where $V_{\mathbf{h}}$ is the volume of cell (**h**) and $\mathbf{S}_n$ is the position, in the nth crystallographic asymmetric unit, of cell (**p**) corresponding to **S** in known cell (**h**). Let

$$A_{p,\,n}\exp(i\gamma_n) = \sum_{\mathbf{h}} \mathbf{F}_{\mathbf{h}} G_{\mathbf{hp}_n} \exp(-2\pi i \mathbf{h}\cdot\mathbf{S}),$$

which are the coefficients of the molecular transform for the known molecule placed into the nth asymmetric unit of the **p** cell. Thus

$$\mathbf{F}_{\mathbf{p}}(\mathbf{S}) = \frac{U}{V_{\mathbf{h}}} \sum_{n=1}^{N} A_{\mathbf{p},\,n} \exp[i(\gamma + 2\pi \mathbf{p}\cdot\mathbf{S}_n)]$$

or

$$\mathbf{F}_{\mathbf{p}}(\mathbf{S}) = \frac{U}{V_{\mathbf{h}}} \sum_{n=1}^{N} A_{\mathbf{p},\,n} \exp[i(\gamma_n + 2\pi \mathbf{p}_n\cdot\mathbf{S})],$$

where $\mathbf{p}_n = [\boldsymbol{C}_n^T]\mathbf{p}$ and $\mathbf{S} = \mathbf{S}_1$. Hence

$$|\mathbf{F}_{\mathbf{p}}(\mathbf{S})|^2 = \left(\frac{U}{V_{\mathbf{h}}}\right)^2 \sum_n \sum_m \big(A_{\mathbf{p},\,n} A_{\mathbf{p},\,m} \times \exp\{i[2\pi(\mathbf{p}_n - \mathbf{p}_m)\cdot\mathbf{S} + (\gamma_n - \gamma_m)]\}\big),$$

and then from (2.3.7.3)

$$T(\mathbf{S}) = \left(\frac{U}{V_{\mathbf{h}}}\right)^2 \sum_{\mathbf{p}} \sum_n \sum_m \Big(|\mathbf{F}_{\mathbf{p},\,\mathrm{obs}}|^2 A_{\mathbf{p},\,n} A_{\mathbf{p},\,m} \times \exp\{i[2\pi(\mathbf{p}_n - \mathbf{p}_m)\cdot\mathbf{S} + (\gamma_n - \gamma_m)]\}\Big), \qquad (2.3.7.4)$$

which is a Fourier summation with known coefficients $\{|\mathbf{F}_{\mathbf{p},\,\mathrm{obs}}|^2 A_{\mathbf{p},\,n} A_{\mathbf{p},\,m} \times \exp[i(\gamma_n - \gamma_m)]\}$ such that $T(\mathbf{S})$ will be a maximum at the correct molecular position.

Terms with $n = m$ in expression (2.3.7.4) can be omitted as they are independent of **S** and only contribute a constant to the value of $T(\mathbf{S})$. For terms with $n \neq m$, the indices take on special values. For instance, if the **p** cell is monoclinic with its unique axis parallel to **b** such that $\mathbf{p}_1 = (p, q, r)$ and $\mathbf{p}_2 = (\bar{p}, q, \bar{r})$, then $\mathbf{p}_1 - \mathbf{p}_2$ would be $(2p,\ 0,\ 2r)$. Hence, $T(\mathbf{S})$ would be a two-dimensional function consistent with the physical requirement that the translation component, parallel to the twofold monoclinic axis, is arbitrary.

Crowther & Blow (1967) show that if $\mathbf{F}_M$ are the structure factors of a known molecule correctly oriented within the cell of the unknown structure at an arbitrary molecular origin, then (altering the notation very slightly from above)

$$T(\mathbf{S}) = \sum_{\mathbf{p}} |\mathbf{F}_{\mathrm{obs}}(\mathbf{p})|^2 \mathbf{F}_M(\mathbf{p}) \mathbf{F}_M^*(\mathbf{p}[\boldsymbol{C}]) \exp(-2\pi i \mathbf{p}\cdot\mathbf{S}),$$

where $[\boldsymbol{C}]$ is a crystallographic symmetry operator relative to which the molecular origin is to be determined. This is of the same form as (2.3.7.4) but concerns the special case where the **h** cell, into which the known molecule was placed, has the same dimensions as the **p** cell.

