Journal of Structural Biology **125,** 196–208 (1999)
Article ID jsbi.1999.4083, available online at http://www.idealibrary.com on IDEAL®

Cryo-Electron-Microscopy Reconstruction of Partially Symmetric Objects

Michael G. Rossmann[1] and Yizhi Tao

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

Received November 10, 1998, and in revised form January 4, 1999

A procedure is described for reconstructing three-dimensional objects from two-dimensional projections. The method is based both on the original Crowther, DeRosier, and Klug (DeRosier, D.J., and Klug, A. (1968), Nature 217, 130–134; Crowther, R.A., DeRosier, D.J., and Klug, A. (1970) *Proc. R. Soc. London A* 317, 319–340) work on image reconstructions of icosahedral viruses and on the concept of noncrystallographic symmetry (Rossmann, M.G., 1995, *Curr. Opin. Struct. Biol.* 5, 650–655). The procedure has been applied so far only to test data where the objective has been the determination of particle orientation, both *ab initio* and through the use of model data. © 1999 Academic Press

Key Words: **electron microscopy; image reconstruction; noncrystallographic symmetry; partially symmetric objects.**

INTRODUCTION

A pair of stereoscopic images is sufficient to view an object in three dimensions. The two images are projections down axes differing in orientation by only a few degrees and can be used both to view the object and to compute its three-dimensional structure (Rossmann and Argos, 1980). However, the accuracy of the three-dimensional reconstruction will depend on the angular separation of the images and would improve if the number of projected images were increased. The same problem is encountered in reconstructing an object from numerous electron-microscopic images representing projections down a random set of axes. The larger number of images permits viewing of the three-dimensional object with increased resolution at a reduced noise level.

Techniques for reconstructing objects from electron-microscopic projection data were originally developed by Crowther, DeRosier, and Klug (DeRosier and Klug, 1968; Crowther *et al.*, 1970a) and independently by Hoppe (1974; Hoppe *et al.*, 1974). These methods have had a tremendous impact on the study of biological macromolecular structure, especially since the advent of cryo-electron microscopy (cryo-EM). A vast amount of information has been gathered on the structure of viruses and their complexes with antibodies (e.g., Prasad *et al.*, 1990; Smith *et al.*, 1993) or receptors (Olson *et al.*, 1993). The study of viruses has depended largely on the early work by Crowther (Crowther *et al.*, 1970a,b; Crowther, 1971; Crowther and Amos, 1972), which utilized icosahedral symmetry to orient the individual images of isometric viruses. One of the most difficult aspects has been the determination of the orientation of individual projected images. Until about 4 years ago, the primary tool was the method of common lines (Crowther, 1971), but recently alternative procedures that depend on the comparison with homologous three-dimensional images have been developed (Baker and Cheng, 1996). The early techniques were based on Fourier analyses of the images ("reciprocal space" methods). A variety of such techniques have also been used even in the absence of particle symmetry. Alternative techniques have been developed (cf. Frank, 1996) based on real space comparisons. These techniques have been especially useful for the reconstruction of objects that either have no symmetry (e.g., ribosomes (Frank *et al.*, 1995)) or have little symmetry (e.g., GroES (Saibil, 1996)).

Recently, cryo-EM reconstructions have been extended to almost 7-Å resolution in the analysis of hepatitis B cores (Böttcher *et al.*, 1997; Conway *et al.*, 1997), where the secondary structure of the capsid protein is starting to emerge. These advances have been possible in part because of improved instrumentation using field emission electron guns to increase spatial coherence and because of the development of methods to correct for the phase-contrast transfer function when using a series of differently focused images.

These advances in cryo-EM imaging are bringing new topics of biological interest within range. For instance, it has become apparent that many icosahedral viruses have a significant amount of their nucleic acid in an icosahedrally ordered form (Chen

[1] To whom correspondence should be addressed. Fax: (765) 496-1189.

1047-8477/99 $30.00

et al., 1989; Larson *et al.*, 1993; Grimes *et al.*, 1998), suggesting that, possibly, much more of the nucleic acid than seen in crystallographic electron density maps has a unique structure. Thus, if the virus could be oriented by virtue of its icosahedral coat and if a selection were made among the 60 possible orientations of any one particle to match the internal nucleic acid of another particle, then it might be possible to obtain extensive information on the nucleic acid structure and fold. In other cases, it has been shown (Al Ani *et al.*, 1979; Casjens and Hendrix, 1988) that nucleic acid is extruded from a single portal of an otherwise highly symmetric particle. It is, therefore, of interest to examine symmetric particles that have local distortions (e.g., at a phage head portal) or in which the higher symmetry exists for only a portion of the structure.

With these problems in mind, we consider here procedures for image reconstruction from projected images in conjunction with experience in noncrystallographic-symmetry electron-density averaging as used for solving crystal structures (Rossmann *et al.*, 1992; Rossmann, 1995; Kleywegt and Read, 1997). We shall assume that the projection data have been corrected for experimental deficiencies such as astigmatism and the effects of the phase-contrast transfer function. Although we borrow heavily from Crowther *et al.* (1970a), we do not use spherical harmonics. Furthermore, while Crowther requires reciprocal space interpolation in only one dimension (to reduce computational complexity), we use a general three-dimensional, Cartesian-coordinate-based interpolation expression.

This study has not yet progressed beyond developing the procedure with theoretical data. Whether there are experimental advantages in these reconstruction techniques will require tests on real data and, indeed, such work is currently in progress using the data for various ϕ29 particles (Tao *et al.*, 1998).

DEFINITIONS

The nomenclature defined below is that used by Rossmann *et al.* (1992) in describing a procedure for electron-density averaging. The object to be reconstructed from a set of projections will be placed into the "*h*-cell." (The "*h*" and "*p*" in the terms "*h*-cell" and "*p*-cell" originate in early attempts to use the known lactate dehydrogenase structure—LDH in the *h*-cell—to solve the unknown structure of glyceraldehyde-3-phosphate dehydrogenase—GAPDH in the *p*-cell.) A point in this cell is defined as being at x,y,z (or $\mathbf{x}$ in vector notation) in fractional coordinates. The complex Fourier coefficients, $\mathbf{F}(h,k,l)$, corresponding to the electron density in the *h*-cell, will be defined by the indices (h,k,l), or $\mathbf{h}$ in vector notation. If the reconstructed object has some symmetry, then it is useful to place the object into the *h*-cell with as many as possible of its symmetry elements coincident with the symmetry of the repeating lattice of the *h*-cell. For instance, if the object has 222 symmetry, then the center of the object can be placed on the origin of the cell with its twofold axes along the orthogonal cell directions. Thus, some or all of the object's symmetry elements can be included in the crystallographic symmetry of the *h*-cell. The remaining symmetry of the object will be noncrystallographic, being valid only locally within a confined envelope around the object.

Each projection image can be defined with respect to its own *p*-cell. Positions within this cell are defined by the fractional coordinates u,v,w (or $\mathbf{u}$ in vector notation) with the *w* direction being perpendicular to the projected image. The indices of the complex Fourier coefficients $\mathbf{Q}_{pq}$, representing the projected density $P(u,v)$ in the *p*-cell, are defined by (p,q) (or $\mathbf{p}$ in vector notation). The orientation of a particular projection relative to the object in the *h*-cell is given by the transformation

$$\mathbf{x} = [D]\mathbf{u}, \tag{1}$$

where $[D]$ is a 3×3 matrix dependent on three rotation angles and the cell dimensions of both the *h*- and *p*-cells (see below).

RECONSTRUCTION OF AN ASYMMETRIC OBJECT FROM PROJECTION IMAGES

Let the three-dimensional electron density of the reconstructed object in the *h*-cell be represented by $\rho(x,y,z)$.

