Lead Article

Acta Cryst. (1990). A**46**, 73–82

The Molecular Replacement Method

BY MICHAEL G. ROSSMANN

Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, *USA*

(*Received* 15 *June* 1989; *accepted* 26 *September* 1989)

Abstract

Molecular replacement can be used for obtaining approximate phasing of an unknown structure from a known related molecule and for phase improvement as well as extension in the presence of noncrystallographic symmetry. Emphasis is placed on the latter procedure. It is shown that the real-space method of iterative electron density averaging and Fourier back transformation corresponds to iterative phase substitution in the right-hand side of expressions to give a set of improved phases. Analysis of these expressions (the 'molecular replacement equations') provides insight into the limits of possible phase extension, and the implications for the use of calculated structure factors when there are no observed amplitudes. It is shown that the percentage of observed data and inaccuracy of the observed amplitudes available for phase extension are compensated by the extent of noncrystallographic redundancy and the fraction of crystal cell volume that may be flattened because it is outside the control of noncrystallographic symmetry.

Introduction

Structural redundancy, a consequence of noncrystallographic symmetry, can be used to solve the phase problem. The significance of this possibility for the solution of virus structures, in part, initiated my interest in virus structure (Rossmann & Blow, 1962). Nevertheless, it has taken until today for such a procedure to be fully implemented to a real structure determination. The Mengo virus (Luo, Vriend, Kamer, Minor, Arnold, Rossmann, Boege, Scraba, Duke & Palmenberg, 1987) and foot-and-mouth disease virus (Acharya, Fry, Stuart, Fox, Rowlands & Brown, 1989) structure determinations are the most powerful applications so far. Yet even now there is no example of a truly *ab initio* phase determination. There was an initially slow, but now ever faster, acceptance and use of the molecular replacement method. Indeed, today, in conjunction with the knowledge of probably most major types of protein folds, molecular replacement is a frequently used technique for the solution of macromolecular structures. The first exposition of noncrystallographic symmetry as a tool for structure determination was given by Rossmann & Blow (1962). The term *Molecular Replacement* was introduced as the name of the book in which the early papers were collected and briefly reviewed (Rossmann, 1972). This book remains a useful reference source. Another review was written in 1980 by Argos & Rossmann (1980), but there exists no review which has been written since noncrystallographic symmetry has been used successfully for phase extension (as opposed to phase improvement at a given resolution limit). I therefore hope that this mini-review, in which I shall emphasize phase determination, may be of some value. Examples are taken primarily from my own experience, but I hope that credit to important advances are correctly acknowledged.

Stages in molecular replacement

The original concept of the molecular replacement method (Rossmann & Blow, 1962) was of a three-stage process:

(1) Determination of relative orientation ('rotation') of identical unknown structures in the same or different crystals.

(2) Use of information from (1) to determine the position of the local noncrystallographic operators relative to the crystallographic symmetry elements ('translation').

(3) Phase determination using a knowledge of the noncrystallographic operators derived in (1) and (2).

With time and experience the types of applications have widened to include any of the following:

(1) *Rotation*

The determination of the relative orientation of unknown structures in the same crystal form might be of two or more molecules in one crystallographic asymmetric unit or between components of an oligomer or virus. The latter problem is one of finding the relative orientation of molecular axes, such as the

* *Editorial note*: This invited paper is one of a series of comprehensive Lead Articles which the Editors invite from time to time on subjects considered to be timely for such treatment.

0108-7673/90/020073-10$03.00

three perpendicular twofold axes in a tetramer with 222 symmetry or the symmetry axes of an icosahedral virus with 532 point symmetry. Thus, it is also possible to establish the point symmetry of an oligomer or biological assembly and hence the number of subunits of which it is composed. It is also possible to determine the relative orientation of the same molecule in different unit cells. This application can be used, for instance, in finding the relative orientation of a known molecular structure in an unknown unit cell. The purpose here is to use the known analogous molecule to solve the structure of the unknown crystal; that is, the known molecule is being used to search the unknown cell for a similar pattern. This is the most widely used application of the molecular replacement technique.

The rotation function is the tool for determining relative orientation. Hoppe (1957) and Huber (1965) had suggested the 'Faltmolekül Methode' for this purpose. However, the rotation function as proposed by Rossmann & Blow (1962), or its fast version (Crowther, 1972), is the usual tool.

(2) *Translation*

Having determined the relative orientation, one may now be able to place the molecule in space with respect to the crystallographic symmetry elements. Thus, for instance, if the orientation of a molecule has already been determined with a known search model, then it is necessary to define its position in the cell with respect to the selected origin. If there is no known or related structure but the molecule or macromolecular assembly has noncrystallographic symmetry, then it is necessary to define the position of a selected origin of the molecule (*e.g.* the center of the noncrystallographic point group, such as the intersection of the twofold axes in an object with 222 symmetry) relative to the chosen crystallographic origin.

Determination of the translation problem is often the most difficult part in the application of the molecular replacement method. When a search model of sufficient size is available, then such techniques as have been described by Crowther & Blow (1967), Argos & Rossmann (1980) and Blow, Rossmann & Jeffery (1964) or, more recently, the powerful procedure described by Read & Schierbeek (1988) and Schierbeek *et al.* (1989) should succeed. The determination of the point-group center of an unknown structure can be done by means of a translation function (Rossmann, Blow, Harding & Coller, 1964) only if there exists a twofold axis in the molecule. This resolves itself into a special case when the molecular twofold axis is parallel to a crystallographic twofold axis [*cf.* the Mengo virus structure determination (Luo *et al.*, 1987)] where inspection of the corresponding Harker section is sufficient. If these special symmetry conditions do not pertain, then it may be necessary to locate heavy atoms in the molecule which are themselves related by the molecular point-group symmetry (*cf.* Buehner, Ford, Moras, Olsen & Rossmann, 1974; Lin, Konno, Abad-Zapatero, Wierenga, Murthy, Ray & Rossmann, 1986).

Solution of the translation problem is often helped by packing considerations. A systematic application of this approach was made by Hendrickson & Ward (1976) or a less quantitative example is the determination of the virus center position of canine parvovirus (Luo, Tsao, Rossmann, Basak & Compans, 1988).

Results from the rotation function can be useful for the detection of heavy-atom positions, for these will (in general) be related by the same noncrystallographic symmetry as the molecule itself. A heavy-atom difference Patterson map (Rossmann, 1960) or, in the presence of anomalous dispersion, a Bijovet difference Patterson map (Rossmann, 1961) can be used for a search of the self vectors between heavy atoms related by the noncrystallographic point group (Argos & Rossmann, 1976; Arnold, Vriend, Luo, Griffith, Kamer, Erickson, Johnson & Rossmann, 1987). Similarly, as mentioned above, the cross vectors between molecules (usually related by crystallographic symmetry) can be used to determine the position of a molecule in crystallographic space.

