SHORT COMMUNICATIONS

Acta Cryst. (1996). A52, 782–784

Combined molecular replacement

LIANG TONG *at Boehringer Ingelheim Pharmaceuticals, Inc., 900 Ridgebury Road, PO Box 368, Ridgefield, CT 06877, USA. E-mail: b4w@cc.purdue.edu*

(Received 14 February 1996; accepted 29 April 1996)

Abstract

Current molecular-replacement methods assume that the correct rotations to be applied to a search atomic model are close to peaks in the rotation function. In addition, generally only the top peaks in the rotation function are examined by the translation function. For difficult structures and for high-symmetry space groups, this assumption may no longer hold true. The more powerful approach is to examine not only the peaks but also other angles that have reasonable values in the rotation function to look for the molecular-replacement solutions. The combined molecular-replacement protocol is therefore a limited six-dimensional search, where the sampling of the rotational degrees of freedom is restricted by the rotation function. A packing check is used to eliminate solutions that cause steric clashes of the search atomic model in the crystal. Several new structures have been determined with this protocol.

1. Introduction

A structure determination by the molecular-replacement method involves finding the orientation and position of a search atomic model in the crystal unit cell (Rossmann, 1990; Lawrence, 1991). This is generally a six-dimensional problem and is extraordinarily expensive in CPU time if the six degrees of freedom are tackled at the same time. Traditionally, the problem is divided into two parts and solved in two steps – a rotation function that determines the orientation of the search model, followed by a translation function that determines its position (Rossmann, 1972). Generally, only the top few peaks in the rotation function are followed up with translation-function calculations, as manual intervention is required to pass the information from the rotation function to the translation function. Moreover, the original translation function based on Patterson correlation (Crowther & Blow, 1967) can only determine the position in two-dimensional projections, requiring further manual intervention to combine the results from different projections. The translation function based on the R factor or correlation coefficient, though able to provide three-dimensional information, requires much more CPU time and is impractical for more than a few rotation angles. As a consequence, considerable emphasis is placed on the rotation function and many methods have been developed to improve the results from it (Brünger, 1990; Yeates & Rini, 1990; Tong, 1996).

An improved formulation of the translation function based on Patterson correlation makes it possible to determine the position of the search model in one step (Harada, Lifchitz, Berthou & Jolles, 1981; Tong, 1993). The amount of CPU time required to evaluate this function by the fast-Fourier-transform (FFT) technique (Ten Eyck, 1973), is relatively small. The program *AMoRe* (Navaza, 1994), which became available recently, can automatically pass the peaks in the rotation function to the Patterson-correlation translation function. The top peaks in this translation function are kept as possible solutions to the molecular-replacement problem. An R factor and a correlation coefficient are calculated for each of the peaks.

The assumption behind the current molecular-replacement methods is that the correct rotation angle is either among or near the top peaks in the rotation function. For more difficult structures and for space groups with high symmetry, this assumption may not hold true. However, it can probably still be assumed that the correct angle will produce a relatively high rotation-function value, even though it may not be close to any of the peaks and its value may be lower than that of many noise peaks in the rotation function. Translation functions can then be used to sift through the angles that have reasonable heights in the rotation function to look for the correct molecular-replacement solutions. This, together with a packing check to eliminate impossible solutions (see below), represents a new protocol for applying the molecular-replacement method. With this protocol, some of the emphasis that was placed on the rotation function is shifted to the translation function. In essence, this protocol represents a limited six-dimensional search, hence the name combined molecular replacement, with the sampling of the rotational degrees of freedom restricted by the rotation function. It can be expected that such a protocol will be computationally more intensive than traditional molecular-replacement techniques. However, the power of modern computers certainly makes it a feasible approach. Moreover, this protocol could prove especially useful for difficult structures where traditional approaches have failed.

Examination of the packing of the search model in the crystal is an important part of a structure determination by the molecular-replacement method. A set of orientation and position parameters that cause steric clashes of the search model in the unit cell cannot be a solution but it often produces high correlation-coefficient values. This will increase the noise level in the molecular-replacement calculation and may make it impossible to discern the correct solution. A packing check is therefore an integral component of this combined molecular-replacement protocol. The packing check can be performed by counting the number of cases where distances between Cα atoms (or P, C4′, N1 and C4 or N9 atoms for nucleic acids) of different molecules in the crystal fall below a cut-off value. Only those orientation and position parameters that have small numbers of crystal-packing clashes should be saved as potential solutions.

