

A Translation-Function Approach for Heavy-Atom Location in Macromolecular Crystallography

ALEXEI VAGIN^a AND ALEXEI TEPLYAKOV^{b*}

^aDepartment of Chemistry, University of York, Heslington, York YO1 5DD, England, and ^bEuropean Molecular Biology Laboratory, DESY, Notkestrasse 85, D-22603 Hamburg, Germany. E-mail: alex@embl-hamburg.de

(Received 27 July 1997; accepted 30 October 1997)

Abstract

A method for locating heavy atoms in the unit cell of macromolecular crystals by a full-symmetry translation function is described. The approach has been implemented in the program *TRAHALO* and tested on experimental isomorphous and anomalous data.

1. Introduction

The multiple isomorphous replacement method is the principal approach for solving the phase problem in macromolecular crystallography. It has been developed significantly during the last few years. Most attention has been attracted to the new methods of incorporation of heavy atoms into the protein, X-ray data collection and processing, calculation of best phases and their improvement by modification of electron density (Carter & Sweet, 1997). In contrast, location of heavy atoms in the unit cell remains underdeveloped and poses significant problems in cases of numerous binding sites or high-symmetry space groups. The existing methods of heavy-atom location are based either on the real-space interpretation of difference Patterson maps with the help of the programs *RSPS* (Knight, 1989) and *GROPAT* (Jones & Stuart, 1991), or on the Patterson superposition or direct methods implemented in *SHELXS96* (Sheldrick, 1997). None of the approaches are free of certain limitations. Direct methods face the difficulty in identifying small E values as the ΔF 's represent lower bounds of their true values. Methods operating in real space have to deal with the large origin peak and ripples around it resulting from the series termination. Additional problems are due to peak overlaps (the number of peaks is a square function of the number of atoms). Those approaches which use the peak-lists derived directly from the Patterson function are sensitive to the absence of peaks at expected positions. Although these approaches proved to be useful in solving many crystal structures, there is a strong need for improvement in this field.

We present here a new method of heavy-atom location using a translation function (TF) approach. It was tested on a number of cases using experimental

isomorphous and anomalous X-ray data and has been successfully applied for solving new crystal structures.

2. Description of the method

The task of location of heavy atoms in the unit cell may be considered as a special case of molecular replacement when a trial model consists of only one atom. Because of a spherical symmetry of the model, only a translational search is required which may be best performed by using a reciprocal-space fast TF.

This approach has been implemented in the program *TRAHALO* (translation function heavy atom location) which uses a full-symmetry TF described in detail elsewhere (Vagin, 1989; Vagin & Teplyakov, 1997). The TF originates from the T2 function of Crowther & Blow (1967) corrected by Harada *et al.* (1981) which has an advantage of the simultaneous use of all symmetry operators. This improves the signal-to-noise ratio and gives directly a position of a model in the unit cell rather than a position with respect to a symmetry element. The TF may be calculated either as a product or as a sum of T_{jk} functions (for a pair of symmetry operators j and k), and may be coupled with a packing function Q ,

$$T(s) = \prod_{j \neq k} T_{jk}(s)Q(s) \\ = \prod_{j \neq k} \int P^o(r)P^c_j(s, r)dr \sum_{j \neq k} \int \rho_j(r, s)\rho_k(r, s)dr,$$

where P^o and P^c are observed and calculated Patterson functions, $\rho_j(r, s)$ is the electron density of a search model calculated for the symmetry operator j . In addition, the TF was modified to allow incorporation of the second model already placed in the cell (Vagin, 1989). In a heavy-atom search, this feature of a TF makes it possible to include all atoms whose positions are already known in the search of additional binding sites.

The procedure of heavy-atom location consists of several steps.

(1) Scaling of the native and derivative structure factors (in the case of isomorphous replacement) by scaling their Patterson origin peaks (Rogers, 1965). This

Table 1. *Heavy atoms in hevamine derivatives (test 1) and in the stromelysin1/TIMP1 complex (test 2)*

Atom†	Difference Fourier‡	TF§	PP¶
2,4-Dichloromercuri-6-nitrophenol			
Hg1†	40.6	0.82 (2)	1.44 (1)
Hg2	10.5	—	—
Trimethyllead acetate			
Pb1†	28.9	1.00 (1)	1.29 (1)
4-Diacetamido-phenyl-mercuric acetate			
Hg1†	15.9	0.72 (25)	1.09 (2)
Hg2†	15.2	0.74 (18)	1.07 (5)
Hg3	14.4	0.83 (9)	1.03 (26)
Hg4	13.0	—	—
AgNO ₃			
Ag1†	19.0	0.62 (8)	1.14 (1)
Ag2†	18.8	0.59 (11)	1.11 (2)
Stromelysin1/TIMP1			
Sc1	15.7	0.33 (10)	1.00 (2)
Sc2	14.3	0.24 (11)	0.96 (10)
Sc3†	12.3	0.44 (8)	0.97 (5)
Sc4	11.4	0.37 (9)	0.97 (7)
Zn1†	13.0	0.60 (3)	1.00 (1)
Zn2	12.6	0.21 (13)	0.95 (11)
Zn3	11.9	0.47 (7)	0.97 (4)
Zn4	10.9	0.16 (14)	0.95 (12)