(3) *Phase determination*

Three major applications exist:

(i) When a search molecule has been used to determine the orientation and position of the noncrystallographic symmetry operation, it can also be used to determine an initial set of phases. These can then be used for computing an electron density map whose structural interpretation should be easy. This structure can be refined using standard techniques which will also include the elucidation of that part of the structure that differs from the search model. Alternatively, the initial phases could be refined either by use of any available noncrystallographic symmetry or by density modification and solvent flattening (Wang, 1985; Bhat & Blow, 1982).

(ii) Phase improvement can be achieved at a given resolution in the presence of noncrystallographic symmetry. Normally this is achieved by averaging of the noncrystallographically related units and then back transformation of the average electron density. The resultant calculated (and presumably improved) phases are then applied, with suitable weighting, to the original F_{obs} for the computation of a new electron density map. The averaging is then repeated for the next cycle of phase refinement. That part of the structure outside the molecular envelope, and hence beyond the limits of the applicability of the local (noncrystallographic) symmetry, is usually flattened to represent solvent. Hence, the special case where the

noncrystallographic redundancy is unity corresponds to the Wang (1985) procedure. The averaged and modified density can then be Fourier back transformed to yield an improved set of phases. The cycle can then be repeated as often as computational resources permit or until convergence appears to have been achieved. Among the early successes of this procedure are the structure determinations of α-chymotrypsin (Matthews, Sigler, Henderson & Blow, 1967), hexokinase (Fletterick & Steitz, 1976) and lobster glyceraldehyde 3-phosphate dehydrogenase (Buehner, Ford, Moras, Olsen & Rossmann, 1974). The determination of hexokinase is particularly noteworthy as it involved a comparison of molecules in different crystal forms. More recently, the structure determination of the influenza virus neuraminidase spike (Varghese, Laver & Colman, 1983) and the major histocompatibility protein (Bjorkman, Saper, Samraoui, Bennett, Strominger & Wiley, 1987) are other examples of the use of a number of crystal forms. The most popular program is one written by Bricogne (1974, 1976), although others exist (*cf.* Johnson, 1978).

(iii) Phase extension was and remains the most controversial aspect of the molecular replacement method. It follows by induction that if phases can be successfully extended gradually from, say, 8 to 3 Å resolution [as was done for the Mengo virus structure determination (Luo *et al.*, 1987)] then there is no reason why phases should not be extended from 10 to 8 Å or from 20 to 10 Å resolution. An initial set of phases can normally be determined at 20 Å either from electron-microscopy studies or by making reasonable assumptions about molecular shape (*e.g.* a virus can usually be considered as a hollow shell of easily determinable outer and inner radii). Thus, in this application, the early ambitions of the molecular replacement method of *ab initio* phase determination have come true.

Phase extension

The procedure for phase extension, as currently used, depends on very gradual phase extension (about one or two reciprocal-lattice units at a time) interleaved with phase improvement at the current resolution. All this is normally done in real space by electron density averaging as briefly described above.

The original concept was of molecular averaging in reciprocal space (Rossmann & Blow, 1962). A notable success was an application to a made-up triclinic structure with fourfold redundancy (Main, 1967). Crowther (1967, 1969) had also made a series of elegant investigations of the procedure in reciprocal space. The molecular replacement equations of Main & Rossmann (1966) correspond closely to the $[H]$ matrix of Crowther. However, lack of computing power, the difficulty of bug-free programming in reciprocal space and a lack of true belief in the success of the method caused a great deal of skepticism as to the possibility of *ab initio* phase determination for 'real' structure determinations.

Early success with phase extension using real-space averaging came with an application to lobster glyceraldehyde 3-phosphate dehydrogenase (Buehner *et al.*, 1974; Argos, Ford & Rossmann, 1975) where the already known molecular envelope was assumed. Then, using 222 noncrystallographic symmetry, phases were obtained at 21·4 Å and extended to 6·3 Å resolution. The remarkable structure determination of polyoma virus at 22·5 Å resolution (Rayment, Baker, Caspar & Murakami, 1982) was a critical success for real-space phase extension. In this case the results contradicted the anticipated properties of quasi-symmetry (Caspar & Klug, 1962) and, thus, the structure determination had to be shown to be correct (Rayment, Baker & Caspar, 1983; Baker, Caspar & Murakami, 1983) in spite of widely held dogmatic views. An important milestone was reached with the phase extension at higher resolution from 4·0 to 3·2 Å in the structure determination of hemocyanin (Gaykema, Hol, Vereijken, Soeter, Bak & Beintema, 1984; Gaykema, Volbeda & Hol, 1986), providing confidence for the 1985 phase extension from 6 to 3 Å in the structure determination of human rhinovirus (Rossmann, Arnold, Erickson, Frankenberger, Griffith, Hecht, Johnson, Kamer, Luo, Mosser, Rueckert, Sherry & Vriend, 1985). Since that time numerous other virus structures have been determined by using an initially poor low-resolution phasing set and then improving and extending it to high resolution (Hogle, Chow & Filman, 1985; Luo *et al.*, 1987; Hosur, Schmidt, Tucker, Johnson, Gallagher, Selling & Rueckert, 1987; Chen, Stauffacher, Li, Schmidt, Bomu, Kamer, Shanks, Lomonossoff & Johnson, 1989; Acharya *et al.*, 1989).

I continue to believe that there may be advantages in computing time and rates of convergence by using reciprocal-space phase extension as originally proposed (Rossmann & Blow, 1963; Main & Rossmann, 1966), and am currently developing such a program. This time, however, I am certain in the knowledge that phase extension does work. Nevertheless, there are those who maintain that phase extension in reciprocal space is a waste of time - both for the computer and the researcher. Here, as on previous occasions (*cf.* Rossmann, 1972), it is difficult to tread a fine line between the frequently useful and important considerations of the pessimists and the perhaps equally convincing opposing and often lonely point of view.

Noncrystallographic symmetry

Crystallographic symmetry, by definition, holds throughout the infinite crystal. For instance, a

fourfold axis is true not only in the immediate neighborhood of a selected tetrad, but also implies that a unit cell some thousand Å from the origin superimposes on an identical unit cell after a 90° rotation. In contrast, noncrystallographic symmetry is true only within a defined envelope. The identical local symmetry will be true in neighboring unit cells, but does not extend from one cell to the next (Fig. 1). It follows that any part of the structure outside the defined envelope does not obey noncrystallographic symmetry. Conversely, if the envelope completely fills the crystallographic asymmetric unit then the presumed noncrystallographic symmetry is space filling and is, in fact, crystallographic symmetry.

