<center>13. MOLECULAR REPLACEMENT</center>

13.2.3.3. *Strategies*

Since the first attempts to detect subunits within the crystal asymmetric unit, a great deal of experience has been gained and the results collected in several treatises on molecular replacement (Rossmann, 1972; Machin, 1985; Dodson *et al.*, 1992; Carter & Sweet, 1997).

The success of the rotation function relies in part on the choice of the domain of integration, *i.e.*, inner and outer radii of the spherical domain, and limits on data resolution. The outer radius is chosen so as to maximize the ratio between the number of intramolecular and intermolecular vectors. A typical value of this radius is 50–75% of the subunit's diameter for spherical molecules. The choice of the inner radius is less crucial, provided that the origin peaks are subtracted from the Patterson functions; it is often set to zero.

The resolution range of the selected diffraction data depends on whether we are computing self- or cross-rotation functions. For self rotations, data up to the highest available resolution may be used. For cross rotations, the high-resolution limit is dictated by the degree of similarity between the probe and the actual molecules that constitute the crystal. The low-resolution limit is usually chosen so as to skip the solvent contribution.

Another parameter of interest concerns the angular resolution, which is rather specific to the fast rotation function. Although it was derived by expanding the interference function to follow the original reciprocal-space formulation as closely as possible, it may also be obtained by expanding the Patterson functions in spherical harmonics and then performing the integration (Crowther, 1972). In this way, the concept of angular resolution is more evident. In fact, the spherical harmonics play the same role in the angular domain as the imaginary exponentials do in Cartesian coordinate space. The analogue of the Miller indexes here is the label ℓ, directly related to the angular resolution. The term with $\ell = 0$ is invariant under rotations and is thus eliminated from the summation in equation (13.2.3.14). It represents the contribution of the best radial functions, in a least-squares sense, that approximate the Patterson functions within Ω. The origin peak is thus properly removed by omitting the $\ell = 0$ term, as well as a substantial component of the original function. Also, by omitting low ℓ's, the angular resolution of peaks is enhanced. These eliminations may also be done in other formulations, but never as efficiently.

The parameters discussed above are not all independent. The relationship between them stems from the behaviour of the spherical Bessel functions for small values of their argument (Watson, 1958),

$$j_\ell(2\pi hb) \simeq (2\pi hb)^\ell / (2\ell + 1)!!$$

[see equation (13.2.3.8)]. So, when omitting low ℓ terms, we are also weighting down low-resolution data.

13.2.3.4. *Symmetry properties of the rotation function*

The overlap integral that defines $\mathcal{R}$ may be calculated by rotating P_t instead of P_s, but with the inverse rotation,

$$\mathcal{R}(\mathbf{R}) = (1/v) \int_\Omega P_t(\mathbf{r}) P_s(\mathbf{R}^{-1}\mathbf{r}) \, \mathrm{d}^3\mathbf{r}$$

$$= (1/v) \int_\Omega P_t(\mathbf{R}\mathbf{r}) P_s(\mathbf{r}) \, \mathrm{d}^3\mathbf{r}. \qquad (13.2.3.18)$$

This property enables the analysis of the consequence of the symmetries of the Patterson functions upon the rotation function (Tollin *et al.*, 1966; Moss, 1985). For example, when the target and search functions are the same, a trivial symmetry of the self-rotation results: its values at $\mathbf{R}$ and $\mathbf{R}^{-1}$ are the same or, in Euler angles, at (α, β, γ) and $(180 - \gamma, \beta, 180 - \alpha)$. More generally, if P_t is invariant under the rotation $\mathbf{T}$, *i.e.*, $P_t(\mathbf{T}^{-1}\mathbf{r}) = P_t(\mathbf{r})$ [and similarly

for the search function, $P_s(\mathbf{S}^{-1}\mathbf{r}) = P_s(\mathbf{r})$], then

$$\mathcal{R}(\mathbf{R}) = (1/v) \int_\Omega P_t(\mathbf{r}) P_s(\mathbf{R}^{-1}\mathbf{r}) \, \mathrm{d}^3\mathbf{r}$$

