Implementation of molecular replacement in *AMoRe*

Jorge Navaza

CNRS-GIF, LGV, 91198 Gif-sur-Yvette, France

Correspondence e-mail:
jorge.navaza@gv.cnrs-gif.fr

An account is given of the molecular replacement method as implemented in the package *AMoRe*. The overall strategy of the method is presented and the main functions used in the package are described. The most important features of *AMoRe* are the quality of the fast rotation and translation functions and the facility of multiple inputs to translation and rigid-body refinement functions, which allow for a fast multiple exploration of crystal configurations with a high level of automation.

1. Introduction

The idea of molecular replacement is to build a tentative crystal structure using known molecular models similar to the actual molecules that constitute the crystal in order to start model building or refinement. The problem is to determine the positions of the models within the crystal cell. This is ultimately performed by comparing observed and calculated structure factors for selected positions of the independent molecules within the cell. In *AMoRe*, the comparison essentially involves the correlation coefficient in terms of amplitudes. This criterion was chosen in the light of the results available one decade ago, results that now may be considered as corresponding to easy or moderately difficult MR problems. At that time, an exhaustive positional search involving in general six variables per independent model using that simple but robust criterion could not be envisaged. Nowadays, a full six-dimensional search would also be too lengthy, although feasible. This explains, perhaps, the fact that the original ideas of Rossmann and Blow, *i.e.* the splitting of the search into two consecutive three-dimensional ones, are still found in filigree in most MR packages.

The main programs in *AMoRe* aim at selecting a certain number of positions, obtained through the exhaustive exploration of three-dimensional domains with fast functions, and computing the correlation coefficients associated with these positions. The idea is to assess many crystal configurations, as *it is the contrast in the values of the criterion that gives one confidence in the solution*. The fast functions, rotation functions and translation functions are either improved versions of already proposed ones or new ones. Accurate and fast algorithms are used throughout the package in order to save computing time. In particular, molecular scattering factors replace coordinates, which are used only once in the whole procedure.

The main stream in *AMoRe* is the set of values of the variables that specify the positions of the independent models within the crystal, from which structure factors and inputs to

the fast functions are calculated. We will first define these variables and their relationship to the calculated structure factors. We will then describe the strategy for the selection of configurations.

2. Positional variables and crystal configurations

The position of the molecular model within the crystal is determined by the rotation $\mathbf{R}$ and the translation $\mathbf{T}$ that move the model from a reference initial position, specified by the atomic vectors $\{\mathbf{r}^o\}$, to the current position, specified by the atomic vectors $\{\mathbf{r}\}$,

$$\mathbf{r} = \mathbf{R}\mathbf{r}^o + \mathbf{T}. \tag{1}$$

The translation $\mathbf{T}$ is usually given in fractional coordinates (x, y, z) in the crystal cell. The rotation $\mathbf{R}$ is parameterized with the Euler angles (φ, θ, ψ) associated with an orthonormal frame $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$. Several conventions exist for the names of angles and definitions of the axes involved in this parameterization. We will follow the convention by which (φ, θ, ψ) denotes a rotation of ψ about the $\mathbf{Z}$ axis, followed by a rotation of θ about the $\mathbf{Y}$ axis and finally a rotation of φ about the $\mathbf{Z}$ axis,

$$\mathbf{R}(\varphi, \theta, \psi) = \mathbf{R}(\varphi, \mathbf{Z})\mathbf{R}(\theta, \mathbf{Y})\mathbf{R}(\psi, \mathbf{Z}). \tag{2}$$

The angles take values within the parallelepiped $\{0 \leq \varphi < 360; 0 \leq \theta \leq 180; 0 \leq \psi < 360°\}$. For $\theta = 0$ or $180°$, only the combinations $\varphi + \psi$ or $\varphi - \psi$ are independent, respectively.

The initial position of the model is usually chosen with its center of mass placed at the origin and its principal axes of inertia parallel to the orthonormal frame, as this leads to an efficient sampling of configurations. A good choice for the orthonormal frame is $\mathbf{Z}$ parallel to the highest crystal symmetry axis (nort = 0 in *AMoRe*). This choice restricts the orientational search to $\{0 \leq \varphi < 36/n\}$, where n is the order of the rotational symmetry around $\mathbf{Z}$.