If a periodic structure such as a crystal is superimposed on itself after operation with a noncrystallographic operator it will superimpose only within the envelope defining the local symmetry. A product of the superimposed periodic structures will be nonperiodic, containing only the point symmetry of the noncrystallographic operators (Fig. 2). This fact can frequently be used to select a molecular envelope where it was not obvious prior to noncrystallographic averaging (*e.g.* Buehner *et al.*, 1974; Lin *et al.*, 1986). Although no knowledge of the crystallographic envelope is needed for this first averaging, it is necessary to have determined it for the averaged molecular structure within the crystallographic cell to permit Fourier back transformation.

Two kinds of noncrystallographic symmetry elements may be defined: proper and improper. The

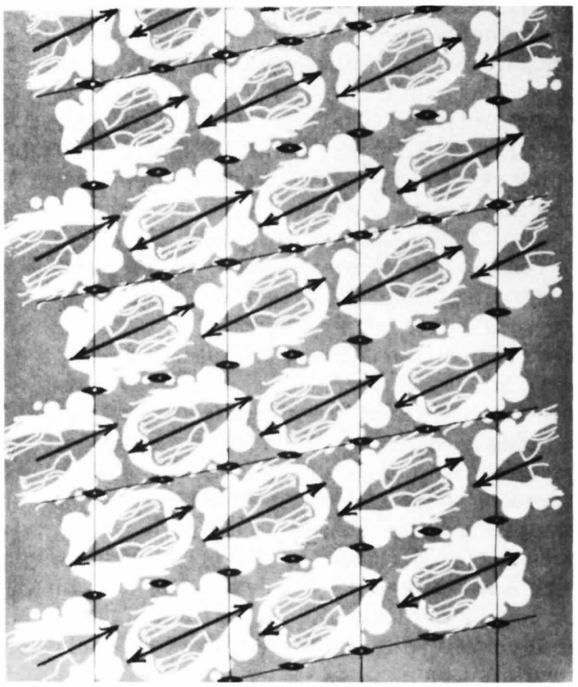

Fig. 1. Two-dimensional periodic design shows crystallographic twofold axis perpendicular to the page and local noncrystallographic rotation axes in the plane of the paper (design by Audrey Rossmann). [Reprinted with permission from Rossmann (1972). Copyright by Gordon & Breach.]

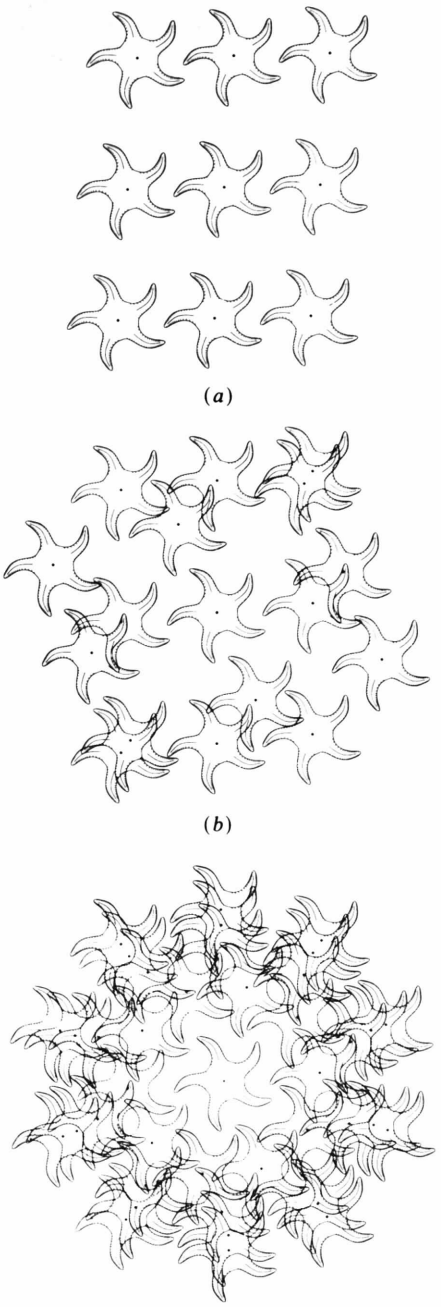

Fig. 2. (*a*) Noncrystallographic symmetry in a triclinic cell. (*b*) Superposition of the pattern in (*a*) on itself after operation with the noncrystallographic fivefold axis. (*c*) Superposition of the pattern in (*a*) on itself after a rotation of 1/5th, 2/5th, 3/5th and 4/5th. Note that the sum or product of periodic patterns is aperiodic and in (*c*) has the point symmetry of the noncrystallographic operation.

former satisfies a closed point group [*e.g.* a 17-fold rotation as occurs in tobacco mosaic virus disc protein (Champness, Bloomer, Bricogne, Butler & Klug, 1976)]. Here it does not matter whether a rotation axis is applied right- or left-handedly. The result is indistinguishable. On the other hand, the relationship between different molecules in a crystallographic asymmetric unit is unlikely to be a closed point group. Thus a rotation one way round (followed, no doubt, by a translation) might achieve superposition of the two molecules, while the other way round would not. This is called an improper noncrystallographic symmetry operator. An operation which takes a molecule in one unit cell to that in another unit cell (initially the cells are lined up with, say, their orthogonalized *a*, *b* and *c* axes parallel) must equally be an improper rotation.

It is irrelevant *where* in space a noncrystallographic rotation-symmetry operator is situated. The rotation operation will orient the two molecules similarly. A subsequent translation, whose magnitude depends on the location of the noncrystallographic symmetry operator, will always be able to superimpose the molecules (Fig. 3). Nevertheless, it is possible to select the position of the noncrystallographic symmetry axis such that the translation is a minimum, and that will occur when the translation is entirely parallel to the noncrystallographic rotation axis.

The position of a noncrystallographic symmetry axis, like everything else in the unit cell, must be defined with respect to the selected origin. Let us consider the noncrystallographic rotation defined by the 3×3 matrix $[C]$. Then, if the point $\mathbf{x}$ is rotated to $\mathbf{x}'$ (both defined with respect to a selected origin and axial system) we may write

$$\mathbf{x}' = [C]\mathbf{x} + \mathbf{d}$$

where $\mathbf{d}$ is a three-dimensional vector which expresses the translational component of the noncrystallographic symmetry operation. The values of $\mathbf{d}$ are quite arbitrary unless the position of the rotation axis is defined. Let us now assume that we are dealing with a proper rotation axis. Hence, there exists a point $\mathbf{x}$ on the axis (if placed to eliminate translation) such that it rotates onto $\mathbf{x}'$. It follows that

$$\mathbf{x} = [C]\mathbf{x} + \mathbf{d}.$$

If the molecular center is known, or if a point on the noncrystallographic symmetry element is known, then it is now possible to determine the components of the vector $\mathbf{d}$. Note that $\mathbf{d} = 0$ when, and only when, the noncrystallographic rotation axis passes through the selected crystallographic origin.

