α^{LW}(H;net) values based on the experimental elastic constants (for neutral atoms the deviations are even larger). However, errors in the elastic constants strongly influence the α^{LW}(H;net) values based on them. Changes of $\pm 25\%$ occur, if the elastic constants are changed by $\pm \sigma$ in a random way. Within this large experimental error α^{LW}(H;net) for the charged-atom model shows reasonable agreement with the LW values calculated from the elastic constants for most reflections. Exceptions are the reflections with H roughly parallel to $7E_1 + 2E_2 + 2E_3$ (for definition of E_1, E_2 and E_3, see Table 1), for which α^{LW}(H;el) is relatively small.

References

ALEKSANDROV, K. S., BELIKOVA, G. S., RYZHENKOV, A. P., TESLENKO, V. R. & KITAIGORODSKII, A. I. (1963). Sov. Phys. Crystallogr. 8, 164–166.
BORN, M. E. & HUANG, K. (1968). Dynamical Theory of Crystal Lattices. Oxford: Clarendon Press.
HELMHOLDT, R. B. (1975). Thesis, Univ. of Groningen.
HELMHOLDT, R. B. & VOS, A. (1977). Acta Cryst. A33, 38–45.
KROON, P. A. (1977). Thesis, Univ. of Groningen.
KROON, P. A. & VOS, A. (1979). Acta Cryst. A35, 675–684.
WILLIAMS, D. E. (1970). Trans. Am. Crystallogr. Assoc. 6, 21–33.
WILLIAMS, D. E. (1971). Acta Cryst. A27, 452–455.
WILLIAMS, D. E. (1974). Acta Cryst. A30, 71–77.
WOOSTER, W. A. (1962). Diffuse X-ray Reflections from Crystals, p. 67. Oxford: Clarendon Press.

Incorporation of Experimental Phases in a Restrained Least-Squares Refinement

BY D. C. REES* AND MITCHELL LEWIS

The Gibbs Chemical Laboratory, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA

Abstract

The least-squares refinement of macromolecular structures is characterized by a low ratio of observed data to refined parameters. Efforts have been made to compensate for this problem by incorporating subsidiary restraints into the observational equations. In this paper, a method is proposed and examples given for the introduction of additional observations into a least-squares refinement in the form of experimental phase information.

Introduction

The reciprocal-space least-squares refinement of atomic coordinates is becoming a routine procedure in the structure determination of macromolecules. The refinement of such structures is handicapped, however, by the comparatively low ratio of observed data to refined parameters. To increase the overdetermination of the refinement, efforts have been made to incorporate additional sources of information into the least-squares equations (Waser, 1963). These extra terms are generally in the form of molecular geometry restraints or constraints, using stereochemical data

---

* Present address: Department of Chemistry, University of California, Los Angeles, CA 90024, USA.

obtained from small-molecule structures. The quantity, Φ, minimized by the least-squares method may then be expressed as the sum of several terms:

$$\Phi = \sum \omega_h [\,|F_o| - |F_c|\,]^2 + \sum \omega_p (p_o - p_{model})^2, \quad (1)$$

where the first summation is the residual between observed and calculated structure factors, while the second summation contains the restraints between observed and model values for various stereochemical parameters. The weighting functions ω_h and ω_p control the contribution of each term to the least squares.

Traditionally, the role of experimental phases in a structure determination was limited to calculation of the electron density. Phases have been excluded from least-squares refinements, although inclusion of the experimental phases in a structure refinement could approximately double the number of observations. In addition, several protein structure refinements suggest that the experimental phases may indeed contribute useful information to the structure refinement. Phases calculated from a refined model agree more closely with the experimental phases than phases calculated from the preliminary model (Watenpaugh, Sieker, Herriot & Jensen, 1973; Rees, Lewis & Lipscomb, 1982).

Experimental phases usually contain significant random errors, but are free of systematic errors due to misinterpretation of the electron density map. We wish to demonstrate in this paper that by using appropriate selection criteria, experimental phases may indeed

provide a useful restraint in macromolecular structure refinements.

Theoretical background

The incorporation of phase restraints into a least-squares refinement is accomplished by introducing an additional term to Φ of the form:

$$\sum \omega_\alpha (\alpha_o - \alpha_c)^2, \qquad (2)$$

where α_o and α_c are the experimental and calculated phases respectively, and ω_α is the weighting function. The derivatives of α_c required for the least-squares refinement have been given by Cruickshank (1952):

$$\frac{\partial \alpha_c}{\partial x_I} = \frac{1}{|F_c|} \left(\cos \alpha_c \frac{\partial B_c}{\partial x_I} - \sin \alpha_c \frac{\partial A_c}{\partial x_I} \right), \qquad (3)$$

where A_c and B_c are the real and imaginary parts of the calculated structure factor, F_c; and x_I represents one of the parameters x,y,z, and B_{iso}, where B_{iso} is the isotropic temperature factor.

