Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

An approach to multi-copy search in molecular replacement

Alexei Vagin[a]* and Alexei Teplyakov[b]

[a]Department of Chemistry, University of York, Heslington, York YO1 5DD, England, and [b]Center for Advanced Research in Biotechnology, Rockville, Maryland 20850, USA

Correspondence e-mail: alexei@ysbl.york.ac.uk

Received 28 June 2000
Accepted 4 October 2000

The molecular-replacement method has been extended to a simultaneous search for multiple copies of the macromolecule in the unit cell. The central point of this approach is the construction of a multi-copy search model from the properly oriented monomers using a special translation function. The multi-copy search method has been implemented in the program *MOLREP* and successfully tested using experimental data.

1. Introduction

The molecular-replacement method (Rossmann, 1972) is one of the principal techniques of macromolecular crystal structure determination. Usually, the process of locating a search model in the unit cell of the crystal of interest is divided into two three-dimensional searches. First, the orientation of the model is found using a rotation function (RF). Then, the properly oriented model is subjected to positional search using a translation function (TF). If some phases from other sources are available, the phased TF (Bentley, 1997) can be used. In spite of significant improvement of algorithms in recent years (Turkenburg & Dodson, 1996; Carter & Sweet, 1997), numerous cases remain unsolved owing to various reasons. The presence of multiple copies of the macromolecule (or several monomers of one complex molecule) in the asymmetric part of the unit cell can be an obstacle to the successful application of molecular replacement. The traditional approach to the structure determination in such cases would be first to locate one molecule, fix it and search for the next molecule. However, the location of the first molecule may appear difficult. In special cases of macromolecular assemblies with a simple point-group symmetry (typical for viruses), the locked RF (Tong & Rossmann, 1990) can be helpful. Its application assumes the *a priori* knowledge of the monomers arrangement in the assembly (the term 'monomer' is used throughout the text to define a protein or nucleic acid molecule or part of it which is approximated by a search model used for molecular replacement).

We have developed a general method for locating multiple copies of the monomer in the unit cell of an unknown crystal structure. The central point of the method is the construction of a multi-copy search model from the properly oriented monomers using a special TF. This model can then be used for a positional search with a conventional TF. The method does not impose any limitation on the oligomeric structure of the protein, either by the number of monomers or by their relative location (*i.e.* the pure rotational symmetry is not required). In principle, monomers may even be of different types. The method was implemented in the program *MOLREP*

(currently with two monomers of the same kind) and tested on a number of cases using experimental X-ray data.

2. Description of the method

Suppose the asymmetric part of the unit cell contains two monomers, which we call a dyad, and a search model of a homologous molecule is available. The process of structure determination using the multi-copy search is carried out in three steps.

In step 1, the two monomers are to be properly oriented. The RF is calculated for an initial search model. Nrf highest peaks of the RF are used to produce a set of oriented monomers. All possible pair combinations of these monomers constitute a set of putative dyads. The total number of these is $\mathrm{Nrf}^2/2$.

In step 2, for every putative dyad, the intermolecular vector relating two monomers of the dyad is to be determined. This is achieved by using a special TF.

The Patterson function calculated for the dyad, in which the centers of mass of the monomers are defined by vectors $\mathbf{s}_1$ and $\mathbf{s}_2$, is

$$\begin{aligned} P(\mathbf{h}) &= F(\mathbf{h})F^*(\mathbf{h}) \\ &= F_1(\mathbf{h})F_1^*(\mathbf{h}) + F_2(\mathbf{h})F_2^*(\mathbf{h}) \\ &\quad + F_1^*(\mathbf{h})F_2(\mathbf{h})\exp[-2\pi i\mathbf{h}(\mathbf{s}_2 - \mathbf{s}_1)] \\ &\quad + F_1(\mathbf{h})F_2^*(\mathbf{h})\exp[-2\pi i\mathbf{h}(\mathbf{s}_1 - \mathbf{s}_2)], \end{aligned}$$

where $F_1(\mathbf{h})$ and $F_2(\mathbf{h})$ are structure factors for the two monomers centered in the origin and $\mathbf{h}$ represents the reciprocal-space vectors.