Therefore, given the models' initial positions, the crystal unit-cell parameters, the space-group symmetry and the orientation of the orthonormal frame, a crystal configuration is uniquely determined by giving the positions of the independent molecular models within the unit cell, expressed in terms of the positional variables,

$$
\begin{array}{ccccccc}
\#m' & \varphi_{m'} & \theta_{m'} & \psi_{m'} & x_{m'} & y_{m'} & z_{m'} \\
\#\dots & \dots & \dots & \dots & \dots & \dots & \dots \\
\#m & \varphi_m & \theta_m & \psi_m & x_m & y_m & z_m.
\end{array}
$$

The labels $m', \dots, m$ identify the molecules and the molecular models. Note that some of these models may coincide.

3. Structure-factor calculation

The calculated structure factors are conveniently written in terms of the individual molecular scattering factors $f_m(\mathbf{s})$, *i.e.* the Fourier transform of the electron density corresponding to the isolated molecule in its initial position. These molecular scattering factors are computed with the *TABLING* program, which translates the model coordinates so that the center of

mass is at the origin and rotates the coordinates so that the model's principal axes of inertia are parallel to the model box. An electron density is then constructed and eventually transformed by Fast Fourier techniques. One feature of *AMoRe* is that the model may well be an electron density or an electron-microscopy reconstruction, as only the Fourier coefficients are used.

If $\mathbf{R}_m$ and $\mathbf{T}_m$ denote the rotation and translation that define the molecule's current position, $\mathbf{M}_g$ and $\mathbf{t}_g$ the space-group transformation matrix and translation vector of the *g*th symmetry operation and $\mathbf{H}$ the coordinates of a crystal reciprocal vector, the contribution of molecule m to the calculated crystal structure factor is

$$\sum_{g=1}^{G} f_m(\mathbf{HM}_g\mathbf{DR}_m\mathbf{O}_m) \exp[2\pi i\mathbf{H}(\mathbf{M}_g\mathbf{T}_m + \mathbf{t}_g)]. \tag{3}$$

$\mathbf{D}$ and $\mathbf{O}_m$ are orthogonalizing and deorthogonalizing matrices. In fact, $\mathbf{DR}_m\mathbf{O}_m$ is simply the rotation matrix $\mathbf{R}_m$ expressed in a mixed basis: it applies (from left to right) to reciprocal coordinates (Miller indices) in the crystal and produces reciprocal coordinates in the model box. If there are M independent molecules we have to add M terms like this. Assuming that the individual molecular scattering factors $f_m(\mathbf{s})$ have been set to a common scale, we have

$$F_\mathbf{H}^{\mathrm{cal}} = \sum_{m=1}^{M}\sum_{g=1}^{G} f_m(\mathbf{HM}_g\mathbf{DR}_m\mathbf{O}_m) \exp[2\pi i\mathbf{H}(\mathbf{M}_g\mathbf{T}_m + \mathbf{t}_g)]. \tag{4}$$

4. Correlation coefficient

As stated in the introduction, the agreement criterion to assess crystal configurations is the (linear) correlation coefficient between observed and calculated amplitudes,

$$\mathrm{CC}_F = \left(\sum_\mathbf{H} \overline{\overline{|F_\mathbf{H}^{\mathrm{obs}}|}} \times \overline{\overline{|F_\mathbf{H}^{\mathrm{cal}}|}}\right) \Big/ \left[\left(\sum_\mathbf{H} \overline{\overline{|F_\mathbf{H}^{\mathrm{obs}}|}}^2\right) \times \left(\sum_\mathbf{H} \overline{\overline{|F_\mathbf{H}^{\mathrm{cal}}|}}^2\right)\right]^{1/2}, \tag{5}$$

where $\overline{\overline{|F_\mathbf{H}|}}$ denotes a 'centered' variable, *e.g.*

$$\overline{\overline{|F_\mathbf{H}|}} = |F_\mathbf{H}| - \langle|F_\mathbf{H}|\rangle, \tag{6}$$

and $\langle\cdots\rangle$ means average over reflections. CC_F takes values in the interval $(1, -1)$.

