Acta Cryst. (1994). D**50**, 675–686

Enhancement of the Method of Molecular Replacement by Incorporation of Known Structural Information

BY XUE-JUN ZHANG AND BRIAN W. MATTHEWS

Institute of Molecular Biology, Howard Hughes Medical Institute, and Department of Physics, University of Oregon, Eugene, Oregon 97403, *USA*

(*Received* 11 *May* 1993; *accepted* 24 *February* 1994)

Abstract

Crystals of macromolecules often have two or more molecules per asymmetric unit, or contain domains of a macromolecule or a macromolecular complex that are structurally independent. In such cases the conventional molecular-replacement method attempts to determine the position of each structural unit independently. Typically, some parts of the structure can be determined more easily or more reliably than other parts. Methods are proposed whereby information from a part of a crystal structure that has been determined can be used to help determine the structure of the remainder. Two different strategies are discussed, 'subtraction' and 'addition'. With 'subtraction' strategy the Patterson function of the known part of the structure is subtracted from the 'observed' Patterson. This approach is found to be most effective in the context of the rotation function in that it eliminates peaks that are irrelevant to the desired solution. With 'addition' strategy the structure factors of the known component are added to those of the search model. This procedure is most effective in the context of the translation function because it brings the structure factors calculated from the search model closer to those observed. Methods of applying the fast Fourier transform to facilitate these calculations are described. A number of examples are provided including structures of mutants of T4 lysozyme that might not have been solved without recourse to the proposed methods. A method of including information from a heavy-atom derivative in a translation function is also developed and shown to be superior in some situations to the conventional translation function.

Introduction

The basic idea of molecular replacement is to use the known structure of a macromolecule in one crystal form to determine the structure of the same or of a related macromolecule in new crystal forms (Hoppe, 1957; Rossmann & Blow, 1962; Huber, 1965; Rossmann, 1972; Machin, 1985). If the molecule of interest can be approximated with a search model, it may be possible to place the search model in the unit cell of the 'new' crystal and so obtain a set of preliminary phases from which the actual structure can eventually be derived.

An 'ideal' search model is one that matches exactly the entire structure that is to be determined. In practice, the search model and the structure to be determined may differ in many respects. For example, there may be changes in conformation of the desired structure relative to the search model or imperfect sequence and structural identity between the crystallized protein and the search model. Also the crystal may include regions or domains for which independent information is available, or there may be multiple copies of the molecule in the asymmetric unit.

In these cases, if the complete structure cannot be solved in one step, it may be possible to use molecular replacement to solve the structure in parts. In principle, if the search model is capable of showing a detectable signal in both the rotation and translation functions, any part of a crystallographic asymmetric unit can be solved independently. In practice, search models for different parts of a crystal structure have different effectiveness, even if they correspond to the same fraction of the structure. For example, it is almost always the case that the rotation function for a crystal containing two or more molecules per asymmetric unit has unequal peak heights for the different molecules. Differences in crystal environment, in the mobility of distinct structural fragments, and in the resemblance of the search model to the crystal structure, can contribute to differences in search power. The type of secondary structures in the molecule of interest can also have a dramatic effect on the ability to determine its location: α-helical structures seem to be more effective as search models than structures dominated by loops and β-sheets (Driessen & White, 1985; Sheriff, Padlan, Cohen & Davies, 1990). Thus, some parts of a crystal structure are usually easier to determine than others. In these situations, inclusion of the partial structural information that has been obtained can be helpful in searching for the remaining part.

In this report, we describe two general methods ('addition' and 'subtraction') used to enhance the effectiveness of molecular replacement by the inclusion of known structural information. In addition, a new translation function based on the incorporation of data from an isomorphous heavy-atom derivative is also proposed. Some aspects of the 'addition' method are available in the program packages *MERLOT* (Fitzgerald, 1988, 1991)

ISSN 0907-4449

and *X-PLOR* (Brünger, Kuriyan & Karplus, 1987) but we are not aware of a comprehensive discussion of this subject.

Theoretical background

In molecular replacement the orientation and position parameters of an appropriate search model(s) are determined by searching over the parameter space in order to minimize the difference between the observed structure amplitudes, $F_o(\mathbf{h})$, and those calculated from the rotated and/or translated model $[F_c(\mathbf{h})]$. Alternatively, the search can be based on functions of $F_o(\mathbf{h})$ such as the Patterson function.

Mathematically, there are two types of functions that can be used to measure the difference between the observed and calculated data, namely the residual function and the correlation function. If $\mathbf{O}$ is an operator (specifying a translation or a rotation or both) acting on a molecular model, the residual function, $R(\mathbf{O})$, can be written as,

$$R(\mathbf{O}) = \Sigma_{\mathbf{h}} \big| |F_o(\mathbf{h})| - k|\mathbf{O}F_c(\mathbf{h})| \big| / \Sigma_{\mathbf{h}} |F_o(\mathbf{h})|. \quad (1)$$

In general, the effectiveness of a residual function depends on the completeness of the search model and the accuracy of the scale factor k (Nixon & North, 1976).

The correlation form of the rotation function (Rossmann & Blow, 1962) can be written as,

$$C_r(\mathbf{R}) = \langle P_o|\mathbf{R}P_c\rangle / [\langle P_o|P_o\rangle \langle \mathbf{R}P_c|\mathbf{R}P_c\rangle]^{1/2}, \quad (2)$$

where $\mathbf{R}$ is a rotation operator, P_o is the Patterson function derived from the observed intensities, P_c the Patterson function corresponding to the search model and $\langle|\rangle$ denotes the point-by-point sum of the product of the two functions enclosed within the parentheses.

In the following paragraphs we define alternative correlation functions and discuss some mathematical ramifications, but the reader who is interested primarily in general principles can skip to the next section.

The correlation function, $C(f,g)$, between two sets of data, f and g, can be written as,

$$C(f,g) = \langle f|g\rangle / [\langle f|f\rangle \langle g|g\rangle]^{1/2}, \quad (3)$$

where the brackets ($\langle|\rangle$) indicate an integral of the product of the two enclosed data sets. There are two forms of the correlation function, one (sometimes called the product moment correlation) in which the average values of the f and g are subtracted, the other without subtraction (normalized inner product). Most of the commonly used rotation and translation functions utilize one form or the other of these correlation functions, and the integral is usually implemented as a summation. The first form of the correlation function (with subtraction), between a data set f, and a data set g acted on by an operator $\mathbf{O}$, can be written as,

$$C_1(f,\mathbf{O}g) = \Sigma(f - f_A)(\mathbf{O}g - \mathbf{O}g_A) \times [\Sigma(f - f_A)^2 \Sigma(\mathbf{O}g - \mathbf{O}g_A)^2]^{-1/2}, \quad (4)$$

where the subscript A indicates the average value of the corresponding data set.

The second form of the correlation function (without subtraction) reduces to,

$$C_2(f,\mathbf{O}g) = \Sigma(f\mathbf{O}g)/[\Sigma ff\, \Sigma \mathbf{O}g\mathbf{O}g]^{1/2}. \quad (5)$$

In the following discussion, we will focus on the correlation function, particularly (5), as the function to be minimized. In many crystallographic applications the averages of f and g are zero, in which case (4) reduces to (5).

In the crystallographic context the volume integral of the rotation function [(2) and (5)] should include the intramolecular vectors while excluding as many intermolecular vectors as possible. For convenience, the integral volume is often chosen as a sphere or a spherical shell, centered at the origin of the Patterson function. In this case, the denominator in (2) is independent of the rotation operator $\mathbf{R}$. Keeping this denominator, however, makes the range of values of $C_r(\mathbf{R})$ physically more meaningful.

Although it can be expressed mathematically in different coordinate systems, the rotation function of (2) is most commonly calculated in either the Eulerian angular system (Crowther's fast rotation function) (Crowther, 1972) or in the spherical polar angular system (Tanaka, 1977) in order to take advantage of the high speed of the integration algorithm with spherical harmonic functions. Crowther's fast rotation function, for example, implements the spherical integral with summations of the series of coefficients (Kabsch, 1986).

$$C_{r,\text{Crowther}}(\alpha,\beta,\gamma) = \Sigma_{l,m,m'} C_{l,m,m'} R_{l,m',m}(\alpha,\beta,\gamma), \quad (6)$$

where $\{C_{l,m,m'}\}$ is the set of coefficients associated with the product of the two Patterson functions and the coefficients $\{R_{l,m',m}\}$ are associated with a rotation specified by (α,β,γ) in an Eulerian angular system.

