Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Improved Fourier coefficients for maps using phases from partial structures with errors

R. J. Read

Acta Cryst. (1986). A**42**, 140–149

Acta Cryst. (1986). A**42**, 140-149

Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors

BY RANDY J. READ*

Medical Research Council of Canada Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7

(*Received* 12 *June* 1985; *accepted* 16 *October* 1985)

Abstract

Unrefined or partially refined models of macromolecules are generally incomplete and typically have large coordinate errors. It is shown that phase probability equations appropriate for a perfect partial structure lead to inaccurate estimates of phase probabilities in such cases. Therefore, it is necessary to use equations that have been derived allowing for errors in the partial structure. A method is given to estimate the parameter σ_A in these phase probability expressions from the observed and calculated structure factor amplitudes. From the variation of σ_A with resolution, one can estimate the mean coordinate error for the model. Electron density maps calculated using partial structure phases are biased towards the partial structure. When there are coordinate errors, a new expression for the non-centric Fourier coefficients $[(2m|\mathbf{F}_N| - D|\mathbf{F}_P^c|)\exp(i\alpha_P^c)]$ is required to suppress this model bias. Judged by correlation coefficients comparing electron density maps with the correct and the partial structure maps, the Fourier coefficients derived here are superior to others currently in use.

1. Introduction

Accurate phase probabilities are important for combining independent sources of phase information (Rossmann & Blow, 1961; Hendrickson & Lattman, 1970) or for calculating probability-weighted electron density maps (Blow & Crick, 1959). Woolfson (1956) and Sim (1959, 1960) derived expressions for phase probabilities from partial structures for centric and non-centric structure factors respectively. Srinivasan and co-workers extended this work to include coordinate errors in the partial structure (Srinivasan & Ramachandran, 1965; Srinivasan, 1966).

Two problems arise when one attempts to apply the results on phase probabilities to the calculation of electron density maps using partial structure phases. The first problem is that of estimating either Σ_Q or σ_A from the observed and calculated structure factor amplitudes, in order to obtain accurate phase probabilities. The parameter Σ_Q, which measures the amount of missing scattering matter, is used in the expressions of Woolfson (1956) and Sim (1959, 1960); σ_A, which is a combined measure of the completeness and the accuracy of the partial structure, is required for the expressions of Srinivasan (1966). (These terms and others are defined in Table 1.) § 2 deals with the estimation of these parameters and with the evaluation of thier associated phase probability distributions. The second problem, which is treated in § 3, is that of minimizing the bias towards the model in a model-phased electron density map.

2. Estimating phase probabilities for partial structures

2.1. *Estimating* Σ_Q

It has been common to assume that the probability expressions for partial structures with no coordinate errors provide a reasonable approximation for the case with errors. These expressions require an estimate for Σ_Q, modified perhaps to include a contribution from errors (Rossmann & Blow, 1961). Because of thermal motion and the finite size of atoms, Σ_Q is a function of resolution and is generally estimated for several resolution ranges. Blundell & Johnson (1976, p. 418) suggest the mean square deviation between the structure factor amplitudes for the complete and partial structures, $\overline{(|\mathbf{F}_N| - |\mathbf{F}_P|)^2}$, as an empirical estimate. However, when $\mathbf{F}_Q$, the missing-atoms structure factor, is small compared to and independent in direction of $\mathbf{F}_P$, $\overline{(|\mathbf{F}_N| - |\mathbf{F}_P|)^2}$ is a measure of the variance of each component of $\mathbf{F}_Q$, the components in phase and out of phase with $\mathbf{F}_P$ (Henderson & Moffat, 1971). Therefore, a better estimate is given by

$$\Sigma_Q = \langle|\mathbf{F}_Q|^2/\varepsilon\rangle \simeq \overline{n(|\mathbf{F}_N| - |\mathbf{F}_P|)^2/\varepsilon}, \qquad (1)$$

where $n = 2$ for non-centric reflections and 1 for centric reflections, which have no out-of-phase component for $\mathbf{F}_Q$. The factor ε corrects for the difference in expected intensity for different reciprocal-lattice zones. Bricogne (1976) suggests

$$\Sigma_Q \simeq \overline{||\mathbf{F}_N|^2 - |\mathbf{F}_P|^2|}.$$

* Present address: Laboratory of Chemical Physics, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands.

0108-7673/86/030140-10$01.50

Table 1. *Definitions of terms and notation*

Some terms are not given explicitly, but are analogous to terms defined here.

$\bar{x}$ = mean value of x

$\langle x \rangle$ = expected value, or probability weighted average, of x

$\mathbf{F}_N = \sum_{j=1}^{P} f_j \exp(2\pi i \mathbf{s} \cdot \mathbf{r}_j) + \sum_{j=P+1}^{N} f_j \exp(2\pi i \mathbf{s} \cdot \mathbf{r}_j)$

where s is the reciprocal-lattice vector ($|\mathbf{s}| = 2\sin\theta/\lambda$) and the $\mathbf{r}_j$ are the atomic coordinates (in Å)

$= \mathbf{F}_P + \mathbf{F}_Q$, where the P atoms constitute the partial structure and the Q atoms the missing structure

$= |\mathbf{F}_N| \exp(i\alpha_N)$

$\mathbf{F}_P^c = \sum_{j=1}^{P} f_j \exp[2\pi i \mathbf{s} \cdot (\mathbf{r}_j + \Delta\mathbf{r}_j)]$, where $\Delta\mathbf{r}_j$ are positional errors

= structure factor of partial structure with errors

$D = \langle \cos(2\pi \mathbf{s} \cdot \Delta\mathbf{r}) \rangle$ (Luzzati, 1952)

ε = correction factor for expected intensity in reciprocal-lattice zone

$\Sigma_N = \sum_{j=1}^{N} f_j^2 = \langle |\mathbf{F}_N|^2/\varepsilon \rangle$

$\mathbf{E}_N = \mathbf{F}_N/(\varepsilon\Sigma_N)^{1/2}$

$\sigma_A = D(\Sigma_P/\Sigma_N)^{1/2}$

$m = \langle \cos(\alpha_N - \alpha_P^c) \rangle$

$= I_1(X)/I_0(X)$ for non-centric reflections, where I_0 and I_1 are the zero and first-order modified Bessel functions, respectively, or

= $\tanh(X/2)$ for centric reflections, where

$X = 2|\mathbf{F}_N||\mathbf{F}_P|/\varepsilon\Sigma_Q$ for a partial structure with no errors (Woolfson, 1956; Sim, 1959, 1960)

$= 2\sigma_A|\mathbf{E}_N||\mathbf{E}_P^c|/(1-\sigma_A^2)$ for a partial structure with errors (Srinivasan, 1966)

With a correction for expected intensity included, this becomes

$$\Sigma_Q \simeq \overline{||\mathbf{F}_N|^2 - |\mathbf{F}_P|^2|/\varepsilon}. \qquad (2)$$

This expression is used most commonly. Finally, for non-centric data

$$\langle |\mathbf{F}_Q|^2 \rangle = |\mathbf{F}_N|^2 + |\mathbf{F}_P|^2 - 2|\mathbf{F}_N||\mathbf{F}_P| I_1(X)/I_0(X),$$

where

$$X = 2|\mathbf{F}_N||\mathbf{F}_P|/\Sigma_Q$$

(Srinivasan, 1968). Nixon & North (1976) note this and solve the equation

$$\sum_h \langle |\mathbf{F}_Q|^2 \rangle = \sum_h \Sigma_Q$$
$$= \sum_h [|\mathbf{F}_N|^2 + |\mathbf{F}_P|^2 - 2|\mathbf{F}_N||\mathbf{F}_P| I_1(X)/I_0(X)]$$

for Σ_Q by numerical methods. If this is extended to include centric data and again the factor ε is included, it becomes

$$\sum_h \langle |\mathbf{F}_Q|^2/\varepsilon \rangle = \sum_h \Sigma_Q$$
$$= \sum_h [(|\mathbf{F}_N|^2 + |\mathbf{F}_P|^2 - 2m|\mathbf{F}_N||\mathbf{F}_P|)/\varepsilon], \qquad (3)$$

where m is the appropriate expression for the figure of merit of centric or non-centric data (see Table 1). It is instructive to test these methods on calculated data where the correct values of Σ_Q and the phase error are known.