For the experimental phases to provide additional information for a least-squares refinement, it is necessary to determine how errors in the experimental phases influence the accuracy with which coordinates may be determined. A probability distribution function to describe the magnitude of the phase error for a given coordinate error has been derived by Parthasarathy & Parthasarathi (1974). This probability distribution may be used to determine the dependence of the average phase error, $\Delta\alpha$, on the r.m.s. coordinate error (Δr) and the fraction of scattering mass present, σ^2. For non-centrosymmetric reflections, the probability distribution $P(|\Delta\alpha|)$ may be written:

$$P(|\Delta\alpha|) = \frac{1 - \sigma_A^2}{\pi(1 - \sigma_A^2 \cos^2 \Delta\alpha)} \left\{ 1 + \frac{\sigma_A \cos \Delta\alpha}{(1 - \sigma_A^2 \cos^2 \Delta\alpha)^{1/2}} \right.$$

$$\left. \times \left[\frac{\pi}{2} + \sin^{-1} (\sigma_A \cos \Delta\alpha) \right] \right\}. \qquad (4)$$

σ_A is a product of two terms:

$$\sigma_A = \sigma D, \qquad (5)$$

where $D = \langle \cos (2\pi h \, \Delta r) \rangle$. Luzzati (1952) has demonstrated that the dependence of D on Δr and resolution may be expressed as:

$$D = \exp (-\pi^3 \, \Delta r^2 \sin^2 \theta / \lambda^2). \qquad (6)$$

For centrosymmetric reflections, the probability $P(s)$, where s is the product of the true and calculated phase sign, is

$$P(s) = \tfrac{1}{2} + (s/\pi) \sin^{-1}(\sigma_A). \qquad (7)$$

The average phase error as a function of σ_A (and consequently of resolution, Δr, and σ) may be determined by calculating the average:

$$\langle |\Delta\alpha| \rangle = \int_0^\pi \Delta\alpha P(|\Delta\alpha|) \, \mathrm{d}(\Delta\alpha). \qquad (8)$$

This integral may be evaluated analytically for centrosymmetric reflections. For non-centrosymmetric reflections, the integral was numerically evaluated using Gauss–Jacobi quadratures (Stroud & Secrest, 1966). Mean phase errors as a function of resolution, Δr and σ, for the non-centrosymmetric and centrosymmetric cases are illustrated in Figs. 1 and 2, respectively. Test calculations based on the structure factors calculated from two models of carboxypeptidase A having a r.m.s. coordinate difference of 0·72 Å show that the theoretical phase-error distribution agrees to within 10° of the observed distribution.

It is now possible to assess the accuracy with which coordinates may be determined from experimental phase information. For isomorphous replacement phases, the figure of merit, m, estimates the cosine of the average phase error for a given reflection (Dickerson, Kendrew & Strandberg, 1961). Consequently, one can determine the coordinate error for a particular value of m at a given resolution from Figs. 1 and 2,
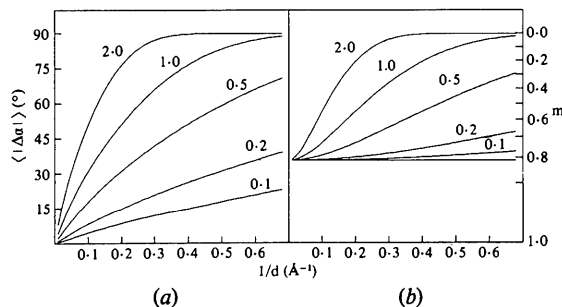

Fig. 1. The average phase difference, $\langle |\Delta\alpha| \rangle$, between correct phases and phases calculated from a non-centrosymmetric model, with r.m.s. coordinate errors of 0·0–2·0 Å, plotted as a function of resolution, $1/d$. The corresponding values for m are indicated on the right-hand side. (a) $\sigma^2 = 1·0$ (complete model), (b) $\sigma^2 = 0·8$ (partial model).

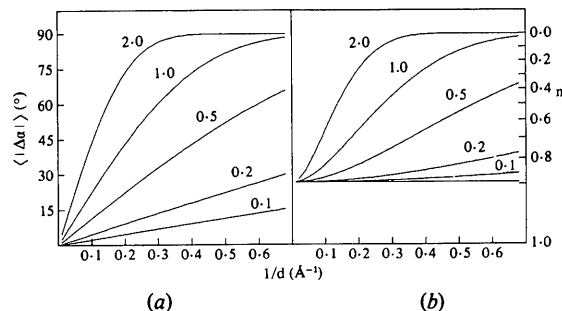

Fig. 2. The same plots as Fig. 1(a) and (b), but for centrosymmetric reflections.

using $\sigma^2 = 1$. For example, a 3 Å resolution reflection with $m = 0 \cdot 8$ would have an associated Δr of $0 \cdot 4$ Å. Luzzati (1952) provided an analogous statistical method for assessing the accuracy of atomic positions based on the discrepancy between observed and calculated structure factors.

