References

BUSING, W. R. & LEVY, H. A. (1957). *Acta Cryst.* **10**, 180–182.
COPPENS, P. & HAMILTON, W. C. (1970). *Acta Cryst.* A**26**, 71–83.
COPPENS, P., LEISEROWITZ, L. & RABINOVICH, D. (1965). *Acta Cryst.* **18**, 1035–1038.
FLACK, H. D. (1974). *Acta Cryst.* A**30**, 569–573.
HAMILTON, W. C. (1963). *Acta Cryst.* **16**, 609–611.
HUBER, R. & KOPFMANN, G. (1969). *Acta Cryst.* A**25**, 143–152.
International Tables for X-ray Crystallography (1972). Vol. II, 3rd ed., pp. 291–312. Birmingham: Kynoch Press.
KATAYAMA, C., SAKABE, N. & SAKABE, K. (1972). *Acta Cryst.* A**28**, 293–295.
KOPFMANN, G. & HUBER, R. (1968). *Acta Cryst.* A**24**, 348–351.
NORTH, A. C. T., PHILLIPS, D. C. & MATHEWS, F. S. (1968). *Acta Cryst.* A**24**, 351–359.
SEILER, P. & DUNITZ, J. D. (1978). *Acta Cryst.* A**34**, 329–336.
STUART, D. & WALKER, N. (1979). *Acta Cryst.* A**35**, 925–933.

Acta Cryst. (1983). A**39**, 166–170

A Method for Refinement of Partially Interpreted Protein Structures Including a Procedure for Scaling Between a Model and an Electron-Density Map

BY T. N. BHAT AND D. M. BLOW

Blackett Laboratory, Imperial College of Science and Technology, London SW7 2AZ, England

(*Received* 22 *February* 1982; *accepted* 26 *August* 1982)

Abstract

A method for the structure-factor least-squares refinement of a poor and incomplete atomic model of a protein molecule is suggested. Usually structure-factor least squares treat calculated phases as an observable to be associated with observed amplitudes of the complete model and assume that the difference between observed and calculated structure factors arises from the inaccuracies of the model. It is proposed that the assumption may be made valid for an incomplete model by including, in the calculation of the structure factors, electron-density features observed in the region outside the existing model. Structure factors obtained from the atomic model and from electron density may be scaled by fitting the radial distribution function in reciprocal space and their relative weights may be determined by an R-factor search procedure.

Introduction

Often an electron-density map for a protein, obtained by the multiple isomorphous replacement technique (MIR), can only be partly interpreted in terms of a molecular model. We recently described a density-modification procedure for the cyclic improvement of a partially interpreted electron-density map (Bhat & Blow, 1982). This procedure makes the map more readily interpretable but it does not include any method of adjusting the model coordinates so as to give a better fit to the improved map. (The only change it makes to the model is to change the 'occupancy' of atoms, depending on the values of the electron density in the vicinity of the atom site.)

It seems very desirable to introduce a procedure for refinement of the atomic model (which is likely to be tentative and incomplete) within each cycle of the procedure. In this paper we suggest how this can be done, increasing the power of the method, and reducing the amount of manual intervention. This provides a general method for refinement of an incompletely interpreted structure which is superior to refinement methods previously used.

The parameters of a diffraction pattern, in an ideal sense, are the amplitudes and the associated phase angles. Because the phases are not observed in X-ray diffraction, a number of different approaches to refinement of atomic parameters are possible, depending on the treatment of phases between cycles. There are three main choices,

(1) to keep the phases fixed, say at values obtained from the data [*e.g.* multiple isomorphous replacement (MIR)] – a phase-locked procedure;

(2) to recompute phases between two refinement cycles based on the refined parameters – a phase-free procedure;

(3) to restrain the recomputed phases towards previously obtained values – a phase-restrained procedure.

A general discussion of these types of approach is given by Diamond (1976).

The phase-locked refinement may be visualized as a technique of moving the model to achieve a better fit to the adjacent electron density. The electron density which defines the target positions for the model remains unchanged because of fixed phase angles. Early application of real-space refinement demonstrated a definite limit to the improvement which can be obtained

0567-7394/83/010166-05$01.50

while the electron density is held fixed by the relatively inaccurate MIR phases. The phase-free method takes advantage of the improvement of calculated phases, in response to shifts from earlier cycles. Deisenhofer & Steigemann (1975) using 1·8 Å data and a complete model showed that refinement converged to a more accurate result in a phase-free approach in which the density was modified by the use of calculated phases in real-space refinement.

