**REPLACE, a suite of computer programs for molecular-replacement calculations.** By LIANG TONG,*
Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA

(Received 15 January 1993; accepted 4 June 1993)

**Abstract**

A suite of computer programs (REPLACE) for crystal-structure determination by the molecular-replacement method has been developed. It currently includes a general (locked) rotation-function program (GLRF) and a translation-function program (TF). The rotation-function program can carry out ordinary as well as locked self- or cross-rotation-function calculations. It can also optimize non-crystallographic symmetry parameters based on the maximal electron-density overlap of the subunits related by the non-crystallographic symmetry. The translation search can be based on R factors, correlation coefficients, Patterson correlation and electron-density-correlation criteria. The packing of the molecular structure in the unit cell can be checked as well.

**Introduction**

The application of the molecular-replacement method to crystal-structure determination usually proceeds in several steps (Rossmann, 1972). The rotation functions are first examined to determine the orientation of non-crystallographic symmetry axes (self-rotation functions) or the orientation of a molecular structure in the crystal unit cell (cross-rotation functions). With the knowledge of the orientation of the molecular structure, translation functions can then be calculated to position the model in the crystal unit cell. Thirdly, in the presence of non-crystallographic symmetry, molecular-replacement averaging can be performed to improve and extend the phase information (Rossmann, 1990).

REPLACE is designed as a suite of computer programs for the application of the molecular-replacement method to crystal-structure determination. A general (locked) rotation-function (Tong & Rossmann, 1990) program (GLRF) and a general translation-function program (TF) are currently available. Other program packages for molecular-replacement applications have been reported in the literature. These include PROTEIN (Steigemann, 1974), MERLOT (Fitzgerald, 1988), CCP4 (1986) and X-PLOR (Brünger, 1992). The GLRF program calculates the rotation function in reciprocal space, whereas the PROTEIN and X-PLOR programs are based on calculations in real space. MERLOT and CCP4 utilize the fast rotation function (Crowther, 1972).

Several new strategies have been developed to speed up the calculations in reciprocal space. The TF program is based on a more general derivation of the translation function.

**Rotation function**

The ordinary rotation function was initially derived in its reciprocal-space form (Rossmann & Blow, 1962),

\[ R([C]) = \sum_p \sum_h F_p F_{h}^{*} G_{ph}, \]  

where

\[ [C] = [a][p][f]. \]  

\([x]\) is the deorthogonalization matrix and \([f]\) is the orthogonalization matrix (Rossmann & Blow, 1962). \([p]\) is the rotation matrix, which is usually defined based on a set of Eulerian or polar angles. The \(G\) function \(G_{ph}\) is the Fourier transform of a sphere with radius \(R\), which represents the longest interatomic vectors to be included in the calculation. The values of the \(G\) function depend on the length of the reciprocal-space vector \(h + p[C]\) (Rossmann & Blow, 1962).

The locked self- and cross-rotation functions have been described previously (Tong & Rossmann, 1990). They take advantage of the presence of multiple copies of a molecular subunit in a crystal unit cell. Another kind of rotation function can be defined that takes advantage of the presence of the same molecular structure in two or more crystal forms. The orientational relationship between the molecular structures in the two crystal forms \((B\ and\ C)\) can be defined with an ordinary cross-rotation function,

\[ y_{C} = [x_{C}][p_{CB}][a_{B}]y_{B}. \]  

Therefore, given a rotational relationship \([a_{B,A}]\) between a molecular model \((A)\) and crystal form \(B\), the relationship to crystal form \(C\) can be calculated:

\[ y_{C} = [x_{C}][p_{CB}][a_{B,A}][a_{A}]. \]

The 'triple' rotation function \(R_{T}\) is defined as the product of the two individual rotation functions:

\[ R_{T}([C_{B,A}]) = R([C_{B,A}]) \times R([C_{BA}][C_{B,A}]). \]

**Positioning and optimization of non-crystallographic symmetry elements**

When phases are available for reflections, a non-crystallographic symmetry axis can be located in the unit cell and its parameters optimized based on the maximal electron-density overlap of the molecular subunits related by the non-crystallographic symmetry (Blow, Rossmann & Jeffrey, 1964; Rossmann, 1972).

