

LIGHTFOOT, P., TREMAYNE, M., HARRIS, K. D. M. & BRUCE, P. G. (1992). *J. Chem. Soc. Chem. Commun.* pp. 1012–1013.
 MCCUSKER, L. (1988). *J. Appl. Cryst.* **21**, 305–310.
 PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. & VETTERLING, W. T. (1986). *Numerical Recipes*, pp. 484–488. Cambridge Univ. Press.

SJÖLIN, L., PRINCE, E., SVENSON, L. A. & GILLILAND, G. L. (1991). *Acta Cryst.* **A47**, 216–233.
 TREMAYNE, M., LIGHTFOOT, P., MEHTA, M. A., BRUCE, P. G., HARRIS, K. D. M., SHANKLAND, K., GILMORE, C. J. & BRICOGNE, G. (1992). *J. Solid State Chem.* **100**, 191–196.

Acta Cryst. (1993). **A49**, 501–504

A Rotation Function with Increased Signal Size

BY JORDI RIUS AND CARLES MIRAVITLLES

CSIC, Institut de Ciència de Materials de Barcelona, 08193 Cerdanyola-Bellaterra, Catalunya, Spain

(Received 12 May 1992; accepted 15 October 1992)

Abstract

The most widespread application of the rotation function is the determination of the relative orientation of a given search fragment in the unit cell of an unknown crystal structure [Rossmann & Blow (1962). *Acta Cryst.* **15**, 24–31]. Here a modification is presented of the rotation function for this specific application, which exploits the information of the intensity data more effectively, thus leading to a higher signal size with the same computing cost.

1. Introduction

Although the primary phasing of intensity data from small equal-atom structures and from macromolecular compounds is normally carried out using either direct methods or the multiple isomorphous-replacement technique, molecular-replacement techniques are increasingly used when a suitable search fragment (or model) is available. Besides the crystal symmetry, the principal factors determining the success of molecular-replacement methods are, on the one hand, the size, the form and the accuracy of the search fragment and, on the other hand, the number and reciprocal-space distribution of the measured intensities. In general, the larger and more accurately known the fragment is, the less drastic are the requirements imposed on the intensity data.

As is well known, the real-space formulation of the rotation function of Rossmann & Blow (1962) for the case where a suitable search model is available is

$$R(\Omega) \propto \int_U P_o(\mathbf{u}) P_{\text{model}}(\Omega\mathbf{u}) \, d\mathbf{u}. \quad (1)$$

The integral in (1) measures the agreement of the Patterson function of the unknown crystal structure (P_o) with the rotated Patterson function of the isolated search model (P_{model}) in a region U around the origin of the unit cell. The symbol Ω denotes a rotation operator that rotates the coordinate system of

the search model with respect to that of the unknown crystal structure. $R(\Omega)$ will have a large maximum when the two Patterson functions are brought into maximum coincidence. If U corresponds to the whole unit cell, (1) can be expressed in reciprocal space as the summation

$$\sum_{\mathbf{H}} |F_o(\mathbf{H})|^2 |S(\mathbf{H}, \Omega)|^2, \quad (2)$$

where $|F_o(\mathbf{H})|^2$ and $|S(\mathbf{H})|^2$ are, respectively, the Fourier coefficients of the observed and the model Patterson functions (Tollin & Cochran, 1964). $|S(\mathbf{H})|^2$ can be written in the form

$$|S(\mathbf{H})|^2 = \sum_{j=1}^n \sum_{k=1}^n Z_j Z_k \cos(2\pi \mathbf{H} \cdot \mathbf{r}_{jk}) \quad (3)$$

with n being the number of atoms of the fragment, Z_j being the atomic number of the j th atom and \mathbf{r}_{jk} denoting the difference vector $\mathbf{r}_j - \mathbf{r}_k$, where \mathbf{r}_j is the position vector of the j th atom referred to a fixed local origin.

Inspection of (2) reveals that the contribution to the \mathbf{H} summation of those terms with small $|F_o(\mathbf{H})|^2$ values is not significant. Consequently, it seems reasonable to expect an increased signal size if the rotation function (2) is modified to include additionally the significant contribution of the weak reflections. In practice, this modification can be useful in those cases where only a small intensity data set is available, as is typical for low-resolution X-ray powder diffraction data of organic compounds (Rius & Miravittles, 1988; Wilson & Wadsworth, 1990; Rius, Miravittles, Molins, Crespo & Veciana, 1990; Amigó, Ochando, Abarca, Ballesteros & Rius, 1992). Owing to the reduced number of available intensities, the most difficult step is the rotation search. The subsequent fragment positioning is greatly simplified with the combined use of translation and packing functions (Harada, Lifchitz, Berthou & Jolles, 1981; Stubbs & Huber, 1991) as well as with the calculation of the R

value for the higher-ranked solutions found with the τ translation function (Rius & Miravittles, 1986).

