Table 1. *Minimal balance surfaces, built up from catenoids with two, three or four spout-like attachments*

Minimal balance surface	Group-subgroup pair	Nets of twofold axes	Genus	Symmetry of surface patches	
				Infinite	Finite
$C(H)$	$P6_3/mmc$-$P\bar{6}m2$	6^3	7	$P(\bar{6})m2$	$\bar{6}m2$
$C(R2)$	$I4/mcm$-$P4/mbm$	48^2	25	$P(4/m)bm$	$m.m2$
$C(R3)$	$P6/mcc$-$P6/m$	46.12	37	$P(6/m)11$	$m..$
$tC(P)$	$I4/mmm$-$P4/mmm$	4^4	9	$P(4/m)mm$	$4/mmm$
$oC(P)$	$Fmmm$-$Cmmm$	4^4	9	$Cmm(m)$	mmm
PT	$Fmmm$-$Cmmm$	4^4	5	$Cm(mm)$	mmm

The existence of $C(H)$, $tC(P)$ and PT surfaces can be proved by soap-film experiments. For $C(R2)$, $C(R3)$ and $oC(P)$ surfaces such experiments are impossible because of the absence of mirror planes that bound the finite surface patches. Probably $C(H)$, $C(R2)$, $C(R3)$ and $tC(P)$ surfaces exist only within a certain range of axial ratios $(c/a)_{\min} \le c/a \le (c/a)_{\max}$. The soap-film experiments suggest that $(c/a)_{\max}$ is larger for $C(H)$ and $tC(P)$ surfaces than for H and tP surfaces, *i.e.* the handles connecting the catenoids stabilize the minimal surfaces for large c/a values. For orthorhombic surfaces the ratios b/a and c/a must be examined. b/a describes the shape of the rectangles and c/a the distance between the nets. In the case of the PT surfaces the soap-film experiment shows that handles in the a direction are stable only for $b/a \ge (b/a)_{\min} > 1$. PT surfaces, therefore, are incompatible with square nets.

References

BOHM, J. & DORNBERGER-SCHIFF, K. (1967). *Acta Cryst.* **23**, 913-933.

FISCHER, W. & KOCH, E. (1987). *Z. Kristallogr.* **179**, 31-52.

FISCHER, W. & KOCH, E. (1989*a*). *Acta Cryst.* A**45**, 166-169.

FISCHER, W. & KOCH, E. (1989*b*). *Acta Cryst.* A**45**, 485-490.

KOCH, E. & FISCHER, W. (1988). *Z. Kristallogr.* **183**, 129-152.

KOCH, E. & FISCHER, W. (1989). *Acta Cryst.* A**45**, 169-174.

NEOVIUS, E. R. (1883). *Bestimmung zweier spezieller periodischer Minimalflächen.* Akad. Abhandlungen, Helsingfors.

SCHOEN, A. H. (1970). *Infinite Periodic Minimal Surfaces Without Self-intersections.* NASA Tech. Note No. D-5541.

SCHWARZ, H. A. (1890). *Gesammelte mathematische Abhandlungen*, Vol. 1. Berlin: Springer.

Acta Cryst. (1989). A**45**, 563-572

A Full-Symmetry Translation Function Based on Electron Density*

BY MIROSLAW CYGLER† AND MARC DESROCHERS

Biotechnology Research Institute, National Research Council of Canada, Montréal, Québec H4P 2R2, Canada

(*Received* 21 *September* 1988; *accepted* 21 *March* 1989)

Abstract

A method for positioning an oriented fragment within the unit cell is presented. It is based on a correlation between a model and observed data which is performed in Fourier rather than Patterson space. Symmetry-related molecules are located in the electron density map calculated in space group $P1$, with the phases derived from a model that is correctly oriented but arbitrarily positioned in the unit cell. It is shown that considering all symmetry elements simultaneously substantially increases the sensitivity of the method and makes it less susceptible to the errors in the model. The procedure also automatically incorporates a penalty for the overlap of symmetry-related molecules, and the stringency of this requirement is easily modified. The method has been tested on two different proteins and the results compare favorably with other translation functions.

Introduction

Analysis of the architecture of proteins with known 3D structures (*e.g.* Rossmann & Argos, 1976; Richardson, 1977, 1981; Chotia, 1984; Janin & Chotia, 1980; Chotia, Levitt & Richardson, 1981) indicates that their folding pattern is conserved to a much higher degree than their amino-acid sequence, and suggests that the number of different structural motifs (patterns of folding units of a polypeptide chain) in globular proteins is relatively limited. Knowing the amino-acid sequence of a particular protein, one can obtain some information about its structure from a possible sequence homology to other proteins with known structures. This knowledge can be utilized either for building a model of the protein (*e.g.* Blundell,

* Issued as NRCC Publication No. 30293.

† To whom all correspondence should be addressed.

0108-7673/89/080563-10$03.00

Sibanda, Sternberg & Thornton, 1987; Greer, 1985, 1988) or, when crystals of this protein are available, for assisting in solving its structure by the application of the molecular replacement (MR) method (Rossmann, 1972, and references cited therein).

The ultimate success of the MR method depends on how well the chosen model corresponds to all or part of the unknown protein molecule. The usefulness of this methodology for determining the structure of new proteins would be much greater if formulations less sensitive to the distortions of the model and/or errors in its orientation could be found. A very desirable property of the MR method, from the practical point of view, is the ability to locate models that comprise only a fraction of the unknown molecule (Cygler & Anderson, 1988*a*).

The orientation of the model is determined by correlating the Patterson function of the unknown crystal with the Patterson function of the model molecule. These calculations are performed either in reciprocal space (Rossmann & Blow, 1962) or in direct space (Huber, 1965) but they are variations of the same mathematical formulation of the problem. The computationally least intensive and most widely used formalism is due to Crowther (1972).

A solution of the translation problem depends not only upon the degree of correspondence between the chosen model and the unknown molecule but also on the precision with which its orientation has been determined. In contrast to the rotation problem, many different formulations have been proposed for the solution of the translation problem [*e.g.* see recent review by Beurskens, Gould, Bruins Slot & Bosman (1987)]. Both real- and reciprocal-space algorithms have been developed. All these algorithms, with the exception of *TRADIR* (Doesburg & Beurskens, 1983), operate in Patterson space.