2. Implementation

This protocol has been implemented in the *TF* program of the *REPLACE* suite (Tong, 1993). The calculation proceeds in the following steps.

(i) Read in the cross-rotation-function map calculated with the program *GLRF* (Tong & Rossmann, 1990, 1996) and select those grid points that have rotation-function values greater than a cut-off. Alternatively, one can select only the peaks in this rotation function, in a mode similar to that available in *AMoRe* (Navaza, 1994), although *AMoRe* does not perform packing checks for translation solutions. These grid points (or peaks) are then sorted in descending order according to their rotation-function values.

(ii) Calculate a Patterson-correlation translation function for each set of the rotation angles. The top peaks in this translation function are identified and those that have packing problems are eliminated. To place different solutions on the same scale, the correlation coefficient and R factor are calculated for each of the remaining peaks. A smaller cut-off value, in the range 0.5–1, is used to select the large terms. The program will periodically output the current top solutions, so that potential solutions can be identified even before the computation is completed. (This is the reason behind sorting the rotation-function grid points.)

(iii) When all the angles have been used in the translation function, the solutions are sorted on the correlation coefficient (or the R factor), merged and then output.

Several new structures have been determined using this protocol. These include a new crystal form of the p56lck SH2 domain (Tong, Warren & Jakes, unpublished data), human tumor necrosis factor in a cubic crystal form (Tong, Pandit, Jancarik & Kim, in preparation), both of which have resisted several attempts with traditional molecular-replacement techniques, and the entire extracellular domain of human CD4 (Wu, Kwong & Hendrickson, in preparation). The SH2 domain case will be described in more detail here. The crystals diffract to 1.9 Å resolution and belong to space group $P4_12_12$ or $P4_32_12$ with $a = b = 58.8$ and $c = 110.5$ Å. Two molecules are expected in the asymmetric unit. The search model was based on the SH2 domain structure at 1.0 Å resolution (Tong, Warren, King, Betageri, Rose & Jakes, 1996). A flexible loop on the surface of the protein was deleted. The highest peak in the cross rotation function, using reflection data between 15.0 and 3.0 Å resolution, was at 4.7σ and the next peak was at 3.5σ and with 74% of the height of the first peak. A clear translation-function solution, using reflection data between 6.0 and 3.0 Å resolution, was obtained with the first rotation-function peak in the space group $P4_32_12$, having a correlation coefficient of 0.345. However, the rotation function did not show any clear indication for the orientation of the second molecule. The Patterson-correlation orientational refinement of *X-PLOR* (Brünger, 1992) did not provide any additional information either.

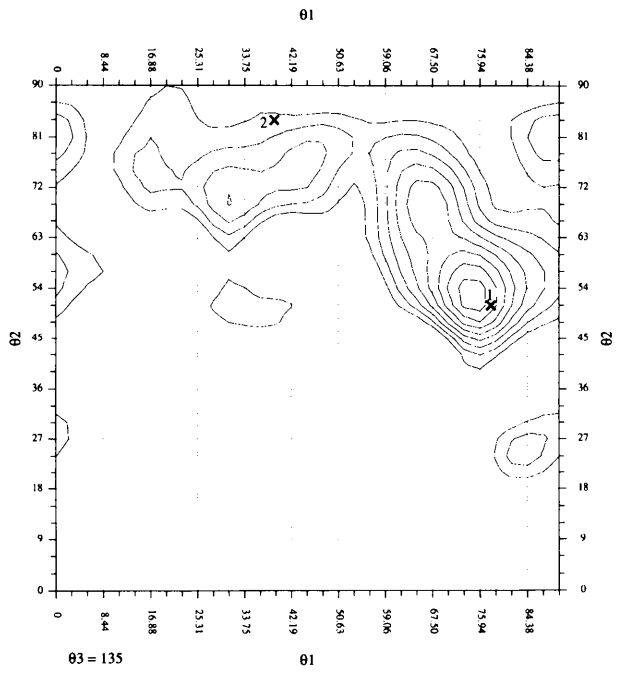

Fig. 1. A section of the cross rotation function for the p56lck SH2 crystal. Reflection data between 15.0 and 3.0 Å resolution were used in the calculation with the fast rotation function. The highest peak in this function was scaled to a height of 1000. The starting contour level is at 250 and the increment is 100. The orientations of the two molecules in the crystal are marked with crosses and labeled. The rotation-function value for the orientation of the second molecule is only about 30% of the highest peak.