There are considerable advantages in the use of proper noncrystallographic symmetry. Consider, for example, a tetramer with 222 symmetry (such as lactate dehydrogenase). It is not necessary to define the chemical limits of one polypeptide chain. The noncrystallographic symmetry is true everywhere within the molecular envelope containing (in this example) four noncrystallographic asymmetric units. The boundaries of the polypeptide chain are irrelevant to the geometrical considerations. The electron density at every point within the molecular envelope (which itself must have 222 symmetry) can be averaged among all four 222 related points. On the other hand, if there is only improper noncrystallographic symmetry then the envelope must define the limits of one noncrystallographic asymmetric unit although the crystallographic asymmetric unit contains two or more such units.

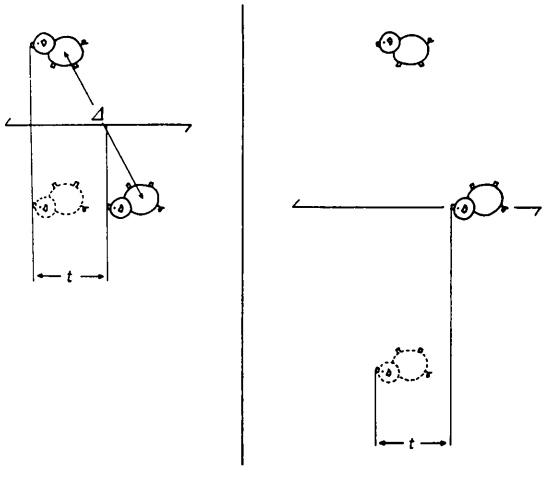

Fig. 3. The position of the twofold rotation axis which relates the two piglets is completely arbitrary. The diagram on the left shows the situation when the translation is parallel to the rotation axis. The diagram on the right has an additional component of translation perpendicular to the rotation axis, but the component parallel to the axis remains unchanged. [Reprinted with permission from Rossmann *et al.* (1964). Copyright by the International Union of Crystallography.]

Analysis of phase determination

No attempt will be made here to review the background of the rotation function or translation function as these procedures have been well reviewed frequently (Rossmann, 1972; Argos & Rossmann, 1980; Rossmann & Arnold, 1989) and are generally in use in many laboratories. On the other hand, phase extension in the presence of noncrystallographic symmetry appears to be poorly understood and it is, therefore, well worth some space here. Apart from the work prior to 1973 (with the exception of the α-chymotrypsin determination), phase extension has been used only in real space. Yet there is an exact equivalence in reciprocal space. It is my contention not only that the reciprocal-space view is more informative as to the mechanism of phase determination, but it is also a general expression of the rotation function, translation functions, density modification

and solvent flattening as well as direct methods such as are implied by Sayre's equations (Sayre, 1952; Arnold & Rossmann, 1986). A simple derivation of the molecular-replacement equations (Main & Rossmann, 1966) in terms of real-space averaging is given here.

The electron density for the averaged copy at $\mathbf{x}$ is given by

$$\rho_{\text{avg}}(\mathbf{x}) = (1/N) \sum_{n=1}^{N} \rho(\mathbf{x}_n), \tag{1}$$

where the averaging is over the N copies $\rho(\mathbf{x}_n)$. The noncrystallographic symmetry relating the different copies is given by

$$\mathbf{x}_n = [C_n]\mathbf{x}_1 + \mathbf{d}_n, \tag{2}$$

where $[C_n]$ is the rotation matrix relating the nth copy to the reference copy, and $\mathbf{d}_n$ is the corresponding translational component with respect to an arbitrarily selected origin. By replacing the electron density $\rho(\mathbf{x}_n)$ by its corresponding Fourier summation, it is seen that

$$\rho_{\text{avg}}(\mathbf{x}) = (1/N) \sum_{n=1}^{N} (1/V) \sum_{h} \mathbf{F}_h \exp(-2\pi i \mathbf{h}.\mathbf{x}_n), \tag{3}$$

where h is the Miller index. Now, recomputing structure factors using the averaged density and assuming zero density outside the molecular envelopes, we have for reflection p

$$\mathbf{F}_p = \sum_{n=1}^{N} \int_{U_n} \rho_{\text{avg}}(\mathbf{x}_1) \exp(2\pi i \mathbf{p}.\mathbf{x}_n).\mathrm{d}\mathbf{x}_n, \tag{4}$$

where U_n bounds the volume containing the nth copy. By substitution of (3) into (4) it follows that

$$\mathbf{F}_p = (1/NV) \sum_{h} \mathbf{F}_h \sum_{n=1}^{N} \exp(-2\pi i \mathbf{h}.\mathbf{d}_n) \times \int_{U_n} \exp\{2\pi i(-\mathbf{h}[C_n] + \mathbf{p}).\mathbf{x}_n\}.\mathrm{d}\mathbf{x}_n. \tag{5}$$

If we now define

$$\mathbf{G}_{hpn} = (1/U) \int_{U_n} \exp\{2\pi i(\mathbf{p} - \mathbf{h}[C_n]).\mathbf{x}_n\}.\mathrm{d}\mathbf{x}_n, \tag{6}$$

where U is the sum of the volumes bounded by U_n $(n = 1, 2, \ldots, N)$, and define

$$\mathbf{T}_{hpn} = \exp(-2\pi i \mathbf{h}.\mathbf{d}_n), \tag{7}$$

then (5) simplifies to

$$\mathbf{F}_p = (U/NV) \sum_{h} \mathbf{F}_h \sum_{n=1}^{N} \mathbf{G}_{hpn}\mathbf{T}_{hpn} \tag{8}$$

or

$$\mathbf{F}_p = \sum_{n=1}^{N} \left[(U/NV) \sum_{h} \mathbf{F}_h \mathbf{G}_{hpn}\mathbf{T}_{hpn} \right] = \sum_{n=1}^{N} \mathbf{F}_{\mathbf{h}'_n} \tag{9}$$

where $\mathbf{h}'_n = [C_n^T]^{-1}\mathbf{p}$ and corresponds to the rotation of $\mathbf{p}$ in reciprocal space equivalent to $[C_n]$ in real space. Thus, $\mathbf{F}_{\mathbf{h}'_n}$ is the structure factor at the nonintegral reciprocal-lattice point $\mathbf{h}'_n$ corresponding to the rotation of $\mathbf{p}$ by the nth noncrystallographic symmetry element. Hence, $\mathbf{F}_p$ represents the complex averaging of structure factors at the N noncrystallographically equivalent positions in reciprocal space.

