Acta Cryst. (1996). A52, 476–479

The Locked Translation Function and Other Applications of a Patterson Correlation Function

LIANG TONG

Department of Inflammatory Diseases, Boehringer Ingelheim Pharmaceuticals, Inc., 900 Ridgebury Road, PO Box 368, Ridgefield, CT 06877, USA

(Received 10 September 1995; accepted 23 January 1996)

Abstract

Another form of the Patterson correlation function is presented. It is based on the overlap of the self vectors throughout the entire unit cell. The dependence of this overlap on the position of a monomer (of known orientation) relative to the center of a non-crystallographic assembly leads to the definition of a locked translation function. This overlap is also dependent on the orientation of a search atomic model and therefore can be applied to the refinement of rotational parameters. Other possible applications of this correlation function are also discussed.

1. Introduction

Patterson correlation lies at the foundation of many molecular-replacement techniques. Rotation functions are based on the correlation of one Patterson map with the rotated version of another within a spherical volume centered at the origin (Rossmann & Blow, 1962). The evaluation of this correlation can take various forms, giving rise to, for example, the slow (Rossmann & Blow, 1962) and fast (Crowther, 1972; Navaza, 1987) rotation functions. Patterson correlation can also be applied to the translation problem (Crowther & Blow, 1967; Harada, Lifchitz, Berthou & Jolles, 1981; Tong, 1993). The resulting translation function can be evaluated much faster than those based on the R factor or the correlation coefficient between structure-factor amplitudes.

When the asymmetric unit of a crystal contains a macromolecular assembly obeying a point-group symmetry, the orientations of the non-crystallographic symmetry (NCS) elements may be determined with an ordinary self-rotation function (Rossmann & Blow, 1962). A locked self-rotation function may be used as well, which can determine the orientations of all the NCS elements at the same time (Tong & Rossmann, 1990). If an atomic model for the monomer of the assembly is available, its orientation in the crystal unit cell can be determined by an ordinary cross-rotation function. Alternatively, a locked cross-rotation function can be used to search for all the molecules of the assembly at the same time (Tong & Rossmann, 1990, 1996). However, a translation function where the positions of all the molecules of the assembly are searched for at the same time (a 'locked translation function') has so far been unavailable.

Patterson correlation refinement (Brünger, 1990) and intensity-based domain refinement (Yeates & Rini, 1990) have been very useful in optimizing the rotational parameters of a search atomic model after a cross-rotation-function calculation. The Patterson correlation refinement (Brünger, 1990) is based on the maximization of the correlation coefficient between squared normalized observed and calculated structure factors. The intensity-based domain refinement (Yeates & Rini, 1990) is based on the minimization of the differences between squared observed and calculated structure factors by a least-squares procedure.

Another form of the Patterson correlation function is presented in this paper. This correlation is based on the overlap of the self vectors between the model and the crystal throughout the entire unit cell. It can be used to refine the rotational parameters of a search model as derived from an ordinary cross-rotation-function calculation. It can also be used to define a locked translation function, where the positions of all the molecules of the assembly (relative to the center of the assembly) can be determined at the same time.

2. Theoretical background

Given a molecule (x_j^0) of fixed orientation and at a reference position, the calculated structure factor after the molecule is translated by $\mathbf{x}_0$ and placed in the crystal unit cell is given by

$$\mathbf{F}_h^c = \sum_m \mathbf{F}_{h,m} \exp\{2\pi i h[T_m]\mathbf{x}_0\}, \qquad (1)$$

where the summation goes over the crystallographic symmetry operators and

$$\mathbf{F}_{h,m} = \sum_j f_j \exp\{2\pi i h([T_m]x_j^0 + t_m)\}$$
$$= \exp(2\pi i h t_m)\mathbf{F}_{h[T_m],1} \qquad (2)$$

is the contribution of the mth crystallographic asymmetric unit to the structure factor at the reference position. The correlation between the observed and calculated Patterson maps throughout the entire unit cell is given by (Tong, 1993)

$$PC(\mathbf{x}_0) = \sum_h \sum_m (F_h^o)^2 |\mathbf{F}_{h,m}|^2 + \sum_h \sum_m \sum_{n \neq m} (F_h^o)^2 \mathbf{F}_{h,m} \mathbf{F}_{h,n}^* \times \exp\{-2\pi i h([T_n] - [T_m])\mathbf{x}_0\}. \quad (3)$$

The first term in (3) represents the overlap of the self vectors between the search model and the crystal, whereas the second term represents the overlap of the cross vectors. Equation (3) has generally been used to determine the translation vector $\mathbf{x}_0$ given the search molecule with a pre-determined orientation. In such a case, the first term in (3) is a constant and is generally ignored.