In this paper we consider the problems which arise when a significant portion of density ρ_2 cannot be interpreted, if an attempt is made to refine the parameters which express the interpretable part of the density ρ_1 as an atomic model using a phase-free approach. In a phase-locked approach, except in regions of contact between ρ_1 and ρ_2, a refinement can treat ρ_1 and ρ_2 separately. In other words, phase-locked refinement permits partitioning of the unit cell into independent volumes for ρ_1 and ρ_2. In phase-free refinement changes in the calculated phases in response to shifts of the model produce changes in electron density throughout the structure and no partitioning is possible.

A refinement, whether phase free or phase locked, which obtains the phases from a model fitted only to ρ_1 ignores the contribution to the structure factors, and especially to the phases, from parts of the structure outside the volume ρ_1. This assumption is not justified when ρ_2 is significant. We describe here a refinement procedure where structural features of ρ_2 are explicitly recognized and included in the calculation of phases for the observed amplitudes. The procedure attempts to use more realistic phases than the conventional phase-free method. It is designed to estimate the contribution of parts of the structure of ρ_2 which have not been interpreted, in calculating shifts of the existing model. The estimate of ρ_2 depends on the assumed phase angles, as well as on the residual between the observed structure amplitude F_o and $|F_1|$ derived from ρ_1. If the refinement procedure improves the calculated structure factors, their phases can be used to obtain an improved estimate of ρ_2.

Estimation of the contribution from the density found in the volume covered by ρ_2 to the calculated structure factors and phases introduces a scaling problem. A similar problem was encountered in merging two electron-density maps in the density-modification procedure (Bhat & Blow, 1982), and the solution we proposed was not entirely satisfactory. We now suggest a better scaling procedure which can be applied over a wider range of conditions.

The method of structure-factor least squares minimizes the residual $\sum W(F_o - F_c)^2$ or a similar one. This function is not suitable when F_c are calculated from a partial model. The F_o used in the equation come from the entire scattering matter of the unit cell. A simple-minded approach to minimize the residual would tend to move the atoms of the existing partial model to account for the complete F_o. In the absence of restraining information like the shape of the atomic scattering curve, and stereochemistry, and without a favourable ratio of observations to parameters the refinement would tend to mask out the missing part of the structure.

With appropriate weights, refinement to minimize $\sum W(F_o - F_c)^2$ is equivalent to moving atoms along gradients of the difference map (Jensen, 1976; Cruickshank, 1952). The convergence of such a refinement depends on the height and width of the atomic peak (Diamond, 1976). Therefore, for a structure-factor least-squares refinement to be convergent, the electron-density map computed from F_c should resemble (at least in the vicinity of the model) the electron-density map computed from F_o. In order to achieve this, it may be necessary to reduce the effect of badly placed atoms of the model on F_c. It is proposed to assign weights to the model atoms on the basis of electron density computed from F_o.

Method

The method is related to the method previously proposed (Bhat & Blow, 1982), but includes a search for electron density not represented by the model and a step of structure-factor least-squares refinement within each cycle. It involves five steps:

(*a*) Determination of an 'occupancy' for each atom or group of atoms in the tentative model, based on the electron density found at the atom sites.

(*b*) Identification of large, well-connected electron-density peaks from regions of the map not covered by the tentative model.

(*c*) Calculation of two sets of structure factors, F_1 from the model coordinates with occupancies determined in step (*a*), and F_2 from the continuous densities identified in step (*b*).

(*d*) Determination of the scale factors s_1, s_2 to bring F_1 and F_2 to a level comparable to the observed structure amplitudes F_o.

(*e*) Structure-factor least-squares refinement of the model parameters defining F_1, so that $F_c = |s_1F_1 + s_2F_2|$ correspond to the observed structure amplitudes F_o as closely as possible.

This series of steps can be incorporated into a cyclic refinement process, described in a later section.

Step (*a*) has already been described by Bhat & Blow (1982). The occupancy is assigned on the basis of the degree of overlap between the volume of high electron density produced by the atomic model ρ_M, and the volume of high electron density observed in the starting electron-density map ρ_0. The occupancies have the effect of giving lower weight to regions of the model where atoms are poorly placed. This avoids distortions which would occur in the environment of the model density, if there were large discrepancies between ρ_M and ρ_0.

Step (*b*), which identifies large, contiguous and well-connected volumes where the electron density is above some assigned minimum value, has also been described by Bhat & Blow (1982). This step, which was optional in the density-modification procedure, assumes crucial importance in the procedure for refinement of a partial atomic model, as it provides a method of estimating the structure-factor contributions F_2 from the uninterpreted parts of the structure.

