A Group Refinement Procedure in Protein Crystallography Using Fourier Transforms

BY ROBERT HUBER AND MONIKA SCHNEIDER

Max-Planck Institut fuer Biochemie, D-8033 Martinsried, Federal Republic of Germany

(Received 14 September 1984; accepted 8 January 1985)

Abstract

A rigid-body refinement method and program for crystallography of macromolecules is described. Orientational and translational parameters are refined by fitting the molecular Fourier transforms to the observed structure-factor amplitudes. The range of convergence of the method has been tested on four examples with known crystal structure: PTI, chymotrypsinogen and two forms of z1 anti-trypsin. It was successfully applied in the structure solution of two unknown crystal structures: a third form of z1 anti-trypsin and C-phycocyanin.

1. Introduction

The number of observed intensities is usually insufficient to allow free-atom refinement in protein crystallography. Several well established refinement procedures exist in which the model is constrained or restrained to a geometry known from accurate small-molecule structures (Diamond, 1971; Jack & Levitt, 1978; Hendrickson & Konnert, 1981). The relative merits of these procedures have been discussed (Deisenhofer, Remington & Steigemann, 1984; Steigemann, 1980). These methods require a model and atomic interpretation of the protein molecule in the crystal cell. Atomic coordinates are required to calculate the derivatives of the structure-factor amplitudes with respect to positions or dihedral and interbond angles or energy terms, respectively. The derivatives and the observed structure-factor amplitudes form the basis of a set of equations to determine the shifts of parameters that minimize the residual. The parameters varied may be atomic positions, temperature parameters, group positions and orientations, dihedral angles and interbond angles depending on the procedure used and the particular problem.

Group refinement has been introduced in protein crystallography and programs are available (Sussman, Holbrook, Church & Kim, 1977). In this procedure a large molecular fragment may have rigid internal geometry. Also this method requires atomic coordinates to calculate derivatives with respect to the group orientational and translational parameters.

However, in the initial steps of a protein crystal structure analysis, an electron density map obtained by isomorphous replacement techniques is available only and a straightforward atomic interpretation may be difficult. Analysis of a second crystal form may be helpful in this case as averaging of the two electron density distributions after optimal superposition will decrease the noise substantially. Cyclic averaging procedures are very powerful means to improve the phases (Bricogne, 1976). Such methods must obviously be applied in two stages: In the first stage an approximate solution for the relative orientation and translation of the molecular electron densities of the different copies has to be found. This may be possible by inspection of the Fourier maps if experimental phases are available. It is also possible to solve the crystal structure of a second crystal structure applying Patterson search techniques on the basis of the electron density distribution of the parent compound. It is obvious that refinement of the approximate orientation and translation is required whatever the means for obtaining a preliminary solution are. For this purpose a stepwise variation of the relevant group parameters and calculation of crystallographic $R$ values has been performed on various problems (Bode, Fehlhammer & Huber, 1976; Grau & Rossmann, 1981) but does not seem a satisfactory way to reach the local minimum in the multidimensional parameter space. We describe here a method and program for a least-squares rigid-body refinement on the basis of electron densities and their associated Fourier transforms.

2. Method and program

The Fourier transform $F(h)$ of the crystal unit cell is decomposed into the sum of the molecular transforms $F^M(h)$,

$$F(h) = \sum_{i,j} F^M(h)_{i,j},$$

where $i$ is the index of summation over the symmetry operations and $j$ is the index of summation over the independently treated molecules or molecular parts within the asymmetric unit. The molecular Fourier transforms $F^M(h)$ are usually defined in arbitrary orientation and translation in a frame different
from the reciprocal lattice of the crystal. De-orthogonalization, rotation and the symmetry operations must therefore be applied to the molecular Fourier transforms. The corresponding conventions and transformation matrices are described in the Appendix and characterized briefly here: The Fourier transform is calculated for the molecular electron density placed in a cubic unit cell with dimensions exceeding at least twice the largest diameter of the molecule to allow interpolation at non-integral lattice points. In the procedure used, the index triples of the observed reflections are transformed into the cubic lattice of the molecular transform and rotated by a set of Eulerian angles characterizing the orientation of the molecule in the cell. Linear interpolation is performed to obtain the structure factors at the non-integral lattice points. Symmetry operations are applied to generate the contributions from the symmetrically related molecules. Phase shifts are applied corresponding to the translation vectors of the molecules in the crystal cell.

