Acta Cryst. (1998). D**54**, 817–821

On the Use of Strong Patterson Function Signals in Many-Body Molecular Replacement

JORGE NAVAZA,* EZEQUIEL HORACIO PANEPUCCI† AND CAROLE MARTIN

ERS 582 du CNRS, Laboratoire de Physique, Centre d'Etudes Pharmaceutiques. 92290 Chatenay-Malabry, France. E-mail: jorge.navaza@cep.u-psud.fr

(*Received 7 November 1997; accepted 15 December 1997*)

Abstract

The symmetry elements detected by the self-rotation and the Patterson functions, associated with strong correlations between the positions of the molecules in the asymmetric unit, are used to reduce the effective number of independent bodies to be located by the molecular replacement method. A distinction is made between 'frustrated' crystallographic symmetries, *i.e.* those that are almost crystallographic ones, and 'standard' non-crystallographic symmetries, which are taken into account by specific techniques. These have been successfully applied to many-body macromolecular crystal structures, with important savings in time and computational effort.

1. Introduction

Crystal structures containing many independent molecules in the asymmetric unit are, in general, difficult problems to solve by molecular replacement (MR). The reasons for the difficulties cannot be ascribed solely to the inherent limitation of the usual formulation of MR, *i.e.* rotation searches performed independently of the translation ones. Indeed, very often the best values of rather robust criteria correspond to incorrect one-body positions. Therefore, it is the overall strategy of the method, based on locating one body after the other, that fails if the first bodies are not correctly placed.

Automation and many-body searches have substantially improved the performance of the method. Automation allows one not only to test many potential orientations but also to write scripts for the many-body translation problem where the contribution of the fixed molecules is alternately chosen from a list of best available positions. This combinatorial approach has proved to be efficient in a number of cases, but may lead to extremely lengthy calculations.

In this paper we discuss how the effective number of independent molecules may be reduced by exploiting the information provided by non-crystallographic symmetries. These are detected by the self-rotation and the Patterson functions, whose prominent peaks indicate rotational and translational symmetries, respectively. We will make a distinction between 'frustrated' crystallographic symmetries (FCS) and 'standard' non-crystallographic symmetries (NCS). As suggested by the name, FCS are almost crystallographic symmetries. Their signals are typically stronger than those of NCS, but this is not a general rule. FCS and NCS are taken into account by specific methods, as described in the following sections.

2. Taking FCS into account

The information provided by an FCS is customarily used to perform a cell transformation which renders the approximate symmetry a crystallographic one.

A rotational FCS may be detected by inspection of the diffraction pattern or, more easily, by means of the self-rotation function. The computation of the $R_{\rm sym}$ factor, assuming the frustrated space group, is needed to assess the quality of the new incorporated symmetry. Indeed, peaks with high values of correlation in the self rotation do not necessarily indicate a frustrated symmetry; peculiar packings may give rise to extremely high correlations. A remarkable example is provided by a hexameric DNA structure (Urpi, 1996) where the molecules are placed in parallel but rather distant layers. Each layer displays the same twofold symmetry axis which gives a correlation of 90% with data between 15 and 3.5 Å and an integration radius equal to that of the molecule. Since the axis does not intercept any grid point of the Bravais lattice, frustration can be ruled out.

A translational FCS shows up as a strong peak in the Patterson function, whose fractional coordinates – simple fractions – indicate the amount of reduction to be applied to the cell parameters. A cell reduction implies disregarding a big fraction of measured intensities and re-indexing. The ratio of the sum of eliminated over the sum of total intensities, within a given resolution range, may be used to measure the acceptability of the forced symmetry. In general, this ratio increases with resolution.