Since $P(u,v)$ is a projection along *w* (perpendicular to *u* and *v*),

$$P(u,v) = \int_{w=-1/2}^{+1/2} \rho(x,y,z)\, \mathrm{d}w. \tag{2}$$

Now, $\rho(x,y,z)$ can be expressed as a Fourier summation in the *h*-cell with coefficients $\mathbf{F}(h,k,l)$. Therefore,

$$\rho(x,y,z) = \frac{1}{V} \sum_{h,k,l} \mathbf{F}(h,k,l) \mathrm{e}^{2\pi i(hx+ky+lz)},$$

where *V* is the volume of the *h*-cell, or

$$\rho(\mathbf{x}) = \frac{1}{V} \sum_{\mathbf{h}} \mathbf{F}_{\mathbf{h}} \mathrm{e}^{2\pi i \mathbf{h}\cdot\mathbf{x}} \tag{3}$$

in vector notation. Substituting (3) in (2),

$$P(u,v) = \int_{w=-1/2}^{+1/2} \left(\frac{1}{V} \sum_{\mathbf{h}} \mathbf{F}_{\mathbf{h}} e^{2\pi i \mathbf{h}\cdot\mathbf{x}} \right) dw.$$

In order to perform the integration over w, it is necessary to express $\mathbf{x}$ in terms of $\mathbf{u}$ by means of the definition (1). Then, recognizing that each term in the summation can be separately integrated,

$$\begin{aligned} P(u,v) &= \frac{1}{V} \sum_{\mathbf{h}} \int_{w=-1/2}^{+1/2} \mathbf{F}_{\mathbf{h}}\, e^{2\pi i \mathbf{h}\cdot[D]\mathbf{u}}\, dw \\ &= \frac{1}{V} \sum_{\mathbf{h}} \mathbf{F}_{\mathbf{h}} \int_{w=-1/2}^{+1/2} e^{2\pi i(\rho' u + q' v + r' w)}\, dw, \end{aligned}$$

where

$$\left.\begin{aligned} p' &= d_{11}h + d_{21}k + d_{31}l \\ q' &= d_{12}h + d_{22}k + d_{32}l \\ r' &= d_{13}h + d_{23}k + d_{33}l \end{aligned}\right\} \tag{4a}$$

and d_{ij} are the elements of $[D]$. Or, in brief,

$$\mathbf{p}' = [D^{\mathrm{T}}]h. \tag{4b}$$

As w is the only variable,

$$\begin{aligned} P(u,v) &= \frac{1}{V} \sum_{\mathbf{h}} \mathbf{F}_{\mathbf{h}} e^{2\pi i(p'u+q'v)} \int_{w=-1/2}^{+1/2} e^{2\pi i r' w}\, dw \\ &= \frac{1}{V} \sum_{\mathbf{h}} \mathbf{F}_{\mathbf{h}} e^{2\pi i(p'u+q'v)} \frac{\sin \pi r'}{\pi r'}. \end{aligned} \tag{5}$$

Expression (5) gives the relationship between the three-dimensional Fourier coefficients $\mathbf{F}_{\mathbf{h}}$ and the projected density $P(u,v)$. Experimentally, it is $P(u,v)$ that is observed and derived from the electron-microscope data, while the Fourier coefficients $\mathbf{F}_{\mathbf{h}}$ are to be determined from the collection of projected images.

Now, the Fourier coefficients $\mathbf{Q}_{pq}$ can be determined by numerical integration of the projection density $P(u,v)$ using the relationship

$$\mathbf{Q}_{pq} = \int_A P(u,v) e^{-2\pi i(pu+qv)}\, du\, dv, \tag{6}$$

where A is the area of the p-cell in the u,v plane.

By substituting $P(u,v)$ in expression (6) with (5), a relationship can be found between the known projection coefficients $\mathbf{Q}_{pq}$ and the unknown three-dimensional Fourier coefficients $\mathbf{F}_{\mathbf{h}}$. It follows that

$$\begin{aligned} \mathbf{Q}_{pq} &= \int_A \left[\frac{1}{V} \sum_{\mathbf{h}} \mathbf{F}_{\mathbf{h}} e^{2\pi i(p'u+q'v)} \frac{\sin \pi r'}{\pi r'} \right] \\ &\quad \times e^{-2\pi i(pu+qv)}\, du\, dv \\ &= \frac{1}{V} \sum_{\mathbf{h}} \mathbf{F}_{\mathbf{h}} \frac{\sin \pi r'}{\pi r'} \int_{u=-1/2}^{+1/2} e^{2\pi i(p'-p)u}\, du \\ &\quad \times \int_{v=-1/2}^{+1/2} e^{2\pi i(q'-q)v}\, dv \end{aligned}$$

or

$$\mathbf{Q}_{pq} = \frac{1}{V} \sum_{\mathbf{h}} \mathbf{F}_{\mathbf{h}} G_{\mathbf{hp}}, \tag{7}$$

where

$$G_{\mathbf{hp}} = \frac{\sin \pi(p'-p)}{\pi(p'-p)} \cdot \frac{\sin \pi(q'-q)}{\pi(q'-q)} \cdot \frac{\sin \pi r'}{\pi r'}. \tag{8}$$

If the orientation of the projected image is known relative to the standard particle orientation in the h-cell and is expressed in terms of the $[D]$ matrix, then (p',q',r') can be calculated for a chosen set of indices (h,k,l) using (4). Hence, $\mathbf{Q}_{pq}$ could be calculated from the summation (7), given the three-dimensional coefficients $\mathbf{F}_{\mathbf{h}}$, for any specific indices (p,q). However, in practice, the coefficients $\mathbf{Q}_{pq}$ are known, while the $\mathbf{F}_{\mathbf{h}}$ coefficients need to be determined. Given a sufficient number of $\mathbf{Q}_{pq}$ coefficients, it is then possible to solve the Eqs. (7) for the required three-dimensional Fourier coefficients $\mathbf{F}_{\mathbf{h}}$. These can then be used to compute a Fourier synthesis that represents the reconstructed image in the h-cell.

It is unnecessary to evaluate all terms in the summation over $\mathbf{h}$ in expression (7) as most of the terms will be negligibly small. Only the terms for which $p' - p \cong 0$, $q' - q \cong 0$, and $r' \cong 0$ will be significant. Thus, using (4), only the (h,k,l) terms that roughly satisfy the relationship

$$\begin{aligned} p &= d_{11}h + d_{21}k + d_{31}l \\ q &= d_{12}h + d_{22}k + d_{32}l \\ 0 &= d_{13}h + d_{23}k + d_{33}l \end{aligned}$$

or

$$\begin{pmatrix} p \\ q \\ 0 \end{pmatrix} = [D^{\mathrm{T}}] \begin{pmatrix} h \\ k \\ l \end{pmatrix}$$

for a given pair of indices (p,q) need to be considered. These terms are those that satisfy the plane $0 = d_{13}h + d_{23}k + d_{33}l$, which is a plane in the reciprocal space of the h-cell that is perpendicular to the projection direction $\mathbf{w}$.

In practice, it is useful to solve for $\mathbf{h}''$ for selected indices (p,q), where

$$\begin{pmatrix} h'' \\ k'' \\ l'' \end{pmatrix} = [D^{\mathrm{T}}]^{-1} \begin{pmatrix} p \\ q \\ 0 \end{pmatrix}. \tag{9}$$

Thus, the (h,k,l) indices that will be associated with significantly large values of $G_{\mathbf{hp}}$ are those integers that lie close to (h'',k'',l''). These can be substituted in (4a) to obtain (p',q',r'), which are necessary to evaluate $G_{\mathbf{hp}}$ from (8).

The summation in (7) is over all coefficients within the limiting sphere of resolution. Thus, the summation can be rewritten as

$$\mathbf{Q}_{pq} = \frac{1}{V} \sum_{\mathbf{h}}^{\text{hemisphere}} G_{\mathbf{hp}}\mathbf{F}_{\mathbf{h}} + G_{\bar{\mathbf{h}}\mathbf{p}}\mathbf{F}_{\bar{\mathbf{h}}}$$

$$= \frac{1}{V} \sum_{\mathbf{h}}^{\text{hemisphere}} [(G_{\mathbf{hp}} + G_{\bar{\mathbf{h}}\mathbf{p}})A_{\boldsymbol{h}} + i(G_{\mathbf{hp}} - G_{\bar{\mathbf{h}}\mathbf{p}})B_{\boldsymbol{h}}],$$

where $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$ are the real and imaginary parts of $\mathbf{F}_{\mathbf{h}}$. Note also that $G_{\bar{\mathbf{h}}\mathbf{p}}$ is unlikely to be significant except for low-order data. In general, therefore,

$$S_{pq} = \sum_{\mathbf{h}}^{\text{hemisphere}} a_{\mathbf{hp}} A_{\mathbf{h}} \text{ and } T_{pq} = \sum_{\mathbf{h}}^{\text{hemisphere}} b_{\mathbf{hp}} B_{\mathbf{h}}, \tag{10}$$

where S_{pq} and T_{pq} are the known real and imaginary parts of $\mathbf{Q}_{pq}$, and where the coefficients $a_{\mathbf{hp}}$ and $b_{\mathbf{hp}}$ can be calculated from the assumed orientation of the projection relative to the standard orientation of the particle in the h-cell.