One common problem in using these angular systems is that a rotation function contoured in the normal way will usually show a 'singularity' when a peak occurs at or close to some special positions. For example, in the polar angular system a peak at or near the origin will appear as a high value on the entire $\kappa = 0$ section. This in turn will affect the apparent average value and standard deviation of the rotation function and will cause these values to change according to the reference orientation specified for the search model. For a rotation function defined with spherical polar angles, this problem can be avoided by representing the rotation-function map on a three-dimensional sphere. In this method, the spherical polar coordinates (φ,θ,r) represent the sampling angles (φ,θ,κ). φ is between 0 and 180°, θ is between 0 and 180° and κ is between −180 and 180°. The standard

deviation (σ) of the $C_r(\varphi,\theta,\kappa)$ map can be defined as follows,

$$\sigma = [\Sigma\kappa^2 \sin\theta C_r^2(\varphi,\theta,\kappa)/\Sigma\kappa^2 \sin\theta]^{1/2}. \quad (7)$$

With this definition, every sample point is weighted by its differential volume, and is thereby less dependent on the initial orientation of the search model. The method is an extension of the two-dimensional projection methods used in the rotation-function program *GLRF* (Tong & Rossmann, 1990). An example of the implementation of this method is given under *Examples*.

Once the correct orientation of the search model has been identified, several types of translation function have been proposed to determine the translation parameters (Crowther & Blow, 1967; Nixon & North, 1976; Harada, Lifchitz & Beathou, 1981; Fujinaga & Read, 1987; Read & Schierbeek, 1988; Driessen *et al.*, 1991). Translation functions are often classified as 'fast' or 'slow', according to whether or not the fast Fourier transform (FFT) can be applied (Rossmann, 1972). This, in fact, depends on the nature of the calculated structure factor $F_c(\mathbf{h},\mathbf{t})$ [$= |F_c|\exp(i\varphi)$]. If F_c is used as a complex number then the FFT can be applied to the translation function calculation. On the other hand if only a part of F_c is used this is no longer the case. The following translation function, for example (8), cannot be evaluated with the FFT because it uses only the amplitude of the calculated structure factor, $|F_c|$, which is not an analytical function of the translation vector $\mathbf{t}$.

$$\begin{aligned} C_t(\mathbf{t}) = {} & \Sigma_{\mathbf{h}}[|F_o(\mathbf{h})| - |F_o(\mathbf{h})|_A][|F_c(\mathbf{h},\mathbf{t})| - |F_c(\mathbf{t})|_A] \\ & \times \{\Sigma_{\mathbf{h}}[|F_o(\mathbf{h})| - |F_o(\mathbf{h})|_A]^2 \\ & \times \Sigma_{\mathbf{h}}[|F_c(\mathbf{h},\mathbf{t})| - |F_c(\mathbf{t})|_A]^2\}^{-1/2}. \end{aligned} \quad (8)$$

As before, the subscript A indicates the average value. $|F_c(\mathbf{h},\mathbf{t})|$ is calculated as,

$$\begin{aligned} |F_c(\mathbf{h},\mathbf{t})|\exp(i\varphi) &= F_c(\mathbf{h},\mathbf{t}) \\ &= \Sigma_j^{\text{sym}} F_{c,P1}(\mathbf{hR}_j)\exp[i2\pi\mathbf{h}(\mathbf{t}_j + \mathbf{R}_j\mathbf{t})] \end{aligned} \quad (9)$$

where $\mathbf{R}_j$ and $\mathbf{t}_j$ are the rotation and translation components of the jth symmetry operator of the crystal; $\mathbf{t}$ is the translational vector of the search model in real space and $F_{c,P1}$ is the structure factor of the correctly oriented search model calculated in the observed crystal cell but with symmetry $P1$.

To take advantage of the FFT, the following translation function can be used,

$$\begin{aligned} C_t(\mathbf{t}) &= \langle P_o|P_c(\mathbf{t})\rangle/[\langle P_o|P_o\rangle\langle P_c(\mathbf{t})|P_c(\mathbf{t})\rangle]^{1/2} \\ &= \Sigma_{\mathbf{h}} I_o(\mathbf{h}) I_c(\mathbf{h},\mathbf{t})/[\Sigma_{\mathbf{h}} I_o^2(\mathbf{h}) \Sigma_{\mathbf{h}} I_c^2(\mathbf{h},\mathbf{t})]^{1/2} \end{aligned} \quad (10)$$

where $I(\mathbf{h})$ is the intensity $[F(\mathbf{h})F^*(\mathbf{h})]$. Both the numerator and denominator can be calculated as functions of the translation vector $\mathbf{t}$ using the FFT. Because of the terms involving $I_c(\mathbf{h},\mathbf{t})$ and $I_c^2(\mathbf{h},\mathbf{t})$, the maximum indices of coefficients for the Fourier transforms can be twice or four times as large, respectively, as those of the input structure factors. Therefore, in practice the coefficients often need to be truncated, especially for crystals with large unit cells. Also the denominator may be replaced with an approximation function.

For example, the overlap function proposed by Harada and coworkers (Harada *et al.*, 1981) uses $P_c(\mathbf{u} = \mathbf{0},\mathbf{t})$ $[= \Sigma_{\mathbf{h}} I_c(\mathbf{h},\mathbf{t})]$ to replace $[\langle P_c(\mathbf{t})|P_c(\mathbf{t})\rangle]^{1/2}$ $\{= [\Sigma_{\mathbf{h}} I_c^2(\mathbf{h},\mathbf{t})]^{1/2}\}$. It actually works as well as (10) in terms of signal-to-noise ratio, especially when a series truncation has to be made in (10). The physical meaning of including $P_c(\mathbf{u} = \mathbf{0},\mathbf{t})$ is to monitor the value at the origin of the Patterson function of the calculated model while translating the model through the unit cell, looking for solutions that avoid interpenetration of different molecules. Similarly, $\langle P_c(\mathbf{t})|P_c(\mathbf{t})\rangle$ considers the overall Patterson function, reducing the weight whenever there is intermolecular overlap. Since it is defined in terms of Patterson space rather than reciprocal space, (10) not only makes it possible to use the FFT, it also is physically more meaningful than (8). This may become clearer in subsequent discussion.

In the following, we will refer to (2) and (10) as the 'ordinary rotation function' and the 'ordinary translation function', respectively.

Enhancement of molecular replacement by the incorporation of known information

Generally speaking, there are two ways to incorporate known structural information into molecular replacement, either by an 'addition' strategy or by a 'subtraction' approach. The objective of the former is to increase the signal while the latter is intended to reduce the noise. In addition strategy, the information from the part of the structure that is known is used to supplement the search model for the remaining part still to be solved. The objective is that the structure factors or Patterson function calculated from the enhanced model will better resemble the observed data. For the addition strategy to be effective, the known structural information should provide an independent signal to the correlation function. This requires that the addition term respond to changes of the operator (*e.g.* a translational operator), rather than being a constant term that is added.

With subtraction strategy, the information from the part of the structure that has been determined is subtracted from the observed data so that the modified observations will more closely represent the part of the structure still to be solved. Use of the subtraction strategy is intended to enhance the desired peaks by reducing noise and by eliminating peaks which correspond to the part of the structure already known. Subtraction strategy can be used only in Patterson space. It cannot be applied in the reciprocal space formulations such as in the 'slow'

translation function (8) which require the amplitudes of the structure factors. This is because a calculated structure factor cannot be subtracted from the amplitude of an observed reflection without knowledge of its phase information. The subtraction strategy requires a reasonably accurate scale factor between the observed and calculated data, which may be estimated with Wilson statistics (Schierbeek, Renetseder, Dijkstra & Hol, 1985) and from the knowledge of the percentage of the overall structure that is being subtracted.

Suppose that a crystal structure consists of two parts, a and b, and that an appropriate search model has been used to determine either the orientation or the orientation plus translation of part a. The structure of part b is still to be determined. It is not necessary that $a + b$ comprise the whole asymmetric unit, although this will be assumed for convenience. The Patterson function of the observed crystal can be written as,

$$\begin{aligned} P_o = P_{a+b} &= P_a + P_b + P_{ab} \\ &= P_{a,\text{intra}} + P_{aa} + P_{b,\text{intra}} + P_{bb} + P_{ab}, \end{aligned} \quad (11)$$

where

$$P_a = P_{a,\text{intra}} + P_{aa}, \quad (12)$$

and

$$P_b = P_{b,\text{intra}} + P_{bb}. \quad (13)$$

$P_{a,\text{intra}}$ and $P_{b,\text{intra}}$ include the intramolecular vectors while P_{aa} and P_{bb} represent the intermolecular vectors between the symmetry-related a fragments and the symmetry-related b fragments, respectively. P_{ab} includes the vectors between all the symmetry-related copies of fragment a and all the symmetry-related copies of fragment b. Decomposing into symmetry-related molecules (or fragments) $P_{a,\text{intra}}$ and P_a can be expressed as follows,

$$\begin{aligned} P_{a,\text{intra}}(\mathbf{u}) &= \Sigma_j^{\text{sym}} \mathbf{S}_j P_{a,P1}(\mathbf{u}) \\ &= \Sigma_{\mathbf{h}} [\Sigma_j^{\text{sym}} I_{a,P1}(\mathbf{h}\mathbf{R}_j)] \exp(-i2\pi \mathbf{h}.\mathbf{u}), \end{aligned} \quad (14)$$

$$P_a(\mathbf{u}) = \Sigma_{\mathbf{h}}\, I_{a,N\text{sym}}(\mathbf{h}) \exp(-i2\pi \mathbf{h}.\mathbf{u}), \quad (15)$$

where $\mathbf{R}_j$ is the rotation matrix of the symmetry operator $\mathbf{S}_j$, $\{I_{a,P1}(\mathbf{h})\}$ is the set of intensities associated with an isolated fragment a (*i.e.* only the unitary symmetry operator is present), and $\{I_{a,N\text{sym}}(\mathbf{h})\}$ is the set of intensities associated with all the symmetry-related copies of fragment a. Similarly, there are corresponding equations for $P_{b,\text{intra}}(\mathbf{u})$ and $P_b(\mathbf{u})$. Consequently, every term in (11) has the same symmetry.