If some heavy-atom derivatives are available, it is possible to combine information from this source with the knowledge of the orientation of the model molecule to facilitate determination of the translation of the model. We have previously shown (Cygler & Anderson, 1988*b*) how the knowledge of the orientation of the model can be used to find simultaneously the positions of heavy atoms and the translation of the molecule, by expanding the space group to $P1$ and exploring Fourier space. A way of solving this problem *via* reciprocal-space calculations was reported by Read & Schierbeek (1988; see references cited therein for earlier attempts at addressing this problem). In this case, one needs to have a set of multiple-isomorphous-replacement (MIR) or single-isomorphous-replacement (SIR) phases in order to apply the procedure, and the method facilitates the recognition of the model in otherwise difficult or impossible to interpret MIR maps.

A recently proposed approach (Cygler & Anderson, 1988*b*) to solve the translation problem for macromolecules is conceptually similar to the one described by Doesburg & Beurskens (1983), with some important differences. The basic idea is to explore Fourier rather than Patterson space. It follows the practice sometimes used in small-molecule crystallography, when a correctly oriented but misplaced partial structure can be found (*e.g.* Karle & Karle, 1971). The space-group symmetry of the crystal is artificially reduced to $P1$, and the electron density (e.d.) map is calculated with phases derived from a model that is correctly oriented but arbitrarily positioned within the unit cell (the 'phasing' molecule). Recognition of a symmetry-related molecule, however distorted, allows for positioning of the symmetry element. In the case of an e.d. map of a protein, recognition of a symmetry-related molecule by visual inspection of a poor-quality map is rather unlikely, and a more systematic approach was developed. A symmetry-related molecule is translated step by step throughout the unit cell, and at each location of the model, a measure of the overlap - the S function - between the positive features of the e.d. map and atomic positions is calculated. The best fit represents the position of a symmetry-related molecule relative to the phasing molecule. From the positions of these two molecules one can determine the position of the symmetry element, which is equivalent to the location of two coordinates of the origin (in the plane perpendicular to the symmetry axis). The coordinate along the symmetry axis remains undetermined.

This approach has been tested for the space group $P2_1$ in the case of the HED10 Fab fragment (Cygler & Anderson, 1988*b*). It was shown that using as little as 10–12% of the unit-cell contents for phasing was sufficient to solve the translation problem, but only when the model was very good and its orientation precisely known. It was also clear that, when 25% of the unit-cell contents was included, this approach was less sensitive to misorientation errors than Patterson-space-based functions [the Crowther–Blow translation function (Crowther & Blow, 1967); F_{obs}/F_{calc} correlation coefficient search, *BRUTE* (Fujinaga & Read, 1987)].

In this paper, we extend the S-function approach to higher-symmetry space groups, and derive the relation of the S function to the correlation coefficient between the 'experimental' and full-symmetry model electron density. We show that the resulting decrease in the percentage of the unit-cell contents used for phasing can be counterbalanced by correlation results for all symmetry elements.

S translation function

In the following discussion it is assumed that the atoms of the model molecule, which has correct orientation but is arbitrarily located in the unit cell, are at positions $\mathbf{r}_i$, for $i = 1, n$. The electron density

corresponding to this molecule, $\rho_m(\mathbf{r})$, vanishes outside the volume U (molecular envelope).

Let $\rho(\mathbf{r})$ be the electron density calculated in space group $P1$ with the observed structure amplitudes (expanded to $P1$ space-group symmetry) and the phases derived from the model molecule.

Let $\mathcal{T}_j = \{\mathbf{A}_j, \mathbf{t}_j\}$, for $j = 1, \ldots, N$, be the symmetry elements of the crystal ($\mathcal{T}_1$ being identity), where $\mathbf{A}_j$ is the point-symmetry transformation and $\mathbf{t}_j$ is the translation component. For non-centrosymmetric space groups (which are the only ones we are concerned with here), $\mathbf{A}_j$ is the rotation matrix.

Let us assume that $\mathbf{O}_v$ is the position of the origin for applying symmetry transformations (in the correct space group) and ${}_v\mathbf{r}^j$ is the transformation of vector $\mathbf{r}$ by the jth symmetry element relative to this origin.

It can be seen from Fig. 1 that

$$\begin{aligned} {}_v\mathbf{r}^j &= {}_v\mathbf{q}^j + \mathbf{v} \\ &= \mathbf{A}_j\mathbf{q} + \mathbf{t}_j + \mathbf{v} \\ &= \mathbf{A}_j(\mathbf{r} - \mathbf{v}) + \mathbf{t}_j + \mathbf{v} \\ &= \mathbf{A}_j\mathbf{r} + \mathbf{t}_j + \mathbf{v} - \mathbf{A}_j\mathbf{v}. \end{aligned} \tag{1}$$

The model electron density, ρ^M, that includes all symmetry-related molecules is then

$$\rho^M(\mathbf{r}, \mathbf{v}) = \sum_{j=1}^{N} \rho_m^j(\mathbf{r}) \tag{2}$$

where

$$\rho_m^j({}_v\mathbf{r}^j) = \rho_m(\mathbf{r}) \qquad (\rho_m^1 \equiv \rho_m) \tag{3}$$

$$U_j = \mathcal{T}_j\{U\} \tag{4}$$

and $\rho_m^j = 0$ outside the volume U_j.

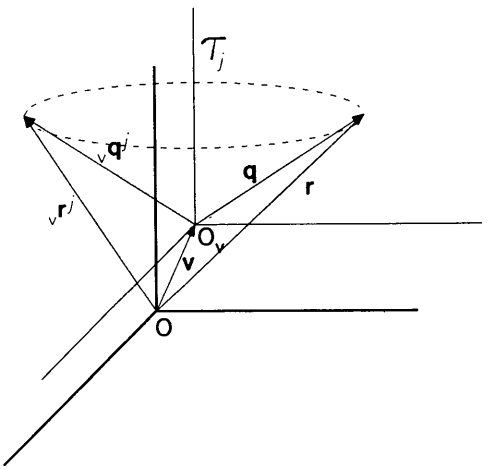

Fig. 1. Relationship between coordinates in two reference systems. Thick lines represent an arbitrary coordinate system with origin at $\mathbf{O}$ relative to which the e.d. map was calculated. Thin lines correspond to a coordinate system with origin at $\mathbf{O}_v$, relative to which the symmetries are applied. Vector $\mathbf{q}$ is transformed by $\mathcal{T}_j$ into ${}_v\mathbf{q}^j$.