LEAST-SQUARES ANALYSIS

Because of the large experimental error in each observational equation of type (10), it is necessary to have a large excess of observational equations over the number of three-dimensional Fourier coefficients $\mathbf{F}_{\mathbf{h}}$ that are to be determined. These observational equations need to be reduced to a set of normal least-squares equations for solution.

For convenience, the unknown $\mathbf{F}_{\mathbf{h}}$ coefficients can be numbered from 1 to m. If the particle has no symmetry that can be incorporated into the h-cell lattice, then the number of unknowns will be equal to the number of $\mathbf{F}_{\mathbf{h}}$ coefficients in a hemisphere of reciprocal space. However, if the particle of the h-cell can be described as having N crystallographic asymmetric units, then the number of unknowns will be reduced by $1/N$. Thus, the observational Eq. (10) can be written as

$$\sum_{i=1}^{m} a_i A_i = S \quad \text{and} \quad \sum_{i=1}^{m} b_i B_i = T.$$

The normal equations will then have the form

$$\begin{pmatrix} \sum_{l=1}^{n} a_{1l}a_{1l} & \sum_{l=1}^{n} a_{1l}a_{2l} & \cdots & \sum_{l=1}^{n} a_{1l}a_{ml} \\ \sum_{l=1}^{n} a_{2l}a_{1l} & \sum_{l=1}^{n} a_{2l}a_{2l} & \cdots & \sum_{l=1}^{n} a_{2l}a_{ml} \\ \vdots & \vdots & & \vdots \\ \sum_{l=1}^{n} a_{ml}a_{1l} & \sum_{l=1}^{n} a_{ml}a_{2l} & \cdots & \sum_{l=1}^{n} a_{ml}a_{ml} \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{pmatrix} = \begin{pmatrix} \sum_{l=1}^{n} S_l a_{1l} \\ \sum_{l=1}^{n} S_l a_{2l} \\ \vdots \\ \sum_{l=1}^{n} S_l a_{ml} \end{pmatrix}, \tag{11}$$

with a similar set for the imaginary set of equations. The summations are carried out over the n coefficients $\mathbf{Q}_{pq}$ in the k different projections. Before inverting the normal equations, it will be necessary to determine whether all Fourier coefficients $\mathbf{F}_{\mathbf{h}}$ have been adequately represented by large enough $G_{\mathbf{hp}}$ values. If there is a nonuniform distribution of projection orientations, some $\mathbf{F}_{\mathbf{h}}$ coefficients may not have significant amplitude in Eq. (7). This can be determined by looking for small diagonal terms in the normal equations. A useful criterion was found by rejecting all equations whose diagonal terms were less than 0.01 of the average value of the diagonal terms. The row and column associated with such a small diagonal term must then be eliminated prior to matrix inversion.

THE NUMBER OF REQUIRED PROJECTIONS

It will be necessary to have more (usually far more) observational equations, $\mathbf{Q}_{pq}$, than the number of unknowns, $\mathbf{F}_{\mathbf{h}}$, whose values are to be deter-

mined. The number of $\mathbf{F_h}$ coefficients will depend on the unit cell size of the *h*-cell and the desired resolution, *R*. Since the *h*-cell can be defined to be only just larger than the largest dimension of the unknown particle and as the *p*-cells represent different projections of the same particle, it is generally convenient to define the *h*- and *p*-cells as being of the same size. For spherical objects such as viruses, it is reasonable to make both the *h*- and *p*-cells cubic in shape with a cell edge of *a*. It may sometimes be convenient to use a trigonally shaped *h*-cell if, for instance, the object can be assumed to contain a threefold axis.

The volume of the reciprocal *h*-cell will be $(1/a)^3$. Thus, for a cubic cell of length *a*, the number of reciprocal unit cells within a resolution of *R* will be $[4/3\ \pi(1/R)^3]/(1/a)^3$. Because of Friedel's law, the number of independent reflections that need to be determined will then be half of the number of reciprocal unit cells, or $[2/3\ \pi(a/R)^3]$. If the object to be reconstructed into the *h*-cell is assumed to have symmetry that can be incorporated into the *h*-cell lattice ("crystallographic symmetry"), this will reduce the number of independent Fourier coefficients $\mathbf{F_h}$. For instance, if an object has a 23-point group and the two- and threefold axes of the object are defined to lie along the axes and body diagonal of the cubic-shaped *h*-cell, then the number of independent $\mathbf{F_h}$'s within the resolution *R* will be reduced by 1/12. Hence, if symmetry elements are defined to create *N* crystallographic asymmetric units in the *h*-cell, then the number of independent terms that need to be determined will be $(1/N)\ (2/3)\ \pi(a/R)^3$. Now, using similar arguments, the number of $\mathbf{Q}_{pq}$ coefficients in a projection will be $(\pi/2)(a/R)^2$. Thus, if *k* different projections are available and considering the large error in the observed $\mathbf{Q}_{hp}$ coefficients, it will be necessary to have

$$k\frac{\pi}{2}\left(\frac{a}{R}\right)^2 >>> \frac{1}{N}\frac{2}{3}\pi\left(\frac{a}{R}\right)^3;$$

that is,

$$k >>> \frac{4}{3}\frac{1}{N}\left(\frac{a}{R}\right).$$

Table I shows the *theoretical* lower limit (in the absence of all noise) of the required number of projected images for a successful reconstruction to *R* Å resolution when $a = 500$ Å and $N = 1$ or 12. The large error associated with low-dose EM images will, however, greatly increase the number of images required for a useful reconstruction. Henderson (1995) has considered the effect of error and concludes that the number of useful images is independent of the size of the object. This result, however, defies common sense as the amount of information necessary to reconstruct a large particle to a given resolution must be more than that required for a small particle.

TABLE I

Theoretical Minimum Number of Projected Images Required for a Successful Reconstruction in the Absence of All Noise[a]

Resolution (Å)	40	30	20	15	10
p-cell					
Minimum no. of projections required when $N = 1$	17	22	33	45	67
Minimum no. of projections required when $N = 12$	2	2	3	4	6
h-cell					
No. of independent unknown $\mathbf{F_h}$'s when $N = 1$	4091	9696	32725	77570	261799
No. of independent unknown $\mathbf{F_h}$'s when $N = 12$	341	808	2727	6464	21817

Note. It is assumed that the particle has a diameter of about 500 Å and has either $N = 1$-fold or $N = 12$-fold redundancy that can be incorporated into the *h*-cell lattice.

[a] The minimum number of required projections for an icosahedral reconstruction will be reduced by 1/5 when the NCS restraints are included in the formation of the normal equations. However, the number of $\mathbf{F_h}$'s to be determined remains unchanged.

RECONSTRUCTION OF OBJECTS THAT HAVE NONCRYSTALLOGRAPHIC SYMMETRY

If the symmetry of the object can be matched with the symmetry of the *h*-cell lattice (crystallographic symmetry), then the number of unknowns can be reduced by the crystallographic redundancy. The orientation matrix [*D*] for each projected image will then have to be determined (see below) with respect to the assumed placement of the object's symmetry axes in the *h*-cell. For example, if an object's assumed twofold axis is aligned with the *h*-cell's *b*-axis, then the Fourier coefficients $\mathbf{F}_{hkl}$ and $\mathbf{F}_{\bar{h}k\bar{l}}$ will be identical. Thus, in evaluating the coefficients $G_{\mathbf{hp}}$ for Eqs. (7) or (10), all Fourier coefficients with indices $(\bar{h},k,\bar{l})$ can be replaced by coefficients with indices (h,k,l).

The highest symmetry of any periodic lattice is 432. Thus, not every symmetry operator of an object can be incorporated into the crystallographic symmetry. For instance, an icosahedral virus has 532 symmetry, but only the 23 tetrahedral symmetry

(with a redundancy of 12) can be incorporated into the lattice. The fivefold symmetry present in the virus cannot be used to reduce the number of unknown Fourier coefficients. Hence, this "noncrystallographic" symmetry (NCS) must be treated in a different manner.

In contrast, Crowther (1971; Crowther *et al.*, 1970a) aligns the highest n-fold rotation axis (e.g., a fivefold axis if an icosahedron) along the z-axis of the standard h-cell, thus reducing the number of variables by a factor of n. Then Crowther uses the remaining symmetry axes (such as the two- and threefold axes of an icosahedron) to increase the number of observations. This implies that in many cases (e.g., an icosahedron) the present technique requires the determination of fewer unknown coefficients. In addition, the method presented here is completely flexible and general.