Figs. 1(b) and 2(b) also illustrate that when only a fraction of the atoms are included in a structure, large errors are present in the calculated phases, even in the absence of positional errors for the included atoms. Under these conditions, which are not uncommon during the early stages of macromolecular refinement, the potential of incorporating experimental phases into a refinement procedure would be enhanced.

To utilize experimental phase information effectively in the least squares, it is necessary to consider the weighting term in (2). Least-squares weights are properly the reciprocal of the variance in an observation. For the phase case, the weight may be estimated by

$$\omega_\alpha = (1/\cos^{-1} m)^2. \tag{9}$$

Although $\cos^{-1} m$ is strictly the average phase error, and not the standard deviation of the phase error, this approximation will not significantly influence the relative weighting of different terms.

Results and discussion

The utility of experimental phase restraints in refinement was tested through model calculations on the dipeptide N-acetyl-L-phenylalanyl-L-tyrosine (NAPT) (Stenkamp & Jensen, 1973). The restrained least-squares program of Hendrickson & Konnert (1981) was modified to include the phase restraints. Relative weights were adjusted so that the structure factors, molecular geometry, and the phase information contributed equally to the diagonal elements of the normal matrix.

The observed structure factors for this test were calculated to 2·0 Å resolution from the model coordinates using an isotropic temperature factor of 10 Å^2 for each atom. In the various test calculations, no errors were introduced into these F_o's, but the observed phases were obtained by introducing r.m.s. errors of 0, 10, 20 and 35° in the calculated phases. The initial model for the refinement tests was generated from the true coordinates by applying a 0·15 Å r.m.s. random shift to the x,y,z coordinates of all the atoms, and an additional 0·15 Å systematic shift to the x coordinate, giving a net r.m.s. deviation of 0·297 Å. The initial crystallographic R factor calculated for this model was 0·24.

Five test refinement runs were generated varying the nature of the incorporated phase error. Seven cycles of least-squares refinement were calculated per test case.

The r.m.s. deviations of atoms from their true position during each test run are shown in Fig. 3. Comparison of test I (perfect experimental phases) and test II (no phase information) demonstrates that phase information can noticeably accelerate convergence of the refinement. Introduction of an average phase error of either 10° (test III) or 20° (test IV) into the experimental phases still improves the rate of convergence relative to test II. With a 35° phase error (test V), however, the phase information actually retards convergence. This illustrates the importance of employing the appropriate criteria for selecting the phases used in the refinement.

If an estimate of the average coordinate error is available (perhaps from a 'Luzzati plot' of the crystallographic R factor as a function of $2 \sin \theta/\lambda$), a plot of the average error in the calculated phases versus $2 \sin \theta/\lambda$ may be constructed from (6) and (8). A family of such curves for a range of coordinate errors is illustrated in Figs. 1 and 2. Reflections whose estimated experimental phase error falls below a particular curve will help improve the refinement. On the other hand, reflections whose estimated phase error falls above a curve will be of little value in the refinement. Using the results of the NAPT test calculations when $\Delta r = 0 \cdot 297$ Å, we see that for values of $2 \sin \theta/\lambda > 0 \cdot 1$ Å^{-1}, reflections whose phase error is less than 10° will improve the refinement. Likewise, reflections with a 20° phase error are useful for data with $2 \sin \theta/\lambda > 0 \cdot 2$ Å^{-1}. Most of the reflections used in the NAPT test calculations have $2 \sin \theta/\lambda$ values above $0 \cdot 2$ Å^{-1}, and the results of test runs III and IV show that phases with 10 and 20° errors improve the refinement relative to test run II (no phase information). With a 35° phase error, however, only phases for reflections with $2 \sin \theta/\lambda > 0 \cdot 4$ Å^{-1} should be used in the refinement; for the NAPT test calculations, this
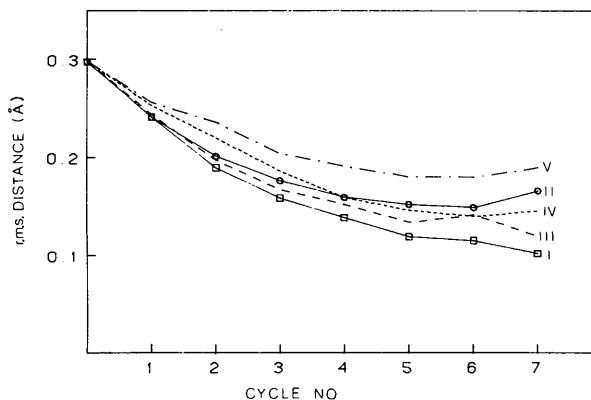
Fig. 3. The r.m.s. difference between refined and correct coordinates of NAPT, as a function of the least-squares cycle number. Details of the refinement are presented in the text. The phase errors for the various test runs are: I 0°; III 10°; IV 20°; V 35°. No phase information was included in run II.

is approximately half the observed data. As test run V illustrates, the phasing information is sufficiently poor for reflections with 2 sin θ/λ <0·4 Å^{-1} that the refinement convergence is actually retarded.

