research papers

Molecular replacement – historical background

Michael G. Rossmann

Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392, USA

Correspondence e-mail:
mgr@indiana.bio.purdue.edu

A review is given of the mathematical procedures required for a molecular-replacement structure determination. These apply equally to the more frequently encountered situations where a known homologous structure can be used as a search model and to phase determination in the presence of non-crystallographic symmetry (NCS). In general, the former represents improper NCS between two different unit cells, whereas the latter occurs when there is proper NCS within one unit cell.

1. Introduction

Although the concept of non-crystallographic symmetry (NCS) was introduced in the first paper (Rossmann & Blow, 1962) on what is now known as the molecular-replacement (MR) method, the first recorded use of the term 'molecular replacement' was, as far as I can remember, in the title of my book on this subject (Rossmann, 1972), published ten years later. The summary that I wrote for this book in 1972 remains relevant and will serve as a component of the following text. I have indicated the original text by using italics. Numerous other reviews on MR have also been written (Lawrence, 1991), including some by myself (Rossmann, 1990, 1995) and by Eddy Arnold and myself (Rossmann & Arnold, 1993). In addition, there have been at least two Daresbury Study Weekends (Machin, 1985; Dodson *et al.*, 1992) devoted to MR. Furthermore, the independent development by Walter Hoppe and colleagues of the Faltmolekülmethode should not be forgotten (Hoppe, 1957*a*,*b*).

Undoubtedly, the most significant change in emphasis since the 1972 publication (Rossmann, 1972) is as a result of the now very large PDB database. Whereas in 1972 most structural studies had to employ NCS for *ab initio* phase determinations, today the most frequent use of MR is to solve a structure when that of an analogous molecule or domain is already known. Indeed, it is to be anticipated that with the high level of funding for structural genomics, that there will soon be available systematic search methods that compare an unknown structure with a representative set of known protein folds.

My intent in coining the description 'molecular replacement' was to cover all methods that utilize NCS whether within or between crystal forms. However, comments by Phil Evans and Gerard Bricogne during the 2001 CCP4 meeting indicated that they preferred to limit the term 'molecular replacement' to the case where an unknown structure is to be solved with a known search molecule. As the computational procedures have much in common in all uses of NCS, I will here continue to employ the term MR in the manner I had first intended (*Appendix A*).

2. Genesis of molecular replacement

By the summer of 1960 only two proteins had been solved: myoglobin at 2.0 Å resolution and hemoglobin at 5.5 Å resolution. A number of other proteins were being studied, but none had given up its secrets. The principal difficulty lay in the preparation of good single-site heavy-atom derivatives. It was necessary either to make some advance in the alchemy used in the preparation of the derivatives or to find an alternative method if the two respiratory proteins were not to stand unchallenged for many years. Some handle was needed which was a natural property of these large molecules and which did not depend on some artificial device as is implied in the isomorphous replacement technique.

David Blow and I had already shown that much could be performed with only one heavy-atom derivative (Blow & Rossmann, 1961), but the enormity of the impasse was still considerable. While discussing these problems with colleagues, at the time of the International Union of Crystallography held in Cambridge in 1960, it occurred to me that both the ability of proteins to crystallize in different space groups and their frequent property of being made up of identically folded polypeptide chains might form a basis for an alternative process in the solution of the phase problem.

Many larger protein molecules are made of identical, or closely similar, subunits. Plausible causes for such aggregates of the protein parts of virus structures were set out by Crick & Watson, 1956). Changes in the amino acid sequence of functionally similar proteins produce only minor structural alterations indicating that the evolution of divergent functions can occur without large changes in structure.

3. Defining NCS

Crystallographic and non-crystallographic symmetry are differentiated by the property that an operator applies throughout the whole infinite crystal for the former, whereas a non-crystallographic symmetry element relates only to a localized volume within a crystal. For instance, in Fig. 1 there are non-crystallographic twofold axes in the plane of the paper valid only in the immediate vicinity of each line. However, the crystallographic twofold axes perpendicular to the plane of the paper apply to every point within the crystal. Two types of non-crystallographic symmetry can occur. A 'proper' element is independent of the sense of rotation. An example of this kind is a fivefold axis in a virus. Clearly a rotation of the crystal either left or right by one-fifth of a revolution will leave all parts of a given virus coat (but not the whole crystal) in equivalent positions. Improper rotation axes are found if two molecules are arbitrarily oriented in the same asymmetric unit or are in two entirely different crystal lattices. It follows that any non-crystallographic operator that possesses an element of translation must be improper. Because proteins consist only of L-amino acids and nucleic acids have only D-ribose rings, NCS in crystals of biological macromolecules consists exclusively of rotation axes (with or without translational components).