Incorporation of 'subtraction' strategy in the rotation function. 'Subtraction' strategy can be applied in a very straightforward manner by subtracting either $P_{a,\text{intra}}$ or P_a (= $P_{a,\text{intra}} + P_{aa}$) from the observed Patterson function, P_o, depending on whether the alignment or the complete location of part a is known. The rotation function, modified to include the subtraction strategy, can be written as follows,

$$\begin{aligned} C_r^S(\mathbf{R}) = &\langle (P_o - kP_a) | \mathbf{R}P_{b,P1} \rangle \\ &\times [\langle (P_o - kP_a) | (P_o - kP_a) \rangle \langle P_{b,P1} | P_{b,P1} \rangle]^{-1/2}, \end{aligned} \quad (16)$$

where $P_{b,P1}$ is the Patterson function of a single search model for fragment b. P_a is calculated from the model for the known structural fragment. The scale factor k relating the observed and the calculated intensities is given by,

$$k = f \Sigma_{\mathbf{h}} I_o(\mathbf{h}) / \Sigma_{\mathbf{h}} I_a(\mathbf{h}), \quad (17)$$

where f is the fraction of the structure that corresponds to fragment a. $\{I_o(\mathbf{h})\}$ and $\{I_a(\mathbf{h})\}$ are the observed and calculated intensity data sets, respectively.

The main difference between the rotation function with the subtraction strategy and the ordinary rotation function is the subtraction of the term kP_a from P_o in (16). In general, this will reduce 'noise' in the rotation function map. For example, in the case where a and b represent two copies of the same molecule in an asymmetric unit, respectively (a known and b unknown), subtraction of P_a can significantly reduce the unwanted peaks at orientations corresponding to molecule a.

Since the rotation function relies primarily on intramolecular vectors, the intermolecular vectors (P_{aa}) are usually less important. For this reason, knowledge of the translational parameters of a may not be necessary. Specifically, if only the orientation of fragment a is known, it may be possible to adjust the symmetry-related fragments away from each other so that most of the calculated (incorrect) intermolecular vectors (P_{aa}) will fall outside of the integral sphere of the rotation function.

Modification of the rotation function to incorporate addition strategy. Direct implementation of addition strategy in the rotation function [*i.e.* replacement of $\mathbf{R}P_c$ with $(P_a + \mathbf{R}P_{b,P1})$ in (2)] leads to,

$$\begin{aligned} C_r^E(\mathbf{R}) = &\langle P_o | (P_a + \mathbf{R}P_{b,P1}) \rangle \\ &\times [\langle P_o | P_o \rangle \langle (P_a + \mathbf{R}P_{b,P1}) | (P_a + \mathbf{R}P_{b,P1}) \rangle]^{-1/2}. \end{aligned} \quad (18)$$

The extra term $\langle P_o | P_a \rangle$ in the numerator does not depend on the rotation of the search molecule b. Thus, it adds a constant but is not expected to improve the signal-to-noise ratio of the rotation function.

A method to use information from crystallographic symmetry to enhance the rotation function was proposed by Nordman (1986) and discussed further by Yeates (1989). Although it is not particularly effective in general, it may be helpful in reducing so-called symmetry bias (Yeates, 1989) when the search model

has local symmetry that is similar to an element of the crystallographic symmetry. In this special situation the use of a model including local symmetry, *e.g.* a dimer or trimer, has the advantage that the intermolecular vectors will enhance the search power of the rotation function if these vectors also exist in the crystal. The technique of simultaneously searching for symmetry-related molecules is discussed in more detail in the *Appendix*.

Similarly, information about local symmetry obtained from the self-rotation function (*i.e.* obtained without the use of a known structural model) can be incorporated into a cross rotation function. This has been called the locked rotation function and is discussed by Tong & Rossmann (1990). Assume, for example, that the self rotation function shows one independent peak corresponding to a local symmetry element, *e.g.* a non-crystallographic rotation from molecule *b* to molecule *a*, which is represented by the rotation operation $\mathbf{R}_{\rm self}$. The modified rotation function that includes this information can be written as,

$$C^E_{r,\rm self}(\mathbf{R}) = \langle P_o|(1 + \mathbf{R}_{\rm self})\mathbf{R}P_{b,P1}\rangle \times [\langle P_o|P_o\rangle\langle(1 + \mathbf{R}_{\rm self})\mathbf{R}P_{b,P1}| (1 + \mathbf{R}_{\rm self})\mathbf{R}P_{b,P1}\rangle]^{-1/2}. \quad (19)$$

With an algorithm similar to that discussed in the *Appendix*, one can show that this equation can be implemented within the framework of the fast rotation function. Equation (19) basically corresponds to the overlay of the ordinary rotation-function map on itself, but rotated by $\mathbf{R}_{\rm self}$. When the maps are superimposed the signal will rise but so will the noise. If the noise is distributed randomly, however, the superposition will improve the signal-to-noise ratio. In a case where the local symmetry includes a large number (n) of copies, *e.g.* in a virus crystal, the relative noise can be expected to be reduced by a factor of $n^{1/2}$ (Tong & Rossmann, 1990).

The translation function including both addition and subtraction strategy. An addition strategy has been used previously in an R-factor search (Bi *et al.*, 1983), including determination of the relative origins of independently solved structural fragments in space groups in which there is a free choice of origin, in a correlation translation function search (Fujinaga & Read, 1987). Similarly, subtraction of the intramolecular vectors of the search model from the observed Patterson map has been considered in some Patterson space translation functions (Rossmann, Blow, Harding & Coller, 1964; Crowther & Blow, 1967).

As before, we assume that part a of the desired structure is known in its entirety and, in addition, the rotation of part b is also known. A translation function (20) that incorporates both subtraction and addition strategies and, at the same time, allows the application of the FFT, can then be defined.

$$C_t(\mathbf{t}) = \langle P'_o|[P_{ab}(\mathbf{t}) + P_{bb}(\mathbf{t})]\rangle / \{\langle P'_o|P'_o\rangle \times \langle[P_{ab}(\mathbf{t}) + P_{bb}(\mathbf{t})]|[P_{ab}(\mathbf{t}) + P_{bb}(\mathbf{t})]\rangle\}^{1/2}. \quad (20)$$

$P_{ab}(\mathbf{t})$ corresponds to the calculated intermolecular vectors from all the symmetry-related a parts to all the symmetry-related b parts, and $P_{bb}(\mathbf{t})$ corresponds to the vectors generated by all the symmetry-related b parts. P'_o is a modified observed Patterson map defined by,

$$P'_o = P_o - k(P_a + P_{b,\rm intra}) \simeq P_{o,ab} + P_{o,bb}. \quad (21)$$

The numerator of (20) has been independently proposed by Driessen *et al.* (1991) as the 'non-crystallographic translation function'.