5. Strategy

The overall strategy of MR as implemented in *AMoRe* is easily understood if we consider the correlation coefficient between intensities

$$\mathrm{CC}_I = \left(\sum_\mathbf{H} \overline{\overline{I_\mathbf{H}^{\mathrm{obs}}}} \times \overline{\overline{I_\mathbf{H}^{\mathrm{cal}}}}\right) \Big/ \left[\left(\sum_\mathbf{H} \overline{\overline{I_\mathbf{H}^{\mathrm{obs}}}}^2\right) \times \left(\sum_\mathbf{H} \overline{\overline{I_\mathbf{H}^{\mathrm{cal}}}}^2\right)\right]^{1/2} \tag{7}$$

as the target function for screening. The calculated total intensity is given by

$$I_{\mathbf{H}}^{\mathrm{cal}} = \sum_{m,m'=1}^{M} \sum_{g,g'=1}^{G} f_m(\mathbf{HM}_g\mathbf{DR}_m\mathbf{O}_m)\overline{f_{m'}(\mathbf{HM}_{g'}\mathbf{DR}_{m'}\mathbf{O}_{m'})}$$
$$\times \ \exp[2\pi i\mathbf{H}(\mathbf{M}_g\mathbf{T}_m+\mathbf{t}_g-\mathbf{M}_{g'}\mathbf{T}_{m'}-\mathbf{t}_{g'})], \tag{8}$$

where the overline means 'complex conjugate'. The positional variables entering into this expression are successively determined by using different approximations to $I_{\mathbf{H}}^{\mathrm{cal}}$ and, accordingly, CC_I. The protocol consists of three main steps.

(i) Rotation search. For each of the M molecules, determine their possible orientations. The calculated intensities are approximated by

$$I_{\mathbf{H}}^{\mathrm{cal}} \simeq \sum_{g=1}^{G} |f_m(\mathbf{HM}_g\mathbf{DR}_m\mathbf{O}_m)|^2 \tag{9}$$

and the highest peaks of CC_I (as a function of $\mathbf{R}_m$) are selected. This function is essentially the direct rotation function (Delano & Brünger, 1995). Even though CC_I, with $I_{\mathbf{H}}^{\mathrm{cal}}$ given by (9), cannot be calculated by standard fast techniques, available computing resources allow for a point-by-point evaluation of CC_I at an average speed of $\sim 3 \times 10^{-7}$ s per reflection and per symmetry operation on a 500 MHz Digital XP1000.

(ii) One-body translation search. For each molecule and for each selected orientation, determine the possible translations. The calculated intensities are approximated by

$$I_{\mathbf{H}}^{\mathrm{cal}} \simeq \sum_{g,g'=1}^{G} f_m(\mathbf{HM}_g\mathbf{DR}_m\mathbf{O}_m)\overline{f_m(\mathbf{HM}_{g'}\mathbf{DR}_m\mathbf{O}_m)}$$
$$\times \ \exp\{2\pi i\mathbf{H}[(\mathbf{M}_g-\mathbf{M}_{g'})\mathbf{T}_m+\mathbf{t}_g-\mathbf{t}_{g'}]\}. \tag{10}$$

CC_I (as a function of $\mathbf{T}_m$) is evaluated by FFT (Navaza & Vernoslova, 1995) within the Cheshire cell. The assessed (partial) crystal configurations are given by

$$\#m \quad \varphi_m \quad \theta_m \quad \psi_m \quad x_m \quad y_m \quad z_m \quad CC_I.$$

The positions corresponding to the top one-body configurations are then refined; that with highest value of CC_I is assumed to be correct.

(iii) n-body translation search. When many independent molecules or molecular fragments have to be positioned, the contribution of already placed models dramatically increases the chances of success of the method. If $M' < M$ molecules are already positioned (say those with labels $m' \in \mathcal{P}$), then for the top orientations of the remaining molecules ($m \notin \mathcal{P}$), determine the possible translations. The calculated intensities are approximated by

