Acta Cryst. (1993). D**49**, 18–23

The Application of Direct Methods and Patterson Interpretation to High-Resolution Native Protein Data

BY GEORGE M. SHELDRICK

Institut für Anorganische Chemie der Universität Göttingen, Tammannstraße 4, W-3400 Göttingen, Germany

ZBIGNIEW DAUTER AND KEITH S. WILSON

European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, W-2000 Hamburg 52, Germany

HÅKON HOPE

Department of Chemistry, University of California, Davis, Davis, CA 95616, USA

AND LARRY C. SIEKER

Department of Biological Structures, SM-20, University of Washington, Seattle, WA 98195, USA

(*Received* 26 *May* 1992; *accepted* 30 *June* 1992)

Abstract

Conventional small-molecule methods of solving the phase problem from native data alone, without the use of heavy-atom derivatives, known fragment geometries or anomalous dispersion, have been tested on 0.9 Å resolution data for two small proteins: rubredoxin, from *Desulfovibrio vulgaris*, and crambin. The presence of three disulfide bridges in crambin and an FeS_4 unit in rubredoxin enabled automated Patterson interpretation as well as direct methods to be tried. Although both structures were already well established, the known structures were not used in the phasing attempts, except for identifying successful solutions. Direct methods were not successful for crambin, although the correct phases were stable to phase refinement and gave figures of merit clearly superior to any obtained in the *ca* 500 000 random starting phase sets that were refined. It appears that the presence of an iron atom in rubredoxin reduces the scale of the search problem by many orders of magnitude, but at the cost of producing 'over-consistent' phase sets that overemphasize the iron atom and involve partial loss of enantiomorph information. However, about 1% of direct-methods trials were successful for rubredoxin, giving mean phase errors of about 56° (for all $E > 1.2$) that could be reduced to about 20° by standard *E*-Fourier recycling methods. Limiting the resolution of the data degraded the quality of the solutions and suggested that the limiting resolution for routine direct-methods solution of rubredoxin is about 1.2 Å. With the 0.9 Å data, automated Patterson interpretation convincingly finds the three disulfide bridges in crambin and the FeS_4 unit in rubredoxin, and in both cases *E*-Fourier recycling starting from these 'heavier' atoms yields almost the complete structure. Whereas crambin could only be solved in this way at very high resolution, rubredoxin could be solved by Patterson interpretation down to 1.6 Å. These results emphasize the benefits of collecting protein data to the highest possible resolution, and indicate that when a few 'heavier' atoms are present, it may prove possible in favorable cases to solve the phase problem from a single native data set collected to 'atomic resolution'.

Introduction

Direct methods have transformed small-molecule crystallography in the course of the last three decades, but as yet have had only a marginal impact on macromolecular crystallography. This could be because macromolecular data rarely extend to 'atomic resolution', which is often assumed in the derivation of the probability distributions which form the basis of direct methods, or it may be simply a consequence of the size of the structure, which inevitably weakens the probability distributions, at least for individual phase relations. Although overshadowed by the success of direct methods, automated Patterson interpretation to locate the heavier atoms also provides an efficient approach to the solution of small-molecule structures, and could in principle be applied to metalloproteins, proteins containing disulfide bridges, and polynucleotides. In this paper we address the question of whether these 'small-molecule' methods are capable of solving the phase problem directly for macromolecular structures that happen to contain a few heavier atoms, given data to a sufficiently high resolution.

Very few protein data sets have been measured to a resolution of better than 1.2 Å, which might be regarded as a threshold for atomic resolution. It has been suggested (Sheldrick, 1990) that direct methods are unlikely to be successful in the solution of a 'small-molecule' structure if fewer than half of the theoretically measurable reflections in the range 1.1–1.2 Å are 'observed' [*i.e.* have $F > 4\sigma(F)$], although this rule can be relaxed a little for structures containing heavier atoms. Improvements in data collection, for example rapid cooling techniques and area detectors, coupled with the use of synchrotron radiation, now

0907-4449/93/010018-06$06.00

make it possible to collect complete data sets from a single protein crystal before radiation damage becomes significant. This suggests that high-resolution data for macromolecules may become available more often in the near future.