Equation (20) can be implemented as follows,

$$C_t(\mathbf{t}) = \Sigma_{\mathbf{h}} I'_o(\mathbf{h})[I_{ab}(\mathbf{h},\mathbf{t}) + I_{bb}(\mathbf{h},\mathbf{t})] \times \{\Sigma_{\mathbf{h}}[I'_o(\mathbf{h})]^2 \Sigma_{\mathbf{h}}[I_{ab}(\mathbf{h},\mathbf{t}) + I_{bb}(\mathbf{h},\mathbf{t})]^2\}^{-1/2}, \quad (22)$$

where

$$I'_o(\mathbf{h}) = I_o(\mathbf{h}) - k[I_{c,a}(\mathbf{h}) + \Sigma_j^{\rm sym} I_{b,P1}(\mathbf{h}\mathbf{R}_j)], \quad (23)$$

$$I_{ab}(\mathbf{h},\mathbf{t}) = F_{c,a}(\mathbf{h})\Sigma_j^{\rm sym} F^*_{b,P1}(\mathbf{h}\mathbf{R}_j) \times \exp[-i2\pi\mathbf{h}(\mathbf{t}_j + \mathbf{R}_j\mathbf{t})] + \text{c.c.}, \quad (24)$$

and

$$I_{bb}(\mathbf{h},\mathbf{t}) = \Sigma_j^{\rm sym} F_{b,P1}(\mathbf{h}\mathbf{R}_j)\exp[i2\pi\mathbf{h}(\mathbf{t}_j + \mathbf{R}_j\mathbf{t})] \times [\Sigma_{j'(>j)}^{\rm sym} F^*_{b,P1}(\mathbf{h}\mathbf{R}_{j'})] \times \exp[-i2\pi\mathbf{h}(\mathbf{t}_{j'} + \mathbf{R}_{j'}\mathbf{t})] + \text{c.c.}, \quad (25)$$

where c.c. stands for the corresponding complex conjugate. (Alternatively one can remember that the sum of a function plus its complex conjugate equals twice the real part of the function.) The parameter k in (23) specifies the amount of subtraction of the constant intra- and intermolecular vectors from the observed Patterson function. The upper limit of k is estimated using Wilson statistics and is modulated by the structural percentage of the included model(s), similar to (17). In the lower limit k can be set equal to zero. Both the numerator and denominator of (22) can be evaluated with the FFT by using coefficients indexed conjugated to the translation vector $\mathbf{t}$. For a crystal that has a space group with a primitive lattice (P), the summations over the symmetries in (23)–(25) should include all of the crystallographic symmetry operators. For space groups that have non-primitive lattices (*e.g.* C or R), the lattice symmetry operators should not be included in the summations in (23)–(25) but should be included in the calculation of $F_{b,P1}(\mathbf{h})$. This procedure is equivalent to performing the translation search in a reindexed primitive (and therefore smaller) unit cell.

The combination of subtraction and addition strategies in a single translation function should significantly enhance its effectiveness, particularly when much of the structure is known and the part that remains to be determined is a relatively small fraction of the unit cell.

Use of heavy-atom derivative information to determine translation

When attempting molecular replacement it is common, whenever possible, to use heavy-atom information to check the results. A difference map with amplitudes $[|F_{PH}(\mathbf{h})| - |F_P(\mathbf{h})|]$ and phases from the proposed molecular-replacement model should show peaks at the heavy-atom sites (Cygler & Anderson, 1988). (F_P is the structure amplitude of the native protein and F_{PH} that of the heavy-atom derivative. Usually F_P is the same as F_o but here we need to distinguish the observed amplitude of the native crystal from that of the heavy-atom derivative.) Even though a single heavy-atom derivative may provide fairly reliable coordinates for the heavy-atom binding sites, it need not give sufficiently good phases to allow determination of the crystal structure. One approach in this situation is to calculate a single isomorphous replacement electron-density map and to use the structural model to search directly in this map. The method is extremely powerful, with a very high signal-to-noise ratio (Bode *et al.*, 1983; Reynolds *et al.*, 1985), although in the most general case can require a lengthy six-dimensional search. If a solution to the rotation function has been obtained it reduces the six-dimensional search to one in three dimensions and the translation search can be performed with the FFT as follows,

$$T(\mathbf{t}) = \int_{\text{cell}} \rho_{\text{obs}}(\mathbf{x})\rho_{\text{model}}(\mathbf{x} - \mathbf{t})\,d\mathbf{x} \quad (26)$$

$$= \Sigma_{\mathbf{h}} m_P |F_P(\mathbf{h})| \exp(i\varphi_P) F_c^*(\mathbf{h}) \exp(-i2\pi \mathbf{h}.\mathbf{t}) \quad (27)$$

$$= \Sigma_{\mathbf{h}} m_P |F_P(\mathbf{h})| \exp(i\varphi_P) F_c^*(\mathbf{h},\mathbf{t}).$$

$|F_P|$ is the observed protein structure amplitude, φ_P the (approximate) protein phase and m_P the figure of merit. In its correlation function form (27) was proposed as the so-called 'phased translation function' by Read & Schierbeek (1988). The structure factor $F_c(\mathbf{h},\mathbf{t})$ may include not only the full crystallographic symmetry, as Read & Schierbeek pointed out, but also information from a known structural fragment as we have suggested in the context of the other translation functions. The phased translation function can be written as

$$C_{t,\text{phased}}(\mathbf{t}) = \Sigma_{\mathbf{h}} m_P |F_P(\mathbf{h})| \exp(i\varphi_P) F_c^*(\mathbf{h},\mathbf{t}) \times [\Sigma_{\mathbf{h}} m_P^2 |F_P^2| \Sigma_{\mathbf{h}} |F_c(\mathbf{h},\mathbf{t})|^2]^{-1/2}. \quad (28)$$

With the inclusion of knowledge of part of the structure, F_c becomes

$$F_c(\mathbf{h},\mathbf{t}) = F_a(\mathbf{h}) + \Sigma_j^{\text{sym}} F_{b,P1}(\mathbf{h}\mathbf{R}_j) \exp[i2\pi \mathbf{h}(\mathbf{t}_j + \mathbf{R}_j\mathbf{t})]. \quad (29)$$

Furthermore, the phase information $\{\varphi_P, m_P\}$ is not limited to that from the heavy-atom derivatives. For example a structure partially solved with molecular replacement may provide similar information (Bentley & Houdusse, 1992).

In a crystal of space group $P1$, a part of a structure that has been correctly oriented can be used to define the origin and to provide approximate phases for the structure factors $F_P(\mathbf{h})$. These partial structure phases may then be used to determine or confirm heavy-atom site(s) relative to the same origin. With the phased translation function technique, such phases can also be used to locate other parts of the crystal structure (Bentley & Houdusse, 1992). For non-$P1$ space groups the same principle can be applied by expanding the observed structure amplitudes $|F_P(\mathbf{h})|$ to a $P1$ space group and using a similar technique to determine the heavy-atom positions. By comparing the 'local' symmetry of the heavy-atom sites in the $P1$ cell with the crystal symmetry, the translation vector of the initial search model (correctly oriented but arbitrarily positioned) may be determined (Cygler & Anderson, 1988).

In the following, we propose an alternative way to use heavy-atom-derivative information to determine the translation vector of a structural fragment, assuming that its rotational parameters are known. This 'heavy-atom' translation function, $T_H(\mathbf{t})$, is defined as follows,

$$T_H(\mathbf{t}) = \Sigma_{\mathbf{h}}[|F_{PH}(\mathbf{h})| - |F_P(\mathbf{h})|]\exp[i\varphi_c(\mathbf{h},\mathbf{t})] \times \Sigma_j \exp(-i2\pi \mathbf{h}\mathbf{x}_j), \quad (30)$$

where the $\varphi_c(\mathbf{h},\mathbf{t})$'s are the phases calculated with (9) from the correctly oriented but translated model and $\mathbf{x}_j$ are the fractional coordinates of the jth heavy atom. The idea behind (30) is to monitor the known heavy-atom site(s) while translating the search model through the unit cell. When the search model is at the correct position, $T_H(\mathbf{t})$ will have a relatively high value which corresponds to the electron densities at the monitored heavy-atom site(s). Because the choice of the heavy-atom coordinates will have fixed the origin, the translation search should, in general, cover the whole unit cell.

For reasons similar to those discussed in the context of the slow translation function, (30) cannot be evaluated with the FFT. A modified version that allows the use of the FFT can be obtained by the introduction of a factor $|F_c(\mathbf{h},\mathbf{t})|/|F_P(\mathbf{h})|$, *i.e.*

$$T'_H(\mathbf{t}) = \Sigma_{\mathbf{h}}\{[|F_{PH}(\mathbf{h})| - |F_P(\mathbf{h})|]/|F_P(\mathbf{h})|\} \times F_c(\mathbf{h},\mathbf{t})\Sigma_j \exp(-i2\pi \mathbf{h}\mathbf{x}_j). \quad (31)$$

Here, $F_c(\mathbf{h},\mathbf{t})$ [equation (9)] is an analytical function of $\mathbf{t}$, and hence $T'_H(\mathbf{t})$ can be written as a series of Fourier syntheses with coefficient indexes conjugated to the translation vector $\mathbf{t}$. Strictly speaking, for T_H to be equivalent to T'_H the introduced factor $|F_c(\mathbf{h},\mathbf{t})|/|F_P(\mathbf{h})|$ should be a constant, which is unlikely. In practice, however, it seems sufficient to simply omit reflections for which $F_P(\mathbf{h})$ is small and thereby avoid the introduction of large errors (see below).