Let us now define the correlation coefficient between the observed and full-symmetry model electron density over the unit-cell volume V:

$$C(\mathbf{v}) = \frac{\int_V [\rho(\mathbf{r}) - \bar{\rho}][\rho^M(\mathbf{r}, \mathbf{v}) - \bar{\rho}^M]\, d\mathbf{r}}{\{\int_V [\rho(\mathbf{r}) - \bar{\rho}]^2\, d\mathbf{r}\}^{1/2} \{\int_V [\rho^M(\mathbf{r}, \mathbf{v}) - \bar{\rho}^M]^2\, d\mathbf{r}\}^{1/2}}. \tag{5}$$

If $F(000)$ terms are not included in the calculations of electron density, then $\bar{\rho}, \bar{\rho}^M = 0$.

We expect that the function $C(\mathbf{v})$ will reach a global maximum for a vector $\mathbf{v}$ corresponding to the position of the origin $\mathbf{O}_v$ in the proper space group.

The electron density ρ can be separated into two terms,

$$\rho = \rho_o + \Delta\rho \tag{6}$$

where

$$\rho_o(\mathbf{r}) = \begin{cases} \rho(\mathbf{r}) & \text{for } \mathbf{r} \in U \\ 0 & \text{for } \mathbf{r} \in V - U. \end{cases} \tag{7}$$

Because the phases used to calculate ρ are derived from the model (ρ_m), the above partition separates large e.d. (ρ_0), corresponding to the image of the phasing molecule, from low e.d. ($\Delta\rho$) containing information about symmetry-related molecules.

In further formulas the variable $\mathbf{r}$ will be omitted from the functions for brevity. The numerator of (5) can be rewritten as

$$\begin{aligned} \int_V \rho\rho^M\, d\mathbf{r} &= \int_U \rho_o\rho_m\, d\mathbf{r} + \int_U \Delta\rho\rho_m\, d\mathbf{r} \\ &\quad + \sum_{j>1}^{N} \int_U \rho_o\rho_m^j\, d\mathbf{r} + \sum_{j>1}^{N} \int_{U_j} \Delta\rho\rho_m^j\, d\mathbf{r}. \end{aligned} \tag{8}$$

Since the first term on the right-hand side does not depend on the position of the origin $\mathbf{O}_v$ and the second term is equal to 0, one obtains

$$\int_V \rho\rho^M\, d\mathbf{r} = K_1 + \sum_{j>1}^{N} \int_U \rho_o\rho_m^j\, d\mathbf{r} + \sum_{j>1}^{N} \int_{U_j} \Delta\rho\rho_m^j\, d\mathbf{r}. \tag{9}$$

Accordingly, the first integral in the denominator of (5) does not depend on the position of the origin for the symmetry elements and represents a constant scale factor K_2. The second integral becomes

$$\begin{aligned} &\int_V (\rho^M)^2\, d\mathbf{r} + N \int_U (\rho_m)^2\, d\mathbf{r} + 2 \sum_{i<j}^{N} \int_V \rho_m^i\rho_m^j\, d\mathbf{r} \\ &\quad = K_3 + 2 \sum_{i<j}^{N} \int_V \rho_m^i\rho_m^j\, d\mathbf{r}. \end{aligned} \tag{10}$$

Equation (5) then takes the form

$$C(\mathbf{v}) = \frac{K_1 + \sum_{j>1}^{N} \int_{U_j} \Delta\rho\rho_m^j\, d\mathbf{r} + \sum_{j>1}^{N} \int_U \rho_o\rho_m^j\, d\mathbf{r}}{K_2\left(K_3 + 2 \sum_{i<j}^{N} \int_V \rho_m^i\rho_m^j\, d\mathbf{r}\right)^{1/2}}. \tag{11}$$

It is easy to show that, in the case when there is no overlap between the symmetry-related molecules and the phasing molecule, the last integrals in the numerator and the denominator vanish and (11) reduces to

$$C(\mathbf{v}) = K_4 + K_5 \sum_{j>1}^{N} \int_{U_j} \Delta\rho\rho_m^j \, \mathrm{d}\mathbf{r}. \quad (12)$$

Let us define the function S as

$$S(\mathbf{v}) = \sum_{j>1}^{N} \int_{V} \Delta\rho(\mathbf{r})\rho_m^j(\mathbf{r}) \, \mathrm{d}\mathbf{r}. \quad (13)$$

Since, for the correct position of the origin $\mathbf{O}_v$, the overlap between the symmetry-related molecules is minimal, we can examine maxima of the S rather than the C function, provided that the no-overlap condition is met. Use of the S function has significant computational advantages over the C function as given by (5) or (11). We show below that the calculations of the S function can be partitioned into two-dimensional projections, decreasing the CPU time by 1–2 orders of magnitude in comparison with calculations for the C function.

A simple way to ensure that the maxima of the S function correspond to a minimal overlap between symmetry-related molecules is to include an additional term acting as an overlap penalty.

$$S(\mathbf{v}) = \sum_{j>1}^{N} \int_{U_j} \Delta\rho\rho_m^j \, \mathrm{d}\mathbf{r} + \sum_{j>1}^{N} \int_{V} \mu\rho_m^j \, \mathrm{d}\mathbf{r} \quad (14)$$

where

$$\mu(\mathbf{r}) = \begin{cases} \leq 0 & \text{for } \mathbf{r} \in U \\ = 0 & \text{for } \mathbf{r} \in V - U. \end{cases} \quad (15)$$

Maxima of the function given by (13) for which there is a significant overlap of symmetry-related molecules are eliminated from the modified S function in (14) by an appropriate choice of the penalty function.

Equation (14) can be rewritten as

$$S(\mathbf{v}) = \sum_{j>1}^{N} S_j(\mathbf{v}) \quad (16)$$

where

$$S_j(\mathbf{v}) = \int_{U_j} [\Delta\rho(\mathbf{r}) + \mu(\mathbf{r})]\rho_m^j(\mathbf{r}) \, \mathrm{d}\mathbf{r}. \quad (17)$$

Let

$$\rho^{\mathrm{mod}}(\mathbf{r}) = \Delta\rho(\mathbf{r}) + \mu(\mathbf{r}). \quad (18)$$

It can be easily shown that (17) becomes

$$S_j(\mathbf{v}) = \int_{U} \rho^{\mathrm{mod}}({}_v\mathbf{r}^j)\rho_m(\mathbf{r}) \, \mathrm{d}\mathbf{r}. \quad (19)$$

