USE OF AN ELEMENTARY GROUP THEORETICAL METHOD IN DETERMINING THE STRUCTURE OF A BIOLOGICAL CRYSTAL FROM ITS PATTERSON FUNCTION

D.B. Litvin

Crystallographers have been using X-rays to investigate the structure of biologically important macromolecules for over forty years[1]. One type of these are the so-called spheroidal or globular macromolecules, such as myoglobin, which is responsible for the storage of oxygen in muscle tissue, and hemoglobin, which is responsible for transporting oxygen in the blood stream. There are other types, such as the fibrous, fiber-like macromolecules found in for example hair and skin, but we will limit our interest to the globular type. To have an idea of the size, weight, and number of atoms in such macromolecules, in Table 1 we compare these characteristics of some typical globular macromolecules with those of some familiar molecules.

Most globular macromolecules can be crystallized. In forming a crystal these macromolecules are not to any large extent distorted, there are in general only a few molecules in the unit cell of the crystal, and the identity of each molecule is preserved. The term "biological crystal" in the title of this talk refers to such a crystal.

The object of investigating the structure of such biological crystals is to determine the structure of the macromolecules. To determine the structure of the crystal one attempts to calculate

TABLE 1

Molecule	Molecular weight	Size	Number of atoms
water H_2O	18	4Å across	3
benzine H_6C_6	78	6Å diameter	12
myaglobin	17,000	spheroid 44Å × 44Å × 25Å	~2,600
hemoglobin	64,000	spheroid 65Å × 55Å × 50Å	~10,000
disk of TMV protein coat	600,000	160Å in diameter	~100,000

the electron density $\rho(r)$ of the crystal via the Fourier expansion.

$$\rho(r) = \frac{1}{V} \sum_k F(k)e^{-2\pi ik \cdot r}$$

where the Fourier coefficients $F(k) = |F(k)|e^{i\alpha(k)}$ are called the complex structure factors, $\alpha(k)$ the phase of the structure factor, and V is the volume of the unit cell of the crystal. Because of the translational symmetry of the crystal, there is a summation over all vectors k of the reciprocal lattice associated with the translation group of the crystal. One then interprets this electron density in terms of the atomic structure of the macromolecule.

From experimental X-ray diffraction data one obtains the magnitude of the structure factors $|F(k)|$. (In practice one does not obtain $|F(k)|$ for all values of k, and in fact uses only an approximation for the above expression for the electron density of the crystal. However, we will not discuss this complication.) Knowing only the magnitude of the structure factors one cannot determine the electron density; one needs also the phases.

To determine the phases the following method has been used: the X-ray data from the crystal containing the molecules whose structure we want to determine, is compared with the X-ray data from the same crystal where in each molecule a small number of atoms have been replaced by a second set of atoms which contains a heavy atom. This method is called the "isomorphous replacement method". The differences in the X-ray data are used to determine the phases. Using this method of determining the phases, the electron density maps of myoglobin and hemoglobin were determined in 1959.

To investigate the structure of these biological macromolecules and bypass the need for artificially altering the structure of the macromolecules, in the past fifteen years the so-called "molecular replacement" method has been developed[2]. This method, which we will discuss here, uses only the magnitude of the

structure factors $|F(k)|$ which are obtained from the X-ray data. This method consists of three steps.

1) The rotation problem: determine the point group of the molecules in the crystal, and their orientation with respect to the translational symmetry vectors of the crystal.

2) The translation problem: determine the vectors between the molecular positions in the crystal. These vectors are called translation vectors.

3) The phase problem: using the information obtained in the first two steps, determine the phases of the structure factors.

With respect to the third step, I will only comment that there has been so far only limited success in determining phases using this method[9]. What has been successful recently is the use of the information obtained in the first two steps in combination with the isomorphous replacement method[10]. I will concentrate on the first two steps. Procedures were introduced over ten years ago to generate data from which one is able to solve the rotation and translation problems. However during this time no general method was derived to analyze this data systematically, and consequently, as I will show, misinterpretations of data have occurred. I will show that a general method to systematically analyze this data can be formulated, and that to do so one needs to introduce only truly elementary group theoretical arguments. It is surprising that these elementary group theoretical arguments were not applied to this problem until very recently.