If \(s_{A}\) and \(s_{B}\) are the centers of two molecules (\(A\) and \(B\)) related by the non-crystallographic symmetry, then

\[ s_{B} = [C]s_{A} + d, \]
where \([C]\) and \(d\) are the rotational and translational components of the non-crystallographic symmetry axis, respectively. The electron-density values within a sphere centered at \(s_A\) will be considered for overlap with those in a sphere centered at \(s_B\).

\[
T = \int \rho(x_A) \rho(x_B) \, dx_A,
\]

where \(U\) is the volume of a sphere of radius \(R\) centered at \(s_A\). Noting \(x_B = [C]x_A + d\) and expanding the electron-density functions as Fourier transforms,

\[
T([C], s_A, s_B) = \sum_{h} \sum_{p} F_h F_p G_{hp} \exp \left( -2\pi i h s_A \right) \times \exp \left( -2\pi i p s_B \right).
\]

This equation can be used in three ways. Given the orientation \(([C]\)) and the center of the first molecule \((s_A)\), \(s_B\) can be determined by a Fourier transform with \(p\) as the indices and \(F_p \sum_{h} F_h G_{hp} \exp \left( -2\pi i h s_A \right)\) as the coefficients. Second, systematic searches can be performed to optimize initial estimates of \([C]\), \(s_A\), \(s_B\) with the above equation. In the derivation above, the two molecules can come from different crystal forms, with \([C]\) in this case determined from cross-rotation functions.

If \(s_A\) and \(s_B\) are related by a non-crystallographic symmetry axis in a crystal, the projection of the vector \(s_B - s_A\) onto the direction of the non-crystallographic symmetry axis will give the translation element along the rotation axis \((d)\). The average of \(s_A\) and \(s_B\) will give the location of the rotation axis in the unit cell. In the derivation of (7) above, if \(s_A\) is assumed to lie on the rotation axis, then \(s_B = [C]s_A + d\). This leads to the third application of (7):

\[
T(s_A) = \sum_{h} \sum_{p} F_h F_p G_{hp} \exp \left( -2\pi i p d \right) \exp \left( -2\pi i h s_A \right).
\]

Therefore, the non-crystallographic symmetry axis can be located by a Fourier transform with \(h + p\) as the indices and \(\sum_{h} F_h F_p G_{hp} \exp \left( -2\pi i p d \right)\) as the coefficients.

**Translation function**

Translation searches are usually carried out in terms of the crystallographic \(R\) factor, the correlation coefficient, Patterson correlation and electron-density correlation criteria. The orientation of the search model is predetermined with rotation functions and is generally not allowed to change. In such a case, the structure factors can be calculated as (Rae, 1977)

\[
F_h^* = \sum_{m} F_{h,m} \exp \left( 2\pi i h [T_m] x_0 \right),
\]

where \(x_0\) is a translation vector from a reference position for the search model. The summation goes over the crystallographic symmetry operators of the space group \((T_m)\) and

\[
F_{h,m} = \sum_{j} f_j \exp \left( 2\pi i h [T_m] x_j + t_m \right)
\]

is the contribution of the search model in the \(m\)th crystallographic asymmetric unit to the structure factor at the reference position. Equation (9) can be used to calculate \(R\) factors and correlation coefficients for the search model at different locations in the unit cell.

Rotation functions are based on the overlap of only a subset of the interatomic vectors in the Patterson map, i.e., only the self-vectors within each crystallographically unique molecule. The correct orientation and position of a molecular structure in the crystal unit cell should lead to the maximal overlap of both the self and the cross vectors, i.e., maximal overlap between the observed and the calculated Patterson maps throughout the entire unit cell. Therefore, a ‘Patterson correlation’ translation function can be defined,

\[
PC(x_0) = \int P(u) P(u) \, du
\]

where the integration goes over the entire unit cell. Noting (9), we have

\[
PC(x_0) = \sum_{h} \sum_{m} (F_h^*)^2 F_{h,m}^* F_{h,m} \exp \left\{ -2\pi i h [T_m - T_h] x_0 \right\}.
\]

This equation is similar to that derived by Crowther & Blow (1967). The Patterson correlation function can be treated as a Fourier transform with \(h[T_h - T_m]\) as the indices and \((F_h^*)^2 F_{h,m}^* F_{h,m}\) as the coefficients.

Similarly, if a correctly oriented atomic model is to be placed into an electron-density map obtained from other sources (e.g., multiple isomorphous replacement phasing), an ‘electron-density correlation’ translation function can be defined,

\[
EDC(x_0) = \sum_{h} F_h^* F_h^* \exp \left\{ -2\pi i h [T_m] x_0 \right\}.
\]

**The computer programs**

Programs **GLRF** and **TF** are written in standard Fortran and have been tested on VAX/VMS, DECstation Ultrix, IBM Risc6000, Silicon Graphics Iris and Indigo, Convex C220 and Cyber205 computers. Minimal modifications are needed to implement the programs on different computers. All the program input commands are free formatted and keyword based. There are extensive documentations for each program describing the function and the usage of each command. The programs and their documentations are available freely upon request from the author.

Many different conventions exist in the literature as to the definition of the Eulerian and polar angles as well as the orthogonalization of crystal unit cells. The programs have been coded to support most of the commonly used conventions and users can easily implement new conventions of their choice.