4. Utilization of NCS

Molecular replacement utilizes the similarity or identity of structure in different parts of the crystallographic asymmetric unit caused by a repetition of the same subunit structure in the formation of a whole molecule. Molecular replacement may also utilize the relationship between differing crystal forms of the same or similar molecules. Three principal stages in determining a structure in the presence of NCS are as follows.

A. Determination of the relative orientation of the independent molecules, or subunits of molecules, within one crystal lattice or between different crystal forms. This may be performed by a systematic inspection of the Patterson function(s), particularly in the region nearer the origin where there are more vectors arising from within subunits than vectors between subunits.

B. Given the relative orientation, the translation of the subunits must be determined with respect to designated crystallographic symmetry elements. This can be performed by inspection of Pattersons in certain special cases, by the use of an homologous search model or by analysis of heavy-atom sites. At the completion of stages A and B, the exact equivalence between point **x** *in the standard molecule or subunit and the point* **x**′ *in any of the other subunits will be known and can be expressed a*s

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

where [C] is the rotation matrix determined in stage A and **d** *is the translation vector determined in stage B. The above relationship will be true only within the volume U of the standard subunit.*

The translation vector **d** must be defined relative to a stated origin, such as a crystallographic symmetry element. If a specific point in the reference molecule is known to be at **S**
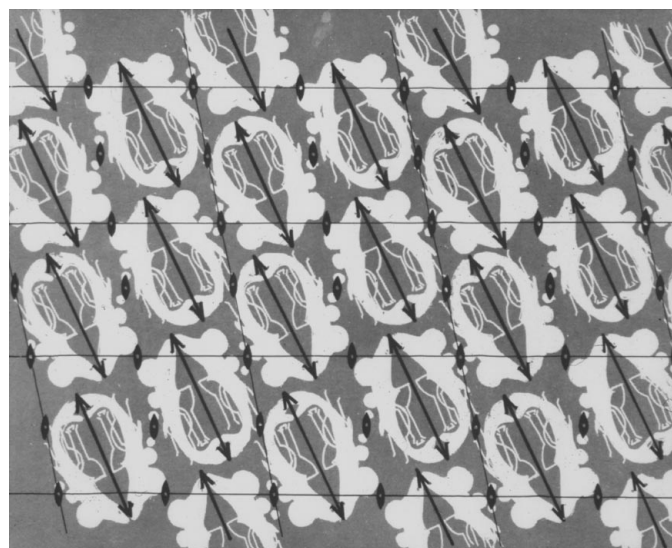

Figure 1
Two-dimensional periodic design shows crystallographic twofold axes perpendicular to the page and local non-crystallographic rotation axes in the plane of the paper. [Reprinted with permission from Rossmann (1972). Copyright (1972) Gordon & Breach.]

relative to the origin of the coordinate system, then the NCS-related position will be at $\mathbf{S}'$. Thus

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

If the NCS applies to a molecule with a closed point group (*e.g.* a virus with 532 symmetry), then if $\mathbf{S}$ defines the center of the molecule, $\mathbf{S}' = \mathbf{S}$, as the center stays in the same position after rotation. Hence it follows that in this common situation,

$$\mathbf{d} = ([1] - [C])\mathbf{S}.$$

C. Determination or refining an initial set of phases by (i) use of one or more known homologous structures, (ii) *ab initio* phase determination using the NCS relationships, (iii) phase refinement of the initial phases or (iv) phase extension of the initial phases to higher resolution.