The structure of a catalytic antibody Fab (Golinelli-Pimpaneau *et al.*, 1994) constitutes an example of translational FCS. The crystal has $P1$ symmetry and eight molecules in the unit cell. The Patterson function shows two massive peaks at $(\frac{1}{2},0,0)$ and $(0,0,\frac{1}{2})$ whose

† Present address: DFQ-IQSC-USP São Carlos, SP, Brazil.

heights are greater than 90% of the origin peak value. The structure was solved by halving **a** and **c**, which reduces the number of molecules to place by a factor of four.

2.1. *Combined rotational and translational FCS*

The crystal of a secreted aspartic protease (Abad-Zapatero *et al.*, 1996) is an example of combined rotational and translational FCS. The space group is $P1$ and there are eight independent molecules in the unit cell. The Patterson function has a prominent peak at $(\frac{1}{2},\frac{1}{2},0)$. Its height is 60% of the origin peak value when using data between 20 and 8 Å, but drops to 30% with data between 20 and 4 Å. The self rotation shows a massive peak corresponding to a twofold axis parallel to **b**−**a**. With a radius of integration of 35 Å – equal to the model's radius – and data in the 15–4.5 Å resolution range, the correlation coefficient is 80%. This twofold axis, together with the 'centering' $(\frac{1}{2},\frac{1}{2},0)$, corresponds in fact to a frustrated $C2$ symmetry, which is displayed in the new cell given by the transformation

$$\mathbf{a}' = (\mathbf{a} + \mathbf{b})/2$$
$$\mathbf{b}' = (\mathbf{b} - \mathbf{a})/2$$
$$\mathbf{c}' = \mathbf{c}$$

$\mathbf{b}'$ turns out to be almost perpendicular to $\mathbf{a}'$ and $\mathbf{c}'$, as required by $C2$. The new cell has half the original volume. Taking into account the twofold axis, the above transformation reduces the number of independent molecules by a factor of four. Two molecules have thus to be positioned by standard MR methods, from which the whole $P1$ cell can be filled.

3. Taking NCS into account

NCS is usually concerned with rotational symmetries. It is used in locked rotation and translation functions (Rossmann *et al.*, 1972; Tong & Rossmann, 1990; Tong, 1996): the former to enhance the signal-to-noise ratio of doubtful peaks of the cross rotation; the latter to position a monomer relative to the center of the NCS assembly.

The values of the locked cross-rotation function are the average of the cross-rotation values at orientations related by the NCS (Tong & Rossmann, 1990). For low-order rotational NCS, as is generally the case for proteins, the averaging may fail to promote the correct peaks to the first ranks. As an example of unsuccessful use we mention the case of Erabutoxin-b (Saludjian *et al.*, 1992). One orientation of the dimer appeared at the top of the cross rotation and the second one, linked by an NCS twofold axis, was ranked 36th. Spurious peaks, though linked by the NCS, gave values of the locked cross rotation higher than those corresponding to the true orientations of the dimer.

3.1. *Computing the locked cross-rotation function*

The locked cross-rotation function is usually calculated by interpolation. By redefining the target function it may be computed as an ordinary cross rotation, leading to more precise results.

Let Φ denote the three parameters representing a rotation, $\mathbf{R}(\Phi)$ its associated matrix, and Φ^{-1} the parameters of the corresponding inverse rotation. For any function $\mathcal{P}$ we define the rotated function $[\mathcal{R}(\Phi)\mathcal{P}]$ as

$$[\mathcal{R}(\Phi)\mathcal{P}](\mathbf{r}) = \mathcal{P}[\mathbf{R}(\Phi^{-1})\mathbf{r}]. \tag{1}$$