† Atoms selected in the main pass (step 6) of the program. ‡ Peak height (in standard deviations from mean density) in the difference Fourier map ($F_{\text{der}} - F_{\text{nat}}$) with phases from the refined model. § The primary TF value relative to the top one which is set to 1.0 and the position of the atom in the list sorted on the TF value (in brackets). ¶ PP and the position of the atom in the primary TF list sorted on the PP (in brackets).

method is advantageous when only low-resolution data is available (which is often true for heavy-atom derivatives) and the estimation of an overall B factor from the Wilson plot may be inaccurate.

(2) Application of the low-resolution cut-off in a form of a Gaussian,

$$F = F_{\text{obs}} \{1 - \exp[-B_{\text{off}}(\sin \theta/\lambda)^2]\}$$

where $B_{\text{off}} = (2d_{\text{max}})^2$ to reduce the weight of low-resolution terms which may be corrupted by the bulk solvent contribution. In our experience, it is better than to exclude those terms completely (Vagin & Teplyakov, 1997).

(3) The primary heavy-atom search using the TF with a one-atom probe model. This gives the so-called 'primary TF' list of atomic positions sorted on the TF value which roughly corresponds to the height of the self-Patterson peaks.

(4) Estimation of the phasing power (PP is defined as the ratio of the r.m.s. heavy-atom structure factor to the r.m.s. lack of closure) for each of the putative heavy atoms. The primary TF list is re-sorted on the PP. Tests show that in most cases the PP is a better selection criterion than the TF value.

(5) Heavy-atom search with a one-atom probe model and one atom from the primary TF list in the fixed

position. This step is repeated for N selected atoms to produce N 'secondary TF' lists. The atoms in these lists are sorted on the PP which is now calculated for a pair of atoms. An important step at this stage is the detection of atoms which give a high PP but which are not present in the primary TF list. Such atoms are also considered as possible candidates for the solution.

(6) Selection of the best single atom and a pair of atoms from the primary and secondary TF lists on the basis of the highest PP. Note that the secondary TF's have brought atoms to a common origin.

(7) Search for additional sites using all accepted atoms as a fixed model. This step is repeated while the addition of new atoms increases the PP and their occupancies are above a certain level.

The program *TRHALO* is completely automatic and requires only X-ray data (isomorphous or anomalous) and unit-cell information on input. The output is a set of heavy-atom positions which may be used for phase calculation.

3. Parameters

3.1. Radius of the probe atom

The trial model is represented by a Gaussian or uniform spherical distribution of electron density of an appropriate radius. Tests show that the algorithm is not sensitive to the type of the heavy atom. Using the probe radius of 2.5 Å, solutions were obtained for a range of heavy atoms from Mn to U.

3.2. Resolution limits

Most cases could be solved using the high-resolution limit of 4–5 Å. Using of higher resolution data might improve the accuracy of atomic positions but would significantly increase the computation time. In contrast the choice of the low-resolution limit appeared to be very important. Tests suggesting a B_{off} of 400 Å², which corresponds to 10 Å resolution, to effectively reduce the contribution of low-resolution terms.

3.3. Number of sites for the secondary TF search in step 5

This parameter determines the CPU time as each TF calculation takes about 10 min on SGI Challenge L for a medium-size unit cell and 4 Å resolution. In fact, there is no need to test a large number of sites as the major sites almost always appear in the top of the primary TF list. The exceptions may occur when numerous sites with low occupancies contribute equally to isomorphous/anomalous differences and the PP of any one of them is obviously not very high. For this reason, few sites with the highest TF values are also tested by the program.

4. Tests

The method of heavy-atom location was tested on a number of cases using experimental isomorphous and anomalous X-ray data. Two of them are described here.