The general (locked) rotation-function program **GLRF** can carry out the following calculations:
1. ordinary or locked self-rotation function;
2. ordinary or locked cross-rotation function;
3. triple cross-rotation function;
4. special translation functions to optimize non-crystallographic symmetry parameters;
5. contouring of the rotation search maps. Stereographic projections are used for searches in polar angles. The plotting commands are in PostScript.

The translation-function program TF can carry out the following calculations:
1. structure-factor calculation from an atomic model by direct summation;
2. R factor and correlation-coefficient translation functions;
3. Patterson correlation translation function;
4. electron-density correlation translation function;
5. examination of molecular packing in a crystal unit cell;
6. contouring of the translation search maps;
7. atomic coordinate handling for the input, output, rotation and translation of molecular models.

Improving the speed performance of the programs

Owing to the necessity of sampling the rotation function at precise orientations for locked-rotation-function calculations, the 'slow' (reciprocal-space) rotation function is used in GLRF. Several strategies have been developed to improve the speed performance of the program. First, given a reflection p, the summation over h is limited to those reflections close to \(-p[C]\). This is feasible as the G function assumes large values only when \(h + p[C]\) is small (Rossmann & Blow, 1962). The size of this integration box is usually chosen as \(3 \times 3 \times 3\). Secondly, the 'large-term' approach is implemented in the program (Tollin & Rossmann, 1966). This limits the summation over p to only the strongest 10 to 20% of the reflections in each resolution shell. Two new approaches have been implemented in the GLRF program. A G-function table is set up for points within the integration box, sampled at 0.1 of the reciprocal-space unit vector. The G-function values are then obtained by table lookup. This generally produces a 30% speedup in program execution. The other new approach is based on the relatively slow variation of the rotation function. If the sampling of the rotation space is sufficiently fine (e.g. at 3 or 4° intervals), it will be unlikely for the 26 close neighbors of a given search grid point to have high rotation-function values if the point itself has a very low value. Consequently, the program is instructed to eliminate these 26 points from further calculations. Experience has shown that this 'neighborhood correlation' criterion can be used to remove 50 to 70% of the search grid points from the calculation, resulting in a two- to threefold speedup in program execution.

Similar approaches have been implemented to improve the speed performance of the translation-function program. For example, the search can be carried out with only the strongest 10 to 50% of the reflections in each resolution shell. The neighborhood correlation criterion is applied to the TF program with a slight modification due to the faster variation of the translation function. Only the six closest neighbors of any given search grid point are considered for removal from further calculations. Simple packing considerations have also been implemented which represent each search molecule as a sphere. Positions that would result in a collision of these spheres are removed from R-factor and correlation-coefficient searches.

Test case

The GLRF and the TF programs were used successfully in the structure determination of Sindbis virus core protein (Choi et al., 1991; Tong, Choi, Minor & Rossmann, 1992). The parameters of the non-crystallographic symmetry elements were optimized with the GLRF program based on solvent-flattened multiple isomorphous replacement phases. The electron-density values in one crystal form were successfully used to solve the structure in the other crystal form by molecular replacement (Tong et al., 1992). At present, the GLRF program has been distributed to about 50 sites.

As a test case, the molecular-replacement solution of Sindbis virus capsid protein in the tetragonal crystal form (space group \(P_4_2_1_2\)) based on an atomic model from the monoclinic crystal form (space group \(P_2_1\)) is presented. The atomic model was placed in a triclinic cell with \(a = b = c = 100\,\text{Å}\) and \(\alpha = \beta = \gamma = 90°\). Structure factors were calculated for 75 596 reflections between 10 and 3 Å resolution using the program TF. There were 1164 atoms in the atomic model and the calculation took 10 CPU min on an Indigo R4000. The cross-rotation function was calculated using reflections between 8.0 and 3.0 Å resolution. The radius of integration was 20 Å. Only 657 out of the 3078 observed reflections (21%) in this resolution range were used in the calculation as large terms. The search was carried out in Eulerian angles, covering the region 0-90° in \(\theta_1\) and \(\theta_2\) and 0-360° in \(\theta_3\) (Rao, Jih & Hartsuck, 1980). The search grid interval was 3° in \(\theta_1\) and \(\theta_2\) and 4° in \(\theta_3\), giving a total of 87 451 grid points. The neighborhood correlation criterion removed 54% of the grid points from the calculation (Table 1). The calculation took 4.5 CPU h on an Indigo. A section of the rotation-function map containing the top peak is shown in Fig. 1. A finer search was carried out, with a 1° interval, around the top peak of the rotation function, locating the peak at (7, 26, 200) in Eulerian angles. As a comparison, similar rotation-search results were obtained with the program MERLOT (9 CPU min on a VAX/VMS 6000) and X-PLOR (40 CPU min on a Convex C220).