5. Early results and problems

It was fortunate that, by the time I had formulated the general concepts of molecular replacement, I had also established a remarkable collaboration with David Blow. His natural caution and skepticism caused me to put rigor into my derivations; his willing ear and astute criticisms gave me sufficient encouragement to proceed. Encouragement came also from Max Perutz who first suggested to Dorothy Hodgkin that the rotation function might determine the relative orientation of the two crystallographically independent molecules in rhombohedral insulin. The peak in the ensuing rotation function of R3 insulin (Fig. 2) *was most pronounced* (Rossmann & Blow, 1962) *and was the beginning of a happy collaboration on the application of the molecular replacement method to insulin.*

There were, however, formidable objections to the solution of the translation problem and possibly even greater obstacles in the solution of the phase problem. Frances Crick was active in pointing out that the amount of translation required to superimpose two identical objects, after they had been similarly oriented, would depend on the position of the axis of rotation. How then could there be a unique solution of the translation problem at all? For a time this appeared to be an insurmountable hurdle, but gradually, as we considered the specific translation problem of R3 insulin, the difficulties became less formidable. Nevertheless, some of the biggest unsolved problems (even today) *still lie in the determination of equivalent origins in differing crystal forms of equivalent molecular species; that is in the solution of the translation problem in the presence of improper non-crystallographic symmetry*, in the absence of a known search model.

Frances Crick and Max Perutz had in addition serious objections to the third stage, the solution of the phase problem. It had been pointed out that the information presented by identical subunits which were different crystallographically was similar to that obtained by sampling the reciprocal lattice at non-integral lattice points. This is readily understood when it is shown (see *Appendix B*) that the rotation, represented by $\mathbf{x}' = [C]\mathbf{x}$, of the position vector $\mathbf{x}$ to $\mathbf{x}'$ in real space corre-

sponds to the rotation of the reciprocal-lattice vector $\mathbf{h}$ to $\mathbf{h}'$ given by $\mathbf{h}' = [C^T]\mathbf{h}$. Thus, the integral reciprocal-lattice point $\mathbf{h}$ will become a non-integral reciprocal-lattice point $\mathbf{h}'$ at the same resolution. *The situation was indeed similar to the hemoglobin shrinkage stages investigated by Perutz (1952)* (Fig. 3). *When the sign of the molecular transform in the centric h0l zone changes, then a discontinuity results in the plot of its magnitude. Since various shrinkage stages permitted the sampling of the transform's magnitude at points far closer than its anticipated minimum wave length, it became possible to determine the position of these discontinuities and hence the sign of the transform at every point. Frances Crick pointed out that, in the more general non-centric case, there would be no discontinuities in the function $\sqrt{A^2 + B^2}$, as might be present separately in the real, A, or imaginary, B, parts. That is, even if the magnitude of a transform of a molecule is known at every*
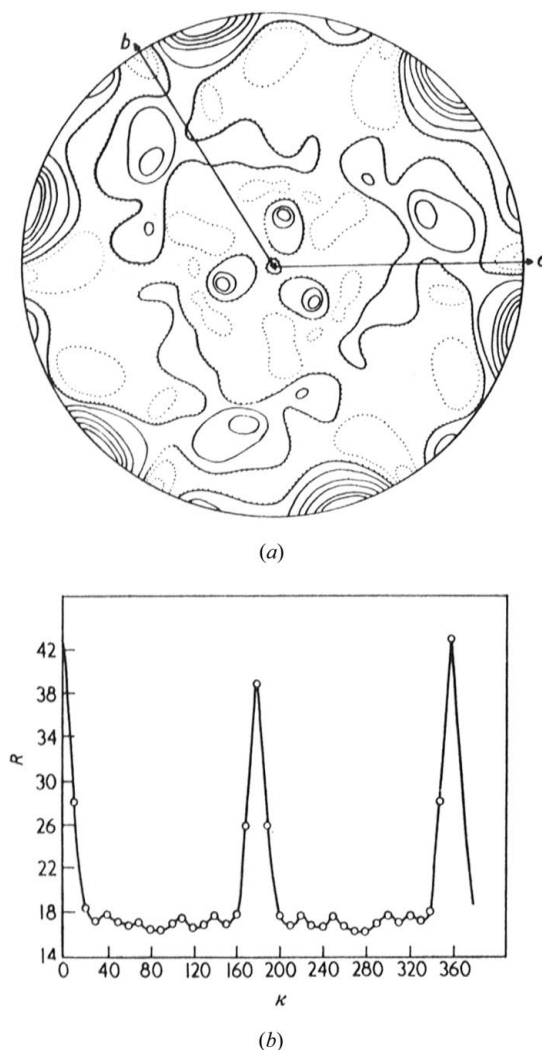

