RESEARCH PAPERS

AMoRe: an Automated Package for Molecular Replacement

BY J. NAVAZA

Laboratoire de Physique (UPR 180 du CNRS), Faculté de Pharmacie, 92290 Chatenay Malabry, France

Abstract

A new molecular-replacement package is presented. It is an improvement on conventional methods, based on more powerful algorithms and a new conception that enables automation and rapid solution.

Introduction

We start with a brief overview of molecular replacement in order to set the scene for the detailed description to follow. The technique has been extensively described in the literature (Rossmann, 1972; Machin, 1985). Three main steps may be distinguished: finding the orientation of the search model; locating its position; assessing the quality of the solution. The first and second steps are carried out by Patterson search techniques: the position of the model is varied until its calculated Patterson function matches the observed one, corresponding to the unknown structure. The orientation is usually obtained with the rotation function (RF) and, given the orientation, the displacement is determined using the translation function (TF). Eventually, the positional parameters are optimized by rigid-body refinement (RBR).

The six parameters that define the best rotational and translational fit are affected by several errors, including:

the intrinsic correlation between the rotational and translational parameters, which, for computational convenience, are treated separately;

poor quality of the data;

lack of similarity between search and target models;

a small size of search model as compared to the unit-cell content.

As a consequence, one ends up with many peaks of the combined RF and TF calculations. When many potential solutions have to be assessed, the procedure becomes difficult practically and requires much user interpretation.

The problems that molecular replacement encounters may be classified into three major types: those for which the solution is rather straightforward, those that are difficult yet solvable and those that are simply unsolvable by the method. A solvable problem may be tentatively defined as one for which an exhaustive six-dimensional search would give the correct solution.

The *AMoRe* (automatic molecular replacement) package is mainly aimed at the second class of problems. The usual situation is one in which the RF and/or TF are poorly contrasted and the correct solutions are not among the first peaks. The critical dependence of the outputs on the parameters that define the rotation and translation functions implies that, in such situations, the peak height is not a reliable criterion to use to select the correct solutions.

AMoRe is based on a new strategy: a compromise between automation, speed and performance. A profound analysis of the mathematical structures of the required functions was needed. Some of them are novel ones, in particular the *n*-body TF. Its success in a great number of cases, including very difficult problems, results from the combined action of the following main characteristics of the package:

(i) adequate functions are computed by powerful and fast algorithms;

(ii) many potential solutions are explored;

(iii) the information coming from models already positioned is automatically incorporated into the procedure;

(iv) the correlation coefficient is used as the main criterion of selection;

(v) there is a high degree of automation.

In the following sections, we describe the main programs of the *AMoRe* suite and illustrate its use in a nontrivial case.

Notation

$(\mathbf{M}_s|\mathbf{t}_s)$, $s = 1, ..., g$ Transformation $\mathbf{M}_s$ and translation vector $\mathbf{t}_s$ corresponding to the *s*th symmetry operation.

$\mathbf{H} = (H, K, L)$	Crystal reciprocal-space (row) vector.
$F_{\mathbf{H}}^{obs}$	Fourier coefficient of the crystal electron density.
$\mathbf{h} = (h, k, l)$	Continuous reciprocal-space vector.
$f(\mathbf{h})$	Continuous Fourier transform of the model electron density.
$\mathbf{R}(\alpha, \beta, \gamma)$	3×3 rotation matrix parametrized by Euler angles.
$\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$	Translation (column) vector in the crystal cell, in fractional coordinates.
$\mathbf{O}$	Matrix to transform fractional into orthogonal coordinates in the model cell.
$\mathbf{D}$	Matrix to transform orthogonal into fractional coordinates in the crystal cell.
$v = |\mathbf{v}|$, $\hat{\mathbf{v}} = \mathbf{v}/v$	Radial and angular parts of vector $\mathbf{v}$ (in direct or reciprocal space).
$j_l(v)$	Spherical Bessel function of order l.
$Y_{lm}(\hat{\mathbf{v}})$	Spherical harmonics of order l and projection m.

Structure-factor calculation

One of the cornerstones of the package is an extremely efficient algorithm to calculate the structure factors of the rotated and translated model by simple interpolation (Lattman & Love, 1970). For given rotation $\mathbf{R}(\alpha, \beta, \gamma)$ and translation $\mathbf{x}$, the calculated structure factor is given by

$$F_{\mathbf{H}}(\alpha, \beta, \gamma, x, y, z) = \sum_s f[\mathbf{HM}_s \mathbf{DR}(\alpha, \beta, \gamma)\mathbf{O}]$$
$$\times \exp(2\pi i \mathbf{H} \mathbf{t}_s) \exp(2\pi i \mathbf{HM}_s \mathbf{x}) \tag{1}$$

(Castellano, Oliva & Navaza, 1992). We see that the values of the continuous Fourier coefficients are needed at the reciprocal vectors $\mathbf{HM}_s \mathbf{DR}(\alpha, \beta, \gamma)\mathbf{O}$, which are not in general sampling points of $f(\mathbf{h})$; therefore, their computation requires interpolation. It was found that satisfactory results may be obtained by using a simple linear interpolation, provided that the model's center of mass is placed at the origin of a cell with linear dimensions about four times the size of the model.