Derivatives of the molecular Fourier transforms with respect to the orientational angles and translation parameters are calculated numerically: orientational angles and translation parameters are changed by small amounts and structure factors calculated. These and the structure factors calculated for the starting parameters provide the partial derivatives with respect to the variables. Derivatives of the structure factors with respect to scale and temperature factor are derived analytically. These derivatives and the observed and calculated structure-factor amplitudes form the basis of a system of linear equations. From these the normal equations are derived and the parameter changes calculated.

The present version of the program allows for three independent molecules or molecular parts in the asymmetric unit of the crystal. Their Fourier transforms are used alternately to calculate the contributions to the Fourier transform of the crystal. Seven variable parameters may be refined per molecule: three translation and three orientation parameters and a temperature factor. The resulting new parameters after a refinement cycle serve as starting parameters for a new cycle. The decrease in crystallographic residual is checked and refinement stopped eventually. The calculated structure factors are saved after each cycle to be used for Fourier or difference Fourier calculations.

3. Results and discussion

Method and program were tested on four examples: pancreatic trypsin inhibitor (PTI), chymotrypsinogen (CH), z1-antitrypsin and C-phycocyanin.

PTI, a molecule consisting of 57 amino-acid residues, forms a highly ordered crystal of space group \( P_{2_1}2_12_1 \) with lattice constants \( a = 43.1, b = 22.9, c = 48.6 \) \( \AA \) (Huber, Kukla, Ruhmann, Epp & Formaneck, 1970; Deisenhofer & Steigemann, 1975). The crystal structure has been refined to 1.5 \( \AA \) resolution. The \( R \) value was 0.165. The resulting model was used to calculate the molecular Fourier transform using a cubic cell of dimensions (70 \( \AA \))^3. The maximal dimension of the model is 33 \( \AA \). Its center of gravity had been shifted to the origin of the cubic unit cell. This model was shifted and reoriented by various amounts from the correct position and orientation to gain experience in the convergence behavior of the procedure. A small subset of the available observed intensities within the resolution shell 10–4 \( \AA \) was used for the calculations. Numerical derivatives were calculated with positional and orientational changes of 0.1 \( \AA \) and 1° respectively.

In Table 1, \( R \) values and convergence behavior are listed. Each cycle involves derivative calculation with the current values of orientational and translational parameters and solution of the resulting normal equations. The temperature parameter was not refined. A fixed value of 10 \( \AA^2 \) had been used for the calculation of the molecular Fourier transform.

Two experiments with orientational shifts of 2 and 5° respectively and translational shifts of 0.5 \( \AA \) were performed and showed that convergence to the correct starting values is achieved within about seven cycles. The calculated standard deviations of orientational and translational parameters are 0.2° and 0.02 \( \AA \) respectively.

The final \( R \) value is around 0.3, substantially higher than the \( R \) value after refinement (Deisenhofer & Steigemann, 1975). This is due to the use of an overall temperature factor instead of individual atomic temperature factors and reflects inaccuracies due to the linear interpolation of the molecular Fourier transform at the reciprocal-lattice points.

Chymotrypsinogen forms highly ordered crystals with space group \( P_{2_1}2_12_1 \) and lattice constants \( a = 59.2, b = 76.9, c = 109.6 \) \( \AA \). There are two independent molecules in the asymmetric unit. The crystal structure was solved by applying Patterson search techniques on the basis of a model of \( \gamma \)-chymotrypsin, omitting segments that were thought to be substantially different in conformation, i.e. about 20% of the polypeptide chain (Wang, Bode & Huber, 1985).