$$= (1/v) \int_\Omega P_t(\mathbf{T}^{-1}\mathbf{r}) P_s(\mathbf{S}^{-1}\mathbf{R}^{-1}\mathbf{r}) \, \mathrm{d}^3\mathbf{r}$$

$$= (1/v) \int_\Omega P_t(\mathbf{r}) P_s(\mathbf{S}^{-1}\mathbf{R}^{-1}\mathbf{T} \ \mathbf{r}) \, \mathrm{d}^3\mathbf{r} = \mathcal{R}(\mathbf{T}^{-1}\mathbf{R}\mathbf{S}).$$

$$(13.2.3.19)$$

However, the actual symmetry displayed by $\mathcal{R}$ depends on the parameterization of the rotations and on the orientation of the orthonormal axes with respect to the crystal ones. Within the Euler parameterization, if the target function has an *n*-fold rotation axis parallel to the orthonormal z axis, then, according to equation (13.2.2.3), $\mathcal{R}$ will have a periodicity of $360/n$ along α. Similarly, a rotation axis of order n along $\mathbf{z}$ of P_s gives rise to a periodicity of $360/n$ along γ. Therefore, the amount of calculation is reduced by choosing $\mathbf{z}$ along the Patterson functions' highest rotational symmetry axes [see equation (13.2.3.11)].

13.2.4. The locked rotation function

The rotational NCS, determined with the help of the self-rotation function, may be used to enhance the signal-to-noise ratio of cross-rotation functions (Rossmann *et al.*, 1972; Tong & Rossmann, 1990). If $\{\mathbf{S}_n, n = 1, \ldots, N\}$ denotes the set of NCS rotations, including the identity, and $\mathbf{R}$ is a correct orientation of the cross rotation, then $\mathbf{S}_n\mathbf{R}$ must also correspond to a correct orientation. Here we are assuming that the rotational NCS forms a group. Otherwise, either $\mathbf{S}_n\mathbf{R}$ or $\mathbf{S}_n^{-1}\mathbf{R}$, but not both, corresponds to another correct orientation. Therefore, a function may be defined, the locked cross rotation, whose values are the average of the cross-rotation values at orientations related by the NCS:

$$\mathcal{R}_{LC}(\mathbf{R}) = \sum_{n=1}^{N} \mathcal{R}(\mathbf{S}_n\mathbf{R})/N. \qquad (13.2.4.1)$$

By redefining the target function, it can be computed as an ordinary cross rotation. Indeed, $\mathcal{R}_{LC}$ may be written in a form similar to equation (13.2.3.3),

$$\mathcal{R}_{LC}(\mathbf{R}) = \sum_{n=1}^{N} (1/v) \int_\Omega P_t(\mathbf{r}) P_s(\mathbf{R}^{-1}\mathbf{S}_n^{-1}\mathbf{r}) \, \mathrm{d}^3\mathbf{r}/N$$

$$= (1/v) \int_\Omega \left[\sum_{n=1}^{N} P_t(\mathbf{S}_n\mathbf{r})/N \right] P_s(\mathbf{R}^{-1}\mathbf{r}) \, \mathrm{d}^3\mathbf{r}, \qquad (13.2.4.2)$$

with the target Patterson function substituted by the average over the NCS of the rotated target functions. The computation of equation (13.2.4.2) is particularly simple in the case of the fast rotation function. The substitution

$$e_{\ell, m, n}^{(t)} \rightarrow \sum_{m'=-\ell}^{\ell} \left[\sum_{n=1}^{N} \mathcal{D}_{m, m'}^{\ell}(\mathbf{S}_n)/N \right] e_{\ell, m', n}^{(t)}, \qquad (13.2.4.3)$$

where we replaced the sum over $\mathbf{S}_n^{-1}$ by a sum over $\mathbf{S}_n$, because of the rearrangement theorem of group theory, gives the required target coefficients.