2. Definition of the rotation function incorporating weak reflections

Let $F(\mathbf{H})$ represent the structure factor of a structure consisting of N point atoms in the unit cell. Assume that some of them are in a correctly oriented but randomly positioned fragment (plus the symmetrically related ones) and that the rest of the atoms are randomly distributed in the cell. On the basis of this prior information, the expected value of $|F(\mathbf{H})|^2$ is given by the expression (see the Appendix)

$$|F(\mathbf{H})|_{\text{exp}}^2 = \varepsilon(\mathbf{H}) \sum_{j=1}^N Z_j^2 + 2\varepsilon(\mathbf{H}) \sum_{p=1}^m \sum_{j=1}^n \sum_{\substack{k=1 \\ k>j}}^n Z_j Z_k \times \cos(2\pi \mathbf{H} \mathbf{R}_p \cdot \mathbf{r}_{jk}), \quad (4)$$

where $\varepsilon(\mathbf{H})$ is an enhancement factor dependent upon the reflection type and the space group (Stewart, Karle, Iwasaki & Ito, 1977) and m is the number of point-group symmetry operators. Expression (4) allows the calculation of the value of $|F(\mathbf{H}, \Omega)|_{\text{exp}}^2$ for each orientation Ω of the model specified by the set of intrafragment vectors.

Now let $D(\mathbf{H}, \Omega)$ be defined as

$$D(\mathbf{H}, \Omega) = |F(\mathbf{H}, \Omega)|_{\text{exp}}^2 / \varepsilon(\mathbf{H}) \quad (5)$$

and let \mathbf{h} and \mathbf{l} represent those reflections having sharpened $|F_o(\mathbf{H})|^2 / \varepsilon(\mathbf{H})$ values greater and smaller, respectively, than the cut-off value F_{min}^2 . With (4) and (5) taken into account, the rotation function (2) can then be approximated as the summation

$$\begin{aligned} & \sum_{\mathbf{h}} |F_o(\mathbf{h})|^2 D(\mathbf{h}, \Omega) \\ &= c \sum_{j=1}^N Z_j^2 + 2m \sum_{j=1}^n \sum_{\substack{k=1 \\ k>j}}^n Z_j Z_k \\ & \quad \times \sum_{\mathbf{h}} |F_o(\mathbf{h})|^2 \cos[2\pi \mathbf{h} \cdot \mathbf{r}_{jk}(\Omega)], \quad (6) \end{aligned}$$

where

$$c = \sum_{\mathbf{h}} |F_o(\mathbf{h})|^2. \quad (7)$$

Similarly, the weighted mean value of $D(\mathbf{l}, \Omega)$ is given by

$$\begin{aligned} & \sum_{\mathbf{l}} w(\mathbf{l}) D(\mathbf{l}, \Omega) / d \\ &= \sum_{j=1}^N Z_j^2 + (2m/d) \sum_{j=1}^n \sum_{\substack{k=1 \\ k>j}}^n Z_j Z_k \\ & \quad \times \sum_{\mathbf{l}} w(\mathbf{l}) \cos[2\pi \mathbf{l} \cdot \mathbf{r}_{jk}(\Omega)], \quad (8) \end{aligned}$$

where

$$d = \sum_{\mathbf{l}} w(\mathbf{l}). \quad (9)$$

Since \mathbf{l} represents the weak reflections, (8) will have a minimum for the correct model orientation, *i.e.* the summation of cosine terms will be predominantly negative. For incorrect orientations, however, this summation will tend to zero. The weights $w(\mathbf{l})$ are assumed to be unity in the present paper, but other weighting schemes can also be contemplated, *e.g.* $w(\mathbf{l}) = \|F_o(\mathbf{l})\|^2 - |F|_{\text{av}}^2$ with $|F|_{\text{av}}^2$ being the average value of $|F_o(\mathbf{H})|^2$ over all \mathbf{H} .

