Table 5. *Coefficient K of the fit between pseudocentric distributions for different types of seminvariants for K*18*JAP*

Seminvariant	Combined K	Non-centrosymmetric triplets	Centrosymmetric triplets	Non-centrosymmetric quartets	$\sum_1$ relation
Sequence no. of the best set	2	2	11	12	20
Refined phases	6·7	3·2	1·6	8·7	3·7
The best set of phases	19·3	8·4	3·5	7·4	7·6
The lowest K	13·0	2·7	0·8	5·1	2·7
The highest K	193·3	161·1	47·8	20·1	24·7

5. Concluding remarks

When DFM's are used as figures of merit for the determination of the best trial set of phases in early stages of the phase determination, the description of the distribution profiles has to be adapted in order to overcome the problems with the bias of the phases in symbolic addition or multisolution methods. Therefore, in the case of symbolic addition, the distributions were calculated in two points only, as in the centrosymmetric case. Naturally, in this way the distribution profiles are neglected and, as a consequence, the discriminating power of DFM's is not fully utilized.

The significantly lower values of the coefficient of the fit for the refined phases, compared with those for pseudocentric solutions, indicate the possibility of obtaining better results by using the method *ab initio, i.e.* when the phases are refined directly by minimization of a criterion based on DFM's as described by Hašek (1985*b*, *c*, *d*) without the preceding step of multisolution or symbolic addition procedures. The main problem of this approach will be to find a sufficiently fast and converging algorithm. Also, theoretical probability distributions have to be used which describe the true distributions of seminvariants more adequately, in particular for quartets and quintets, where the existing formulae (*e.g.* Hauptman, 1975; Giacovazzo, 1976) are not sufficiently exact for the description of the profiles (Peschar 1987; Peschar & Schenk, 1986, 1987).

JH thanks all members of the Laboratory for Crystallography at the University of Amsterdam for their cooperation and Dr K. Huml (IMC) for supporting this research.

References

GIACOVAZZO, C. (1976). *Acta Cryst.* A**32**, 91-99.
GIACOVAZZO, C. (1980). *Direct Methods in Crystallography*, p. 316. New York: Academic Press.
HAŠEK, J. (1974). *Acta Cryst.* A**30**, 576-579.
HAŠEK, J. (1984*a*). *Acta Cryst.* A**40**, 338-340.
HAŠEK, J. (1984*b*). *Acta Cryst.* A**40**, 340-346.
HAŠEK, J. (1984*c*). *Acta Cryst.* A**40**, 346-350.
HAŠEK, J. (1984*d*). *Acta Cryst.* A**40**, 350-352.
HAŠEK, J., HUML, K., SCHAGEN, J. D. & SCHENK, H. (1983). 8th Eur. Crystallogr. Meet., Liege, Belgium. Abstracts, p. 4.04-P.
HAŠEK, J., SCHENK, H., KIERS, C. TH. & SCHAGEN, J. D. (1985). *Acta Cryst.* A**41**, 333-340.
HAUPTMAN, H. (1975). *Acta Cryst.* A**31**, 680-687.
MAIN, P., FISKE, S. J., HULL, S. E., LESSINGER, L., GERMAIN, G., DECLERCQ, J.-P. & WOOLFSON, M. M. (1980). *MULTAN*80. *A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data.* Univs. of York, England, and Louvain, Belgium.
PESCHAR, R. (1980). Unpublished results.
PESCHAR, R. (1987). Thesis. Univ. of Amsterdam, The Netherlands.
PESCHAR, R. & SCHENK, H. (1986). *Acta Cryst.* A**42**, 309-317.
PESCHAR, R. & SCHENK, H. (1987). *Acta Cryst.* A**43**, 84-92.
SCHENK, H. (1983). *Recl Trav. Chim. Pays-Bas*, **102**, 1-8.

Acta Cryst. (1988). A**44**, 485-495

The Isomorphous Pseudo-Derivative Technique for Phase Refinement by Density Modification

BY CH. ZELWER

Centre de Génétique Moléculaire, CNRS, 91190 *Gif-sur-Yvette, France*

(*Received* 24 *December* 1987; *accepted* 26 *February* 1988)

Abstract

Density modification techniques try to improve the phases of poorly resolved electron density maps given by isomorphous replacement by correcting the systematic errors of the maps according to known physical properties. The phases computed from the corrected maps are combined with the observed moduli through a suitable weighting scheme. A new refinement strategy is proposed which considers the observed moduli and the moduli of the Fourier coefficients of the 'best' map as isomorphous pairs, the Fourier transform of the known systematic errors being a 'heavy-atom contribution'. The lack of closure

0108-7673/88/040485-11$03.00

is assumed to have a Gaussian probability law and gives the basis of the weighting scheme as in a true single isomorphous replacement. The application of the technique to the multiple isomorphous replacement map of *E. coli* methionyl-tRNA-synthetase yields dramatic improvements.

Notations and abbreviations

IR	Isomorphous replacement
SIR	Single isomorphous replacement
MIR	Multiple isomorphous replacement
CSF	Classical solvent flattening
IPD	Isomorphous pseudo-derivative
$\rho(\mathbf{x})$	Electron density of the cell
$F(\mathbf{h})$	Fourier transform of $\rho(\mathbf{x})$
$\alpha_{g,q}$	Phase of $F_{g,q}(\mathbf{h})$
F_o	Observed structure factor
F_c	Fourier transform of the modified map
F_{mir}	Fourier transform of the MIR 'best' map
ε	Lack of closure
E	Mean lack of closure
pw	Phasing power

Introduction

MIR electron density maps have given the starting point for atomic model building of most protein structures. However, since the efficiency of isomorphous replacement (IR) decreases as the molecular weight of the asymmetric unit increases, a number of structures need additional techniques to improve the maps obtained from isomorphous replacement, prior to construction of an atomic model. Some molecules may also have a well ordered structure core with less ordered loops at the periphery. This is especially true for proteins which bind RNA; no such structure has at present been entirely solved. For all these reasons, new techniques are needed to optimize contrast and structure determination.

A decisive improvement of IR phases has been possible with phase refinement methods based on the invariance of the electron density through simple mathematical transformations allowed by geometrical redundancies or multiplication by a mask defined by the molecular envelope (Bricogne, 1974, 1976). In his theoretical paper Bricogne (1974) has demonstrated that when more than two copies of the same molecule occur in the asymmetric unit the phases are overdetermined, given the structure-factor moduli at any resolution. However, at least one isomorphous derivative is necessary to find the geometrical relationships between identical molecules in the asymmetric unit and to design molecular envelopes. In Bricogne's (1976) paper, an iteration process consisting in alternate steps of modification of the maps and combination of the phases from the modified maps with the observed moduli is presented. To this end a probabilistic treatment for weighting the phases from these modified maps was proposed, allowing their combination with the starting IR phases.