Then, given the Patterson functions $\mathcal{P}^{(t)}$ and $\mathcal{P}^{(s)}$, corresponding to the target and search crystals, the cross-rotation function is defined as the overlap of $\mathcal{P}^{(t)}$ and $[\mathcal{R}(\Phi)\mathcal{P}^{(s)}]$ (Rossmann & Blow, 1962),

$$\frac{1}{v}\int_{\Omega} \mathcal{P}^{(t)}(\mathbf{r})[\mathcal{R}(\Phi)\mathcal{P}^{(s)}](\mathbf{r})\mathrm{d}^3\mathbf{r} = \langle \mathcal{P}^{(t)} \| \mathcal{R}(\Phi)\mathcal{P}^{(s)} \rangle. \tag{2}$$

v is the volume of the region Ω of integration, generally chosen as a spherical shell. Now, if $\{\Phi_i; i = 1...g\}$ denotes the set of NCS rotations, including the null one, and using the unitary property of rotation operators, the locked cross-rotation function is given by

$$\begin{aligned} &\Big[\sum_{i=1}^{g} \langle \mathcal{P}^{(t)} \| \mathcal{R}(\Phi_i)\mathcal{R}(\Phi)\mathcal{P}^{(s)} \rangle\Big]/g \\ &= \Big\langle \Big\{\Big[\sum_{i=1}^{g} \mathcal{R}(\Phi_i^{-1})\mathcal{P}^{(t)}\Big]/g\Big\} \| \mathcal{R}(\Phi)\mathcal{P}^{(s)} \Big\rangle \end{aligned} \tag{3}$$

which is of the same form as (2), with the target Patterson function substituted by the average over the NCS of the rotated target functions.

The computation of the locked cross rotation is particularly simple in the case of the fast rotation function (Crowther, 1972). Indeed, the cross-rotation function takes the factorized form

$$\sum_{l=0}^{\infty} \sum_{m,m'=-l}^{l} C^{l}_{mm'} \mathcal{D}^{l}_{mm'}(\Phi), \tag{4}$$

which is an expansion in terms of the matrices $\mathcal{D}^{l}_{mm'}$ of the irreducible representations of the rotation group. The coefficients of the expansion depend on the intensities of both crystals and on the size of Ω; they are given by the expression

$$C^{l}_{mm'} = 12\pi \sum_{n=1}^{\infty} e^{(t)}_{lmn}\, e^{(s)*}_{lm'n} \tag{5}$$

in terms of target and search contributions (Navaza, 1993),

$$e_{lmn} = [2(l+2n)-1]^{1/2} \sum_{\mathbf{H}} |F_{\mathbf{H}}|^2 Y_{lm}(\hat{\mathbf{H}}) \frac{j_{l+2n-1}(2\pi HR)}{2\pi HR}. \tag{6}$$

Here, R denotes the radius of Ω, Y_{lm} the spherical harmonic, j_l the spherical Bessel function, and

$\hat{\mathbf{H}} = \mathbf{H}/H$, the angular part of vector $\mathbf{H}$. The substitution

$$e^{(t)}_{lmn} \rightarrow \sum_{m'=-l}^{l} \left\{ \left[\sum_{i=1}^{g} \mathcal{D}^{l}_{mm'}(\Phi_i^{-1}) \right] / g \right\} e^{(t)}_{lm'n} \qquad (7)$$

gives the locked cross rotation.

The transformation (7) is a minor modification of the fast rotation function, where most of the computing time is spent in the calculation of the coefficients e_{lmn}. It is applied, optionally, in the *ROTING* program of the *AMoRe* package (Navaza, 1994).

If the rotational NCS forms a group, we can replace the sum over $\{\Phi_i^{-1}\}$ by a sum over $\{\Phi_i\}$ because of the re-arrangement theorem of group theory. Only in this case peaks of the cross-rotation function related by the NCS will appear in the locked cross rotation with equal heights.

3.2. *Translational NCS*

Translational NCS has already been described in the literature (*e.g.* Yuhasz *et al.*, 1989). Here we will consider the particular translational NCS which arises when independent molecules have the same orientation. The crystal of an Fab structure (Abergel & Padlan, 1994) is an example of this kind of NCS. It has $P2_1$ symmetry and four independent molecules in the unit cell (a = 115.58, b = 116.424, c = 70.3 Å, β = 97.84°). Its Patterson function shows a peak at $\mathbf{T}$ = (0.206,0,0.482) having 26% of the origin peak value, which is rather insensitive to resolution. All other peaks have values less than 7%. The independent molecules are associated in pairs; each pair is formed by almost equally oriented molecules whose centers of mass are separated by the translation $\mathbf{T}$.