The solutions of the rotation and translation problems are based on the analysis of a function which can be calculated from the magnitude of the structure factors $|F(k)|$. This is the Patterson function introduced in 1934.

$$P(r) = \frac{1}{V} \sum_{k} |F(k)|^2 e^{2\pi i k \cdot r}.$$

This function is the self-convolution (self-correlation) of the electron density of the crystal.

$$P(r) = \int \rho(r')\rho(r'+r)dr'.$$

Because in the biological crystal the macromolecules preserve their identity, the electron density of the crystal can be written as a sum of the electron densities of the molecules, and the Patterson function can be written as

$$P(r) = \sum_{j,k,t} P_{jkt}(r)$$

where j and k index the molecules in the unit cell of the crystal, t the translations of the crystal, and

$$P_{jkt}(r) = \int \rho_{j0}(r')\rho_{kt}(r'+r)dr'$$

the convolution of the electron density of the jth molecule in the unit cell with the kth molecule in the tth unit cell. The convolution function $P_{jkt}(r)$, in general, is called a cross-Patterson function, and in the special case where j = k and t = 0, $P_{jj0}(r)$ is called a self-Patterson.

The two problems:

1. The rotation problem

Here one wants to determine the point group and orientation of the molecules in a biological crystal. The crystals which we consider are assumed to be made up of only one kind of biological macromolecule.

To determine this, one looks for relations between the point group and orientation of the molecules in the crystal and properties of the Patterson function. Consider the electron density $\rho_{j0}(r)$ of the jth molecule in the unit cell. This electron density is localized because of the finite dimension of the molecule,

about the center of mass r_j of the jth molecule. The symmetry
point group of the molecule is the set of all proper rotation mat-
rices P such that

$$\rho_{j0}(r_j + Pu) = \rho_{j0}(r_j + u).$$

(Because the biological macromolecules are made up of "left-handed"
amino acids, the symmetry point group of the molecule consists
only of proper rotation matrices.) We choose a coordinate system
in the crystal; then the group of matrices P, which is the symme-
try point group of the jth molecule, is defined with respect to
this coordinate system. The orientation of the molecule is the
orientation of the rotation axes of the rotations which are re-
presented by these matrices. Another molecule in the crystal has
as its symmetry point group a group of matrices P' also defined
with respect to this coordinate system. The two groups of mat-
rices are in general different but equivalent, i.e. they both be-
long to the same class of point groups and are denoted by the same
symbol in, e.g., international notation. When referring to the
point group of a molecule we will sometimes mean a specific group
of matrices defined in the chosen coordinate system, and at other
times, the class of point groups to which this group of matrices
belong. The meaning should be clear from the context.

It follows that if P is the symmetry point group of the jth
molecule, then the self-convolution of the electron density of the
j^{th} molecule, i.e. the jth self-Patterson function

$$P_{jj0}(r) = \int \rho_{j0}(r')\rho_{j0}(r'+r)dr$$

is invariant under all rotations of P, i.e., $P_{jj0}(r) = P_{jj0}(Pr)$.
$P_{jj0}(r)$ is also invariant under inversion, and consequently $P \times \bar{1}$
is an invariance point group of $P_{jj0}(r)$. We shall assume that
$P \times \bar{1}$ is the symmetry point group of $P_{jj}(r)$.

As this self-Patterson is localized in a volume about the

origin of the Patterson function, Rossmann and Blow[3], in order to
determine from the Patterson function the point group and orienta-
tion of the molecules, introduced in 1962 the rotation function
R(A):

$$R(A) = \int_U P(r)P(Ar)dr$$

where A is a proper rotation, and the integration is over a volume
about the origin of the Patterson function. This is an overlap
integral of a volume about the origin of the Patterson function
with a rotation image of the same volume. Relative maxima of this
rotation function, as a function of A, are called peaks of the ro-
tation function. Obviously if $P_{jj0}(r)$ is invariant under a rota-
tion P then there will be a peak in the rotation function at A = P.
Consequently by determining the peaks of this rotation function,
one can obtain information on the point group of the molecules in
the crystal. While this rotation function has been successfully
used, a general method has been available for less than a year to
determine systematically the point group symmetry of the molecules
from information provided by the rotation function[4]. To derive
this method, one needs to apply only very elementary group theo-
retical arguments.

Consider a biological crystal consisting of identical mole-
cules generated by a space group G from a single molecule at posi-
tion r_1. Let T denote the translational subgroup of G, and
$(R_j|\tau_j)$ j=1,...,n, the coset representatives of T in G. We will
consider the case where r_1 is a general position, i.e. the n vec-
tors $r_j = (R_j|\tau_j)r_1$ are distinct. We then have n molecules in the
unit cell, and the electron density of the molecule at r_j is re-
lated to the electron density of the molecule at r_1 by

$$\rho_{j0}(r_j+u) = \rho_{10}(r_1+R_ju)$$

where $r_j = (R_j|\tau_j)r_1$. That is, we have n identical molecules in
different orientations in the unit cell, and their mutual orienta-

tion is determined by the rotations of the space group of the crystal. It also follows that the self-Pattersons $P_{jj0}(r)$, j=1,...,n, which are all localized about the origin of the Patterson, are identical, in different orientation, and their mutual orientations are also determined by the rotations of the space group of the crystal.

