Acta Cryst. (1995). D**51**, 347–353

Reciprocal-Space Molecular-Replacement Averaging

BY LIANG TONG* AND MICHAEL G. ROSSMANN†

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

(*Received* 27 *April* 1994; *accepted* 19 *September* 1994)

Abstract

The molecular-replacement equations, in which electron-density averaging and skew averaging have been unified, were used in reciprocal space to refine and extend the resolution of phased reflections. A procedure has been developed for the treatment of molecular envelopes of general shape. The equations were successfully applied to the reflection data of bacteriophage φX174 (60-fold redundancy). Truncation of the G diffraction function beyond the first few nodes did not have a significant effect on the quality of the molecular-replacement equations. Reciprocal-space molecular-replacement averaging should prove to be a useful alternative to real-space averaging. Strategies are discussed that are possible only in reciprocal space.

Introduction

Phase determination in the presence of non-crystallographic redundancy generated by local molecular symmetry is now a well established technique (Lawrence, 1991). It has been used to extend phases derived from crude, 20 Å resolution, hollow-shell models to at least 3.5 Å resolution in a number of viral structures [cowpea chlorotic mottle virus (Speir, Munshi, Baker & Johnson, 1993), canine parvovirus (Tsao, Chapman, Wu, Agbandje, Keller & Rossmann, 1992), φX174 (McKenna, Xia, Willingmann, Ilag & Rossmann, 1992), foot-and-mouth disease virus (Acharya, Fry, Stuart, Fox, Rowlands & Brown, 1989) and MS2 (Valegård, Liljas, Fridborg & Unge, 1990)]. The power of non-crystallographic restraints on phase determination was first fully demonstrated by real-space averaging of structurally equivalent sections of electron density (Matthews, Sigler, Henderson & Blow, 1967; Buehner, Ford, Moras, Olsen & Rossmann, 1974; Fletterick & Steitz, 1976). While initially real-space averaging was used only for phase improvement, its use for phase extension to higher resolution is more recent. Gaykema, Hol, Vereijken, Soeter, Bak & Beintema (1984) extended the phases from 4.0 to 3.2 Å resolution in the structure determination of hemocyanin. Subsequently, a rather poor phase set for human rhinovirus 14 at 6 Å resolution was extended to 3.5 Å resolution (Rossmann, Arnold, Erickson, Frankenberger, Griffith, Hecht, Johnson, Kamer, Luo, Mosser, Rueckert, Sherry & Vriend, 1985; Arnold, Vriend, Luo, Griffith, Kamer, Erickson, Johnson & Rossmann, 1987). Similar results were also reported shortly afterwards in the structure determination of poliovirus (Hogle, Chow & Filman, 1985). Nowadays, it is common to use non-crystallographic symmetry for phase extension where the starting phase set is based merely on a crude model, such as a hollow sphere representing a virus. A variety of computer programs have been written that can cope with anticipated real-space problems (Bricogne, 1976; Johnson, 1978; Jones, 1992; Rossmann, McKenna, Tong, Xia, Dai, Wu, Choi & Lynch, 1992; Rossmann, McKenna, Tong, Xia, Dai, Wu, Choi, Marinescu & Lynch, 1992).

Although electron-density averaging realizes the original hope of using local non-crystallographic symmetry for *ab initio* phase determination (Rossmann & Blow, 1962, 1963), the process was originally conceived in reciprocal space (Main & Rossmann, 1966; Crowther, 1967, 1969). It has been shown (Colman, 1974; Rossmann, 1990; Main, 1967; Bricogne, 1974) that real-space electron-density averaging is equivalent to solving the simple algebraic molecular-replacement equations (Main & Rossmann, 1966; Crowther, 1967, 1969; Rossmann, 1990) for the unknown phases. Furthermore, the reciprocal-space equations provide insight into the nature of the non-crystallographic averaging process, whether it be in reciprocal or real space.

We first derive a general reciprocal-space molecular-replacement averaging procedure and then describe its application to the reflection data of φX174. We also suggest how the reciprocal-space technique might be further improved as it provides opportunities not so easily implemented in real space. We have the psychological advantage over earlier reciprocal-space attempts (Main, 1967; Crowther, 1969) in that we know that non-crystallographic symmetry is, indeed, an exceedingly powerful phase-determining tool and can readily yield accurate phases.

