This work was supported in part by USPHS grant GM 30902.

References

ADMAN, E. T., SIEKER, L. C. & JENSEN, L. H. (1973). *J. Biol. Chem.* **248**, 3987-3996.
FAN HAI-FU & GU YUAN-XIN (1985). *Acta Cryst.* **A41**, 280-284.
FAN HAI-FU, HAN FU-SON, QIAN JIN-ZI & YAO JIA-XING (1984). *Acta Cryst.* **A40**, 489-495.
FORTIER, S., MOORE, N. J. & FRASER, M. E. (1985). *Acta Cryst.* **A41**, 571-577.
GIACOVAZZO, C. (1983). *Acta Cryst* **A39**, 585-592.
HAUPTMAN, H. (1982*a*). *Acta Cryst.* **A38**, 289-294.
HAUPTMAN, H. (1982*b*). *Acta Cryst.* **A38**, 632-641.
KARLE, J. (1980). *Int. J. Quantum Chem. Symp.* **7**, 357-367.
KARLE, J. (1983). *Acta Cryst.* **A39**, 800-805.
KARLE, J. (1984*a*). *Acta Cryst.* **A40**, 4-11.
KARLE, J. (1984*b*). *Acta Cryst.* **A40**, 366-373, 374-379.
KARLE, J. (1984*c*). *Acta Cryst.* **A40**, 526-531.
KARLE, J. (1985*a*). *Acta Cryst.* **A41**, 182-189.
KARLE, J. (1985*b*). *Acta Cryst.* **A41**, 387-394.
PONTENAGEL, W. M. G. F., KRABBENDAM, H., PEERDEMAN, A. F. & KROON, J. (1983). Eighth Eur. Crystallogr. Meet., Liège, Belgium, 8-12 August 1983. Abstract 4.01-P.
RAMASESHAN, S. & ABRAHAMS, S. C. (1975). Editors. *Anomalous Scattering.* Copenhagen: Munksgaard.
TIMKOVICH, R. & DICKERSON, R. E. (1976). *J. Biol. Chem.* **251**, 4033-4046.
YAO JIA-XING & FAN HAI-FU (1985). *Acta Cryst.* **A41**, 284-285.

Acta Cryst. (1986). A**42**, 253-256

Propagation of Phase Errors During Phase Refinement and Extension

BY LEE MAKOWSKI

Department of Biochemistry and Molecular Biophysics, Columbia University, 630 W 168th Street, New York, NY 10032, USA

(*Received* 17 *September* 1985; *accepted* 14 *January* 1986)

Abstract

A formalism is developed for estimating phase errors during refinement using non-crystallographic symmetry, solvent flattening or density modification. This formalism, based on a separation of all structure factors into known (best estimate) and unknown (random variable) parts, leads directly to simple expressions for the propagation of phase errors during a refinement process. Phase extension and treatment of unmeasured reflections are readily incorporated into this scheme. The formulation provides a direct method for evaluating the success of a refinement process. This may be useful in cases where examination of the resulting electron density map does not provide a quantitative evaluation of the calculations, such as at low to moderate resolution or when phase extension has been used. It may also provide a basis for designing optimal refinement strategies.

Introduction

It is becoming common in crystallographic studies of macromolecules to use density modification, non-crystallographic symmetry and solvent flattening to improve electron density maps or to extend and refine initial phase sets. These refinement strategies benefit from a weighting scheme in which accurately phased reflections are included in the calculation of the electron density map with higher weights than poorly phased reflections (*e.g.* Sim, 1959; Bricogne, 1976; Rayment, Baker & Caspar, 1983). Estimation of errors in the initial phase sets as obtained, for instance, by isomorphous replacement, can be calculated taking into account observational errors and lack of closure (*e.g.* Blow & Crick, 1956; Dickerson, Kendrew & Strandberg, 1961). Errors during refinement have generally been estimated by some variation of the method put forth initially by Sim (1959) who calculated the phase errors resulting from using a partial structure for the calculation of structure factors. In most applications the electron density map being refined cannot be divided into known and unknown regions. Consequently, the phase error is estimated by some measure of the mean discrepancy between calculated and observed intensities (*e.g.* Hendrickson & Lattmann, 1970; Bricogne, 1976). Although these measures provide good relative estimates of the decrease in phase error during refinement (as judged from the corresponding electron density maps), they are not absolute measures, and the success of a refinement procedure is usually based on the interpretability of the resulting electron density map. This leaves open the question of the reliability of the results, particularly at low resolution or after phase extension.

In this paper, the reciprocal-space formalism first put forward by Crowther (1967, 1969) is used as a starting point for deriving simple expressions for

0108-7673/86/040253-04$01.50

errors in phase estimates made during refinement using density modification, solvent flattening or non-crystallographic symmetry. The expressions for phase errors are based not on partial information in real space [as was Sim's (1959) formulation], but on partial information in reciprocal space.

