remaining 5% (1208 reflections) were randomly excluded from the data set and used as a cross-validation test for the free R factor (Brünger, 1992). The side chains were adjusted using $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ electron-density maps displayed with *XFIT* (McRee, 1993). In a later stage of refinement, the positions of water molecules were located using the program *SHELXL* and were checked in the corresponding $|F_o| - |F_c|$ difference map. In the same way, the two conformations of the side chains Asn59 and Val109 were

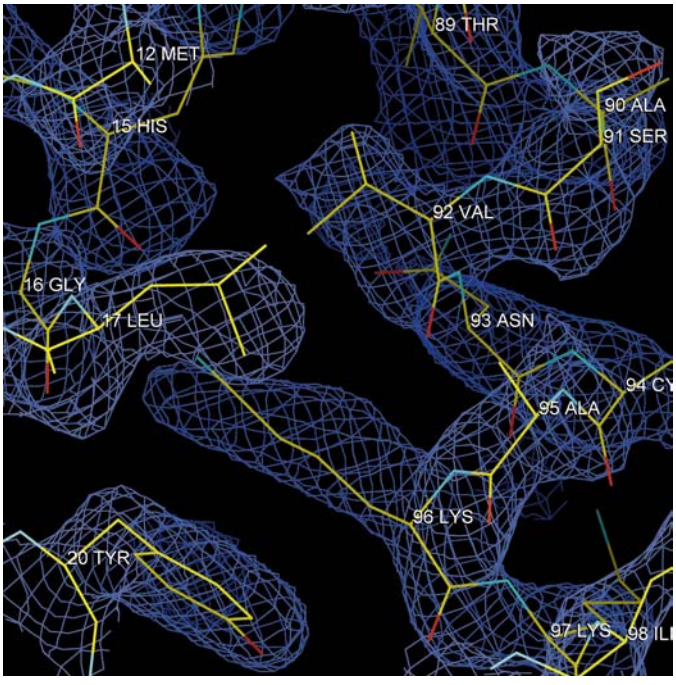

Figure 3
ME electron-density map based on 752 structure-factor phases and the structure model from PDB entry 193l. Region *B*: part of the α-helix.

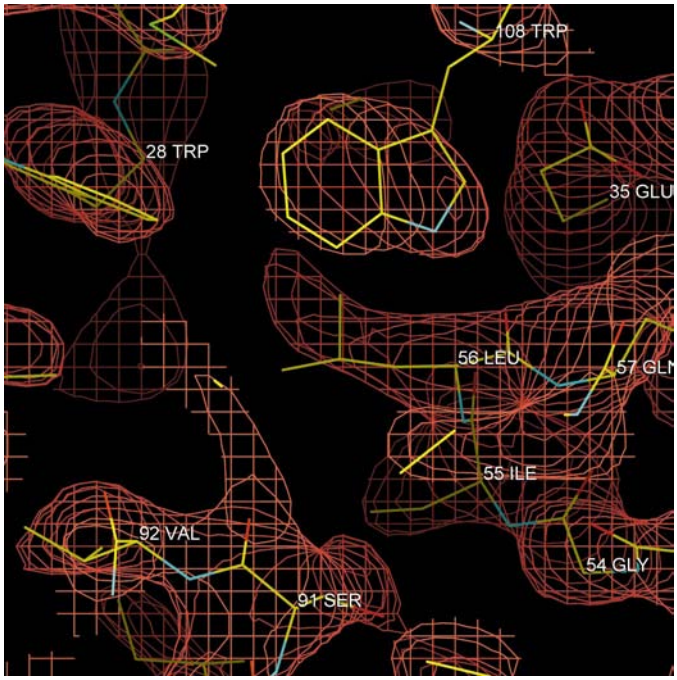

Figure 4
ME electron-density map based on 548 structure-factor phases with structure-factor moduli larger than 360 and a resolution lower than 3.5 Å and the structure model from the PDB entry 193l. Region *A*: compared with Fig. 2, some of the details are lost and a chain break has occurred between the main chain and side chain of Trp108.

determined [R_{free} = 20.1% for $|F_o| > 4\sigma(|F_o|)$ and R_{free} = 21.4% for all data]. During the final steps of refinement, all data were included without further updating of the model. The final R factor was 16.5% for 19 013 reflections with $F_o > 4\sigma(F_o)$ and 17.4% for all data (24 111 reflections).

The final model for the 1001 protein atoms of the HEW lysozyme consists in total of 1007 protein-atom sites (the side chains Asn59 and Val109 were modelled in two conformations), 114 water molecules, one Na^+ atom and one Cl^- atom. Compared with PDB entry 193l, the main chains and nearly all side chains of both models agree, except for the side chains Trp62, Arg73, Asn77, Arg125, Arg128 and Leu129, where the electron density was not well defined. In the starting ME electron-density map, the electron density of these side chains was not defined or was only poorly defined; therefore, large differences between the starting model and the model 193l occurred and have been retained nearly unchanged during refinement. Between the refined model and entry 193l, the r.m.s. deviation of the atomic positions of the residues with similar configuration is 0.12 Å; it is 1.7 Å for those built differently. The residues of both models fall in the same regions of the Ramachandran plot. The observed differences in some side chains also can be observed between various other PDB entries.