Averaging electron density

Three different coordinate systems will be used.

* Present address: Department of Medicinal Chemistry, Boehringer Ingelheim Pharmaceuticals, Inc., 900 Ridgebury Road, PO Box 368, Ridgefield, CT 06877, USA.

† To whom correspondence should be addressed.

ISSN 0907-4449

(1) The h cell, with fractional coordinates $\mathbf{x}$, contains the *input* electron density. The orthogonalized Cartesian coordinates in this cell will be referred to as $\mathbf{X}$. Structure factors of this cell are defined as $\mathbf{F}_h$.

(2) The p cell, with fractional coordinates $\mathbf{y}$, contains the *output* (usually averaged) electron density. The orthogonalized Cartesian coordinates in this cell will be referred to as $\mathbf{Y}$. Structure factors of this cell are defined as $\mathbf{F}_p$.

(3) A reference Cartesian coordinate system $\mathbf{U}$ with respect to which the symmetry elements of the macromolecular assembly will be defined. For example, an icosahedral object might have its orthogonal twofold axes along the three principal directions of this coordinate system.

The averaging process will take the electron density in the h cell and place it at equivalent points in the p cell for the purpose of computing structure factors. The structure factors in the p cell are given by,

$$\mathbf{F}_p = \int_{V_p} \rho_p(\mathbf{y}) \exp(2\pi i \mathbf{p}.\mathbf{y}) d\mathbf{y}, \tag{1}$$

where V_p is the volume of the p cell and $\rho_p(\mathbf{y})$ is the density distribution in the p cell. If the p cell contains one molecular assembly in each of m crystallographic asymmetric units, and if the intervening electron density is assumed to be zero, then

$$\mathbf{F}_p = \sum_m \int_{U_m} \rho_p(\mathbf{y}_m) \exp(2\pi i \mathbf{p}.\mathbf{y}_m) d\mathbf{y}, \tag{2}$$

where U_m is the volume of the molecular assembly and $\mathbf{y}_m$ is a position vector in the mth crystallographic asymmetric unit. The implied simplifying assumption that the density is zero between particles can be readily modified to be a non-zero constant. Similarly, the derivation can be extended to include two or more particles per crystallographic asymmetric unit related by improper symmetry.

The density in the p cell, ρ_p, can be replaced by the equivalent density ρ_h of the h cell such that,

$$\mathbf{F}_p = \sum_m \int_{U_m} \rho_h(\mathbf{x}) \exp(2\pi i \mathbf{p}.\mathbf{y}_m) d\mathbf{y}. \tag{3}$$

The input h cell might have the same shape and symmetry as the output cell, but the output cell is based on solvent flattening and averaging of the non-crystallographically related densities in the h cell. Alternatively, the input cell might be different crystal forms of the same molecular substance which are to be averaged into a common orientation in the p cell. The relationship between the h and p cell is given by,

$$\mathbf{x} = [\Delta_m]\mathbf{y}_m + \delta_m. \tag{4}$$

Here $[\Delta_m]$ and δ_m are the rotational and translational relationships between the mth asymmetric unit in the p cell with the reference particle in the h cell. If the particle in the mth asymmetric unit of the p cell is centered at $\mathbf{s}_{p,m}$ and if the reference particle in the h cell is at $\mathbf{s}_h$, then it follows from (4),

$$\delta_m = \mathbf{s}_h - [\Delta_m]\mathbf{s}_{p,m}. \tag{5}$$

Now, if there are N non-crystallographically equivalent densities in the particle in the h cell, and if the nth position is at $\mathbf{x}_n$, then the average of all these density values can be used to calculate $\mathbf{F}_p$. From (3) it then follows that

$$\mathbf{F}_p = \sum_m \int_{U_m} \left\{ \left[\sum_n \rho_h(\mathbf{x}_n) \right] / N \right\} \exp(2\pi i \mathbf{p}.\mathbf{y}_m) d\mathbf{y}. \tag{6}$$

But the electron density in the h cell can be expressed as the Fourier summation,

$$\rho_h(\mathbf{x}) = (1/V_h) \sum_h \mathbf{F}_h \exp(-2\pi i \mathbf{h}.\mathbf{x}), \tag{7}$$

where V_h is the volume of the h cell.