Step (*c*) uses standard fast Fourier transform routines. In step (*e*), any appropriate structure-factor least-squares procedure may be adapted to include the contribution F_2 from the continuous density. Step (*d*) is described at length below.

Scaling the transform of the continuous density to the atomic model

Bhat & Blow (1982) describe a method of obtaining a scale factor between ρ_M and ρ_0, but mention that the method is not entirely satisfactory. A much more satisfactory method will now be described, generalized so that the scale factor between the two contributions, F_1 and F_2, is also obtained. The chosen criterion is to find scale factors s_1, s_2 which minimize the conventional R factor

$$R = \sum |F_o - F_c| / \sum F_o = \sum \big|F_0 - |s_1 F_1 + s_2 F_2|\big| / \sum F_o. \tag{1}$$

In this equation F_o and F_c represent observed and calculated structure amplitudes; F_1 the calculated structure factors derived from the atomic model and occupancies in step (*a*), F_2 the calculated structure factors from the uninterpreted density identified in step (*b*), both on an arbitrary scale. As F_1 and F_2 are likely to have different form factors from F_o, it is more appropriate to assume scale factors $s_{\theta 1}$, $s_{\theta 2}$, which are dependent on $\sin\theta/\lambda$ and to calculate separate values over a number of ranges of $\sin\theta$:

$$R_\theta = \sum_\theta \big|F_o - |s_{\theta 1} F_1 + s_{\theta 2} F_2|\big| / \sum_\theta F_o. \tag{2}$$

Within each range of $\sin\theta$, assume

$$\sum_\theta F_o^2 = \sum_\theta |s_{\theta 1} F_1 + s_{\theta 2} F_2|^2. \tag{3}$$

We can define scale factors $c_{\theta 1}$, $c_{\theta 2}$, which can be applied to F_1, F_2 so that F_o^2, $|F_1|^2$, $|F_2|^2$ have the same radial distribution function:

$$\sum_\theta F_o^2 = c_{\theta 1}^2 \sum_\theta |F_1|^2 = c_{\theta 2}^2 \sum_\theta |F_2|^2. \tag{4}$$

Using such scale factors applied to F_1, F_2, (3) can be rewritten as

$$\sum_\theta F_o^2 = \sum_\theta |W_\theta c_{\theta 1} F_1 + (1 - W_\theta) c_{\theta 2} F_2|^2. \tag{5}$$

Here W_θ is a fraction between 0 and 1, the fraction of the total structure represented by the structure factor F_1 in this $\sin\theta$ range. Equation (2) can now be rewritten

$$R_\theta = \sum_\theta \big|F_o - |W_\theta c_{\theta 1} F_1 + (1 - W_\theta) c_{\theta 2} F_2|\big| / \sum_\theta F_o. \tag{6}$$

Thus by the use of (3), the scaling problem has been reduced to one unknown, W_θ, in each range of $\sin\theta$. In practice it is possible to make an estimate of W from a comparison of the map volume covered by the model density with the map volume covered by other continuous density identified in step (*b*). An accurate value of each W_θ is found by a search procedure, in which R_θ is calculated for different values of W_θ. If the estimation of the two contributions F_1, F_2 varies with $\sin\theta$, then the value of W_θ will vary for different values of θ.

Inclusion of continuous electron-density contribution F_2

In estimating the calculated structure factor, F_2 is included as an approximation to the contribution of the missing part of the structure. Since F_2 is derived from an earlier electron-density distribution, it tends to restrain changes in the calculated phases. Inclusion of F_2 helps to make $(F_o - F_c)$ a more realistic estimate of the structure-factor difference due to errors in the atomic parameters of the partial model. Crystallographic least-squares refinement has a particular difficulty over other types of least-squares analysis since only the structure-factor amplitudes F_o are observable. In minimizing the difference of amplitudes $(F_o - F_c)$ one implicitly assumes the same phase for the observed and calculated structure factors. In this type of refinement, if the contributions F_2 to F_c are omitted, not only may the magnitudes of the shifts be wrongly estimated, but the directions of shifts may be seriously incorrect.

The phases from a partial model may be so inaccurate that conventional structure-factor least-squares refinement cannot correct it. Under these circumstances independent phase information (isomorphous replacement) may improve the phases. The phase continuation techniques of Hendrickson & Lattman (1970) and Bricogne (1976) provide statistical methods to accommodate uncertainty in phase angles in a way which attempts to be free of bias. But we consider the more definite phase contributions provided by F_2 should be included in the calculation of F_c, and the phase combination then performed so as to take account of the remaining uncertainties.