If we simplify (8) by putting

$$\mathbf{a}_{hp} = \sum_{n=1}^{N} \mathbf{G}_{hpn}\mathbf{T}_{hpn},$$

then

$$\mathbf{F}_p = (U/NV) \sum_{h} \mathbf{F}_h \mathbf{a}_{hp}. \tag{10}$$

These are the molecular replacement equations defined by Main & Rossmann (1966), whose coefficients are equivalent to the $[H]$ matrix of Crowther (1967, 1969). Here the complex coefficients $\mathbf{a}_{hp}$ are determined entirely from a knowledge of the orientation, position and extent of the noncrystallographic symmetry elements. The molecular-replacement equations are exact other than the assumptions that the noncrystallographic symmetry holds to within the resolution limits of the available data and that the solvent regions of the cell can be approximated by a constant level of electron density.

Substitution of currently available approximate phases on the right-hand side of (10) will produce an improved set of phases on the left-hand side in a process which is entirely equivalent and the same as the real-space averaging and back-transformation procedure. Approximation enters in as far as many terms must be neglected in setting up the molecular replacement equations because they are deemed too small in magnitude to matter. The same approximation occurs in calculating a rotation function (Rossmann & Blow, 1962). In real space the approximations relate to linear (Bricogne, 1974) or nonlinear (Nordman, 1980) interpolation to obtain the value of electron density at nonintegral grid points. The elegance of merely substituting phases in a set of complex linear equations is self-apparent (at least to me!) and is also highly suitable for rapid arithmetic in parallel processing computers.

Just as is the case for the G function in its application to the rotation function, so here also the largest coefficient $\mathbf{a}_{hp}$ will be between terms of about the same resolution. Thus, the interactions represented by (10) will be significant only in a relatively thin shell at the same resolution as that of the structure factor $\mathbf{F}_p$. If we approximate the molecular envelope to be spherical with a radius R and if it is, say, in a unit cell with cell dimensions $4R$ (about eight particles in the cell), then the argument $\mathbf{H}.\mathbf{R}$ of G is given by $(n/4R)\boldsymbol{R}$ where n is the number of reciprocal-lattice points represented by the difference $\mathbf{p} - \mathbf{h}[C^T]$ [see (6)]. Now G for a spherical envelope (Fig. 4)

becomes zero for the first time when its argument is 0·7. Thus, the thickness of a shell required to include the larger interactions in the molecular-replacement equations (10) has a half-width when $n/4 = 0{\cdot}7$ or $n = 2{\cdot}8$ that is about three reciprocal-lattice units. The interpolation for each value of $\mathbf{F}_{\mathbf{h}'_n}$ within a radius of three reciprocal-lattice points around $\mathbf{h}'_n$ is equivalent in real space to the interpolation required to find the electron density at a nonintegral grid point.

Omission of a significant coefficient $\mathbf{F}_h$ on the right-hand side of (10) will cause an error in the value of $\mathbf{F}_p$. It is clearly more prudent to include an estimate of that value if there is no observed value available. This can be obtained by the calculation of $\mathbf{F}_h$ from the molecular replacement equation when $p = h$. This process is identical to the inclusion of $\mathbf{F}_{calc}$ values obtained by back transformation of an averaged electron density map in the computation of a new and improved map. Rayment (1983) and Arnold *et al.* (1987) have shown that such a procedure leads to a truer phase determination. Here is, then, the theoretical reason. Furthermore, it is now seen that if there were, say, no observed amplitudes, then the value of $\mathbf{F}_p$ on substituting all the $\mathbf{F}_{calc}$ values on the right would not change from the previous $\mathbf{F}_{calc}$ value. Thus, absence of some observed amplitudes slows down convergence but does not entirely stop progress towards a phase solution.

Similarly, omission of structure factors outside the current resolution limit leads to a decrease of satisfaction of the molecular replacement equations. An equation on the exact limit of resolution will have half its terms missing and cannot give a reasonable estimate of $\mathbf{F}_p$. The lack of agreement between observed and calculated amplitudes at the limit of resolution is typical (Fig. 5) and its necessity is easily appreciated. Since the sum of the terms on the right-hand side is essentially the sum of a random set of vectors, omission of some terms will cause an overall reduction of calculated $\mathbf{F}_p$ values. Hence, calculated structure factors will require progressively further up-scaling as they approach the limit of current resolution, as is indeed found in practice (Arnold *et al.*, 1987; Luo, Vriend, Kamer & Rossmann, 1989).

The correlation coefficient C is defined as

$$C = \frac{\sum (\langle F_{obs}\rangle - F_{obs})(\langle F_{calc}\rangle - F_{calc})}{[\sum (\langle F_{obs}\rangle - F_{obs})^2 \sum (\langle F_{calc}\rangle - F_{calc})^2]^{1/2}}$$

where $\langle\ \rangle$ denotes averages in local resolution ranges. It has been found to be a particularly useful way of checking the progress of refinement (Fig. 5). This was first introduced in the structure determination of tomato bushy stunt virus (TBSV) (Harrison, Olson, Schutt, Winkler & Bricogne, 1978; Harrison, Olson & Bricogne, 1977) and has been adopted generally. In Fig. 5 it is seen that the initial phase determination at 8 Å for Mengo virus using the crudely homologous HRV14 structure gave $C = 0{\cdot}54$. Refinement at that resolution rapidly improved the

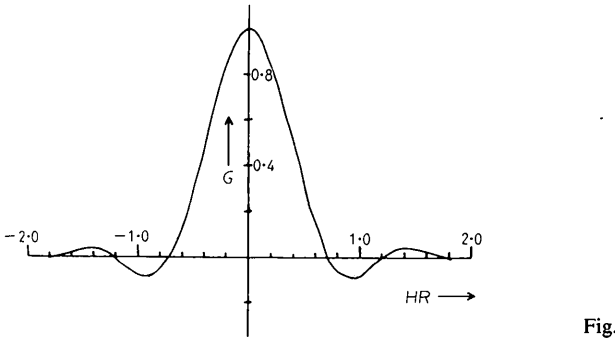

Fig. 4. Shape of the interference function G for a spherical envelope of radius R at a distance H from the reciprocal-space origin. [Reprinted with permission from Rossmann & Blow (1962). Copyright by the International Union of Crystallography.]