All peaks of the rotation function correspond to rotations which

1) Leave a self-Patterson $P_{jj0}(r)$, for some j, invariant; or

2) Rotate a self-Patterson $P_{jj0}(r)$ into the orientation of a self-Patterson $P_{kk0}(r)$, where j ≠ k.

Let P denote the symmetry point group of the molecule at r_1:

1) The group of rotations $\{R(jj)\} \equiv \{R_j PR_j^{-1}\}$ is the symmetry point group of the self-Patterson $P_{jj0}(r)$; and

2) The set of rotations $\{R(jk)\} = \{R_k PR_j^{-1}\}$ is the set of all rotations which rotates $P_{jj0}(r)$ into the orientation of $P_{kk0}(r)$.

Therefore, all peaks of the rotation function R(A) correspond to all the rotations contained in the set of rotations[4]:

$$[\{R(jk)\}|j,k=1,\ldots,n]. \tag{1}$$

One now has a systematic method to analyze the data obtained from the rotation function.

1) From the rotation function calculate all rotations which correspond to all peaks.

2) Determine the point groups such that the set of distinct rotations in (1) is identical with the set of rotations corresponding to peaks of the rotation function. If P is such a point group, the conjugate point groups $R_j PR_j^{-1}$, j=1,...,n, where R_j is a rotation of the space group of the crystal, are also such point groups. If there is only one set of such conjugate subgroups $R_j PR_j^{-1}$, j=1,...,n, we will say that P is the symmetry point group

of the molecules. The orientation of the rotation axes of these
conjugate subgroups, with respect to the translational vectors of
the crystal, determine the orientation of the molecules in the
crystal. If there is more than one set of such conjugate sub-
groups, the solution of the rotation function problem is not
uniquely determined by this method. However, in practice it is
impracticable to search for all peaks of the rotation function,
i.e. to calculate R(A) for all possible rotations A. One must
then use an alternate method:

1) From the rotation function calculate the rotations corre-
sponding to some of the peaks.

2) Determine the point groups such that the set of distinct
rotations of (1) includes all those rotations found in step 1 from
the rotation function.

3) Determine if there are peaks of the rotation function cor-
responding to the additional rotations of (1).

Example. Satellite Tobacco Necrosis Virus (STNV). A few
years ago there was a debate as to whether the protein coat of
this "spherical" virus was of cubic O(432) or icosahedral J(532)
point group symmetry. A rotation function study was then made of
a crystal containing two STNV molecules in the unit cell[5]. The
crystal was monoclinic of space group symmetry C_2^3(C2), and
the orientations of the two STNV molecules were related by a ro-
tation of 180°. It was felt that a rotation function study of
this crystal would easily determine the point group since there
are 4-fold rotations contained in the cubic point group and none
in the icosahedral, and 5-fold rotations in the icosahedral and
none in the cubic.

A set of strong peaks were found with corresponding rotations
which were exactly those proper rotations of a cube, and this was
interpreted as meaning that the STNV molecules were of cubic sym-
metry. There were peaks corresponding to 5-fold rotations, which

are characteristic of icosahedral symmetry, but these peaks were much lower than the cubic peaks.

This interpretation was immediately challenged and it was shown that all the peaks of the rotation function corresponding to 5-fold rotations could be interpreted as two molecules of icosahedral symmetry in two different orientations related by the rotation of 180° of the space group of the crystal[6]. There is general agreement that this is the correct interpretation and that the STNV molecules do have icosahedral symmetry. However in this reinterpretation the stronger cubic peaks were explained away in an argument that approximated this monoclinic crystal as being cubic!

All these peaks can be explained using the above formalism and taking the point group of the STNV molecules as being icosahedral[4]: If one calculates the set of rotations in (1) taking $j,k=1,2$, R_2 the rotation of 180° of the space group, and P as the icosahedral symmetry point group of one of the molecules, one finds 240 rotations. These include the icosahedral rotations of both molecules, and a set of cubic rotations, exactly that set of cubic rotations determined from the rotation function. Each of these cubic rotations either leaves both molecules invariant or interexchanges the two orientations, explaining the high corresponding peaks, since all other rotations either leave only one molecule invariant, or rotate one molecule into the orientation of the other. One finds also that this set (1) of rotations contains additional rotations which have not yet been determined. However, even without determining peaks of the rotation function corresponding to these rotations, it does seem that the symmetry point group of the STNV molecules is icosahedral.