(*a*)

(*b*)

Figure 2
The rotation function for 2 Zn insulin at 6 Å resolution. The space group was *R*3 with $a = 82.5$ and $c = 34.0$ Å in the hexagonal setting. (*a*) A stereographic projection down the *c* axis showing the $R = 180°$ section. The *a* and *b* axes are indicated by the straight lines. The NCS peak is at $\psi = 90$, $\varphi = 44°$. (*b*) The line through the rotation function at $\psi = 90$, $\varphi = 44°$ as a function of κ. [Reprinted with permission from Dodson *et al.* (1966). Copyright (1966) Academic Press.]

point in space, Crick argued, the structure of the molecule could not be determined. These arguments were powerful, indeed sufficiently so, that I found myself working alone for some time.

In hindsight, it is easy to see that the use of the hemoglobin shrinkage stages to phase (sign) determination is an example of MR in the presence of multiple crystal forms. Thus, Max Perutz should be credited with the first successful application of the MR method.

The first ideas on how to use NCS for phase determination were based on equating the electron density, ρ, at points in the asymmetric unit which were equivalent chemically. Thus, if the non-crystallographic relationship is expressed as

$$\mathbf{x}'_n = [C_n] + \mathbf{d}_n \text{ with } \mathbf{n} = 1, N \text{ for each NCS operator}$$

then

$$\rho(\mathbf{x}_1) = \rho(\mathbf{x}_2) = ... = \rho(\mathbf{x}_N).$$

Each relationship *can be expressed as a Fourier summation where the unknowns are the phase angles,* α_h.

As many equations of this type can be written as there are unknown phases to be determined. Each expresses the equality of electron density at a different pair of points $\mathbf{x}$ *and* $\mathbf{x}'$. *The problem is then one of solving this non-linear system of equations for the unknown phases. From here it was only a small step to employ a continuous integration over the volume of one molecule* (see *Appendix B*). *Yet no success was encountered in the solution of the equations.*

The cause of failure was not in the physical content of the equations, but in the method of their solution. All attempts so far had been based on separating real and imaginary parts and then solving these two families of equations separately for $\cos\alpha$ *and* $\sin\alpha$. *Solutions arrived at in this way did not in general satisfy the condition* $\cos^2\alpha + \sin^2\alpha = 1$.
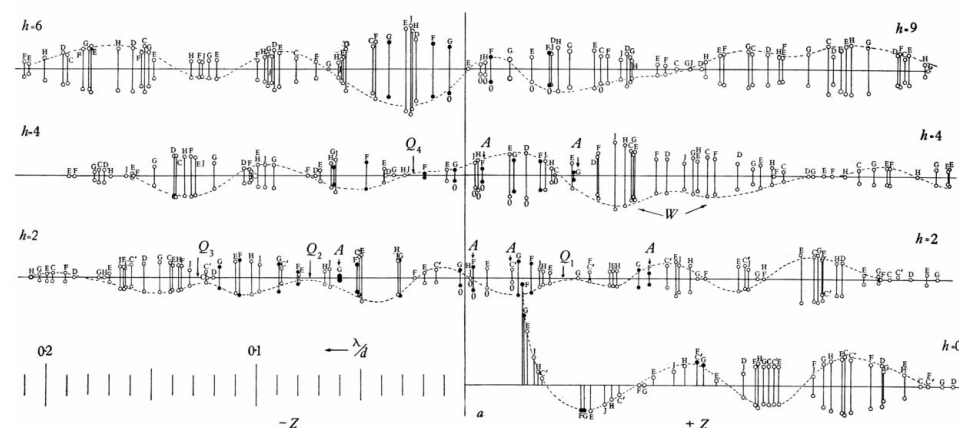
6. Eventual success in the use of NCS for phase determination

Details of the early attempts to solve the resultant MR equations (Main & Rossmann, 1966; Crowther, 1967, 1969; Main, 1967) are recorded in my book, but are no longer important for us today. A major breakthrough came with the solution of the structure of glyceraldehyde-3-phosphate dehydrogenase (Buehner *et al.*, 1973) in which the 222 NCS of the molecule was used to improve the MIR phasing by using real-space averaging. The equivalence of real-space averaging to the reciprocal-space solution of the MR equations had been hinted at by Main & Rossmann (1966), Colman (1974) and Bricogne (1974) (see *Appendix B*). The principal advantage of real-space averaging was that it was easier to program and the results were easy to appreciate and understand. Reciprocal-space phase determination in the presence of NCS was revisited by Tong & Rossmann (1995). Although reciprocal space provides insight to possible problems and can be represented in a simple and elegant mathematical form, experience in imposing NCS in real space makes it likely that most or all future phase determination or refinement using NCS and solvent flattening will be by real-space density averaging.