The same ideas may be applied to the self-rotation function. Here the NCS is assumed beforehand, with elements $\{\mathbf{I}_n, n = 1, \ldots, N\}$ in a given reference orientation. They are related to the actual NCS elements by

$$\mathbf{S}_n = \mathbf{R}_n \mathbf{I}_n \mathbf{R}_n^{-1} \qquad (13.2.4.4)$$

[see equation (13.2.2.2)]. Since each $\mathbf{S}_n$ should correspond to a local

<center>272</center>

maximum of the self rotation, the function

$$\mathcal{R}_{LS}(\mathbf{R}) = \sum_{n=1}^{N} \mathcal{R}(\mathbf{R}\mathbf{I}_n \mathbf{R}^{-1})/N \qquad (13.2.4.5)$$

should also display a maximum for each $\mathbf{R}_n$, but with a noise level reduced by $(N)^{1/2}$. Equation (13.2.4.5) defines the locked self-rotation function. Its computation by fast rotation techniques is less straightforward than in the locked cross-rotation case.

13.2.5. Other rotation functions

The rotation function was hitherto described in terms of self- and cross-Patterson vectors. This is perhaps inevitable in the self-rotation case, but the problem of determining the absolute orientation of the subunits when a model structure is available may be formulated in a different way. We may try to compare directly the observed and calculated intensities or structure factors by using any criterion analogous to those employed in refinement procedures, *e.g.*, the crystallographic R factor or correlation coefficients.

When the space-group symmetry is explicitly exhibited, the structure factor corresponding to a crystal with M independent molecules in the unit cell takes the form

$$F(\mathbf{h}) = \sum_{m=1}^{M} \sum_{g=1}^{G} f_m(\mathbf{h}\mathbf{M}_g) \exp[2\pi i \mathbf{h}(\mathbf{M}_g \mathbf{r}_m + \mathbf{t}_g)], \qquad (13.2.5.1)$$

where $\mathbf{M}_g$ and $\mathbf{t}_g$ denote, respectively, the transformation matrix and the translation associated with the gth symmetry operation of the crystal space group. The corresponding intensity is

$$I(\mathbf{h}) = \sum_{m, m'=1}^{M} \sum_{g, g'=1}^{G} \overline{f_m(\mathbf{h}\mathbf{M}_g)} f_{m'}(\mathbf{h}\mathbf{M}_{g'})$$
$$\times \exp[2\pi i \mathbf{h}(\mathbf{M}_g \mathbf{r}_m + \mathbf{t}_g - \mathbf{M}_{g'} \mathbf{r}_{m'} - \mathbf{t}_{g'})]. \qquad (13.2.5.2)$$

For criteria based on amplitudes, the calculated structure factor will contain only the contribution of the rotated model,

$$F_{\mathbf{h}}^{\text{calc}}(\mathbf{R}) = f_m(\mathbf{h}\mathbf{R}), \qquad (13.2.5.3)$$

i.e., the Fourier transform of a single molecule in the crystal cell, assuming $P1$ symmetry. For criteria based on intensities, some symmetry information may be introduced,

$$I_{\mathbf{h}}^{\text{calc}}(\mathbf{R}) = \sum_{g=1}^{G} |f_m(\mathbf{h}\mathbf{M}_g \mathbf{R})|^2. \qquad (13.2.5.4)$$

A criterion often considered is the correlation coefficient on intensities,

$$\mathcal{R}_{\mathcal{D}}(\mathbf{R}) = \langle (I^{\text{obs}} - \langle I^{\text{obs}} \rangle)(I^{\text{calc}} - \langle I^{\text{calc}} \rangle) \rangle$$
$$\times [\langle (I^{\text{obs}} - \langle I^{\text{obs}} \rangle)^2 \rangle \langle (I^{\text{calc}} - \langle I^{\text{calc}} \rangle)^2 \rangle]^{1/2}, \qquad (13.2.5.5)$$

where $\langle \ldots \rangle$ means 'average over reflections'. It may be calculated within reasonable computing time provided that

(1) the structure factors are computed by interpolation from the Fourier transform of the isolated molecule's electron density; and

(2) an efficient sampling set of $\mathcal{R}_{\mathcal{D}}$ is defined.