Extending (5), and taking $[\Delta_{m,n}]$ to be the rotation matrix which equivalences the point $\mathbf{y}_m$ in the p cell to the reference non-crystallographic asymmetric unit in the h cell, using (6) and (7),

$$\mathbf{F}_p = (1/NV_h) \sum_m \sum_n \sum_h \mathbf{F}_h \exp\left[-2\pi i \mathbf{h}.(\mathbf{s}_h - [\Delta_{m,n}]\mathbf{s}_{p,m})\right] \times \int_{U_m} \exp\left[2\pi i (\mathbf{p} - \mathbf{h}[\Delta_{m,n}]\mathbf{y}_m)\right] d\mathbf{y}.$$

Now setting $\mathbf{y}' = \mathbf{y}_m - \mathbf{s}_{p,m}$, that is changing origin for the integral to the center of the mth particle, and if U is the volume of one particle ($U = U_1 = U_2 = \ldots$),

$$\mathbf{F}_p = (U/NV_h) \sum_h \mathbf{F}_h \sum_m \sum_n \mathbf{G}_{hpmn} \times \exp\left[2\pi i (\mathbf{p}.\mathbf{s}_{p,m} - \mathbf{h}.\mathbf{s}_h)\right] \tag{8}$$

where

$$\mathbf{G}_{hpmn} = (1/U) \int_{U'_m} \exp\left[2\pi i (\mathbf{p} - \mathbf{h}[\Delta_{m,n}].\mathbf{y}')\right] d\mathbf{y}'. \tag{9}$$

The volume U'_m is centered at the origin.

Equation (8) can be written as

$$\mathbf{F}_p = \sum_h \mathbf{F}_h \mathbf{a}_{hp}, \tag{10}$$

where

$$\mathbf{a}_{hp} = (U/NV_h) \sum_m \sum_n \mathbf{G}_{hpmn} \times \exp\left[2\pi i (\mathbf{p}.\mathbf{s}_{p,m} - \mathbf{h}.\mathbf{s}_h)\right]. \tag{11}$$

Here the complex coefficients $\mathbf{a}_{hp}$ are dependent only on the non-crystallographic symmetry operators and the limits of the integration volume U_m.

Phase refinement proceeds by substituting the observed structure amplitudes and current phases in the

right-hand side of (10) and, hence, computing a new set of 'calculated' amplitudes and phases, $\mathbf{F}_p$. The new calculated phases are then used in the next refinement iteration in the same manner as has frequently been described for real-space averaging (Arnold, Vriend, Luo, Griffith, Kamer, Erickson, Johnson & Rossmann, 1987; Rossmann, 1990; Rossmann, McKenna, Tong, Xia, Dai, Wu, Choi, Marinescu & Lynch, 1992). (10) shows that the computation time for reciprocal-space molecular-replacement averaging is proportional to the number of reflections in the p cell, the number of symmetry operators of the molecular assembly, the number of symmetry operators of the p cell (excluding Bravais centering) and the number of reflections in the h cell that are used in the summation.

The reciprocal-space procedure can be used for:

(1) skew averaging where the p cell is defined as one with its axes parallel and perpendicular so as to orient the structure in the desired direction.

(2) Averaging between non-crystallographically related units within the same cell, in which case the h and p cells have the same dimensions.

(3) Averaging the electron density in a variety of different h cells and placing these into a common p cell.

Matrix algebra

Orthogonalizing and deorthogonalizing matrices will be defined in accordance with the notation of Rossmann & Blow (Rossmann & Blow, 1962).

Thus,

$$\left.\begin{aligned}\mathbf{y} &= [\alpha_p]\mathbf{Y},\ \mathbf{Y} = [\beta_p]\mathbf{y}\\ \mathbf{x} &= [\alpha_h]\mathbf{X},\ \mathbf{X} = [\beta_h]\mathbf{x}\end{aligned}\right\}. \qquad (12)$$

It will be assumed that the molecular assembly which is present in all the crystal forms has a closed point-group symmetry. If the center of the point group is placed at the origin of a Cartesian coordinate system **U**, then the symmetry operators of the point group can be represented as a set of rotation matrices

$$\mathbf{U}_n = [I_n]\mathbf{U}_1 \ (n = 1, 2, \ldots, N), \qquad (13)$$

where N (defined above) is the total number of asymmetric objects in the point group. A 'standard' orientation of the point group can be chosen such that some of the symmetry operators lie along special directions of the coordinate system. The three twofolds of the point group 222, for example, could be aligned along the Cartesian coordinate axes. The rotation matrices would then all be diagonal.