The integral in (19) can be approximated by a sum over all the atoms in the model,

$$S_j(\mathbf{v}) \simeq \sum_{i=1}^{n} \rho^{\mathrm{mod}}({}_v\mathbf{r}_i^j)\rho_m(\mathbf{r}_i). \quad (20)$$

Vector $\mathbf{v}$ can be decomposed into components perpendicular ($\mathbf{v}_{\perp_j}$) and parallel ($\mathbf{v}_{\parallel_j}$) to the direction of the rotation axis of the jth symmetry element. Then, since $\mathbf{A}_j\mathbf{v}_{\parallel_j} = \mathbf{v}_{\parallel_j}$, combination of (1) and (20) gives

$$S_j(\mathbf{v}) = \sum_{i=1}^{n} \rho_m(\mathbf{r}_i)\rho^{\mathrm{mod}}[\mathbf{A}_j\mathbf{r}_i + \mathbf{t}_j + (\mathbf{I} - \mathbf{A}_j)\mathbf{v}_{\perp j}]$$
$$= S_j(\mathbf{v}_{\perp_j}), \quad (21)$$

where $\mathbf{I}$ is the identity matrix.

This indicates that each of the functions S_j needs to be calculated in only one plane, *i.e.* the calculations are two- rather than three-dimensional.

It must be stressed here that the decrease of the dimensionality of the problem follows from the fact that the position of one molecule (model) in the e.d. map is known.

Similar arguments have also been used to show that Patterson-based translation-function formulae, including all symmetry elements simultaneously, can be reduced to a summation of properly scaled contributions from each symmetry element at a time (Beurskens, Gould, Bruins Slot & Bosman, 1987).

The function $S(\mathbf{v})$ is obtained by summing S_j's at vectors corresponding to projections of $\mathbf{v}$ onto the planes perpendicular to the respective symmetry axes. Finally, since the value of $\rho_m(\mathbf{r}_i)$ is proportional to Z_i, the number of electrons of atom i, with a good approximation (neglecting a change in a scale factor) one can replace ρ_m by 1 to obtain

$$S(\mathbf{v}) = \sum_{j>1}^{N} \sum_{i=1}^{n} \rho^{\mathrm{mod}}[\mathbf{A}_j\mathbf{r}_i + \mathbf{t}_j + (\mathbf{I} - \mathbf{A}_j)\mathbf{v}_{\perp_j}]. \quad (22)$$

The S function can be evaluated in the real space *via* (22). An equivalent reciprocal-space formulation could be derived, following Doesburg & Beurskens (1983), in which structure factors corresponding to the modified map would have to be used. In order to avoid ripples due to limited resolution of the data, great care would have to be taken to ensure that ρ^{mod} is a smoothly varying function. It seemed to us that a direct-space approach gives us more flexibility in choosing the form of the penalty function and we followed that route for current applications.

The algorithm presented above has been coded in Fortran (program *RTRANS*) and tested for two different proteins. The e.d. map was calculated in the range 0–1 in all three directions, scaled to fall within the range −127 to 127 and stored as integer values. The map was modified around the phasing molecule as follows: each atom was centered on the closest grid point in the map and the density in a $3\times3\times3$ box was modified to either 0, if the point was on the surface of the protein, or a low negative value, if it was inside the protein. A value in the range −15 to −5 gave good results. Choice of a much larger nega-

tive value resulted in missing the correct solution (see *Discussion*). Next, for each symmetry element (except identity), a two-dimensional search was done and the values of $S_j(\mathbf{v}_{\perp_j})$ were stored. Finally, $S(\mathbf{v})$ was calculated according to (22).

Results

In order to test the above formalism and estimate its robustness against model errors and the size of the model, *RTRANS* was applied to two different proteins, both of which have been crystallized in orthorhombic space groups, with four molecules in the unit cell (one molecule in the asymmetric unit). In the first test, a serine protease, the model corresponded to the whole molecule, while in the second test, an Fab fragment of immunoglobulin IgG, the model comprised roughly half of the molecule.

Parallel calculations, using the same sets of data and the same models, were performed with *BRUTE* [Fujinaga & Read (1987), correlation coefficient between $|F_{obs}|^2$ and $|F_{calc}|^2$]. Fujinaga & Read (1987) compared this method with the Crowther–Blow (CB) translation function, with the $TO(r)/O(r)$ function of Harada, Lifchitz, Berthou & Jolles (1981), and with the systematic R-factor search, for three different proteins, and concluded that *BRUTE* performed comparably or slightly better than the others. We have included one of these proteins in our tests, in order to compare the *RTRANS* results with other methods.

RTRANS was implemented on the IBM-3090/200 while the correlation coefficient searches were run on an FPS-264 array processor with a version of *BRUTE* optimized for this machine. To compare the speed of these two computers we have compiled on both machines a version of *BRUTE* in which all FPS-specific library subroutines were substituted by their Fortran equivalents. Under these conditions the execution times were roughly the same. However, the optimized FPS version was faster by a factor of ~4. The timing information in Tables 1, 3 and 5 refer to the FPS optimized version.

In the previous applications of *BRUTE* (Fujinaga & Read, 1987; Read & James, 1988; Cygler & Anderson, 1988*b*) the data were limited to a relatively narrow shell of resolution. The reason for that was the limitation of the memory of the FPS-164 array processor and long execution times. Access to a faster array processor with a larger memory allowed us to increase the resolution shell and compare the two methods using the same data.

A full cycle of *RTRANS* calculations includes computing structure factors in space group $P1$, computing an e.d. map and computing the S function. The total CPU time for all these steps using 2·5 Å resolution data and all atoms of the model for searches is less than 4 min on an IBM-3090/200. This compares to 2–3 h of FPS-264 time for *BRUTE* calculations.

SGT protease

Streptomyces griseus trypsin (SGT; Read & James, 1988) has been crystallized in space group $C222_1$. Its cell dimensions are $a = 72{\cdot}3$, $b = 51{\cdot}0$, $c = 120{\cdot}1$ Å. The model used in structure determination was bovine trypsin (Chambers & Stroud, 1979; Chambers, Stroud & Finer-Moore, 1988). The orientation of the model was originally found (Read & James, 1988) by Crowther's (1972) fast rotation function algorithm. The translation was determined with the help of *BRUTE* using data in a 4–5 Å resolution shell (Fujinaga & Read, 1987).