Three different translation searches were carried out, based on the correlation coefficient, R factor and Patterson correlation criteria. 446 out of 766 observed reflections (58%) between 5 and 4Å resolution were used as large terms in the correlation-coefficient and R-factor searches. The search covered the region 0-1 in \(a\) and 0-0.5 in \(b\) and \(c\) and the search-grid interval was 0.01. The atomic model was represented by a sphere of 9Å radius for packing calculation, which removed 19% of the search-grid points from the calculation. The neighborhood correlation criterion removed about 40% of the grid points from the calculation. Reflection data between 8 and 4 Å resolution were used in the Patterson correlation translation search, saving 20% of the observed reflections as large terms. The Fourier transform was sampled on a grid of 90 \(\times\) 90 \(\times\) 180. The Patterson correlation calculation is much faster than the correlation coefficient or the R-factor search (Table 1).
### Table 1. Rotation and translation search results from the test case

<table>
<thead>
<tr>
<th>Resolution (Å)</th>
<th>Number of reflections</th>
<th>Number of large terms</th>
<th>Number of search grid points</th>
<th>% grid points removed by neighborhood correlation</th>
<th>% grid points removed by packing calculation</th>
<th>% grid points actually searched</th>
<th>Rotation function</th>
<th>Correlation coefficient</th>
<th>R factor</th>
<th>Patterson correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 ± 3</td>
<td>3078</td>
<td>657</td>
<td>87451</td>
<td>-</td>
<td>19</td>
<td>48</td>
<td>270</td>
<td>66.9 (10.6)</td>
<td>27.2 (12.3)</td>
<td>222 (17.0)</td>
</tr>
<tr>
<td>7 ± 3</td>
<td>766</td>
<td>446</td>
<td>262701</td>
<td>-</td>
<td>19</td>
<td>36</td>
<td>446</td>
<td>46.8 (6.8)</td>
<td>36.2 (7.5)</td>
<td>1298 (9.9)</td>
</tr>
<tr>
<td>6 ± 3</td>
<td>446</td>
<td>246</td>
<td>262701</td>
<td>-</td>
<td>19</td>
<td>36</td>
<td>298</td>
<td>45.3 (6.5)</td>
<td>37.9 (6.6)</td>
<td>1258 (9.6)</td>
</tr>
<tr>
<td>5 ± 4</td>
<td>298</td>
<td>262701</td>
<td>262701</td>
<td>-</td>
<td>19</td>
<td>43</td>
<td>298</td>
<td>38.0 (6.5)</td>
<td>38.0 (6.5)</td>
<td>1206 (9.2)</td>
</tr>
<tr>
<td>4 ± 5</td>
<td>1206 (9.2)</td>
<td></td>
<td></td>
<td>-</td>
<td>19</td>
<td>43</td>
<td>298</td>
<td>1999* (6.4)</td>
<td>66.9 (12.3)</td>
<td>2221 (17.0)</td>
</tr>
<tr>
<td>3 ± 6</td>
<td>612 (3.1)</td>
<td></td>
<td></td>
<td>-</td>
<td>19</td>
<td>36</td>
<td>298</td>
<td>46.8 (6.8)</td>
<td>36.2 (7.5)</td>
<td>1298 (9.9)</td>
</tr>
<tr>
<td>2 ± 7</td>
<td>590 (2.9)</td>
<td></td>
<td></td>
<td>-</td>
<td>19</td>
<td>36</td>
<td>298</td>
<td>38.0 (6.5)</td>
<td>38.0 (6.5)</td>
<td>1206 (9.2)</td>
</tr>
<tr>
<td>1 ± 8</td>
<td>581 (2.9)</td>
<td></td>
<td></td>
<td>-</td>
<td>19</td>
<td>36</td>
<td>298</td>
<td>1999* (6.4)</td>
<td>66.9 (12.3)</td>
<td>2221 (17.0)</td>
</tr>
</tbody>
</table>

* The peak height. The highest peak was arbitrarily scaled to a height of 999. The minimum of the map was -285. The average of the map values was 244 and the standard deviation was 118.

† The numbers in parentheses represent the peak heights in terms of the standard deviation of the map values, calculated as peak height - average/standard deviation from average.

‡ The correct solution.

§ Peak height on arbitrary scale. The minimum of the map was -499. The average was 0 and the standard deviation from average was 131.

Fig. 1. Contouring of a section of the rotation-function map. The starting contour level was 500 (the highest peak was scaled to a height of 999). The increment was 50. This plot was output directly from the program GLRF.

I thank Professor Michael G. Rossmann for encouragement and helpful discussions and Dr Joseph Gilboa for suggesting the name of the program suite. I thank Joseph Gilboa and many members of the crystallography groups at Purdue for trying out earlier versions of the programs.

### References