The orientation of the reference molecular assembly in each input h cell can be determined with an ordinary (Rossmann & Blow, 1962) or locked (Rossmann, Ford, Watson & Banaszak. 1972; Tong & Rossmann, 1990) rotation function. Similarly, the orientation of the reference molecular assembly in the output p cell can be arbitrarily defined for skew averaging and it will be the same as in the h cell for averaging within a given unit cell. Let

$$\mathbf{X}_n = [E_h]\mathbf{U}_n \text{ and } \mathbf{Y}_1 = [E_p]\mathbf{U}_1. \qquad (14)$$

Furthermore, let the crystallographic symmetry in the p-cell be represented by

$$\mathbf{y}_m = [T_m]\mathbf{y}_1 + \mathbf{t}_m, \qquad (15)$$

where $[T_m]$ and $\mathbf{t}_m$ are the crystallographic rotation and translation operators, respectively. Neglecting, temporarily, the translational elements, from (12), (13), (14) and (15),

$$\mathbf{x}_{m,n} = [\alpha_h][E_h][I_n][E_p^{-1}][\beta_p][[T_m^{-1}]\mathbf{y}_m.$$

Comparison with (5) shows that

$$[\Delta_{m,n}] = [\alpha_h][E_h][I_n][E_p^{-1}][\beta_p][[T_m^{-1}]. \qquad (16)$$

Evaluation of the G function

If the molecular mask is spherical with a radius of r, the G function can be calculated analytically (Rossmann & Blow, 1962) from

$$G = \left[3(\sin 2\pi Hr - 2\pi Hr\cos 2\pi Hr)\right]/(2\pi Hr)^3, \qquad (17)$$

where H is the length of the reciprocal-space vector $(\mathbf{p} - \mathbf{h}[\Delta_{m,n}])$.

If a more accurate definition is available for the envelope of the molecular assembly, an easy way to define the mask is by reference to the Cartesian coordinate system **U**. If (U, V, W) defines an integral grid point in this system, the mask can be defined as

$$U,V\left|(W_1,W_2),(W_3,W_4), \ldots, (W_{2k-1},W_{2k})\right],$$

where for every grid point, (u,v) are k pairs of numbers (w_1,w_2), *etc.*, which specify the initial and final grid-point ranges along w that are protein rather than solvent. Similar to the derivation (16), the fractional u,v,w coordinates can be mapped into the p cell. Then from (9)

$$\mathbf{G}_{hpmn} = (D^3/U)\sum_k \exp\left[2\pi i(\mathbf{p} - \mathbf{h}[\Delta_m]) \cdot [T_m][\alpha_p][E_p][\beta_u]\mathbf{u}_k\right], \qquad (18)$$

where β_u is the fractionalization matrix in the **U** cell and D is the grid interval in the mask cell. Note that the translational components can be ignored when mapping **y** to **u**, as the limits of the integration volume U'_m are centered at the origin. Now define

$$\mathbf{h}' = (\mathbf{p} - \mathbf{h}[\Delta_{m,n}])\cdot[T_m][\alpha_p][E_p][\beta_u].$$

Hence,

$$\mathbf{G}_{hpmn} = (D^3/U)\sum_u\sum_v\sum_w \text{mask}(U,V,W) \times \exp\left[-2\pi i(h'u + k'v + l'w)/L\right], \quad (19)$$

where L is the length of the edges of the cell **U** in which the mask is defined.