One advantage of the $T_H(\mathbf{t})$ function is that it makes use of additional information (*i.e.* the heavy-atom positions) which is independent of the intermolecular vectors. Therefore, a solution provided by this approach is independent of any solution obtained from the ordinary translation function. Another useful feature of the function $T_H(\mathbf{t})$ is that it is less sensitive than the correlation translation functions to the incompleteness of the search model. In test examples (see below), both (30) and (31) were always superior to the ordinary translation function (10) and especially so in cases where the search model constituted a relatively small fraction of the content of the asymmetric unit. The T_H function works well if all coefficients ($|F_{PH}| - |F_P|$) are included. In the case of the T'_H function the signal-to-noise ratio was progressively improved by deleting more and more of the weaker reflections ($|F_P|$) up to about 50% of the observed data (Table 4).

The program

The rotation–translation function package consists of a suite of three major programs designed to interface to the general-purpose macromolecular structure refinement package *TNT* (Tronrud, Ten Eyck & Matthews, 1987). The first program *ALMN* calculates fast rotation function coefficients, $A_{l,m,n}$, from structure factors F_c or F_o. The second program, *ROTFUN*, calculates the rotation function between two Patterson functions. The third program, *FASTRAN*, calculates a modified version of the fast translation function of Harada *et al.* (1981). All input is in a keyword-leading free-formatted form. The programs are coded to run on DEC machines under VAX/VMS, but should be readily transportable. Copies of the programs with write-ups and example command files are available on request from the authors (e-mail: CHK@UOXRAY.UOREGON.EDU).

Examples

Tests of the potential and the limitations of the proposed methods include examples of the rotation function with the subtraction strategy (16), the translation function with both addition and subtraction strategies (20), and the two translation functions using heavy-atom derivative information [(30) and (31)]. Crystallographic data for wild-type T4 lysozyme and a mutant of the enzyme that crystallizes in a non-isomorphous form are used as representative examples.

Wild-type T4 lysozyme crystallizes in space group $P3_221$ with one molecule per asymmetric unit. The structure was solved using multiple isomorphous replacement (Matthews & Remington, 1974) and refined to an R factor of 0.165 with data to 1.7 Å resolution (Weaver & Matthews, 1987; Bell *et al.*, 1991). The protein consists of one peptide chain of 164 amino-acid residues, folded into two domains that are connected by a 20-residue

Table 1. *Rotation-function tests based on wild-type T4 lysozyme*

Model $T4L_{1\ 162}$ includes the whole T4 lysozyme molecule, $T4L_{1-60}$ includes just the N-terminal domain (residues 1–60) and $T4L_{61\ 162}$ includes just the C-terminal domain (residues 61–162). $T4L_{1\ 60}$[61–162]A uses residues 1–60 as the search model but also assumes that the orientation of residues 61–162 is known, *i.e.* the intramolecular vectors contributed by these atoms are subtracted from the Patterson function. (The center of residues 61–162 was moved ~10 Å from its correct position to the center of the molecule as a whole.) $T4L_{1\ 60}$[61–162]B again uses residues 1–60 as the search model but assumes that both the orientation and the translation of resudes 61–162 are known. The rotation functions were calculated using (16) and sampled at increments of (3,3,3°). Observed structure amplitudes between 3.5 and 7 Å resolution were included, and an 18 Å radius was used for the volume integration. $(p - a)/\sigma$ gives the number of standard deviations (σ) that the peak height (p) is above the average value (a) (which is close to zero for this calculation). The same measures are used in the following tables. A peak that is more than 10° away from the correct soution is considered as a noise peak. The 'angular error' is the discrepancy between the apparent angular orientation indicated in the rotation function and that corresponding to the actual refined structure.

Search model	Magnitude of peak closest to correct solution $(p - a)/\sigma$	Ranking of peak height	Magnitude of highest noise peak $(p - a)/\sigma$	σ	Angular error (°)
$T4L_{1\ 162}$	8.2	1	3.9	0.036	0.0
$T4L_{1-60}$	4.1	28	4.8	0.035	7.6
$T4L_{61\ 162}$	7.1	1	3.7	0.038	3.1
$T4L_{1\ 60}$[61–162]A	4.2	1	4.0	0.035	7.2
$T4L_{1\ 60}$[61–162]B	5.0	1	4.0	0.036	4.2

α-helix extending from residues 60 to 80. The amino-terminal domain includes residues 1–60 and consists of several β-sheet strands and a few loops as well as two helices. The carboxyl terminal domain (residues 80–164) is dominated by helical secondary structure.

To test the rotation function including, in particular, the effectiveness of the subtraction strategy (16), the refined structure of wild-type T4 lysozyme (*i.e.* the whole molecule) as well as the N-terminal and C-terminal domains considered separately, were used as representative search models. The results, summarized in Table 1, indicate that the segment consisting of residues 1–60 alone is not sufficient to give a distinct solution in that the desired peaks are not the highest in the rotation function map (Fig. 1*a*). However, when information derived from knowledge of the orientation and translation of the fragment comprising residues 61–162 is included the desired solution emerges as the highest peak (Fig. 1*b*). Subtraction of the contribution of the known part of the structure from the observed Patterson function not only makes the solution detectable but also reduces the rotational error from 7.6° (if the solution could be recognized) to 4.2°.

Successively smaller segments of the structure of T4 lysozyme were also used as search models to test the translation function (20). The results are shown in Table 2 and Fig. 2. Even a relatively small fraction

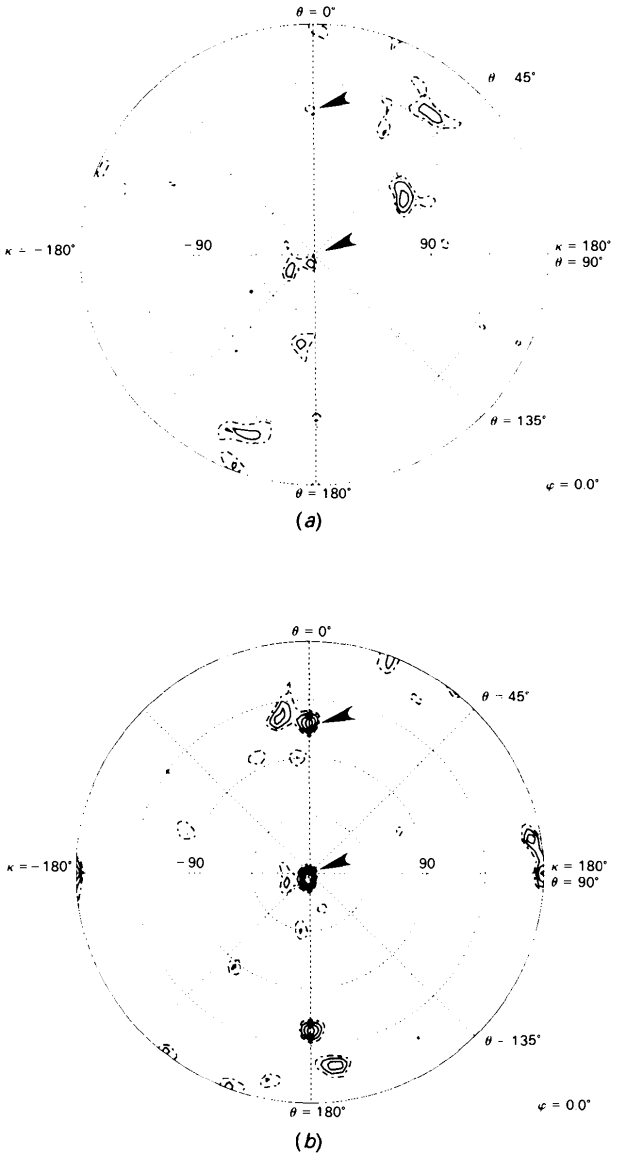

Fig. 1. Illustration of the use of the subtraction method in the rotation function. The example is for T4 lysozyme (space group $P3_221$) as in Table 1. The crystallographic a axis is horizontal (positive to the right) and the c axis is vertical. Section $\varphi = 0.0°$ is shown. A rotation, κ, about an axis (φ,θ) should be indicated by a peak with coordinates (φ,θ,κ). The distance of the peak from the center of the figure gives the angle of rotation (κ). In this case the search model is residues 1–60 of T4 lysozyme aligned relative to the same axes as in the actual crystal structure. Therefore, in the section $\varphi = 0.0°$ two peaks are anticipated at the positions indicated by the arrowheads. One peak is expected at the origin corresponding to the identity superposition and the second peak is expected at $\theta = 0$, $\kappa = 120°$ (plus the symmetry mate at $\theta = 180$, $\kappa = 120°$) corresponding to superposition of the search domain on the lysozyme molecule related by the crystallographic 3_2 symmetry operator. The first contour (broken line) is at 2.0σ and subsequent contours are at increments of 0.5σ where σ is the standard deviation of the map calculated using equation (7). (*a*) Search using residues 1–60 alone. (*b*) Search using the same model (residues 1–60) but subtracting from the observed Patterson function the Patterson function of fragment 61–162 (see Table 1). In this case peaks at the desired positions (arrowheads) are clearly seen.