When the number of equivalent molecules is too small to have an overdetermination or when there is a single copy of the molecule in the asymmetric unit, the density modification can still be used in a statistical sense: although not fixed algebraically, a probability law for each phase may be computed from the modified maps and this probability law may be combined with the IR phase probability law. Schevitz, Podjarmy, Zwick, Hughes & Sigler (1981), using the weighting scheme proposed by Bricogne (1976), showed that drastic improvements of a tRNA map at 4·0 Å resolution from crystals containing 70% solvent are obtained by solvent flattening and by reduction of the densities below the solvent level in the molecular domain. Wang (1985) has used solvent flattening in the same way to solve the phase ambiguity from a single isomorphous replacement. Wang's algorithm determines automatically the molecular envelope from a SIR map and refines it progressively while the phases are improved. Leslie (1987) proposed a fast algorithm to find the molecular envelope, making use of the differences between the mean square deviation in the solvent domain and for the molecule.

Other techniques of phase refinement have been proposed. Rice (1981) used the phases from an atomic partial model including parts for which the sequence had not been assigned to improve the initial MIR phases. In a similar way, Bhat & Blow (1982) selected from a noisy map the features likely to belong to the molecule and considered it as a partially known density. In both cases the combinations of the computed phase probabilities with the initial MIR ones allowed the building of an atomic model.

Although a number of structures have been determined with the help of solvent flattening, it must be emphasized that solvent flattening cannot by itself solve the phase problem. At least one good SIR derivative is necessary to initiate the refinement. In our case, *i.e.* methionyl-tRNA-synthetase (Met-RS), the classical solvent flattening process (CSF) did not enable us to construct the entire molecule (Zelwer, Risler & Brunie, 1982; Brunie, Mellot, Zelwer, Risler, Blanquet & Fayat, 1987). The maps that result from this classical process are less noisy than the MIR maps but also lack continuity. In the areas where the electron density is weaker for physical reasons and hard to distinguish from the solvent, the process does not improve the maps even with a conservative envelope. Furthermore, an automatic procedure may yield a non-conservative envelope and a biased map.

We will show that the Sim phasing and weighting scheme, proposed by Bricogne when geometrical redundancies occur in the asymmetric unit, is not

relevant in the absence of such redundancies. This paper proposes a new way of estimating and weighting the phases obtained from density modification techniques. This new way of handling the phase information has been applied successfully to the structure of Met-RS.

I. Theoretical background

I.1. *The classical solvent flattening method*

Let us call $\rho(\mathbf{x})$ the true electron density and ρ_s the value of the solvent density supposed to be constant. If $\gamma(\mathbf{x})$ is a mask whose value is 1 inside the molecular domain and 0 outside it, we have

$$\rho(\mathbf{x}) \geq 0 \tag{1}$$

and

$$[\rho(\mathbf{x}) - \rho_s]\gamma(\mathbf{x}) = \rho(\mathbf{x}) - \rho_s. \tag{2}$$

In reciprocal space we have

$$\sum F(\mathbf{h})M(\mathbf{k}-\mathbf{h}) = F(\mathbf{h}). \tag{3}$$

$M(\mathbf{h})$ is the Fourier transform (FT) of $\gamma(\mathbf{x})$ and $F(\mathbf{h})$ the FT of $\rho(\mathbf{x})$. $\rho(\mathbf{x})$ and $[\rho(\mathbf{x}) - \rho_s]$ have the same FT, except for $|\mathbf{h}| = 0$.

Given a set of $|F|$'s, the solvent flattening method can be said to find a map whose FT fulfils the linear system (3) and whose moduli are equal to the observed ones. In fact, the M's are weak outside a sphere of $10\,\text{Å}^{-1}$, especially for hand-made envelopes. A convolution equation creates an overdetermination for the phases only if the number of terms of the summation is large enough. This is not the case in macromolecular crystallography and the phases given by (3) have only a statistical meaning. More precisely, the use of (3) in iterations like the classical 'tangent formula', starting with imprecise phases and weighted $|F|$'s, does not increase the number of terms of the summation as in small-molecule crystallography at atomic resolution.

If we now add the condition that the phases must also be compatible with one or several isomorphous derivatives, the constraints are stronger and experience shows that a unique solution can be found for the phases even in the case of a single isomorphous replacement.

To determine the 'best' phases by combining probability laws from independent experiments, it is necessary to give a weighting scheme to each phase solution. The Blow & Crick (1959) treatment of errors from isomorphous replacement gives for each reflection a probability law based on the estimation of the lack of closure $\varepsilon(\alpha)$ between the observed derivative structure factor and the sum of the parent structure factors with the computed contribution of the heavy atoms:

$$p(\alpha) = k \exp[-\varepsilon(\alpha)^2/2E^2] \tag{4}$$

where E^2 is the variance of ε and k a normalization constant.

In order to allow us to combine these probabilities with the information given by partial models, Hendrickson & Lattman (1971) proposed an expansion of the exponential argument of (4) into a Fourier series limited to the first four terms:

$$p(\alpha) = k' \exp(A \cos\alpha + B \sin\alpha + C \cos 2\alpha + D \sin 2\alpha). \tag{5}$$

When two probabilities concerning the same random variable are relative to *independent* experiments, the total probability is given by the product of the probabilities from each experiment. Hence combining the probabilities from different isomorphous derivatives can be done by adding the corresponding coefficients A, B, C and D. When a partial model is known, the probability law is given by the Sim (1959) formula. This probability may also be put in the form of expression (5) with

$$C = D = 0$$

in this special case.

Probability from density modification. In order to find a probability law corresponding to averaging of the densities from identical molecules and to the solvent flattening, Bricogne (1976) makes the assumption that the density values within the molecular envelope are known. Hence the structure factors computed from the modified densities are considered as the FT of the known part of the electron density and the Sim formula can be used to weight the corresponding phases:

$$p(\alpha) = k \exp[X \cos(\alpha - \alpha_H)] \tag{6}$$

with $X = 2|F_o||F_c|/\sum_2$ and α_H the phase of F_c. $\sum_2$ is the mean value of the squared structure-factor moduli of the unknown atoms. In the formula proposed by Bricogne, $\sum_2$ is estimated as the average of $|F_o^2 - |F_c|^2|$.