$$I_{\mathbf{H}}^{\mathrm{cal}} \simeq \sum_{m'\in\mathcal{P}} \sum_{g',g=1}^{G} f_{m'}(\mathbf{HM}_{g'}\mathbf{DR}_{m'}\mathbf{O}_{m'})\overline{f_m(\mathbf{HM}_g\mathbf{DR}_m\mathbf{O}_m)}$$
$$\times \ \exp[2\pi i\mathbf{H}(\mathbf{M}_{g'}\mathbf{T}_{m'}+\mathbf{t}_{g'}-\mathbf{M}_g\mathbf{T}_m-\mathbf{t}_g)]$$
$$+ \ f_m(\mathbf{HM}_g\mathbf{DR}_m\mathbf{O}_m)\overline{f_{m'}(\mathbf{HM}_{g'}\mathbf{DR}_{m'}\mathbf{O}_{m'})}$$
$$\times \ \exp[2\pi i\mathbf{H}(\mathbf{M}_g\mathbf{T}_m+\mathbf{t}_g-\mathbf{M}_{g'}\mathbf{T}_{m'}-\mathbf{t}_{g'})]. \tag{11}$$

CC_I (as a function of $\mathbf{T}_m$) is evaluated by FFT, now within the whole crystal cell. The assessed (partial) crystal configurations are now given by

$$\#m' \quad \varphi_{m'} \quad \theta_{m'} \quad \psi_{m'} \quad x_{m'} \quad y_{m'} \quad z_{m'}$$
$$\#\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$
$$\#m \quad \varphi_m \quad \theta_m \quad \psi_m \quad x_m \quad y_m \quad z_m \quad CC_I,$$

where CC_I corresponds to the whole $(M' + 1)$-body configuration. The positions corresponding to the best configurations are then refined and that with highest value of CC_I is assumed to be a correct one.

(iv) Repeat the preceding step until $M' = M$.

The actual protocol in *AMoRe* differs from the one above mainly in the rotational search. The *ROTING* program, based on the fast rotation function proposed by Crowther, is used to determine the possible orientations of the models (Crowther, 1972). Also, as previously stated, the crystal configurations are assessed with CC_F instead of CC_I. The translations of the oriented models (one-body and n-body searches) are determined with the *TRAING* program. Several translation functions have been incorporated, among which the one described in the above protocol, *i.e.* CC_I as a function of $\mathbf{T}_m$. The refinement of the positional variables is performed with the fast rigid-body refinement program *FITING* (Castellano *et al.*, 1992). These fast functions will be described in the following section.

A situation where this protocol fails is often one in which a six-dimensional search fails too. As a rule, this corresponds to a poor quality of the search model or a small size of the search fragment with respect to the asymmetric unit content.

The fast structure-factor calculation algorithm (4), the performance of *ROTING* and the facility of multiple inputs to *TRAING* and *FITING* allow for a fast multiple exploration. A link between the input/output of the above programs allows for automation. In fact, three levels of automation may be distinguished.

Level I: automation is here reduced to the possibility of multiple inputs to the main programs.

Level II: inputs to the main programs are created by selecting and combining putative solutions.

Level III: a protocol for the whole MR problem is created, starting from general purpose control parameters.

6. Description of the fast search programs

6.1. The *ROTING* program

It is possible to determine the rotations $\mathbf{R}$ that superimpose a search molecule upon the homologous ones within the target crystal by calculating the overlap within a conveniently chosen region Ω of volume v of the observed Patterson function (the target function P_t) and a rotated version of the Patterson function corresponding to the isolated search molecule (the search function P_s),

$$\mathcal{R}(\mathbf{R}) = \frac{1}{v} \int_{\Omega} P_t(\mathbf{r})P_s(\mathbf{R}^{-1}\mathbf{r})\,\mathrm{d}^3\mathbf{r} \tag{12}$$

(Rossmann & Blow, 1962). $\mathcal{R}$ should display a local maximum for the sought rotations. Note that when we rotate the search function P_s by $\mathbf{R}$, its argument contains $\mathbf{R}^{-1}$.