These considerations may be of more widespread utility for the general problem of combination of phases from different sources. Phase combination has typically been accomplished by multiplication of the appropriate phase distribution curves, and then calculating the best phase (Rossmann & Blow, 1961; Hendrickson & Lattmann, 1970). In general, the relative weights for different phase sources are determined empirically, and are not sensitive to the resolution of a reflection. The preceding discussion illustrates that a reflection with $m = 0·8$ may be much more useful at 2 Å resolution than at 6 Å resolution, even though the experimental phase probability distribution curves would be similar. Attention to such considerations may be of some value in improving the power of phase combination methods.

Phase information may also be particularly appropriate in the treatment of low-resolution reflections in structure refinement. Low-resolution reflections (2 sin θ/λ <0·1 Å^{-1}) are typically excluded from refinements due to neglect of the solvent in the structure model (Watenpaugh et al., 1973). The effect of the solvent is to decrease systematically the F_o's relative to the F_c's. Applying Babinet's principle to this problem indicates, however, that the solvent primarily decreases the amplitude of the reflection, without changing the phase (Moews & Kretsinger, 1975). This behavior is consistent with the observations of Lipscomb, Reeke, Hartsuck, Quiocho & Bethge (1970), who noted that even though F_o and F_c agreed poorly for low-resolution reflections of carboxypeptidase A (when the scale factor between the two data sets was calculated primarily from high-resolution data), the phase sets agreed quite well. As a result, utilization of phases in a structure refinement would permit incorporation of information from strong low-resolution reflections, which are now generally excluded.

Experimental phases do contain useful structural information, and should not be arbitrarily dismissed from a role in macromolecular structure refinement. The critical indicator for the utility of phase infor-

mation in refinement is the quality of the experimental phases. Recent progress in direct refinement of experimental phases (especially with density modification methods, although direct methods have also been used) has generated phase sets of remarkable quality from rather poor initial experimental phases (Harrison, Olson, Schutt, Winkler & Bricogne, 1978). Incorporation of this phase information into a least-squares structure refinement would provide a powerful method for increasing the number of experimental observations included in the refinement.

References

CRUICKSHANK, D. W. J. (1952). Acta Cryst. **5**, 511–518.
DICKERSON, R. E., KENDREW, J. C. & STRANDBERG, B. E. (1961). Acta Cryst. **14**, 1188–1195.
HARRISON, S. C., OLSON, A. J., SCHUTT, C. E., WINKLER, F. K. & BRICOGNE, G. (1978). Nature (London), **276**, 368–373.
HENDRICKSON, W. A. & KONNERT, J. (1981). In Biomolecular Structure Function, Conformation & Evolution, Vol. 1, edited by R. SRINIVASAN, pp. 43–47. Oxford: Pergamon Press.
HENDRICKSON, W. A. & LATTMANN, E. E. (1970). Acta Cryst. B**26**, 136–143.
LIPSCOMB, W. N., REEKE, G. N., HARTSUCK, J. A., QUIOCHO, F. A. & BETHGE, P. H. (1970). Philos. Trans. R. Soc. London Ser. B, **257**, 177–214.
LUZZATI, V. (1952). Acta Cryst. **5**, 802–810.
MOEWS, P. C. & KRETSINGER, R. H. (1975). J. Mol. Biol. **91**, 201–228.
PARTHASARATHY, S. & PARTHASARATHI, V. (1974). Acta Cryst. A**30**, 43–46.
REES, D. C., LEWIS, M. & LIPSCOMB, W. N. (1982). In preparation.
ROSSMANN, M. G. & BLOW, D. M. (1961). Acta Cryst. **14**, 631–647.
STENKAMP, R. E. & JENSEN, L. H. (1973). Acta Cryst. B**29**, 2872–2878.
STROUD, A. H. & SECREST, D. (1966). Gaussian Quadrature Formulas. Englewood Cliffs: Prentice Hall.
WASER, J. (1963). Acta Cryst. **16**, 1091–1094.
WATENPAUGH, K. D., SIEKER, L. C., HERRIOT, J. R. & JENSEN, L. H. (1973). Acta Cryst. B**29**, 943–956.