The assumption that the contents of the molecular envelope are known has been justified in the case of geometrical redundancies by the fact that the contents of the molecular envelope have density values close to the true ones after a small number of refinement cycles. In the absence of such redundancies which overdetermine the phases, this assumption does not make sense.

In the absence of non-crystallographic symmetry, the phase refinement requires IR phase probabilities since solvent flattening cannot by itself determine the phases as seen before. The combination of phase probabilities from different origins is critical here.

Contrary to the usual assumptions, the density of the solvent area is the only one known *a priori*. Conversely, the structure factor of the envelope contents is a random variable which has to be determined even if an 'independent' experiment (isomorphous

replacement) gives a mean value (the Fourier coefficient of the 'best' map). If one tries to apply the process described by Schevitz *et al.* (1981) without smoothing the solvent and without removing the densities below the solvent level, one will find the phases rarely changed but with all figures of merit close to 1. This means that the iteration process is not self-consistent. Another argument to change the weighting scheme of the method proposed by Schevitz and co-workers lies in the fact that the efficiency of solvent flattening is expected to increase with the volume of the solvent area (Leslie, 1987). Improper use of the Sim formula gives weights, however, which increase with the volume of the molecular domain.

In the following sections we propose other formulas to handle the phase information from density modifications.

I.2. *Relationship between the phase shifts and the density modifications*

The initial map from MIR may be considered as the sum of two terms,

$$\rho_{\text{mir}}(\mathbf{x}) = \rho_1(\mathbf{x}) + \zeta(\mathbf{x}), \tag{7}$$

ρ_1 being the corrected density map and ζ the density, which is considered as an error and removed from the map. In reciprocal space, the difference

$$g(\mathbf{h}) = [F_{\text{mir}}(\mathbf{h}) - F_{c1}(\mathbf{h})]$$

is the FT of ζ. The phase shifts due to the density modification therefore depend exclusively on ζ.

More generally we may consider the density ρ_{mir} as the sum of two terms,

$$\rho_{\text{mir}}(\mathbf{x}) = \rho(\mathbf{x}) + \tau(\mathbf{x}). \tag{8}$$

ρ is the true (unknown) electron density and τ an error which is generally unknown except for certain areas of the cell. In these areas we have

$$\tau(\mathbf{x}_s) = \zeta(\mathbf{x}_s).$$

In reciprocal space

$$F_{\text{mir}}(\mathbf{h}) = F(\mathbf{h}) + t(\mathbf{h}). \tag{9}$$

This is an exact relation where $t(\mathbf{h})$ is unknown. If we replace $t(\mathbf{h})$ by $g(\mathbf{h})$, which is known, $F_{\text{mir}}(\mathbf{h})$ has to be replaced by a random variable $\xi(\mathbf{h})$, and we have

$$F(\mathbf{h}) = \xi(\mathbf{h}) - g(\mathbf{h}). \tag{10}$$

Since $g(\mathbf{h})$ is the FT of $\zeta(\mathbf{x})$ which is equal to 0 for $\mathbf{x} \notin \mathbf{x}_s$, the contribution to $\xi(\mathbf{h})$ is given by the structure factor of the atoms and by the FT of the error $\tau(\mathbf{x})$ in the molecular domain. If we assume that the random variable ξ follows the Wilson probability law,

$$p(|\xi|, \alpha)\, \mathrm{d}|\xi|\, \mathrm{d}\alpha = (1/\pi \textstyle\sum_2) \exp[-|\xi|^2/\textstyle\sum_2] |\xi|\, \mathrm{d}|\xi|\, \mathrm{d}\alpha,$$

(the variance $\sum_2$ being computed from $\langle |F_{\text{mir}}|^2 \rangle$), the line of argument developed by Sim (1959) may be applied to our case and we obtain

$$p(\alpha) \propto \exp[X' \cos(\alpha - \alpha_g + \pi)] \tag{11}$$

with $X' = 2|g||F_o|/\langle |F_{\text{mir}}|^2 \rangle$. We may notice that if $F_{\text{mir}}(\mathbf{h}) = 0$ the most probable phase given by (11) is the same as the most probable phase given by (6) but the weight of this phase is in principle much smaller. When the IR figure of merit differs from zero, the most probable phase given by (6) is the phase of the vector, $(F_{\text{mir}} - g)$, whereas the most probable phase given by (11) is the phase of $(-g)$.

Equation (6) tends therefore to strengthen artificially the weight of the MIR phases. Equation (11) has been used to improve the MIR phases of Met-RS (Brunie *et al.*, 1987). The resulting map was better than the one obtained by the classical method in the helical domain and was equivalent or less good in other parts of the molecule. We know now that the scaling between F_o and F_c is critical and should have been changed. This point will be discussed later.

The assumption that $\xi(\mathbf{h})$ follows the Wilson law is convenient but approximate. Luzzati (1955) has shown that since the atoms of protein crystals are not randomly distributed in the cell, the corresponding intensities do not follow a Gaussian law [his relation (22)]. The intensity law proposed from different stochastic models is the product of a Gaussian term with a correcting term which involves the FT of the atomic distribution law (*i.e.* the FT of the molecular envelope). Beyond a certain resolution (10 Å typically), this correcting term is likely to be constant and the Gaussian term gives a good approximation of the probability density of the molecular part of the structure factor. We may also assume that the FT of the error $\tau(\mathbf{x})$ is the sum of a large number of independent random terms and this should keep for $\xi(\mathbf{h})$ the behaviour of a Gaussian two-dimensional random variable.

I.3. *The isomorphous pseudo-derivative technique (IPD)*

Equation (11) does not account for the fact that the F_c computed from the modified maps have to be as close as possible to the observed structure-factor moduli. Relation (10) concerns the moduli as well as the phases. If we try to determine the phase $\alpha(\mathbf{h})$ of $F(\mathbf{h})$ given $|F(\mathbf{h})|$, $|F_{\text{mir}}(\mathbf{h})|$ and $t(\mathbf{h})$ we are in a case similar to the phasing by a single isomorphous derivative. $t(\mathbf{h})$ is only approximately known, as in the early stages of a multiple isomorphous replacement process. We may therefore use $g(\mathbf{h})$ instead of $t(\mathbf{h})$ and apply the statistical treatment of the lack of closure ε, following Blow & Crick (1959):

$$p(\alpha) = k \exp[-\varepsilon(\alpha)^2/2E^2] \tag{12}$$

with

$$\varepsilon(\alpha) = \big|\, |F(\mathbf{h})| \exp(i\alpha) + g(\mathbf{h})| - |F_{\text{mir}}(\mathbf{h})| \,\big|. \tag{13}$$