Any reliable method for estimating Σ_Q should work in the ideal case of a perfect partial structure. Such a case was modelled by taking as $\mathbf{F}_N$ the calculated structure factors for *Streptomyces griseus* trypsin (SGT) at cycle 78 of least-squares refinement when the R factor was 0.159 (Read, Brayer, Jurášek & James, 1984). About 30% of the atoms were removed randomly to give a partial structure from which $\mathbf{F}_P$ was calculated. This test data set will be referred to as TD1. The correct value of Σ_Q was calculated as a function of resolution from the scattering factors of the missing atoms. Estimates of Σ_Q were calculated using (1), (2) and (3) (with the appropriate values of ε for the different zones of the space group $C222_1$). In Fig. 1(*a*), the correct values of Σ_Q and the three sets of estimates are shown; in Fig. 1(*b*), $\overline{\cos(\alpha_N - \alpha_P)}$ is compared to $\bar{m}$ calculated from each set of estimates. Equation (3) gives the best results, while (1) and (2) lead to a slight underestimate and a large overestimate respectively of Σ_Q.

By the time one has accurate coordinates in protein crystallography, however, there is comparatively little need for phase probabilities. In the early stages of developing a structure, there are large coordinate errors. A more realistic set of test data (referred to as TD2) was constructed, using as $\mathbf{F}_P^c$ the structure factors calculated for SGT at cycle 7 of refinement. At cycle 7, the R factor was 0.455 to 1.7Å resolution. Parts of the structure were missing, including all of the solvent molecules, and the model was similar to bovine trypsin in places where SGT is not. Individual thermal motion parameters had not yet been introduced, and the parts of the model that were essentially correct were inaccurate. Fig. 2 demonstrates that, for TD2, neither the method of Bricogne [(2)] nor even that of Nixon & North [(3)] gives reliable estimates of phase probabilities. The problem is not one of being unable to estimate Σ_Q correctly [even when the phases are used to estimate Σ_Q *via* $\overline{|\mathbf{F}_N - \mathbf{F}_P^c|^2/\varepsilon}$, $\bar{m}$ does not agree with $\overline{\cos(\alpha_N - \alpha_P^c)}$]; rather the problem is that the phase probability expressions are no longer valid.

2.2. *Estimating* σ_A

Since one may not safely ignore coordinate errors, it is necessary to use the phase probability distributions of Srinivasan and co-workers. Srinivasan & Ramachandran (1965) showed that, when the probability distributions are cast in terms of normalized structure factors, the effects of missing structure and of coordinate errors are formally equivalent. The parameter σ_A in these expressions varies from zero, when the partial structure provides no phase informa-

tion (no atoms in the partial structure, or an unrelated partial structure), to one, when the partial structure is perfect and complete. The factor D in σ_A (see Table 1) varies strongly with resolution when there are significant coordinate errors. As a result, σ_A should generally be estimated for several resolution ranges.

Hauptman (1982) has derived joint probability distributions for structure factors from isomorphous pairs of structures. In these equations, his parameter α plays the same role as σ_A in the expressions of Srinivasan. Hauptman suggests that σ_A can be estimated as the square root of the correlation coefficient between $|\mathbf{E}_1|^2$ and $|\mathbf{E}_2|^2$.

$$\sigma_A \simeq \left[\frac{\sum (|\mathbf{E}_1|^2 - \overline{|\mathbf{E}_1|^2})(|\mathbf{E}_2|^2 - \overline{|\mathbf{E}_2|^2})}{[\sum (|\mathbf{E}_1|^2 - \overline{|\mathbf{E}_1|^2})^2 \sum (|\mathbf{E}_2|^2 - \overline{|\mathbf{E}_2|^2})^2]^{1/2}}\right]^{1/2}. \quad (4)$$

Lunin & Urzhumtsev (1984) have proposed that parameters defining phase probabilities for partial structures with errors can be estimated from non-centric structure factors by maximizing a likelihood function. In the Appendix, this approach is extended to include centric data, and it is shown to correspond to estimating σ_A by finding the zero of the residual function

$$R = \sum w(\sigma_A - m|\mathbf{E}_N||\mathbf{E}_P^c|). \quad (A8)$$

In (A8), $w = 1$ for centric and 2 for non-centric reflections, m is the appropriate function of σ_A, and the structure factors are normalized so that $\sum w|\mathbf{E}|^2/\sum w = 1$. Equation (A8) is consistent with the result of Srinivasan & Chandrasekaran (1966) that

$$\sigma_A = \langle|\mathbf{E}_1||\mathbf{E}_2| \cos(\alpha_1 - \alpha_2)\rangle / (\langle|\mathbf{E}_1|^2\rangle\langle|\mathbf{E}_2|^2\rangle)^{1/2}. \quad (5)$$

Newton's method is used to solve for the zero of the residual function R; the initial estimate of σ_A is calculated from (4).

$$dR/d\sigma_A = \sum w(1 - |\mathbf{E}_N||\mathbf{E}_P^c| dm/d\sigma_A). \quad (6)$$

The expression for m differs for centric and non-centric data.

For centric data,

$$dm/d\sigma_A = \tfrac{1}{2}(1 - m^2)\, dX/d\sigma_A. \quad (7)$$

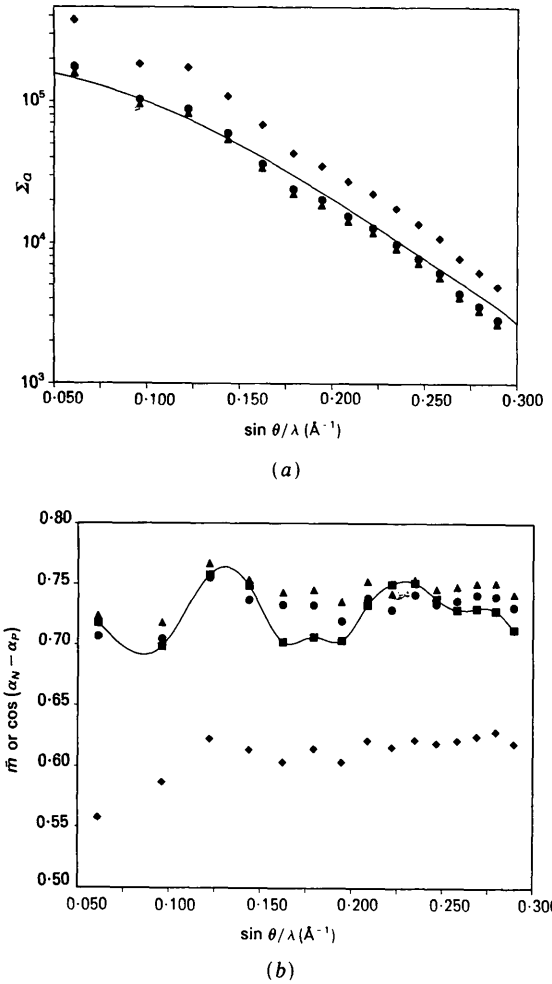

Fig. 1. Evaluation of methods for estimating Σ_Q in the case of a perfect partial structure (TD1). Estimates were calculated for 15 ranges of equal width in $(\sin\theta/\lambda)^2$. ▲ = estimate calculated according to (1) (Henderson & Moffat, 1971). ◆ = estimate from (2) (Bricogne, 1976). ● = estimate from (3) (Nixon & North, 1976). (*a*) Estimates of Σ_Q. The correct value, calculated as a function of $\sin\theta/\lambda$, is shown by the curve. The three sets of points show the three sets of estimates. (*b*) Estimates of $\overline{\cos(\alpha_N - \alpha_P)}$. The smooth curve connects the mean values of $\cos(\alpha_N - \alpha_P)$ for each resolution range, shown by ■. The other points are the mean values of m calculated from the estimates of Σ_Q shown in (*a*).