The continuous Fourier coefficients are customarily computed by Fourier transformation of the electron density constructed on a Cartesian grid, based on the model's atomic coordinates. However, (1) applies to an arbitrary electron density as well.

Fast rotation function

The other cornerstone of the package is the ROTING program, which calculates the fast rotation function (Crowther, 1972). It uses expansions in spherical harmonics as in Crowther's formulation, but the radial variables are handled differently, leading to more accurate results. The basic formulas used in ROTING were derived in previous articles (Navaza, 1987, 1990), to which the reader is referred for details.

The rotation function takes the factorized form

$$R(\alpha, \beta, \gamma) = \sum_{l=0}^{\infty} \sum_{m,m'=-l}^{l} C_{mm'}^l D_{mm'}^l(\alpha, \beta, \gamma), \tag{2}$$

where $D_{mm'}^l$ are the matrices of the irreducible representations of the rotation group. They are written, in Eulerian angles, as

$$D_{mm'}^l(\alpha, \beta, \gamma) = d_{m,m'}^l(\beta) \exp[-i(m\alpha + m'\gamma)]; \tag{3}$$

the reduced matrices $d_{m,m'}^l(\beta)$ are calculated with the very stable recurrence relation (Navaza, 1990):

$$[(l - m' + 1)(l + m')]^{1/2} d_{m,m'-1}^l(\beta)$$
$$+ [(l + m' + 1)(l - m')]^{1/2} d_{m,m'+1}^l(\beta)$$
$$+ 2[m - m' \cos(\beta)] \sin(\beta)^{-1} d_{m,m'}^l(\beta) = 0, \tag{4}$$

starting from the expression

$$d_{m,l}^l(\beta) = [(2l)!/(l + m)!(l - m)!]^{1/2}$$
$$\times \sin(\beta/2)^{l-m} \cos(\beta/2)^{l+m}. \tag{5}$$

The $C_{mm'}^l$ depend on the intensities of the crystal and model and on the definition of the spherical domain of integration but not on angular variables. They are given in terms of the radial functions

$$c_{lm}(r) = 4\pi(-i)^l \sum_{\mathbf{H}} |F_{\mathbf{H}}|^2 j_l(2\pi Hr) Y_{lm}(\hat{\mathbf{H}})^* \tag{6}$$

by the one-dimensional integral

$$C_{mm'}^l = (3/4\pi)(b^3 - a^3)^{-1} \int_a^b c_{lm}^{(t)}(r) c_{lm'}^{(s)}(r)^* r^2 \, dr \tag{7}$$

[a and b are the inner and outer radii of the spherical domain and the labels (t) and (s) refer to the target crystal and the search model, respectively]. A closed analytical expression can be obtained by substituting (6) into (7) and using the well known properties of Bessel functions.

In practice, however, the $C_{mm'}^l$ are calculated by numerical integration. There are no severe limitations concerning the ratio of radius to resolution and the results are more accurate than those obtained with other algorithms (Alzari & Navaza, 1991; Strynadka, Adachi, Jensen, Johns, Sielecki, Betzel, Sutoh & James, 1992). The program ROTING first computes expression (6) for the crystal and the model, sampling the radial functions at the points of

a Gauss–Legendre quadrature formula. The spherical harmonics are computed with (4) and (5), putting $m = 0$. The spherical Bessel functions are computed using the relation

$$j_{l-1}(v) = [(2l+1)/v]j_l(v) - j_{l+1}(v). \tag{8}$$

Then, the $C^l_{mm'}$ are evaluated with the formula

$$C^l_{mm'} = \sum_{n=1}^{N} c^{(l)}_{lm}(r_n)c^{(s)}_{lm}(r_n)^* r_n^2 w_n, \tag{9}$$

where w_n and r_n are the integration weights and points, respectively. An optimal Gauss–Legendre quadrature formula is determined for each l; it depends on the data resolution and the interval of integration; changing any of these parameters implies changing the sampling of the radial functions.