We know that this treatment assumes that all the errors arise from the moduli of the derivative ($|F_{mir}|$), although in our case the main souce of error comes from $g(\mathbf{h})$. However we expect that, as in the refinement of the heavy-atom sites in the MIR process, the iterations consisting of alternate density modification and phasing steps will allow us to improve $g(\mathbf{h})$. This expectation relies on the fact that the differences between the new map and the IR map in the molecular domain allow an extrapolation of $\zeta(\mathbf{x})$. Fig. 1 gives the Harker construction for our case. Since the IR figure of merit is always smaller than or equal to 1, for many reflexions the circles will have no intersection, and this is particularly true at the beginning of the refinement. In the case of a SIR, the 'best' phase is always the phase of $-g(\mathbf{h})$, but the weight of this phase differs significantly from the one given by (11). The weight given by IPD only depends on the discrepancy between the computed and observed moduli of the 'derivative'. The weight given by (11) increases with the magnitudes $|g(\mathbf{h})|$ and $|F_o|$. When the two circles have an intersection, $p(\alpha)$ becomes bimodal and its combination with the MIR probability density will shift the phase mean value towards one of the two maxima. $p(\alpha)$ has the sharpest shape when the two circles are tangential and when the two radii have the maximum difference (*i.e.* when $|F_{mir}| = 0$). During the iteration process, our knowledge of $\tau(\mathbf{x})$ is improved since we may extrapolate the known errors towards the molecular domain. The IPD technique is therefore able to refine more efficiently the initial IR phases and eventually cure systematic errors due to a lack of isomorphism of one derivative.

For the mean square deviation E of ε we take

$$E^2 = \langle |F_o| - |F_c||^2 \rangle \tag{14}$$

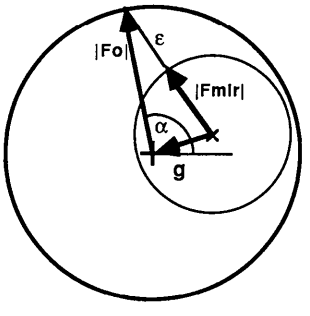


Fig. 1. Argand diagram of the isomorphous pseudo-derivative. The parent circle is given by the heavy line and its radius is $|F_o|$; the pseudo-derivative is given by the light line and its radius is $|F_{mir}|$. g is the 'heavy-atom' contribution computed from the known errors; its origin is the centre of the derivative circle and its end is the centre of the parent circle. For a given α, the lack of closure ε is computed from the difference between $|F_{mir}|$ and $|g + |F_o| \exp(i\alpha)|$.

which is equivalent to

$$E^2 = \langle (|F(\mathbf{h}) + g(\mathbf{h})| - |F_{mir}(\mathbf{h})|)^2 \rangle.$$

We may also define a phasing power,

$$\text{pw} = (\langle |g(\mathbf{h})|^2 \rangle / E^2)^{1/2} \tag{15}$$

and an R factor,

$$R = \langle |F_o - |F_c|| \rangle / \langle |F_o| \rangle. \tag{16}$$

I.4. *Phase extension with IPD*

The use of non-crystallographic symmetry is not limited by resolution contrary to the IR phasing. In the absence of non-crystallographic symmetry, although it is always possible to evaluate the phases by smoothing the solvent area, these evaluated phases are mean values of poor quality. The IPD technique is also able to compute high-resolution phases. The examination of the maps showed an enhanced protein-solvent contrast but without a real improvement of the maps at places where the interpretation was ambiguous. Low-resolution reflexions in the small-angle scattering range may also be phased from modified maps but with more rewarding results. These reflexions are directly related to the contrast between the solvent density and the average protein density. When data from this low-angle range are available, the corresponding phases cannot be estimated from solvent smoothing but from a structure-factor calculation of the contents of the envelope. Indeed Luzzati (1955) demonstrated that in protein crystals the intensities do not follow Wilson's statistics because the distribution of the atoms is non-uniform. Luzzati gave a two-dimensional probability law $p(|F|, \beta) \mathrm{d}|F| \mathrm{d}\beta$ which is very similar to the one proposed by Sim (1959) for partial structures. In this formula, the FT of the probability density of an atom in the cell replaces the structure factor of the partial atomic model in the Sim formula [expression (19), p. 797 of Luzzati (1955)]. The derivation of $p(\beta)$ is straightforward:

$$p(\beta)\, \mathrm{d}\beta = k \exp[2 \textstyle\sum_1 |F_q||F_o| \cos \beta / \sum_2] \tag{17}$$

with $\beta = \alpha - \alpha_q$; F_q is the structure factor of the atomic probability law. The molecular volume defined by the envelope gives us a probability law for the atomic distribution in the cell if we assume that they are uniformly distributed within this volume. Unlike the usual Sim formula, (17) is useless at a resolution higher than 10 Å since the FT of the molecular volume $G(\mathbf{h})$ is weak beyond this resolution (Leslie, 1987). In practice, we may replace $\sum_1 |F_q|$ by the structure-factor moduli computed from the envelope contents. Since the electron density of the solvent ρ_s only modifies the magnitudes of the structure factors in a first approximation, the attribution of computed phases to the low-resolution structure factors will reciprocally determine ρ_s.

II. The practice of the IPD technique

II.1. *The different steps of the iteration process*

Any density modification technique requires prior knowledge of the molecular boundaries. These boundaries may be determined automatically (Wang, 1985; Leslie, 1987) or manually. The result of this determination is a map whose values are 0 in the solvent domain and 1 in the molecular domain. For convenience the file containing this mask should be written in an electron density map format. The cycles of computation including alternate density modification and phasing steps are similar to those described by Schevitz *et al.* (1981): (1) density modification of the MIR map $\rho_0(\mathbf{x})$ or the map $\rho_{n-1}(\mathbf{x})$; (2) fast Fourier transform (FFT) of the modified density to obtain the computed structure factor F_c; (3) phasing step to obtain new weighted Fourier coefficients; and (4) computation of the map $\rho_n(\mathbf{x})$ and return to step 1. We will now describe for each step the specific features of the IPD technique.