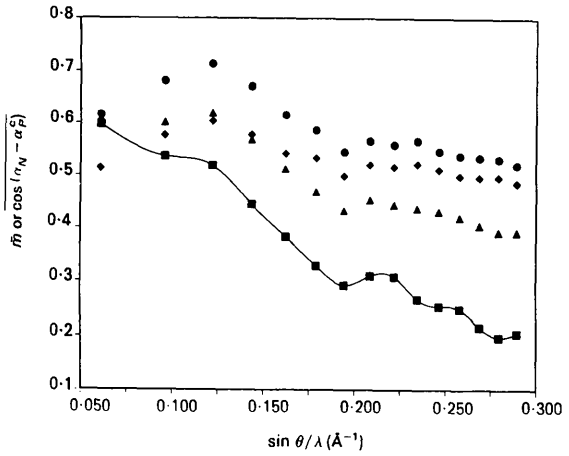

Fig. 2. Evaluation of phase probabilities calculated from estimates of Σ_Q in the case of a partial structure with coordinate errors (TD2). The smooth curve connects the mean values of $\cos(\alpha_N - \alpha_P^c)$ for each resolution range (■). ◆ = mean values of m calculated from Σ_Q estimated by (2). ● = mean values of m calculated from Σ_Q estimated by (3). ▲ = mean values of m calculated from Σ_Q estimated from $\overline{|\mathbf{F}_N - \mathbf{F}_P^c|^2/\varepsilon}$, *i.e.* using a knowledge of the phases.

For non-centric data,

$$dm/d\sigma_A = [1-(m/X)-m^2]\,dX/d\sigma_A. \quad (8)$$

Finally,

$$dX/d\sigma_A = 2|\mathbf{E}_N||\mathbf{E}_P^c|(1+\sigma_A^2)/(1-\sigma_A^2)^2. \quad (9)$$

Equations (6) through (9) define $dR/d\sigma_A$, and the next estimate of σ_A is given by

$$\sigma_{A,i+1} = \sigma_{A,i} - R/(dR/d\sigma_A).$$

The refinement generally converges in three or four cycles.

2.2.1. *Practical considerations in the algorithm.* The parameter σ_A varies with resolution. Therefore, σ_A is evaluated in shells of equal width in $(\sin\theta/\lambda)^2$. When there are between 500 and 1000 reflections in each shell, the estimates of σ_A seem to vary fairly smoothly and reliably. When there are too few reflections in a shell and the correct value of σ_A is small, there is sometimes a negative correlation between $|\mathbf{E}_N|^2$ and $|\mathbf{E}_P^c|^2$, in which case any starting value of σ_A refines to zero. [Note from (*A*8) that $\sigma_A = 0$ is always a zero of the residual *R*.] In general, the algorithm is less stable and the results less reliable when the correct value of σ_A is small. In this case, it is necessary to have a larger number of reflections in each resolution shell.

If the method used to normalize the structure factors does not result in the weighted mean of $|\mathbf{E}|^2$ being precisely unity, the simplifications made in the Appendix are not valid. The easiest solution is to renormalize within each resolution shell. Otherwise, one must return to (*A*6) and (*A*7) in the Appendix.

2.2.2. *Results of σ_A estimation.* Figures of merit calculated using σ_A are much more reliable than those calculated using Σ_Q (compare Fig. 3 with Figs. 1 and 2). From Fig. 3 one sees that the refined estimates of σ_A are slightly better than the estimates from (4), and that they lead to somewhat more reliable figures of merit for both sets of test data. (In implementing these methods, one might decide that the increased accuracy does not justify the increased programming effort.)

Lunin & Urzhumtsev (1984) observed that the accuracy of phases obtained from models refined in reciprocal space is overestimated. This effect can be seen in the TD2 data. The first seven cycles of refinement for SGT used 6.0 to 2.8Å data, and the figures of merit are systematically overestimated only within these resolution limits. In some way structure refinement must alter the distribution of errors; even the values of σ_A calculated using the phase differences *via* (5) give slightly high figures of merit.

2.2.3. *Omission of small $|\mathbf{F}_N|$.* In macromolecular crystallography, many low-intensity measurements are quite unreliable. The practice of discarding these observations has been criticized (*e.g.* Hirshfeld & Rabinovich, 1973), but is still quite common. This is probably due in part to concerns of cost and computer memory; in addition, the resulting bias is less pronounced for positional than for thermal motion parameters.

When low-intensity reflections are discarded, the distribution of $|\mathbf{F}_N|$ is altered, so that the joint distribution of $|\mathbf{F}_N|$ and $|\mathbf{F}_P^c|$ is altered. Therefore, under

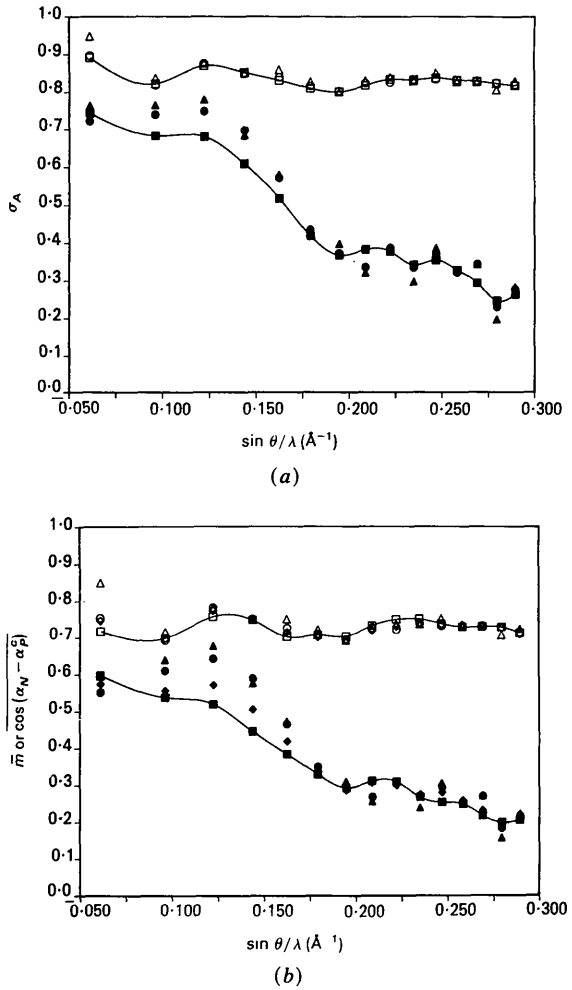

Fig. 3. Evaluation of methods for estimating σ_A. Open symbols are used for the test with TD1, closed symbols for TD2. (*a*) Estimates of σ_A. The smooth curves connect the values of σ_A calculated from (5) for each resolution range (□ and ■). The triangles (△ and ▲) show the estimates of σ_A from (4); the circles (○ and ●) show the refined estimates of σ_A. (*b*) Comparison of $\overline{\cos(\alpha_N - \alpha_P^c)}$ and $\bar{m}$. The smooth curves connect the mean values of $\cos(\alpha_N - \alpha_P^c)$. Other points indicate the mean values of *m* calculated from the estimates of σ_A shown with the same symbols as in (*a*). In addition, the diamonds (◇ and ◆) show the mean values of *m* calculated from the values of σ_A determined with (5).

these circumstances, phase probabilities determined using (4) are unreliable. On the other hand, the distribution of $|\mathbf{F}_P^c|$ conditional on $|\mathbf{F}_N|$ is unaffected, so that the equations derived in the Appendix are still valid. Numerical tests (results not shown) confirm this conclusion. However, the reduction in the number of observations can aggravate the instability in the algorithm when σ_A is small, so that figures of merit determined using all of the data are more reliable. In addition, the inclusion of data that were not used in structure refinement would be expected to reduce the overestimation of σ_A.