A Fourier transformation of (19), in which all the U,V,W points inside the mask are set to unity and all other points are set to zero, will sample G sufficiently finely. Although the quantities $\mathbf{h}'$ will be non-integral values, if L is chosen to be eight to ten times larger than the longest molecular dimension, only integral values of $\mathbf{h}'$ need be considered. Alternatively, the summations over W between pairs of values W_{2k-1}, W_{2k} can be expressed analytically as the sum of a geometric series. From (18),

$$\mathbf{G}_{hpmn} = (D^3/U)\sum_u\sum_v \exp\left[-2\pi i(h'U + k'V)/L\right] \times \sum_k \{\exp\left[(-2\pi i l' W_{2k-1})/L\right] + \exp\left[-2\pi i l'(W_{2k-1} + 1)/L\right] + \ldots\}.$$

The series in braces is a geometric series with a ratio $\exp(-2\pi i l'/L)$. Now, the sum of a geometric series with ratio x is given by

$$\sum_{n=1}^{n=j} x^n = (x^{j+1} - x^i)/(x - 1).$$

Thus, it follows that

$$\mathbf{G}_{hpmn} = (D^3/U)\sum_u\sum_v \exp\left[-2\pi i(h'U + k'V)/L\right] \times \sum_k \left[-i/2\sin(\pi l'/L)\right] \times \{\exp\left[(-2\pi i l'/L)(W_{2k-1} + \tfrac{1}{2})\right] - \exp\left[(-2\pi i l'/L)(W_{2k-1} - \tfrac{1}{2})\right]\}. \quad (20)$$

Further calculations using (20) can be carried out either by direct summation or by a two-dimensional fast Fourier transformation over the U and V directions.

To test the validity of the treatment above and to check the effect of different mask grid sampling on the mask transform, a sphere of radius 35 Å was placed on mask grids with 1, 2, 3 and 4 Å intervals, respectively. The resulting spherical mask was then transformed using (20), sampling the G function to its third node. The G-function values calculated based on this transform were then compared with those calculated directly from the analytical expression (17) (Table 1 and Fig. 1). The agreement (expressed as a conventional crystallographic R factor) between the two sets of G values was only 2% if 1 Å grid intervals were used to define the mask. The calculations also show that a relatively coarse grid can be used in defining the mask, resulting in considerable savings in computer time needed for the transform.

Table 1. *G-function calculation by mask transform of a spherical envelope*

The G-function was calculated for a sphere of radius 35 Å.

Grid interval (Å)	1	2	3	4
No. of grid points in mask	183191	23473	6989	3067
R factor* (%)	2.0	4.4	5.0	8.7
CPU time (min)	24	7	3	2

* R factor is calculated between the G function value from the mask transform and that from the analytical equation.

Masks are inherently low-resolution objects. Therefore, the G function will assume large values only when the length of H is small. For a spherical mask, the G function reaches its first zero when Hr is 0.72 and its value never exceeds 0.09 after the first node. For a general mask, the G function will no longer assume real values unless the mask has a center of symmetry. The amplitude of the G function, nevertheless, will be significant only in the region where $(\mathbf{p} - \mathbf{h}[\Delta_{m,n}])$ is small. Therefore, to speed up the calculation, the summation over the h-cell reflections can be restricted to only the first few nodes of the G function. This results in a truncation of the G function and introduces errors in the molecular-replacement equation. Such an approximation is used successfully in reciprocal-space rotation-function calculations (Rossmann & Blow, 1962).

Comparison with real-space molecular-replacement averaging

The time required for calculations that are involved in real-space and reciprocal-space molecular-replacement averaging are compared in Table 2. The sampling of the real space (electron density) can be varied, though usually a grid size between 1/2.5 and 1/6 of the resolution of the reflection data is used. In practice, sampling

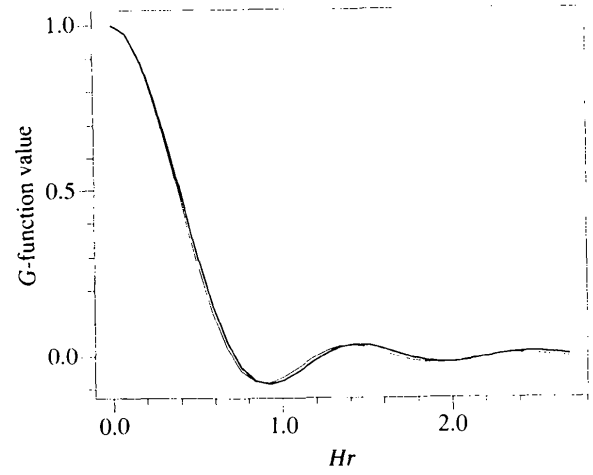

Fig. 1. Comparison of the calculation of the G function for a perfect sphere of radius 25 Å using the analytical function (17) (thick line) or as the transform of a mask (20) (thin line) on a 4 Å grid.