Expression (2) is itself an expansion in terms of the complete and orthogonal set of functions $D^l_{mm'}$. The label l here plays a role similar to the frequency in Fourier synthesis: the $l = 0$ term is the average value of the RF – this simply adds a constant contribution; high-l terms imply high angular resolution. These properties are exploited by *ROTING* to enhance the resolution of the rotation peaks and to cast the RF output in the form of a correlation coefficient.

Translation function

The package uses the translation function defined as the overlap of the observed and calculated Patterson functions, for a given orientation of the model (Crowther & Blow, 1967; Harada, Lifchitz, Berthou & Jolles, 1981):

$$T(\mathbf{x}) = \sum_{\mathbf{H}} m_{\mathbf{H}}\Delta|F^{obs}_{\mathbf{H}}|^2|F_{\mathbf{H}}(\alpha,\beta,\gamma,x,y,z)|^2$$

$$= \sum_{\mathbf{H}} m_{\mathbf{H}}\Delta|F^{obs}_{\mathbf{H}}|^2\sum_{s,s'}\tilde{f}(\mathbf{h}_s)^*\tilde{f}[\mathbf{h}_{s'}]$$

$$\times \exp[-2\pi i\mathbf{H}(\mathbf{t}_s - \mathbf{t}_{s'})]$$

$$\times \exp[-2\pi i\mathbf{H}(\mathbf{M}_s - \mathbf{M}_{s'})\mathbf{x}]. \tag{10}$$

with $\Delta|F^{obs}_{\mathbf{H}}|^2 = |F^{obs}_{\mathbf{H}}|^2 - \langle|F^{obs}_{\mathbf{k}}|^2\rangle_{\mathbf{k}}$. $\tilde{f}(\mathbf{h}_s)$ stands for $\tilde{f}[\mathbf{HM}_s\mathbf{DR}(\alpha,\beta,\gamma)\mathbf{O}]$ and $m_{\mathbf{H}}$, the multiplicity of reflection $\mathbf{H}$, is introduced to restrict the sum to the asymmetric unit (including Friedel's symmetry). The vectors $\{\mathbf{H}(\mathbf{M}_s - \mathbf{M}_{s'})\}$ define a Bravais lattice that does not in general coincide with the crystal one; it displays the translational symmetry of $T(\mathbf{x})$. The corresponding direct cell is precisely the Cheshire cell (Hirshfeld, 1968).

The preceding formulas may be easily modified to incorporate the information concerning already placed models. We have simply to add to $F_{\mathbf{H}}(\alpha,\beta,\gamma,x,y,z)$ their contribution $F'''_{\mathbf{H}}$. This gives the extra term (skipping translation-invariant contributions),

$$\sum_{\mathbf{H}} m_{\mathbf{H}}\Delta|F^{obs}_{\mathbf{H}}|^2\sum_s \{\tilde{f}(\mathbf{h}_s)F'''^*_{\mathbf{H}} \exp[2\pi i\mathbf{H}(\mathbf{M}_s\mathbf{x} + \mathbf{t}_s)]$$

$$+ \tilde{f}[\mathbf{h}_s]^*F'''_{\mathbf{H}} \exp[-2\pi i\mathbf{H}(\mathbf{M}_s\mathbf{x} + \mathbf{t}_s)]\}. \tag{11}$$

The *TRAING* program incorporated in *AMoRe* computes the T function using fast Fourier transform (FFT) techniques. In the one-body case, the search is performed on the Cheshire cell. Otherwise (the n-body TF), the search is performed on the whole crystal cell. The values of T are only used as a means of selecting reasonable peaks. In fact, the output is the R factor

$$\mathrm{Rf} = \sum_{\mathbf{H}} m_{\mathbf{H}}||F^{obs}_{\mathbf{H}} - |F_{\mathbf{H}}(\alpha,\beta,\gamma,x,y,z)||$$

$$\times \left(\sum_{\mathbf{H}} m_{\mathbf{H}}|F^{obs}_{\mathbf{H}}|\right)^{-1} \tag{12}$$

and the correlation coefficient

$$\mathrm{Cc} = \sum_{\mathbf{H}} m_{\mathbf{H}}\Delta|F^{obs}_{\mathbf{H}}||\Delta|F_{\mathbf{H}}(\alpha,\beta,\gamma,x,y,z)|$$

$$\times \left\{\left[\sum_{\mathbf{H}} m_{\mathbf{H}}(\Delta|F^{obs}_{\mathbf{H}}|)^2\right]\right.$$

$$\times \left.\left[\sum_{\mathbf{H}} m_{\mathbf{H}}[\Delta|F_{\mathbf{H}}(\alpha,\beta,\gamma,x,y,z)|]^2\right]\right\}^{-1/2} \tag{13}$$

($\Delta|F_{\mathbf{H}}|$ stands for $|F_{\mathbf{H}}| - \langle|F_{\mathbf{K}}|\rangle_{\mathbf{K}}$), corresponding to the highest peaks of T.