It may be useful to compare rotation functions obtained under different conditions. For this, some kind of normalization is needed. In fact, $\mathcal{R}$ is cast into the form of a correlation coefficient by dividing (12) by the norms of the truncated Patterson functions,

$$\mathcal{R}_N(\mathbf{R}) = \int_\Omega P_t(\mathbf{r})P_s(\mathbf{R}^{-1}\mathbf{r})\,\mathrm{d}^3\mathbf{r} \bigg/ \left[\int_\Omega P_t(\mathbf{r})^2\,\mathrm{d}^3\mathbf{r} \int_\Omega P_s(\mathbf{r})^2\,\mathrm{d}^3\mathbf{r}\right]^{1/2}.$$

(13)

The reciprocal-space formulation of (12) is obtained by replacing the Patterson functions by their Fourier summations

$$P(\mathbf{r}) = \sum_\mathbf{h} \frac{I(\mathbf{h})}{V}\exp(-2\pi i\mathbf{h}\mathbf{r}).$$

(14)

Taking into account that $I(-\mathbf{h}) = I(\mathbf{h})$, we obtain

$$\mathcal{R}(\mathbf{R}) = \sum_\mathbf{h}\sum_\mathbf{k} \frac{I_t(\mathbf{h})}{V_t}\frac{I_s(\mathbf{k})}{V_s}\frac{1}{v}\int_\Omega \exp[2\pi i(\mathbf{h} - \mathbf{k}\mathbf{R}^{-1})\mathbf{r}]\,\mathrm{d}^3\mathbf{r}$$
$$= \sum_\mathbf{h}\sum_\mathbf{k} \frac{I_t(\mathbf{h})}{V_t}\frac{I_s(\mathbf{k})}{V_s}\chi_\Omega(\mathbf{h} - \mathbf{k}\mathbf{R}^{-1}).$$

(15)

χ_Ω is the Fourier transform of the function that takes the value 1 within Ω and 0 outside. In principle, the domain of integration could have any shape. However, in order to take full advantage of the properties of the rotation group, Ω is usually chosen as a spherical domain of radius b. Letting $\mathbf{s} = \mathbf{h} - \mathbf{k}\mathbf{R}^{-1}$ for short, we have

$$\chi_b(\mathbf{s}) = \frac{3}{4\pi b^3}\int_0^b\int_0^\pi\int_0^{2\pi} \exp(2\pi i\mathbf{s}\mathbf{r})r^2\sin(\theta)\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\varphi$$
$$= 3\frac{\sin(2\pi sb) - 2\pi sb\cos(2\pi sb)}{(2\pi sb)^3}.$$

(16)

Although simple, the resulting expression for the rotation function has the disadvantage of containing entangled $\mathbf{h}$, $\mathbf{k}$ and $\mathbf{R}$ contributions, which renders its computation time consuming if the whole domain of rotations has to be explored. The difficulty may be overcome by expanding the exponentials entering into (15) in spherical harmonics, $Y_{l,m}$. Taking advantage of their transformation under rotations and using recurrence relationships between spherical Bessel functions j_l, we obtain

$$\chi_b(\mathbf{h} - \mathbf{k}\mathbf{R}^{-1})$$
$$= \sum_{l=0}^\infty \frac{j_l(2\pi hb)j_{l-1}(2\pi kb)2\pi kb - j_l(2\pi kb)j_{l-1}(2\pi hb)2\pi hb}{(2\pi hb)^2 - (2\pi kb)^2}$$
$$\times \sum_{m,m'=-l}^l \overline{Y_{l,m}(\mathbf{h}/h)}Y_{l,m'}(\mathbf{k}/k)D_{m,m'}^l(\mathbf{R})$$
$$= \sum_{l=0}^\infty\left\{\sum_{n=1}^\infty 12\pi[2(l+2n)-1]\frac{j_{l+2n-1}(2\pi hb)}{2\pi hb}\frac{j_{l+2n-1}(2\pi kb)}{2\pi kb}\right\}$$
$$\times \sum_{m,m'=-l}^l \overline{Y_{l,m}(\mathbf{h}/h)}Y_{l,m'}(\mathbf{k}/k)D_{m,m'}^l(\mathbf{R}),$$

(17)

where $D_{m,m'}^l$ are the matrices of the irreducible representations of the rotation group. The awkwardness of (17) is apparent rather than real.

(i) The expression separates angular from crystal variables. It also separates target from search contributions.