R-factor calculations are sometimes used to determine the position of a known molecular fragment in an unknown cell, particularly if only one parameter is being searched. Such calculations are computationally less convenient than the Fourier methods described above, but can be more sensitive. All these methods can be improved by simultaneous consideration of packing requirements of the molecular fragments (Harada *et al.*, 1981; Hendrickson & Ward, 1976; Rabinovich & Shakked, 1984). Indeed, packing considerations can frequently limit the search volume very considerably.

2.3.7.4. *Position of a noncrystallographic symmetry element in a poorly defined electron-density map*

If an initial set of poor phases, for example from an SIR derivative, are available and the rotation function has given the orientation of a noncrystallographic rotation axis, it is possible to search the electron-density map systematically to determine the translation axis position. The translation function must, therefore, measure the quality of superposition of the poor electron-density map on itself. Hence $\mathbf{S}_x = \mathbf{S}_{x'} = \mathbf{S}$ and the function (2.3.7.1) now becomes

$$T(\mathbf{S}) = \frac{2}{V_{\mathbf{h}}^2} \sum_{\mathbf{h}} \sum_{\mathbf{p}} |\mathbf{F}_{\mathbf{h}}||\mathbf{F}_{\mathbf{p}}| G_{\mathbf{hp}} \cos[\alpha_{\mathbf{h}} + \alpha_{\mathbf{p}} - 2\pi(\mathbf{h} + \mathbf{p})\cdot\mathbf{S}].$$

This real-space translation function has been used successfully to determine the intermolecular dyad axis for α-chymotrypsin (Blow *et al.*, 1964) and to verify the position of immunoglobulin domains (Colman & Fehlhammer, 1976).

2.3.8. Molecular replacement

2.3.8.1. *Using a known molecular fragment*

The most straightforward application of the molecular-replacement method occurs when the orientation and position of a known molecular fragment in an unknown cell have been previously determined. The simple procedure is to apply the rotation and translation operations to the known fragment. This will place it into one 'standard' asymmetric unit of the unknown cell. Then the crystal operators (assuming no further noncrystallographic operators are present in the unknown cell) are applied to generate the complete unit cell of the unknown structure. Structure factors can then be calculated from the rotated and translated known molecule into the unknown cell. The resultant model can be refined in numerous ways.

More generally, consider a molecule placed in any crystal cell (**h**), within which coordinate positions shall be designated by **x**. Let the corresponding structure factors be $\mathbf{F}_{\mathbf{h}}$. It is then possible to compute the structure factors $\mathbf{F}_{\mathbf{p}}$ for another cell (**p**) into which the same molecule has been placed N times related by the crystallographic symmetry operators $[\boldsymbol{C}_1], \mathbf{d}_1;\ [\boldsymbol{C}_2], \mathbf{d}_2;\ \ldots;\ [\boldsymbol{C}_N], \mathbf{d}_N$. Let the electron density at a point $\mathbf{y}_1$ in the first crystallographic asymmetric unit be spatially related to the point $\mathbf{y}_n$ in the nth asymmetric unit of the **p** crystal such that

$$\rho(\mathbf{y}_n) = \rho(\mathbf{y}_1), \qquad (2.3.8.1)$$

where

$$\mathbf{y}_n = [\boldsymbol{C}_n]\mathbf{y}_1 + \mathbf{d}_n. \qquad (2.3.8.2)$$

From the definition of a structure factor,

$$\mathbf{F}_{\mathbf{p}} = \sum_{n=1}^{N} \int_U \rho(\mathbf{y}_n) \exp(2\pi i \mathbf{p}\cdot\mathbf{y}_n)\, \mathrm{d}\mathbf{y}_n, \qquad (2.3.8.3)$$

where the integral is taken over the volume U of one molecule. But since each molecule is identical as expressed in equation (2.3.8.1) and since (2.3.8.2) can be substituted in equation (2.3.8.3), we have

$$\mathbf{F}_{\mathbf{p}} = \sum_{n=1}^{N} \int_U \rho(\mathbf{y}_1) \exp[2\pi i \mathbf{p}\cdot([\boldsymbol{C}_n]\mathbf{y}_1 + \mathbf{d}_n)]\, \mathrm{d}\mathbf{y}_1. \qquad (2.3.8.4)$$

Now let the molecule in the **h** crystal be related to the molecule in the first asymmetric unit of the **p** crystal by the noncrystallographic symmetry operation

$$\mathbf{x} = [\boldsymbol{C}]\mathbf{y} + \mathbf{d}, \qquad (2.3.8.5)$$

which implies

$$\rho(\mathbf{x}) = \rho(\mathbf{y}_1) = \rho(\mathbf{y}_2) = \ldots. \qquad (2.3.8.6)$$