In this paper we have applied small-molecule direct and Patterson methods in the form of the program *SHELXS*92 (Sheldrick, 1990, 1992) to the solution of two known protein structures, rubredoxin (from *Desulfovibrio vulgaris*) and crambin, for which we have collected high-quality X-ray diffraction data to about 0.9 Å resolution. The effect of resolution has been investigated by simply truncating the data to the desired resolution; in practice the data near the resolution limit would be noisier than in these tests. Both proteins happen to crystallize in the space group $P2_1$; there are 46 residues including three disulfide bridges in crambin, and 52 residues including an FeS_4 unit in rubredoxin. Although there are two other sulfur atoms in rubredoxin (in the *N*-terminal methionine and a bisulfate ion) they both exhibit high thermal motion, and so do not function as 'heavier atoms' for the purpose of Patterson interpretation *etc*. Small-molecule crystallographers usually describe the size (and hence difficulty) of structures solved by direct methods in terms of the number of non-hydrogen atoms in the asymmetric unit; the largest unknown equal-atom structure that has been solved to date by direct methods is probably gramicidin A, with 334 unique atoms (Langs, 1988). Proteins contain disordered solvent and side chains, and it is not quite clear how many atoms should be counted; if we count individual atomic sites, crambin would rank as roughly a 500-atom structure and rubredoxin as 600 atoms. Probably it is more realistic to count only fully occupied protein and solvent sites, which would be about 350 for crambin and about 400 for rubredoxin. In fact the factor N in the direct-methods probability formulas is the number of atoms per (primitive) cell, so the space group $P2_1$ is relatively advantageous; similarly it is the number of atoms per cell that dictates the extent of overlap in the Patterson function.

Data preparation

The measurement of the crambin data has been described already (Hope, 1988), as has the refinement of the structure based on these data (Teeter & Hope, 1986). Data were collected at 130 (2) K from a flash-cooled crystal on a Syntex $P2_1$ four-circle diffractometer with Cu $K\alpha$ radiation, graphite monochromator and locally modified Syntex LT-1 low-temperature attachment. A fast ω scan was used with backgrounds based on 200 points distributed over reciprocal space, so that *ca* 32 000 reflections were measured in about 60 h. The resulting data set of 28 727 unique reflections is complete out to 1.1 Å and 92.5% complete out to the limiting resolution of 0.83 Å. In the critical 1.1–1.2 Å range (see above) 81.7% of the theoretically possible data have $F > 4\sigma(F)$.

The rubredoxin data collection and refinement will also be described in detail elsewhere. A wavelength of 0.70 Å was selected at the EMBL beamline X31 at the DESY synchrotron and a MAR-Research imaging plate scanner employed as detector. The short wavelength not only lengthened the life of the crystal but also enabled all the data to be collected with the plate perpendicular to the incident beam, speeding up the data collection. Although the crystal was at room temperature it showed negligible decay during the total data collection time of about 18 h. 74 995 full and 17 712 partial reflections were processed to give 26 237 unique reflections that are 98.5% complete to the limiting resolution of 0.92 Å. These figures refer to the merged data set in which Friedel opposites were also averaged; the corresponding R_{int} was 0.037. In the 1.1–1.2 Å range, 92.4% of the theoretically possible reflections were 'observed' [$F > 4\sigma(F)$], so both protein data sets very comfortably pass the 'atomic resolution' test.

