SHORT COMMUNICATIONS

Acta Cryst. (1997). D53, 789-791

A weighted rotation function

D. I. SVERGUN† *at European Molecular Biology Laboratory, Hamburg Outstation, EMBL c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany. E-mail: svergun@embl-hamburg.de*

(*Received 25 September 1996; accepted 22 April 1997*)

Abstract

A simple probability expression is derived to discriminate the intermolecular vectors in the cross rotation function. The radial weighting function introduced takes into account both particle anisometry and crystal packing and can be incorporated into existing molecular replacement packages.

1. Introduction

The cross-rotation function (CRF, Rossmann & Blow, 1962) is the main tool for orienting a known (search) molecule in a crystal unit cell, the key step in solving macromolecular structures by the molecular replacement technique (Rossmann, 1972). This technique is based on the assumption that the structure of the search molecule is similar to that of the actual (target) molecule. The CRF is evaluated as a correlation between the Patterson function of the target molecule $P_{obs}(\mathbf{r})$ and that of the rotating search molecule $P_{cal}(\mathbf{r})$,

$$R(\Omega) = \int_{r < R_{cut}} P_{obs}(\mathbf{r}) P_{cal}(\mathbf{r}, \Omega) d\mathbf{r}, \qquad (1)$$

and the maximization of $R(\Omega)$ yields the rotational parameters [*e.g.* Euler angles, $\Omega = (\alpha, \beta, \gamma)$] providing the optimum orientation of the search molecule in the unit cell. The correlation, should, strictly speaking, be based only on the intramolecular vectors and the integration volume is thus restricted by a cutoff value $R_{cut} \leq D$ where D is the particle diameter. $P_{cal}(\mathbf{r})$ is usually evaluated in a sufficiently large artificial $P1$ cell and any cutoff with $R_{cut} \leq D$ eliminates the intermolecular contributions. For the experimental $P_{obs}(\mathbf{r})$, the discrimination between the intra- and intermolecular vectors is, however, not trivial. A compromise must be found taking into account that R_{cut} should be kept small to exclude the unwanted intermolecular vectors, but that by lowering R_{cut} one reduces the number of Patterson peaks in the integration volume and thus worsens the signal-to-noise ratio in the CRF. The choice of R_{cut} can be critical for the entire procedure of molecular replacement, but there is no consensus on how to determine this value. Its estimates range from the geometrical mean of the ellipsoid semiaxes (Lifchitz, 1983) to 0.8 times the diameter of the molecule (Blow, 1985). A rule of thumb proposed by Joynson, North, Sarma, Dickerson & Steinrauf (1970) and further recommended as a starting estimate by Lifchitz (1983) and Tickle & Driessen (1996) is to use half the diameter of the molecule.

A distinct cutoff between the intra- and intermolecular vectors in $P_{obs}(\mathbf{r})$ does not exist as each vector of magnitude $0 < r \leq D$ can with some probability join points belonging either to the same or to two different particles. As the orientation of the target molecule is unknown, this probability cannot depend on the direction of the vector, but only on its magnitude. The probability for the given vector $\mathbf{r}$ to be intramolecular decreases with r and depends both on the particle shape and packing conditions. Below a simple expression for this probability is derived which allows to discriminate the intermolecular vectors in the CRF.

2. Probability discrimination

The Patterson function itself can be treated in probability terms. Consider a uniform particle with a dimensionless unitary density,

$$\rho(\mathbf{r}) = \begin{cases} 1, & \mathbf{r} \text{ inside the particle} \\ 0, & \text{elsewhere} \end{cases}, \qquad (2)$$

representing the molecular envelope of the search particle. The spherically averaged normalized self-Patterson function of the envelope,

$$\gamma_{cal}(r) = \langle P_{cal}(\mathbf{r}) \rangle_{\Omega} / P_{cal}(0) = \frac{1}{V} \langle \int \rho(\mathbf{u}) \rho(\mathbf{u} + \mathbf{r}) d\mathbf{u} \rangle_{\Omega}, \qquad (3)$$

gives the probability that a point at a distance r from a randomly chosen point inside the envelope also belongs to the particle (Feigin & Svergun, 1987, ch. 2). Here, $\langle \rangle_{\Omega}$ denotes the average over all particle orientations and $P_{cal}(0) = V$ where V is the volume of the envelope. The function $\gamma_{cal}(r)$ decreases from 1 at $r = 0$ to 0 at $r = D$ with a slope depending on the envelope geometry [the more anisometric the molecule, the steeper the slope – this was also noted by Blow (1985)].