(ii) The equation is accurate, even when truncating the summations on l and n to reasonable values. The upper limit for l is of the order of the highest argument of the spherical Bessel functions,

$$l_{\max} \simeq 2\pi b/d_{\min},$$

where $d_{\min}$ is the resolution of the data. The upper limit for n depends on the current value of l,

$$n_{\max}(l) \simeq (l_{\max} - l + 2)/2.$$

(iii) When the rotations are parameterized in Euler angles (φ, θ, ψ), the matrices $D_{m,m'}^l$ take the form

$$D_{m,m'}^l(\varphi, \theta, \psi) = d_{m,m'}^l(\theta)\exp[i(m\varphi + m'\psi)],$$

(18)

which enables the computation of $\mathcal{R}$ for each given value of θ by means of two-dimensional fast Fourier transforms. This formulation is referred to as the fast rotation function (Crowther, 1972).

6.2. Computing the fast rotation function

The calculations are organized as follows.

(i) Given the search and the target diffraction data, compute

$$e_{l,m,n} = \{12\pi[2(l+2n)-1]\}^{1/2}$$
$$\times \sum_\mathbf{h}\frac{I(\mathbf{h})}{V}Y_{l,m}(\mathbf{h}/h)\frac{j_{l+2n-1}(2\pi hb)}{2\pi hb}$$

(19)

and normalize (to compute $\mathcal{R}_N$ instead of $\mathcal{R}$)

$$e_{l,m,n} \rightarrow e_{l,m,n}\bigg/\left(\sum_{l=2}^{\max}\sum_{m=-l}^l\sum_{n=1}^{n_{\max}(l)}|e_{l,m,n}|^2\right).$$

(20)

Odd l terms disappear because the Friedel related reflections contribute with opposite signs,

$$I(-\mathbf{h})Y_{l,m}(-\mathbf{h}/h) = (-1)^l I(\mathbf{h})Y_{l,m}(\mathbf{h}/h).$$

Also, if the Patterson function has an n-fold rotation axis along $\mathbf{Z}$, only the terms with m equal to a multiple of n survive.

(ii) Given the $e_{l,m,n}$ values, perform the sums

$$C_{m,m'}^l = \sum_{n=1}^{n_{\max}(l)} \overline{e_{l,m,n}^{(t)}}e_{l,m',n}^{(s)}.$$

(21)

(iii) For each θ value, calculate the reduced matrix elements $d_{m,m'}^l$ and compute

$$S_{m,m'}(\theta) = \sum_{l=2}^{l_{\max}} C_{m,m'}^l d_{m,m'}^l(\theta).$$

(22)

Then evaluate the β section of $\mathcal{R}_N$ by FFT,

$$\mathcal{R}_N(\varphi, \theta, \psi) = \sum_{m,m'=-l_{\max}}^{l_{\max}} S_{m,m'}(\theta)\exp[i(m\varphi + m'\psi)].$$

(23)

The sampling in φ and ψ here is dictated by the standard FFT requirements.

$\mathcal{R}_N$ is used in *AMoRe* just to select a certain number of peaks. The output of *ROTING* contains, besides the values of $\mathcal{R}_N$, those of the correlation coefficients (CC_F and CC_I as in *P*1) for each of the selected orientations. CC_F is more efficient, in general.

6.3. 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 values of $\mathcal{R}$ at orientations related by the NCS,

$$\mathcal{R}_L(\mathbf{R}) = \sum_{n=1}^{N} \mathcal{R}(\mathbf{S}_n\mathbf{R})/N. \tag{24}$$

By redefining the target function, it can be computed as an ordinary cross rotation. Indeed, $\mathcal{R}_L$ may be written in a form similar to (12),

$$\mathcal{R}_L(\mathbf{R}) = \sum_{n=1}^{N} \frac{1}{v} \int_{\Omega} P_t(\mathbf{r}) P_s(\mathbf{R}^{-1}\mathbf{S}_n^{-1}\mathbf{r}) \, \mathrm{d}^3\mathbf{r}/N$$
$$= \frac{1}{v} \int_{\Omega} \left[\sum_{n=1}^{N} P_t(\mathbf{S}_n\mathbf{r})/N \right] P_s(\mathbf{R}^{-1}\mathbf{r}) \, \mathrm{d}^3\mathbf{r}, \tag{25}$$

with the target Patterson function substituted by the average over the NCS of the rotated target functions. The computation of (25) is particularly simple in the case of the fast rotation function. The substitution

$$e_{l,m,n}^{(t)} \rightarrow \sum_{m'=-l}^{l} \left[\sum_{n=1}^{N} D_{m,m'}^{l}(\mathbf{S}_n)/N \right] e_{l,m',n}^{(t)}, \tag{26}$$

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.