Direct methods

The 'phase-annealing' approach has proved to be one of the more effective approaches for large 'small-molecule' structures, and so was employed here. Only the essential details are given here; for a detailed description of the method see Sheldrick (1990). A large number of sets of random initial phases are refined iteratively, so that each step involves a sum over a large number of phase relations. At no stage is it necessary to rely on the correctness of a single phase relation: the progressive weakening of the probabilities of individual relations as the structure becomes larger is compensated, at least to a first approximation, by the increased number of phase relations involved in each summation. The primary phase refinement formula is:

$$\alpha_{\mathbf{h}} = 2|E_{\mathbf{h}}|\Sigma_{\mathbf{k}}\mathbf{E}_{\mathbf{k}}\mathbf{E}_{\mathbf{h}-\mathbf{k}}/N^{1/2},$$

which is closely related to Sayre's equation (Sayre, 1952) and in some form or other still forms the basis of most small-molecule direct methods; *e.g.* see Karle & Karle (1966) and Germain, Main & Woolfson (1970). We shall consider $\alpha_{\mathbf{h}}$ to be a complex number, *i.e.* it has a phase as well as a magnitude. The 'tangent formula' simply consists of iteratively replacing the phase of $\mathbf{E}_{\mathbf{h}}$ by its expectation value, which is the phase of $\alpha_{\mathbf{h}}$. We can also define a figure of merit based on the agreement of the magnitudes of the left and right sides of this equation:

$$R_{\alpha} = \Sigma_{\mathbf{h}} w_{\mathbf{h}}(\alpha_{\mathbf{h}} - \langle\alpha_{\mathbf{h}}\rangle)^2/\Sigma_{\mathbf{h}} w_{\mathbf{h}}\alpha_{\mathbf{h}}^2,$$

where $w_{\mathbf{h}}$ is a weight and where $\langle\alpha_{\mathbf{h}}\rangle$ is the expected value of α (Karle & Karle, 1966; Cascarano, Giacovazzo & Viterbo, 1987). Weaker but independent phase information is provided by a summation over 'negative quartets' (Schenk, 1974; Hauptman, 1974; Giacovazzo, 1976):

$$\eta_{\mathbf{h}} = 2|E_{\mathbf{h}}|\Sigma_{\mathbf{k},\mathbf{l}}[\mathbf{E}_{\mathbf{k}}\mathbf{E}_{\mathbf{l}}\mathbf{E}_{\mathbf{h}-\mathbf{k}-\mathbf{l}}(2 - |E_{\mathbf{h}-\mathbf{k}}|^2 - |E_{\mathbf{l}-\mathbf{h}}|^2 - |E_{\mathbf{k}+\mathbf{l}}|^2)]/N,$$

which is restricted to contributors in which $|E_{h-k}|$, $|E_{l-h}|$ and $|E_{k+l}|$ are all much less than unity. The phase of η_h should theoretically be shifted by 180° from that of α_h (and $\mathbf{E}_h$), so the figure of merit

$$\text{NQUAL} = \Sigma_h \text{Real}[\alpha_h.\eta_h]/\Sigma_h|\alpha_h.\eta_h|$$

should approach the limiting value of -1 for the correct phases. In practice this value is only reached for very small structures, but since typical false solutions obtained by tangent formula refinement tend to have positive NQUAL values, NQUAL still provides a useful filter for eliminating false solutions.

Unfortunately, iterative application of the tangent formula often tends not to a minimum of R_α, but instead to a 'uranium-atom solution' in which the individual vector terms in the summation for α_h line up, *i.e.* are over-consistent, giving a value of α greater than its expected value. This is a particularly severe problem in symmorphic space groups, *e.g.* $P1$, $C2$ and $R3$. A closely related problem that especially affects polar space groups such as $P2_1$ is the tendency to refine towards a false centrosymmetric solution. We should like to retain the computational efficiency of the tangent formula so as to be able to explore a large number of phase sets efficiently, but without these side effects. In the phase-annealing approach, the following correction is applied to the tangent formula phase φ:

$$|\Delta\varphi| = \cos^{-1}\{[4\alpha/kT + \ln(R)]/[4\alpha/kT - \ln(R)]\},$$

where R is a random number between 0 and 1, and kT describes a thermodynamic analogy in which the phases should achieve a Boltzmann distribution with a fictitious 'temperature' T. The twofold sign ambiguity is resolved by choosing the sign of $\Delta\varphi$ so that the resulting phase of $\mathbf{E}_h$ is closest to that of $-\eta_h$. This approach has the advantage of avoiding over-consistency and also introduces an effective search algorithm. A phase set with a small mean α will make large excursions until a region of higher mean α is reached, and will then tend to be more stable because the phase shifts are smaller. As the 'temperature' T is gradually reduced, the well defined phases (high α) will be subject to smaller fluctuations than those with low α. Thus, phases linked by strong phase relations will tend to become established earlier in the phase determination procedure, while the remaining phases are still free to explore phase space until the later stages.