Furthermore, in the **h** cell

$$\rho(\mathbf{x}) = \frac{1}{V_{\mathbf{h}}} \sum_{\mathbf{h}} \mathbf{F}_{\mathbf{h}} \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}), \qquad (2.3.8.7)$$

and thus, by combining with (2.3.8.5), (2.3.8.6) and (2.3.8.7),

$$\rho(\mathbf{y}_1) = \frac{1}{V_{\mathbf{h}}} \sum_{\mathbf{h}} \mathbf{F}_{\mathbf{h}} \exp[-2\pi i(\mathbf{h}[\boldsymbol{C}] \cdot \mathbf{y}_1 + \mathbf{h} \cdot \mathbf{d})]. \qquad (2.3.8.8)$$

Now using (2.3.8.4) and (2.3.8.8) it can be shown that

$$\mathbf{F}_{\mathbf{p}} = \frac{U}{V_{\mathbf{h}}} \sum_{\mathbf{h}} \mathbf{F}_{\mathbf{h}} \sum_{n=1}^{N} G_{\mathbf{hp}_n} \exp[2\pi i(\mathbf{p} \cdot \mathbf{S}_n - \mathbf{h} \cdot \mathbf{S})], \qquad (2.3.8.9)$$

where

$$UG_{\mathbf{hp}_n} = \int_U \exp[2\pi i(p[\boldsymbol{C}_n] - \mathbf{h}[\boldsymbol{C}]) \cdot \mathbf{u}] \, \mathrm{d}\mathbf{u}. \qquad (2.3.8.10)$$

S is a chosen molecular origin in the **h** crystal and $\mathbf{S}_n$ is the corresponding molecular position in the nth asymmetric unit of the **p** crystal.

2.3.8.2. *Using noncrystallographic symmetry for phase improvement*

The use of noncrystallographic symmetry for phase determination was proposed by Rossmann & Blow (1962, 1963) and subsequently explored by Crowther (1967, 1969) and Main & Rossmann (1966). These methods were developed in reciprocal space and were primarily concerned with *ab initio* phase determination. Real-space averaging of electron density between noncrystallographically related molecules was used in the structure determination of deoxyhaemoglobin (Muirhead *et al*., 1967) and of α-chymotrypsin (Matthews *et al*., 1967). The improvement derived from the averaging between the two noncrystallographic units was, however, not clear in either case. The first obviously successful application was in the structure determination of lobster glyceraldehyde-3-phosphate dehydrogenase (Buehner *et al*., 1974; Argos *et al*., 1975), where the tetrameric molecule of symmetry 222 occupied one crystallographic asymmetric unit. The improvement in the essentially SIR electron-density map was considerable and the results changed from uninterpretable to interpretable. The uniqueness and validity of the solution lay in the obvious chemical correctness of the polypeptide fold and its agreement with known amino-acid-sequence data. In contrast to the earlier reciprocal-space methods, noncrystallographic symmetry was used as a method to improve poor phases rather than to determine phases *ab initio*.

Many other applications followed rapidly, aided greatly by the versatile techniques developed by Bricogne (1976). Of particular interest is the application to the structure determination of hexokinase (Fletterick & Steitz, 1976), where the averaging occurred both between different crystal forms and within the same crystal.

The most widely used procedure for real-space averaging is the 'double sorting' technique developed by Bricogne (1976) and also by Johnson (1978). An alternative method is to maintain the complete map stored in the computer (Nordman, 1980*b*). This avoids the sorting operation, but is only possible given a very large computer or a low-resolution map containing relatively few grid points.

Bricogne's double sorting technique involves generating real-space non-integral points (D_i) which are related to integral grid points (I_i) in the cell asymmetric unit by the noncrystallographic symmetry operators. The elements of the set D_i are then brought back to their equivalent points in the cell asymmetric unit (D_i') and sorted by their proximity to two adjacent real-space sections. The set I_i', calculated on a finer grid than I_i and stored in the computer memory two sections at a time, is then used for linear interpolation to determine the density values at D_i' which are successively stored and summed in the related array I_i. A count is kept of the number of densities received at each I_i, resulting in a final averaged aggregate, when all real-space sections have been utilized. The density to be assigned outside the molecular envelope (defined with respect to the set I_i) is determined by averaging the density of all unused points in I_i. The grid interval for the set I_i' should be about one-sixth of the resolution to avoid serious errors from interpolation (Bricogne, 1976). The grid point separation in the set I_i need only be sufficient for representation of electron density, or about one-third of the resolution.