This example illustrates several features and traps of many-body searches. Although the correct orientations appear at the top of the cross-rotation function (first and third peaks), the one-body translation function and subsequent rigid-body refinement gives a wrong first position. An exhaustive six-dimensional search would have failed in this case. Starting from this wrong first position, up to three Fab's may be placed – the fourth one clashes. It is the second best one-body solution that leads to a correct result. Also, the particular packing of this structure leads to a peculiar result: the first self-rotation peak is a spurious one (correlation of 28% with real data and 50% with calculated data). Its use in a locked rotation would thus lead to incorrect results.

Based on this last example we tested the efficiency of the Patterson function in detecting translations relating two pairs of equally oriented molecules. Initially a data set was calculated for a crystal with four Fab molecules in the asymmetric unit, two of them with a given orientation and their centers of mass separated by the vector $\mathbf{T}$, and the other two in a different orientation but related by the same vector.

We applied random rotations ranging from 1 to 20° to one of the molecules and we calculated a Patterson function for each configuration. The same procedure was performed for the other pair of molecules. A third calculation was performed applying random rotations to one molecule of each pair. We see in Fig. 1 that the signal decreases much more rapidly when both pairs are being changed simultaneously. We also note that a strong Patterson signal is already present when only two of the molecules are aligned.

3.3. *Including translational NCS into the translation functions*

A simple modification of the input to the standard translation functions allows one to incorporate into the search procedure the information provided by a trans-

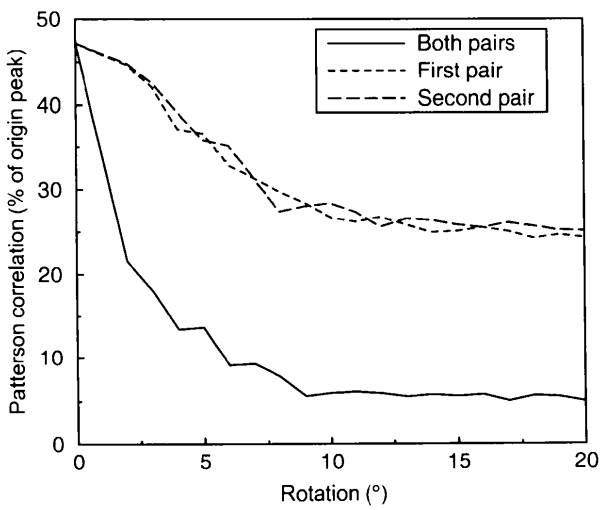

Fig. 1. Loss of the Patterson signal as a function of the mis-orientation of pairs of molecules.

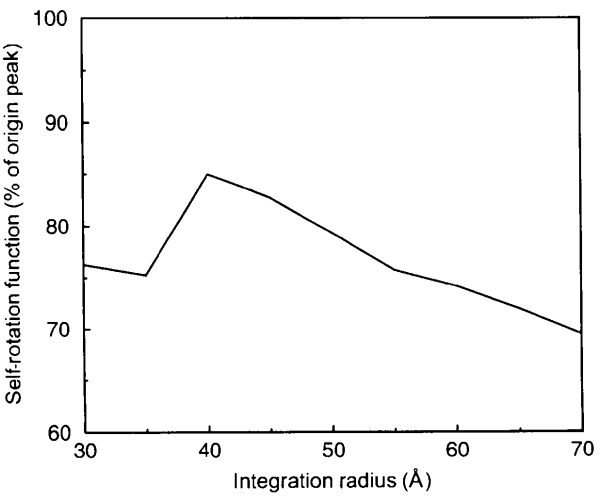

Fig. 2. Height of the highest non-crystallographic peak of the self-rotation function as a function of the integration radius, with data between 15 and 3.5 Å.