Phase extension using NCS remained a highly contentious subject. However, in 1984, Hol and colleagues (Gaykema *et al.*, 1984) extended phases from 4.0 to 3.2 Å resolution in their structure determination of hemocyanin. Soon afterwards, encouraged by Hol's results, we used real-space averaging to extend phases from 6.0 to 3.5 Å resolution in our structure determination of human rhinovirus 14 (Rossmann *et al.*, 1985) and since then quite a number of virus structures have been solved by phase extension from about 20 Å. It therefore took 25 years (from 1960 to 1985) to see the complete success of the proposal I made at the time of the International Union of Crystallography meeting in Cambridge in 1960.

7. Babinet inversions and other problems encountered in MR

If very low resolution data is missing from a data set and phase determination has been initiated with a low-resolution model (*e.g.* with a cryo-EM structure), then there is a danger of obtaining a Babinet inverted phase solution where the phase solution is $\alpha + \pi$ instead of α. This means that positive density is negative and *vice versa*. The cause of such an inversion can be readily understood by examination of Fig. 3. If the low-resolution terms were absent from the (00*l*) reflections, then it would be easy to see how the continuous hemoglobin Fourier transform

Figure 3
Variation of the hemoglobin molecular transform for the centric (*h*0*l*) reflections derived from the structure amplitudes of seven shrinkage stages. Note that the data from the shrunken unit cells plots at non-integral reciprocal-lattice points of the fully wet crystals allowing for the interpolation of the molecular transform between integral reciprocal-lattice points. [Reprinted with permission from Perutz (1954). Copyright (1954) the Royal Society of London.]

could be followed while interpreting all negative regions as positive and positive regions as negative. A similar situation could have occurred in the initial phasing of the low-resolution terms in the structure determination of southern bean mosaic virus (Fig. 4). The correct assumption that the virus particle radius is 150 Å shows that the lowest recorded resolution data, between 64 and 51 Å, have a phase of 0, whereas an assumption of a radius of 200 Å would make these phases have a value of π. Yet either assumption would have given an almost equally good fit to the experimentally observed data.

A related problem occurred in the structure solution of the bacterial phage φ29 (Simpson *et al.*, 2000). Here, the NCS 12-fold rotation axis was roughly perpendicular to the $a*b*$ reciprocal-lattice planes. Thus, the process of electron-density averaging (see *Appendix B*) only generated interpolations between the reciprocal-lattice points within each $l = 0, 1, 2, \ldots$ reciprocal-lattice plane, with little interaction between planes. Hence, the Babinet solution and handedness of the phases of each plane were mostly independent of each other, although consistent within a plane. As a result, the electron density could not be interpreted until a better overall phasing start had been obtained from an isomorphous derivative.

8. Conclusions

I had imagined that it might be difficult for the inexperienced crystallographer to use MR without a sufficient mathematical foundation. However, I am delighted that there are now a number of easy-to-use programming packages that allow even the most inexperienced student to use MR without tears. Foremost among these programs are *AMoRe* by Navaza (1994), *DM* (Cowtan, 1994), *CNS* (Brunger et al., 1998) and *Patsol* (Tong, 1993). These automatic procedures are likely to be of progressively more help in solving structures as the number of depositions in the Protein Data Bank (PDB) steadily increases. In addition, the use of MR for NCS
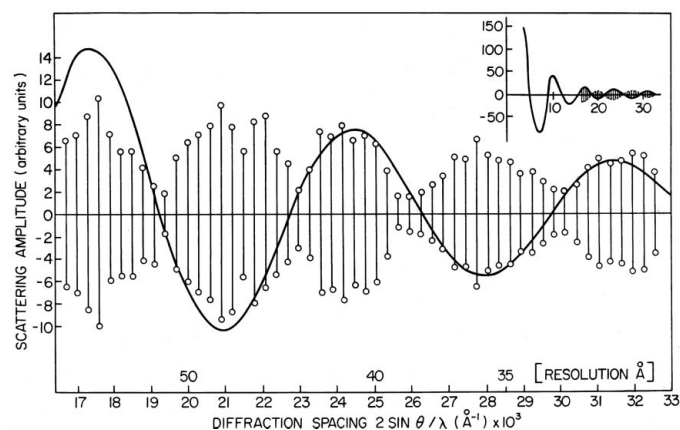