6.4. The *TRAING* program

The possible translations of an oriented model are selected in *AMoRe* by means of fast translation functions computed with the *TRAING* program. The output of this program contains, besides the values of the fast translation function, those of CC_F, CC_I and the R factor for each of the selected translations. Several fast translation functions may be calculated. If we write the Fourier coefficient of the oriented model, rotated by a given $\mathbf{R}_m$ and placed at $\mathbf{T}$, as

$$F_{\mathbf{H}}^{\mathrm{cal}}(\mathbf{T}) = \sum_{g=1}^{G} [f_m(\mathbf{HM}_g\mathbf{DR}_m\mathbf{O}_m) \exp(2\pi i\mathbf{Ht}_g)] \exp(2\pi i\mathbf{HM}_g\mathbf{T})$$
$$= \sum_{g=1}^{G} u_g^m(\mathbf{H}) \exp(2\pi i\mathbf{HM}_g\mathbf{T}) \tag{27}$$

(see equation 3) and the corresponding intensity as

$$I_{\mathbf{H}}^{\mathrm{cal}}(\mathbf{T}) = \sum_{g,g'=1}^{G} f_m(\mathbf{HM}_g\mathbf{DR}_m\mathbf{O}_m)\overline{f_m(\mathbf{HM}_{g'}\mathbf{DR}_m\mathbf{O}_m)}$$
$$\times \exp\{2\pi i\mathbf{H}[(\mathbf{M}_g-\mathbf{M}_{g'})\mathbf{T}+\mathbf{t}_g-\mathbf{t}_{g'}]\}$$
$$= \sum_{g,g'=1}^{G} u_g^m(\mathbf{H})\overline{u_{g'}^m(\mathbf{H})} \exp[2\pi i\mathbf{H}(\mathbf{M}_g-\mathbf{M}_{g'})\mathbf{T}] \tag{28}$$

(see equation 10), then the options are (same notation as in equations 6 and 7)

(i) centered overlap,

$$\mathrm{CO}(\mathbf{T}) = \sum_{\mathbf{H}} \overline{\overline{I_{\mathbf{H}}^{\mathrm{obs}}}} \times \overline{\overline{I_{\mathbf{H}}^{\mathrm{cal}}(\mathbf{T})}} \tag{29}$$
$$\propto \sum_{g,g'=1}^{G} \sum_{\mathbf{H}} \overline{\overline{I_{\mathbf{H}}^{\mathrm{obs}}u_g^m(\mathbf{H})}}u_{g'}^m(\mathbf{H}) \exp[-2\pi i\mathbf{H}(\mathbf{M}_g-\mathbf{M}_{g'})\mathbf{T}],$$

(ii) Harada–Lifchitz (Harada *et al.*, 1981)

$$\mathrm{HL}(\mathbf{T}) = \left(\sum_{\mathbf{H}} \overline{\overline{I_{\mathbf{H}}^{\mathrm{obs}}}} \times \overline{\overline{I_{\mathbf{H}}^{\mathrm{cal}}(\mathbf{T})}} \right) \Big/ \sum_{\mathbf{H}} I_{\mathbf{H}}^{\mathrm{cal}}(\mathbf{T}), \tag{30}$$

(iii) correlation coefficient

$$\mathrm{CC}(\mathbf{T}) = \left[\sum_{\mathbf{H}} \overline{\overline{I_{\mathbf{H}}^{\mathrm{obs}}}} \times \overline{\overline{I_{\mathbf{H}}^{\mathrm{cal}}(\mathbf{T})}} \right] \Big/ \left\{ \left(\sum_{\mathbf{H}} \overline{\overline{I_{\mathbf{H}}^{\mathrm{obs}}}}^2 \right) \times \left[\sum_{\mathbf{H}} \overline{\overline{I_{\mathbf{H}}^{\mathrm{cal}}(\mathbf{T})}}^2 \right] \right\}^{1/2}, \tag{31}$$