lational NCS. The idea is very simple: if the model to translate consists of N identical molecules having equal orientations, but separated by given vectors $\mathbf{T}_n, n = 1, N$, its Fourier coefficients are given in terms of those of the individual molecule by

$$F_H \rightarrow F_H \sum_{n=1}^{N} \exp(2\pi i \mathbf{H}\mathbf{T}_n).$$

Although this may be considered as a sort of locked-translation function, we are not adding the values taken by the functions at positions related by the NCS. This is merely a computationally economic way of changing the search model to incorporate the translational NCS.

4. A case with 16 independent molecules

Barstar, the natural intracellular inhibitor of the ribonuclease Barnase from *Bacillus amyloliquefaciens* (Guillet *et al.*, 1993), crystallizes in many different crystal forms under almost identical conditions. A tetragonal form has been characterized, with unit-cell dimensions $a = b = 109.89$, $c = 300.55$ Å, space group $I4_1$ and 16 molecules in the asymmetric unit.

The self rotation shows a strong peak corresponding to a twofold axis perpendicular to **c** and 35° off **a**, for almost any value of the integration radius (see Fig. 2). We could expect this behavior from a rotational FCS but, in this case, it is incompatible with the lattice. It is in fact a standard rotational NCS: as observed in all preceding crystallographic studies, Barstar crystallizes in dimeric form.

Using a dimer of Barstar as a search model, the structure could be solved straightforwardly by using the standard protocol of *AMoRe*. There were eight independent molecules to be positioned. The output of the one-body translation function left only one possible candidate, with correlation coefficient (Cf) of 0.313 and R factor (Rf) of 0.613; the next best position had values of 0.252 and 0.627, respectively. After rigid-body refinement, the whole configuration had Cf = 0.842 and Rf = 0.341. The eight independent molecules are almost parallel (we observed a dispersion of 5° about the average orientation) and are separated by about **c**/8. Fig. 3 (produced with *molpack*; Wang *et al.*, 1991) shows the crystal packing, projected along **c**.

When using a monomer as a search model, correct one-body positions were ranked third and fourth after rigid-body refinement, with Cf values of 0.25 and 0.246. A lengthy combinatorial approach gave the 16 positions corresponding to the correct structure.

The repeat vector **c**/8 gives four independent strong peaks in the Patterson function. With data from 15 to 5 Å resolution, their positions are (0,0,0.124), (0,0,0.247), (0,0,0.371) and (0,0,0.500). By using data at a higher resolution (15 to 3.5 Å) the last peak splits into (0,0,0.492) and its symmetry-related (0,0,0.508). Their heights are functions of the resolution, as shown in Fig. 4.

It was assumed that these Patterson peaks corresponded to a translational FCS, so that the **c** axis was divided by 8. Only one dimer had then to be positioned. The discarded reflections – about 7/8 of the total – represented 50% of the sum of intensities. In order to keep the same **a** and **b** axes, the space group was taken as $C4$. One solution was found with Cf = 0.664 and Rf = 0.447, and a good contrast: the corresponding values for the next best position were 0.477 and 0.514, respectively. The packing was essentially the same as in $I4_1$.

It is always unpleasant to discard large amounts of data, particularly in the present case where they repre-