Fast rigid-body refinement

This procedure was first proposed by Huber & Schneider (1985); the algorithms used in *AMoRe* are those of the *FITING* program described by Castellano *et al.* (1992). A quadratic misfit is minimized with respect to the positional parameters, the overall scale factor λ and the overall temperature factor B,

$$Q(\lambda,B,\alpha,\beta,\gamma,x,y,z) = \sum_{\mathbf{H}} m_{\mathbf{H}}\{|F^{obs}(\mathbf{H})|\exp(-B|\mathbf{H}|^2)$$

$$- \lambda|F_{\mathbf{H}}(\alpha,\beta,\gamma,x,y,z)|\}^2. \tag{14}$$

If there are N search models, the calculated Fourier coefficient $F_{\mathbf{H}}$ is the sum of N contributions like (1). If the individual model Fourier transforms f_n have been set to a common scale, they are

$$F_{\mathbf{H}} = \sum_n \sum_s f_n[\mathbf{HM}_s\mathbf{DR}(\alpha_n,\beta_n,\gamma_n)\mathbf{O}]$$
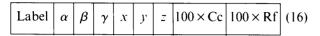
$$\times \exp(2\pi i\mathbf{Ht}_s)\exp(2\pi i\mathbf{HM}_s\mathbf{x}_n). \tag{15}$$

The minimization procedure (least-squares) involves the derivatives of Q, which are calculated starting from analytical expressions. At each cycle, the optimal λ and B are determined and the minimization with respect to the positional parameters is

alternately performed for each search model while the others are kept fixed.

Strategy

The package allows for multiple inputs to the individual programs. The outputs are lists of potential solutions (see the tables described later). Each output line

Label	α	β	γ	x	y	z	$100 \times Cc$	$100 \times Rf$	(16)

represents a potential position corresponding to a particular model; it contains all the information required by the package: a label to identify the model, three angles and three translations to position the model within the crystal unit cell, the correlation coefficient [(13)] and R factor [(12)]. In the case of n-body outputs, the last two items refer to a whole n-body configuration. In the case of RF outputs, Cc measures the correlation between Patterson functions restricted to the spherical domain of integration; translations and R factors are set to zero. The label is, in general, the logical number of the file containing the model's continuous Fourier coefficients.

Two preliminary programs must be run in order to cast the data into a suitable representation: SORT-ING, which sorts, packs and assesses the quality of measured data, and TABLING, which calculates the continuous Fourier coefficients corresponding to the different search models. These coefficients are obtained by Fourier transforming the electron density based on the atomic coordinates of the molecule, which is placed within a cell with linear dimensions about four times the size of the model. To avoid using a big cell resulting from an unfavorable position of the molecule, the center of mass of the atoms is shifted to the origin and the coordinates are rotated in order to align their principal axes of inertia with the cell axes; the resulting position is taken as the initial reference one.

The ROTING program is then used to calculate the cross-RF. The integration radius is usually taken as the distance from the center of mass to the farthest atom (the inner radius is systematically set to zero). The structure factors corresponding to the search model are computed by interpolation, taking a $P1$ cell of size equal to the smallest box containing the model, plus the integration radius, plus the requested resolution (this cell is substantially smaller than that used to compute the continuous Fourier coefficients). The standard output contains all peaks greater than 50% of the maximum value of the RF.

The TRAING program is then used to compute the one-body TF. The input is essentially the whole RF output. For each line (16), the model, identified

by its label, is rotated by the specified angles and translated within the Cheshire cell. The standard output contains the top ten peaks greater than 50% of the maximum value of the TF, sorted in descending order of their correlation coefficient.

If more than one molecule is to be placed within the asymmetric unit, a series of n-body TFs are executed. At each stage, a potential solution of the preceding step (in general that with greatest correlation coefficient) provides the fixed contribution to the n-body TF [(11)].

Finally, FITING is used to refine the positional parameters of the potential solutions. The lines of its output contain a further column with the values of the overall B factors [(14)]. Although this program is commonly used at the final stage of the whole procedure, it may be profitable to run it after each n-body TF.

AMoRe may be executed in a completely automatic way; although decisions are worth taking by the user in order to avoid unnecessary computations. In fact, three levels of automation may be distinguished. At the lowest level, the inputs to the individual programs are prepared by the crystallographer. Automation is here reduced to the possibility of multiple inputs to the n-body TF and RBR programs. The code distributed by the CCP4 Working Group (SERC Daresbury Laboratory, 1979) essentially works at this level; most parameters defining the individual functions are controlled by the user.