2.2.4. *Estimation of coordinate error from* σ_A. The Luzzati (1952) plot, which is commonly used to estimate coordinate errors in macromolecular structures, is based on the variation of D (see Table 1) with resolution. Srinivasan & Ramachandran (1965) note that, if the variation in σ_A with resolution is ascribed to the factor D, the resolution dependence of σ_A can also be used, in principle, to estimate the mean coordinate error of the atoms comprising the partial structure. However, the approach developed in later papers (*e.g.* Srikrishnan & Srinivasan, 1968) is to calculate an overall normalized R factor that, in a comparison with theoretical values, leads to a value for the mean coordinate error; this approach requires one to make an accurate *a priori* estimate of the ratio (Σ_P/Σ_N).

If the coordinate errors are assumed to be normally distributed, then

$$D = \exp\left[-\pi^3(\langle|\Delta\mathbf{r}|\rangle)^2(\sin\theta/\lambda)^2\right],$$

where $\langle|\Delta\mathbf{r}|\rangle$ is the expected value of the coordinate error (in Å) (Luzzati, 1952). Therefore,

$$\ln\sigma_A = (1/2)\ln(\Sigma_P/\Sigma_N) - \pi^3(\langle|\Delta\mathbf{r}|\rangle)^2(\sin\theta/\lambda)^2. \quad (10)$$

If the ratio (Σ_P/Σ_N) is constant, then a plot of $\ln\sigma_A$ *vs* $(\sin\theta/\lambda)^2$ should give a straight line with a slope of $[-\pi^3(\langle|\Delta\mathbf{r}|\rangle)^2]$ and an intercept of $(1/2)\ln(\Sigma_P/\Sigma_N)$. An example of a σ_A plot is shown in Fig. 4.

By assuming that (Σ_P/Σ_N) is constant, one assumes that the missing atoms are of the same type and have the same overall temperature factor as the atoms included in the partial structure. Clearly, this is invalid for the disordered solvent in a protein crystal; in applying (10) it will be necessary to ignore data to which the disordered solvent atoms contribute significantly, *i.e.*. reflections at lower than 5 or 6Å resolution (see Fig. 4). Atoms missing from partial structures often come from the less-well-ordered regions. This will hinder the use of (10) at intermediate stages of structure refinement. The σ_A plot will also be affected by the overestimation of σ_A that results from structure refinement (Lunin & Urzhumtsev, 1984).

For highly-refined structures, the observation errors in $|\mathbf{F}_N|$ will contribute significantly to the disagreement with $|\mathbf{F}_P^c|$. Since the proportional error in $|\mathbf{F}_N|$ generally increases with resolution, measurement errors will tend to lead to an overestimate of the coordinate error. The nonlinearity at high resolution of the σ_A plot shown in Fig. 4 can probably be attributed in part to measurement errors. In addition, only a minority of the highest-resolution data are above the cutoff used in refinement, so that the overestimation of σ_A due to structure refinement might be reduced at high resolution.

Similar considerations affect the use of the method of Luzzati (1952), except that the condition on the ratio (Σ_P/Σ_N) is more restrictive: it is assumed implicitly that Σ_P and Σ_N are equal. In a Luzzati plot, one compares a family of theoretical curves to R factors calculated in resolution shells; the σ_A plot defined by (10) requires the determination only of the slope of a line. Since σ_A is calculated from normalized structure factors, the σ_A plot, unlike the Luzzati plot, is unaffected by scaling errors. Though a σ_A plot must be interpreted with due care, it is therefore preferable in several respects to the Luzzati plot.

3. Removing model bias from maps

Structural information in the Fourier synthesis is contained to a great extent in the phase angles (Ramachandran & Srinivasan, 1970; Oppenheim, 1981). Therefore, electron density maps phased with

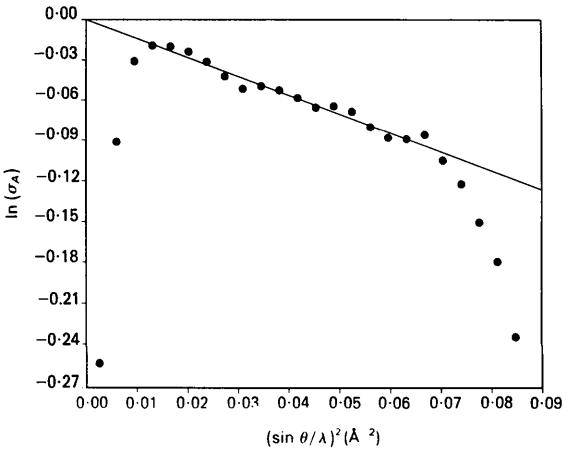

Fig. 4. Estimating coordinate errors with a σ_A plot for SGT at cycle 78 of least-squares refinement. To show the effects of disordered solvent and measurement errors, observed structure factors are used as $|\mathbf{F}_N|$. (For reflections with negative measured intensities, $|\mathbf{F}_N|$ is set to zero. In the highest-resolution range, 20·9% of the reflections fall into this category.) The line is a least-squares fit to all the points excluding the first three and the last four. From the intercept (0·000), $(\Sigma_P/\Sigma_N) = 1{\cdot}000$, and from the slope ($-1.410$Å^2), $\langle|\Delta\mathbf{r}|\rangle = 0.213$Å.

model phases are biased towards the model. Several suggestions have been made for Fourier coefficients that reduce model bias.

Luzzati (1953) showed that, for an almost complete structure, a map with non-centric coefficients $|\mathbf{F}_N| \exp(i\alpha_P)$ will show the missing atoms at half weight, but at less than half weight when more of the structure is missing. Because of this effect, maps with the commonly used coefficients $(2|\mathbf{F}_N| - |\mathbf{F}_P^c|) \exp(i\alpha_P^c)$ bring missing atoms up towards full weight. For a generalized version of these coefficients, $[n|\mathbf{F}_N| - (n-1)|\mathbf{F}_P|] \exp(i\alpha_P)$, Vijayan (1980) determined the value of n appropriate for different amounts of missing structure. Following a somewhat different approach, Main (1979) showed that, for non-centric data,

$$m|\mathbf{F}_N| \exp(i\alpha_P) \simeq \tfrac{1}{2}\mathbf{F}_N + \tfrac{1}{2}\mathbf{F}_P$$

so that non-centric coefficients that reduce model bias are given by $(2m|\mathbf{F}_N| - |\mathbf{F}_P|) \exp(i\alpha_P)$. Main's approach will be extended here to the case of a partial structure with errors.

