13.3. Translation functions

By L. Tong

13.3.1. Introduction

A structure determination by the molecular-replacement method traditionally proceeds in two steps (Rossmann, 1972, 1990). The first step involves the determination of the orientation of the search model in the unknown crystal unit cell by the rotation functions (see Chapter 13.2). Once the orientation of the search model is known, translation functions are employed in the second step to determine the location of the model in the crystal unit cell. This essentially reduces a six-dimensional problem (three rotational and three translational degrees of freedom) to two three-dimensional problems, which are computationally more manageable. With the speed of modern computers, a strict division between the rotational and the translational components of a molecular-replacement structure solution may no longer be necessary (see Section 13.3.7).

Translation functions are normally formulated to achieve minimum or maximum values when the search molecule is at its correct position in the crystal unit cell. As with the rotation problem, the translation problem is solved as a search. The positional parameters of the model are varied in the unit cell, generally on a grid. Translation functions are evaluated at these search grid points in order to identify those that minimize or maximize the functions.

Most translation functions involve a comparison between the observed structure-factor amplitudes (or squared amplitudes) and those calculated based on the search model. The R factor and the correlation coefficient can be used as indicators for translation searches (Section 13.3.2). The correlation between the observed Patterson map and that which is calculated based on the search model is the foundation of another translation function (Section 13.3.3). If phase information is available from other sources, the correlation between the electron-density maps is the basis for the phased translation function (Section 13.3.4). The power of the translation functions can be enhanced in the presence of noncrystallographic symmetry (Section 13.3.8). Proper packing of the search model in the crystal unit cell is an essential component of a solution to the translation problem (Section 13.3.5).

13.3.2. *R*-factor and correlation-coefficient translation functions

The crystallographic R factor is often used as an indicator in translation searches. It is a measure of the percentage difference between the observed (F_h^o) and the calculated (F_h^c) structure-factor amplitudes,

$$R_F = 100 \times \sum_h \left| F_h^o - k_F F_h^c \right| \Big/ \sum_h F_h^o. \tag{13.3.2.1}$$

A similar R factor can be defined based on the square of the structure-factor amplitudes, *i.e.* an R factor based on intensity,

$$R_I = 100 \times \sum_h \left| I_h^o - k_I I_h^c \right| \Big/ 2\sum_h I_h^o. \tag{13.3.2.2}$$

A factor of 2 is introduced in equation (13.3.2.2) to make R_I values fall in the same range as R_F. In equations (13.3.2.1) and (13.3.2.2) k_F and k_I are scale factors that bring the observed and the calculated structure factors to the same level. These scale factors are generally calculated in shells of equal reciprocal volume, which can compensate for differences in the displacement factors between the observed and the calculated structure-factor amplitudes.

In addition to the R factor, the correlation coefficient between the observed and the calculated structure factors is also used in translation searches. Like the R factors, correlation coefficients can be defined based on the amplitude or the intensity of the reflections, and that based on the amplitude is shown below.

$$cc_F = \frac{\sum_h \left(F_h^o - \langle F_h^o \rangle \right)\left(F_h^c - \langle F_h^c \rangle \right)}{\left[\sum_h \left(F_h^o - \langle F_h^o \rangle \right)^2 \left(F_h^c - \langle F_h^c \rangle \right)^2 \right]^{1/2}}$$

$$= \frac{\sum_h F_h^o F_h^c - \sum_h F_h^o \sum_h F_h^c / N}{\left\{ \left[\sum_h \left(F_h^o \right)^2 - \left(\sum_h F_h^o \right)^2 \Big/ N \right] \left[\sum_h \left(F_h^c \right)^2 - \left(\sum_h F_h^c \right)^2 \Big/ N \right] \right\}^{1/2}} \tag{13.3.2.3}$$

In (13.3.2.3), N is the number of reflections that are used in the calculation and $\langle F_h \rangle$ denotes the average structure-factor amplitude over the reflections. Unlike the R factors, the correlation coefficients do not depend on the overall scale factor between the observed and the calculated structure factors. However, they can be affected by large differences in the overall displacement factors between the observed and the calculated structure factors.

