Acta Cryst. (1998). D**54**, 151-153

R_{omit} profile analysis for molecular replacements

DAISUKE TSUCHIYA and AKIO TAKENAKA* *at Department of Life Science, Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama 226, Japan. E-mail: atakenak@bio.titech.ac.jp*

(*Received 12 November 1996; accepted 3 June 1997*)

Abstract

A new procedure for evaluation of molecular replacement has been examined by using an R factor calculated from the probe structure with some omitted parts (R_{omit}). It has been demonstrated that changes in R_{omit} from the conventional R factor for the whole structure are sensitive to the local fitness in the omitted region even for large molecules such as proteins. Their profile, plotted against residues, is effective for distinguishing the most probable one from several solutions. In addition, this profile analysis exhibits useful information for model building.

1. Introduction

In recent years the molecular-replacement procedure has been established as a useful method for solving crystal structures of macromolecules. An advantage of this method is that the initial phases can be estimated without further experiments if an appropriate probe structure similar to that of the target molecule is available. Since the method was proposed, the major problem has been the required time-consuming computation, but this is being addressed by extensive efforts to improve the algorithms (Crowther, 1972; Crowther & Blow, 1967; Navaza, 1993) as well as recent developments in hardware and software (Fitzgerald, 1988; Brünger, 1992; Navaza, 1994). However, there is a problem in its practical application.

The structure which gives a unique solution in molecular replacement is easily solved but as the differences in tertiary structure between the probe and the target molecules become larger, ambiguities result. In such cases, several candidates for the solution require careful examination by R-factor calculations, electron-density maps, molecular packing and so on. Since these criteria cannot always distinguish the correct solution, the final judgment must await successful completion of structure refinement and the resultant reasonable structural parameters. Much time and effort are wasted during such processes. A more effective index or criterion is required to detect the most probable solution.

For this purpose, we have examined the usefulness of R_{omit} calculated from the probe structure with some omitted parts, to evaluate molecular-replacement solutions. This R_{omit} should be sensitive in principle to the local differences when the solution is totally correct. In addition, the change in R_{omit} as a function of residues also gives us useful information for model building. The procedure and its effectiveness are discussed below with some examples.

2. Definition and method

R_{omit} is defined as,

$$R_{omit}(r, n) = \sum ||F_{obs}| - |F_{calc}(r, n)|| / \sum |F_{obs}|,$$

where F_{obs} is an observed structure factor and $F_{calc}(r,n)$ is a structure factor calculated from a probe structure with omitted n residues around a residue r. The change in R_{omit} from R_{all} calculated from the non-omitted probe will reflect the differences between the probe and the target structures. When $R_{omit} > R_{all}$, the local structure around the residue r will be fitted. When R_{omit} is reduced ($R_{omit} < R_{all}$), there must be some differences around the residue r between the two structures. This criterion can be modified for expressing the local fitness (Lf),

$$\mathrm{Lf}(r, n) = R_{omit}(r, n) - R_{all}.$$

The local structure is the same or similar if Lf > 0 and different if Lf < 0.

The Lf values can be calculated for any omitted residues. It is better to normalize the effect of omissions by modifying the probe with alanines for all residues. A profile of Lf values as a function of changing r must show that contiguous regions with Lf > 0 have the same or a similar structure, and that those with Lf < 0 are different between the two structures. Therefore, this overview is useful in evaluating the overall fitness. In molecular-replacement solutions, the most probable one should have the highest fitness, which can be easily judged from its characteristic pattern with the long contiguous positive Lf's.

The R_{omit} profile analysis was examined by applying it to molecular-replacement solutions of 3-isopropylmalate dehydrogenase from *Bacillus coagulans* ($P3_121$ form; Tsuchiya *et al.*, 1996). This structure is now solved (Tsuchiya, Sekiguchi & Takenaka, submitted) using the structure of the enzyme from *Thermus thermophilus* (Imada *et al.*, 1991) as a probe, and refined to the final R factor of 0.180 using 10–3.0 Å resolution data. The asymmetric unit contains one dimeric enzyme of the two identical subunits, each of which has a molecular weight of 39 808 with 366 amino-acid residues. To evaluate this analysis, we investigated further how many differences could be detected by Lf between the initial probe and the final structure. The program *AMoRe* (Navaza, 1994) was used for the molecular replacement. Calculations of structure factors and R factors were performed with the program *X-PLOR* (Brünger, 1992).

3. Results and discussion

The probe molecule has a sequence identity of 51% with the target. Although their core structures can be well superimposed, large differences are found in the flexible loop regions. In particular, the three loops which have different lengths with gaps in amino-acid sequence between the two structures are markedly different in conformation (Tsuchiya, Sekiguchi & Takenaka, submitted).

Fig. 1 shows two profiles of Lf values for molecular-replacement solutions with the best R factor (0.453; correct solution) and the next best one (0.537; incorrect solution). The

R_{omit} values were calculated by omitting every three residues along the amino-acid sequence using 30–3.5 Å resolution data.† The Lf values for the incorrect solution are randomly distributed around zero, but for the correct one almost all are above zero. These features clearly indicate that the correct solution is easily distinguished by looking at the profile.

As shown in Fig. 1, the Lf < 0 regions in the correct solution exactly correspond to gaps (*g*) in the amino-acid sequence or loops (*l*) in the tertiary structure. In fact, these parts differ largely between the two structures. It shows that the R_{omit} profile also has an ability to anticipate the real differences in the local structure.