The electron density $\rho(\mathbf{x})$ in the h-cell is equal at all S NCS-related positions. Hence, $\rho(\mathbf{x}_1) = \rho(\mathbf{x}_2) = \cdots = \rho(\mathbf{x}_s)$. Now, $\mathbf{F_h}$ is given by

$$\mathbf{F_h} = \int_V \rho(\mathbf{x}) e^{-2\pi i \mathbf{h}\cdot\mathbf{x}}\, d\mathbf{x}.$$

However, $\rho(\mathbf{x})$ can be divided into S parts, each of which can be separately integrated within the h-cell and then summed to give $\mathbf{F_h}$. That is,

$$\mathbf{F_h} = \int_U \left\{\sum_{s=1}^{S} \rho(\mathbf{x}_s)\right\} e^{-2\pi i \mathbf{h}\cdot\mathbf{x}_s}\, dx,$$

where the integral is taken only over the volume U of the object and assuming the rest of the cell has zero density. Similar to the definition of (1) for $[D]$ (which relates a point in the p-cell to a point in the h-cell), the relationship between NCS points within the h-cell can be given as

$$\mathbf{x}_s = [C_s]\mathbf{x}, \tag{12}$$

where $[C_s]$ is the sth NCS operator relating a point $\mathbf{x}_s$ in the h-cell to another point $\mathbf{x}$ also in the h-cell. Then,

$$\mathbf{F_h} = \sum_{s=1}^{S} \int_U \rho(\mathbf{x}) e^{-2\pi i \mathbf{h}\cdot[C_s]\mathbf{x}}\, d\mathbf{x}. \tag{13}$$

However, the density $\rho(\mathbf{x})$ in the h-cell can be expressed as a Fourier summation such that

$$\rho(\mathbf{x}) = \frac{1}{V}\sum_{\mathbf{t}} \mathbf{F_t} e^{+2\pi i \mathbf{t}\cdot\mathbf{x}}, \tag{14}$$

where the Fourier coefficients have indices $\mathbf{s}(s,t,u)$. Substituting (14) in (13)

$$\begin{aligned}\mathbf{F_h} &= \frac{1}{V}\sum_{s=1}^{S}\sum_{\mathbf{t}} \mathbf{F_t} \int_U e^{+2\pi i(\mathbf{t}-[C_s^{\mathrm{T}}]\mathbf{h})\cdot\mathbf{x}}\, d\mathbf{x} \\ &= \frac{1}{V}\sum_{s=1}^{S}\sum_{\mathbf{t}} \mathbf{F_t} G_{\mathbf{ht}},\end{aligned} \tag{15}$$

where $G_{\mathbf{ht}}$ is the integral in the above expression. If, for instance, the object is a spherical virus, then the limits of the integral can be taken as a sphere, in which case $G_{\mathbf{ht}}$ can be expressed analytically as in expression (8) with an argument of $\theta = 2\pi H\mathscr{R}$, where $H = |(\mathbf{t} - [C_s^{\mathrm{T}}]\mathbf{h})|$ and $\mathscr{R}$ is the radius of the limiting sphere representing the viral envelope. Note that the coefficients, $G_{\mathbf{ht}}$, of these equations are always the same, independent of what orientation had been assumed to derive Eq. (10).

Equation (15) can be written for all m independent Fourier coefficients $\mathbf{F_h}$ within the limit of the desired resolution. The number of significant terms in each equation will depend on the NCS redundancy S. The greater the redundancy, the greater will be the number of terms with significantly large (approaching unity) $G_{\mathbf{ht}}$ terms and, hence, restricting the possible relationships among the Fourier coefficients. Each Eq. (15) will be an observational equation relating the m unknown $\mathbf{F_h}$ coefficients. These are of the same form as Eqs. (7) or (10) relating the known projection Fourier coefficients $\mathbf{Q}_{pq}$ with the unknown three-dimensional Fourier coefficients $\mathbf{F_h}$. Hence, Eq. (15) can be added to the normal Eqs. (11) prior to inversion. Thus, while the crystallographic symmetry elements can be constrained to determine only the crystallographic symmetry-independent $\mathbf{F_h}$ terms, the NCS is applied as a series of restraints to impose a desired relationship among the $\mathbf{F_h}$'s.

FINDING THE ORIENTATION OF AN OBJECT

The observational Eqs. (7) or (10) depend upon a knowledge of the object's orientation given by $[D]$ (Eq. (1)). If a false value of $[D]$ were used, it would not be possible to find a set of $\mathbf{F_h}$ coefficients that would satisfy Eqs. (7) or (10). Thus, a technique for finding the orientation of an object is to test all possible orientations of the particle in the h-cell. The normal equations are then set up for each orientation and solved for $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$ from which $\mathbf{F_h}$ is calculated. The $\mathbf{F_h}$ values can be substituted back in the right-hand sides of (10) to obtain a calculated set of $\mathbf{Q}_{pq,\mathrm{calc}}$. The

calculated $\mathbf{Q}_{pq,\mathrm{calc}}$ values are then compared with the observed values $\mathbf{Q}_{pq,\mathrm{obs}}$ (derived from a Fourier transform of the projected densities).

The search algorithm can be stated in terms of the following instructions.

1. Define the arbitrarily chosen asymmetric unit of the N crystallographic symmetry elements in the h-cell.

2. Select a point given by polar coordinates (κ,ψ,ϕ) within the reference crystallographic asymmetric unit of the h-cell and calculate the orientation matrix [D] relating the p- to the h-cell.

3. Determine the coefficients $a_{\mathbf{hp}}$ and $b_{\mathbf{hp}}$ of the observational Eqs. (10) for each two-dimensional Fourier coefficient $\mathbf{Q}_{pq,\mathrm{obs}}$ of the selected projection image. Add the contribution of each observational equation to the normal equations as shown by (11).

4. Add the NCS restraints given by (15) if appropriate. Note that the coefficients of the NCS restraining equations are the same for every orientation. The NCS restraints merely specify relationships among the $\mathbf{F}_{\mathbf{h}}$ Fourier coefficients.

5. Solve the normal equations for the unknown, three-dimensional, h-cell Fourier coefficients $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$.

6. Substitute the calculated values of $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$ into the right-hand sides of (10) to obtain calculated values of $\mathbf{Q}_{pq,\mathrm{calc}}$.

7. Compare $\mathbf{Q}_{\mathrm{obs}}$ with $\mathbf{Q}_{\mathrm{calc}}$ by using the four different criteria

$$R_1 = \frac{\sum_n |(|\mathbf{Q}_{\mathrm{obs}}| - k|\mathbf{Q}_{\mathrm{calc}}|)|}{\sum_n |\mathbf{Q}_{\mathrm{obs}}|} \times 100 \text{ (factor based on amplitude),}$$

where k is a scale factor and the sum is over all n $\mathbf{Q}_{\mathrm{obs}}$ values for the projection within the chosen resolution limits;

$$R_2 = \frac{\sum_n |(\alpha_{\mathrm{obs}} - \alpha_{\mathrm{calc}})|}{n} \text{ (factor based on phases),}$$

where α_{obs} and α_{calc} are the phases of $\mathbf{Q}_{\mathrm{obs}}$ and $\mathbf{Q}_{\mathrm{calc}}$, and their difference is taken as being the shortest distance around a circle of phases from 0° to 360°;

$$R_3 = \frac{\sum_n [|\mathbf{Q}_{\mathrm{obs}}|^2 + |\mathbf{Q}_{\mathrm{calc}}|^2 - 2|\mathbf{Q}_{\mathrm{obs}}||\mathbf{Q}_{\mathrm{calc}}| \cos(\alpha_{\mathrm{calc}} - \alpha_{\mathrm{obs}})]}{\left(\sum_n |\mathbf{Q}_{\mathrm{obs}}|^2 \cdot \sum_n |\mathbf{Q}_{\mathrm{calc}}|^2\right)^{1/2}} \times 100 \text{ (factor based on amplitudes and phase vectors);}$$

and

$$\mathrm{CC} = \frac{\sum_n (\langle|\mathbf{Q}_{\mathrm{obs}}|\rangle - |\mathbf{Q}_{\mathrm{obs}}|)(\langle|\mathbf{Q}_{\mathrm{calc}}|\rangle - |\mathbf{Q}_{\mathrm{calc}}|)}{\left[\sum_n (\langle|\mathbf{Q}_{\mathrm{obs}}|\rangle - |\mathbf{Q}_{\mathrm{obs}}|)^2 \times \sum_n (\langle|\mathbf{Q}_{\mathrm{calc}}|\rangle - |\mathbf{Q}_{\mathrm{calc}}|)^2\right]^{1/2}} \text{ (based on amplitudes),}$$

where CC is the correlation coefficient.