In order to evaluate R factors and correlation coefficients for a translation search, structure factors need to be calculated for the search model, with a given orientation, at different positions in the unit cell. For this special case, where only the positional parameters of the search model are varied, the calculation of the structure factors can be simplified (Nixon & North, 1976; Rae, 1977). The structure-factor equation can be written as a double summation – first over the atoms in one asymmetric unit of the unit cell and then over all the asymmetric units,

$$\overline{F}_h^c = \sum_n \sum_j f_j \exp\{2\pi i h([T_n] + t_n)\}, \tag{13.3.2.4}$$

where j goes over all the atoms in the asymmetric unit and n goes over all the crystallographic asymmetric units. The nth crystallographic symmetry operator is given by

$$x_n = [T_n]x + t_n, \tag{13.3.2.5}$$

where $[T_n]$ is the rotational component and t_n is the translational component of the symmetry operator.

For simplicity, first consider the case where there is only one molecule in the asymmetric unit. In the translation search, the model will be placed at different positions in the unit cell,

$$x_j = x_j^0 + v_0, \tag{13.3.2.6}$$

where v_0 is a translation vector which is applied to move the model from its starting position (x_j^0). Substituting x_j in (13.3.2.6) into (13.3.2.4) gives

$$\overline{F}_h^c = \sum_n \overline{f}_{h,n} \exp(2\pi i h[T_n]v_0), \tag{13.3.2.7}$$

where $\overline{f}_{h,n}$ is the structure factor calculated based only on the nth symmetry-related molecule

$$\overline{f}_{h,n} = \sum_j f_j \exp\left\{ 2\pi i h\left([T_n]x_j^0 + t_n\right) \right\}. \tag{13.3.2.8}$$

It can be calculated by placing the search model in a $P1$ unit cell having the same cell dimensions as the unknown-crystal unit cell. The structure factors calculated for this $P1$ cell are related to $\overline{f}_{h,n}$ by

$$\overline{f}_{h,n} = \exp(2\pi i h t_n)\overline{f}_{h[T_n],1}. \tag{13.3.2.9}$$

Therefore, the summation over the atoms in the structure-factor calculation, a rather time-consuming process, needs to be performed only once, for the search model at the starting position. Subsequent structure-factor calculations after translation of the model are no

longer dependent on the number of atoms present in the unit cell [equation (13.3.2.7)]. The starting position is usually chosen such that the centre of the search model is at (0, 0, 0). Then the vector that is determined from the translation searches will define the centre of the model in the unit cell.

Equation (13.3.2.7) can be generalized to allow for the presence of other molecules that are to remain stationary during the translation search:

$$\overline{F}_h^c = \overline{A}_h + \sum_n \overline{f}_{h,n} \exp(2\pi i h[T_n]v_0), \qquad (13.3.2.10)$$

where $\overline{A}_h$ is the contribution from the stationary molecules. This formulation is useful if there is more than one molecule in the asymmetric unit. The position of one of the molecules can be determined first, and the model is then included as a stationary molecule for the position search of the next molecule.

Evaluation of the R factor and the correlation coefficient [equations (13.3.2.1) and (13.3.2.3)] in a translation search is generally rather slow. A method has been developed to calculate the correlation coefficient by the fast Fourier transform (FFT) technique, which involves reciprocal vectors up to four times the resolution of the reflection data (Navaza & Vernoslova, 1995).

Equation (13.3.2.7) can also be generalized to allow for the presence of two (or more) search models that are to move independently of each other during the translation search. However, this will generally lead to a six- (or more) dimensional problem and is extremely expensive in computation time. With recent improvements in computer technology (especially parallel processing), it might be feasible to carry out such searches in special cases. However, this aspect of translation functions will not be discussed further here.

13.3.3. Patterson-correlation translation function

The most commonly used translation-search indicator is based on the correlation between the observed and the calculated Patterson maps (Crowther & Blow, 1967). Rotation functions are based on the overlap of only a subset of the interatomic vectors in the Patterson map, *i.e.* only those near the origin of the unit cell, which generally contain the self vectors within each crystallographically unique molecule. The correct orientation and position of a search model in the crystal unit cell should lead to the maximal overlap of both the self and the cross vectors, *i.e.* maximal overlap between the observed and the calculated Patterson maps throughout the entire unit cell (Tong, 1993),

$$PC = \int_\Omega P^o(u)P^c(u) \, du$$
$$= \sum_h (F_h^o)^2 |\overline{F}_h^c|^2. \qquad (13.3.3.1)$$