To simulate various degrees of error in the model, the calculations were performed for four sets of orientation angles. The first set ($\alpha = 66{\cdot}3$, $\beta = -42{\cdot}6$, $\gamma = -82{\cdot}4°$) corresponded to the best orientation (Table 3, Read & James, 1988), the second one ($\alpha = 67{\cdot}0$, $\beta = -44{\cdot}1$, $\gamma = -86{\cdot}0°$) to a 3·5° error in the orientation, the third one ($\alpha = 72{\cdot}0$, $\beta = -46{\cdot}0$, $\gamma = -91{\cdot}0°$) to a 6·9° error in the orientation (Table 2, Read & James, 1988), and the fourth one ($\alpha = 71{\cdot}0$, $\beta = -47{\cdot}0$, $\gamma = -92{\cdot}0°$) to an 8·3° error in the orientation.

All 1629 atoms from the trypsin molecule (approximately 25% of unit-cell content) were used as a model. There was no attempt to remove parts of the chain where differences in the conformation occur.

Following Fujinaga & Read (1987), calculations with *BRUTE* were performed on a grid of $1{\cdot}0 \times 1{\cdot}0 \times 1{\cdot}0$ Å and were repeated with a diagonal offset of 0·5 Å in all directions. The influence of the resolution of the data on the peak-to-noise ratio was investigated. Table 1 shows the result for the above sets of angles.

For *RTRANS* calculations, the e.d. map was computed in superspace group $C1$ rather than $P1$, with the data from 15 Å to at most 2·5 Å resolution, on a grid of approximately $0{\cdot}8 \times 0{\cdot}8 \times 0{\cdot}8$ Å. Appropriate provisions were made for modifying the electron density within the region corresponding to the translationally related molecule. Table 2 summarizes the results.

As with *BRUTE*, the influence of the resolution of the e.d. map on the peak-to-noise ratio of the S function was investigated. Another parameter tested was the completeness of the model used for searches (as opposed to the phasing molecule). Three levels of completeness were tried: all atoms, backbone and C_β atoms, and C_α atoms. The influence on the results of the shape of the overlap penalty function μ [(15)] was also investigated.

*JEL*318 *Fab fragment*

JEL318 Fab fragment has been crystallized in space group $P2_12_12_1$ with cell dimensions $a = 82{\cdot}0$, $b = 139{\cdot}2$, $c = 41{\cdot}3$ Å (Boodhoo, Mol, Lee & Anderson, 1988). It has been solved by the MR method (Cygler,

Table 1. *Results of BRUTE for Streptomyces griseus trypsin*

Resolution (Å)	Number of reflections	Orientation angles α (°)	β (°)	γ (°)	Orientation error (°)	S/N*	Computational time† (min)
8-4·0	1734	66·3	−42·6	−82·4	0·0	1·65	27
15-3·0	4585	66·3	−42·6	−82·4	0·0	1·73	65
8-4·0	1734	67·0	−44·1	−86·0	3·5	1·61	26
15-3·0	4585	67·0	−44·1	−86·0	3·5	1·67	68
8-4·0	1734	72·0	−46·0	−91·0	6·9	0·91	27
15-4·0	1971	72·0	−46·6	−91·0	6·9	1·02	30
12-3·0	4542	72·0	−46·0	−91·0	6·9	0·98	68
15-3·0	4585	72·0	−46·0	−91·0	6·9	1·16	67
15-2·5	7760	72·0	−46·0	−91·0	6·9	1·09	113
8-4·0	1734	71·0	−47·0	−92·0	8·3	0·70	27
15-3·0	4585	71·0	−47·0	−92·0	8·3	1·00	69
15-2·5	7760	71·0	−47·0	−92·0	8·3	1·09	115

* S/N is calculated as the ratio of $(C - \bar{C})$ for the correct and the highest spurious peak, where C is the correlation coefficient.
† This column shows the real running time on the FPS-264 in a single task regime.

Table 2. *Results of RTRANS for Streptomyces griseus trypsin*

Resolution (Å)	Error (°)	Model*	Symmetry† $x\,\bar{y}\,\bar{z}$	$\bar{x}\,y\,\frac{1}{2}-z$	$\bar{x}\,\bar{y}\,\frac{1}{2}+z$	Combined
15-2·7	0·0	All atoms	1·97	2·00	1·95	2·31
15-2·7	0·0	Backbone	2·31	1·77	2·00	2·12
15-2·7	0·0	C_α	1·63	1·11	1·10	1·76
15-3·0	3·5	All atoms	1·52	1·36	1·89	2·08
15-3·0	3·5	Backbone	1·86	1·54	2·04	2·09
15-3·0	3·5	C_α	1·35	1·36	1·31	1·64
15-4·0	6·9	All atoms	0·83	<0·7	0·95	0·71
15-4·0	6·9	Backbone	0·78	<0·7	1·09	1·03
15-4·0	6·9	C_α	<0·8	0·66	0·90	0·81
15-3·0	6·9	All atoms	<0·7	<0·7	0·93	1·06
15-3·0	6·9	Backbone	0·78	0·83	1·26	1·11
15-3·0	6·9	C_α	0·85	<0·7	1·07	1·15
15-2·5	6·9	All atoms	0·89	<0·7	1·06	1·10
15-2·5	6·9	Backbone	0·82	0·85	1·20	1·17
15-2·5	6·9	C_α	0·70	<0·7	1·06	1·19
15-3·0	8·3	All atoms	<0·7	0·75	0·69	0·86
15-3·0	8·3	Backbone	0·75	0·89	0·92	1·12
15-3·0	8·3	C_α	<0·7	1·06	<0·8	0·68
15-2·5	8·3	All atoms	0·76	0·62	<0·7	0·95
15-2·5	8·3	Backbone	0·83	0·72	<0·75	1·01
15-2·5	8·3	C_α	0·60	<0·7	<0·6	0·80

* Backbone includes N, C_α, C, O and C_β atoms.
† The values shown in the table are the ratio of the *S*-function maximum corresponding to the correct solution to the highest spurious maximum. For larger orientation errors, the projection of the position of the combined peak on a plane perpendicular to a symmetry axis may be slightly off (1 grid point) from the position of the peak for this symmetry element.