Direct-methods attempts

For crambin, roughly 500 000 random starting phase sets were refined under a wide range of values for all possible tuning parameters, but no 'correct' solution could be identified. The number of phases refined was varied between 1000 and 2000. Typically, 1500 phases were refined using all possible contributors to the right-hand side of the expression for α, corresponding to an average of 137 terms for each summation. The summations involving negative quartets were restricted to the 12 408 strongest negative quartets out of a total of 78 706 that had been generated. However, when phases calculated from correct positions for the six sulfur atoms were input to the phase refinement, the phase refinement proved stable and gave good figures of merit (R_α = 0.16, NQUAL = -0.31); the best value of R_α refining from random phases was 0.18 for solutions with negative NQUAL. The E maps obtained starting from six sulfur atoms in this way revealed essentially the complete structure. The figures of merit calculated from the correct phases without any phase refinement were R_α = 0.11 and NQUAL = -0.13, and the mean α was 0.90 times its estimated value. The difficulty in solving this structure from random starting phases appears to be primarily one of finding the correct minimum, which may well also be the global minimum, in a multi(1500!)-dimensional phase space that clearly contains a very large number of incorrect local minima.

For rubredoxin a completely different picture emerges. When 1600 phases were refined with an average of 167 terms in each α summation, and the 13 208 strongest out of 152 710 unique negative quartets were employed for η, about 1% of the random starting sets led, after the usual E-Fourier recycling (Sheldrick, 1982), to essentially the complete structure. Although the figures of merit successfully identified the correct solutions (R_α = 0.17, NQUAL = -0.17), there was some evidence of over-consistency; the mean α was about 1.2 times its estimated value. For the true phases without refinement, R_α was 0.11 and NQUAL was -0.18, with a mean value of α 1.19 times its estimated value. As mentioned above, there is some doubt as to how to calculate the value for N for use in the probability formulas in the case of a protein. For these tests we wished to avoid using the known structures to 'fine tune' the direct-methods parameters, so we used the approximately known unit-cell contents (as we would have done for a small molecule) to estimate the effective N, and we used all very low-angle reflections, even though these are often difficult to model in protein refinements.

To clarify the effect of resolution, we have calculated the mean phase error after the phase refinement and after each cycle of E-Fourier recycling at various resolutions; Table 1 gives the values for 0.92 and 1.10 Å. As the resolution becomes worse, the pseudo-mirror plane through the iron atom becomes more pronounced, leading to 'over-consistent' phase sets, and the enantiomorph resolving power of the E-Fourier recycling is greatly impaired. This imposes an effective resolution limit for the successful application of routine direct methods to the solution of rubredoxin of about 1.2 Å. Clearly the 'heavy' atom greatly simplifies the search problem, but there is a price to pay in terms of degradation of the quality of the resulting solution.

We have also made a similar analysis of the 0.98 Å data for avian pancreatic polypeptide, a 36-residue hormone that crystallizes in the polar and symmorphic space group $C2$, and contains one zinc atom in the asymmetric unit

Table 1. *Mean phase errors* (°) *for all* $E > 1.2$ *for rubredoxin based on the full* 0.92 Å *data (first two columns) and data truncated to* 1.10 Å *(last two columns)*

$\Delta\varphi_+$ and $\Delta\varphi_-$ are the mean phase errors for the correct and incorrect enantiomorphs, respectively.