Figure 4
Mean amplitude of radial averaged structure amplitudes of southern bean mosaic virus follow the transform of a sphere with a radius of 150 Å. However, in the absence of very low resolution data, a slight change in the assumption of the radius would have resulted in the interpretation of the positive parts of the transform as negative and *vice versa*. [Reprinted with permission from Johnson *et al.* (1976). Copyright (1976) Academic Press.]

electron-density averaging is much better understood and will allow more *ab initio* determinations as the complexity of structures being determined continues to increase.

As progressively more structures are solved by MR, the danger of major errors increases owing to bias introduced by the initial phasing model. This is a significant problem when phasing is initiated from an assumed structure as is necessary both when using a homologous model or even with '*ab initio*' phasing. Notwithstanding the term '*ab initio*', it is minimally necessary to assume a very low resolution model (such as a solid sphere) to initiate phasing. It will be essential to observe that the MR procedure has produced anticipated structural features (*e.g.* a Trp where there was a Gly in the phasing model) that are consistent with information not present in the phasing model.

APPENDIX *A*
The equivalence of different MR procedures

The rotation function $R(\theta_1\theta_2\theta_3)$ can be defined as a match between two Patterson functions (Rossmann & Blow, 1962), such that

$$R(\theta_1\theta_2\theta_3) = \int_U P_1(\mathbf{x})P_2(\mathbf{x}') \, d\mathbf{x}, \qquad (1)$$

where $R(\theta_1\theta_2\theta_3)$ represents the rotation function when the Patterson $P_1(\mathbf{x})$ is rotated by the angles θ_1, θ_2, θ_3 from an arbitrarily defined original orientation and then superimposed onto another Patterson $P_2(\mathbf{x}')$. The superimposed Pattersons are then integrated over the volume U. A large value of R indicates a good match between the two Pattersons. It can readily be shown (Rossmann & Blow, 1962) that expression (1) is equivalent to the corresponding reciprocal-space expression

$$R(\theta_1\theta_2\theta_3) = \sum_h \sum_p \mathbf{F}_h^2 \mathbf{F}_p^2 G_{hp}, \qquad (2)$$

where $\mathbf{F}_h^2$ and $\mathbf{F}_p^2$ are the Patterson Fourier coefficients for the Patterson P_1 and P_2, respectively. G_{hp} is a diffraction function of the form

$$G_{hp} = \int_U \exp[2\pi i(\mathbf{h} + [C^T]\mathbf{p}) \cdot \mathbf{x}] \, dx \qquad (3)$$

(see the body of this paper for the definition of $[C]$).

Two types of MR situations were mentioned in this paper. The first is when a homologous phasing structure is available. In that case, the first Patterson P_1 (with Fourier coefficients $\mathbf{F}_h^2$) would be those of the known molecule and the second Patterson P_2 (with coefficients $\mathbf{F}_p^2$) would be those of the unknown molecule placed in an arbitrary unit cell, generally of space group $P1$. A 'cross-rotation function' (equation 2 above) then determines the orientation of the known molecule with respect to the unknown molecule in each of the asymmetric units of the crystal under investigation.

The second type of MR situation utilizes the NCS relating different copies of the unknown molecule (or unknown subunits within a larger molecule). In that case, a 'self-rotation function' (equation 2 above) determines these relationships by

comparing the Patterson P_1 (with coefficients $\mathbf{F}_h^2$) with itself (the same coefficients now identified by $\mathbf{F}_p^2$).

Thus, the first step in the MR process for either of these two types of situations requires an identical computational procedure. Similar arguments apply to the more difficult translational problem.