Next, we can define the rotation function $\mathcal{R}(\Omega)$ actively incorporating the weak reflections. It results from the combination of (6) with (8):

$$\begin{aligned} \mathcal{R}(\Omega) &= \sum_{\mathbf{h}} |F_o(\mathbf{h})|^2 D(\mathbf{h}, \Omega) \\ & \quad - (c/d) \sum_{\mathbf{l}} w(\mathbf{l}) D(\mathbf{l}, \Omega) \quad (10a) \end{aligned}$$

$$\begin{aligned} &= 2m \sum_{j=1}^n \sum_{\substack{k=1 \\ k>j}}^n Z_j Z_k \left\{ \sum_{\mathbf{h}} |F_o(\mathbf{h})|^2 \right. \\ & \quad \times \cos[2\pi \mathbf{h} \cdot \mathbf{r}_{jk}(\Omega)] - (c/d) \sum_{\mathbf{l}} w(\mathbf{l}) \\ & \quad \left. \times \cos[2\pi \mathbf{l} \cdot \mathbf{r}_{jk}(\Omega)] \right\} \quad (10b) \end{aligned}$$

and, like $R(\Omega)$, this function will have a maximum for the correct model orientation. The higher signal size comes from the larger number of observations involved in the calculation of $\mathcal{R}(\Omega)$ as compared with $R(\Omega)$.

Finally, if δP_o is a modified Patterson function with Fourier coefficients $|F_o(\mathbf{h})|^2$ and $-w(\mathbf{l})c/d$, then (10b) reduces, when proportionality constants are ignored, to

$$\mathcal{R}(\Omega) = \sum_{j=1}^n \sum_{\substack{k=1 \\ k>j}}^n Z_j Z_k \delta P_o[\mathbf{r}_{jk}(\Omega)]. \quad (10c)$$

The symbol δ has been introduced to emphasize the 'difference' nature of this Patterson function. Since δP_o vanishes at the origin, the difficulty arising from the presence of intrafragment vectors longer than a lattice constant, and the subsequent possible overlap of these vectors with the Patterson origin peaks of neighbouring cells, is minimized.

3. Test calculations

This modification has been tested with the structures of cortisone (Declercq, Germain & Van Meerse, 1972), $C_{21}H_{28}O_5$, space group $P2_12_12_1$, $Z = 4$; loganin (Jones, Sheldrick, Glösenkamp & Tietze, 1980), $C_{17}H_{26}O_{10}$, space group $P2_12_12_1$, $Z = 4$; and NEWQB (Grigg, Kemp, Sheldrick & Trotter, 1978),

$C_{24}H_{20}N_2O_5$, space group $P\bar{1}$, $Z = 4$. In all three cases, the δP_o function was calculated using sharpened intensity data at 1.00 Å resolution and with a grid size of 0.25 Å to avoid the need for interpolation. The systematically absent reflections have been removed in each case. All calculations were performed with an adapted version of *ROTSEARCH* (Rius & Miravittles, 1987).

The signal size H of the tested rotation function \mathcal{R} is defined as $[\mathcal{R}(\Omega) - \bar{\mathcal{R}}]/\sigma$, where $\bar{\mathcal{R}}$ and σ^2 are, respectively, the mean value and the associated variance obtained from averaging $\mathcal{R}(\Omega)$ over all the 33 552 scanned model orientations [mean rotation increment = 7°].

The cortisone model (scattering power 42%) was obtained by arbitrary rotation of the unrefined coordinates taken from an E map of the structure. The atomic coordinates of the loganin and NEWQB fragments with respective scattering powers of 28 and 12% were derived by application of an arbitrary rotation to the published ones. The minimum and maximum lengths of the selected intrafragment vectors are 1.9 and 6.0 Å. 58 vectors were studied for the cortisone model, 26 for loganin and 19 for NEWQB.

Since the search fragments of cortisone and loganin are relatively large, the first rotation-function solutions produced by (6) and (10c) are the correct ones. For NEWQB, the correct orientation always corre-

sponds to one of the first three solutions, depending on the value of F_{\min} used. The rotation angles Ω obtained with (10c) are not significantly different from those obtained with (6). Inspection of Fig. 1 clearly indicates that, as expected, the rotation function (10c) yields a higher signal size than the rotation function (6).

This work was supported by the DGICYT (Project PB89-0036).