Muir, Lee & Anderson, 1988) with the help of Crowther's fast rotation function, and the *BRUTE* correlation coefficient search. The latter calculations included data in the 4-6 Å resolution shell only. A database of commonly oriented domains from Fab fragments was created. For this purpose the KOL Fab fragment (Marquart, Deisenhofer, Huber & Palm, 1980) was positioned with its long axis along x and the 'elbow' axis along y and was used as a template for orienting domains from other Fab fragments. The rotation-function results indicated that the best model for the variable (V) domain of the JEL318 Fab fragment was the appropriate domain of the J539 Fab fragment (Suh *et al.*, 1986), and for the constant (C) domain the corresponding domain of the HED10 Fab fragment (Cygler, Boodhoo, Lee & Anderson, 1987). The numbers of atoms included in the models were 1721 and 1441 (~13 and ~11% of the unit-cell content) for the V and C domains respectively. Tables 3 to 6 summarize the results of calculations for the V and C domains using *BRUTE* and *RTRANS*. The calculations with *BRUTE* were performed on a grid of 0·8×0·8×0·8 Å. For *RTRANS*, the e.d. map was calculated on a grid of approximately the same size. As with SGT, the influence of the resolution of the data, errors in the orientation of the model and the size of the search model were investigated.

Table 3. *Results of BRUTE for JEL318 Fab fragment with V-domain model*

Resolution (Å)	Number of reflections	Orientation angles α (°)	β (°)	γ (°)	Orientation error (°)	S/N*	Computational time† (min)
6-4·0	2895	195·7	21·9	254·9	0·0	1·25	43
15-3·0	8842	195·7	21·9	254·9	0·0	1·49	138
6-4·0	2895	190·0	25·0	260·0	3·8	1·11	44
15-3·0	8842	190·0	25·0	260·0	3·8	1·46	133
6-4·0	2895	190·0	27·0	260·0	5·6	0·89	43
15-3·0	8842	190·0	27·0	260·0	5·6	1·18	133
6-4·0	2895	190·0	27·5	260·0	6·1	<0·77	43
15-4·0	4058	190·0	27·5	260·0	6·1	1·04	62
15-3·0	8842	190·0	27·5	260·0	6·1	1·08	132
15-2·4	12000	190·0	27·5	260·0	6·1	1·08	183
6-4·0	2895	190·0	28·5	260·0	7·1	<0·73	44
15-3·0	8842	190·0	28·5	260·0	7·1	0·91	134
15-2·4	12000	190·0	28·5	260·0	7·1	0·91	184

* As in Table 1.
† As in Table 1.

Table 4. *Results of RTRANS for JEL318 Fab fragment with V-domain model*

Resolution (Å)	Error (°)	Model*	Symmetry† $\frac{1}{2}+x\,\frac{1}{2}-y\,\bar{z}$	$\bar{x}\,\frac{1}{2}+y\,\frac{1}{2}-z$	$\frac{1}{2}-x\,\bar{y}\,\frac{1}{2}+z$	Combined
15-3·0	0·0	All atoms	1·59	2·23	1·35	1·73
15-3·0	0·0	Backbone	1·84	2·16	1·28	1·63
15-3·0	0·0	C_α	1·61	2·11	1·23	1·65
15-3·0	3·8	All atoms	1·53	1·31	1·05	1·66
15-3·0	3·8	Backbone	1·54	1·59	1·27	1·64
15-3·0	3·8	C_α	1·21	1·38	0·99	1·47
15-3·0	5·6	All atoms	0·94	1·06	0·83	1·11
15-3·0	5·6	Backbone	1·12	1·50	0·91	1·18
15-3·0	5·6	C_α	1·06	1·24	<0·7	1·20
15-4·0	6·1	All atoms	<0·7	1·15	<0·7	0·94
15-4·0	6·1	Backbone	<0·6	1·33	<0·8	1·12
15-4·0	6·1	C_α	0·74	1·36	<0·8	0·96
15-2·4	6·1	All atoms	<0·6	0·93	<0·7	0·91
15-2·4	6·1	Backbone	0·75	1·32	0·85	1·20
15-2·4	6·1	C_α	1·11	1·15	<0·8	1·08
15-3·0	7·1	All atoms	<0·7	0·72	<0·7	<0·7
15-3·0	7·1	Backbone	<0·7	0·93	<0·6	0·82
15-3·0	7·1	C_α	0·92	1·11	<0·6	<0·8

* Backbone includes N, C_α, C, O and C_β atoms.
† As in Table 2.

Table 5. *Results of BRUTE for JEL318 Fab fragment with C-domain model*

Resolution (Å)	Number of reflections	Orientation angles α (°)	β (°)	γ (°)	Orientation error (°)	S/N*	Computational time† (min)
6-4·0	2895	145·3	16·2	302·0	0·0	1·42	42
15-3·0	8842	145·3	16·2	302·0	0·0	1·28	136
6-4·0	2895	145·0	20·0	300·0	4·4	1·03	42
15-3·0	8842	145·0	20·0	300·0	4·4	1·20	137
6-4·0	2895	145·0	22·0	300·0	6·2	0·96	42
15-3·0	8842	145·0	22·0	300·0	6·2	0·91	134
6-4·0	2895	145·0	23·0	300·0	7·2	0·80	42
15-3·0	8842	145·0	23·0	300·0	7·2	0·82	132
15-2·4	12000	145·0	23·0	300·0	7·2	0·81	183

* As in Table 1.
† As in Table 1.

Table 6. *Results of RTRANS for JEL318 Fab fragment with C-domain model*

Resolution (Å)	Error (°)	Model*	Symmetry†			
			$\frac{1}{2}+x\,\frac{1}{2}-y\,\bar{z}$	$\bar{x}\,\frac{1}{2}+y\,\frac{1}{2}-z$	$\frac{1}{2}-x\,\bar{y}\,\frac{1}{2}+z$	Combined
15-3·0	0·0	All atoms	1·51	2·40	1·67	1·57
15-3·0	0·0	Backbone	1·50	2·53	1·47	1·47
15-3·0	0·0	C_α	1·13	1·77	1·28	1·37
15-3·0	4·4	All atoms	1·21	1·23	1·03	1·48
15-3·0	4·4	Backbone	1·43	1·31	0·80	1·37
15-3·0	4·4	C_α	1·08	1·41	<0·7	1·20
15-3·0	6·2	All atoms	1·27	0·90	1·04	1·49
15-3·0	6·2	Backbone	1·24	1·05	0·86	1·41
15-3·0	6·2	C_α	0·80	1·21	<0·7	1·13
15-4·0	7·2	All atoms	0·77	1·07	0·96	0·99
15-4·0	7·2	Backbone	<0·8	1·11	0·88	0·97
15-4·0	7·2	C_α	<0·7	0·75	<0·8	<0·8
15-3·0	7·2	All atoms	0·79	0·92	0·86	1·11
15-3·0	7·2	Backbone	0·71	0·89	0·81	1·23
15-3·0	7·2	C_α	0·72	0·73	<0·7	0·87
15-2·4	7·2	All atoms	<0·7	0·92	0·95	1·05
15-2·4	7·2	Backbone	0·76	0·86	0·83	1·08
15-2·4	7·2	C_α	0·67	0·68	<0·6	0·92

* Backbone includes N, C_α, C, O and C_β atoms.
† As in Table 2.