In the third stage of MR, a known structure is used for initiating phasing. The known structure can be a homologous molecule or it can be a simple assumption about the nature of the molecule, such that it is a roughly spherical object. The latter is a minimal assumption when NCS is used for 'ab initio' phasing. Subsequently, these phases are usually refined using NCS electron-density averaging (if available), plus solvent flattening (NCS implies that part of the cell has no NCS and must, therefore, be solvent flattened) or by only solvent flattening. Thus, again all the MR applications discussed in this paper require identical calculations, with the case of the NCS redundancy of 1 being a special situation. Once atomic positions can be picked (possibly after phase extension during the density-modification procedure), normal crystallographic refinement can take over.

APPENDIX *B*
The equivalence of real- and reciprocal-space phase determination in the presence of NCS restraints

It will be assumed that the electron density $\rho(\mathbf{x})$ at the NCS-related positions $\mathbf{x}_n(n = 1,N)$ should be equal. Hence, the averaged electron density at $\mathbf{x}$ is given by

$$\rho_{\mathrm{avg}}(\mathbf{x}) = \frac{1}{N} \sum_{n=1}^{N} \rho(\mathbf{x}_n). \qquad (4)$$

The NCS relating the different densities is given by

$$\mathbf{x}_n = [C_n]\mathbf{x}_1 + \mathbf{d}_n, \qquad (5)$$

where $[C_n]$ is the rotation matrix relating the nth density to the reference density 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

$$\rho_{\mathrm{avg}}(\mathbf{x}) = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{V} \sum_{h} \mathbf{F}_h \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}_n), \qquad (6)$$

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

$$\mathbf{F}_p = \sum_{n=1}^{N} \int_{U_n} \rho_{\mathrm{avg}}(\mathbf{x}_1) \exp(2\pi i \mathbf{p} \cdot \mathbf{x}_n) \, \mathrm{d}\mathbf{x}_n, \qquad (7)$$

where U_n bounds the volume containing the nth molecular subunit. By substitution of (6) into (7), it follows that

$$\mathbf{F}_p = \frac{1}{NV} \sum_{h} \mathbf{F}_h \sum_{n=1}^{N} \exp(-2\pi i \mathbf{h} \cdot \mathbf{d}_n)$$
$$\times \int_{U_n} \exp\{2\pi i(-\mathbf{h}[C_n] + \mathbf{p}) \cdot \mathbf{x}_n\} \, \mathrm{d}\mathbf{x}_n. \qquad (8)$$

If we now define

$$G_{hpn} = \frac{1}{U} \int_{U_n} \exp\{2\pi i(\mathbf{p} - \mathbf{h}[C_n]) \cdot \mathbf{x}_n\} \, \mathrm{d}\mathbf{x}_n, \qquad (9)$$

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

$$\mathbf{T}_{hpn} = \exp(-2\pi i \mathbf{h} \cdot \mathbf{d}_n), \qquad (10)$$

then (8) simplifies to

$$\mathbf{F}_p = \frac{U}{NV} \sum_{h} \mathbf{F}_h \sum_{n=1}^{N} G_{hpn} \mathbf{T}_{hpn} \qquad (11)$$

or

$$\mathbf{F}_p = \sum_{n=1}^{N} \left[\frac{U}{NV} \sum_{h} \mathbf{F}_h G_{hpn} \mathbf{T}_{hpn} \right] = \frac{U}{NV} \sum_{n=1}^{N} \mathbf{F}_{\mathbf{h}'_n}, \qquad (12)$$

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 non-integral reciprocal-lattice point $\mathbf{h}'_n$ corresponding to the rotation of $\mathbf{p}$ by the nth NCS element. Hence, $\mathbf{F}_p$ represents the complex averaging of structure factors at the N non-crystallographically equivalent positions in reciprocal space.

If we simplify (11) by putting

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

then

$$\mathbf{F}_p = \frac{U}{NV} \sum_{h} \mathbf{F}_h \mathbf{a}_{hp}. \qquad (13)$$

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 dependent only on a knowledge of the orientation, position and extent of the NCS elements. The molecular-replacement equations are exact other than the assumptions that the NCS 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 (13) will produce an improved set of phases on the left-hand side in a process that is entirely equivalent and the same as the real-space averaging and back-transformation procedure (Arnold *et al.*, 1987). 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 non-linear (Nordman, 1980) interpolation to obtain the value of electron density at non-integral 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 (13) 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 $\mathbf{H} \cdot \mathbf{R}$ of G is given by $(n/4R)R$, where n is the number of reciprocal-lattice points represented by the difference $\mathbf{p} - \mathbf{h}[C^T]$. Now, G for a spherical envelope, 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 (13) has a half-width when $n/4 = 0.7$ or $n = 2.8$, which 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 non-integral grid point.