of 1/3 is ample if an eight-point interpolation is used between grid points without loss of accuracy or rate of convergence (Rossmann, McKenna, Tong, Xia, Dai, Wu, Choi & Lynch, 1992). (10) can be taken as the reciprocal-space equivalent of real-space interpolation. The G function is the weighting factor for this interpolation for the h reciprocal lattice points around the non-integral reciprocal lattice position given by $\mathbf{p}[\Delta_{m,n}]$. Due to the coarse sampling in reciprocal space, the interpolation must involve more points; for example, truncation of G at the first node requires at least a box size of $3 \times 3 \times 3$ or 27 terms. On the other hand, there are roughly only $(1/3)^3$ as many structure factors as there are electron-density points to sample.

With reciprocal-space averaging, it is possible to consider subsets of reflections in the calculation. This is the major advantage of the reciprocal-space method over the real-space method, where all the reflections must be considered simultaneously. As a simple example, only those reflections near the edge of the resolution range need to be considered in a phase extension. The phase information for lower resolution reflections need be updated only periodically by including all reflections in the calculation.

Application to bacteriophage φX174

Crystals of φX174 belong to the space group $P2_1$ with a = 305.6, b = 360.8, c = 299.5 Å and β = 92.89° (McKenna, Xia, Willingmann, Ilag, Krishnaswamy, Rossmann, Olson, Baker & Incardona, 1992; McKenna, Xia, Willingmann, Ilag & Rossmann, 1992). The virus particle is situated at (0.2505, 0.2500, 0.2505) in the asymmetric unit. The orientation of the particle at this position is related to the defined standard orientation by an Eulerian rotation of (82.15, 92.35, 81.64°) (McKenna, Xia, Willingmann, Ilag & Rossmann, 1992). The phased reflection data between 30 and 9 Å resolution, after real-space averaging at 9 Å resolution, was used as the starting phase set. A molecular mask was determined by using the φX174 electron density in the p cell and skew averaging it into an h cell with cell dimensions of a = b = c = 330 Å and $\alpha = \beta = \gamma = 90°$. The breakdown of the local icosahedral symmetry reduces the height of the averaged density beyond the confines of the virus particle and provides a basis for recognizing the limits of the molecular envelope. Details of the structure were removed by local smearing (Wang, 1985). The solvent content of the crystal cell based on the resultant molecular mask was 55%. As the particle possesses 532 symmetry, the mask is centrosymmetric and the G function is entirely real.

Phases were extended from 9.0 to 7.9 Å resolution in six steps. Several cycles were carried out at each step. The solvent region was assumed to have zero density. (In the real-space averaging, the solvent region was reset to its average value for each cycle.) At each step the resolution was extended by the length of the shortest reciprocal lattice vector ($\mathbf{b}^*$). To speed up the calculations, only those phases were determined that were in a thin shell at the highest current resolution limit. The thickness of the shell was set equal to the width of the integration box. When the integration box was $3 \times 3 \times 3$, the shell contained about 13 500 reflections. After extension to 7.9 Å, one pass was carried out using all 70 404 reflections. Extension beyond 7.9 Å resolution was not pursued as the structure of φX174 had been previously determined using real-space averaging, and as the general properties of the reciprocal-space method had then been reasonably explored.

Four factors were varied to monitor their effects on the results of the calculation: the number of cycles at each extension step, the size of the artificial mask U cell, the size of the integration box, and the removal of small G-function values (Table 2). The mask was placed in U cells with 3300 or 6600 Å on edge. The calculation of the transform by direct summation to $\mathbf{h}'$of 55 (a total of 348 254 reflections) took 3 CPU hours on an SGI Indigo R4000. Using a larger U cell made little difference in this case. A larger U cell may prove to be more advantageous in other problems as the sampling of the G function would be more accurate. There is little penalty on computing time for using a larger U cell. Greater truncation of the G function led to a slight speeding up of the calculation but also worsened the results. Using a larger integration box produced slightly more accurate phase angles for the phase-extended reflections. However, the correlation coefficient was lower and the execution time was much longer. Significant improvements were obtained by running more cycles of averaging at each extension step. The average phase change at the last cycle was 1° after ten cycles and 0.2° after 20 cycles. A comparison with the phase angles after real-space averaging at 30–7.9 Å resolution showed that reciprocal-space calculations gave essentially the same results (Table 2 and Fig. 2). The greater the care taken in reciprocal space (the larger the integration box size, the less truncation of the G function and the number of iteration cycles), the better is the phase agreement. Differences between the reciprocal- and real-space method are within the range expected by comparing with phases computed from atomic positions. The somewhat lower quality of test E remains unexplained, although it may be related to the point of truncation of the G function.