At the second level of automation, an intermediate code, OIC, prepares first-level inputs in terms of the available outputs. The code performs a number of operations, including a rudimentary packing analysis, in order to retain only the most promising solutions. At this level, the user controls the criteria of selection and some options of the individual programs.

A third level of automation was also developed: all the inputs to OIC are generated by the JOB code. This code produces, besides the second-level inputs, a user-friendly tentative command file (only for Unix and VMS operating systems) for the whole procedure, following the strategy presented in this section. The feasibility of such highly automated calculations is illustrated in the following section.

Example

We present here a succinct description of the automated procedure used to solve the crystal structure of barstar (Guillet, Lapthorn, Fourniat, Benoit, Hartley & Mauguen, 1993). Barstar crystallizes in space group $P6$ ($a = b = 143.6$, $c = 35.24$ Å, $\gamma = 120$), with four similar molecules in the asymmetric unit. Data were used between 10 and 4 Å resolution.

Table 1. RF output

The two numbers in the line after the header indicate, respectively, that this is a one-body output and that there are 31 potential orientations.

ROTING: **BARSTAR HEXAGONAL**

		1	31						
#	11	32.64	61.90	208.33	0.00000	0.00000	0.00000	17.4	0.0
	11	-0.52	66.05	147.87	0.00000	0.00000	0.00000	15.7	0.0
#	11	5.12	63.76	208.59	0.00000	0.00000	0.00000	15.1	0.0
	11	1.85	124.28	164.77	0.00000	0.00000	0.00000	14.4	0.0
	11	8.09	114.23	330.97	0.00000	0.00000	0.00000	13.9	0.0
#	11	36.16	115.37	27.61	0.00000	0.00000	0.00000	13.5	0.0
#	11	55.41	109.06	33.53	0.00000	0.00000	0.00000	13.5	0.0
	11	59.13	118.18	23.66	0.00000	0.00000	0.00000	12.8	0.0
	11	29.73	113.81	328.75	0.00000	0.00000	0.00000	12.5	0.0
	11	50.32	141.66	268.46	0.00000	0.00000	0.00000	12.3	0.0
...									

Table 2. One-body-TF output

Each group of potential solutions is preceded by an indication of the type of output (one-body in this case) and the number of solutions in the group, respectively. For the correct orientations, two peaks are displayed in order to show the contrast in correlation coefficient. For the misoriented monomer, none of the proposed solutions correspond to the true position.

TRAING: **BARSTAR HEXAGONAL**

		1	10						
#	11	32.64	61.90	208.33	0.46385	0.70478	0.00000	19.3	52.5
	11	32.64	61.90	208.33	0.92108	0.97570	0.00000	16.8	53.3
...									
	11	59.48	66.05	147.87	0.02526	0.99573	0.00000	12.7	54.5
...									
#	11	5.12	63.76	208.59	0.01380	0.34271	0.00000	22.7	51.5
	11	5.12	63.76	208.59	0.06824	0.03270	0.00000	17.4	53.7
...									
	11	1.85	124.28	164.77	0.11464	0.09517	0.00000	14.4	55.1
...									
	11	8.09	114.23	330.97	0.68550	0.25260	0.00000	14.6	54.2
...									
#	11	36.16	115.37	27.61	0.89118	0.10618	0.00000	17.0	53.2
	11	36.16	115.37	27.61	0.64838	0.36248	0.00000	15.7	53.1
	11	36.16	115.37	27.61	0.96577	0.97482	0.00000	15.5	54.7
	11	36.16	115.37	27.61	0.12592	0.20634	0.00000	15.5	54.1
	11	36.16	115.37	27.61	0.97993	0.03868	0.00000	15.4	54.1
	11	36.16	115.37	27.61	0.95748	0.50029	0.00000	14.8	54.2
	11	36.16	115.37	27.61	0.73938	0.35214	0.00000	14.8	54.0
	11	36.16	115.37	27.61	0.92334	0.95002	0.00000	13.6	55.4
	11	36.16	115.37	27.61	0.10519	0.07609	0.00000	13.3	54.5
	11	36.16	115.37	27.61	0.02499	0.99745	0.00000	12.2	57.6
#	11	55.41	109.06	33.53	0.88621	0.41959	0.00000	21.3	51.8
	11	55.41	109.06	33.53	0.00830	0.95063	0.00000	17.0	53.1
...									
	11	59.13	118.18	23.66	0.65398	0.33444	0.00000	15.4	54.1
...									
	11	29.73	113.81	328.75	0.55199	0.96929	0.00000	13.1	54.1
...									
	11	50.32	141.66	268.46	0.93459	0.99270	0.00000	15.0	53.6
...									