$\mathcal{R}_{\mathcal{D}}$ is referred to as the direct-rotation function (DeLano & Brünger, 1995). A major advantage of this formulation is that the information stemming from already-positioned subunits may be taken into account, just by adding their contribution to the calculated intensities.

13.2.6. Concluding remarks

Each formulation of the rotation function described above has its advantages and disadvantages. The direct-space formulation [equation (13.2.3.3)] offers the possibility of modifying the Patterson function or selecting the strongest peaks to be used in the overlap integral. Also, the domain of integration may have any shape, as in the reciprocal-space formulation [equation (13.2.3.5)]. However, in both formulations, the numerical results can be somewhat imprecise because of the approximations introduced to save computing time.

When the domain of integration is spherical, as is usually the case, then the fast rotation function [equation (13.2.3.15)] is faster, more accurate and allows for angular-resolution enhancement. Moreover, since most of the computing time is spent in the calculation of the coefficients $e_{\ell, m, n}$, a library of these coefficients may be compiled to assess the models in a given molecular-replacement problem more rapidly.

The direct rotation [equation (13.2.5.5)] seems preferable to the cross-rotation function, as it includes all self-Patterson terms of the search model. However, when the order of the space group is high, the calculated intensity represents only a small fraction of the observed one, and the discriminative power of the function drops, as compared with the cross-rotation function.

Patterson searches now benefit from new supercomputers. The real problem of molecular replacement, split for convenience into rotation and translation searches, is beginning to be tackled by genuine six-dimensional searches where all orientations are tested. This may represent the end of cross-rotation and direct-rotation functions.

Appendix 13.2.1.
Formulae for the derivation and computation of the fast rotation function

This appendix aims to present a complete set of formulae which allow the derivation and computation of the fast rotation function. They involve a particular convention for the definition of the irreducible representations of the rotation group suitable for crystallographic computations.

A13.2.1.1. *Euler parameterization*

By applying the group property of rotations [equation (13.2.2.2)], the Euler parameterization may be expressed as rotations around fixed axis (see Fig. 13.2.2.1b):

$$\mathbf{R}(\alpha, \beta, \gamma) = \mathbf{R}(\gamma, \mathbf{p})\mathbf{R}(\beta, \mathbf{n})\mathbf{R}(\alpha, \mathbf{z})$$
$$= \left[\mathbf{R}(\beta, \mathbf{n})\mathbf{R}(\gamma, \mathbf{z})\mathbf{R}(\beta, \mathbf{n})^{-1} \right] \mathbf{R}(\beta, \mathbf{n})\mathbf{R}(\alpha, \mathbf{z})$$
$$= \mathbf{R}(\beta, \mathbf{n})\mathbf{R}(\gamma, \mathbf{z})\mathbf{R}(\alpha, \mathbf{z})$$
$$= \left[\mathbf{R}(\alpha, \mathbf{z})\mathbf{R}(\beta, \mathbf{y})\mathbf{R}(\alpha, \mathbf{z})^{-1} \right] \mathbf{R}(\gamma, \mathbf{z})\mathbf{R}(\alpha, \mathbf{z})$$
$$= \mathbf{R}(\alpha, \mathbf{z})\mathbf{R}(\beta, \mathbf{y})\mathbf{R}(\gamma, \mathbf{z}). \qquad (A13.2.1.1)$$

A13.2.1.2. *The $D_{m, m'}^{\ell}$ matrices*

A linear representation of dimension n of the rotation group is a correspondence between rotations and matrices of order n. The matrices $D_{m, m'}^{\ell}(\mathbf{R})$, with $-\ell \leq m, m' \leq \ell$, are associated with the irreducible representation of dimension $2\ell + 1$ ($0 \leq \ell < \infty$). They have the following properties (Brink & Satchler, 1968):

273

references