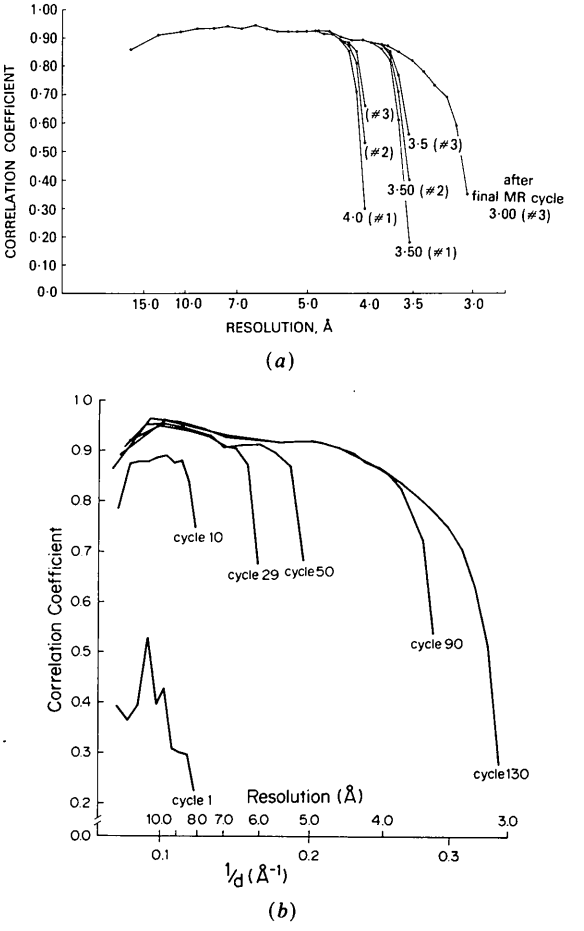

Fig. 5. Correlation coefficients plotted against resolution for selected phase extension steps: (*a*) For HRV14 from 4·3 to 3·0 Å resolution. [Reprinted with permission from Arnold *et al.* (1987). Copyright by the International Union of Crystallography.] (*b*) For Mengo virus from 8·0 to 3·0 Å. [Reprinted with permission from Luo *et al.* (1987). Copyright by the American Association for the Advancement of Science.]

correlation coefficient to 0·87. The phase extensions maintained this correlation coefficient until high resolution was attained where the lower accuracy and fewer data caused a diminution of the correlation coefficient (matched by an increase in the R factor between F_{obs} and F_{calc}).

Ab initio phase determination

Phase extension from rather low (8 Å) to high (3 Å or better) resolution has had its greatest success in the structure determinations of Mengo virus (Luo *et al.*, 1987) and foot-and-mouth disease virus (Acharya *et al.*, 1989). In both cases a rather poor low-resolution phasing model based on other, crudely homologous, picorna virus structures was used as a starting phasing set. It would, however, be useful to initiate phasing from a model at, say, about 30 Å resolution, which could be based on an electron-microscopy study or on simple assumptions such as a hollow shell for a spherical virus. Indeed, the radial distribution of structure amplitudes usually follows that of a shell or sphere at very low resolution and, hence, can be used to assign starting phases. This was indeed done in a 22·5 Å resolution study of southern bean mosaic virus (SBMV; Johnson, Akimoto, Suck, Rayment & Rossmann, 1976; Fig. 6) where phases had been extended from 30 to 22·5 Å. Similarly, the analysis of polyoma virus (Rayment *et al.*, 1982) started with a very simple low-resolution model.

If the model contains a center of symmetry, as is the case for a hollow shell, and if the distribution of particles in the cell is also centric (*e.g.* in the $R32$ cell of SBMV which has one particle per rhombohedral cell), then the initial phases will also be centric. However, if the symmetry elements of the averaged structure are not coincident with the crystallographic symmetry elements, then, when the averaged structure is put back into the crystal cell, the center of symmetry will be broken. Thus, the phases from the back-transformed averaged map will lack a center of symmetry, permitting phase improvement and then phase extension. The hand will have been chosen when the averaged structure was replaced into the cell, for there will have been two ways of doing this. The alternative ways are related by the center of symmetry in the crystal lattice.

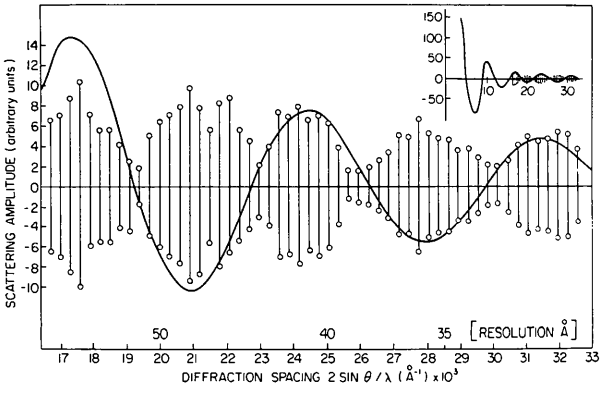

Fig. 6. The averaged radial distribution of structure amplitudes of $R32$ southern bean mosaic virus corresponds to that of a solid sphere. The radius of the sphere can be determined from the distance between nodes. Note that here there is only one virus particle per rhombohedral unit cell. Where there is more than one particle, the distribution will be the vector combination of the structures of all particles. In that case it will be necessary to know, at least roughly, the particle positions in order to extract a transform of the sphere (Johnson & Hollingshead, 1981). [Reprinted with permission from Johnson *et al.* (1976). Copyright by Academic Press Inc.]

An example where the center of symmetry can be broken in the manner described above is in the structure determination of $P2_1$ canine parvovirus (Luo *et al.*, 1988 and work in progress). Here there are two particles in the cell related by a 2_1 axis. If spherical particles are placed at (0·25, 0·25, 0·25) and (−0·25, −0·25, −0·25) there will be a center of symmetry at (000). However, one of the icosahedral axes is inclined by 2·5° to the crystallographic axis. Thus, the averaged particle, derived from an electron density map based on assigning signs according to the variation of the observed spherical transforms, will contain a mirror plane inclined by 2·5° to the crystallographic mirror plane perpendicular to the 2_1 axis. Hence, the electron density map produced by placing two averaged particles in the cell related by the 2_1 crystallographic axis will lack a center of symmetry at least at resolutions sufficient to be able to differentiate the 2·5° inclination. This procedure has worked in a test case and is now waiting to be applied to real data.

An example where the center of symmetry cannot be broken in this manner would have been rhombohedral SBMV. Here the 32 point symmetry of the crystal and icosahedron are coincident. Nevertheless, it may be possible to break the symmetry even in this case. A low-resolution large-intensity reflection might be arbitrarily assigned a phase of 90 or 270° (thus selecting a hand). If the arbitrarily selected reflection has a phase near 0 or 180° then the rest of the procedure will not work. However, if the reflection has a phase well away from 0 or 180° its presence in the Fourier map will perturb the electron density slightly away from centricity. Thus, the subsequent molecular replacement should gradually refine into that hand consistent with the selected phase angle for the arbitrarily selected reflection (*cf.* Rossmann & Blow, 1963).