The real-space averaging and phase extension from 9 to 7.9 Å resolution took about 30 CPU hours on a Cyber 205. A direct comparison of the CPU performance of the two techniques is difficult as the real-space averaging program had been extensively vectorized and optimized for the Cyber. The CPU performance of the reciprocal-space approach in the current implementation is probably inferior to the real-space one, though many improvements are possible in the execution of the reciprocal-space calculations.

Table 2. *Quality of phasing under different computational conditions*

Test calculations with the φX174 data.

(i) Definition of the test conditions

Test	*U* cell *a* (Å)	$\|G\|_{min}$*	No. of cycles	Integration box size	CPU (h)†
A	6600	0.01	10	3 × 3 × 3	33
B	6600	0.01	20	3 × 3 × 3	66
C	6600	0.10	10	3 × 3 × 3	27
D	3300	0.01	10	3 × 3 × 3	33
E	3300	0.01	10	5 × 5 × 5	183

(ii) Phase comparison among the different tests

		Average phase difference (°)					
Test	⟨CC⟩‡	*B*	*C*	*D*	*E*	30–7.9 Å§	9–7.9 Å§
A	0.809	30	25	5	44	27	45
B	0.827	—	48	30	33	22	27
C	0.796	—	—	24	53	32	60
D	0.807	—	—	—	44	27	46
E	0.742	—	—	—	—	29	38

* All terms were removed for which $G < |G|_{min}$.

† CPU time on an SGI Indigo R4000 computer.

‡ Average correlation coefficient, CC, over the resolution range 30–7.9 Å, where $CC = [\sum(\langle F_o\rangle - F_o)(\langle F_c\rangle - F_c)]/\{[\sum(\langle F_o\rangle - F_o)^2 \times \sum(\langle F_c\rangle - F_c)^2]^{1/2}\}$. $\langle F_o\rangle$ is the mean of the observed structure amplitudes, F_o, divided into ten resolution ranges. $\langle F_c\rangle$ is the mean of the calculated structure amplitudes from equation (10) divided into similar resolution ranges.

§ Average difference with respect to phase angles obtained from real-space averaging.

Concluding remarks

A number of earlier, partially successful, attempts have been made to determine phases using non-crystallographic symmetry in reciprocal space (Crowther, 1967, 1969; Main, 1967; Johnson, 1978). These were made prior to the extensive success and experience with real-space averaging. The formulation of the reciprocal-space process has a great deal of elegance and can be expressed very simply in terms of (10). In contrast, real-space averaging requires a series of procedures: averaging, Fourier analysis and Fourier synthesis. Hence, in this paper we have re-examined the reciprocal-space method. However, what we report here is merely the beginning. We again demonstrate the validity of the method and the effect of the necessary approximation in reciprocal space caused by the truncation of the *G* function. We also, for the first time, apply a careful molecular mask in reciprocal space, as opposed to a simple spherical envelope.

The reciprocal-space method has some intrinsic advantages, particularly when extending phases. Only those equations need be used for which the terms on the right-hand side are based on good phases for large amplitude coefficients. Thus, phase extension can proceed along chosen pathways in reciprocal space which permit good phase determination. In contrast, in real space phase extension has to proceed in shells of reciprocal space representing small increments of resolution. It may be possible to proceed with reciprocal-space determinations by using only a subset of structure factors with large amplitudes. Also it may be possible to weigh each coefficient independently in (10) with a weight representing the current quality of the phase information. All these techniques suggest that the reciprocal-space method is more flexible than the real-space method and may be made more powerful by its ability to determine each phase sequentially. This also then permits updating of the current phase information after each new phase has been determined from the current phase set and, therefore, should lead to more rapid convergence.