The combined-molecular-replacement protocol was then applied to this problem. First, as a test, all the peaks in the rotation function with heights greater than 10% of the top peak were used for translation searches. The first molecule was included in the calculation. The distance cut-off for the packing check was 3 Å and the maximum allowed number of contacts was three. The top ten peaks in each Patterson-correlation translation function were identified and those without packing problems were saved as possible solutions. There was a total of 237 rotation function peaks, and the best solution from the translation calculations had a correlation-coefficient value of 0.310, which was lower than that based on the first molecule alone. This calculation took 3.6 h CPU on a SGI Indigo R4000. Then rotation-function grid points with heights greater than 25% of the top peak were used in the translation functions. The calculation was divided into three separate jobs, one using all grid points with heights greater than 40% of the top peak (a total of 4280 grid points), one using grid points between 30 and 40% (5580 points), and one using grid points between 25 and 30% (4644 points). After 10% of the rotation angles were completed (taking 8.5 h CPU on a SGI Indigo R4000), a clear solution was observed from the second job, with a correlation coefficient of 0.378 and good packing in the unit cell. After 40% (26 h CPU) and 20% (13.5 h CPU) of the rotation-grid points, the same solution was found in the first and the third job as well, with correlation coefficients of 0.357 and 0.397, respectively. Subsequent refinement with *X-PLOR* (Brünger, 1992) showed that this was the correct solution. A section of the cross rotation function close to the orientation of the second molecule is shown in Fig. 1. The correct orientation has a rotation-function value that is only 30% that of the top peak and is more than 12° away from the nearest rotation-function peak, which was apparently too large for traditional molecular-replacement techniques. The packing check removed more than 90% of the possible solutions.

A similar approach has recently been reported for very low resolution structure determination (Urzhumtsev & Podjarny, 1995), though no packing check was performed in that procedure. Rabinovich & Shakked (1984) proposed a full six-dimensional search procedure with ellipsoidal packing check.

The calculation was limited to low resolution owing to computation requirements. The program *BRUTE* (Fujinaga & Read, 1987) can perform local modifications of rotation angles. However, the calculation could not be extended to a large number of rotation angles as the correlation coefficient is used.

3. Conclusions

The protocol presented here encompasses and extends the current methods for molecular-replacement calculations. The limit on this six-dimensional search is imposed mostly to reduce the amount of computing time. In special cases, a more complete six-dimensional search could be attempted. The CPU performance of the current implementation of this protocol is certainly not optimal. It should also be noted that these calculations are perfectly suited for parallel computers. Of course, the search model has to be reasonably close to the crystal structure for any of these searches to produce meaningful results.

I thank Hao Wu for many useful discussions, Thomas Warren for preparing the crystal of the p56lck SH2 domain, and Paul Swepston for helpful suggestions.

References

Brünger, A. T. (1990). *Acta Cryst.* A**46**, 46–57.
Brünger, A. T. (1992). The *X-PLOR* manual, version 3.0. Yale University, New Haven, CT, USA.
Crowther, R. A. & Blow, D. M. (1967). *Acta Cryst.* **23**, 544–548.
Fujinaga, M. & Read, R. J. (1987). *J. Appl. Cryst.* **20**, 517–521.
Harada, Y., Lifchitz, A., Berthou, J. & Jolles, P. (1981). *Acta Cryst.* A**37**, 398–406.
Lawrence, M. C. (1991). *Q. Rev. Biophys.* **24**, 399–424.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Rabinovich, D. & Shakked, Z. (1984). *Acta Cryst.* A**40**, 195–200.
Rossmann, M. G. (1972). *The Molecular Replacement Method.* New York: Gordon and Breach.
Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
Ten Eyck, L. F. (1973). *Acta Cryst.* A**29**, 183–191.
Tong, L. (1993). *J. Appl. Cryst.* **26**, 748–751.
Tong, L. (1996). *Acta Cryst.* A**52**, 476–479.
Tong, L. & Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 783–792.
Tong, L. & Rossmann, M. G. (1996). *Methods in Enzymology*, edited by C. W. Carter Jr & R. W. Sweet. Orlando, FL: Academic Press. In the press.
Tong, L., Warren, T. C., King, J., Betageri, R., Rose, J. & Jakes, S. (1996). *J. Mol. Biol.* **256**, 601–610.
Urzhumtsev, A. & Podjarny, A. (1995). *Acta Cryst.* D**51**, 888–895.
Yeates, T. O. & Rini, J. M. (1990). *Acta Cryst.* A**46**, 352–359.