Table 2. *Fast translation-function tests based on wild-type T4 lysozyme*

Equation (20) with maximum index cutoff = 50 was used to calculate the translation function including data between 3.5 and 7.0 Å resolution. The conventions used to define the search models are as used in Table 1. When search model $T4L_{10-30}$ is used, *i.e.* residues 10–30 of T4 lysozyme, the desired peak is not detected above the noise. When knowledge of the orientation and translation of, for example, residues 1–9 are included, as in $T4L_{10-30}[1–9]$, the desired peak is the highest in the map.

Search model	Magnitude of peak closest to correct solution $(p-a)/\sigma$	Ranking of peak height	Magnitude of highest noise peak $(p-a)/\sigma$
$T4L_{1-162}$	25.5	1	12.5
$T4L_{1-60}$	11.5	1	7.3
$T4L_{1-40}$	9.9	1	7.0
$T4L_{1-30}$	8.0	1	5.6
$T4L_{10-30}$	5.0	21	6.5
$T4L_{10-30}[1–9]$	6.7	1	6.0
$T4L_{10-30}[60–80]$	8.3	1	5.1
$T4L_{10-30}[1–9,31–160]$	19.9	1	5.1

of the whole molecule (*e.g.* residues 1–30) can give a correct determination of the desired translation. It appears that the effectiveness of the translation function is less dependent on the completeness of the search model than is the rotation function (compare Table 1 with Table 2). When a search model alone is being used (20), which reduces to (10), works at least as well in terms of signal to noise as Harada's *TO/O* function (Harada *et al.*, 1981) (data not shown). As shown in Table 2 the inclusion of knowledge of even a small fragment (13%) of the molecule (*e.g.* residues 1–9 or 60–80) significantly improves the power of the translation function. In the case that the search model itself is small (residues 10–30), this improvement can be dramatic.

A second example is provided by the polyalanine mutant lysozyme 9001A, or E128A/V131A/N132A/K135A/S136A/R137A/Y139A/N140A/Q141A, in which the nine amino acids Glu128, Val131, *etc.* are all substituted with alanine. The mutant was crystallized from 0.1 *M* phosphate, 20% PEG, pH 6.5 with space group $P2_1$ and cell dimensions $a = 40.4$, $b = 112.3$, $c = 135.2$ Å and $\beta = 91.7°$. The crystal solvent parameter, V_M (Matthews, 1968), suggested that there might be five ($V_M = 3.4$ Å^3 Da^{-1}) or six ($V_M = 2.8$ Å^3 Da^{-1}) molecules per asymmetric unit. 80% of the possible data were collected in the range 4.0–12.0 Å resolution.

A self-rotation calculation with data between 4.5 and 9.0 Å resolution and 20 Å radius of integration showed a local fivefold axis of symmetry, perpendicular to the 2_1 axis, b, and approximately parallel to a. The highest self-rotation peak, corresponding to the fivefold axis, was 58% that of the origin peak. This strongly suggested that there were five molecules in the asymmetric unit and limited possible solutions of the cross-rotation function.

The general course of the subsequent determination of the structure is outlined in Table 3. In the initial cross-rotation function search (step 2, also performed with data from 4.5 to 9.0 Å resolution and a 20 Å integration radius), a variety of different lysozyme search models with different hinge-bending angles were tested (Faber & Matthews, 1990; Dixon, Nicholson, Shewchuk, Baase & Matthews, 1992; Zhang, Baase & Matthews, 1992). Mutant $I3P_A$ (Dixon *et al.*, 1992) seemed best and was used for all subsequent searches. (Later, it was found that the backbone atoms of the five monomers in the 9001A structure differ, respectively, from $I3P_A$ by root-mean-square values of 0.80, 0.51, 0.72, 0.80 and 0.79 Å.) In the initial cross-rotation function calculation (Table 3, step 2) the top three peaks corresponded to desired solutions (molecules *A*, *E*, *B*), followed by a noise peak. By subsequently making use of the determined location of molecule *A* (step 4), the rotation function revealed the rotations corresponding to the remaining four molecules above the next highest noise peak. It should be noted, however, that the benefit of including the located molecules does not go on indefinitely. By step 8, for example, at which point molecules *A*, *B* and *C* have been located and included in the calculation, the peak for molecule *E* is not the highest and no peak at all was apparent for molecule *D*. The fifth molecule was located by applying the fivefold rotation followed by rigid-body refinement. Another approach, perhaps more general, would have been to first reduce accumulated error by refining the structures of the known parts of the structure (molecules *A*, *B*, *C*, *E*) and repeating step 10 with this improved information. This is illustrated in step 10′, which is identical to step 10 except that molecules *A*, *B*, *C* and *E* were first subjected to rigid-body refinement. Following location of all five molecules, refinement with data between 6.0 and 3.0 Å resolution gave a crystallographic residual of 16.1% with discrepancies of bond lengths and bond angles from expected values of 0.015 and 2.4°, respectively.

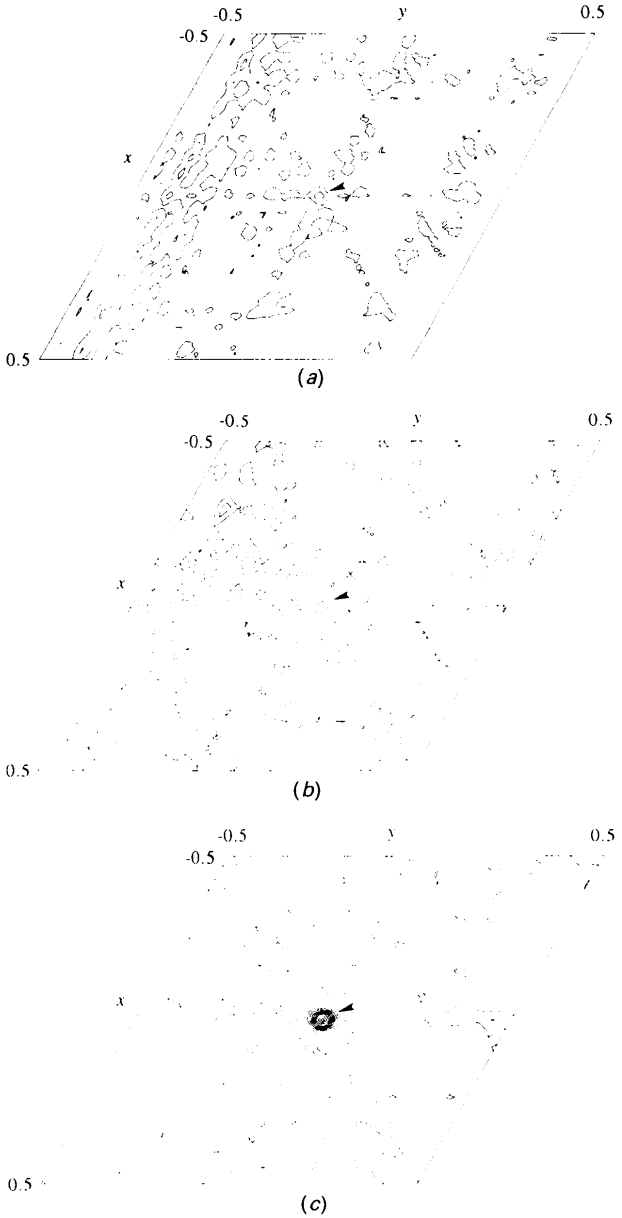

Fig. 2. Illustration of the use of addition strategy (20) in the translation function. The tests are based on T4 lysozyme using as a search model the fragment consisting of residues 10–30 (Table 2). The desired translation peak is expected at the origin which is at the center of the map section shown. The first contour is drawn at 2σ and subsequent contours are at increments of σ. (*a*) Search model consists of residues 10–30 alone. The peak at the desired position (arrowhead) is the 21st highest in the whole map. (*b*) Same search model (residues 10–30) but incorporating knowledge of the position of residues 1–9. The peak at the desired position (arrowhead) is now the highest in the whole map. (*c*) Same search model (residues 10–30) but assuming that the rest of the lysozyme structure (residues 1–9 plus 31–162) is known. The peak at the desired position (arrowhead) is obvious.

In summary, by the use of the 'subtraction' method the signal-to-noise ratio in the subsequent cross-rotation function searches was improved and the errors of the derived rotation angles were also reduced. Experience suggests that an improvement of even a couple of degrees in the orientation inferred from a rotation function can significantly improve the subsequent translation function search.