Next, we consider the performance (P_{Lf}) of Lf in detecting local differences in the probe structure. P_{Lf} is defined as a fraction of residues, the Lf values of which correctly anticipate the local similitude or difference between the probe and the refined structure,

$$P_{Lf} = \frac{N_{Lf}}{N_{all}}.$$

N_{all} is the total number of residues. N_{Lf} is the number of the residues of which the local structures with Lf > 0 are close to the refined one or those with Lf < 0 are different. When the final refined structure is compared with the initial structure which is molecular replaced in the unit cell, if the displacement of the Cα atom is beyond *D* Å, the region around the residue *r* is regarded as different from the refined structure. Fig. 2 shows P_{Lf} with different *D* (cutoff distance) and *n* (number of omitted residues). When $n \geq 3$ and $D \geq 2$ (Å), P_{Lf} reaches a plateau, suggesting that the Lf value with omission of more than three residues can detect differences in the probe at least by 2 Å at 90% probability. However, it should be noted that all R_{omit} values with large *n* are always greater than R_{all}, because a decrease in the *R* factor by omitting the different parts becomes smaller than the increase by omitting the similar parts. Therefore, the Lf value with large *n* makes it difficult to detect the local differences.

The R_{omit} profile analysis has been applied to other cases. The crystal structure of yeast lipoamide dehydrogenase (Toyoda, Sekiguchi & Takenaka, 1997) was solved by molecular replacement using a polyalanine model constructed from the tertiary structure of human glutathione reductase (Karplus & Shultz, 1987). The solution was correctly ascertained by the R_{omit} profile even though the sequence identity (28%) between the target and the probe is conspicuously lower. Another crystal form ($P2_12_12_1$) of 3-isopropylmalate dehydrogenase (Tsuchiya *et al.*, 1996) contains four dimeric enzymes in its asymmetric unit. Although it is generally difficult to solve such a structure with molecular replacement, the R_{omit} profiles made it easy to select the correct solution from several candidates (Tsuchiya, Sekiguchi & Takenaka, in preparation).

For practical application of this profile analysis, care should be taken with regard to the two following points. One is that the value of *n* may depend on the size of the molecules to be determined. For the present case, Fig. 2 suggests a suitable range of *n*. The second point to be noted is that all parts with Lf > 0 do not always represent local fitness, because such parts appear even in incorrect solutions as random distribution of Lf. Therefore, those parts should not be considered until the solution is judged to be correct.

The present profile analysis would be a reliable criterion for judging the most probable solution in molecular replacement. By incorporation of this analysis, it may be possible to realize a fully automated program for molecular replacement.

This work was supported in part by Grant-in-Aid for Scientific Research on Priority Areas (No. 08214203) from the Ministry of Education, Science, Sport and Culture of Japan, and by the Sakabe project of TARA (Tsukuba Advanced Research Alliance), University of Tsukuba, Japan.

† An apparent scale of $|F_{obs}|$'s was adjusted to $|F_{calc}(r,n)|$'s in each omission.

Fig. 1. R_{omit} profiles for solutions with the best *R* factor (upper, correct solution) and the next one (lower, incorrect solution). Only a profile calculated from one subunit is shown because the other subunit has a similar profile. Regions with Lf < 0 correspond to the parts containing gaps in the amino-acid sequence (*g*) and those containing loops with large differences from the refined structure (*l*). Data were taken from a trigonal crystal of 3-isopropylmalate dehydrogenase from *Bacillus coagulans*.

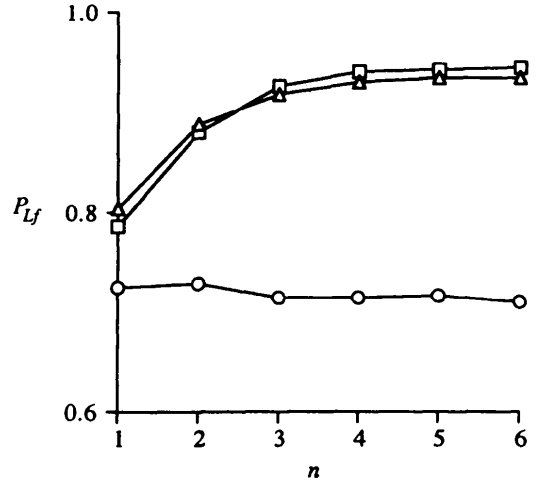

Fig. 2. P_{Lf} values calculated from different numbers (*n*) of omitted residues and different distance thresholds [*D* = 1.0 (○), 2.0 (△) and 3.0 (□)]. Data were taken from a trigonal crystal of 3-isopropylmalate dehydrogenase from *Bacillus coagulans*.

References

Brünger, A. T. (1992). *X-PLOR. Version 3.1. A System for X-ray Crystallography and NMR*. Yale University, Connecticut, USA.

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.

Fitzgerald, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273–278.

Imada, K., Sato, M., Tanaka, N., Katsube, Y., Matsuura, Y. & Oshima, T. (1991). *J. Mol. Biol.* **222**, 725–738.

Karplus, P. A. & Shultz G. E. (1987). *J. Mol. Biol.* **195**, 701–729.

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

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

Toyoda, T., Sekiguchi, T. & Takenaka, A. (1997) *J. Biochem. (Tokyo)*, **121**, 1–4.

Tsuchiya, D., Matsumoto, O., Gorai, T., Sekiguchi, T., Nosoh, Y. & Takenaka, A. (1996). *Acta Cryst.* D**52**, 1030–1032.