II.2. *Density modification*

(*a*) *Positivity of the electron density.* In CSF, the density values below the solvent level are multiplied with an attenuation factor (0·1 typically). For the IPD technique this treatment is forbidden. Indeed, in our case modification of the map is used to evaluate the error $\tau(\mathbf{x})$ rather than to define the 'known' part of the structure. It is especially important that features of the true electron density $\rho(\mathbf{x})$ should not be considered as belonging to $\tau(\mathbf{x})$. For this reason, identification of negative densities implies that the map should be computed on an absolute scale and the F_{000} term included. The determination of the scale and of the constant F_{000} makes use of the fact that the low-resolution reflection magnitudes (between ∞ and 20 Å) are sensitive to the contrast between the average density in the molecular domain and ρ_s. We also know that the average electron density of a hydrated protein is about 0·41 e Å^{-3}. The condition that the R factor for the low-resolution reflexions be as low as possible and that the average density must be close to its theoretical value in the molecular domain allows us to find suitable values for F_{000} and the scale. Finding these values is equivalent to identifying the negative densities and the maximum contrast allowed between molecular and solvent average densities. In practice, at the beginning of the refinement the low-resolution reflexions have a figure of merit of 0 and the scale is arbitrary. The F_{000}/V term is adjusted so as to put the average density of the molecular domain close to 0·41 e Å^{-3} and added to the MIR map. If the scale factor is underestimated, the amount of removed 'negative' densities will be excessive and the contrast between protein and solvent overestimated. This overestimate produces in the following cycles a strong discrepancy between $|F_o|$ and $|F_c|$ in the low-angle scattering range. A new scale factor can then be computed from these low-angle reflexions which reduces the amount of negative density. The scale may therefore be adjusted progressively. Another way of computing the absolute scale can be found by using the FT of the molecular envelope as F_c, the scale of the F's being proportional to the difference $(0{\cdot}41 - \rho_s)$.

(*b*) *Solvent smoothing.* The mask corresponding to the molecular volume is defined by a file of logical values arranged like a map file. The first envelope has been designed by contouring the areas of density likely to contain meaningful features. After several cycles of iteration the envelope may be adjusted to the improved map automatically. The result of this adjustment is a volume whose FT has non-zero values at a higher resolution. Since the multiplication of the map by the mask is equivalent in reciprocal space to convoluting the F's with the Fourier coefficients of the mask, the higher the resolution of the mask, the stronger the constraints on the phases.

II.3. *Phasing step*

(*a*) *Phase probabilities.* From Fig. 1, the probability for a given α value is given by (12) and (13). This probability density is put in the form proposed by Hendrickson & Lattman (1970) to allow its combination with the MIR probability density. As seen before, the low-resolution reflexions (in the range ∞–8 Å) are phased according to (6) if the MIR figure of merit is smaller than a preset value (0·3 typically).

(*b*) *Estimation of the discrepancy between the MIR best map and the electron density.* The structure factors $g(\mathbf{h})$ of the known errors are given by

$$g(\mathbf{h}) = F_{\text{mir}}(\mathbf{h}) - F_c(\mathbf{h}). \qquad (18)$$

This means that if in the first cycle the errors are only estimated from the solvent domain and from the negative values of the scaled MIR map, in the following cycles $g(\mathbf{h})$ also includes the improvements of the map in the molecular domain which result from phase combinations from the preceding cycles. Thus, equation (18) allows an extrapolation of the errors towards the areas where they are not measured. Of course, the extension of the knowledge of the errors towards the molecular domain does not have the same reliability as the estimation of these errors from the departure of the map from known physical properties. The only control we have on these estimations is the minimization of the mean lack of closure E during the refinement process, or of the R factor between F_c and F_o. We may, however, note that this situation does not differ significantly from the one occurring when one attempts to interpret a heavy-atom deriva-

tive starting from a main site and looking up difference maps to include the weaker sites in the refinements. The analogy between the IPD technique and the heavy-atom refinement in isomorphous replacement is complete.

(*c*) *Low-resolution reflexions.* As stated before (§ I.4), the low-resolution reflexions (in the range ∞–8 Å), which have a figure of merit below a preset value, are phased according to the Sim formula (6).

(*d*) *Scaling between* $|F_o|$ *and* $|F_c|$. In current practice (Bricogne, 1976; Schevitz *et al.*, 1981; Wang, 1985) the scaling of the structure factors computed from modified maps is done assuming that they are similar to structure factors computed from a partial atomic model. According to this assumption, $\langle|F_c|\rangle$ would have to be close to $\langle|F_o|\rangle$ at any resolution and a temperature factor would have to compensate for the decay of the mean figure of merit with the resolution. By contrast, for the IPD technique the MIR structure factors and the observed structure factors are considered as isomorphous pairs and the difference $(F_{\rm mir} - F_c)$ represents the 'heavy-atom' contribution. The scaling between F_c and F_o must therefore only include reflexions whose figures of merit are close to 1. Our experience has shown that the usual scaling results in an underestimate of the mean square error E and an overestimate of $g(\mathbf{h})$. The maps computed with the weighted structure factors resulting from this usual scaling associated with the IPD technique are damaged in the areas where the density is especially weak. Makowski (1986) emphasized the need to take into account the unphased reflexions (with a figure of merit of 0) for the estimation of the errors. This requirement is in agreement with our experience.

II.4. *Implementation*

The set of programs written in Fortran 77 has been implemented on the IBM 3090 computer of CIRCE (Orsay) under the MVS environment. At present the use of the programs is restricted to space groups whose centrosymmetric reflexions have either 0 or π as phase values. The different steps of a single cycle can be run as separate and successive jobs. The FT subroutines are those written by Ten Eyck (1973), all the map files being compatible with this set of programs. The reflexion files are written in a '12A2' format, one item containing the Miller indices, a flag for centrosymmetric reflexions, the $|F_o|$, the four Hendrickson & Lattman coefficients, the figure of merit and the phase. The reflexion file used to compute the electron density map to be modified contains all measured reflexions, the phase coefficients (A, B, C, D) and the figure of merit (set to zero if no IR phase information is available). There are three possible items for each reflexion; the first one corresponds to the IR information, the second one to the information resulting from the previous cycle, and the third one to the FT of the modified map arranged in the 12A2 format. Since at the first cycle there are two identical items for each reflexion, the second item contains either the phase information from the phase combination program or the IR phase information if the reflexion has not been treated by the combination program. This file organization allows us to improve the phases by shells of increasing resolution. The structure factors F_c of the modified maps are put in the same format, sorted and merged with the initial file, by means of the IBM sort/merge package.