3.1. *Non-centric case*

Following Main (1979), we start with the cosine law

$$|\boldsymbol{\delta}|^2 = |\mathbf{E}_N|^2 + \sigma_A^2|\mathbf{E}_P^c|^2 - 2\cos(\alpha_N - \alpha_P^c)\sigma_A|\mathbf{E}_N||\mathbf{E}_P^c|. \tag{11}$$

We use $\sigma_A\mathbf{E}_P^c$ instead of $\mathbf{E}_P^c$ because the expected magnitude and direction of $\boldsymbol{\delta}$ $(= \mathbf{E}_N - \sigma_A\mathbf{E}_P^c)$ are uncorrelated with those of $\mathbf{E}_P^c$. Thus, $\mathbf{E}_N$, $\sigma_A\mathbf{E}_P^c$ and $\boldsymbol{\delta}$ are interrelated in the same way as $\mathbf{F}_N$, $\mathbf{F}_P$ and $\mathbf{F}_Q$, respectively. Replacing both sides of (11) by expected values,

$$\langle|\boldsymbol{\delta}|^2\rangle = |\mathbf{E}_N|^2 + \sigma_A^2|\mathbf{E}_P^c|^2 - 2m\sigma_A|\mathbf{E}_N||\mathbf{E}_P^c|.$$

Noting that

$$|\mathbf{E}_N|^2 = \mathbf{E}_N\mathbf{E}_N^* = \mathbf{E}_N(\sigma_A\mathbf{E}_P^{c*} + \boldsymbol{\delta}^*),$$

we can rearrange and multiply both sides by $\exp(i\alpha_P^c)$ to get

$$\begin{aligned} &m|\mathbf{E}_N| \exp(i\alpha_P^c) \\ &\quad = \mathbf{E}_N/2 + \sigma_A\mathbf{E}_P^c/2 + \mathbf{E}_N\boldsymbol{\delta}^*/2\sigma_A\mathbf{E}_P^{c*} - \langle|\boldsymbol{\delta}|^2\rangle/2\sigma_A\mathbf{E}_P^{c*} \\ &m|\mathbf{E}_N| \exp(i\alpha_P^c) \\ &\quad = \mathbf{E}_N/2 + \sigma_A\mathbf{E}_P^c/2 + (\boldsymbol{\delta}^*/2)\exp(2i\alpha_P^c) \\ &\quad + (|\boldsymbol{\delta}|^2 - \langle|\boldsymbol{\delta}|^2\rangle)/2\sigma_A\mathbf{E}_P^{c*}. \end{aligned} \tag{12}$$

In the Fourier transform of $m|\mathbf{E}_N| \exp(i\alpha_P^c)$, third and fourth terms of (12) will lead to background noise [*cf.* Main (1979)]. Therefore, ignoring the noise contributions,

$$\begin{aligned} \mathbf{E}_N &\simeq (2m|\mathbf{E}_N| - \sigma_A|\mathbf{E}_P^c|) \exp(i\alpha_P^c) \\ \mathbf{F}_N &\simeq (\varepsilon\Sigma_N)^{1/2}(2m|\mathbf{E}_N| - \sigma_A|\mathbf{E}_P^c|) \exp(i\alpha_P^c). \end{aligned} \tag{13}$$

Substituting into (13) the expression for σ_A from Table 1, we get

$$\mathbf{F}_N \simeq (2m|\mathbf{F}_N| - D|\mathbf{F}_P^c|) \exp(i\alpha_P^c). \tag{14}$$

Thus, when there are no errors in the coordinates of the partial structure, $D = 1$ and (14) simplifies to give the Fourier coefficients derived by Main (1979). An advantage to working with normalized structure factors is that the scale factor and overall B value used in calculating structure factors do not affect the values of $\mathbf{E}_P^c$, so that errors in these quantities have no effect on Fourier coefficients calculated from (13). The information on the scale and B parameters relative to those of $|\mathbf{F}_N|$ is contained in the values of σ_A (see Table 1), which vary as a function of resolution.

3.1.1. *Effect of truncated data on non-centric map coefficients.* The estimation of σ_A from the equations derived in the Appendix depends on the conditional probability $P(\mathbf{E}_P^c; \mathbf{E}_N)$, whereas the derivation of (13) depends on $P(\mathbf{E}_N; \mathbf{E}_P^c)$. As long as the structure factors are normalized to be on the same scale ($\langle|\mathbf{E}_P^c|^2\rangle = \langle|\mathbf{E}_N|^2\rangle$), these expressions will be symmetrical (Srinivasan & Ramachandran, 1965), so that the value of σ_A in $P(\mathbf{E}_P^c; \mathbf{E}_N)$ is the same as in $P(\mathbf{E}_N; \mathbf{E}_P^c)$. This will be true when the data are complete. When low-intensity observations are discarded, normalization over the truncated data changes the relative scale of $|\mathbf{E}_N|$ and $|\mathbf{E}_P^c|$.

If we use a prime to indicate variables derived from truncated data, and if we let $|\mathbf{E}_N'| = k_N|\mathbf{E}_N|$ and $|\mathbf{E}_P^{c\prime}| = k_P|\mathbf{E}_P^c|$, then $\sigma_A' = (k_P/k_N)\sigma_A$ (as discussed in the Appendix) and $\Sigma_N' = \Sigma_N/k_N^2$. If one uses these values in (13),

$$\begin{aligned} &(\varepsilon\Sigma_N')^{1/2}(2m|\mathbf{E}_N'| - \sigma_A'|\mathbf{E}_P^{c\prime}|) \\ &\quad = 2m|\mathbf{F}_N| - (k_P/k_N)^2 D|\mathbf{F}_P^c|. \end{aligned} \tag{15}$$

Therefore, if not accounted for, data truncation will introduce a systematic error into the map coefficients. Knowing the relative size of k_P and k_N, however, one can make the appropriate correction.

It is necessary to determine only one of k_P and k_N; the other can then be estimated. From (*A*6),

$$\sum w(|\mathbf{E}_P^{c\prime}|^2 - \sigma_A'^2|\mathbf{E}_N'|^2)/\sum w = 1 - \sigma_A'^2, \tag{16}$$

where the sums are taken over the truncated data. Since the difference vector $(\mathbf{E}_P^{c\prime} - \sigma_A'\mathbf{E}_N')$ is independent of $\mathbf{E}_N'$, (16) would be valid even if the missing data were added. Therefore,

$$k_P^2 - k_N^2\sigma_A'^2 = 1 - \sigma_A'^2. \tag{17}$$

Thus, if k_N is determined by comparing Σ_N with Σ_N', k_P can be calculated from (17) and the correction implied by (15) can be made to the non-centric map coefficients.

3.2. *Centric case*

In the centric case, we start from (12) and note that

$$(\delta^*/2)\exp(2i\alpha_P^c) = \delta/2$$

because

$$\alpha_P^c = \alpha_\delta + n\pi.$$

Therefore,

$$m|\mathbf{E}_N|\exp(i\alpha_P^c) = \mathbf{E}_N + (|\delta|^2 - \langle|\delta|^2\rangle)/2\sigma_A\mathbf{E}_P^{c*}. \quad (18)$$

As in the case treated by Main (1979), therefore, the appropriate Fourier coefficients for centric data are simply $m|\mathbf{F}_N|\exp(i\alpha_P^c)$.

3.3. *Evaluating the map coefficients*

It is difficult to do objective visual comparisons of electron density maps calculated with different coefficients. One common quantitative measure for comparing maps is the root-mean-square value of the difference electron density (Blow & Crick, 1959). A related measure, which has the virtue of being unaffected by scaling errors, is the coefficient of correlation between electron density maps. For an electron density map omitting the contribution of the F_{000} term, the mean density is zero, so that the correlation coefficient is defined by

$$r = \frac{\int_v \rho_1(\mathbf{x})\rho_2(\mathbf{x})\,d\mathbf{x}}{[\int_v \rho_1(\mathbf{x})^2\,d\mathbf{x}\int_v \rho_2(\mathbf{x})^2\,d\mathbf{x}]^{1/2}}.$$

Application of the convolution theorem gives

$$\mathbf{r} = \frac{\sum|\mathbf{F}_1||\mathbf{F}_2|\cos(\alpha_1-\alpha_2)}{[\sum|\mathbf{F}_1|^2\sum|\mathbf{F}_2|^2]^{1/2}}, \quad (19)$$

where the sums are taken over a hemisphere of reciprocal space. One might note, comparing (19) and (5), that the correlation coefficient between two E maps is equivalent to an overall value of σ_A.