The current extensive success with real-space averaging has discouraged the development of the complementary reciprocal-space method. Here, however, we have made a new attempt at developing phase determination in reciprocal space.

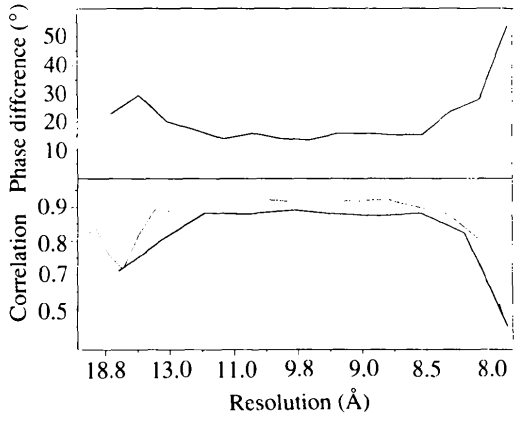

Fig. 2. Top, mean difference between phases computed by real- and reciprocal-space methods. Bottom, comparison of real-space (thin) and reciprocal-space (thick) calculated amplitudes with observed amplitudes.

We are grateful to Robert McKenna, Leodevico L. Ilag and others at Purdue University for help in the use of the φX174 reflection data. We also thank Cheryl Towell and Sharon Wilder for help in the preparation of this manuscript. The work was supported by a National Science Foundation grant to MGR and a Lucille P. Markey Award for the expansion of Structural Biology at Purdue University.

References

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

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.

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.

COLMAN, P. M. (1974). *Z. Kristallogr.* **140**, 344-349.

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

CROWTHER, R. A. (1969). *Acta Cryst.* B**25**, 2571-2580.
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.
HOGLE, J. M., CHOW, M. & FILMAN, D. J. (1985). *Science*, **229**, 1358-1365.
JOHNSON, J. E. (1978). *Acta Cryst.* B**34**, 576-577.
JONES, T. A. (1992). *Molecular Replacement. Proceedings of the CCP4 Study Weekend*, 31 January–1 February 1992, edited by E. DODSON, S. GOVER & W. WOLF, pp. 91-105. Daresbury: SERC.
LAWRENCE, M. C. (1991). *Quart. Rev. Biophys.* **24**, 399-424.
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.
MCKENNA, R., XIA, D., WILLINGMANN, P., ILAG, L. L., KRISHNASWAMY, S., ROSSMANN, M. G., OLSON, N. H., BAKER, T. S. & INCARDONA, N. L. (1992). *Nature (London)*, **355**, 137-143.
MCKENNA, R., XIA, D., WILLINGMANN, P., ILAG, L. L. & ROSSMANN, M. G. (1992). *Acta Cryst.* B**48**, 499-511.
ROSSMANN, M. G. (1990). *Acta Cryst.* A**46**, 73-82.
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., FORD, G. C., WATSON, H. C. & BANASZAK, L. J. (1972). *J. Mol. Biol.* **64**, 237-249.
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., MCKENNA, R., TONG, L., XIA, D., DAI, J., WU, H., CHOI, H. K. & LYNCH, R. E. (1992). *J. Appl. Cryst.* **25**, 166-180.
ROSSMANN, M. G., MCKENNA, R., TONG, L., XIA, D., DAI, J., WU, H., CHOI, H. K., MARINESCU, D. & LYNCH, R. E. (1992). *Molecular Replacement. Proceedings of the CCP4 Study Weekend*, 31 January–1 February 1992, edited by E. DODSON, S. GOVER & W. WOLF, pp. 33-48. Daresbury: SERC.
SPEIR, J., MUNSHI, S., BAKER, T. S. & JOHNSON, J. E. (1993). *Virology*, **193**, 234-241.
TONG, L. & ROSSMANN, M. G. (1990). *Acta Cryst.* A**46**, 783-792.
TSAO, J., CHAPMAN, M. S., WU, H., AGBANDJE, M., KELLER, W. & ROSSMANN, M. G. (1992). *Acta Cryst.* B**48**, 75-88.
VALEGÅRD, K., LILJAS, L., FRIDBORG, K. & UNGE, T. (1990). *Nature (London)*, **345**, 36-41.
WANG, B. C. (1985). *Methods Enzymol.* **115**, 90-112.