An additional program is used to update the envelope from an initial mask map. This program requires the prior definition of a minimum density level and of a number of sections of the electron density map to be averaged. The averaging assigns a flag to each grid point, taking into account the values of the neighbouring sections for the same grid point. If $\rho > \rho_m$ for any of the set of sections, the flag is set to (1). Otherwise the flag is set to (0) outside the previous envelope and (−1) inside it. The boundaries of the molecule are shifted so as to include in the molecular volume the maximum number of (1) pixels and to avoid confusion of van der Waals interspacing with solvent. The averaging of several sections is necessary to separate empty areas in the molecular domain from solvent areas. The refined envelope follows the molecular contours more accurately, thus leading to stronger constraints on the computed phases [*cf.* relation (3)].

This version of the programs has been mainly used to set the method. In order to have a more general and transportable set of programs, we are modifying the *SCALE* and *COMBINE* steps of the Wang algorithm to make them compatible with the IPD phasing. This work is being done on our local microVAX computer. A further advantage of including IPD phasing in existing programs lies in the automatic procedures to design the envelopes.

III. Application of the IPD technique to the Met-RS structure

The crystallized methionyl-tRNA-synthetase is a molecule of 64 000 daltons obtained by mild proteolysis of the native dimer ($2\times 76\,000$; Waller, Risler, Monteilhet & Zelwer, 1971; Dardel, Fayat & Blanquet, 1984). The space group is $P2_1$ with one molecule per asymmetric unit. The first model was obtained from multiple isomorphous replacement followed by a solvent flattening process according to Schevitz *et al.* (1981). The map allowed the description of the topology of the N-terminal domain consisting of a Rossmann fold with a large inclusion between strands C and D (Risler, Zelwer & Brunie, 1981; Zelwer *et*

Table 1. *Comparison between similar significant test values for the initial MIR refinement, the usual refinement with the use of the Sim formula and the IPD refinement*

	Initial	Classical	IPD
R factor	46%	26%	21%
$\langle m \rangle$	0·63	0·89	0·89
Number of reflexions	16 000	17 000	21 000
pw	—	—	2·89
r ratio	0·89	0·46	0·46

al., 1982). The C-terminal domain consists of several α helices, the connexions of which were not given as certain. Areas of ill defined densities were present in the map and it was not clear whether these areas belonged to the C terminus or were parts of the inclusions between strands of the Rossmann fold. The tentative model published before the sequence was available did not account for the totality of the molecule. Cross-linking experiments done with oxidized tRNA suggested that 55 residues had to be included between strands *D* and *E* of the Rossmann fold in the N-terminal domain (Hontoudji, Blanquet & Lederer, 1985). Following this information a reinterpretation of the maps has been undertaken. The present available structure appears compatible now with all the independently obtained biochemical and genetic data (Brunie *et al.*, 1987).

III.1. *The application of the IPD technique*

The data consists of 21 000 observed moduli representing a complete 2·5 Å sphere. These reflexions have been recorded with a rotation camera except for the low-angle reflexions recorded with a diffractometer. 16 000 reflexions only have been phased by MIR. The 5000 additional ones include 3000 reflexions excluded from MIR because they were too weak on our first photographs, and 2000 reflexions from the cusp added later.

The initial envelope was drawn by hand and was the same as the one used for the classical solvent flattening technique. After a preliminary set of alternate steps of density modification and phase combination, the map looked better than with the classical solvent flattening technique and the envelope was refined using the program described above. The whole process was then resumed from the beginning with the new envelope for 14 cycles. The molecular domain represented 58% of the cell, the initial R factor was 46%. The initial resolution limit was 6 Å and this limit was extended progressively during the 14 cycles. Table 1 gives the values of some significant variables used to follow the refinement: the mean phasing power of the IPD, the mean figure of merit, the R factor, and the ratio r given by

$$r = \langle \rho - \langle \rho \rangle \rangle^2_{\text{solvent}} / \langle \rho - \langle \rho \rangle \rangle^2_{\text{cell}}. \qquad (19)$$

Further information on the refinement is given in Tables 2 and 3.

The initial scale of the F's and the F_{000} term put the zero level of the map at a midpoint between the mean level in the molecular domain and the minimum of the density. At cycle 14 the mean density was 0·43 e Å^{-3} for the molecule and 0·32 e Å^{-3} for the solvent. These values are unrealistic and the figure of merit of the first 30 reflexions between ∞ and 20 Å was bad (0·63), reflecting a strong discrepancy between $|F_o|$ and $|F_c|$. Dividing this initial scale by 2·2 led after four refinement cycles to a figure of merit of 0·70 for these reflexions and a mean difference of 0·03 e Å^{-3} between the molecule and the solvent areas. The corresponding displacement of the zero level of the map (the negative densities representing only 0·25 of the densities below the solvent level) resulted in a weaker contribution of negative densities to $g(\mathbf{h})$. We may wonder whether the low-resolution reflexion moduli are reliable enough to adjust the scale since their measurements are affected by strong Lorentz and polarization corrections as well as by strong absorption factors. In fact, the interpretation of the resulting maps showed that with the previous scale some parts of the molecule surrounded by the solvent are easier to distinguish from the solvent than with the second one although few sharp irregularities

Table 2. *Statistics of the IPD refinement per resolution shell*

N_{Sim} is the number of reflexions for which the MIR phase has been replaced by the one given by the Sim formula (6). The phase shift $\Delta\alpha$ has been taken with respect to the initial MIR phase.

Resolution (Å)	25	12·5	8·3	6·2	5·0	4·2	3·6	3·1	2·8	2·5	Total
Number of reflexions	30	178	446	814	1295	1931	2654	3535	4517	5619	21 019
N_{Sim}	30	91	164	22	0	0	0	0	0	0	307
$\langle m \rangle$	0·70	0·90	0·94	0·90	0·90	0·93	0·91	0·90	0·89	0·88	0·90
pw	1·5	2·1	3·2	2·5	2·6	3·3	3·2	3·2	3·0	3·1	2·9
$\langle \Delta\alpha \rangle$ (°)	63	45	50	43	45	46	51	57	62	72	59

Table 3. *Repartition of the reflexions per figure of merit and corresponding mean phase shift for the final IPD cycle*