Two correlation coefficients will be used as objective criteria to judge electron density maps. The first is the correlation with the correct map (*i.e.* the Fourier transform of $\mathbf{F}_N$), which should be as close as possible to unity. Following a similar argument to that of Blow & Crick (1959), one can show that the maximum correlation between the correct map and one with Fourier coefficients $(w_1|\mathbf{F}_N| + w_2|\mathbf{F}_P^c|)\exp(i\alpha_P^c)$ is obtained when $w_1 = m$ and $w_2 = 0$; any coefficients designed to compensate for model bias will lower the correlation with the correct map. Nonetheless, model bias in an electron density map makes it difficult to detect and correct errors in the model. For this reason, coefficients that reduce model bias lead to maps that are subjectively (even if not objectively) improved. Optimal map coefficients will reduce model bias at only a small cost in the correlation with the correct map.

Table 2. *Correlation coefficients between electron density maps for TD2*

Non-centric Fourier coefficients of tested map*	Correct map (coefficients $\mathbf{F}_N$)	Model map (coefficients $\mathbf{F}_P^c$)
$\mathbf{F}_N$	1.0	0.585
$\|\mathbf{F}_N\|\exp(i\alpha_P^c)$	0.640	0.851
$m\|\mathbf{F}_N\|\exp(i\alpha_P^c)$	0.698	0.833
$(2m\|\mathbf{F}_N\| - D\|\mathbf{F}_P^c\|)\exp(i\alpha_P^c)$	0.663	0.659
$(2m\|\mathbf{F}_N\| - D\|\mathbf{F}_P^c\|)\exp(i\alpha_P^c)$†	0.666	0.681
$(2m\|\mathbf{F}_N\| - \|\mathbf{F}_P^c\|)\exp(i\alpha_P^c)$	0.570	0.397
$m(2\|\mathbf{F}_N\| - \|\mathbf{F}_P^c\|)\exp(i\alpha_P^c)$	0.660	0.682
$(2m_B\|\mathbf{F}_N\| - \|\mathbf{F}_P^c\|)\exp(i\alpha_P^c)$‡	0.588	0.479
$(2\|\mathbf{F}_N\| - \|\mathbf{F}_P^c\|)\exp(i\alpha_P^c)$	0.573	0.630
$(3\|\mathbf{F}_N\| - 2\|\mathbf{F}_P^c\|)\exp(i\alpha_P^c)$	0.490	0.446

* Centric Fourier coefficients were $m|\mathbf{F}_N|\exp(i\alpha_P^c)$ or $|\mathbf{F}_N|\exp(i\alpha_P^c)$ for all figure-of-merit-weighted and unweighted maps, respectively.

† These data were truncated by setting to zero all coefficients having $|\mathbf{F}_N| < 150e$.

‡ m_B refers to figures of merit calculated by the method of Bricogne (1976).

The second correlation coefficient is that with the model map. For a model-biased map this correlation will be higher than that of the correct map with the model map. Map coefficients that reduce model bias should lower this correlation, but not excessively. A correlation lower than that between the correct and model maps would indicate that correct features of the model were being eliminated.

Table 2 shows the results of some test calculations with TD2. The two correlation coefficients were evaluated, using (19), for several types of map coefficients that have been suggested previously. These results indicate that the coefficients described here are superior in reducing model bias with little cost in the resemblance to the correct map.

All of the map coefficients evaluated in Table 2 are phased by the model. The greatest model bias is found for the unweighted coefficient, $|\mathbf{F}_N|$. As expected, the highest correlation is given by figure-of-merit weighting ($m|\mathbf{F}_N|$), but this map shows considerable model bias. The non-centric coefficients $(2m|\mathbf{F}_N| - D|\mathbf{F}_P^c|)$ give a large reduction in model bias with little cost in the correlation to the correct map. Some of this reduction in model bias comes from reflections with small $|\mathbf{F}_N|$, for which the map coefficient will often be negative; omitting reflections with small $|\mathbf{F}_N|$ leads to a slight increase in model bias. Though the factor D might seem counter-intuitive, its omission in the coefficients $(2m|\mathbf{F}_N| - |\mathbf{F}_P^c|)$ leads to an excessive reduction in both correlation coefficients because the negative $|\mathbf{F}_P^c|$ component is too large. The coefficients $[m(2|\mathbf{F}_N| - |\mathbf{F}_P^c|)]$ are fairly successful in the case of TD2, but with an accurate partial structure, the figure of merit would not provide a fortuitous compensation for the factor D. When figures of merit calculated by the method of Bricogne (1976) are used in the coefficients $(2m_B|\mathbf{F}_N| - |\mathbf{F}_P^c|)$, both correlations are

quite low; they are not as low as for the related coefficients $(2m|\mathbf{F}_N|-|\mathbf{F}_P^c|)$ because the overestimation of m_B relative to m compensates in part for the omission of the factor D. Finally, the unweighted coefficients $(2|\mathbf{F}_N|-|\mathbf{F}_P^c|)$ and $(3|\mathbf{F}_N|-2|\mathbf{F}_P^c|)$ both lead to a low correlation with the correct map.

Fig. 5 allows a more subjective comparison. Electron density is shown for a part of SGT where two phenylalanine side chains were positioned incorrectly. The map using the coefficients derived here is compared to a map computed with the non-centric coefficients $(2m_B|\mathbf{F}_N|-|\mathbf{F}_P^c|)$, a figure-of-merit-weighted $(m|\mathbf{F}_N|)$ map, and the correct map. The figure-of-merit-weighted map displays serious model bias. In contrast, both the $(2m|\mathbf{F}_N|-D|\mathbf{F}_P^c|)$ and the $(2m_B|\mathbf{F}_N|-|\mathbf{F}_P^c|)$ maps indicate that the model is incorrect, and both show, at least in part, the correct positions of the side chains. However, the map computed with the coefficients $(2m|\mathbf{F}_N|-D|\mathbf{F}_P^c|)$ is somewhat superior in clarity and in connectivity of the density. The loss of connectivity in the $(2m_B|\mathbf{F}_N|-|\mathbf{F}_P^c|)$ map, which is consistent with the correlation coefficients in Table 2, would seriously impede the interpretation of some parts of the map.

4. Conclusion

It has been shown that, if one takes into account errors in the partial structure, more accurate estimates of phase probabilities can be made. It should be noted that a partial structure can be an atomic model or a density-modified electron density map. The values of σ_A in the phase probability expressions (see Table 1) are estimated most easily with (4) (Hauptman, 1982). However, estimates from (4) are not reliable when low-intensity observations are omitted. The effect of errors in the observed structure factor magnitudes has not been considered. However, these errors are likely to be small compared to the difference between the correct and calculated structure factors in a case for which phase probabilities are needed, *i.e.* when the model is bad.

Main's (1979) work on map coefficients for partial structures has been extended to encompass model errors. In a simulation, the coefficients derived here $[(2m|\mathbf{F}_N|-D|\mathbf{F}_P^c|)\exp(i\alpha_P^c)$ for non-centric, $m|\mathbf{F}_N|\exp(i\alpha_P^c)$ for centric] are objectively superior to coefficients that are currently in general use.