Almost all the information required by *JOB* is contained in the preceding paragraph: cell parameters, space-group symmetry, resolution range, number of different models, number of times each model is expected to appear within the asymmetric unit. *JOB* further requests the number of orientations to be tested (ten in this example) and a distance cut-off (10 Å in this case) to eliminate unphysical solutions.

The true orientations were among the top ten peaks of the RF (Table 1); one of them was more than 15° off its final value (*i.e.* the final rotation may be obtained from the original one by turning 15° about a certain axis). The positions, within the two-dimensional Cheshire cell, of the three more accurate orientations were determined by the one-body TF (Table 2). The contrast of the correlation coefficient was significant, although some solutions were not at the top of the overlap function. The correctness of these solutions was confirmed and their relative positions within the crystallographic cell determined by the subsequent *n*-body TF (*n* = 2, 3). The true translation of the misoriented monomer appeared only in a poorly contrasted four-body-TF output (Table 3). Eventually, the RBR program revealed the correct solution (Table 4). A final program, *MR2IC*, produced the rotation matrices and translations to apply to the original input atomic model. The execution of the whole procedure took about 36 min of CPU time, on a MicroVAX3100/80 (about 9 min on the IBM3090 of CIRCE). First-level automation would have required less CPU time but would have required manual interventions at intermediate steps.

The solution of the barstar structure makes clear some of the advantages of *AMoRe* and illustrates its new philosophy: instead of changing the parameters that define the different functions or improving the search model in order to promote the correct peaks to the top of the RF/TF outputs, we explore many potential solutions with adequate functions and incorporate pertinent information as it is acquired.

The model is changed only in the absence of a detached solution. In our experience, the notion of 'significance level of RF and TF peaks' is not a useful one for moderately complicated problems.

Completely automated calculations are recommended only for small-size problems. In the general case, a safe approach is to split the command file produced by *JOB* into its individual steps and execute them one after the other, modifying the second-level inputs (selection criteria, number of solutions to test) according to the intermediate results.

Miscellaneous results

AMoRe has solved a number of difficult structures. The first nontrivial case was the antibody–antigen complex F9.13.7–GEL (Lescar, Riottot, Souchon, Chitarra, Bentley, Navaza, Alzari & Poljak, 1993),

Table 3. *Four-body-TF output*

The top positions in each group correspond to the fixed molecules. The Cc and Rf values take into account their contributions.

TRAING: **BARSTAR HEXAGONAL**

		4	10						
#	11	5.12	63.76	208.59	0.01380	0.34271	0.00000		
#	11	55.41	109.06	33.53	0.88588	0.41925	0.67790		
#	11	32.64	61.90	208.33	0.46875	0.70060	0.90209		
#	11	36.16	115.37	27.61	0.73311	0.90005	0.78587	41.3	46.2
	11	36.16	115.37	27.61	0.66756	0.67290	0.89507	38.9	47.6

		4	10						
	11	5.12	63.76	208.59	0.01380	0.34271	0.00000		
	11	55.41	109.06	33.53	0.88588	0.41925	0.67790		
	11	32.64	61.90	208.33	0.46875	0.70060	0.90209		
	11	59.13	118.18	23.66	0.62428	0.31454	0.05270	39.6	47.1

		4	10						
	11	5.12	63.76	208.59	0.01380	0.34271	0.00000		
	11	55.41	109.06	33.53	0.88588	0.41925	0.67790		
	11	32.64	61.90	208.33	0.46875	0.70060	0.90209		
	11	50.32	141.66	268.46	0.49528	0.34422	0.69027	39.5	46.7

		4	10						
	11	5.12	63.76	208.59	0.01380	0.34271	0.00000		
	11	55.41	109.06	33.53	0.88588	0.41925	0.67790		
	11	32.64	61.90	208.33	0.46875	0.70060	0.90209		
	11	8.09	114.23	330.97	0.39918	0.32265	0.12672	39.5	46.8

		4	10						
	11	5.12	63.76	208.59	0.01380	0.34271	0.00000		
	11	55.41	109.06	33.53	0.88588	0.41925	0.67790		
	11	32.64	61.90	208.33	0.46875	0.70060	0.90209		
	11	1.85	124.28	164.77	0.38716	0.91789	0.80365	39.4	46.6

		4	10						
	11	5.12	63.76	208.59	0.01380	0.34271	0.00000		
	11	55.41	109.06	33.53	0.88588	0.41925	0.67790		
	11	32.64	61.90	208.33	0.46875	0.70060	0.90209		
	11	29.73	113.81	328.75	0.31496	0.75751	0.79134	38.8	47.1

		4	10						
	11	5.12	63.76	208.59	0.01380	0.34271	0.00000		
	11	55.41	109.06	33.53	0.88588	0.41925	0.67790		
	11	32.64	61.90	208.33	0.46875	0.70060	0.90209		
	11	59.48	66.05	147.87	0.61046	0.35358	0.85019	38.9	47.2

Table 4. *RBR output*

The last column is the supplementary overall B factor (not refined in this case).