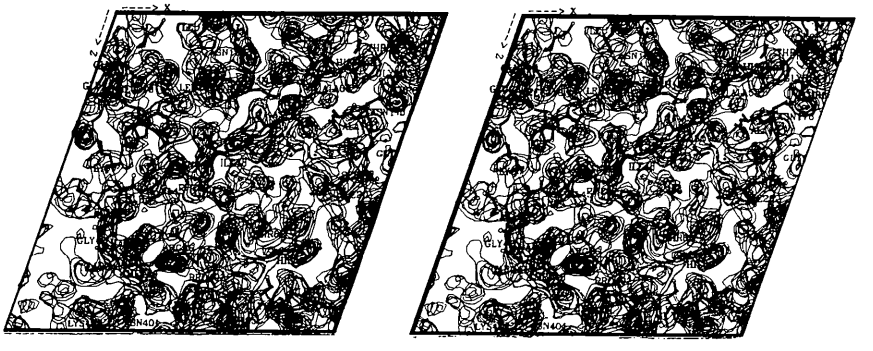
m less than	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
MIR	7334	488	647	871	966	1130	1443	1948	2924	3268
IPD	60	76	104	144	199	345	704	1276	3371	14 740
$\langle \Delta\alpha \rangle$ (°)	89	78	82	76	74	69	58	43	26	10

appear in the main-chain path. At cycle 18 the R factor was 21%, the overall figure of merit was 0·89 for the 21 000 reflexions lying in the 2·5 Å sphere, and the ratio r was 0·47 against 0·88 at the beginning of the refinement. The final phasing power of the IPD was 2·9.

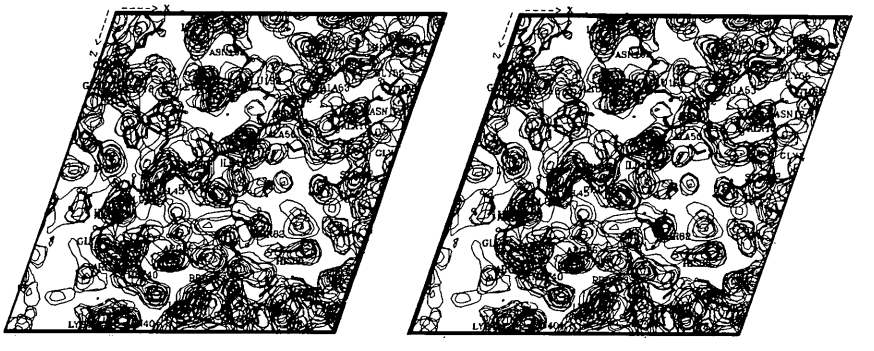
III.2. *Comparison of the maps from different methods*

Figs. 2 and 3 give comparisons between the initial MIR map (*a*), the CSF map (*c*) and the IPD map (*b*) for some large areas of the molecule. The MIR map is obviously noisy and lacks resolution. Map (*c*)

(*a*)

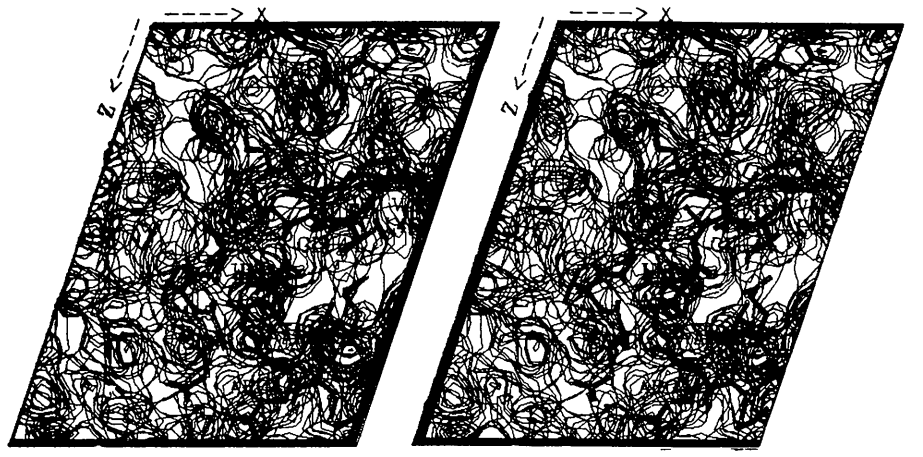


(*b*)

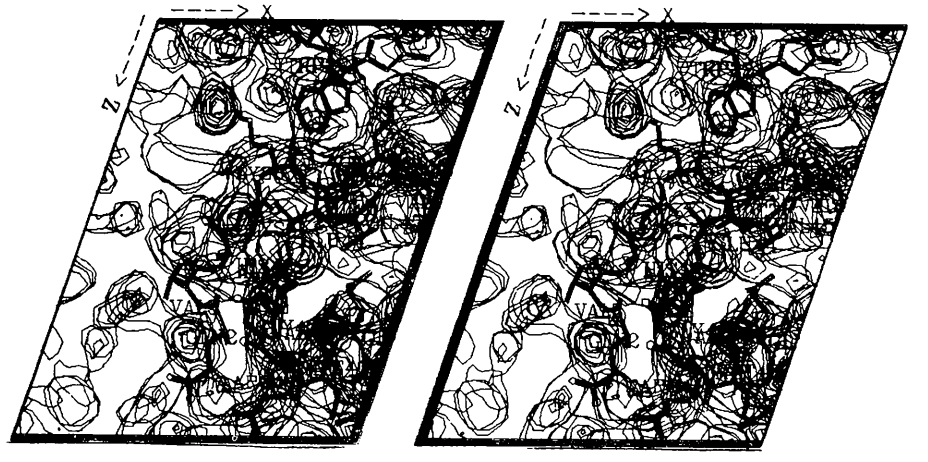


(*c*)

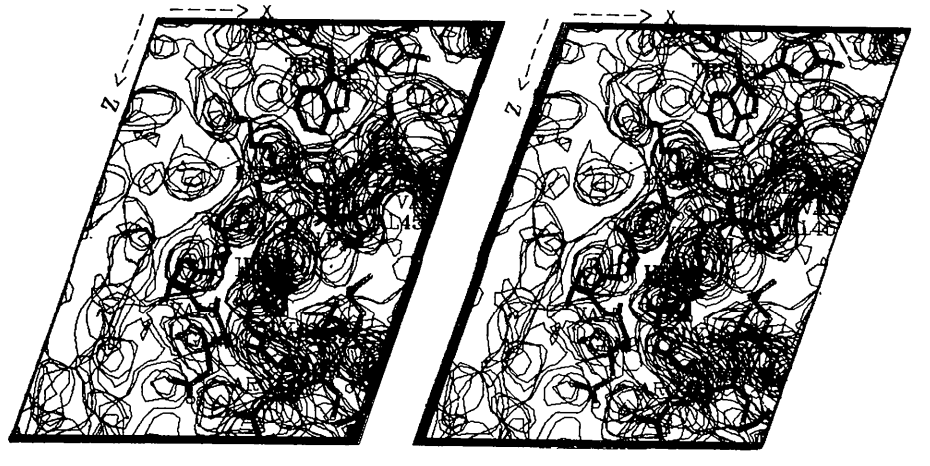
Fig. 2. Stereo pairs of superimposed sections of the map for (*a*) MIR, (*b*) IPD and (*c*) CSF of an area exhibiting strand B of the β sheet of the nucleotide binding domain; map (*c*) exhibits several sharp irregularities in the main-chain path which are not present in (*a*) and (*b*). Several side chains are better defined in map (*b*) than in map (*c*). Trp 221 belongs to a peptide which has been reinterpreted during the refinement of the atomic model. Different scales have been applied to each map so as to ensure that the density values vary in the same range. The superimposition of the density contours and of the atomic skeleton (corresponding to an R factor of 22%) has been done with the program *PLUTO* (Motherwell & Clegg, 1978).