The calculated structure factor is a function of the translation vector v_0 [equation (13.3.2.10)]. Combining equations (13.3.2.10) and (13.3.3.1) gives

$$PCTF(v_0) = \sum_h (F_h^o)^2 |A_h|^2 + \sum_h \sum_n (F_h^o)^2 |\overline{f}_{h,n}|^2$$
$$+ 2\sum_h \sum_n (F_h^o)^2 \overline{A}_h \overline{f}_{h,n}^* \exp(-2\pi i h[T_n]v_0)$$
$$+ \sum_h \sum_n \sum_{m \neq n} (F_h^o)^2 \overline{f}_{h,n} \overline{f}_{h,m}^* \exp\{-2\pi i h([T_m] - [T_n])v_0\}.$$
$$(13.3.3.2)$$

The first two terms in equation (13.3.3.2) contribute a constant to the correlation and are generally ignored in this calculation (but see Section 13.3.8). The Patterson-correlation translation function is therefore a Fourier transform, with $h[T_n]$ (from the third term) and

$h([T_m] - [T_n])$ (from the fourth term) as the indices. This function can be evaluated quickly by the FFT technique. The term $h([T_m] - [T_n])$ often leads to the doubling of the original reflection indices. For example, $h([T_1] - [T_2])$ will give rise to $(2h, 0, 2l)$ for monoclinic space groups (*b*-unique setting). Therefore, the Patterson-correlation Fourier transform should normally be sampled with a grid size about 1/6 of the maximum resolution of the reflection data used in the calculation.

A disadvantage of the Patterson-correlation translation function as formulated in equation (13.3.3.2) is that the results from the calculation are on an arbitrary scale. This makes it difficult to compare results from different calculations. The R factor or the correlation coefficient can be calculated for the top peaks in the Patterson-correlation translation function [equation (13.3.3.2)] to place the results on an 'absolute' scale. Alternatively, the correlation, as defined by equation (13.3.3.1), can be normalized to become a Patterson-correlation coefficient (Harada *et al.*, 1981),

$$PC = \frac{\sum_h (F_h^o)^2 |\overline{F}_h^c|^2}{\left[\sum_h (F_h^o)^4 \sum_h |\overline{F}_h^c|^4\right]^{1/2}}. \qquad (13.3.3.3)$$

This correlation coefficient is equivalent to that defined by equation (13.3.2.3) for reflection intensities (Fujinaga & Read, 1987). The FFT technique can be used to evaluate the $\left(\sum_h |\overline{F}_h^c|^4\right)^{1/2}$ term in the denominator of equation (13.3.3.3), which involves reciprocal vectors up to four times the data resolution (Navaza & Vernoslova, 1995). Alternatively, this term can be approximated by an expression which also measures the packing of the search molecules in the crystal (see Section 13.3.5).

The intramolecular vectors can be removed (Crowther & Blow, 1967) from the Patterson maps by subtracting, after appropriate scaling, the structure factors calculated from individual molecules [equation (13.3.2.8)].

The translation functions, as defined above, are based on the structure-factor amplitudes. Normalized structure factors (E's) may provide better results under certain circumstances, since they increase the weight of the high-resolution reflections in the translation function (Harada *et al.*, 1981; Tickle, 1985; Tollin, 1966). Since the Patterson-correlation translation function is also based on reflection intensities, the 'large term' approach can be used to accelerate the calculation (Tollin & Rossmann, 1966).

13.3.4. Phased translation function

If an atomic model needs to be placed in an electron-density map that has been obtained through other methods (*e.g.* the multiple-isomorphous-replacement method or partial model phases), the phased translation function can be used (Bentley & Houdusse, 1992; Colman *et al.*, 1976; Read & Schierbeek, 1988; Tong, 1993). It is essentially based on the correlation between observed and calculated electron-density values throughout the unit cell:

$$PTF(v_0) = \sum_h \sum_n \overline{F}_h^o \overline{f}_{h,n}^* \exp(-2\pi i h[T_n]v_0). \qquad (13.3.4.1)$$

Therefore, the phased translation function can also be evaluated by the FFT technique. As with the Patterson-correlation translation function, the phased translation function can be placed on an 'absolute' scale by introducing appropriate normalizing factors or by converting the results to R factors or correlation coefficients. It should be noted that the prior phase could be in the wrong hand, so both $\overline{F}_h^o$ and $(\overline{F}_h^o)^*$ may need to be tried in the phased translation function.

The stationary molecules contribute a constant to the phased translation function and are not shown in equation (13.3.4.1). However, the phase information from the stationary molecules can