This Patterson function contains three clusters of peaks. One is in the origin and arises from the intramolecular vectors of both monomers,

$$P(0) = F_1(\mathbf{h})F_1^*(\mathbf{h}) + F_2(\mathbf{h})F_2^*(\mathbf{h}).$$

Two other clusters represent the intermolecular vectors between the monomers. They are located in positions $(\mathbf{s}_2 - \mathbf{s}_1)$ and $(\mathbf{s}_1 - \mathbf{s}_2)$, which are related by the center of symmetry. These vectors can be determined using a phased TF. For this purpose, the Patterson function is considered to be an electron density in which the fictitious model represented by the corresponding cluster is to be located. To determine vector $(\mathbf{s}_2 - \mathbf{s}_1)$, the search model will be that described by the structure factors $F_1{}^*(\mathbf{h})F_2(\mathbf{h})$. The model with the structure factors $F_1(\mathbf{h})F_2{}^*(\mathbf{h})$ will define vector $(\mathbf{s}_1 - \mathbf{s}_2)$. No crystal symmetry is considered at this stage, *i.e.* the calculations are performed in the space group $P1$.

As a result of this procedure, we find the intermolecular vectors that relate monomers in the dyad. Step 2 is performed for each putative dyad. Several (Np) top solutions of the phased TF shall be considered to ensure that the correct dyad is not missed. Therefore, the total number of dyads selected for the final positional search will be $\mathrm{Nrf}^2/2 \times \mathrm{Np}$.

In step 3, a positional search for each dyad is performed using a conventional TF. The results are estimated on the basis of the TF value and a correlation coefficient

$$\mathrm{CC} = \frac{\left\langle |F_o||F_c| - \langle |F_o|\rangle\langle |F_c|\rangle \right\rangle}{\left(\left\langle |F_o|^2 - \langle |F_o|\rangle^2 \right\rangle \left\langle |F_c|^2 - \langle |F_c|\rangle^2 \right\rangle\right)^{1/2}}.$$

The dyad search can be extended to the triad search by including the third monomer in the search model. The phased TF will then be used to find three vectors describing the triad. On the other hand, the dyad search seems to be sufficient in most cases, as the main problem is usually the location of the first pair of monomers. When this task is fulfilled, the search can be repeated for the third monomer or the second dyad with the first being fixed. This approach proved to be efficient for solving structures with multiple copies in the asymmetric part (see tests).

3. Applications

The dyad search has been incorporated in the program *MOLREP* (Vagin & Teplyakov, 1997), a fully automated program for molecular replacement. It utilizes the RF of Crowther (1972) and an original full-symmetry TF combined with a packing function (Vagin, 1989).

Although there is no limitation on the relative position and orientation of the two monomers constituting a dyad, the search space can be limited by imposing restraints on the oligomeric structure, *e.g.* by defining the pure rotational symmetry. This feature may be particularly useful in the case of molecular dimers and higher oligomers.

Prior knowledge of the molecular symmetry may not only reduce the computational time, but may also facilitate the search by selecting the functionally meaningful solutions. For this purpose, the self-RF can be calculated by the program in the course of the search or can be introduced as an external source of information.

It is important to consider all symmetry-related dyads, *i.e.* those generated by the application of the crystallographic symmetry to the monomers composing the dyad. Owing to the overlapping of inter- and intramolecular vectors in the Patterson function, the TF may be more sensitive to some dyads and less to the others.

The multi-copy search method was tested on a number of cases with experimental X-ray data and gave satisfactory results. Three of them are described here. The first is the case of four molecules in the pseudo-centered unit cell. The second test presents a high-symmetry densely packed crystal that poses a task of selecting a unique solution from a number of alternatives. In the third example, problems arose because of a poor search model. The tests showed the power of the method in solving difficult cases and helped to define optimal parameters for the search.

3.1. Test 1

The structure of the Fab fragment CC49 (Navaza *et al.*, 2000) presents a complicated case of non-crystallographic symmetry that could not be solved by conventional molecular replacement. Crystals are monoclinic, space group $P2_1$, with unit-cell parameters $a = 115.6$, $b = 116.4$, $c = 70.3$ Å, $\beta = 97.8^\circ$.

There are four molecules in the asymmetric part, arranged in two pairs which are related by a vector in the *ac* plane (0.21, 0, 0.49). The structure of Hy10 was used as a search model. The RF shows two clear peaks, 8.6σ and 4.2σ (the next highest peak is 3.8σ), which correspond to molecules *A* and *B* of each pair. However, when using a single-molecule model, either *A* or *B*, the positional search fails. In contrast, the dyad search finds the correct solution unambiguously. The four crystallographically independent molecules constitute five different dyads: *A*1–*A*2, *A*1–*B*1, *A*1–*B*2, *B*1–*B*2 and *B*1–*A*2 (*A*2–*B*2 is equal to *A*1–*B*1). The crystallographic twofold axis doubles the total number of dyads in the unit cell. The search with the phased TF finds dyads of type *A*–*A* and *B*–*B*, *i.e.* those formed by equally oriented molecules, as top solutions. Their positioning with the TF is characterized by the highest peaks and CCs of 0.354 for the dyad *A*1–*A*2 and 0.295 for the dyad *B*1–*B*2. Subsequent search with the first dyad (*A*1–*A*2) being fixed results in the top solution for the second dyad (*B*1−*B*2) with a CC of 0.426. Calculations were performed at 4 and 5 Å resolution and gave similar results.