4.1. Test 1

Isomorphous data for the crystals of heveamine, space group $P2_12_12_1$, $a = 52.3$, $b = 57.7$, $c = 82.1$ Å, one 30 kDa protein molecule in the asymmetric unit (Terwisscha van Scheltinga *et al.*, 1994). Binding sites for the four derivatives are listed in Table 1. Their relative occupancies can be estimated from the height of the peaks on difference Fourier maps ($F_{\text{der}} - F_{\text{nat}}$) with phases from the refined model. The heavy-atom search was performed at 4 Å resolution with $B_{\text{off}} = 400$ Å². All major binding sites were located in the main pass of the program, *i.e.* by inspecting the primary and secondary TF lists (these sites are marked in Table 1). Minor sites in Hg derivatives were located in the additional passes of the program (step 7).

The major sites of all derivatives were characterized by the highest PP in the primary TF lists. In contrast, the TF values for these sites almost never were the highest. This test clearly shows the advantage of using PP as a criterion for selection of putative sites.

The Ag and the four-site Hg (4-diacetamido-phenyl-mercuric acetate) derivatives could not be solved by the real-space Patterson search due to different reasons. In the Hg derivative, there are four sites with equally low occupancies which do not produce significant self-Patterson peaks. In the Ag derivative, the cross-vector between the two sites lies in the Harker section causing misinterpretation of the difference Patterson map. Both derivatives could be solved by the TF approach.

4.2. Test 2

Anomalous data for the crystal of the stromelysin1/TIMP1 complex, space group $P4_3$, $a = b = 80.6$, $c = 157.6$ Å (Gomis-Rüth *et al.*, 1997). The asymmetric unit contains two 40 kDa complex molecules with two Zn and two Se atoms in each. The data have been collected at a wavelength of 0.979 Å which corresponds to the K absorption edge of Se and which is shorter than that of Zn (1.284 Å) so that the anomalous signal from Zn is also significant. An additional complication of the case was the pseudo I -centering of the cell due to the vector relating the two crystallographically independent molecules. This vector caused the highest peak of the primary TF (0.25, 0.25, 0) which was however listed the third when sorted on the PP. All heavy atoms were among the top 12 sites in the primary TF list (Table 1). What is more important, the best six pairs of atoms in the secondary TF lists contained only the correct sites. The pair Zn1–Se3 with the highest PP (1.05) was selected as a solution and used as a fixed model for the

subsequent search. Addition of the third atom (Se2) increased the PP to 1.13. Finally, all eight atoms were located.

These and other tests demonstrated the power of the TF approach for heavy-atom location using either isomorphous or anomalous data. The obvious advantage of this method is the ability to find heavy atoms in a completely automatic way which makes it very attractive even for simple cases. The TF search is quite stable under a set of default parameters which need not to be optimized for a particular case. This saves a lot of time to the users and helps to avoid errors. The TF approach also proved to be very helpful in solving difficult cases like multi-site derivatives, high-symmetry space groups and pseudosymmetry.

5. Distribution

The program *TRAHALO* is written in standard FORTRAN 77 and could be run under UNIX, VMS and Windows. It is available free as part of the package *BLANC* or independently from the author by e-mail (alexey@yorvic.york.ac.uk).

We thank Drs Fred Antson, Wolfram Bode, Misha Isupov, Anke Terwisscha van Scheltinga and Eleanor Dodson for providing experimental data for testing the program, and Gleb Bourenkov for the data and fruitful discussions. The work was supported by the EU BIOTECH grant BIO4CT-96-0189 (AV).

References

- Carter, C. W. & Sweet, R. M. (1997). *Methods Enzymol.* **276**, 286–494.
- Crowther, R. A. & Blow, D. M. (1967). *Acta Cryst.* **23**, 544–548.
- Gomis-Rüth, F.-X., Maskos, K., Betz, M., Bergner, A., Huber, R., Suzuki, K., Yoshida, N., Nagase, H., Brew, K., Bourenkov, G. P., Bartunik, H. D. & Bode, W. (1997). *Nature (London)*, **389**, 77–81.
- Harada, Y., Lifchitz, A., Berthou, J. & Jolles, P. (1981). *Acta Cryst.* **A37**, 398–406.
- Jones, Y. & Stuart, D. (1991). In *Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 39–47. Warrington: Daresbury Laboratory.
- Knight, S. (1989). PhD thesis, Swedish University of Agricultural Sciences, Sweden.
- Rogers, D. (1965). In *Computing Methods in Crystallography*, edited by J. S. Rollett, pp. 133–148. Oxford University Press.
- Sheldrick, G. M. (1997). *Methods Enzymol.* **276**, 628–641.
- Terwisscha van Scheltinga, A. C., Kalk, K. H., Beintema, J. J. & Dijkstra, B. W. (1994). *Structure*, **2**, 1181–1189.
- Vagin, A. A. (1989). *CCP4 Newslett. Protein Crystallogr.* **29**, 117–121.
- Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.