(iv) phased translation
(*a*) without 'external' phases

$$\mathrm{PT}(\mathbf{T}) = \sum_{g,g'=1}^{G} \sum_{\mathbf{H}} [|F_{\mathbf{H}}^{\mathrm{obs}}/u_g^m(\mathbf{H})| + |F_{\mathbf{H}}^{\mathrm{obs}}/u_{g'}^m(\mathbf{H})| - 2\mathrm{SCAL}]$$
$$\times \overline{u_g^m(\mathbf{H})}u_{g'}^m(\mathbf{H}) \exp[-2\pi i\mathbf{H}(\mathbf{M}_g-\mathbf{M}_{g'})\mathbf{T}], \tag{32}$$

(*b*) with 'external' phases $\varphi_{\mathbf{H}}^{\mathrm{ext}}$

$$\mathrm{PTF}(\mathbf{T}) = \sum_{\mathbf{H}} |F_{\mathbf{H}}^{\mathrm{obs}}| \exp(i\varphi_{\mathbf{H}}^{\mathrm{ext}}) \times \overline{F_{\mathbf{H}}^{\mathrm{cal}}(\mathbf{T})}, \tag{33}$$

(v) *n*-body translation with fixed contribution $F_{\mathbf{H}}^{\mathrm{fix}}$
(*a*) phased translation

$$\mathrm{PTN}(\mathbf{T}) = \sum_{\mathbf{H}} \left(\left| \frac{F_{\mathbf{H}}^{\mathrm{obs}}}{F_{\mathbf{H}}^{\mathrm{fix}}} \right| - \mathrm{SCAL} \right) F_{\mathbf{H}}^{\mathrm{fix}} \times \overline{F_{\mathbf{H}}^{\mathrm{cal}}(\mathbf{T})}, \tag{34}$$

(*b*) for all others, replace

$$F_{\mathbf{H}}^{\mathrm{cal}}(\mathbf{T}) \rightarrow F_{\mathbf{H}}^{\mathrm{cal}}(\mathbf{T}) + F_{\mathbf{H}}^{\mathrm{fix}}. \tag{35}$$

SCAL is a scale factor to subtract the contribution of the phasing position. The complex exponentials in (29) to (32) depend on reciprocal vectors $\mathbf{H}(\mathbf{M}_g - \mathbf{M}_{g'})$, which are in the Cheshire reciprocal cell (Hirshfeld, 1968).

6.5. The *FITING* program

Although *FITING* is not a search program, we include it here as it is one of the main molecular-replacement programs. It performs rigid-body refinement by a fast technique first proposed by Huber & Schneider (1985). The quadratic misfit

$$\sum_{\mathbf{H}} \left\{ \left| |F_{\mathbf{H}}^{\text{obs}}| - \frac{\exp(B|\mathbf{H}|^2)}{\lambda} \right| \sum_{m=1}^{M} f_m(\mathbf{H}\mathbf{M}_g \mathbf{D}\mathbf{R}_m \mathbf{O}_m) \right.$$
$$\left. \times \exp[2\pi i \mathbf{H}(\mathbf{M}_g \mathbf{T}_m + \mathbf{t}_g)] \right|^2 \right\} \tag{36}$$

is minimized with respect to the positional variables $\{\mathbf{R}_m, \mathbf{T}_m\}$, the overall scale factor λ and the overall temperature factor B.

References

Castellano, E., Oliva, G. & Navaza, J. (1992). *J. Appl. Cryst.* **25**, 281–284.

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

DeLano, W. L. & Brünger, A. T. (1995). *Acta Cryst.* D**51**, 740–748.

Harada, Y., Lifchitz, A., Berthou, J. & Jolles, P. (1981). *Acta Cryst.* A**37**, 398–406.

Hirshfeld, F. L. (1968). *Acta Cryst.* A**24**, 301–311.

Huber, R. & Schneider, M. (1985). *J. Appl. Cryst.* **18**, 165–169.

Navaza, J. & Vernoslova, E. (1995). *Acta Cryst.* A**51**, 445–449.

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

Rossmann, M. G., Ford, G. C., Watson, H. C. & Banaszak, L. J. (1972). *J. Mol. Biol.* **64**, 237–245.

Tong, L. & Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 783–792.