The molecular Fourier transform was calculated in a cubic unit cell of dimensions (90 \( \AA \))^3. The maximal diameter of the molecule is 45 \( \AA \). The center of gravity of the molecule was shifted to the origin of the model cell. The model used was the truncated \( \gamma \)-chymotrypsin model used for the Patterson search calculations. The starting parameters used for the test calculations described in Table 1 were those obtained from the Patterson search solution, which was quite accurate as shown by entry 4 of Table 1. The two
Table 1. Test of convergence for PTI, chymotrypsinogen, antitrypsin AC, AB and AH and C-phycocyanines

Shifts in orientation and position were applied to the starting values obtained from the final solution of the crystal structure and the range and speed of convergence was observed.

For antitrypsin AB the starting values were those of the related AC form. For C-phycocyanin the starting values were obtained by Patterson search techniques.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Resolution</th>
<th>Cycles</th>
<th>Starting values</th>
<th>Applied shifts and refined values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \Delta \theta )</td>
</tr>
<tr>
<td>PTI</td>
<td>10-4 Å</td>
<td>4</td>
<td>\begin{align*} \theta_1 &amp; = 0.3388 \ \theta_2 &amp; = -0.9214 \ \theta_3 &amp; = 0.0844 \end{align*}</td>
<td>( \Delta \theta_1 = -2 )</td>
</tr>
<tr>
<td></td>
<td>10-4 Å</td>
<td>7</td>
<td>\begin{align*} \theta_1 &amp; = 0.3388 \ \theta_2 &amp; = -0.9214 \ \theta_3 &amp; = 0.0844 \end{align*}</td>
<td>( \Delta \theta_1 = -5 )</td>
</tr>
</tbody>
</table>

Chymotrypsinogen

<table>
<thead>
<tr>
<th>Entry</th>
<th>Resolution</th>
<th>Cycles</th>
<th>Starting values</th>
<th>Applied shifts and refined values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6-4 Å</td>
<td>6</td>
<td>\begin{align*} \theta_1 &amp; = 0.62 \ \theta_2 &amp; = 0.65 \ \theta_3 &amp; = 1.6 \end{align*}</td>
<td>( \Delta \theta_1 = -0.60 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>\begin{align*} \theta_1 &amp; = 0.34 \ \theta_2 &amp; = 0.58 \ \theta_3 &amp; = 1.1 \end{align*}</td>
<td>( \Delta \theta_1 = -0.34 )</td>
</tr>
</tbody>
</table>

Antitrypsin AC form

<table>
<thead>
<tr>
<th>Entry</th>
<th>Resolution</th>
<th>Cycles</th>
<th>Starting values</th>
<th>Applied shifts and refined values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10-4 Å</td>
<td>3</td>
<td>\begin{align*} \theta_1 &amp; = 0.6299 \ \theta_2 &amp; = 0.2332 \ \theta_3 &amp; = 0.6370 \end{align*}</td>
<td>( \Delta \theta_1 = 0.06 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>\begin{align*} \theta_1 &amp; = 0.06 \ \theta_2 &amp; = -0.12 \end{align*}</td>
<td>( \Delta \theta_1 = -0.06 )</td>
</tr>
</tbody>
</table>

Antitrypsin AH form

<table>
<thead>
<tr>
<th>Entry</th>
<th>Resolution</th>
<th>Cycles</th>
<th>Starting values</th>
<th>Applied shifts and refined values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10-4 Å</td>
<td>5</td>
<td>\begin{align*} \theta_1 &amp; = 0.10 \end{align*}</td>
<td>( \Delta \theta_1 = 0.10 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>\begin{align*} \theta_1 &amp; = 0.10 \end{align*}</td>
<td>( \Delta \theta_1 = 0.10 )</td>
</tr>
</tbody>
</table>