(*a*)



(*b*)



(*c*)

Fig. 3. Stereo pairs of superimposed sections of the map for (*a*) MIR, (*b*) IPD and (*c*) CSF of an area exhibiting the loop connecting the first α helix to strand *B* and a part of the N-terminal peptide. The first residues of the N terminus are surrounded by the solvent and have a weak density. Maps (*b*) and (*c*) are better resolved for the N-terminal peptide.

has a better contrast but numerous sharp irregularities which are not present in the initial MIR map are observed in the main-chain path. In map (*b*) the side chains are better defined than in the other ones and no sharp irregularities occur in the main-chain path for areas which are not in close contact with the molecular envelope.

III.3. *Interpretation of a mini-map*

A mini-map at a scale of 2·4 mm $Å^{-1}$ has been traced on transparent sheets. This IPD mini-map is good enough to define a peptide chain for most of the structure, the density for the peptides 125–185 and 235–285 being improved although the resolution is still lower than in the other parts of the structure.

Further comparison between the IPD map and the refined model will be detailed when the structure is published in the near future.

IV. Concluding remarks

The IPD technique refines the phases while constraining the map to follow some physical properties. The calculated moduli are constrained to be as close as possible to the observed ones. In this sense IPD seems to be more powerful in refining the phases than the classical refinement techniques based on the Sim formula. This technique is not restricted to solvent flattening and may be used with non-crystallographic symmetry. The IPD technique makes a clear separation between the starting information and the unknown density, which is that of the molecular domain. The Blow & Crick weighting scheme has the additional advantage of being simple and cheap in computing time. Obviously the distribution law of the lack of closure, assuming that the errors are coming from the 'heavy-atom derivative' structure-factor moduli (here the $|F_{mir}|$), is not correct since the errors arise in our case from $g(\mathbf{h})$. It is not certain, however, that modification of the weighting scheme will significantly improve the resulting maps. Our experience of the IPD technique has taught us that attempts to minimize the mean square errors artificially or to increase the molecule/solvent contrast artificially by removing the densities below the solvent level may damage the maps in an unpredictable way.

It is in principle possible to evaluate the phases at any resolution with the IPD technique, as with the other density modification methods. In the Met-RS case, 16 000 reflexions only were phased by isomorphous replacement while we kept the 5000 additional ones to obtain a complete 2·5 Å sphere. It is obvious that solvent flattening is equivalent to a single derivative; however, the phases estimated from solvent flattening alone lack precision. Our experience of extending the phases below the limit where the heavy-atom derivatives work shows that the interpretation of the resulting maps is not easier than at 2·5 Å resolution.

The structure determination of the tryptic fragment of Met-RS was well advanced when the IPD mini-map was traced. It is now necessary to check the method on structures which have not been previously modelled at an atomic level. Since the IPD technique may easily be included in existing algorithms, its efficiency in refining the phases can be verified in the near future on various new examples.

I am indebted to Dr V. Luzzati for critical and helpful discussions during the accomplishment of this work. The coordinates of the Met-RS partial model were kindly supplied by Dr S. Brunie (Laboratoire de Biochimie, Ecole Polytechnique, CNRS, Palaiseau) at an intermediate stage of the refinement. I acknowledge Dr L. Sperling for correcting the English version of the manuscript. The phase refinement was carried out at the Centre Interdisciplinaire de Calcul Electronique (CIRCE) in Orsay and the stereo maps with the superimposed atomic model have been computed with the VAX 780 from the Laboratoire pour l'Utilisation du Rayonnement Electromagnétique (LURE, Orsay).

References

BHAT, T. N. & BLOW, D. M. (1982). *Acta Cryst.* A**38**, 22-29.
BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794-802.
BRICOGNE, G. (1974). *Acta Cryst.* A**30**, 395-405.
BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832-847.
BRUNIE, S., MELLOT, P., ZELWER, C., RISLER, J. L., BLANQUET, S. & FAYAT, G. (1987). *J. Mol. Graphics*, **5**, 18-21.
DARDEL, F., FAYAT, G. & BLANQUET, S. (1984). *J. Bacteriol.* **160**, 1115-1122.
HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B **26**, 136-143.
HONTOUDJI, C., BLANQUET, S. & LEDERER, F. (1985). *Biochemistry*, **24**, 1175-1180.
LESLIE, A. G. W. (1987). *Acta Cryst.* A**43**, 134-136.
LUZZATI, V. (1955). *Acta Cryst.* **8**, 795-806.
MAKOWSKI, L. (1986). *Acta Cryst.* A**42**, 253-256.
MOTHERWELL, W. D. S. & CLEGG, W. (1978). *PLUTO.* Program for plotting molecular and crystal structures. Univ. of Cambridge, England.
RICE, D. W. (1981). *Acta Cryst.* A**37**, 491-500.
RISLER, J. L., ZELWER, C. & BRUNIE, S. (1981). *Nature (London)*, **292**, 383-384.
SCHEVITZ, R. W., PODJARNY, A. D., ZWICK, M., HUGHES, J. J. & SIGLER, P. B. (1981). *Acta Cryst.* A**37**, 669-677.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813-815.
TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 183-191.
WALLER, J. P., RISLER, J. L., MONTEILHET, C. & ZELWER, C. (1971). *FEBS Lett.* **16**, 186-188.
WANG, B. C. (1985). In *Methods in Enzymology*, Vol. 115: *Diffraction Methods for Biological Macromolecules*, edited by H. WYCKOFF, C. H. W. HIRS & S. N. TIMASHEFF. New York: Academic Press.
ZELWER, C., RISLER, J. L. & BRUNIE, S. (1982). *J. Mol. Biol.* **155**, 63-81.