The $G_{\mathbf{hp}}$ values in (7), which give rise to the $a_{\mathbf{hp}}$ and $b_{\mathbf{hp}}$ coefficients in (10), are dependent only on the orientation of the projection direction $\mathbf{w}$ in the h-cell. However, the coefficients $\mathbf{Q}_{\mathbf{pq}}$ in (7), corresponding to the S_{pq} and T_{pq} terms in (10), are determined from the projected image data. Thus, the normal Eqs. (11) need only be inverted once for each direction $\mathbf{w}$. Search for the rotation, κ, about $\mathbf{w}$ can be achieved by interpolating $\mathbf{Q}_{pq}$ at rotated values of the two-dimensional projected p-cell lattice, giving rise to different S_{pq} and T_{pq} terms for each value of κ. Thus, the search over κ for a given orientation $\mathbf{w}$ is fast.

8. Go back to step 2 until all points have been explored. Map the result and look for minima in R_1, R_2, and R_3 or maxima in CC to obtain the object's orientation.

The different criteria may not always give the same discrimination (see, for instance, Table VII). However, in general R_3 appears to be the most sensitive criterion, being dependent on both amplitudes and phases.

The above procedure is useful for finding the orientation of a particle with ample symmetry in the absence of any other structural knowledge. It is, therefore, akin to the "common-lines" procedure of Crowther *et al.* (1970a) and could be compared to a "self-rotation function" in crystallography (Rossmann and Blow, 1962). Although these calculations are quite time-consuming, the orientation of any number of projections can be determined simultaneously because the right-hand sides of (10) are

dependent only on (κ,ψ,ϕ) and not on $\mathbf{Q}_{pq}$. Thus, the right-hand sides of (10) need only be computed and inverted once and then used to examine the orientation of each projection.

When a reasonable three-dimensional model of the object is already available, say by combining the results of a few initial projections, the resultant $\mathbf{F}_\mathbf{h}$ values can be used as in steps 6, 7, and 8 above to determine which projection of the current model gives two-dimensional $\mathbf{Q}_{pq}$ coefficients closest to the observed set. This procedure is not only much faster, but gives better results provided a reasonable starting $\mathbf{F}_\mathbf{h}$ set is available. It is equivalent to the polar Fourier transform (PFT) method (Baker and Cheng, 1996) or a cross-rotation function (Rossmann and Blow, 1962) against a known model in crystallography.

DEFINING A COMMON ORIGIN IN EACH PROJECTED IMAGE

The phases of the observed $\mathbf{Q}_{pq}$ Fourier coefficients will depend on the selected origin within the image. Furthermore, on combining data from each image (each $\mathbf{Q}_{pq}$ from every image gives rise to an observational Eq. (7) or (10)), it was assumed that the orientational relationship (1) places the common projection origin onto the origin of the *h*-cell. A reasonable common projection origin can be defined moderately easily for spherical particles by estimating the site of the center of the particle, although this may need a little refinement. A search over a much larger area will be required for an asymmetric object.

A shift in the projection origin to the point $(\Delta u,\Delta v)$ will change the phase of the $\mathbf{Q}_{pq}$ coefficients by $2\pi(p\Delta u + q\Delta v)$. This change will affect only the values of *S* and *T* in the left-hand sides of Eqs. (10). Since the matrix need only be inverted once for the given (κ,ψ,ϕ) value, the additional search over Δu and Δv is fast.

MATRIX ALGEBRA

Let the *h*-cell coordinate axes be *x*,*y*,*z*. These will now be rotated to correspond to the projected image of the object in the *p*-cell with axes *u*,*v*,*w*. First, rotate the *x*,*y*,*z* system about *z* by ψ to place the axes at x',y',z'. Then, rotate about x' by ϕ to place the axes at x'',y'',z''. Finally, rotate by κ about z'' to place the axes at x''',y''',z''' where these axial directions are now coincident with *u*,*v*,*w*. Hence,

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} \cos\kappa_3 & \sin\kappa_3 & 0 \\ -\sin\kappa_3 & \sin\kappa_3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi_2 & \sin\phi_2 \\ 0 & -\sin\phi_2 & \cos\phi_2 \end{pmatrix} \times \begin{pmatrix} \cos\psi_1 & \sin\psi_1 & 0 \\ -\sin\psi_1 & \cos\psi_1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

$$= [D^{-1}]$$

by comparison with Eq. (1).

If the object to be reconstructed is an icosahedron with 532 symmetry, it is convenient to orient the icosahedral axes along the *x*,*y*,*z* axes of the *h*-cell. However, that still allows two different orientations of the icosahedron related by a 90° rotation about any one of the *h*-cell axes. The definition of the standard icosahedral orientation used in the results given here is shown in Fig. 1. This icosahedron can be generated by applying in order the operators given in Table II.

DIFFERENTIATING BETWEEN CRYSTALLOGRAPHIC AND NONCRYSTALLOGRAPHIC SYMMETRY

The assumed symmetry of the object to be reconstructed from projection data can be given in terms

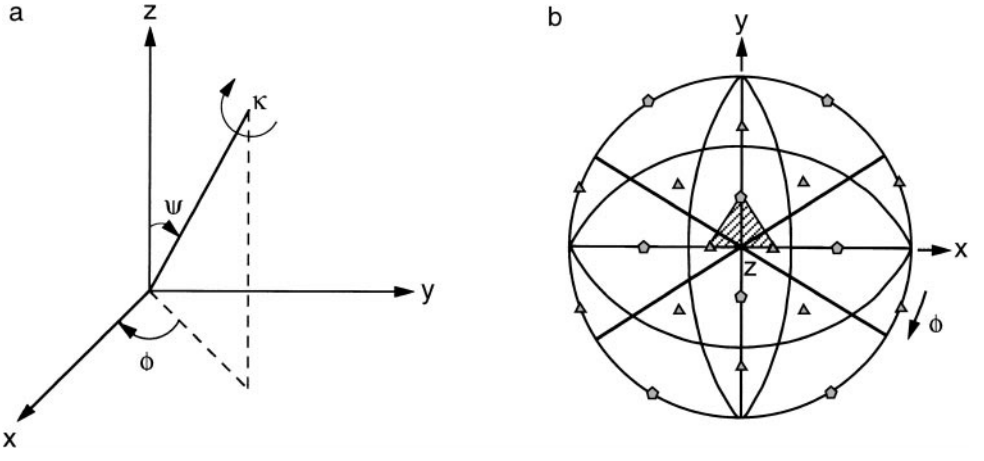

FIG. 1. Definition of the limits of an icosahedral asymmetric unit. (a) The polar coordinates (κ,ψ,ϕ). (b) Stereographic projection showing great circles through icosahedral twofold axes and the limits of one, arbitrarily selected, asymmetric unit (shaded).

TABLE II
Sequential Operators That Generate 532 Symmetry[a]

Number	κ	ψ	ϕ
1	72.0	31.71747	−90.0
2	180.0	36.0	−148.28253
3	180.0	36.0	148.28253
4	120.0	57.7356	−45.0

[a] The definition of the polar coordinates is the same as for those defined by Rossmann and Blow (1962).

of a sequential series of symmetry operations defined by the polar coordinates (κ,ψ,ϕ). Table II gives an example for generating an icosahedral distribution of 60 points in the *h*-cell. If the *h*-cell is defined as being of cubic shape, these 60 points can be separated into 5 sets of 12 each, where the 12 points within a set are related by the crystallographic symmetry of the *h*-cell lattice. In order to compute the coefficients in Eqs. (10), it will be necessary to select only the crystallographic symmetry operators, whereas the coefficients in Eq. (15) depend on the noncrystallographic symmetries. Thus, the operators that define the full symmetry of the object in the *h*-cell must be separated into crystallographic and noncrystallographic operators with respect to the cell parameters that were used to define the *h*-cell.