Molecular replacement in real space consists of the following steps (Table 2.3.8.1): (*a*) calculation of electron density based on a starting phase set and observed amplitudes; (*b*) averaging of this density among the noncrystallographic asymmetric units or molecular copies in several crystal forms, a process which defines a molecular envelope as the averaging is only valid within the range of the noncrystallographic symmetry; (*c*) reconstructing the unit cell based on averaged density in every noncrystallographic asymmetric unit; (*d*) calculating structure factors from the reconstructed cell; (*e*) combining the new phases with others to obtain a weighted best-phase set; and (*f*) returning to step (*a*) at the previous or an extended resolution. Decisions made in steps (*b*) and (*e*) determine the rate of convergence (see Table 2.3.8.1) to a solution (Arnold *et al*., 1987).

The power of the molecular-replacement procedure for either phase improvement or phase extension depends on the number of

Table 2.3.8.1. *Molecular replacement: phase refinement as an iterative process*

(*A*)	$\mathbf{F}_{\rm obs}, \alpha_n', m_n' \rightarrow \rho_n$
(*B*)	$\rho_n \rightarrow \rho_n$(modified) (i) Use of noncrystallographic symmetry operators (ii) Definition of envelope limiting volume within which noncrystallographic symmetry is valid (iii) Adjustment of solvent density* (iv) Use of crystallographic operators to reconstruct modified density into a complete cell
(*C*)	ρ_n(modified) $\rightarrow \mathbf{F}_{\rm calc,\,n+1}$; $\alpha_{\rm calc,\,n+1}$
(*D*)	$(\mathbf{F}_{\rm calc,\,n+1}, \alpha_{\rm calc,\,n+1}) + (\mathbf{F}_{\rm obs}, \alpha_0) \rightarrow \mathbf{F}_{\rm obs}, \alpha_{n+1}', m_{n+1}'$ (i) Assessment of reliability of new phasing set α_{n+1} in relation to original phasing set $\alpha_0(w)$ (ii) Use of figures of merit m_0, m_{n+1} and reliability w to determine modified phasing set α_{n+1}', m_{n+1}'† (iii) Consideration of α_{n+1} and m_{n+1} where there was no prior knowledge of (*a*) $\mathbf{F}_{\rm obs}$ (*e.g.* very low order reflections or uncollected data) (*b*) α_0 (*e.g.* no isomorphous information or phase extension)
(*E*)	Return to step (*A*) with α_{n+1}', m_{n+1}' and a possibly augmented set of $\mathbf{F}_{\rm obs}$.

* Wang (1985); Bhat & Blow (1982); Collins (1975); Schevitz *et al*. (1981); Hoppe & Gassmann (1968).
† Rossmann & Blow (1961); Hendrickson & Lattman (1970).

noncrystallographic asymmetric units, the size of the excluded volume expressed in terms of the ratio $(V - UN)/V$ and the magnitude of the measurement error on the structure amplitudes. Crowther (1967, 1969) and Bricogne (1974) have investigated the dependence on the number of noncrystallographic asymmetric units and conclude that three or more copies are sufficient to ensure convergence of an iterative phase improvement procedure in the absence of errors on the structure amplitudes. As with the analogous case of isomorphous replacement in which three data sets ensure reasonable phase determination, additional copies will enhance the power of the method, although their usefulness is subject to the law of diminishing returns. Another example of this principle is the sign determination of the *h*0*l* reflections of horse haemoglobin (Perutz, 1954) in which seven shrinkage stages constituted the sampling of the transform of a single copy.

Procedures for real-space averaging have been used extensively with great success. The interesting work of Wilson *et al*. (1981) is noteworthy for the continuous adjustment of molecular envelope with increased map definition. Furthermore, the analysis of complete virus structures has only been possible as a consequence of this technique (Bloomer *et al*., 1978; Harrison *et al*., 1978; Abad-Zapatero *et al*., 1980; Liljas *et al*., 1982). Although the procedure has been used primarily for phase improvement, apparently successful attempts have been made at phase extension (Nordman, 1980*b*; Gaykema *et al*., 1984; Rossmann *et al*., 1985). *Ab initio* phasing of glyceraldehyde-3-phosphate dehydrogenase (Argos *et al*., 1975) was successfully attempted by initially filling the known envelope with uniform density to determine the phases of the innermost reflections and then gradually extending phases to 6.3 Å resolution. Johnson *et al*. (1976) used the same procedure to determine the structure of southern bean mosaic virus to 22.5 Å resolution. Particularly impressive was the work on polyoma virus (Rayment *et al*., 1982; Rayment, 1983; Rayment *et al*., 1983) where crude initial models led to an entirely unexpected breakdown of the Caspar & Klug (1962) concept of quasi-symmetry. *Ab initio* phasing has also been used by combining the electron-diffraction projection data of two different crystal forms of bacterial rhodopsin (Rossmann & Henderson, 1982).