Discussion

In a previous paper (Cygler & Anderson, 1988*b*) we introduced the *S* function as a measure of the correspondence between the shape of a model protein molecule and the positive features of the electron density map. Here we have extended the formulation to include all symmetry-related molecules. We have shown that, under the condition of no overlap between the symmetry-related molecules, the *S* function is linearly related to the correlation coefficient between the 'experimental' and full-symmetry model electron density maps. The no-overlap condition is easily incorporated in the formula for the *S* function by adding an appropriate overlap penalty term.

We have also shown that, in the case when the position of one of the molecules in the e.d. map is known, the *S* function can be determined by summation of two-dimensional projections corresponding to each symmetry element. The method, being based on exploring Fourier rather than Patterson space, takes full advantage of phase information that can be obtained from an oriented molecular fragment. It is similar in general outline to the *TRADIR* procedure (Doesburg & Beurskens, 1983), although there are differences that are important for the application of this strategy to proteins. Indeed, the knowledge of the scale factor to bring the F_{obs} data to the absolute scale is not required, since this procedure does not depend on the difference Fourier coefficients but relies instead on a modification of the e.d. map in the real space. A precise determination of the scale factor for proteins is difficult for at least two reasons: limited resolution of the data and often only approximate knowledge of the unit-cell content (partially known sequence, no *a priori* information on the number of ordered solvent molecules). For the same reasons, all calculations are performed using *F* rather than *E* values. The evaluation of the *S* function in the *RTRANS* program is done in real rather than reciprocal space. The algorithm is very efficient and the CPU time is of the same order as for FFT. The methods for positioning a model molecule based on the search of correlation coefficient or *R* factor between F_{obs} and F_{calc} are much more computer intensive.

Since the *S*-function method separates the phasing step from the calculations of $S(\mathbf{v})$, the phases used to calculate the 'experimental' e.d. map do not necessarily have to come from the model used for searches. A different model from the phasing molecule can be used for the latter step, *e.g.* a subset of atoms such as the backbone atoms or even only C_α atoms.

It is important to note here that any other phase information can be used to calculate the e.d. map, *e.g.* the MIR or SIR phases can be utilized. In such a case the e.d. map does not have to be modified, but the *S* function cannot be partitioned. The calculations cannot be separated into two-dimensional tasks and would require much longer computations. A reciprocal-space approach of Read & Schierbeek (1988) seems to be computationally more advantageous in such a case (see below).

The *S* function has been tested for two proteins that crystallize in orthorhombic space groups. Despite the fact that only between 11 and 25% of the unit-cell content has been used to determine phases, the correct translation was easily identified even when orientation errors as large as 6–7° were imposed on the model. The discrimination of the correct solution from the noise was very good for small orientation errors and

depended on the size of the model. For SGT, where the model represented the whole molecule, the peak-to-noise ratio was 2·3. For the Fab fragment, where the model represents only half of the molecule (~12% of the unit-cell content), this ratio is ~1·6. For large orientation errors of the model it was often observed that the first incorrect maximum had only one wrongly determined coordinate.

As expected, the S function is less sensitive to errors in the model than the individual S_j functions. This is expecially pronounced when there are large errors in the model. In such cases, combining the results for all symmetries reveals all three components of the translation vector, even if some of the S_j functions do not have local maxima at the expected position. For example, for the orientation error of 7·2° the correct solution for the C domain of JEL318 Fab fragment did not correspond to the highest maximum of any of the individual symmetries but was the only consistent solution between different symmetry elements and was the highest maximum of the full-symmetry S function (Table 6).

We have made parallel calculations using the *BRUTE* program. These computations are 1–2 orders of magnitude more intensive than *RTRANS*. An immediate conclusion regarding this procedure is that inclusion of data from a broad resolution shell significantly increases tolerance of this method to orientation errors. For example, for SGT when 4–8 Å resolution data are used, the correct solution was not found as the top correlation coefficient peak for the third orientation [6·9° error; see also Fujinaga & Read (1987) and Read & James (1988)]. Inclusion of data in 15–2·5 Å resolution range moves the error limit to ~8° (Table 1). For smaller fragments, such as the V or the C domain of the Fab fragment, the orientation error limit seems to be about 6°.

The comparison of the two methods indicates that - at least for monoclinic and orthorhombic symmetries - the S function is somewhat less sensitive to orientation errors than the correlation coefficient between $|F^2_{obs}|$ and $|F^2_{calc}|$. Interestingly, this is most evident for the C domain of the Fab fragment, for which the model comprises the smallest fraction of the whole molecule of all tested proteins.

The fact that the S function calculated for the C domain of JEL318 Fab fragment gives a better signal-to-noise ratio for large misorientations than the model of the V domain, despite containing fewer atoms, confirms the results of the rotation function and *BRUTE* calculations (Cygler, Muir, Lee & Anderson, 1988), which suggested that the former is overall a better model of the corresponding domain of the JEL318 Fab fragment than the latter. This again is an indication that in practice it is better to use a smaller but more correct model for molecular-replacement procedures.

Read & Schierbeek (1988) have recently introduced a so-called phased translation function, which combines the known orientation of the model with available MIR phases to determine its translation in the unit cell. They have shown that even with poor phases, when the e.d. map is not interpretable, the MIR phase information can be used successfully to position the model. This method is a reciprocal-space equivalent of the present method (for one symmetry element), in which the MIR phases, rather than phases derived from the model itself, are used with the observed structure amplitudes. This additional experimental information may lead to a better discrimination, though the results would depend on the quality of MIR phases. This phased translation function was tested for SGT, using the same model and orientation parameters as in the present study (Read & Schierbeek, 1988). The MIR phases used in this test had a figure of merit of 0·68. The obtained peak-to-noise ratio for 3·5° orientation error was 2·16, and for 6·9° error it was 1·26. The values obtained by our method, 2·09 and 1·11 respectively (Table 2, search with backbone atoms), are not very different from the above results and yet no additional experimental information was used in the process.