Let $[R_s]$ ($s = 1, 2, \ldots, S$) be the rotation matrices generated by the sequential symmetry operators such as those given in Table II. Let [α] and [β] be deorthogonalization and orthogonalization matrices for the *h*-cell (see Rossmann and Blow, 1962). Then,

$$\mathbf{x}_s = [\alpha][R_s][\beta]\mathbf{x} \tag{16}$$

(Rossmann and Blow, 1962) and

$$[C_s] = [\alpha][R_s][\beta].$$

Now, if this represents a crystallographic operation and if $\mathbf{x}$ is the position of a lattice point, then $\mathbf{x}_s$ should also be a lattice point. Thus, if $\mathbf{x}$ is the point (100), then $\mathbf{x}_s$ should have coordinates that are integers to within rounding error. Testing the points (100), (010), and (001) in turn will, therefore, establish whether the operation is a crystallographic operation. All other operators will be noncrystallographic.

DEFINING THE ASYMMETRIC UNIT IN REAL AND RECIPROCAL SPACE

When exploring the possible orientation of a projected object relative to the orientation of the symmetry elements in the *h*-cell, it is necessary only to explore the asymmetric unit of the object. There will be an equivalent orientation for particular (κ,ψ,ϕ) angles in each of the object's asymmetric units. For instance, if the object has icosahedral symmetry, then the asymmetric unit can be defined as a triangle limited by a fivefold axis and two threefold axes (Fig. 1). It will, therefore, be necessary to generate a set of (κ,ψ,ϕ) that are limited to an asymmetric unit. This can be readily achieved by exploring all angles in the range $0 \le \kappa < 360°$, $0 \le \psi < 180°$, $0 \le \phi < 360°$, from which those in the asymmetric unit are then selected as follows. The *h*-cell coordinates are calculated for each set of angles (κ,ψ,ϕ) from

$$\left.\begin{aligned} X &= \sin\psi\cos\phi \\ Y &= -\sin\psi\sin\phi \\ Z &= \cos\psi \end{aligned}\right\}. \tag{17}$$

All *S* symmetry-related positions are then calculated using the matrices $[R_s]$ ($s = 1, 2, \ldots, S$) and sorted in terms of *Z* (most significant) and *X* (least significant). Then, arbitrarily always selecting the top coordinates from the sorted list is equivalent to selecting the coordinate within the asymmetric unit of the object. Hence, if the original *X*,*Y*,*Z* is the same as that at the top of the sorted list, then the chosen (κ, ψ,ϕ) angles must be within the asymmetric unit of the object.

The same technique can be used to define the asymmetric unit of reciprocal space. However, in this case, the symmetry operators will be the crystallographic operators doubled by virtue of an additional center of symmetry at the origin of reciprocal space. As the operators were defined for real space (Table II), the operators in reciprocal space, when expressed as a rotation matrix, will be the transpose of those in real space. The Fourier coefficients (limited to a selected resolution limit) found to be in the asymmetric unit of reciprocal space are numbered in the sequence in which they are generated. That will then be the sequence of unknowns (1 to *m*) used for Eqs. (11).

REAL SPACE VERSUS RECIPROCAL SPACE

Although the above procedures have been described in reciprocal space, very similar procedures can be used in real space. In reciprocal space, the two-dimensional Fourier coefficients $\mathbf{Q}_{pq}$, corresponding to various projections $P(u,v)$, were used to determine the Fourier coefficients $\mathbf{F}_\mathbf{h}$ representing the unknown object in three dimensions. It would be equally possible to take the projected densities $P(u,v)$ and solve directly for the densities $\rho(x,y,z)$. The *h*-cell can be divided into a three-dimensional grid using a raster size commensurate to the resolution of the data. The asymmetric unit in the *h*-cell can be

defined as above, and the grid points within the asymmetric unit can be numbered sequentially. The problem is then to determine the electron density at each of the numbered grid points.

Any projection point $P(u,v)$ corresponds to the sum of the electron density along a line, w, normal to the projection as given by (2). The density along any given line along w can be represented by densities at a series of grid points separated by a distance that is about the same as the distance between the grid points in the h-cell. Each of these points will be surrounded by numbered grid points in the h-cell grid. Thus, if the density in the h-cell were known, then the density at the points along the line w could be interpolated from the, say, eight surrounding h-cell grid points. For instance, the density at the point u,v,w can be expressed as

$$\rho(u,v,w) = \sum_{i=1}^{8} a_i \rho(x_i,y_i,z_i), \qquad (18)$$

where i signifies the eight surrounding points used for interpolation and a_i is the interpolant whose value will depend on the distance of the point u,v,w from the point x_i,y_i,z_i. Thus, combining (2) and (18) gives an expression of the density $P(u,v)$ in terms of the known interpolants a_i and the unknown densities $\rho_i(x,y,z)$. Whenever a density $\rho(x,y,z)$ is required for interpolation, but is outside the asymmetric unit of the h-cell, then the symmetry-equivalent density must be used from within the asymmetric unit. The expression (18) represents the observational equations, which can then be solved for $\rho(x,y,z)$ in the same manner as was used for solution of the equivalent reciprocal space Eqs. (7) or (10).

There is, however, an important difference between real and reciprocal space, namely the number of unknowns that need to be determined. Because in real space it is necessary to sample the density sufficiently closely to obtain a realistic continuous distribution, it is generally necessary to sample density at intervals equivalent to at least one-third of the resolution. Thus, in three dimensions the number of unknowns that need to be solved will be about 27 times greater in real space than in reciprocal space. This creates an enormous computational burden, even for problems at very low resolution. For this reason, it is better to employ reciprocal space, although real space is easier to comprehend.

PHASE EXTENSION TO HIGHER RESOLUTION

The number of unknowns that need to be determined rises in proportion to the cube of the resolution. For an object of 500-Å diameter with icosahedral symmetry, it is necessary to solve a linear 341^2 diagonally symmetric matrix to determine the unknown $\mathbf{F_h}$ coefficients to 40-Å resolution (Table I). This is not an especially severe problem on even a slow modern workstation. However, at 10-Å resolution the problem has escalated to a $21\,817^2$ symmetric matrix.

Experience with phase extension for crystallographic electron density averaging (Rossmann, 1990) suggests a technique for reducing the computing effort. Fourier coefficients $\mathbf{F_h}$ at low resolution can be determined initially, and these are unlikely to change significantly as the resolution is extended. Hence, their values can be assumed and taken to the right-hand side of Eqs. (7) or (10), thus reducing the number of unknowns that need to be solved simultaneously. This procedure can be used to further extend the resolution in steps. The lowest resolution phases can be redetermined, this time accepting all the already determined coefficients at higher resolution. Unless the orientation of the individual projections is refined, the matrices for determining $\mathbf{F_h}$ at each resolution extension will not change from cycle to cycle. Thus, it will be fast to repeatedly cycle over successive resolution ranges for phase improvements between phase extensions.

TEST CALCULATIONS

The icosahedral atomic structure of human rhinovirus 16 (HRV16) (Oliveira *et al.*, 1993) was used as a test model for the reconstruction procedure described above. The maximum external diameter of this virus is 315 Å. The atomic model (Protein Data Bank Accession No. 1AYN) was placed into a cubic cell with a = 400 Å, with the icosahedral axes oriented as given in Table II and shown in Fig. 1b. Structure factors were calculated to 10-Å resolution, although the current tests used only data to 40-Å (in some cases) or 35-Å (in other cases) resolution.

TABLE III

The Rotation Function Peaks[a] for $\kappa = 72°$ at 40-Å Resolution

Peak no.	ψ	ϕ	Rotation function value	Number of σ above mean background[b]
1	30.0	270.00	69.3	3.64
2	15.0	258.76	14.8	0.00
3	16.5	241.72	14.4	−0.03
4	18.0	232.12	13.9	−0.06
5	18.0	345.80	10.2	−0.31

Note. The values are for the HRV16 structure placed into the h-cell in a standard orientation.

[a] A total of five peaks were found in a search with 1.5° intervals in ψ.

[b] Calculated from all points at 1.5° intervals within the icosahedral asymmetric unit.

TABLE IV
Polar Angles Defining the Orientations of Four Arbitrary Projections

Projection	κ	ψ	ϕ	Number of independent coefficients, $\mathbf{Q}_{pq}$, at 35-Å resolution
1	0.0	16.5	308.57	206
2	0.0	12.0	244.80	206
3	0.0	10.5	270.00	206
4	0.0	1.5	225.00	206

The first test was to check that the virus orientation was correct by calculating a rotation function (Rossmann and Blow, 1962) evaluated at 1.5° intervals in ψ and correspondingly equally spaced values in ϕ within the icosahedral asymmetric unit. This was easily performed with the routine used for setting the noncrystallographic constraints with Eqs. (15). A general peak-picking routine selected explicit peaks, as opposed to merely seeking grid points with large values of the function. The results showed peaks that were well above background at the anticipated NCS positions (Table III).