Omission of a significant coefficient $\mathbf{F}_h$ on the right-hand side of (13) 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 calculations of $\mathbf{F}_h$ from the molecular-replacement equation when $p = h$. This process is identical to the inclusion of 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 and faster phase determination. Furthermore, it is now seen that if there were no observed amplitudes, then the value of $\mathbf{F}_p$ on substituting all the F_{calc} values on the right would not change from the previous 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 for a structure factor $\mathbf{F}_p$ at the exact limit of resolution will have half its terms missing and cannot give a good estimate of $\mathbf{F}_p$. The lack of agreement between observed and calculated amplitudes at the limit of resolution is thus 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 indeed is found in practice (Arnold *et al.*, 1987; Luo *et al.*, 1989).

References

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.

Blow, D. M. & Rossmann, M. G. (1961). *Acta Cryst.* **14**, 1195–1202.

Bricogne, G. (1974). *Acta Cryst.* A**30**, 395–405.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. & Rossmann, M. G. (1973). *Proc. Natl Acad. Sci. USA*, **70**, 3052–3054.

Colman, P. M. (1974). *Z. Kristallogr.* **140**, 344–349.

Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.

Crick, F. H. C. & Watson, J. D. (1956). *Nature (London)*, **177**, 473–475.

Crowther, R. A. (1967). *Acta Cryst.* **22**, 758–764.

Crowther, R. A. (1969). *Acta Cryst.* B**25**, 2571–2580.

Dodson, E. J., Gover, S. & Wolf, W. (1992). Editors. *Proceedings of the CCP4 Study Weekend. Molecular Replacement.* Warrington: Daresbury Laboratory.

Dodson, E., Harding, M. M., Hodgkin, D. C. & Rossmann, M. G. (1966). *J. Mol. Biol.* **16**, 227–241.

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.

Hoppe, W. (1957*a*). *Acta Cryst.* **10**, 750–751.

Hoppe, W. (1957*b*). *Z. Elektrochem.* **61**, 1076–1083.

Johnson, J. E., Akimoto, T., Suck, D., Rayment, I. & Rossmann, M. G. (1976). *Virology*, **75**, 394–400.

Lawrence, M. C. (1991). *Quart. Rev. Biophys.* **24**, 399–424.

Luo, M., Vriend, G., Kamer, G. & Rossmann, M. G. (1989). *Acta Cryst.* B**45**, 85–92.

Machin, P. (1985). Editor. *Proceedings of the CCP4 Study Weekend. Molecular Replacement.* Warrington: Daresbury Laboratory.

Main, P. (1967). *Acta Cryst.* **23**, 50–54.

Main, P. & Rossmann, M. G. (1966). *Acta Cryst.* **21**, 67–72.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Nordman, C. E. (1980). *Acta Cryst.* A**36**, 747–754.

Perutz, M. F. (1952). *Proc. R. Soc. London Ser. A*, **213**, 425.

Perutz, M. F. (1954). *Proc. R. Soc. London Ser. A*, **225**, 264–286.

Rayment, I. (1983). *Acta Cryst.* A**39**, 102–116.

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

Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.

Rossmann, M. G. (1995). *Curr. Opin. Struct. Biol.* **5**, 650–655.

Rossmann, M. G. & Arnold, E. (1993). *International Tables for Crystallography*, edited by U. Shmueli, Vol. B, pp. 230–263. Dordrecht: Kluwer Academic Publishers.

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.

Simpson, A. A., Tao, Y., Leiman, P. G., Badasso, M. O., He, Y., Jardine, P. J., Olson, N. H., Morais, M. C., Grimes, S., Anderson, D. L., Baker, T. S. & Rossmann, M. G. (2000). *Nature (London)*, **408**, 745–750.

Tong, L. (1993). *J. Appl. Cryst.* **26**, 748–751.

Tong, L. & Rossmann, M. G. (1995). *Acta Cryst.* D**51**, 347–353.