3. Refinement of rotational parameters

The correct orientation and position of a search molecule in the crystal unit cell should lead to the maximal overlap between the observed and calculated Patterson maps, *i.e.* should lead to a maximum of (3). The first term in (3) in this case depends only on the orientation of the molecule. It may therefore be expected that the correct orientation of the search molecule should produce the maximal value for this term. From (2), it can be seen that the summation over the crystallographic symmetry operators can be ignored. This leads to the following Patterson correlation function:

$$P1PC([\rho]) = \sum_h (F_h^o)^2 |\mathbf{F}_{h,1}|^2, \quad (4)$$

where the summation goes over all reflections.

This Patterson correlation function can be used to refine the rotational parameters of a search molecule. The calculated structure-factor amplitude $|\mathbf{F}_{h,1}|$ is expressed as a function of the orientations and the positions of the rigid groups in the search molecule. As the calculation is carried out in space group $P1$, the position of the first group may be held fixed. A conjugate-gradient least-squares procedure is used to find the parameters that maximize this correlation.

4. The locked translation function

Given the atomic model X_j^0 (in Cartesian coordinates) for a monomer of the macromolecular assembly, the rotation $[F]$ that brings it into the same orientation as that of one of the monomers of the assembly in the standard orientation can be determined with the locked cross-rotation function (Tong & Rossmann, 1990, 1996). If $\mathbf{X}_0$ is the translation vector that brings the correctly oriented search model into the same position as the monomer in the standard orientation,

$$\mathbf{X}_j = [F]X_j^0 + \mathbf{X}_0. \quad (5)$$

Assume that $[I_n]$ $(n = 1, \ldots, N)$ is the set of rotation matrices for the NCS point group in the standard orientation and $[E]$ is the rotation matrix that brings the standard orientation to that of the assembly in the crystal. The set of fractional atomic coordinates for the assembly centered at the origin of the crystal unit cell is given by

$$x_{j,n} = [\alpha][E][I_n]([F]X_j^0 + \mathbf{X}_0) \quad (6)$$

(noting that the assembly in the standard orientation is centered at the origin). $[\alpha]$ is the de-orthogonalization matrix (Rossmann & Blow, 1962). Similar to the derivation of (4), the correct translation vector $\mathbf{X}_0$ should give rise to calculated structure factors that will maximize the first term in (3). The locked translation function is defined as

$$GLTF(\mathbf{X}_0) = \sum_h (F_h^o)^2 |\mathbf{F}_h^c|^2 = \sum_h \sum_n (F_h^o)^2 |\mathbf{f}_{h,n}|^2 + \sum_h \sum_n \sum_{m \neq n} (F_h^o)^2 \mathbf{f}_{h,n} \mathbf{f}_{h,m}^* \times \exp\{-2\pi i h([\theta_m] - [\theta_n])\mathbf{X}_0\}, \quad (7)$$

where

$$[\theta_n] = [\alpha][E][I_n] \quad (8)$$

and

$$\mathbf{f}_{h,n} = \sum_j f_j \exp\{2\pi i h[\theta_n][F]X_j^0\}. \quad (9)$$

The locked translation function therefore determines the position of a search molecule (of known orientation) relative to the center of the NCS point-group symmetry. Once this position is known, the structure of the entire assembly, centered at the origin of the unit cell, is given by (6). The center of this assembly in the crystal unit cell can then be determined with an ordinary translation function, using the structure of the entire assembly as the search model.

As can be expected, (7) bears much resemblance to (3), by the interchange of the crystallographic quantities $([T_m], \mathbf{F}_{h,m})$ with the NCS quantities $([\theta_n], \mathbf{f}_{h,n})$. While (3) can be evaluated directly by the fast-Fourier-transform (FFT) technique (Ten Eyck, 1973), (7) cannot as the elements of the $[\theta_n]$ matrices are generally

non-integral. The FFT technique can be applied indirectly, however, by carrying out one separate transform for each pair of n and m parameters in (7). This would require the calculation of $N(N-1)/2$ transforms of the type $\sum_h (F_h^o)^2 \mathbf{f}_{h,n} \mathbf{f}^*_{h,m} \exp(-2\pi i h x_{mn})$. The summation over n and m for each $\mathbf{X}_0$ can then be evaluated by interpolating among the values obtained from these transforms, noting that $x_{mn} = ([\theta_m] - [\theta_n])\mathbf{X}_0$. Alternatively, (7) can be evaluated by direct summation, which is much slower but may produce better results. Packing of the monomers in the assembly can be checked to remove those translation vectors that cause steric clashes of the monomers.