The Fourier coefficients for four different projections (Table IV) were calculated using expression (7). In this calculation, as well as in all subsequent calculations, the $G_{\mathbf{hp}}$ coefficients were used with a $3 \times 3 \times 3$ interpolation box with its center closest to the nonintegral point (h'',k'',l'') (see expression (9)). The outside radius of the virus was assumed to be 170 Å for these calculations. This will truncate the expansion in G just before the first node.

The four sets of two-dimensional Fourier coefficients were then used in various combinations to re-create the 35-Å resolution $\mathbf{F}_\mathbf{h}$ data set. The "reconstituted" 308 independent $\mathbf{F}_\mathbf{h}$ coefficients could then be compared with the original "true" coefficients to determine how well the procedure had succeeded (Table V). With four projections, using no NCS restraints, the results were very satisfactory. There was little difference when three projections were used, although the R-factors were very slightly worse. However, with two projections, the number of observations was distinctly worse, particularly when certain rather similar projections were used. Using only one projection, where the number of $\mathbf{Q}_{pq}$ coefficients was less than the number of $\mathbf{F}_\mathbf{h}$ coefficients, the results were essentially random (Table V). However, on applying NCS restraints, one projection was ample for producing a good reconstruction of the original $\mathbf{F}_\mathbf{h}$ coefficients.

The next test was to determine whether the orientation of a projection could be determined given at least an approximate set of $\mathbf{F}_\mathbf{h}$ values. For this, a set of $\mathbf{F}_\mathbf{h}$ values was reconstructed using projections 2, 3, and 4 with NCS restraints (Table IV), and these were used to find the orientation of projection 1. A search was conducted over ψ,ϕ within the icosahedral asymmetric unit using 1.5° intervals in ψ. ϕ was computed (as explained for the rotation function) to give 1.5° between projection normals for any

TABLE V
Reconstruction of 308 Three-Dimensional HRV16 $\mathbf{F}_\mathbf{h}$ Coefficients to 35-Å Resolution

(a) Using 4 projections (1, 2, 3, and 4 in Table IV)										
(i) With fivefold NCS restraints (number of solvable[a] $\mathbf{F}_\mathbf{h}$ values was 308)										
Resolution (Å)	110.7	78.3	63.9	55.3	49.5	45.2	41.8	39.1	36.9	35.0
Scale (k)	0.969	1.008	0.991	1.023	1.049	1.093	1.143	1.152	1.329	1.031
R_1 (amplitudes, %)	22.1	12.3	7.3	7.0	9.5	11.1	21.5	16.4	26.5	26.8
R_2 (phases, °)	11.7	12.2	1.9	15.8	11.0	9.2	16.4	11.2	20.0	22.0
(ii) Without restraints (number of solvable[a] $\mathbf{F}_\mathbf{h}$ values was 259)										
Scale (k)	1.002	0.994	1.000	1.012	1.003	0.948	0.982	1.001	1.001	0.992
R_1 (amplitudes, %)	0.7	2.1	2.3	2.8	2.0	9.5	6.8	5.2	2.7	2.0
R_2 (phases, °)	4.0	0.5	0.4	2.1	2.7	2.2	3.9	1.3	1.6	2.6
(b) Using 3 projections (1, 2, and 3 in Table IV)										
(i) With fivefold NCS restraints (number of solvable[a] $\mathbf{F}_\mathbf{h}$ values was 308)										
Resolution (Å)	110.7	78.3	63.9	55.3	49.5	45.2	41.8	39.1	36.9	35.0
Scale (k)	0.970	1.034	1.027	1.042	1.065	1.171	1.131	1.229	1.321	1.104
R_1 (amplitudes, %)	22.1	15.6	9.1	7.8	9.5	12.4	21.8	22.5	27.6	33.9
R_2 (phases, °)	11.7	19.8	2.3	10.5	12.3	13.9	20.4	15.4	21.1	26.3
(ii) Without restraints (number of solvable[a] $\mathbf{F}_\mathbf{h}$ values was 245)										
Scale (k)	1.002	1.001	0.989	1.039	1.019	0.954	0.971	1.014	1.010	0.988
R_1 (amplitudes, %)	0.9	3.2	7.2	8.4	5.1	16.7	10.7	7.2	4.4	2.9
R_2 (phases, °)	4.2	1.4	0.8	4.0	4.2	5.1	6.2	1.2	2.0	4.0

TABLE V—*Continued*

(c) Using 2 projections (1 and 2 in Table IV)

(i) With fivefold NCS restraints (number of solvable[a] $\mathbf{F_h}$ values was 308)

Resolution (Å)	110.7	78.3	63.9	55.3	49.5	45.2	41.8	39.1	36.9	35.0
Scale (k)	0.964	1.039	1.052	1.057	1.087	1.194	1.144	1.266	1.348	1.146
R_1 (amplitudes, %)	23.2	18.4	11.1	9.3	10.7	15.1	23.8	22.1	29.6	34.6
R_2 (phases, °)	11.8	25.3	2.6	13.9	13.6	14.7	24.1	16.5	20.9	26.8

(ii) Without restraints (number of solvable[a] $\mathbf{F_h}$ values was 244)

Scale (k)	1.005	1.006	0.989	1.052	1.034	0.924	0.952	1.032	0.987	0.964
R_1 (amplitudes, %)	1.4	4.5	9.0	10.8	10.4	19.7	17.8	17.1	7.7	7.1
R_2 (phases, °)	5.5	2.7	0.9	7.2	5.5	9.5	8.6	3.5	8.3	5.1

(d) Using 1 projection only (1 in Table IV)

(i) With fivefold NCS restraints (number of solvable[a] $\mathbf{F_h}$ values was 308)

Resolution (Å)	110.7	78.3	63.9	55.3	49.5	45.2	41.8	39.1	36.9	35.0
Scale (k)	0.935	1.035	1.133	1.073	1.145	1.232	1.216	1.255	1.417	1.171
R_1 (amplitudes, %)	26.0	27.4	18.4	15.8	11.8	18.2	26.3	25.2	30.8	37.7
R_2 (phases, °)	11.9	35.4	12.7	23.0	15.8	22.5	22.0	17.9	25.5	28.5

(ii) Without restraints (number of solvable[a] $\mathbf{F_h}$ values was 85)

Scale (k)	0.974	0.562	0.185	0.038	0.148	0.054	0.081	0.071	0.122	0.076
R_1 (amplitudes, %)	5.8	63.2	85.4	143.9	75.1	92.9	111.1	90.2	91.6	133.3
R_2 (phases, °)	22.4	45.3	75.6	74.2	91.1	75.6	88.2	77.7	83.5	60.8

Note. The two-dimensional Fourier projection coefficients $\mathbf{Q}_{pq,\text{obs}}$ were used to compute the three-dimensional Fourier $\mathbf{F_h}$ coefficients. These were then used to recalculate the two-dimensional $\mathbf{Q}_{pq,\text{calc}}$ coefficients for comparison with the original $\mathbf{Q}_{pq,\text{obs}}$.

[a] The number of solvable $\mathbf{F_h}$ values equals those whose diagonal terms in the normal equations were greater than 0.01 of the average value of all the diagonal terms.

specific value of ψ. The satisfactory results are given in Table VI.

Finally, a search was calculated where no prior information on the $\mathbf{F_h}$ values was assumed to be available. The data for projection 1 were used to calculate a set of $\mathbf{F_h}$ values using NCS restraints. These were then used to obtain a calculated set of $\mathbf{Q}_{pq}$ that were compared with the corresponding observed values. The results are given in Table VII and also show acceptable results, but not as good as

TABLE VI

Orientation Search for Projection 1 Given Three-dimensional $\mathbf{F_h}$ Coefficients Calculated from Projections 2, 3, and 4

Peak No.	ψ	ϕ	R_1 (amp) (%)	R_2 (phases) (°)	R_3 (vec) (%)	CC[a]	σ[b]
1	16.5	308.57	8.7	24.1	13.4	0.9749	4.32
2	19.5	245.85	23.7	51.6	74.7	0.7839	2.56
3	22.5	250.43	29.5	67.6	131.4	0.6637	0.94
4	4.5	360.00	33.3	73.6	139.1	0.6269	0.72
5	30.0	271.48	33.4	79.3	139.7	0.5781	0.70
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
14	10.5	212.73	30.1	84.3	194.0	0.6235	−0.86

Note. Shown are individual peaks sorted by R_3 (vec) (%). A total of 14 peaks were found in the search conducted with 1.5° intervals in ψ.