The example used to test the T_H and T'_H functions is based on the crystal structure of the methionine aminopeptidase from *E. coli* (Roderick & Matthews, 1993). The crystal has space group $P2_1$, with cell parameters a = 39.0, b = 61.7, c = 54.5 Å and β = 107.3° and one molecule per asymmetric unit. The structure was solved by multiple isomorphous replacement and refined to an R factor of 18.2% at 2.4 Å resolution. The search model used for the translation function T_H was that of the refined model which contains 261 amino-acid residues. The heavy-atom compound was a Pb derivative with two binding sites per asymmetric unit. The average difference between the observed structure amplitudes for

Table 3. *Molecular-replacement determination of the structure of mutant lysozyme 9001A*

See text for an explanation of the various steps. (16) and (20) were used for the rotation and correlation translation-function searches, respectively. In all cases except step 10′ the search model(s) corresponds to mutant T4 lysozyme $I3P_A$. The 'rotational error of solution' is the error determined following final refinement of the crystal structure. Unless stated otherwise, data between 4.5 and 9 Å resolution were used in all calculations, and a 20 Å integral radius was used for the rotation-function search.

	Function	Molecule sought in rotation or translation function	Molecules included in the calculation	Ranking of peak height	Magnitude of peak closest to correct solution $(p-a)/\sigma$	Rotational error of solution (°)	Magnitude of highest noise peak $(p-a)/\sigma$
Step 1	Self-rotation	—	—	1	58% of origin	—	—
Step 2	Rotation	*A*		1	4.9	2.4	3.9
		E		2	4.2	4.8	
		B		3	4.0	4.8	
		D		5	3.8	9.3	
		C		7	3.6	2.9	
Step 3	Translation	*A*		1	3.4	—	2.8
Step 4	Rotation	*B*	*A*	1	4.1	1.3	3.6
		E		2	4.1	4.5	
		D		3	3.8	11.0	
		C		4	3.6	3.7	
Step 5	Translation	*B*	*A*	1	8.2	—	6.2
Step 6	Rotation	*C*	*A* + *B*	1	4.2	3.3	4.1
		E		4	4.0	4.7	
		D		5	3.8	10.3	
		D′		9	3.6	6.5	
Step 7	Translation	*C*	*A* + *B*	1	8.1	—	4.7
Step 8	Rotation	*E*	*A* + *B* + *C*	2	4.0	3.8	4.3
		E′		5	3.7	2.5	
		D		—	—	—	
Step 9	Translation	*E*	*A* + *B* + *C*	1	11.0	—	5.3
Step 10	Rotation	*D*	*A* + *B* + *C* + *E*	—	—	—	—
Step 10′	Rotation	*D*	*A* + *B* + *C* + *E*	2	4.3	6.0	4.4
Step 11	Translation	*D*	*A* + *B* + *C* + *E*	1	8.6	—	4.7

Table 4. *Comparison of T_H, T'_H and C_t functions*

The T_H, T'_H and C_t functions were calculated using (30), (31) and (10), respectively. Two independent heavy-atom sites in a lead derivative of methionine aminopeptidase (Roderick & Matthews, 1993) were monitored for the T_H and T'_H functions. Data from 4 to 10 Å resolution were included. The search model incorporated successively smaller fractions of the complete protein, starting in all cases from the amino terminus. In the case of the T'_H function it is expected that the inclusion of observed structure amplitudes $|F_p|$ that are small may introduce errors. Therefore, the calculation was carried out two ways, (*a*) including all $|F_p|$, and (*b*) including only the strongest 60% of the amplitudes, $|F_p|$. Deletion of the weakest 40% of the data greatly enhances the T'_H function and makes it almost comparable with T_H. Both T'_H and T_H are superior to the conventional translation function C_t.

	T_H			T'_H						C_t		
Fraction of protein used as search model (%)	Magnitude of peak closest to correct solution $(p-a)/\sigma$	Ranking of peak	Highest noise peak $(p-a)/\sigma$	Magnitude of peak closest to correct solution $(p-a)/\sigma$ (*a*)	(*b*)	Ranking of peak (*a*)	(*b*)	Highest noise peak $(p-a)/\sigma$ (*a*)	(*b*)	Magnitude of peak closest to correct solution $(p-a)/\sigma$	Ranking of peak	Highest noise peak $(p-a)/\sigma$
100	13.5	1	4.9	8.7	14.8	1	1	3.8	5.0	9.8	1	2.3
80	10.3	1	4.4	5.2	11.8	1	1	4.5	4.8	7.3	1	3.1
60	9.1	1	4.2	5.7	9.4	1	1	4.4	4.6	4.3	1	2.8
40	7.4	1	3.7	5.0	8.0	1	1	4.1	4.1	2.1	13	2.5
20	4.5	1	3.6	3.3	5.8	25	1	4.3	4.5	1.8	11	3.2

the native and Pb data is 13.7% and the Cullis *R* value for centric reflections 62% (Roderick & Matthews, 1993). Table 4 shows a comparison of the relative effectiveness of the heavy-atom translation function, T_H (26) and T'_H (27) as well as the ordinary translation function (10). When the whole protein molecule is used as the search model all methods give a very clear-cut result. As the search molecule reduces to successively smaller fractions of the asymmetric unit, however, both the T_H and T'_H functions become much superior to the normal translation function.

Concluding remarks

A variety of different examples indicate that the quality of rotation and translation functions can be improved by

the use of orientation information, or orientation plus positional information, from a part of a crystal structure that is already known. In addition to the examples given, a number of mutants of T4 lysozyme, non-isomorphous with wild-type, and having two or more molecules in the asymmetric unit, have been solved (Heinz, Baase, Dahlquist & Matthews, 1993; Blaber, Zhang & Matthews, 1993; X.-J. Zhang, M. Blaber & D. W. Heinz, unpublished results). In some cases the correct solutions of rotation and translation functions that were otherwise obscure became apparent. In other cases signal to noise was improved.

Among a number of possible approaches, the rotation function with subtraction strategy and the correlation translation function with addition strategy were found to be most successful. The reason for the former is that the ordinary rotation function includes noise arising from unwanted correlations between different parts of the crystal structure. Subtraction of a part of a structure that is known will delete peaks and noise due to this part, allowing the remainder to become more significant. The reason for the success of the addition strategy in the translation function is that the total model becomes more complete which in turn increases the resemblance of the Patterson function $P_c(\mathbf{u},\mathbf{t})$ to $P_o(\mathbf{u})$.

Suppose one has a crystal with multiple copies of a protein in the asymmetric unit. With the availability of the modified rotation and translation functions, one suggested procedure for structure determination is as follows. An ordinary rotation function is used first to detect (some of) the molecular orientations, at least roughly. This is followed by a rotation function incorporating subtraction strategy to refine the solution(s). The second step can be repeated if and when additional information becomes available. Knowledge of the orientations and/or the positions of different parts of the structure can then be used in translation function searches. In this way the known structural information can be used to help determine or refine both the orientational and translational parameters of individual fragments of the overall structure.

Discussions with Drs S. Jim Remington and Melinda Dixon are greatly appreciated. We also thank Drs Dale Tronrud and Robert DuBose for their help in programming and Drs Steven Roderick, Dirk Heinz and Michael Blaber for helpful discussions and for providing unpublished data to test the various algorithms. This work was supported in part by grants from the NIH (GM21967; GM20066) and the Lucille P. Markey Charitable Trust.

APPENDIX

A modified algorithm to simultaneously search for symmetry-related molecules in a rotation function

Some attempts have been made to use the information from crystal symmetry to improve the potential of rotation functions. A simultaneous search for symmetry-related molecules in a rotation function was proposed by Nordman (1986). The rotation function, $N(\mathbf{R})$, proposed by Nordman is,

$$N(\mathbf{R}) = \langle P_o|\Sigma_j^{\text{sym}}\mathbf{S}_j\mathbf{R}P_c\rangle \times [\langle P_o|P_o\rangle\langle\Sigma_j^{\text{sym}}\mathbf{S}_j\mathbf{R}P_c|\Sigma_j^{\text{sym}}\mathbf{S}_j\mathbf{R}P_c\rangle]^{-1/2}. \quad (A1)$$

It has been reported (Yeates, 1989) not to be particularly effective because the information in an observed Patterson map is redundant. Only the information from one asymmetric unit is required to determine the orientation of the search model. The simultaneous search of symmetry-related molecules will in general increase both the signal peaks and the noise peaks, leaving the ratio approximately the same. The numerator of (A1) can be written as,

$$\langle P_o|\Sigma_j^{\text{sym}}\mathbf{S}_j\mathbf{R}P_c\rangle = \Sigma_j^{\text{sym}}\langle\mathbf{S}_j^{-1}P_o|\mathbf{R}P_c\rangle = n_{\text{sym}}\langle P_o|\mathbf{R}P_c\rangle, \quad (A2)$$

which is simply the ordinary rotation function (5) multiplied by a constant. Usually, the denominator of (A1) is not sensitive to rotation $\mathbf{R}$. In one special situation, however, Nordman's approach may be helpful in reducing the so-called symmetry bias in the rotation function. This happens when the search model has local symmetry that is similar to the symmetry of the crystal and the search model is oriented so that the two symmetry operators coincide. In this case, the denominator becomes significantly larger than usual.