A phase-locked refinement, like the real-space refinement of Diamond (1974), is not affected in the same way by changes in the model. It only takes account of local density, and can converge correctly with an incomplete model. One could devise a structure-factor least-squares procedure with precisely the same properties, which would require the addition to the model F_c of a contribution F_2 which represented all the difference density which is not in the vicinity of the part of the model being refined. Step (*b*) of our procedure selects from this difference density only those parts which have the properties of a real protein structure. Elimination of the remainder results in a degree of solvent flattening and noise suppression which are believed to improve refinement.

Application

The above procedure has been applied in the refinement of tyrosyl-tRNA synthetase (Bhat & Blow, 1982; Bhat, Blow, Brick & Nyborg, 1982). The electron-density map of this protein shows an ordered domain of 320 amino acids for which an atomic model consistent with the amino-acid sequence has been made, leaving a disordered domain of 99 residues which could not be interpreted. A model of the ordered domain alone was used at first for structure-factor least-squares refinement by the method of Konnert & Hendrickson (1980). The progress of the refinement is summarized in Table 1.

In the first attempts at refinement, structure factors were calculated with a full occupancy for all atoms, and the uninterpreted part of the electron-density map was ignored. Refinement was initially limited to 4 Å and then attempts were made to extend the resolution. As the R factor at 4 Å remained at 0·47 (cycle 12) and inclusion of higher-resolution reflections always caused the R factor to increase (cycles 6 to 10) it was decided to abandon the conventional constrained structure-factor least-squares procedure. The refinement was continued with variable occupancies, and a contribution F_2 to the calculated structure factor was calculated from a single piece of connected density of about 1500 Å^3 from the uninterpreted region of the map, identified by step (*b*). This represents a fraction $(1 - W)$ of about 10% of the total calculated structure factor. These two changes caused a large decrease in R and after four cycles (cycle 16) R continued to drop at about the same rate as at cycle 12. The difference in R factor between cycle 12 and cycle 16 (0·09) was probably mainly due to the introduction of occupancies and F_2. The energy restraints were left almost unaltered between these cycles. At cycle 18 the resolution was increased to 3·5 Å, R continued to fall, and the resolution was finally extended to 3·0 Å, with $R = 0{\cdot}307$.

Table 1. *Progress of refinement*

Cycle	R factor	$\sin\theta/\lambda$
5	0·524	0·125
6	0·547	0·152
7	0·533	0·135
8	0·516	0·135
9*	0·510	0·135
10	0·528	0·167
11	0·490	0·125
12	0·470	0·125
13†	0·471	0·125
14	0·426	0·125
15	0·401	0·125
16	0·379	0·125
17	0·362	0·125
18	0·388	0·143
19	0·369	0·143
32	0·307	0·167

* Attempts to extend resolution converged at a high R value (0·51). Refinement by the suggested method led to a final (cycle 32) R value of 0·28 at this resolution.

† From this cycle onwards 'occupancies' and contribution from F_2 are introduced.

A cyclic refinement and density-modification process

This example does not fully demonstrate the power of the refinement procedure, because an improvement of protein phase angles had already been made by cycles of density modification (Bhat & Blow, 1982) in which the structure-factor refinement (step *e*) was not included. Considerable improvement of the structure had already been made by rounds of rebuilding the model before any benefit was gained from automatic shifts in the refinement process. We believe that the structure-factor refinement step could have been introduced from an earlier stage, and this would have produced the same result with much less effort.

The refinement steps (*a*) to (*e*) above may be used in place of step (1) of the density-modification method of Bhat & Blow (1982). Thus, in addition to 'occupancy' determination for atoms of the model structure, a round of least-squares refinement of the model parameters would be performed, taking into account the existence of continuous regions of uninterpreted electron density. The remaining steps (2)–(7) of the density-modification procedure are unaltered. These identify an extended model density, which includes regions of continuous, well-connected density, and generate structure factors and a modified electron density. The scaling operation required in step (4) of our earlier paper should be performed by the superior method described in this paper.

This proposed cyclic procedure has not yet been used, but we hope to apply it to a suitable problem in the near future.

We thank Dr R. Diamond and Dr B. W. Matthews for helpful criticism of the manuscript. Financial support for TNB was provided by the Daresbury Laboratory of the Science and Engineering Research Council.