[a] CC is correlation coefficient.

[b] σ is the number of standard deviations above the mean for all calculated values of R_3 (vec) within the noncrystallographic asymmetric unit.

TABLE VII

Orientation Search for Projection 1 Given the Projection Two-Dimensional $\mathbf{Q}_{pq}$ Fourier Coefficients to 35-Å Resolution

Peak No.	ψ	ϕ	R_1 (amp) (%)	R_2 (phases) (°)	R_3 (vec) (%)	CC[a]	σ[b]
1	16.5	308.57	9.6	31.6	13.1	0.9747	2.52
2	19.5	302.93	15.5	46.6	29.5	0.9218	1.40
3	18.0	331.58	17.9	33.6	29.7	0.9077	1.39
4	15.0	219.38	15.3	38.8	31.2	0.9282	1.28
5	19.5	245.85	14.6	43.7	32.4	0.9319	1.20
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
24	30.0	271.48	23.0	62.0	69.5	0.8237	−1.32

Note. Shown are individual peaks sorted by R_3 (vec) (%). A total of 24 peaks were found in the search conducted with 1.5° intervals in ψ.

[a] CC is correlation coefficient.

[b] σ is the number of standard deviations above the mean for all calculated values of R_3 (vec) within the noncrystallographic asymmetric unit.

those obtained when an approximate set of $\mathbf{F_h}$ values was already available.

CONCLUSIONS

The experiments with HRV16 established that procedures discussed here are successful in the case of perfect data. Testing with real experimental data (Tao *et al.*, 1998) has only just begun. The procedure discussed here requires further development. A positional peak search has not yet been implemented, nor has the stepwise phase extension been tested. Least of all has there been an attempt to find a specific orientation of the RNA or DNA component in viruses or to examine other partially asymmetric objects using parts of the structure with higher symmetry. However, the techniques are completely general and can be made to accept any arbitrary symmetry.

We thank Tim Baker for the inspiration he has given us, without which this work would not have happened. We are grateful for helpful discussions with Terje Dokland, who suggested the simultaneous determination of orientations for sets of projection data. We thank Sharon Wilder and Cheryl Towell for their help in the preparation of the manuscript. The work was supported by grants from the National Institutes of Health and the National Science Foundation to M.G.R.

REFERENCES

Al Ani, R., Pfeiffer, P., Lebeurier, G., and Hirth, L. (1979) The structure of cauliflower mosaic virus. I. pH-induced structural changes, *Virology* **93,** 175–187.

Baker, T. S., and Cheng, R. H. (1996) A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy, *J. Struct. Biol.* **116,** 120–130.

Böttcher, B., Wynne, S. A., and Crowther, R. A. (1997) Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy, *Nature* **386,** 88–91.

Casjens, S., and Hendrix, R. (1988) Control mechanisms in dsDNA bacteriophage assembly, *in* Calendar, R. (Ed.), The Bacteriophages, pp. 15–75, Plenum, New York.

Chen, Z., Stauffacher, C., Li, Y., Schmidt, T., Bomu, W., Kamer, G., Shanks, M., Lomonossoff, G., and Johnson, J. E. (1989) Protein–RNA interactions in an icosahedral virus at 3.0 Å resolution, *Science* **245,** 154–159.

Conway, J. F., Cheng, N., Zlotnick, A., Wingfield, P. T., Stahl, S. J., and Steven, A. C. (1997) Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy, *Nature* **386,** 91–94.

Crowther, R. A. (1971) Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs, *Philos. Trans. R. Soc. London B* **261,** 221–230.

Crowther, R. A., and Amos, L. A. (1972) Three-dimensional image reconstructions of some small spherical viruses, *Cold Spring Harbor Symp. Quant. Biol.* **36,** 489–494.

Crowther, R. A., DeRosier, D. J., and Klug, A. (1970a) The reconstruction of a three-dimensional structure from projections and its application to electron microscopy, *Proc. R. Soc. London A* **317,** 319–340.

Crowther, R. A., Amos, L. A., Finch, J. T., DeRosier, D. J., and Klug, A. (1970b) Three-dimensional reconstructions of spherical viruses by Fourier synthesis from electron micrographs, *Nature* **226,** 421–425.

DeRosier, D. J., and Klug, A. (1968) Reconstruction of three dimensional structures from electron micrographs, *Nature* **217,** 130–134.

Frank, J. (1996) Three-Dimensional Electron Microscopy Assemblies, Academic Press, New York.

Frank, J., Zhu, J., Penczek, P., Li, Y., Srivastava, S., Verschoor, A., Radermacher, M., Grassucci, R., Lata, R. K., and Agrawal, R. K. (1995) A model of protein synthesis based on cryo-electron microscopy of the *E. coli* ribosome, *Nature* **376,** 441–444.

Grimes, J. M., Burroughs, J. N., Gouet, P., Diprose, J. M., Malby, R., Ziéntara, S., Mertens, P. P. C., and Stuart, D. I. (1998) The atomic structure of the bluetongue virus core, *Nature* **395,** 470–478.

Henderson, R. (1995) The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules, *Quart. Rev. Biophys.* **28,** 171–193.

Hoppe, W. (1974) Towards three-dimensional 'electron microscopy' at atomic resolution, *Naturwissenshaften* **61,** 239–249.

Hoppe, W., Gassmann, J., Hunsmann, N., Schramm, H. J., and Sturm, M. (1974) Three-dimensional reconstruction of individual negatively stained yeast fatty-acid synthetase molecules from tilt series in the electron microscope, *Hoppe-Seyler Z. Physiol. Chem.* **355,** 1483–1487.

Kleywegt, G. J., and Read, R. J. (1997) Not your average density, *Structure* **5,** 1557–1569.

Larson, S. B., Koszelak, S., Day, J., Greenwood, A., Dodds, J. A., and McPherson, A. (1993) Double-helical RNA in satellite tobacco mosaic virus, *Nature* **361,** 179–182.

Oliveira, M. A., Zhao, R., Lee, W. M., Kremer, M. J., Minor, I., Rueckert, R. R., Diana, G. D., Pevear, D. C., Dutko, F. J., McKinlay, M. A., and Rossmann, M. G. (1993) The structure of human rhinovirus 16, *Structure* **1,** 51–68.

Olson, N. H., Kolatkar, P. R., Oliveira, M. A., Cheng, R. H., Greve, J. M., McClelland, A., Baker, T. S., and Rossmann, M. G. (1993) Structure of a human rhinovirus complexed with its receptor molecule, *Proc. Natl. Acad. Sci. USA* **90,** 507–511.

Prasad, B. V. V., Burns, J. W., Marietta, E., Estes, M. K., and Chiu, W. (1990) Localization of VP4 neutralization sites in rotavirus by three-dimensional cryo-electron microscopy, *Nature* **343,** 476–479.

Rossmann, M. G. (1990) The molecular replacement method, *Acta Crystallogr. Sect. A* **46,** 73–82.

Rossmann, M. G. (1995) *Ab initio* phase determination and phase extension using non-crystallographic symmetry, *Curr. Opin. Struct. Biol.* **5,** 650–655.

Rossmann, M. G., and Argos, P. (1980) Three-dimensional coordinates from stereodiagrams of molecular structures, *Acta Crystallogr. Sect. B* **36,** 819–823.

Rossmann, M. G., and Blow, D. M. (1962) The detection of sub-units within the crystallographic asymmetric unit, *Acta Crystallogr.* **15,** 24–31.

Rossmann, M. G., McKenna, R., Tong, L., Xia, D., Dai, J., Wu, H., Choi, H. K., and Lynch, R. E. (1992) Molecular replacement real-space averaging, *J. Appl. Crystallogr.* **25,** 166–180.

Saibil, H. (1996) The lid that shapes the pot: Structure and function of the chaperonin GroES, *Structure* **4,** 1–4.

Smith, T. J., Olson, N. H., Cheng, R. H., Chase, E. S., and Baker, T. S. (1993) Structure of a human rhinovirus-bivalently bound antibody complex: Implications for viral neutralization and antibody flexibility, *Proc. Natl. Acad. Sci. USA* **90,** 7015–7018.

Tao, Y., Olson, N. H., Xu, W., Anderson, D. L., Rossmann, M. G., and Baker, T. S. (1998) Assembly of a tailed bacterial virus and its genome release studied in three dimensions, *Cell* **95,** 431–437.