Formulation of the problem

Formally, the relationship among structure factors due to the presence of non-crystallographic symmetry axes can be expressed in reciprocal space as (Crowther, 1967, 1969)

$$\mathbf{F}' = H\mathbf{F}, \tag{1}$$

where $\mathbf{F}$ is a vector containing the structure factors and H is a Hermitian matrix. $\mathbf{F}'$ is the vector of structure factors after application of H which, in the case of well determined amplitudes, is equal to $\mathbf{F}$. With appropriate modification of the components of H, (1) is also applicable to molecular replacement or to any solvent flattening or density modification schemes that can be expressed as

$$\rho'(\mathbf{x}) = b(\mathbf{x})\rho(\mathbf{x}), \tag{2}$$

where $\rho(\mathbf{x})$ is the density before modification, $\rho'(\mathbf{x})$ is the density after modification and $b(\mathbf{x})$ is the modification function. For instance, in solvent flattening, when the density map has been scaled to make solvent density equal to zero, $b(\mathbf{x})$ will be equal to unity inside the molecular envelope and zero outside the envelope. In that case, (1) represents a convolution in which H is the Fourier transform of $b(\mathbf{x})$ and (1) is the Fourier transform of (2).

If (1) is rewritten in terms of the vector and matrix components

$$F'_j = \sum_k H_{jk}F_k, \tag{3}$$

where F_j is the value of the structure factor of the jth reflection and the sum is over all structure factors, k. In any useful case no diagonal component, H_{jj}, will be equal to unity. The question being addressed here is: Given a set of structure factors containing errors, to what extent will the application of the matrix H decrease those errors?

Errors in the estimation of amplitudes

In order to investigate the propagation of errors during refinement it is useful to break up each structure factor into a known part and an unknown part. This is analogous to the real-space procedure of breaking up the structure into a known part and an unknown part (Sim, 1959). The known part of a structure factor can be taken to be the best current estimate F_{je} of the amplitude F_j. For instance, this estimate might be taken to be the centroid of the probability distribution for F_j in the complex plane (Blow & Crick, 1959). In general, the magnitude of the current best estimate will be less than the observed magnitude

$$|F_{je}| < |F_{jo}|. \tag{4}$$

The unknown part, δF_j, is a random complex number.

The value of F'_{je} calculated from the application of H to the current best estimates, F_{ke}, of the structure factors F_k is given by

$$F'_{je} = \sum_k H_{jk}F_{ke} \tag{5}$$

and the errors in these calculated values will be

$$\delta F'_j = \sum_k H_{jk}\,\delta F_k. \tag{6}$$

The error in the calculated value, $\delta F'_j$, is a random variable, different from δF_j as shown in Fig. 1. In particular the distribution of δF_j explicitly takes into account the measured amplitude, $|F_{jo}|$. The probability distribution of $\delta F'_j$ can be calculated since it is the sum of a large number of random variables, δF_k, with known distributions. When the number of reflections in the sum is large, the probability distribution of δF_j can be estimated accurately from the standard deviations of the distributions of all reflections. The sum in (6) can be thought of as a two-dimensional

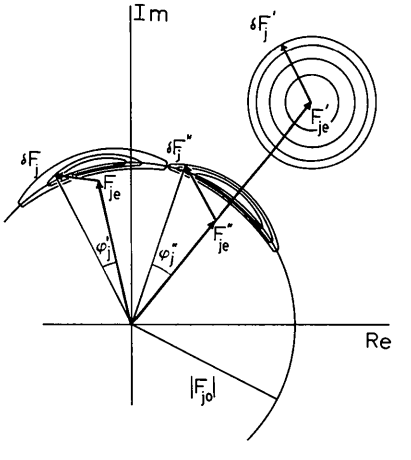

Fig. 1. Diagram of the complex plane showing hypothetical distributions for F_j, F'_j and F''_j in terms of best estimates of these values, F_{je}, F'_{je} and F''_{je}, and their uncertainties, δF_j, $\delta F'_j$ and $\delta F''_j$, for a reflection of measured amplitude $|F_{jo}|$. One round of refinement involves calculation of F''_{je} and $\delta F''_j$ from the values of F_{je} and δF_j, with intermediate calculation of F'_{je} and $\delta F'_j$. F'_{je} is calculated from all F_{ke} as in (5). This calculation will result in a structure factor F'_{je} with a phase different from F_{je} and a magnitude inconsistent with the measured $|F_{jo}|$. The uncertainty in this new estimate is $\delta F'_j$ and has a probability distribution as given in (7). To make this estimate consistent with the measured $|F_{jo}|$, the probability distribution for $\delta F'_j$ is projected onto the circle of radius $|F_{jo}|$ and the probability of a phase error φ''_j is calculated. From this phase-error distribution both the variance in the amplitude, $\delta F''_j$, and the best estimate, F''_{je}, for the reflection can be calculated.