	$\Delta\varphi_+$ (0.92)	$\Delta\varphi_-$ (0.92)	$\Delta\varphi_+$ (1.10)	$\Delta\varphi_-$ (1.10)
Phase refinement	56	68	59	67
E-Fourier recycling, cycle 1	46	72	56	72
E-Fourier recycling, cycle 2	36	74	52	73
E-Fourier recycling, cycle 3	27	77	48	75
E-Fourier recycling, cycle 4	21	78	46	76
E-Fourier recycling, cycle 5	20	79	45	77

(Glover, Haneef, Pitts, Wood, Moss, Tickle & Blundell, 1983). Refinement of 1000 phases with an average of 75 contributors to each α summation, and a total of 5234 negative quartets gave a 2% success rate, with the 'correct' solutions clearly identified by the combination of R_α (*ca* 0.09) and NQUAL (*ca* 0.0, but +0.3 or greater for all wrong solutions with low R_α values). The enantiomorph mixing problem was even more severe than for rubredoxin - there are no sulfur atoms to alleviate it - but again the *E*-Fourier recycling was able to resolve it. The mean phase errors (incorrect enantiomorph in parentheses) were 50° (56°) after phase refinement and 28° (72°) after *E*-Fourier recycling. Avian pancreatic polypeptide can also be solved using the same data with the program *SAYTAN* (Woolfson & Yao, 1990).

Patterson interpretation

For the automated location of heavy atoms in small-molecule structures we have found that a computerized interpretation of the Patterson vector superposition minimum function is very effective. This method was suggested in the early 1950's, and a review of the early literature may be found in Buerger's (1959) book, where it is referred to as the 'vector-shift' method. At the time, hand application of this method did not prove very effective at solving unknown structures, and it was relatively little used until it was revived recently by Richardson & Jacobson (1987), who showed that computer analysis of a single vector superposition could solve quite complex problems. We adopt a similar general strategy to that of Richardson & Jacobson. Since details have been presented elsewhere (Sheldrick, 1992), they will only be summarized briefly. First a sharpened Patterson is calculated using coefficients $(E^3F)^{1/2}$ instead of F^2, and between 1 and 20 suitable peaks selected as superposition vectors. For each of these vectors **u**, a superposition minimum function is calculated by overlaying two copies of the sharpened Patterson that have been displaced from the origin by +**u**/2 and -**u**/2. If a single-weight vector has been chosen for the superposition, this map should theoretically contain only one image of the structure plus its inverse, *i.e.* $2N - 2$ peaks rather than the $N^2 - N$ of the original Patterson. The next stage is the analysis of the peak list to locate potential origin shifts that will move one of the two images so that its constituent atoms conform to the symmetry of the space group; this also enables it to be separated from the other image (that after the move will in general not obey the space-group symmetry). Both this stage and the choice of the original superposition vector generate multiple solutions. Richardson & Jacobson performed the origin search in reciprocal space, but we employ a real-space algorithm. Finally, all solutions are presented in the form of a 'crossword table' that contains one row and one column per potential atom. For each row/column combination two numbers are given: the minimum distance between two atoms and the minimum Patterson density at all vectors between the two atoms, taking symmetry into account in both cases. The most effective figure of merit for comparing the different solutions appears to be the correlation coefficient of Fujinaga & Read (1987).

The automatic selection of 20 superposition vectors longer than 1.8 Å for crambin led to three solutions that showed all three disulfide bridges and two others that contained five of the six sulfur atoms. The correct solutions also had the highest correlation coefficients; Table 2 shows an extract from one of them. In the space group $P2_1$, the column marked 'self' gives the lengths (from the Patterson origin) and Patterson density associated with the Harker vectors $2x$, 0.5, $2z$. There are two independent contributors to each 'cross-vector': $x_1 - x_2$, $y_1 - y_2$, $z_1 - z_2$ and $x_1 + x_2$, $y_1 - y_2 + 0.5$, $z_1 + z_2$. The entries in Table 2 consist of the minimum length and the minimum Patterson density associated with these two vectors. In a higher symmetry space group of course more self- and cross-vectors would contribute to each entry in Table 2.