APPENDIX

The expected value of $|F(\mathbf{H})|^2$, given a well oriented but randomly positioned group

Let us assume, for simplicity, that the unit cell contains only one symmetrically independent molecular group, from which the orientation but not the position \mathbf{r} is known. If $\varepsilon(\mathbf{H})$ is the enhancement factor, which is the number of rotation matrices for which $\mathbf{H}\mathbf{R} = \mathbf{H}$ and $\mathbf{H} \cdot \mathbf{t} = 0 \pmod{1}$ (Hovmöller, 1980), then the squared amplitude of the structure factor can be written in the form

$$|F(\mathbf{H})|^2 = \left| \varepsilon(\mathbf{H}) \sum_{p=1}^{m/\varepsilon_{\mathbf{H}}} S_p(\mathbf{H}) \times \exp [i2\pi\mathbf{H} \cdot (R_p\mathbf{r} + \mathbf{t}_p)] \right|^2, \quad (\text{A.1})$$

where S_p is the structure factor of the p th molecular group and m is the number of point-group symmetry operators. For general reflections, $\varepsilon(\mathbf{H})$ is unity. Equation (A.1) then becomes

$$|F(\mathbf{H})|^2 = \varepsilon(\mathbf{H})^2 \sum_{p=1}^{m/\varepsilon_{\mathbf{H}}} \sum_{q=1}^{m/\varepsilon_{\mathbf{H}}} S_p(\mathbf{H}) S_q(\mathbf{H})^* \times \exp [i2\pi\mathbf{H} \cdot (\mathbf{t}_p - \mathbf{t}_q)] \exp [i2\pi\mathbf{H}_{pq} \cdot \mathbf{r}] \quad (\text{A.2})$$

with $\mathbf{H}_{pq} = \mathbf{H}(R_p - R_q)$. Expression (A.2) (Harada *et al.*, 1981; Rius & Miravittles, 1986) allows the computation of $|F(\mathbf{H})|^2$ as a continuous function from \mathbf{r} . Hence, the expected value of $|F(\mathbf{H})|^2$ can be obtained from

$$|F(\mathbf{H})|^2_{\text{exp}} = (1/V) \int_V |F(\mathbf{H}, \mathbf{r})|^2 d\mathbf{r} \\ = \varepsilon(\mathbf{H})^2 \sum_{p=1}^{m/\varepsilon_{\mathbf{H}}} \sum_{q=1}^{m/\varepsilon_{\mathbf{H}}} S_p(\mathbf{H}) S_q(\mathbf{H})^* \times \exp [i2\pi\mathbf{H} \cdot (\mathbf{t}_p - \mathbf{t}_q)] I(\mathbf{H}_{pq}) \quad (\text{A.3})$$

with

$$I(\mathbf{H}_{pq}) = (1/V) \int_V \exp [i2\pi\mathbf{H}_{pq} \cdot \mathbf{r}] d\mathbf{r}. \quad (\text{A.4})$$

The integral $I(\mathbf{H}_{pq})$ is always zero unless $\mathbf{H}_{pq} = 0$, i.e. when $p = q$, in which case it is unity. Consequently,

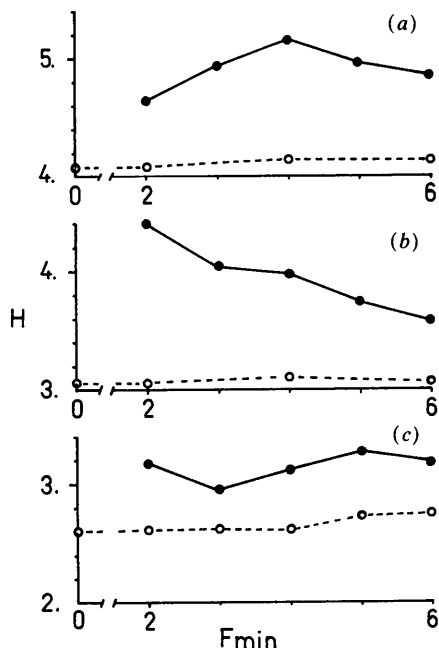


Fig. 1. Application to three test examples of the rotation function computed with (10c) (solid line) and with (6) (dashed line) for different cut-off values F_{\min} . H : signal size for the correct solution given in units of σ above the mean. (a) Cortisone, model scattering power (m.s.p.) 42%; (b) loganin, m.s.p. 28%; (c) NEWQB, m.s.p. 12%. Notice that in all three examples the signal size of (10c) is the highest.

(A.3) reduces to

$$\begin{aligned} |F(\mathbf{H})|_{\text{exp}}^2 &= \varepsilon(\mathbf{H})^2 \sum_{p=1}^{m/\varepsilon_{\mathbf{H}}} |S_p(\mathbf{H})|^2 \\ &= \varepsilon(\mathbf{H}) \sum_{p=1}^m |S_p(\mathbf{H})|^2. \end{aligned} \quad (\text{A.5})$$

Now imagine that the complete molecular group of size N/m is replaced by a smaller fragment of size n , so that the unit cell also contains $m(N/m - n)$ randomly distributed atoms. If (A.5) is modified to include their contribution, it follows that

$$|F(\mathbf{H})|_{\text{exp}}^2 = \varepsilon(\mathbf{H}) \sum_{j=1}^{N-nm} Z_j^2 + \varepsilon(\mathbf{H}) \sum_{p=1}^m |S_p(\mathbf{H})|^2. \quad (\text{A.6})$$

Finally, expression of $|S_p(\mathbf{H})|^2$ in terms of the interatomic vectors $\mathbf{r}_j - \mathbf{r}_k$ results in the desired expression (4).