An exciting recent result is that of the determination of the structure of the RNA phage MS2 (K. Valegård, L. Liljas and others, unpublished) and should be read by all those interested in the subject as soon as possible. A very-low-resolution structure of SBMV at 13 Å resolution was used as a starting phasing model. At this resolution the model is little more than a spherical

Table 1. *Tabulation of phasing power, P, for typical values of redundancy (N), solvent content, U/V, and errors on amplitude expressed as R values*

	N							
R	1	2	4	5	10	20	30	60
0·05	40	56	80	90	126	179	219	310
0·10	20	28	40	45	63	89	110	155
0·15	13	19	27	30	42	60	73	103
0·20	10	14	20	22	32	45	55	78

The error R can conveniently be expressed as

$$\sum_h \sum_i |(\mathbf{F}_h^2 - \mathbf{F}_{hi}^2)| / \sum_h \sum_i \mathbf{F}_h^2,$$

where $\mathbf{F}_h^2$ is the mean of i observations of reflections $\mathbf{F}_{hi}^2$. U/V is assumed to be equal to 0·5.

envelope. The structure, obtained by phase extension, turned out to be totally different from SBMV and any other virus structure. Furthermore, the density had the opposite Babinet solution - the negative density was the structure.

The quality of phase determination

Phase determination by molecular replacement depends entirely on the accuracy of the observed amplitudes in much the same way as phase determination using the isomorphous replacement method is largely dependent on the accuracy of the differences in amplitude between the various heavy-atom derivatives and native data. It was, therefore, not clear whether a unique solution to the phase problem could be achieved where the only physical input is the noncrystallographic symmetry and solvent flattening. This problem was considered by Crowther (1972), but a semi-quantitative relationship was also derived by Arnold & Rossmann (1986). The latter concluded that the 'power', P, of phase determination could be related to the noncrystallographic redundancy, N; the ratio of the volume to be averaged, U, to the total volume of the unit cell, V; the accuracy of structure-factor amplitudes, R; and the proportion of data, f, measured in a thin resolution shell by the expression

$$P = (Nf)^{1/2}/[R(U/V)]. \qquad (11)$$

It is assumed that the noncrystallographic symmetry is known absolutely correctly. While this might be almost true for the rotational and translational parameters, it is unlikely that the limits of the molecular envelope are known particularly well although they can be refined as the structure emerges. An example of the consequence of this relationship is given in Table 1.

The formula (11) also shows the relative importance of noncrystallographic symmetry in relation to the effect of solvent flattening. When there is no noncrystallographic symmetry, $N = 1$. If there is a lot of solvent, (U/V) (usually about 0·5) decreases and P increases proportionately. The effect of partiality of available data (a problem with crystals that are difficult to grow) is also seen to be less drastic than might be supposed. Indeed, the structure of human rhinovirus 1A was determined with only 55% of the observed amplitudes in the presence of tenfold redundancy ($N = 10$) (Kim, Smith, Chapman, Rossmann, Pevear, Dutko, Felock, Diana & McKinlay, 1989).

Someone inexperienced in phase extension might be wise to read one of the technique papers (Rossmann, 1989; Arnold *et al.*, 1987; Luo *et al.*, 1989). For instance, weighting coefficients using either the Rayment (1983) exponential factor or Sim (1959, 1960) weighting is useful. Checking the overall distribution of the F_{calc}'s for unobserved structure factors - they should follow the same distribution as F_{obs} - is useful. Checking phase changes is helpful to determine convergence.

Concluding remarks

The molecular replacement method has been shown to be extremely powerful. Phase determination in the presence of high noncrystallographic symmetry has been seen to be very accurate. Mean phase differences between molecular replacement phases and those calculated from an atomic model can differ by only about 10°, while multiple isomorphous replacement phases are usually accurate to only 60° (Arnold *et al.*, 1987). Thus, maps of virus structures are frequently of great beauty and extraordinarily easy to interpret.

Some major program packages are available for structural analysis using molecular replacement. Steigemann's *PROTEIN* package (Steigemann, 1974) includes E. Lattman's rotation function and translation function programs (*cf.* Crowther & Blow, 1967). Fitzgerald's *MERLOT* package (Fitzgerald, 1988) also includes Lattman's translation function and Crowther's fast rotation function (Crowther, 1972). Read's *BRUTE* package is particularly useful for translation functions (Read & Schierbeek, 1988). Bricogne's programs (Bricogne, 1974) are frequently used for phase improvement by molecular averaging, as is also Johnson's program (Johnson, 1978). A vectorized version of the original rotation function (Rossmann & Blow, 1962) has excellent versatility for exploring self- and cross-rotation functions. However, there are many other programs available too numerous to catalog here, many of which have probably never been mentioned in publications.

It has taken almost three decades to see the full acceptance and power of the molecular replacement method both as a technique for phase determination from a homologous model and for *ab initio* phase determination. I had originally thought it would take only six months to arrive at this point! The success is due to the enthusiastic and original work in many laboratories, but I would particularly like to express my appreciation to the many postdoctoral fellows who have participated in these studies over many

years. Among them, I am especially anxious to mention Cele Abad-Zapatero, Patrick Argos, Eddy Arnold, Manfred Buehner, John Erickson, Ignacio Fita, Geoffrey Ford, Jack Johnson, Greg Kamer, Andrew Leslie, Ming Luo, Peter Main, M. R. N. Murthy, Ivan Rayment, Tomitake Tsukihara and Gert Vriend. I also thank David Blow for many conversations in the early days of the development of the molecular replacement method. I am also greatly indebted to Sharon Wilder for making the manuscripts readable over the past 16 years as well as to Helene Prongay for help in the preparation of this manuscript. The work has been supported by grants from the National Institutes of Health and the National Science Foundation.

References

ACHARYA, R., FRY, E., STUART, D., FOX, G., ROWLANDS, D. & BROWN, F. (1989). *Nature (London)*, **337**, 709-716.

ARGOS, P., FORD, G. C. & ROSSMANN, M. G. (1975). *Acta Cryst.* A**31**, 499-506.

ARGOS, P. & ROSSMANN, M. G. (1976). *Acta Cryst.* B**32**, 2975-2979.

ARGOS, P. & ROSSMANN, M. G. (1980). In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. LADD & R. A. PALMER, pp. 361-417. New York: Plenum.

ARNOLD, E. & ROSSMANN, M. G. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 5489-5493.

ARNOLD, E., VRIEND, G., LUO, M., GRIFFITH, J. P., KAMER, G., ERICKSON, J. W., JOHNSON, J. E. & ROSSMANN, M. G. (1987). *Acta Cryst.* A**43**, 346-361.

BAKER, T. S., CASPAR, D. L. D. & MURAKAMI, W. T. (1983). *Nature (London)*, **303**, 446-448.

BHAT, T. N. & BLOW, D. M. (1982). *Acta Cryst.* A**38**, 21-29.

BJORKMAN, P. J., SAPER, M. A., SAMRAOUI, B., BENNETT, W. S., STROMINGER, J. L. & WILEY, D. C. (1987). *Nature (London)*, **329**, 506-512.

BLOW, D. M., ROSSMANN, M. G. & JEFFERY, B. A. (1964). *J. Mol. Biol.* **8**, 65-78.

BRICOGNE, G. (1974). *Acta Cryst.* A**30**, 395-405.

BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832-847.