FITING: **BARSTAR HEXAGONAL**

		4	1							
#	11	2.86	64.51	208.40	0.01529	0.34310	0.00463	67.9	34.3	0.
#	11	57.69	107.71	34.89	0.88582	0.41786	0.66325	67.9	34.3	0.
#	11	28.23	62.18	208.23	0.46716	0.70382	0.91609	67.9	34.3	0.
#	11	31.11	104.00	32.24	0.73025	0.89819	0.76588	67.9	34.3	0.
		4	1							
	11	2.27	64.32	209.22	0.01542	0.34452	0.00511	46.8	43.4	0.
	11	58.12	107.58	35.58	0.88635	0.41779	0.67724	46.8	43.4	0.
	11	28.17	62.37	208.57	0.46700	0.70261	0.91900	46.8	43.4	0.
	11	48.57	142.36	265.93	0.49666	0.34349	0.68978	46.8	43.4	0.
		4	1							
	11	2.43	64.07	208.59	0.01470	0.34401	0.00369	48.6	43.0	0.
	11	58.26	107.70	35.87	0.88573	0.41741	0.66942	48.6	43.0	0.
	11	28.42	62.71	208.96	0.46765	0.70396	0.91571	48.6	43.0	0.
	11	1.05	125.93	162.44	0.38609	0.91945	0.79012	48.6	43.0	0.
		4	1							
	11	1.76	64.29	208.36	0.01551	0.34404	0.00147	47.7	44.1	0.
	11	58.75	108.21	35.29	0.88578	0.41700	0.66075	47.7	44.1	0.
	11	28.77	62.77	208.82	0.46752	0.70271	0.91269	47.7	44.1	0.
	11	59.12	66.68	149.44	0.61092	0.35477	0.83413	47.7	44.1	0.

which crystallizes in the space group $P2_1$ with two complex molecules in the asymmetric unit (a.u.). The constant and the variable dimers of FabD1.3 (19% of the a.u. content) and triclinic hen lysozyme (13% of the a.u. content) were used as separate search models. Another example is the structure of the antibody–antigen complex D44–HEL, which crystallizes in the space group $P2_12_12_1$ ($a = 99.7$, $b = 167.3$, $c = 84.7$ Å) with two complex molecules in the asymmetric unit. The search models were different Fab molecules solved in the Pasteur Institute and triclinic hen lysozyme. The correct orientations of lysozyme were the 24th and 76th peaks of the RF. This was revealed by the n-body TF when both Fab molecules were already correctly placed.

Two nontrivial cases which constitute a nice illustration of the role of molecular replacement in crystal structure determination are the barnase–barstar complex (Guillet, Lapthorn, Hartley & Mauguen, 1994) and barstar (Guillet *et al.*, 1993). Only the barnase molecule (Mauguen, Hartley, Dodson, Dodson, Bricogne, Chothia & Jack, 1982) was known. Barnase–barstar crystallizes in space group $C2$ ($a = 205.4$, $b = 44.4$, $c = 88.3$ Å, $\beta = 110.5$), with three complex molecules in the asymmetric unit. The structure was solved by molecular replacement using the barnase model (20% of the a.u. content) and density averaging, exploiting the noncrystallographic symmetry. This provided the model to solve the barstar structure, as explained in the preceding section.

Some other examples include structures where standard software failed but which eventually were easy cases for *AMoRe*:

(1) A mutant of the tumor necrosis factor (TNF) (Saludjian, Prangé, Kahn, Fourme, Tavernier, Van Ostade, Fiers & Navaza, 1994), which crystallizes in space group $P4_12_12$ ($a = b = 166.2$, $c = 93.8$ Å, $\gamma = 120$), with two trimers of TNF in the asymmetric unit. Here the search model was a complete trimer.

(2) A catalytic antibody (M. Bizebard, T. Knossow, P. Saludjian & J. Navaza, 1992, unpublished), which crystallizes in space group $P1$ ($a = 99.4$, $b = 68.1$, $c = 83.7$ Å, $\alpha = 71.9$, $\beta = 112.1$, $\gamma = 119.6$), with four Fab molecules in the a.u.; the search models were the constant and variable parts of a Fab molecule.