2.3.8.3. *Equivalence of real- and reciprocal-space molecular replacement*

Let us proceed in reciprocal space doing exactly the same as is done in real-space averaging. Thus

$$\rho_{\rm AV}(\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N}\rho(\mathbf{x}_n),$$

where

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

Therefore,

$$\rho_{\rm AV}(\mathbf{x}) = \frac{1}{N}\sum_{N}\frac{1}{V}\left[\sum_{\mathbf{h}}\mathbf{F}_{\mathbf{h}}\exp(2\pi i\mathbf{h}\cdot\mathbf{x}_n)\right].$$

The next step is to perform the back-transform of the averaged electron density. Hence,

$$\mathbf{F}_{\mathbf{p}} = \int_U \rho_{\rm AV}(\mathbf{x})\exp(-2\pi i\mathbf{p}\cdot\mathbf{x})\,{\rm d}\mathbf{x},$$

where U is the volume within the averaged part of the cell. Hence, substituting for $\rho_{\rm AV}$,

$$\mathbf{F}_{\mathbf{p}} = \int_U\left[\frac{1}{NV}\sum_{N}\sum_{\mathbf{h}}\mathbf{F}_{\mathbf{h}}\exp(2\pi i\mathbf{h}\cdot\mathbf{x}_n)\right]\exp(-2\pi i\mathbf{p}\cdot\mathbf{x})\,{\rm d}\mathbf{x},$$

which is readily simplified to

$$\mathbf{F}_{\mathbf{p}} = \frac{U}{NV}\sum_{\mathbf{h}}\mathbf{F}_{\mathbf{h}}\sum_{N}G_{\mathbf{hp}n}\exp(2\pi i\mathbf{h}\cdot\mathbf{d}_n).$$

Setting

$$\mathbf{B}_{\mathbf{hp}} = \frac{U}{NV}\sum_{N}G_{\mathbf{hp}n}\exp(2\pi i\mathbf{h}\cdot\mathbf{d}_n),$$

the molecular-replacement equations can be written as

$$\mathbf{F}_{\mathbf{p}} = \sum_{\mathbf{h}}\mathbf{B}_{\mathbf{hp}}\mathbf{F}_{\mathbf{h}} \qquad (2.3.8.11)$$

(Main & Rossmann, 1966), or in matrix form

$$\mathbf{F} = [\boldsymbol{B}]\mathbf{F},$$

which is the form of the equations used by Main (1967) and by Crowther (1967). Colman (1974) arrived at the same conclusions by an application of Shannon's sampling theorem. It should be noted that the elements of $[\boldsymbol{B}]$ are dependent only on knowledge of the noncrystallographic symmetry and the volume within which it is valid. Substitution of approximate phases into the right-hand side of (2.3.8.11) produces a set of calculated structure factors exactly analogous to those produced by back-transforming the averaged electron density in real space. The new phases can then be used in a renewed cycle of molecular replacement.

Computationally, it has been found more convenient and faster to work in real space. This may, however, change with the advent of vector processing in 'supercomputers'. Obtaining improved phases by substitution of current phases on the right-hand side of the molecular-replacement equations (2.3.8.1) seems less cumbersome than the repeated forward and backward Fourier transformation, intermediate sorting, and averaging required in the real-space procedure.

2.3.9. Conclusions

Complete interpretation of Patterson maps is no longer used frequently in structure analysis, although most determinations of heavy-atom positions of isomorphous pairs are based on Patterson analyses. Incorporation of the Patterson concept is crucial in many sophisticated techniques essential for the solution of complex problems, particularly in the application to biological macromolecular structures. Patterson techniques provide important physical insights in a link between real- and reciprocal-space formulation of crystal structures and diffraction data.

2.3.9.1. *Update*

This article was originally completed in January 1986. Since then, some advances have occurred. In particular, the use of real-space averaging between noncrystallographically related electron density within the crystallographic asymmetric unit has become an accepted way of extending phase information to higher resolution, particularly for complex structures such as viruses (Gaykema *et al*., 1984; Rossmann *et al*., 1985; Hogle *et al*., 1985; Arnold *et al*., 1987; Hosur *et al*., 1987; Luo *et al*., 1987; Acharya *et al*., 1989). The power of this procedure has been examined theoretically by Arnold & Rossmann (1986).

The availability of fast computers with large random access memories and even larger disk storage also makes many of the techniques considered here commonplace and no longer subject to

references