Antitrypsin AB form

<table>
<thead>
<tr>
<th>Entry</th>
<th>Resolution</th>
<th>Cycles</th>
<th>Starting values</th>
<th>Applied shifts and refined values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10-4 Å</td>
<td>7</td>
<td>\begin{align*} \theta_1 &amp; = 0.00 \ \theta_2 &amp; = -0.12 \end{align*}</td>
<td>( \Delta \theta_1 = -0.07 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>\begin{align*} \theta_1 &amp; = 0.00 \ \theta_2 &amp; = -0.12 \end{align*}</td>
<td>( \Delta \theta_1 = -0.07 )</td>
</tr>
</tbody>
</table>

C-Phycocyanin Agmenellum quadruplicatum

<table>
<thead>
<tr>
<th>Entry</th>
<th>Resolution</th>
<th>Cycles</th>
<th>Starting values</th>
<th>Applied shifts and refined values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25-6 Å</td>
<td>10</td>
<td>\begin{align*} \theta_1 &amp; = 0.15 \end{align*}</td>
<td>( \Delta \theta_1 = -0.97 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>\begin{align*} \theta_1 &amp; = 0.14 \end{align*}</td>
<td>( \Delta \theta_1 = -0.97 )</td>
</tr>
</tbody>
</table>

molecules were then shifted from the parameters of entry 4 in Table 1 by 3, 5° and 1 Å respectively. These test calculations converged and the deviations from the starting parameters are compatible with the standard errors of 0.14° and 0.023 Å. In entries 6 and 7 shifts of 2 Å and 10° were applied. These calculations did not converge. A subset of reflections in a shell of resolution of 4 to 6 Å had been used.
α1-antitrypsin crystallizes in two modifications with space group P4₁2₁2 and lattice constants a = 120.5 Å, b = 120.5 Å, c = 113.5 Å and P6₃22 and lattice constants a = 119.7 Å, b = 119.7 Å, c = 216.3 Å, γ = 120° (Loebermann, Tokuoka, Deisenhofer & Huber, 1984).

Both crystal forms were analysed by isomorphous replacement techniques. The resulting electron density maps were averaged within the molecular envelope and this process repeated in cycles. The final electron density map was interpretable in terms of the known chemical amino-acid sequence. This electron density map, truncated by the molecular envelope, was placed in a cubic cell with dimension (110 Å)³. Its center of gravity was shifted to the origin and the Fourier transform calculated. The test calculations were performed with a data set within resolution limits 10 to 4 Å. The molecule was shifted by 3°, 1 Å and 2° and 1 Å respectively. Entries 9 to 11 in Table 1 refer to the tetragonal data set and show that convergence was achieved within a few cycles. A similar experiment was performed with the hexagonal data set (entries 13 to 15).

The tetragonal crystal modification of α1-antitrypsin undergoes a transition to a related crystal form upon exchange of the external potassium, sodium phosphate solution by ammonium phosphate. This crystal form has lattice constants a = b = 120.5 Å, c = 104.3 Å and space group P4₁2₁2 (AB form). It is related to the AC form but the c axis has shrunk by about 10%. Its crystal structure was solved by applying the procedure and program described above. The Fourier transform used was the same that had been calculated from the trimer PHML model placed in a cubic box with a 220 Å cell edge length. The dimensions of the trimer are 110 x 110 x 30 Å. The shifts obtained are given in Table 1. Phases were calculated for a difference Fourier map of a mercury derivative. Peaks numbered 1, 24, 34, sorted according to their heights, were consistent with an interpretation of the difference Patterson map.