5. Test calculations

Both applications of this Patterson correlation function have been implemented as new options in the program *GLRF* (Tong & Rossmann, 1990, 1996). As both functions are expected to be dominated by strong reflections, the large-term approach can be applied to their calculations (Tollin & Rossmann, 1966). For the refinement of rotational parameters, the program first treats the entire search model as a rigid body and optimizes its orientation. Subsequently, the orientations and positions of the individual rigid-body groups in the search model, as defined by user input, are optimized.

For the locked translation function, a locked cross-rotation function is calculated first. The program can then carry out a locked-translation-function calculation for each of the top few peaks in the rotation function. The output of the calculation is a list of the peaks in this translation function. Translation vectors that cause steric clashes of the monomers in the assembly are ignored either during the calculation itself or during the peak-search process. This packing check is somewhat time consuming and is better avoided during the calculation by the FFT method. If direct summation is used, a packing check should be performed during the calculation. To minimize errors from the interpolation, the Fourier transforms are sampled with a grid size that is $\frac{1}{4}$ of the highest resolution. The program can also output the atomic coordinates for the entire assembly based on the locked-translation-function results. This model of the complete assembly can then be input to an ordinary translation-function program to locate the center of the assembly.

Test calculations with the locked translation function are presented in more detail here. The atomic models for two structures, deoxy-β_4-hemoglobin (entry 1CBL) (Borgstahl, Rogers & Arnone, 1994) and asparaginase (entry 3ECA) (Swain, Jaskolski, Housset, Rao & Wlodawer, 1993), were taken from the Brookhaven Protein Data Bank. Both crystals are monoclinic (space group $P2_1$) with a 222 tetramer in the asymmetric unit. Structure factors to 3 Å resolution were calculated for the crystal with the program *TF* (Tong, 1993). The orientations of the three non-crystallographic twofold axes of the tetramer were determined with the ordinary self-rotation function and confirmed with the locked self-rotation function. Reflection data between 10 and 3.5 Å resolution were used in all the test calculations described here. The large-term cut-off was 1.5, saving about 20% of the reflections. The first monomer in each atomic model was re-oriented, centered at the origin and placed in a large $P1$ cell. Structure factors to 3 Å resolution were calculated for the monomers in the $P1$ cells. The locked cross-rotation functions, calculated with the fast rotation function followed by interpolation (Tong & Rossmann, 1996), contained one significant peak for each model, corresponding to the correct orientation.

A locked cross-rotation search was then carried out with the slow rotation function, covering a small region around the peak and using 1° grid intervals. The top peak from this fine search was used for the locked translation function. The search region covered −30 to 30 Å in X, Y and Z coordinates, with a grid interval of 1 Å. The minimum and maximum lengths of the translation vector were 10 and 30 Å, respectively, removing 52% of the grid points in the box from the calculation. The locked translation function was evaluated in three different ways – by the FFT method without packing check during the calculation, by the FFT method with packing check and by direct summation with packing check (Table 1). The maximum number of C_α contacts (<3 Å) allowed by the packing check was specified as 5, which removed about 36 and 45% of the grid points from the calculation for the test cases 1CBL and 3ECA, respectively. Performing the packing check during the calculation with the FFT method lowered the background noise but it did not change the relative ranking of the peaks. The crystal for test case 1CBL has more compact packing, causing more mixing of the self and cross vectors. Consequently, the results for the 1CBL test case are inferior to those for 3ECA. The calculation by direct summation produced slightly better results for the 1CBL test case, although the computation time was significantly longer.

6. Discussion

This paper describes two applications of another form of the Patterson correlation function. Other applications of this function are possible as well. For example, (4) can be used as the basis of a rotation function. However, a fast way of evaluating this equation for many different sets of rotation angles needs to be developed to make this rotation function practical.