BUEHNER, M., FORD, G. C., MORAS, D., OLSEN, K. W. & ROSSMANN, M. G. (1974). *J. Mol. Biol.* **82**, 563-585.

CASPAR, D. L. D. & KLUG, A. (1962). *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1-24.

CHAMPNESS, J. N., BLOOMER, A. C., BRICOGNE, G., BUTLER, P. J. G. & KLUG, A. (1976). *Nature (London)*, **259**, 20-24.

CHEN, Z., STAUFFACHER, C., LI, Y., SCHMIDT, T., BOMU, W., KAMER, G., SHANKS, M., LOMONOSSOFF, G. & JOHNSON, J. E. (1989). *Science*, **245**, 154-159.

CROWTHER, R. A. (1967). *Acta Cryst.* **22**, 758-764.

CROWTHER, R. A. (1969). *Acta Cryst.* B**25**, 2571-2580.

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

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

FITZGERALD, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273-278.

FLETTERICK, R. J. & STEITZ, T. A. (1976). *Acta Cryst.* A**32**, 125-132.

GAYKEMA, W. P. J., HOL, W. G. J., VEREIJKEN, J. M., SOETER, N. M., BAK, H. J. & BEINTEMA, J. J. (1984). *Nature (London)*, **309**, 23-29.

GAYKEMA, W. P. J., VOLBEDA, A. & HOL, W. G. J. (1986). *J. Mol. Biol.* **187**, 255-275.

HARRISON, S. C., OLSON, A. J. & BRICOGNE, G. (1977). In CECAM Report of Workshop on Virus Crystallography, pp. 1-15.

HARRISON, S. C., OLSON, A. J., SCHUTT, C. E., WINKLER, F. K. & BRICOGNE, G. (1978). *Nature (London)*, **276**, 368-373.

HENDRICKSON, W. A. & WARD, K. B. (1976). *Acta Cryst.* A**32**, 778-780.

HOGLE, J. M., CHOW, M. & FILMAN, D. J. (1985). *Science*, **229**, 1358-1365.

HOPPE, W. (1957). *Acta Cryst.* **10**, 750-751.

HOSUR, M. V., SCHMIDT, T., TUCKER, R. C., JOHNSON, J. E., GALLAGHER, T. M., SELLING, B. H. & RUECKERT, R. R. (1987). *Proteins*, **2**, 167-176.

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

JOHNSON, J. E. (1978). *Acta Cryst.* B**34**, 576-577.

JOHNSON, J. E., AKIMOTO, T., SUCK, D., RAYMENT, I. & ROSSMANN, M. G. (1976). *Virology*, **75**, 394-400.

JOHNSON, J. E. & HOLLINGSHEAD, C. (1981). *J. Ultrastruct. Res.* **74**, 223-231.

KIM, S., SMITH, T. J., CHAPMAN, M. S., ROSSMANN, M. G., PEVEAR, D. C., DUTKO, F. J., FELOCK, P. J., DIANA, G. D. & MCKINLAY, M. A. (1989). *J. Mol. Biol.* In the press.

LIN, Z., KONNO, M., ABAD-ZAPATERO, C., WIERENGA, R., MURTHY, M. R. N., RAY, W. J. JR & ROSSMANN, M. G. (1986). *J. Biol. Chem.* **261**, 264-274.

LUO, M., TSAO, J., ROSSMANN, M. G., BASAK, S. & COMPANS, R. W. (1988). *J. Mol. Biol.* **200**, 209-211.

LUO, M., VRIEND, G., KAMER, G., MINOR, I., ARNOLD, E., ROSSMANN, M. G., BOEGE, U., SCRABA, D. G., DUKE, G. M. & PALMENBERG, A. C. (1987). *Science*, **235**, 182-191.

LUO, M., VRIEND, G., KAMER, G. & ROSSMANN, M. G. (1989). *Acta Cryst.* B**45**, 85-92.

MAIN, P. (1967). *Acta Cryst.* **23**, 50-54.

MAIN, P. & ROSSMANN, M. G. (1966). *Acta Cryst.* **21**, 67-72.

MATTHEWS, B. W., SIGLER, P. B., HENDERSON, R. & BLOW, D. M. (1967). *Nature (London)*, **214**, 652-656.

NORDMAN, C. E. (1980). *Acta Cryst.* A**36**, 747-754.

RAYMENT, I. (1983). *Acta Cryst.* A**39**, 102-116.

RAYMENT, I., BAKER, T. S. & CASPAR, D. L. D. (1983). *Acta Cryst.* B**39**, 505-516.

RAYMENT, I., BAKER, T. S., CASPAR, D. L. D. & MURAKAMI, W. T. (1982). *Nature (London)*, **295**, 110-115.

READ, R. J. & SCHIERBEEK, A. J. (1988). *J. Appl. Cryst.* **21**, 490-495.

ROSSMANN, M. G. (1960). *Acta Cryst.* **13**, 221-226.

ROSSMANN, M. G. (1961). *Acta Cryst.* **14**, 383-388.

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

ROSSMANN, M. G. (1989). In *Improving Protein Phases*, edited by S. BAILEY, E. DODSON & S. PHILLIPS, pp. 49-56. Daresbury, Warrington: SERC.

ROSSMANN, M. G. & ARNOLD, E. (1989). In *International Tables for Crystallography*, Vol. B, edited by U. SHMUELI. Dordrecht: Kluwer. In the press.

ROSSMANN, M. G., ARNOLD, E., ERICKSON, J. W., FRANKENBERGER, E. A., GRIFFITH, J. P., HECHT, H. J., JOHNSON, J. E., KAMER, G., LUO, M., MOSSER. A. G., RUECKERT, R. R., SHERRY, B. & VRIEND, G. (1985). *Nature (London)*, **317**, 145-153.

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

ROSSMANN, M. G. & BLOW, D. M. (1963). *Acta Cryst.* **16**, 39-45.

ROSSMANN, M. G., BLOW, D. M., HARDING, M. M. & COLLER, E. (1964). *Acta Cryst.* **17**, 338-342.

SAYRE, D. (1952). *Acta Cryst.* **5**, 60-65.

SCHIERBEEK, A. J., SWARTE, M. B. A., DIJKSTRA, B. W., VRIEND, G., READ, R. J., HOL, W. G. J., DRENTH, J. & BETZEL, C. (1989). *J. Mol. Biol.* **206**, 365-379.

SIM, G. A. (1959). *Acta Cryst.* **12**, 813-815.

SIM, G. A. (1960). *Acta Cryst.* **13**, 511-512.

STEIGEMANN, W. (1974). PhD thesis, Technische Univ., München, Federal Republic of Germany.

VARGHESE, J. N., LAVER, W. G. & COLMAN, P. M. (1983). *Nature (London)*, **303**, 35-40.

WANG, B. C. (1985). *Methods Enzymol.* **115**, 90-112.