We also tested the sensitivity of the S function to the resolution of the e.d. map. For small errors in the model the resolution at 15–4 Å was sufficient for recognizing the correct solution, but the higher-resolution data were always helpful. For larger errors in the model, inclusion of data to 3·0 Å resolution was essential, and further extension to 2·5 Å resolution led usually to a slightly better peak-to-noise ratio (Tables 2, 4 and 6).

An important part of the *RTRANS* algorithm is the choice of the overlap penalty function. The modifications of the e.d. map within the volume taken by the phasing molecule is necessary since the density is much larger there than in the other parts of the map, and could easily lead to incorrect solutions (Cygler & Anderson, 1988*b*). Since one has to expect that the side chains on the surface of the model might have different conformations from the protein being investigated, there might be some overlap between correctly placed symmetry-related model molecules and the phasing molecule. In fact our tests showed that when the penalty for overlap was too high, the correct solution was ranked very low for that very reason. The choice of the μ function that worked well in our hands was such that the values of μ were set to 0 on the surface of the phasing molecule and to a low negative value (between −12 and −4% of the maximum positive density) inside the phasing-molecule volume.

We have investigated an alternative of using as a search model a subset of the phasing molecule in which all side-chain atoms beyond C_β are removed. The reasoning behind this is as follows. One might

expect that the orientation errors in the model affect the outer parts of the molecule most, where the atoms are shifted farthest from their correct positions. Inside the molecule the displacement of atoms caused by the orientation error is smaller. A truncated model contains all the information about the global shape of the molecule, but loses the information about the details of its surface. One could be more specific and remove only surface side chains from the model, but the easiest way is to eliminate all side chains and that is what we have tried. The results (Tables 2, 4 and 6) show that, for large orientation errors, the S function calculated with a search model that included only backbone and C_β atoms was superior to the one calculated with the whole model.

One can further simplify the search model by including only one point *per* residue, approximately at its gravity center (well represented by the position of a C_α atom) to ensure that the centers of mass of the full and truncated models coincide. The global shape of the molecule is still retained and, since the computing time depends on the number of atoms, there is some gain in the speed of calculations. Even such a simplistic approach worked very well for SGT, but was less effective when the model comprised only a small part of the molecule.

In summary, for larger orientation errors, the best results were obtained with the model in which the side chains beyond C_β atoms were truncated, and with the data extending to at least 3·0 Å resolution.

A greater robustness of the S function to errors in the model over other methods [*e.g. BRUTE*; see also Cygler & Anderson (1988*b*)] occurs because the Fourier map is a simpler representation of the molecule than the Patterson map. The former contains fewer peaks and different parts of the molecule are separated in space. The Patterson map contains contributions from many more peaks within the same volume and suffers from overlap of intra- and intermolecular atom–atom vectors.

We would like to point out that the S function can also be straightforwardly used to locate one domain of the protein relative to the other as described previously (Cygler & Anderson, 1988*b*).

The *RTRANS* program is available from the authors upon request.

The authors would like to thank R. Read for providing the structure factors for SGT used in the tests.

References

BEURSKENS, P. T., GOULD, R. O., BRUINS SLOT, H. J. & BOSMAN, W. P. (1987). *Z. Kristallogr.* **179**, 127-159.

BLUNDELL, T. L., SIBANDA, B. L., STERNBERG, M. J. E. & THORNTON, J. M. (1987). *Nature (London)*, **326**, 347-352.

BOODHOO, A., MOL, C. D., LEE, J. S. & ANDERSON, W. F. (1988). *J. Biol. Chem.* **263**, 18578-18581.

CHAMBERS, J. L. & STROUD, R. M. (1979). *Acta Cryst.* B**35**, 1861-1874.

CHAMBERS, J. L., STROUD, R. M. & FINER-MOORE, J. (1988). Entry 4PTP in Brookhaven Protein Data Bank.

CHOTIA, C. (1984). *Annu. Rev. Biochem.* **53**, 537-572.

CHOTIA, C., LEVITT, M. & RICHARDSON, D. (1981). *J. Mol. Biol.* **145**, 215-250.

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173-178. New York: Gordon and Breach.

CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544-548.

CYGLER, M. & ANDERSON, W. F. (1988*a*). *Acta Cryst.* A**44**, 38-45.

CYGLER, M. & ANDERSON, W. F. (1988*b*). *Acta Cryst.* A**44**, 300-308.

CYGLER, M., BOODHOO, A., LEE, J. S. & ANDERSON, W. F. (1987). *J. Biol. Chem.* **262**, 643-648.

CYGLER, M., MUIR, A. K., LEE, J. S. & ANDERSON, W. F. (1988). Abstr. Am. Crystallogr. Assoc. Meet., 26 June-1 July 1988, p. 100.

DOESBURG, H. M. & BEURSKENS, P. T. (1983). *Acta Cryst.* A**39**, 368-376.

FUJINAGA, M. & READ, R. J. (1987). *J. Appl. Cryst.* **20**, 517-521.

GREER, J. (1985). *Ann. NY Acad. Sci.* **439**, 44-63.

GREER, J. (1988). *Proteins Struct. Funct. Genet.* **3**, 139-145.

HARADA, Y., LIFCHITZ, A., BERTHOU, J. & JOLLES, P. (1981). *Acta Cryst.* A**37**, 398-406.

HUBER, R. (1965). *Acta Cryst.* **19**, 353-356.

JANIN, J. & CHOTIA, C. (1980). *J. Mol. Biol.* **143**, 95-128.

KARLE, I. L. & KARLE, J. (1971). *Acta Cryst.* B**26**, 1891-1898.

MARQUART, M., DEISENHOFER, J., HUBER, R. & PALM, W. (1980). *J. Mol. Biol.* **141**, 369-391.

READ, R. J. & JAMES, M. N. G. (1988). *J. Mol. Biol.* **200**, 523-551.

READ, R. J. & SCHIERBEEK, A. J. (1988). *J. Appl. Cryst.* **21**, 490-496.

RICHARDSON, J. S. (1977). *Nature (London)*, **268**, 495-500.

RICHARDSON, J. S. (1981). *Adv. Protein Chem.* **31**, 167-339.

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

ROSSMANN, M. G. & ARGOS, P. (1976). *J. Mol. Biol.* **105**, 75-96.

ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24-31.

SUH, S. W., BHAT, T. N., NAVIA, M. A., COHEN, G. H., RAO, D. N., RUDIKOFF, S. & DAVIES, D. R. (1986). *Proteins Struct. Funct. Genet.* **1**, 74.