The procedure presented here for the refinement of rotational parameters represents an alternative to those proposed earlier (Brünger, 1990; Yeates & Rini, 1990). The methods are expected to be generally equivalent, although there may be applications where one method

Table 1. *Summary of locked-translation-function test calculations*

Structure	1CBL			3ECA		
No. of C_α atoms per monomer	146			326		
Test[a]	*A*	*B*	*C*	*A*	*B*	*C*
Peak no. 1 height[b]	1000*	1000*	1000*	1000*	1000*	1000*
Peak no. 1 signal-to-noise[c]	5.60	7.38	7.72	6.51	9.92	10.03
Peak no. 2 height	463	463	496	508	508	477
Peak no. 2 signal-to-noise	2.52	3.41	3.82	3.28	5.04	4.78
CPU time (min)[d]	8	89	277	12	73	199

(*a*) The three tests are: *A* calculation by FFT without packing check; *B* calculation by FFT with packing check; *C* calculation by direct summation with packing check. (*b*) The highest value of each locked translation function was scaled to 1000. When a packing check is not performed during the calculation, the position with the highest value may have packing problems and may be rejected by the packing check in the peak-search process. (*c*) The signal-to-noise is defined as (peak height − map average)/standard deviation. (*d*) All calculations were performed on a SGI Indigo R4000. * The correct solution.

proves to be more advantageous than others. It can be seen that (4) is equivalent to the cross term in the least-squares error function of Yeates & Rini (1990). The contribution of crystallographic symmetry is included implicitly in the function proposed here as the summation goes over all reflections, whereas that in the function of Yeates & Rini (1990) goes over the unique ones. The intensity-based domain refinement (Yeates & Rini, 1990) also includes a term that radially dampens the Patterson maps. This will down-weight the noise due to the cross vectors and was shown to produce somewhat better results (Yeates & Rini, 1990). However, the contribution of the cross vectors to (4) is expected to be small as the calculated Patterson map is based on a single molecule in the crystal unit cell and hence lacks any contributions from the cross vectors.

Equation (4) does not consider the presence of NCS in the crystal, although it is relatively straightforward to modify it to utilize the NCS. The first term in (7), which is ignored during the calculation of the locked translation function, actually gives the relevant expression for the refinement of rotational parameters in the presence of NCS.

Similar to the Patterson correlation refinement (Brünger, 1990), the locked translation function as presented here can be re-formulated to be based on the correlation coefficient. The test calculations (Table 1) show that such a correlation coefficient needs to be evaluated by the FFT method to be practical.

The method for the determination of the position of an assembly based on an atomic model for the monomer is similar to that described for the interpretation of a Patterson map in the presence of NCS (Tong & Rossmann, 1993). The self-vector search in the Patterson interpretation is equivalent to the locked translation function. This can be expected as both are based on the overlap of the self vectors.

I thank Lynn Ten Eyck and Duncan McRee for making available routines for FFT calculation.

References

Borgstahl, G. E. O., Rogers, P. H. & Arnone, A. (1994). *J. Mol. Biol.* **236**, 831–843.
Brünger, A. T. (1990). *Acta Cryst.* A**46**, 46–57.
Crowther, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. Rossmann, pp. 174–178. New York: Gordon and Breach.
Crowther, R. A. & Blow, D. M. (1967). *Acta Cryst.* **23**, 544–548.
Harada, Y., Lifchitz, A., Berthou, J. & Jolles, P. (1981). *Acta Cryst.* A**37**, 398–406.
Navaza, J. (1987). *Acta Cryst* A**43**, 645–653.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Swain, A. L., Jaskolski, M., Housset, D., Rao, J. K. M. & Wlodawer, A. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 1474–1478.
Ten Eyck, L. F. (1973). *Acta Cryst.* A**29**, 183–191.
Tollin, P. & Rossmann, M. G. (1966). *Acta Cryst.* **21**, 872–876.
Tong, L. (1993). *J. Appl. Cryst.* **26**, 748–751.
Tong, L. & Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 783–792
Tong, L. & Rossmann, M. G. (1993). *J. Appl. Cryst.* **26**, 15–21.
Tong, L. & Rossmann, M. G. (1996). *Methods in Enzymology*, edited by C. W. Carter Jr & R. W. Sweet. In the press. Florida: Academic Press.
Yeates, T. O. & Rini, J. M. (1990). *Acta Cryst.* A**46**, 352–359.