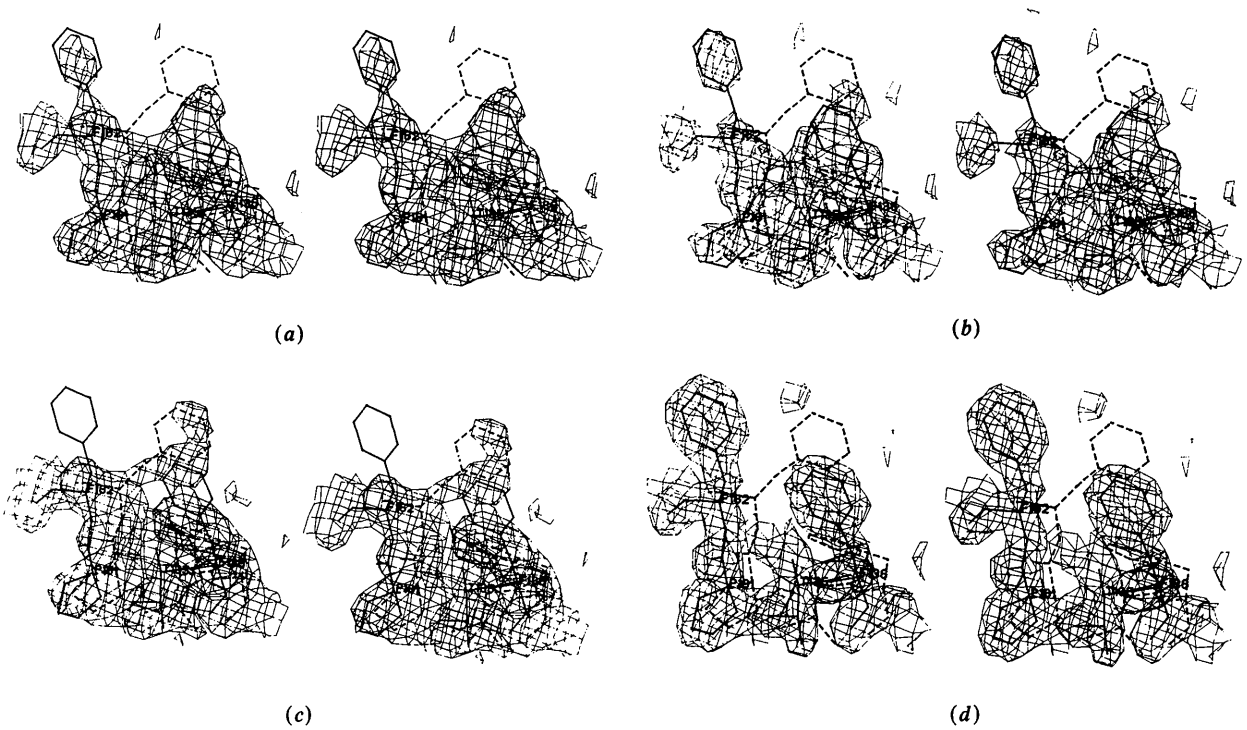

Fig. 5. Comparison of electron density maps. The correct structure (SGT at refinement cycle 78) is shown with solid lines; the partial structure with errors (SGT at cycle 7) is shown with dashed lines. All three maps are computed using the full 1·7 Å TD2 data set. Each map is contoured at 1·25 times the r.m.s. electron density of the map. For clarity, only contours within 1·7 Å of an atom in the figure are shown. (*a*) Map calculated with non-centric coefficients $(2m|\mathbf{F}_N|-D|\mathbf{F}_P^c|)\exp(i\alpha_P^c)$ and centric coefficients $m|\mathbf{F}_N|\exp(i\alpha_P^c)$. Contoured at 0·45 e Å^{-3}. (*b*) Map calculated with non-centric coefficients $(2m_B|\mathbf{F}_N|-|\mathbf{F}_P^c|)\exp(i\alpha_P^c)$ and centric coefficients $m_B|\mathbf{F}_N|\exp(i\alpha_P^c)$, where m_B is the figure of merit calculated by the method of Bricogne (1976). Contoured at 0·42 e Å^{-3}. (*c*) Map calculated with coefficients $m|\mathbf{F}_N|\exp(i\alpha_P^c)$ and contoured at 0·35 e Å^{-3}. (*d*) Map calculated with coefficients $\mathbf{F}_N$, *i.e.* the correct map, and contoured at 0·47 e Å^{-3}.

Accurate phase probabilities for partial structures are important for purposes other than the calculation of electron density maps. The neglect of coordinae errors leads to a significant overestimation of phase accuracy, which would cause combined phases to be skewed towards model phases. In addition, any attempts to improve or extend model phases by direct methods or maximum entropy methods (Bricogne, 1984) should benefit from more accurate phase probabilities.

M. N. G. James encouraged me to pursue the work presented here, while emphasizing the need for practical relevance. Discussions with M. Fujinaga and J. Moult have often been invaluable in helping me to clarify my ideas and their presentation. This research was supported by the Medical Research Council of Canada. During this work, I held a Medical Research Council of Canada Studentship, an Alberta Heritage Foundation for Medical Research Allowance and a University of Alberta Dissertation Fellowship.

APPENDIX

Maximum likelihood equations to determine σ_A

Srinivasan & Ramachandran (1965) derived the probability densities of $|\mathbf{E}_N|$ conditional on $|\mathbf{E}_P^c|$. For the non-centric case,

$$P(|\mathbf{E}_N|;|\mathbf{E}_P^c|) = [2|\mathbf{E}_N|/(1-\sigma_A^2)] \times \exp\{-[(|\mathbf{E}_N|^2+\sigma_A^2|\mathbf{E}_P^c|^2)/(1-\sigma_A^2)]\} \times I_0[2\sigma_A|\mathbf{E}_N||\mathbf{E}_P^c|/(1-\sigma_A^2)].$$

Since the relationships are symmetrical for the normalized structure factors (Srinivasan & Ramachandran, 1965), the roles of $|\mathbf{E}_N|$ and $|\mathbf{E}_P^c|$ can be interchanged. The structure factors are put on an absolute scale by changing variables:

$$P(|\mathbf{F}_P^c|;|\mathbf{F}_N|) = (2|\mathbf{F}_P^c|/\varepsilon\beta) \times \exp\{-[(|\mathbf{F}_P^c|^2+\alpha^2|\mathbf{F}_N|^2)/\varepsilon\beta]\} \times I_0(2\alpha|\mathbf{F}_P^c||\mathbf{F}_N|/\varepsilon\beta), \quad (A1)$$

where $\alpha = \sigma_A(\Sigma_P/\Sigma_N)^{1/2}$ and $\beta = \Sigma_P(1-\sigma_A^2)$. Equation (A1) is the same as equation (7) of Lunin & Urzhumtsev (1984), except for the inclusion of the factor ε and two apparent misprints in that paper (μ for the first β and α for α^2 in the argument of the exponential). For the centric case, one can similarly derive from the results in Srinivasan & Ramachandran (1965) that

$$P(|\mathbf{F}_P^c|;|\mathbf{F}_N|) = (2/\pi\varepsilon\beta)^{1/2} \times \exp\{-[(|\mathbf{F}_P^c|^2+\alpha^2|\mathbf{F}_N|^2)/2\varepsilon\beta]\} \times \cosh(\alpha|\mathbf{F}_P^c||\mathbf{F}_N|/\varepsilon\beta). \quad (A2)$$