Consider now the probability function of the envelope of the target molecule in the crystal $\gamma_{obs}(r) = \langle P_{obs}(\mathbf{r}) \rangle_{\Omega} / P_{obs}(0)$, where $P_{obs}(0) = kV$ (k is the number of molecules in the unit cell). As the envelopes of the target and search molecules are similar, $\gamma_{obs}(r)$ will coincide with $\gamma_{cal}(r)$ at small r (intramolecular vectors). With increasing r, $\gamma_{obs}(r)$ decays more slowly than $\gamma_{cal}(r)$ due to the intermolecular vectors, tending to $1 - v$ rather than to zero (here, v the is fraction of the unit cell occupied by the solvent). The probability of a vector of magnitude r in $P_{obs}(\mathbf{r})$ to be intramolecular is given by the ratio $W(r) = \gamma_{cal}(r)/\gamma_{obs}(r)$. Thus, introducing $W(r)$ in the CRF as a weighting function,

$$R(\Omega) = \int_{r < R_{cut}} W(r) P_{obs}(\mathbf{r}) P_{cal}(\mathbf{r}, \Omega) d\mathbf{r}, \qquad (4)$$

would yield a 'probability discrimination' between the intra- and intermolecular vectors.

† On leave from the Institute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, 117333 Moscow, Russia.

3. Example

Fig. 1 presents the probability distributions of ribonuclease A evaluated for an isolated molecule and in the crystal environment. The protein of molecular weight 13.7 kDa and of the maximum diameter 50.4 Å, crystallizes in the space group $P2$ with $a = 30.4$, $b = 38.4$ $c = 53.2$ Å, $\beta = 105.7°$. The molecular envelope was calculated from its structure refined at 1.1 Å resolution (F. Sica, R. Berisio, Universit di Napoli 'Federico II' and EMBL, Hamburg Outstation, personal communication) using the program *CRYSOL* (Svergun, Barberato & Koch, 1995). The probability function of the isolated envelope $\gamma_{cal}(r)$ is represented by curve 1. Patterson functions of the molecule in a large $P1$ unit cell and in the actual crystal were computed from the complete set of reflections down to 3 Å resolution on a real space grid with a step of 1 Å. Their spherical averages were evaluated and normalized to the corresponding $P(0)$ values according to (2). Curve 2 (large $P1$ cell) agrees well with the theoretical $\gamma_{cal}(r)$ of the envelope whereas curve 3 (actual cell) gives an estimate of $\gamma_{obs}(r)$. The latter lies systematically above $\gamma_{cal}(r)$ and oscillates around $(1-\nu) = 0.56$ starting from $r \simeq 10$ Å. Oscillations at small r in the averaged Patterson maps arise from the atomicity of the structure (note that this range is usually omitted in the CRF to ensure the removal of the origin peak). The weighting function $W(r)$ (curve 4) was evaluated as the ratio of curve 3 and curve 2.

To estimate the influence of the weighting function a test rotation search using the two Patterson functions was performed. The refined model was taken as the search molecule so that the true orientation corresponded to the Euler angles $\alpha = \beta = \gamma = 0°$. The search and target Patterson functions were represented using spherical harmonics (Crowther, 1972). The radial functions of this expansion were evaluated numerically from the three-dimensional Pattersons taking into account the spherical harmonics of orders $4 \le l \le 26$. To remove the origin effects, the integration was performed in the range $R_{min} \le r \le R_{cut}$ with $R_{min} = 6$ Å. A qualitative illustration in Fig. 2 presents the one-dimensional standard (1) and weighted (4) CRF as functions of the Euler rotation angles α and β for $R_{cut} = 25$ (the value suggested by the rule of thumb above). Each CRF was normalized to the magnitude of its true peak A_0 at $\alpha = \beta = \gamma = 0°$. Negative excursions of the CRF are caused by the omission of the harmonics with $l = 0$ and 2, whereas the peak at $\beta = \pi$ is because of the symmetry of the space group. It is seen that the standard CRF contains a noticeable intermolecular contribution leading to false peaks which are suppressed by the weighting procedure.