3.2. Test 2

The human calcium-binding protein S100B was crystallized in the rhombohedral space group *R*3, with unit-cell parameters $a = b = 99.7$, $c = 64.3$ Å and two molecules in the asymmetric part (Moroz, 2000). The search model was its bovine homologue (PDB code 1mho; Kilby *et al.*, 1996). The rotational parameters of the two molecules were obtained from the top solutions of the RF (4.9σ and 4.5σ, compared with the next peak of 4.0σ). The dyad search resulted in the correct solution characterized by a high contrast as reflected by the CC: 0.278 for the correct dyad *versus* 0.131 for the best wrong dyad. The solution appeared as the highest peak of the TF. However, it was found for the dyads ranked sixth to eighth on the basis of the phased TF calculations. The first five dyads appeared to be incorrect. This test illustrates the necessity for including a relatively large number of possible solutions in the calculations.

3.3. Test 3

This test demonstrates the power of the method in the case of a NMR model used for molecular replacement, which is often considered problematic. The crystals of SpoIIAA from *Bacillus sphaericus* belong to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters $a = 36.8$, $b = 53.5$, $c = 101.3$ Å and two molecules in the asymmetric part. The search model was one of the 24 NMR models of SpoIIAA from *B. subtilis* (PDB code 1auz; Kovacs *et al.*, 1998). The RF gave orientations of the molecules as the second and third highest peaks, both 2.8σ. The top peak, 3.2σ, was false and led to a number of false solutions with a CC as high as 0.442 (at 3 Å resolution). The correct solution nevertheless had the highest CC (0.465) and evolved as the second best dyad and the second highest peak of the TF. Two other symmetry-related dyads ranked fifth and eighth in the phased TF produced the next best solutions in terms of the CC (both 0.455). This structure could not be solved by a conventional approach and was finally determined by the MAD technique (Lewis, 2000). Obviously, the multi-copy search broadens the horizons of molecular-replacement applications.

The program *MOLREP* is written in standard Fortran77 and can be run under UNIX, Linux and Windows. It is available free as part of the packages *BLANC* (Vagin *et al.*, 1998) and *CCP*4 (Collaborative Computational Project, Number 4, 1994) or as a stand-alone version *via* the anonymous ftp account ftp.yorvic.york.ac.uk. Inquiries about the program should be addressed to AV at alexei@ysbl.york.ac.uk.

We thank Jorge Navaza, Olga Moroz, Richard Lewis and Misha Isupov for providing experimental data for testing the program. AV is supported by the Collaborative Computational Project, Number 4.

References

Bentley, G. A. (1997). *Methods Enzymol.* **276**, 611–619.
Carter, C. W. & Sweet, R. M. (1997). *Methods Enzymol.* **276**, 558–619.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Crowther, R. A. (1972). *The Molecular Replacement Method*, edited by M.G. Rossmann, pp. 173–178. New York: Gordon & Breach.
Kilby, P. M., Van Eldik, L. J. & Roberts, G. C. (1996). *Structure*, **4**, 1041–1052.
Kovacs, H., Comfort, D., Lord, M., Campbell, I. D. & Yudkin, M. D. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 5067–5071.
Lewis, R. J. (2000). Personal communication.
Moroz, O. (2000). Personal communication.
Navaza, J., Abergel, C. & Padlan, E. (2000). Personal communication.
Rossmann, M. G. (1972). Editor. *The Molecular Replacement Method.* New York: Gordon & Breach.
Tong, L. & Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 783–792.
Turkenburg, J. P. & Dodson, E. J. (1996). *Curr. Opin. Struct. Biol.* **6**, 604–610.
Vagin, A. A. (1989). *CCP4 Newslett. Protein Crystallogr.* **29**, 117–121.
Vagin, A. A., Murshudov, G. N. & Strokopytov, B. V. (1998). *J. Appl. Cryst.* **31**, 98–102.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.