In the next step the αβ units were refined by allowing variation of three orientational and three translational parameters. The shifts obtained are shown in Table 1. Phases and a difference Fourier map of the mercury derivative were calculated. This showed three outstanding peaks in accord with the interpretation of the difference Patterson map. The next highest peak had a height of one half of these maxima. It is obvious that the αβ monomer refinement improved the phases very considerably.

The applications described show that the procedure converges rapidly in a few cycles at the local minimum of the residual. The range of convergence is about 5° in orientation and 1 Å in translation for the examples chosen.

A further application of the method may be the refinement of non-crystallographic symmetry operators within one crystal form. This problem has been attacked by the stepwise variation of orientation and translation and calculation of the overlap between the related electron density fragments (Ladenstein, Epp, Bartels, Jones, Huber & Wendel, 1979; Fletterick & Steitz, 1976).

It is obvious that the program described here can be used to refine the Fourier transforms of the symmetry-related fragments after mutual exchange and to obtain improved local symmetry operators.

The program is available from the authors upon request.
APPENDIX

The crystal electron density \( \rho(x) \) is built up from the sum of molecular electron densities \( \rho(x') \) contained in the unit cell. The molecular electron densities \( \rho(x') \) are considered to be in arbitrary orientation and translation. \( x' \) is defined in an orthonormal frame. \( \rho(x') \) contributes to the crystal electron density \( \rho(x) \) at

\[
x = S_i O^{-1} E x' + S_i t,
\]

where \( E \) is an Eulerian matrix to orient properly the molecule in the cell, \( O \) is an orthogonalization matrix and \( O^{-1} \) the corresponding deorthogonalization matrix, \( S_i \) is the crystal symmetry operator \( i \), and \( t \) is a translation vector defined in the crystal frame to translate properly the molecule in the cell. There are \( n \) such contributions, if \( n \) is the number of crystal symmetry operations.

The orthogonalization matrix \( O \) is defined with the following convention: \( a' \) coincides with \( a \), \( b' \) is in the \( a, b \) plane and \( c' \) is orthogonal to \( a', b' \)

\[
O : \begin{pmatrix}
a & b \cos \gamma & c \cos \beta \\
0 & b \sin \gamma & -c \sin \beta \cos \alpha* \\
0 & 0 & c \sin \beta \sin \alpha*
\end{pmatrix}.
\]

The Fourier transform of \( \rho(x') \), \( F^M(h') \), contributes to the Fourier transform of the crystal \( F(h) \) at

\[
h = S_i^{-1} O^{-1} E^{-1} h'.
\]

In the program \( F(h) \) is evaluated by linear interpolation in the molecular Fourier transform, \( h' \) must therefore be expressed as a function of \( h \):

\[
h' = E'O^{-1} S_i h.
\]

The crystal Fourier transform is the sum over the \( n \) symmetry-related molecular Fourier transforms:

\[
F(h) = \sum_i \exp 2\pi i h S_i t^* F^M(E'O^{-1} S_i h).
\]

If there is more than one molecule in the asymmetric unit, the summation must include these also.

\( O^{-1} \) can be calculated from \( O \), but is given separately:

\[
O^{-1} : \begin{pmatrix}
a^* \sin \gamma \sin \beta^* & 0 & 0 \\
-a^* \sin \beta^* \cos \gamma & b^* \sin \alpha^* & 0 \\
a^* \cos \beta^* & b^* \cos \alpha^* & c^*
\end{pmatrix}.
\]

\( E \) is the Eulerian matrix and \( E' \) its transpose with the following convention: angle 1 rotates around \( c' \), angle 2 around the resulting \( a' \) and angle 3 around the resulting \( b' \).

\[
E' : \begin{pmatrix}
\cos 1 \cos 3 & \sin 1 \cos 3 & -\cos 2 \sin 3 \\
-(\sin 1 \sin 3 \sin 2) & +\cos 1 \sin 2 \sin 3 & \\
\sin 1 \cos 2 & \cos 1 \sin 2 & \sin 2
\end{pmatrix}.
\]

References