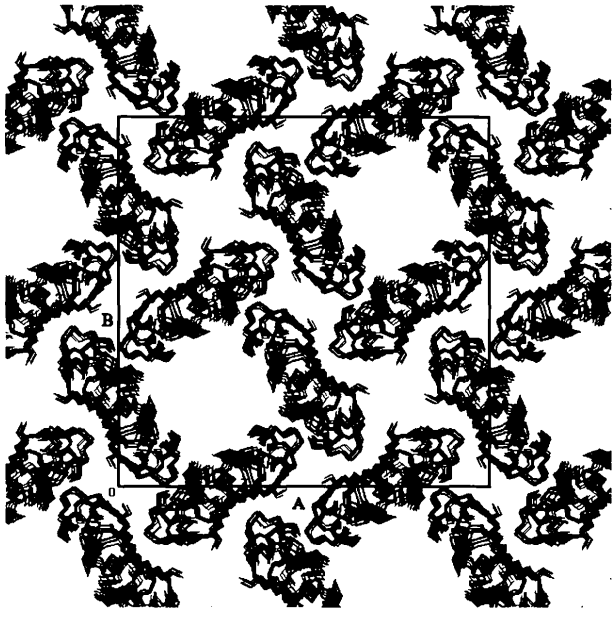

Fig. 3. Packing of the barstar crystal structure, projected along **c**.

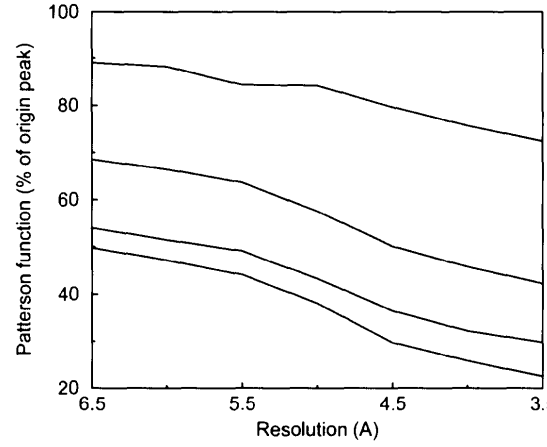

Fig. 4. Heights of the four highest Patterson peaks as a function of the resolution. The low-resolution limit is always 15 Å.

sent half of the diffracted intensity. However, translational FCS should be used in the determination procedure. Note that translational FCS may be handled as a standard translational NCS. This has the advantage of reducing the number of independent positions to be found without discarding any data. Taking the dimer as the search model, there is only one 'one-body' translation function to compute. The saving in time is important since the Cheshire cell for $I4_1$ is two dimensional. With the monomer, there is a further 'two-body' (three-dimensional) translation function to compute.

5. Conclusions

The examples discussed clearly show that the incorporation of FCS or NCS into the molecular replacement technique avoids some of the traps often encountered in many-body searches, and considerably reduces the computational effort.

One of us (EHP) has been supported by a fellowship from CNPq, Brazil.

References

Abad Zapatero, C., Goldman, R., Muchmore, S. W., Hutchins, C., Stewart, K., Navaza, J., Payne, C. D. & Ray, T. L. (1996). *Protein Sci.* **5**, 640–652.

Abergel, C. & Padlan, E. (1994). Personal communication.

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

Golinelli-Pimpaneau, B., Gigant, B., Bizebard, T., Navaza, J., Saludjian, P., Zemel, R., Tawfik, D., Eshhar, Z., Green B. & Knossow, M. (1994). *Structure*, **2**(3), 175–183.

Guillet, V., Lapthorn, A., Fourniat, J., Benoit, J.-P., Hartley, R. W. & Mauguen, Y. (1993). *Proteins Struct. Funct. Genet.* **17**, 325–328.

Navaza, J. (1993). *Acta Cryst.* D**49**, 588–591.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

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.* **21**, 872–876.

Saludjian, P., Prange, T., Navaza, J., Menez, R., Guilloteau, J.P. & Ducruix, A. (1992). *Acta Cryst.* B**48**, 520–531.

Tong, L. (1996). *Acta Cryst.* A**52**, 476–479.

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

Urpi, L (1996). Personal communication.

Wang, D., Driessen, H. P. C. & Tickle, I. J. (1991). *J. Mol. Graphics*, **9**(38), 50–52.

Yuhasz, S.C., Ysern, X., Strand, M. & Amzel, M. (1989). *J. Mol. Biol.* **209**, 319–321.