random flight with step sizes ($H_{jk}\,\delta F_k$). This situation has been analyzed extensively (Chandrasekhar, 1943; Watson, 1944). It can be shown that the probability of the sum in (6) having a magnitude $|\delta F'_j|$ and a phase, ψ, is approximately equal to

$$p(|\delta F'_j|, \psi) = [\pi \sum H_{jk}^2 \sigma^2(\delta F_k)]^{-1} \times \exp\left[-|\delta F'_j|^2 \Big/ \sum_k H_{jk}^2 \sigma^2(\delta F_k)\right], \quad (7)$$

where $\sigma(\delta F_k)$ is the standard deviation (radius of gyration) of the distribution of δF_k. This distribution is, of course, independent of ψ. Fig. 1 shows the relationship between the unknown phase of the reflection F_j, the calculated value F'_{je}, and the calculated uncertainty $\delta F'_j$. If this calculated distribution is combined with the measured amplitude, the probability that the phase of reflection F_j deviates from that calculated from (5) by an angle φ''_j is given by

$$p(\varphi''_j) \propto \exp\,[2|F_{jo}||F'_{je}| \cos \varphi''_j / \sum_k H_{jk}^2 \sigma^2(\delta F_j)]. \quad (8)$$

For small phase errors, where the approximation $\cos \varphi \simeq 1 - \varphi^2/2$ is valid over much of the distribution, the variance in the phase distribution can be calculated to be

$$\sigma^2(\varphi''_j) = \sum_k H_{jk}^2 \sigma^2(\delta F_j)/2|F_{jo}|\,|F'_{je}|. \quad (9)$$

With the assumption of accurate measurement of $|F_{jo}|$, this results in a new estimate for the variance in the amplitude $\delta F''_j$ (Blow & Crick, 1959),

$$\sigma^2(\delta F''_j) = |F_{j0}|^2\left\{1 - \exp\left[\sum_k H_{jk}^2 \sigma^2(\delta F_k) \Big/ 2|F_{jo}|\,|F_{je}|\right]\right\}. \quad (10)$$

With this relationship it is possible to calculate a set of variances for the estimates of amplitudes obtained by a single application of the matrix H. Consequently, this equation can be used to monitor the propagation of errors with the application of H.

Equation (8) is similar in form to that due to Sim (1959) which, in this notation, can be written (Bricogne, 1976)

$$p(\varphi_j) \propto \exp\,[2|F_{jo}|\,|F'_{je}| \cos \varphi_j / \langle I_u \rangle], \quad (11)$$

where $\langle I_u \rangle$ is the mean intensity contributed by the unknown part of the structure. In the formulation presented here, the mean intensity contributed by the unknown part of the structure is replaced by the mean intensity contributed by the unknown parts of the reflections. By moving the analysis from real space to reciprocal space, the separation of the structural information into known and unknown parts becomes easier for many applications.

Phase refinement

A phase refinement scheme can be constructed using an iterative application of the real-space constraints expressed in H such that

$$F_j^{(n+1)} = |F_{jo}|\left\{\sum_k H_{jk} W_k^{(n)} F_k^{(n)} \Big/ \left|\sum_k H_{jk} W_k^{(n)} F_k^{(n)}\right|\right\}. \quad (12)$$

In other words, the $(n+1)$th estimate of F_j will have a magnitude as measured and a phase as estimated from the sum $\sum_k H_{jk} W_k^{(n)} F_k^{(n)}$ based on the nth estimates of the F_k. This is quite standard. The $W_k^{(n)}$ are weighting factors. They should be chosen to minimize the random (unknown) part of the structure factor. The variance, $\sigma^2(\delta F_k)$, of the distribution of δF_k is minimized if $W_k^{(n)}$ is chosen such that the product $W_k^{(n)} F_k^{(n)}$ is at the centroid of the probability distribution for $F_k^{(n)}$. This is the choice for which (10) is most accurate. In this case, from (9),

$$\sigma^2(\varphi_j^{(n)}) = \sum_j H_{jk}^2 \sigma^2(\delta F_k^{(n-1)})/2|F_{jo}|\,|F_j^{(n)}| \quad (13)$$

and (following Blow & Crick, 1959)

$$W_k^{(n)} = \exp\,[-(1/2)\sigma^2(\varphi_j^{(n)})] \quad (14)$$

$$W_k^{(n)} \simeq \exp\left[-\sum_k H_{jk}^2 \sigma^2(\delta F_k^{(n-1)}) \Big/ 4|F_{jo}|\,|F_j^{(n)}|\right]. \quad (15)$$

With these weighting factors, the propagation of errors in amplitude in the refinement scheme (12) is expressed, as in (10), as

$$\sigma^2[\delta F_j^{(n+1)}]^2 = |F_{jo}|^2 \times\left\{1 - \exp\left[-\sum_k H_{jk}^2 \sigma^2(\delta F_k^{(n)}) \Big/ 2|F_{jo}|\,|F_j^{(n+1)}|\right]\right\}. \quad (16)$$

Although the expression for amplitude errors does not explicitly depend on weighting factors, the presence of $|F_j^{(n+1)}|$ in the expression indicates that they will have an effect. Experience indicates that this is so.