Yeates (1989) rewrites the Nordman's rotation function $N(\mathbf{R})$ in a form similar to the following

$$N(\mathbf{R}) = cC_r(\mathbf{R})/Q(\mathbf{R}) \quad (A3)$$

where c is a constant, $C_r(\mathbf{R})$ is the ordinary rotation function, and $Q(\mathbf{R})$ is a sum of self-rotation functions of the Patterson function of the search model.

$$Q(\mathbf{R}) = (1/n_{\text{sym}})\Sigma_j^{\text{sym}}\langle\mathbf{S}_j\mathbf{R}P_c|\Sigma_{j'}^{\text{sym}}\mathbf{S}_{j'}\mathbf{R}P_c\rangle = \langle P_c|\mathbf{R}^{-1}(\Sigma_j^{\text{sym}}\mathbf{S}_j)\mathbf{R}P_c\rangle \quad (A4)$$

where n_{sym} is the number of symmetry operators. Yeates proposed an interpolation algorithm to calculate this function. It requires calculating $C_r(\mathbf{R})$ and the self-rotation function of $\langle P_c|\mathbf{R}P_c\rangle$ separately, followed by n_{sym} interpolations at every sampling point. In the following, an alternative approach is described which is conceptually simpler. Although no direct comparison is available the following approach also seems easier to implement.

In the Eulerian angular system, a rotation function can be written (Kabsch, 1986) as,

$$C_{r,\text{Crowther}}(\alpha,\beta,\gamma) = \Sigma_{l,m,m'} C_{l,m,m'} R_{l,m',m}(\alpha,\beta,\gamma) \quad (A5)$$

where the $C_{l,m,m'}$ and $R_{l,m',m}$ are the coefficients corresponding to the Patterson functions and the rotation operator **R**, respectively. The coefficients are calculated with spherical harmonics and spherical Bessel functions. Since a Patterson function has a center of symmetry, $C_{l,m,m'}$ in (*A*5) vanishes whenever l is odd. Mathematically, the summation in (*A*5) is a matrix contraction.

By definition, in the Eulerian angular system a rotation **R** specified with angles (α,β,γ) can be divided into three consecutive rotations about the y and z axes.

$$\mathbf{R}(\alpha,\beta,\gamma) = R_z(\alpha)R_y(\beta)R_z(\gamma). \qquad (A6)$$

The corresponding spherical harmonics coefficients (Crowther, 1972) can be written as,

$$R_{l,m',m}(\alpha,\beta,\gamma) = \exp(im'\gamma)d_{l,m',m}(\beta)\exp(im\alpha). \qquad (A7)$$

With spherical harmonics, a series of consecutive rotation operations can be represented as a matrix multiplication. Therefore, the $Q(\mathbf{R})$ function in (*A*4) can be written as,

$$\begin{aligned} Q(\alpha,\beta,\gamma) &= \langle P_c|\mathbf{R}(-\gamma,-\beta,-\alpha)(\Sigma_j\mathbf{S}_j)\mathbf{R}(\alpha,\beta,\gamma)P_c\rangle \\ &= \Sigma_{l,m,m',n,n'}C'_{l,m,m'}R_{l,m',n}(-\gamma,-\beta,-\alpha) \\ &\quad \times S_{l,n,n'}R_{l,n',m}(\alpha,\beta,\gamma), \end{aligned} \qquad (A8)$$

where $C'_{l,m,m'}$ are the coefficients corresponding to the Patterson function of the search model and $S_{l,n,n'}$ is the coefficient associated with the crystal symmetry operators $\{R(\alpha_{sj},\beta_{sj},\gamma_{sj}),\ j = 1,\ n_{\rm sym}\}$.

$$S_{l,n,n'} = \Sigma_j^{\rm sym}R_{l,n,n'}(\alpha_{sj},\beta_{sj},\gamma_{sj}). \qquad (A9)$$

An implementation of this algorithm is available as an option in our molecular-replacement program package.

References

Bell, J. A., Wilson, K. P., Zhang, X.-J., Faber, H. R., Nicholson, H. & Matthews, B. W. (1991). *Proteins Struct. Funct. Genet.* **10**, 10–21.

Bentley, G. A. & Houdusse, A. (1992). *Acta Cryst.* A**48**, 312–322.

Bi, R.-C., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Giordano, F., Reynolds, C. D. & Tolley, S. P. (1983). *Acta Cryst.* B**39**, 90–98.

Blaber, M., Zhang, X.-J. & Matthews, B. W. (1993). *Science*, **260**, 1637–1640.

Bode, W., Chen, Z., Bartels, K., Kutzbach, C., Schmidt-Kastner, G. & Bartunik, H. (1983). *J. Mol. Biol.* **164**, 237–282.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.

Crowther, R. A. (1972). In *The Molecular Replacement Method*, edited by M. G. Rossmann. New York: Gordon and Breach.

Crowther, R. A. & Blow, D. M. (1967). *Acta Cryst.* **23**, 544–548.

Cygler, M. & Anderson, W. F. (1988). *Acta Cryst.* A**44**, 300–308.

Dixon, M. M., Nicholson, H., Shewchuk, L., Baase, W. A. & Matthews, B. W. (1992). *J. Mol. Biol.* **227**, 917–933.

Driessen, H., Bax, B., Slingsby, C., Lindley, P. F., Mahadevan, D., Moss, D. S. & Tickle, I. J. (1991). *Acta Cryst.* B**47**, 987–997.

Driessen, H. & White, H. (1985). In *Molecular Replacement, Proceedings of the Daresbury Study Weekend, February* 15–16, edited by P. A. Machin. Warrington: SERC Daresbury Laboratory.

Faber, H. R. & Matthews, B. W. (1990). *Nature (London)*, **348**, 263–266.

Fitzgerald, P. M. D. (1988). *Acta Cryst.* A**44**, 273–278.

Fitzgerald, P. M. D. (1991). *Crystallographic Computing* 5: *From Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 333-347. Oxford Univ. Press.

Fujinaga, M. & Read, R. J. (1987). *J. Appl. Cryst.* **20**, 517–521.

Harada, Y., Lifchitz, A. & Beathou, J. (1981). *Acta Cryst.* A**37**, 398–406

Heinz, D. W., Baase, W. A., Dahlquist, F. W. & Matthews, B. W. (1993). *Nature (London)*, **362**, 561–564.

Hoppe, W. (1957). *Elektrochem.* **61**, 1076–1079.

Huber, R. (1965). *Acta Cryst.* **19**, 353–356.

Kabsch, W. (1986). *Rotation Function* program.

Machin, P. A. (1985). Editor. *Molecular Replacement, Proceedings of the Daresbury Study Weekend, February* 15–16. Warrington: SERC Daresbury Laboratory.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Matthews, B. W. & Remington, S. J. (1974). *Proc. Natl Acad. Sci. USA*, **71**, 4178–4182.

Nixon, P. E. & North, A. C. T. (1976). *Acta Cryst.* A**32**, 320–325.

Nordman, C. E. (1986). Proc. Am. Crystallogr. Assoc. Annu. Meet., Hamilton, Ontario, Abstract P36.

Read, R. J. & Schierbeek, A. J. (1988). *J. Appl. Cryst.* **21**, 490–495.

Reynolds, R. A., Remington, S. J., Weaver, L. H., Fisher, R. G., Anderson, W. F., Ammon, H. C. & Matthews, B. W. (1985). *Acta Cryst.* B**41**, 139–147.

Roderick, S. L. & Matthews, B. W. (1993). *Biochemistry*, **32**, 3907–3912.

Rossmann, M. G. (1972). Editor. *The Molecular Replacement Method.* New York: Gordon and Breach.

Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.

Rossmann, M. G., Blow, D. M., Harding, M. M. & Coller, E. (1964). *Acta Cryst.* **17**, 338–342.

Schierbeek, A. J., Renetseder, R., Dijkstra, B. W. & Hol, W. G. J. (1985). *Molecular Replacement, Proceedings of the Daresbury Study Weekend, February* 15–16, edited by P. A. Machin. Warrington: SERC Daresbury Laboratory.

Sheriff, S., Padlan, E. A., Cohen, G. H. & Davies, D. R. (1990). *Acta Cryst.* B**46**, 418–425.

Tanaka, N. (1977). *Acta Cryst.* A**33**, 191–193.

Tong, L. & Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 783–792.

Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* A**43**, 489–503.

Weaver, L. H. & Matthews, B. W. (1987). *J. Mol. Biol.* **193**, 189–199.

Yeates, T. O. (1989). *Acta Cryst.* A**45**, 309–314.

Zhang, X.-J., Baase, W. A. & Matthews, B. W. (1992). *Protein Sci.* **1**, 761–776.