The computer output shown in abbreviated form in Table 2 reveals the three S—S bonds (2.10, 2.02 and 2.11 Å) that have been marked with asterisks. The minimum Patterson densities involving all six self-vectors and all 15 pairs of different sulfur atoms are high except for one value of 0.0 and one of 1.1; the Patterson densities involving the spurious atoms include several zero values. *E*-Fourier recycling starting from these sulfur-atom coordinates revealed all but 14 of the protein atoms, but it should be noted that crambin has two side chains that are disordered because of sequence inhomogeneity! However, tests showed that even truncating the resolution by a small amount leads to the loss of two or more sulfur atoms in the Patterson interpretation, which would have made the solution of the structure by this method very difficult had it been unknown.

For rubredoxin with the resolution limit set to values in the range 0.9-1.5 Å, the solution with the best correlation coefficient (*ca* 0.25) can be interpreted easily to find the iron atom and the four sulfur atoms to which it is bonded, usually as the top five potential atoms, but not the methionine or bisulfate sulfur atoms. The number of correct solutions, however, decreases with worsening res-

Table 2. *Minimum distances* (Å) *and minimum Patterson densities for potential atoms for crambin*

The atomic numbers Z are estimated from the heights of the symmetry-averaged peaks after applying the origin shift to the vector-superposition map. The first double column gives the self-vectors, the remaining columns and rows correspond to the atoms marked at their intersections. There are three spurious atoms between S5 and S6. Three S—S bonds are indicated by asterisks.

				Self-vectors	Cross-vectors								
Z	x	y	z										
17.1	0.8932	0.0208	0.4592	12.82									
				10.6	S1								
12.8	0.8004	−0.0241	0.0984	18.49	8.90								
				7.3	9.3	S2							
11.9	0.9361	−0.0211	0.4212	11.16	2.10*	9.04							
				7.3	9.5	31.1	S3						
11.7	0.5768	0.0819	0.0477	11.35	15.80	9.38	16.89						
				1.1	7.5	13.0	12.4	S4					
10.9	0.7571	−0.0775	0.1016	18.49	9.84	2.02*	10.21	8.00					
				19.2	16.3	6.5	8.1	15.1	S5				
10.7	0.8975	0.0993	0.2539	16.56	4.81	5.71	4.61	13.82	7.39				
				0.5	6.5	1.7	8.8	0.0	10.0	?			
9.3	0.8161	0.0187	−0.1343	18.49	9.64	5.30	11.14	10.67	6.08	9.37			
				0.0	10.9	0.0	6.8	1.0	4.8	0.0	?		
8.4	1.0450	0.0196	0.6901	13.05	8.01	11.52	7.46	18.32	11.89	8.21	10.16		
				0.9	2.9	0.0	5.4	0.0	0.0	0.0	4.8	?	
8.1	0.6090	0.1713	0.0526	13.02	14.91	8.65	16.00	2.11*	7.66	12.62	9.87	16.49	
				19.0	16.3	8.5	3.7	0.0	5.0	3.6	0.0	0.0	S6

olution. At 1.6 Å there is only one recognizable solution (Fe + 3S) in trials with ten superposition vectors, and it has the fourth best correlation coefficient (0.15). As the data are truncated further, no solutions can be found, even when a good superposition vector (from a successful solution at higher resolution) is used.

The real test of the Patterson-interpretation approach will come when we attempt to solve an unknown metalloprotein or protein containing disulfide bridges. Very recently we had occasion to collect a synchrotron/image plate data set on an octapeptide with one disulfide bridge per molecule and three octapeptide molecules and about 20% water in the asymmetric unit in the space group $P2_12_12_1$. For all intents and purposes the structure behaves like a protein with only 24 residues. This structure had defeated exhaustive attempts at solution using our 150 K diffractometer data that extended to about 1.4 Å. The room-temperature synchrotron data extended to 1.1 Å; the solution from the automated Patterson interpretation with the best correlation coefficient found the six sulfur atoms and E-Fourier recycling revealed most of the rest of the structure (Kallen, Pohl, Sheldrick, Dauter & Wilson, 1992). Direct methods have so far not succeeded in solving the structure from the synchrotron data.