Acta Cryst. (1993). **A49**, 504–513

Molecular Dynamics in Refinement against Fiber Diffraction Data

BY HONG WANG AND GERALD STUBBS

Department of Molecular Biology, Vanderbilt University, Nashville, TN 37235, USA

(Received 5 June 1992; accepted 22 October 1992)

Abstract

The molecular dynamics (MD) method has been adapted for refinement of the structures of helical macromolecular aggregates against X-ray fiber diffraction data. To test the effectiveness of the method, refinements of the tobacco mosaic virus structure were carried out against a set of simulated fiber diffraction intensities using the MD method as well as the conventional restrained least-squares (RLS) method. The MD refinement converged to a very low R factor and produced a structure with generally satisfactory stereochemistry, while the RLS refinement was trapped at a local energy minimum with a larger R factor. Results suggest that the effective experimental radius of convergence of the MD method is significantly greater than that of the RLS method. Even when the initial structure is too far from the true structure to allow direct refinement, the MD method is able to find local minima that resemble the true structure sufficiently to allow improved phasing and thus lead to interpretable difference maps for model rebuilding.

Introduction

Fiber diffraction has been a very effective method for the determination of the molecular structures of filamentous macromolecular assemblies such as viruses, cytoskeletal elements, nucleic acids and polysaccharides. The component parts of these assemblies are often difficult or impossible to crystallize because of their natural tendency to form filaments and, even if they can be crystallized, the crystal structures do not usually reveal the important intermolecular interactions that are often the most biologically significant aspect of the molecular structure. Fiber diffraction is therefore the preferred method of analysis for these systems.

The defining property of a fiber diffraction specimen is that the diffracting units are randomly oriented about an axis, the fiber axis. Specimens may in fact be fibers, oriented gels or even stacks of sheet-like structures such as membranes. As a result of the random orientation about the axis, the fiber diffraction pattern is the cylindrical average of the diffraction pattern to be expected from one particle (in the

References

- AMIGÓ, J. M., OCHANDO, L., ABARCA, B., BALLESTEROS, R. & RIUS, J. (1992). *Mater. Sci. Forum.* In the press.
- DECLERCQ, J. P., GERMAIN, G. & VAN MEERSSCHE, M. (1972). *Cryst. Struct. Commun.* **1**, 13–15.
- GRIGG, R., KEMP, J., SHELDRIK, G. & TROTTER, J. (1978). *J. Chem. Soc. Chem. Commun.* pp. 1109–1111.
- HARADA, Y., LIFCHITZ, A., BERTHOU, J. & JOLLES, P. (1981). *Acta Cryst.* **A37**, 398–406.
- HOVMÖLLER, S. (1980). *Rotation Matrices and Translation Vectors.* IUCr/Univ. College Cardiff Press.
- JONES, P. G., SHELDRIK, G. M., GLÜSENKAMP, K. H. & TIETZE, L. F. (1980). *Acta Cryst.* **B36**, 481–483.
- RIUS, J. & MIRAVITLLES, C. (1986). *Acta Cryst.* **A42**, 402–404.
- RIUS, J. & MIRAVITLLES, C. (1987). *J. Appl. Cryst.* **20**, 261–264.
- RIUS, J. & MIRAVITLLES, C. (1988). *J. Appl. Cryst.* **21**, 224–227.
- RIUS, J., MIRAVITLLES, C., MOLINS, E., CRESPO, M. & VECIANA, J. (1990). *Mol. Cryst. Liq. Cryst.* **187**, 155–161.
- ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24–31.
- STEWART, J. M., KARLE, J., IWASAKI, H. & ITO, T. (1977). *Acta Cryst.* **A33**, 519.
- STUBBS, M. & HUBER, R. (1991). *Acta Cryst.* **A47**, 521–526.
- TOLLIN, P. & COCHRAN, W. (1964). *Acta Cryst.* **17**, 1322–1324.
- WILSON, C. C. & WADSWORTH, J. W. (1990). *Acta Cryst.* **A46**, 258–262.