References

BHAT, T. N. & BLOW, D. M. (1982). *Acta Cryst.* A**38**, 22–29.
BHAT, T. N., BLOW, D. M., BRICK, P. & NYBORG, J. (1982). *J. Mol. Biol.* **158**, 699–709.
BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832–847.
CRUICKSHANK, D. W. J. (1952). *Acta Cryst.* **5**, 511–518.
DEISENHOFER, J. & STEIGEMANN, W. (1975). *Acta Cryst.* B**31**, 238–250.
DIAMOND, R. (1974). *J. Mol. Biol.* **82**, 371–391.
DIAMOND, R. (1976). In *Crystallographic Computing Techniques*, edited by F. R. AHMED, K. HUML & B. SEDLAČEK, pp. 291–301. Copenhagen: Munksgaard.
HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B**26**, 136.
JENSEN, L. H. (1976). In *Crystallographic Computing Techniques*, edited by F. R. AHMED, K. HUML & B. SEDLAČEK, pp. 307–316. Copenhagen: Munksgaard.
KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A**36**, 344–350.

SHORT COMMUNICATIONS

Contributions intended for publication under this heading should be expressly so marked; they should not exceed about 1000 *words; they should be forwarded in the usual way to the appropriate Co-editor; they will be published as speedily as possible.*

Acta Cryst. (1983). A**39**, 170–171

Refinement of cell and orientation parameters in four-circle diffractometry: the selection of orientation parameters.

By WILLIAM CLEGG, *Institut für Anorganische Chemie der Universität, Tammannstrasse* 4, *D*-3400 *Göttingen, Federal Republic of Germany*

(*Received* 11 *May* 1982; *accepted* 26 *August* 1982)

Abstract

A reliable and automatic method of selecting three orientation parameters is presented. The refinement of these and of the symmetry-constrained unit-cell parameters from four-circle diffractometer data is described.

Two basic methods are generally used for the refinement of cell and orientation parameters in four-circle diffractometry. The method of Tichý (1970) is a linear least-squares refinement of the nine independent elements of the orientation matrix **UB**. Busing & Levy (1967*a*) (hereafter BL) refine instead the six unit-cell parameters (a, b, c α, β, γ) and three orientation parameters; the refinement is non-linear, and normally requires two–three cycles. Both methods use as observations the optimum setting angles measured for a selection of reflections. The first method has the advantages of speed and simplicity; the BL method, although slower, allows for the application of symmetry constraints on the cell parameters, permits a simple and sensible empirical weighting scheme (*e.g.* χ values are inherently less precise than ω values, and can be assigned a lower weight), and can be used with partial information (*e.g.* 2θ values only, if orientation is not refined). It is generally accepted that the Tichý method, or an equivalent procedure, gives the 'best' orientation matrix for intensity data collection (allowing for some alignment errors), but the BL method provides more reliable cell parameters (Sparks, 1976). An alternative method described by Shoemaker & Bassi (1970) is similar to the Tichý method, but also allows for symmetry constraints: these, however, are much less simple than in the BL method and considerably complicate the otherwise linear refinement.

The orientation parameters for BL are three of the angles ω_1, χ_1, φ_1, ω_2, χ_2, φ_2 for two specific reflections (Busing, 1970). The choice of these reflections, and of the three parameters from among the possible six, is discussed by Busing & Levy (1967*b*), and summarized by them in a table. An automatic selection of these parameters by the refinement program would be much more convenient. In the BL method, the orientation matrix **UB** is made up of two components: **B** is an orthogonalization matrix, which relates the reciprocal-cell axes to an arbitrary orthogonal Cartesian axis set, and depends only on the cell parameters; **U** is an orthogonal rotation matrix, which relates this crystal Cartesian axis set to an axis set fixed to the diffractometer φ axis, and which can be calculated from the unit-cell parameters, together with the indices and setting angles for the two orienting reflections (for the basic definitions and equations, see BL; we use the same convention here, by which $\omega = 0$ for a reflection in bisecting geometry).

Hamilton (1974) has suggested an alternative method of selecting three orientation parameters: they are the ω_s, χ_s, and φ_s angles through which the crystal must be rotated from the setting with $\omega = \chi = \varphi = 0$ to an orientation in which the crystal Cartesian-axis set and the φ-axis set are coincident. Hamilton demonstrates how initial values for these three parameters are extracted from the known (preliminary) **UB** matrix, so that the procedure can be automated. [The preliminary **UB** matrix can itself be set up from known approximate cell parameters and the indices and setting angles of any two non-colinear reflections, from the angles of three indexed reflections (BL), or by various more automatic procedures, as summarized, for example, by Gabe (1980).] The drawback of Hamilton's procedure is the high correla-

0567-7394/83/010170-02$01.50