To obtain quantitative estimates, three-dimensional searches were performed at different R_{cut} values. The CRF were evaluated with an angular step of 4° excluding the range of 20° around the true peak. The values of contrast (ratio of the true peak A_0 to the highest false peak A_1) and signal-to-noise (ratio of the true peak to the r.m.s. of the CRF) are presented in Table 1. The weighted CRF (WCRF) provides both better

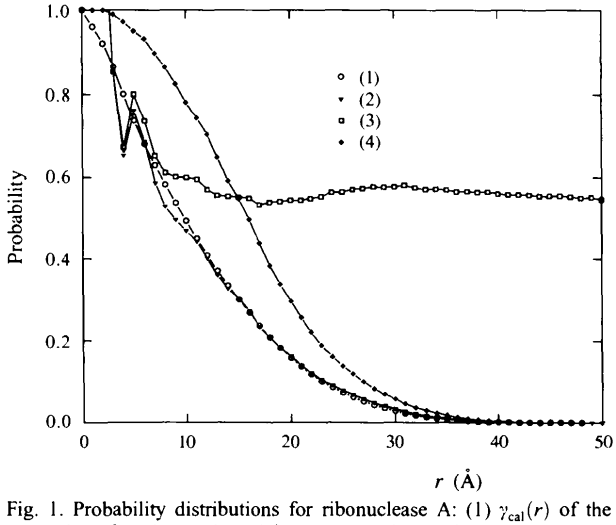

Fig. 1. Probability distributions for ribonuclease A: (1) $\gamma_{cal}(r)$ of the envelope function evaluated from the atomic coordinates, (2) and (3) spherically averaged Patterson functions in a large $P1$ unit cell and in the actual one, respectively; (4) weighting function $W(r)$.

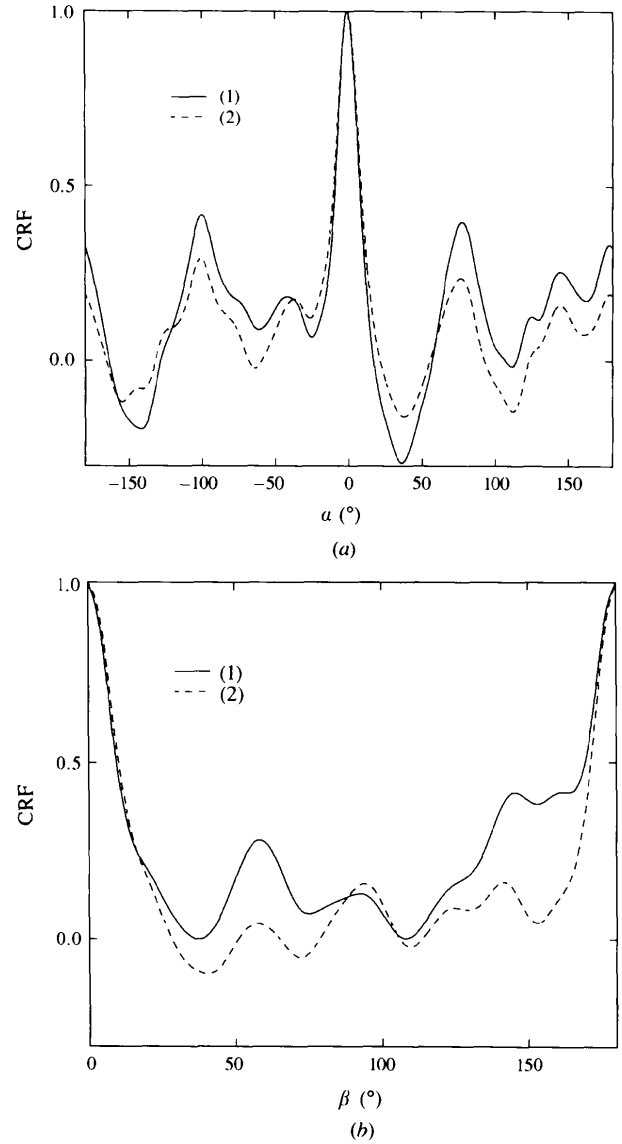

Fig. 2. Comparison of the normalized standard (1) and weighted (2) CRF of ribonuclease A at $R_{cut} = 25$ Å: (*a*) rotation around the Euler angle α at $\beta = \gamma = 0°$, (*b*) rotation around β at $\alpha = \gamma = 0°$.