6. Discussion and conclusions

This investigation shows clearly that it is possible to derive a structural model from a limited number of phases belonging to reflections with large structure-factor moduli. Using maximum-entropy methods to calculate the density map for HEW lysozyme, about 550 phases proved to be sufficient for a visually interpretable map. In this case, reflections above 3.5 Å resolution are not necessary, since only secondary-structure features are of interest at that stage. A further improvement of an existing electron-density map can be achieved by measuring triplet phases that will determine the phase of reflections with the largest discrepancy between the ME extrapolated and measured moduli. Since a quite high resolution intensity-data set was available, the refined molecule is nearly independent from the starting model in those regions where the molecule is well defined. Differences of the refined model from other available models occur in less well defined regions which are strongly biased by the starting model. It is probably worth investigating the active use of experimental information such as the measured triplet phases in least-squares or maximum-likelihood refinement procedures applied to the protein model. However, this requires isomorphism between the crystal(s) from which the intensity data were taken and from those used to determine the triplet phases.

It might be possible to significantly reduce the number of experimental triplet phases necessary to obtain an interpretable electron-density map by a further development of the statistical methods with respect to the specific nature of the triplet phases that can be measured and/or by the application of an automatic refinement protocol (Perrakis *et al.*, 1997). Systematic investigations of these two possibilities are the subject of present studies.

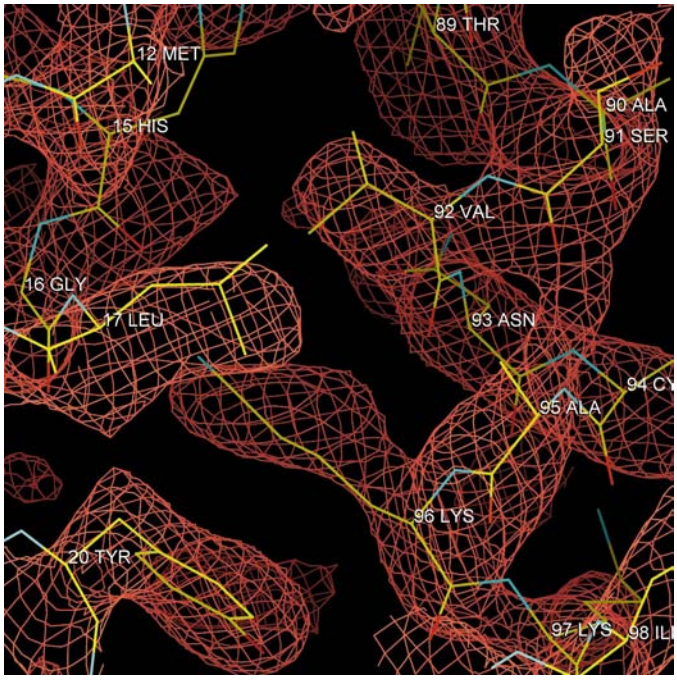

Figure 5
ME electron-density map based on the same phases as in Fig. 4. Region *B*: compared with Fig. 3, an additional connection has occurred between the electron density of Ser91 and Ala95.

We thank M. C. Vaney for making available the structure-factor file r193lsf (PDB entry 193l) prior to publication. We are grateful to Ch. Betzel (Universität Hamburg, Germany) and J. Köpke (MPI für Bio-Physik, Frankfurt, Germany) for advice during the interpretation of the electron-density maps. We thank Robert M. Sweet (Biology Department, BNL, USA) for reading the manuscript. This work has been funded by the German Federal Minister of Education, Science, Research and Technology under contract numbers 05 647VKA and 05 641VKA.

References

Bricogne, G. (1984). *Acta Cryst.* A**40**, 410–415.
Bricogne, G. (1991). *Maximum Entropy in Action*, edited by B. Buck & V. A. Macauley, ch. 8, pp. 187–216. Oxford: Clarendon Press.
Bricogne, G. & Gilmore, C. J. (1990). *Acta Cryst.* A**46**, 284–297.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Chang, S. L., King, H. E., Huang, M. T. & Gao, Y. (1991). *Phys. Rev. Lett.* **67**, 3113–3116.
Giacovazzo, C. (1980). *Direct Methods in Crystallography*. New York: Academic Press.
Hümmer, K., Schwegle, W. & Weckert, E. (1991). *Acta Cryst.* A**47**, 60–62.
Ladd, M. F. C. & Palmer, R. A. (1980). *Structure Determination in X-ray Crystallography*. New York: Plenum Press.
McRee, D. (1993). *Practical Protein Crystallography*. New York: Academic Press.
Main, P. (1977). *Acta Cryst.* A**33**, 750–757.
Main, P. (1978). *Acta Cryst.* A**34**, 31–38.

Mo, F., Mathiesen, R. H., Alzari, P. M., Lescar, J. & Rasmussen, B. (1998). *Materials Structure in Chemistry, Biology, Physics and Technology*, Vol. 5, edited by R. Kuzel, J. Lhotka & L. Dobiasova, p. 484. Praha: Czech and Slovak Crystallographic Association.
Perrakis, A., Sixma, T. K., Wilson, K. & Lamzin, V. S. (1997). *Acta Cryst.* D**53**, 448–455.
Sheldrick, G. M. (1997). *SHELXL97. Crystallographic Refinement Program.* Universität Göttingen, Germany.
Vaney, M. C., Maignan, S., Riès-Kautt, M. & Ducruix, A. (1996). *Acta Cryst.* D**52**, 505–517.
Weckert, E., Hölzer, K., Schroer, K., Zellner, J. & Hümmer, K. (1999). *Acta Cryst.* D**55**, 1320–1328.
Weckert, E. & Hümmer, K. (1997). *Acta Cryst.* A**53**, 108–143.
Weckert, E., Schwegle, W. & Hümmer, K. (1993). *Proc. R. Soc. London Ser. A*, **442**, 33–46.
White, P. S. & Woolfson, M. M. (1975). *Acta Cryst.* A**31**, 53–56.