Concluding remarks

Even if data can be collected to 'atomic resolution' for an essentially equal-atom macromolecule, the difficulty of finding the correct solution in multidimensional phase space in the presence of an astronomical number of false local minima may tax the next few generations of computers. It seems that the presence of a heavier atom (such as the iron atom in rubredoxin) may make the search problem tractable at the cost of producing an 'over-consistent' phase set. This presents problems well understood in the small-molecule field that can be overcome if the resolution of the data is adequate.

In the presence of 'heavier' atoms - even disulfide bridges - automated Patterson interpretation may well prove successful, and there are indications that it is not quite so sensitive to the resolution as are direct methods. Methionine sulfur atoms are not nearly as suitable as disulfide bridges because they tend to have higher B values and because there are no useful distances that could be recognized on examining the results of the Patterson interpretation. The phosphorus atoms in polynucleotides probably fall in between these two extremes.

We thank Dr I. Moura at the Centro de Tecnologia Quimica e Biologica, Oeiras, Portugal, and Dr J. Le Gall at the Department of Biochemistry, University of Georgia, Athens, GA, USA, for the sample of rubredoxin, Dr M. M. Teeter, Boston College, Boston, MA, USA, for the sample of crambin, and Dr I. Tickle, Birkbeck College, University of London, England, for the avian pancreatic polypeptide data. GMS thanks the Leibniz Program of the Deutsche Forschungsgemeinschaft for providing an IBM RISC/6000-32H workstation, which was used for most of the calculations.

References

BUERGER, M. J. (1959). *Vector Space*, ch. 11. New York: Wiley.

CASCARANO, G., GIACOVAZZO, C. & VITERBO, D. (1987). *Acta Cryst.* A**43**, 22-29.

FUJINAGA, M. & READ, R. J. (1987). *J. Appl. Cryst.* **20**, 517-521.

GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* B**26**, 274-285.

GIACOVAZZO, C. (1976). *Acta Cryst.* A**32**, 100-104.

GLOVER, I., HANEEF, I., PITTS, J., WOOD, S., MOSS, D., TICKLE, I. & BLUNDELL, T. (1983). *Biopolymers*, **22**, 293-304.

HAUPTMAN, H. (1974). *Acta Cryst.* A**30**, 822-829.

HOPE, H. (1988). *Acta Cryst.* **B44**, 22-26.
KALLEN, J., POHL, E., SHELDRICK, G. M., DAUTER, Z. & WILSON, K. S. (1992). In preparation.
KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849-859.
LANGS, D. (1988). *Science*, **241**, 188-191.
RICHARDSON, J. W. & JACOBSON, R. A. (1987). *Patterson and Pattersons*, edited by J. P. GLUSKER, B. K. PATTERSON & M. ROSSI, pp. 310-317. Oxford Univ. Press.
SAYRE, D. (1952). *Acta Cryst.* **5**, 60-65.
SCHENK, H. (1974). *Acta Cryst.* **A30**, 477-481.
SHELDRICK, G. M. (1982). *Computational Crystallography*, edited by D. SAYRE, pp. 506-514. Oxford: Clarendon Press.
SHELDRICK, G. M. (1990). *Acta Cryst.* **A46**, 467-473.
SHELDRICK, G. M. (1992). *Crystallographic Computing* 5, edited by D. MORAS, A. D. PODJARNY & J. C. THIERRY, pp. 145-157. Oxford Univ. Press.
TEETER, M. M. & HOPE, H. (1986). *Ann. N. Y. Acad. Sci.* pp. 163-165.
WOOLFSON, M. M. & YAO, J.-X. (1990). *Acta Cryst.* **A46**, 409-413.