Phase extension

Phase extension is readily incorporated into this scheme of phase refinement by noting that the probability distribution for an unphased reflection has a centroid at the origin and a standard deviation (radius of gyration) equal to $|F_{jo}|$. That is, for an unphased reflection $\sigma^2(\delta F_k) = |F_{ko}|^2$ and $W_k = 0$. This means that unphased reflections do not contribute to the estimates for the amplitudes of other reflections (since $W_k = 0$), but they do contribute to estimates for the errors in the amplitudes of other reflections. This is most important: Exclusion of measurable unphased reflections from the refinement process will result in

underestimation of the phase errors of other reflections. Including *all* measurable reflections in a refinement process is the only way to ensure correct estimation of phase errors.

Unmeasured reflections

It has been found that ignoring unmeasured reflections may severely inhibit a phase refinement (Rayment *et al.*, 1983). In the formulation presented here, the best estimate for the amplitude of an unmeasured reflection is F'_{je} as given by (5), the probability distribution for the uncertainty, δF_j, of this estimate is given by (7). The variance of this distribution is

$$\sigma^2[\delta F_j^{(n+1)}] = (1/2) \sum_k H_{jk}^2 \sigma^2(\delta F_k^{(n)}), \qquad (17)$$

which for small errors is the same relationship as (10) with $|F_j^{(n+1)}| = |F_{jo}|$. These values can be used in the refinement process for unmeasured reflections in the same way that the estimates from (12) and (16) are used for measured reflections.

Concluding remarks

It has been shown here that a simple method exists for breaking up the structure factors into known and unknown parts in order to make estimates of the errors of all reflections during a refinement. The known part of a structure factor is the current best estimate for that structure factor and the unknown part is a random variable about the best estimate. Because a useful refinement technique will always express a given reflection in terms of a sum of many other reflections, the detailed probability distribution for each reflection usually need not be considered. This formulation provides an estimate of the phase probability distribution [(8)] exactly analogous to that of Sim (1959) but using reciprocal-space quantities more readily determined in many crystallographic applications. This expression for phase errors can then be used in a manner completely consistent with most phase-refinement techniques to monitor propagation of error during refinement. The structure of the equations expressing the error propagation is such that completely unphased reflections and unmeasured reflections can be readily incorporated into the process. In fact, the equations estimating phase errors indicate that ignoring unphased reflections will always lead to underestimates of phase errors.

The ability to monitor propagation of phase errors during refinement leads inevitably to the question of how little initial phase information is required for a particular constraint matrix, *H*. Since the weighting factors for unphased reflections are necessarily zero, the form of the equations explicitly precludes *ab initio* phasing. But the possibility of bootstrapping from a very limited phase set is not necessarily precluded. By monitoring phase errors during refinement it may also be possible to design optimal strategies for phase extension. For instance, it will be possible to bring additional reflections into the calculation by slowly increasing their weights based on estimates of errors in their amplitude. This is likely to produce a more efficient phase extension scheme than stepping out an arbitrary number of reciprocal-lattice lengths every few rounds of refinement. The formulation presented here makes possible a quantitative approach to designing optimal refinement strategies.

I thank Dr Wayne Hendrickson for helpful discussions. This work was supported by NIH grant GM 29829 and an Irma T. Herschl Career Scientist Award.

References

BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832–847.
CHANDRASEKHAR, S. (1943). *Rev. Mod. Phys.* **15**, 1–89.
CROWTHER, R. A. (1967). *Acta Cryst.* **22**, 758–764.
CROWTHER, R. A. (1969). *Acta Cryst.* B**25**, 2571–2580.
DICKERSON, R. E., KENDREW, J. C. & STRANDBERG, B. E. (1961). *Acta Cryst.* **14**, 1188–1195.
HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B**26**, 136–143.
RAYMENT, I., BAKER, T. S. & CASPAR, D. L. D. (1983). *Acta Cryst.* B**39**, 505–516.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.
WATSON, G. N. (1944). *A Treatise on the Theory of Bessel Functions*, pp. 419–421. Cambridge Univ. Press.