(3) A proteinase, eglin C (Betzel, Dauter, Genov, Lamzin, Navaza, Schnebli, Visanji & Wilson, 1993), which crystallizes in space group $P4_3$ ($a = b = 42.0$, $c = 35.9$ Å) with one molecule in the a.u.

(4) Fasciculine II (Le Du, 1992), which crystallizes in space group $P4_12_12$ ($a = b = 49.0$, $c = 82.0$ Å), with one molecule in the a.u.; the search model was fasciculine I.

(5) A mutant of phosphoglycerate kinase, crystallizing in space group $P2_12_12_1$ with one molecule in

the a.u.; 36% of the data were almost uniformly missing in the 15 to 3.5 Å resolution range. Although poorly contrasted, the correct orientation was at the top of the RF.

(6) Many other structures solved after the distribution of the package and during two workshops on molecular replacement held at CIRCE, Orsay, in April and September 1992.

Future developments

The rotation function is perhaps the crucial step of the whole procedure. Some improvements are undergoing testing and development, in particular concerning the peak-search subroutine. Sometimes a correct orientation is not in the RF output because it is placed within the profile of a massive peak. Situations like this are rather common in multifragment search problems where the difficulty cannot be ascribed to the search model. The RF should include some measure of the quality of the peak, for example its extension, to allow sampling within the peak profile. This means using the distance in the rotation-parameter space, which poses some problems because the rotation group is not simply connected (Navaza, 1987).

Concerning the translation function, a program has been developed that computes the correlation coefficient of intensities by fast-Fourier-transform techniques (M. Sarrazin, 1992, unpublished). The computations are more time consuming because three FFTs, at four times the data resolution, are needed.

Some options of *AMoRe*, although available, have not been extensively tested, in particular the locked RF and the phased TF.

References

ALZARI, P. & NAVAZA, J. (1991). *Crystallographic Computing* 5, edited by D. MORAS, A. D. PODJARNY & J. C. THIERRY, pp. 348–360. IUCr/Oxford Univ. Press.

BETZEL, C., DAUTER, Z., GENOV, N., LAMZIN, V., NAVAZA, J., SCHNEBLI, H. P., VISANJI, M. & WILSON, K. (1993). *FEBS Lett.* 317, N3, 185-188.

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 and Breach.

CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* 23, 544-548.

GUILLET, V., LAPTHORN, A., FOURNIAT, J., BENOIT, J.-P., HARTLEY, R. W. & MAUGUEN, Y. (1993). *Prot. Struct. Funct. Genet.* In the press.

GUILLET, V., LAPTHORN, A., HARTLEY, R. W. & MAUGUEN, Y. (1994). In preparation.

HARADA, Y., LIFCHITZ, A., BERTHOU, J. & JOLLES, P. (1981). *Acta Cryst.* A37, 398–406.

HIRSHFELD, D. (1968). *Acta Cryst.* 24, 301.

HUBER, R. & SCHNEIDER, M. (1985). *J. Appl. Cryst.* 18, 165 169.

LATTMAN, E. E. & LOVE, W. E. (1970). *Acta Cryst.* B26, 1854 1857.

LE DU, M. H. (1992). Thesis, Univ. of Aix-Marsille II, France.

LESCAR, J., RIOTTOT, M.-M., SOUCHON, M., CHITARRA, V., BENTLEY, G., NAVAZA, J., ALZARI, P. & POLJAK, R. (1993). *Prot. Struct. Funct. Genet.* 15, 209–212.

MACHIN, P. A. (1985). Editor. *Molecular Replacement.* Proceedings of the Daresbury Study Weekend, 15–16 February 1985. SERC Daresbury Laboratory, Warrington, England.

MAUGUEN, Y., HARTLEY, R. W., DODSON, E. J., DODSON, G. G., BRICOGNE, G., CHOTHIA, C. & JACK, A. (1982). *Nature (London)*, 297, 162 164.

NAVAZA, J. (1987). *Acta Cryst.* A43, 645 653.

NAVAZA, J. (1990). *Acta Cryst.* A46, 619-620.

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

SALUDJIAN, P., PRANGÉ, T., KAHN, R., FOURME, R., TAVERNIER, J., VAN OSTADE, X., FIERS, W. & NAVAZA, J. (1994). In preparation.

SERC Daresbury Laboratory (1979). *CCP4. A Suite of Programs for Protein Crystallography.* SERC Daresbury Laboratory, Warrington, England.

STRYNADKA, C. J., ADACHI, H., JENSEN, S. E., JOHNS, K., SIELECKI, A., BETZEL, C., SUTOH, K. & JAMES, M. N. G. (1992). *Nature (London)*, 359, 700 705.