Lunin & Urzhumtsev maximize the likelihood function

$$\psi = \Pi P(|\mathbf{F}_P^c|;|\mathbf{F}_N|), \quad (A3)$$

where the expression for $P(|\mathbf{F}_P^c|;|\mathbf{F}_N|)$ is their equation (7). By using (A1) for non-centric and (A2) for centric reflections, all of the data can be used in (A3). Estimates of α and β that maximize the likelihood function ψ occur when the partial derivatives of $\ln\psi$ with respect to α and β are both zero. When the appropriate expressions for the figure of merit are used [where $X = 2\alpha|\mathbf{F}_P^c||\mathbf{F}_N|/(\varepsilon\beta)$], the partial derivatives for the non-centric and centric terms differ only by a weighting factor ($w = 2$ for non-centric and 1 for centric).

$$\partial\ln\psi/\partial\alpha = \sum w(m|\mathbf{F}_P^c||\mathbf{F}_N| - \alpha|\mathbf{F}_N|^2)/\varepsilon\beta = 0 \quad (A4)$$

$$\partial\ln\psi/\partial\beta = \sum w(|\mathbf{F}_P^c|^2+\alpha^2|\mathbf{F}_N|^2 - 2\alpha m|\mathbf{F}_P^c||\mathbf{F}_N| - \varepsilon\beta)/2\varepsilon\beta^2 = 0. \quad (A5)$$

Equations (A4) and (A5) can be solved for β to get

$$\beta = \sum[w(|\mathbf{F}_P^c|^2 - \alpha^2|\mathbf{F}_N|^2)/\varepsilon]/\sum w. \quad (A6)$$

Because of the definitions of α and β, this is not a surprising result. From (A4), the parameter α is determined by finding the zero of

$$R = \sum[w(\alpha|\mathbf{F}_N|^2 - m|\mathbf{F}_P^c||\mathbf{F}_N|)/\varepsilon]. \quad (A7)$$

Note that α and β will adjust to compensate for an arbitrary change of scale. If $|\mathbf{F}_P^c|$ is scaled by a factor k_P and $|\mathbf{F}_N|$ by k_N, then values of α scaled by the factor (k_P/k_N) and of β scaled by k_P^2 will satisfy (A4) and (A5) while leaving the figures of merit unchanged. Therefore, if we use structure factors normalized so that $\sum w|\mathbf{E}|^2/\sum w = 1$, α is equivalent to σ_A, (A6) simplifies to

$$\beta = 1 - \sigma_A^2$$

and (A7) simplifies to

$$R = \sum w(\sigma_A - m|\mathbf{E}_P^c||\mathbf{E}_N|). \quad (A8)$$

References

BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794-802.

BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography.* London: Academic Press.

BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832-847.

BRICOGNE, G. (1984). *Acta Cryst.* A**40**, 410-445.

HAUPTMAN, H. (1982). *Acta Cryst.* A**38**, 289-294.

HENDERSON, R. & MOFFAT, J. K. (1971). *Acta Cryst.* B**27**, 1414-1420.

HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B**26**, 136-143.

HIRSHFELD, F. L. & RABINOVICH, D. (1973). *Acta Cryst.* A**29**, 510-513.

LUNIN, V. Y. & URZHUMTSEV, A. G. (1984). *Acta Cryst.* A**40**, 269-277.

LUZZATI, V. (1952). *Acta Cryst.* **5**, 802-810.

LUZZATI, V. (1953). *Acta Cryst.* **6**, 142-152.

MAIN, P. (1979). *Acta Cryst.* A**35**, 779–785.
NIXON, P. E. & NORTH, A. C. T. (1976). *Acta Cryst.* A**32**, 325–333.
OPPENHEIM, A. V. (1981). *Proc. IEEE*, **69**, 529–541.
RAMACHANDRAN, G. N. & SRINIVASAN, R. (1970). *Fourier Methods in Crystallography.* New York: Wiley.
READ, R. J., BRAYER, G. D., JURÁŠEK, L. & JAMES, M. N. G. (1984). *Biochemistry*, **23**, 6570–6575.
ROSSMANN, M. G. & BLOW, D. M. (1961). *Acta Cryst.* **14**, 641–647.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.
SIM, G. A. (1960). *Acta Cryst.* **13**, 511–512.
SRIKRISHNAN, T. & SRINIVASAN, R. (1968). *Z. Kristallogr.* **127**, 427–441.
SRINIVASAN, R. (1966). *Acta Cryst.* **20**, 143–144.
SRINIVASAN, R. (1968). *Z. Kristallogr.* **126**, 175–181.
SRINIVASAN, R. & CHANDRASEKARAN, R. (1966). *Indian J. Pure Appl. Phys.* **4**, 178–186.
SRINIVASAN, R. & RAMACHANDRAN, G. N. (1965). *Acta Cryst.* **19**, 1008–1014.
VIJAYAN, M. (1980). *Acta Cryst.* A**36**, 295–298.
WOOLFSON, M. M. (1956). *Acta Cryst.* **9**, 804–810.

Acta Cryst. (1986). A**42**, 149–156

On the Number of Ambiguities in Direct Methods – Anomalous Scattering Estimates of the Two- and Three-Phase Structure Invariants

BY SUZANNE FORTIER, MARIE E. FRASER AND NANCY J. MOORE

Department of Chemistry, Queen's University, Kingston, Canada K7L 3N6

(*Received* 4 *June* 1985; *accepted* 4 *November* 1985)

Abstract

The theoretical basis for the integration of direct methods into the single-wavelength anomalous dispersion technique is reexamined. The analysis shows that the approximations responsible for the ability to obtain unique estimates of the two- and three-phase structure invariants [Hauptman (1982). *Acta Cryst.* A**38**, 632–641; Giacovazzo (1983). *Acta Cryst.* A**39**, 585–592] or twofold estimates of the three-phase structure invariants [Kroon, Spek & Krabbendam (1977). *Acta Cryst.* A**33**, 382–385] are also responsible for the substantial errors observed in the applications. It is shown that, in the general case, the method of joint probability distributions leads to twofold estimates of the two-phase invariants and eightfold estimates of the three-phase invariants. Finally, it is shown that more accurately determined three-phase invariant estimates can be obtained by the use of anomalous scatterer substructure information, when available, and the use of a strategy that recognizes cases in which the eight estimates are clustered around one or two values. These cases are then distinguished from those where the eight estimates are widely scattered by a weighting function.

1. Introduction

It has been known for some thirty years that structure amplitude differences due to anomalous scattering can be used to obtain phase information. Reviews on the various proposed phasing techniques based on this approach can be found in several publications (Ramaseshan & Abrahams, 1975; Sayre, 1982; Ramachandran, 1964). Until recently, it was generally believed that, as in the single isomorphous replacement case, the single-wavelength anomalous dispersion experiment could yield only estimates of phases bearing a twofold ambiguity. Using the method of joint probability distributions, Hauptman (1982) and, subsequently, Giacovazzo (1983) obtained formulae which give unique estimates of the two-phase and three-phase structure invariants and thus unique estimates of the phases themselves. These results differ from those reported by Kroon, Spek & Krabbendam (1977), who obtained an estimate of the three-phase sine invariant, which implies of course twofold ambiguity in the estimate of the invariant itself.

While there is, at the moment, a substantial theoretical base for the integration of direct methods into the anomalous dispersion phasing technique, a number of points remain unclear. Firstly, the reasons why the one approach yields unique estimates of the invariants while the other results in a twofold ambiguity are not well understood. Secondly, in the initial applications made by Hauptman (1982) and Giacovazzo (1983), large errors persist in the invariant estimates, even when the calculations are done using error-free data. In addition, as noted by Giacovazzo (1983), the formulae tend to underestimate systematically the variance of the distributions. This suggests the presence of systematic errors in the proposed formulae and thus suggests that better estimates can be obtained once the errors are characterized and corrected.

In the present paper, the theoretical bases used by Hauptman (1982), Giacovazzo (1983) and Kroon,

0108-7673/86/030149-08$01.50