Table 1. *Contrasts and signal/noise ratios for the standard and weighted CRF*

	Standard CRF		Weighted CRF	
R_{cut} (Å)	A_0/A_1	A_0/r.m.s.	A_0/A_1	A_0/r.m.s.
15.0	2.00	8.06	2.13	8.47
17.5	2.29	9.00	2.32	9.35
20.0	2.56	9.17	2.70	9.80
22.5	2.04	7.52	2.56	9.43
25.0	1.37	4.33	2.17	8.00
27.5	0.89	2.80	1.96	7.04
30.0	0.21	0.75	1.79	5.85

contrast and better signal-to-noise ratio at all cutoffs, and these parameters for the WCRF are much less dependent on R_{cut} than those for the standard CRF.

4. Discussion

Practical application of the WCRF requires reliable procedures to evaluate the probability distributions. Given the search molecule, $\gamma_{cal}(r)$ is readily computed either from the isolated envelope (programs available from the author), or by averaging the Patterson function in a large $P1$ unit cell. Alternatively, it can be calculated as a Fourier transform from the spherically averaged intensity (D. Blow, personal communication). Direct evaluation of $\gamma_{obs}(r)$ from the experimental data is more difficult because of termination effects caused by the omission of the low-resolution reflections [in particular, $P(0)$ is ill defined]. However, as seen from Fig. 1, a reasonable estimate of $\gamma_{obs}(r)$ is obtained assuming that,

$$\gamma_{obs}(r) = \begin{cases} \gamma_{cal}(r) & \text{if } \gamma_{cal}(r) > 1 - \nu \\ 1 - \nu & \text{if } \gamma_{cal}(r) \leq 1 - \nu \end{cases} . \qquad (5)$$

For the above example this would mean that $W(r) = 1$ for $r < 10$ Å which yields only marginal differences.

Theoretically, the integration in (4) can be extended to $R_{cut} = D$. Practically, although the WCRF is much less susceptible to the choice of R_{cut}, a proper cutoff still improves the contrast and also speeds up the calculations. The weighting function $W(r)$ provides quantitative information on the probability level for the selection of R_{cut}. From the above example (Fig. 1 and Table 1) one can judge that the integration volume containing the vectors with $W(r) \geq 0.3$ (this level corresponds to $R_{cut} \simeq 20$ Å) yields the optimum cutoff.

The probability discrimination in the WCRF (4) accounts both for the geometry of the search molecule and for the packing conditions. This is more adequate than the simple limitation on the integration volume in (1). Clearly, the theoretical advantages of the WCRF have to be verified in practice. As the function $W(r)$ is radially symmetric, it can easily be incorporated in the existing molecular replacement packages (*e.g.* Fitzgerald, 1988; Brünger, 1992; Navaza, 1994) as a monopole contribution.

The author thanks Drs M. H. J. Koch and V. S. Lamzin for useful discussions and Dr F. Sica and R. Berisio for providing the data on the ribonuclease A prior to their deposition in the Protein Data Bank (Bernstein *et al.*, 1977).

References

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Blow, D. M. (1985). In *Proceedings of the Daresbury Study Weekend February 15–16*, edited by P. A. Machin, pp. 2–7. Warrington: Daresbury Laboratory.

Brünger, A. T. (1992). *X-PLOR, Version* 3.1, *A System for X-Ray Crystallography and NMR*. New Haven, Connecticut, USA.

Crowther, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. Rossmann, pp. 173–178. New York: Gordon and Breach.

Feigin, L. A. & Svergun, D. I. (1987). *Structure Analysis by Small-Angle X-ray and Neutron Scattering*. New York: Plenum Press.

Fitzgerald, P. M. (1988). *J. Appl. Cryst.* **21**, 273–278.

Joynson, M. A., North, A. C. T., Sarma, V. R., Dickerson, R. E. & Steinrauf, L. K. (1970). *J. Mol. Biol.* **50**, 137–142.

Lifchitz, A. (1983). *Acta Cryst.* **A39**, 130–139.

Navaza, J. (1994). *Acta Cryst.* **A50**, 157 163.

Rossmann, M. G. (1972). *The Molecular Replacement Method*. New York: Gordon and Breach.

Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.

Svergun, D. I., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.

Tickle, I. J. & Driessen, H. P. C. (1996). *Methods in Molecular Biology*, Vol. 56, *Crystallographic Methods and Protocols*, edited by C. Jones, B. Mullow & M. Sanderson, pp. 173-203. Totowa, NJ: Humana Press.