2. The translation function

Information on the point group and orientation of the molecules in a biological crystal is found, using the rotation function, by considering that volume of the Patterson about the origin

of the Patterson function. Information on the translation vectors between molecules is found using a similar method, but considering other parts of the Patterson. One uses a so-called translation function $T(x,A)$:[7]

$$T(x,A) = \int_U P(x+r)P(x+Ar)dr.$$

The translation function like the rotation function is an overlap integral of a volume U of the Patterson function with a rotated image of the same volume, but unlike the rotation function, the centre of the volume is now a variable, and not restricted to be at the origin of the Patterson function.

This translation function is non-zero when

1) The volume U intersects with a cross-Patterson $P_{jkt}(r)$ and the intersection is left invariant by the rotation A about x; or

2) The volume U intersects two cross-Patterson $P_{jkt}(r)$ and $P_{j'k't'}(r)$ and one is transformed into the other by the rotation A about x. The relative maxima of a translation function $T(x,A)$ as a function of x, for constant A, are called the peaks of the translation function, and the positions of these peaks are related to the translation vectors between molecules in the crystal. In Figure 1 we have drawn a model crystal, in Figure 2 the corresponding Patterson function, and in Table 2, the positions of the peaks of the corresponding translation function $T(x,A)$. The group theoretical arguments which enter into determining the relationship between the peaks of the translation function and the translation vectors between the molecules are similar to those used in predicting the peaks of the rotation function:

Necessary and sufficient conditions that the translation function $T(x,A)$ has non-zero values associated with the transformation of $P_{jkt}(r)$ into $P_{j'k't'}(r)$ are:
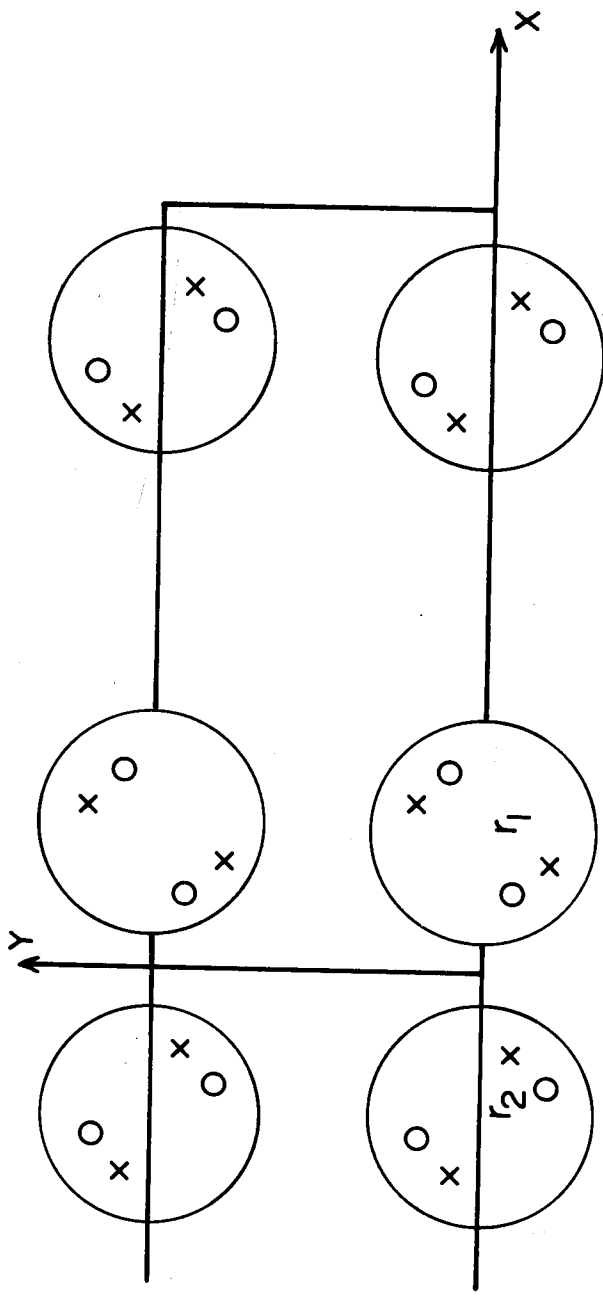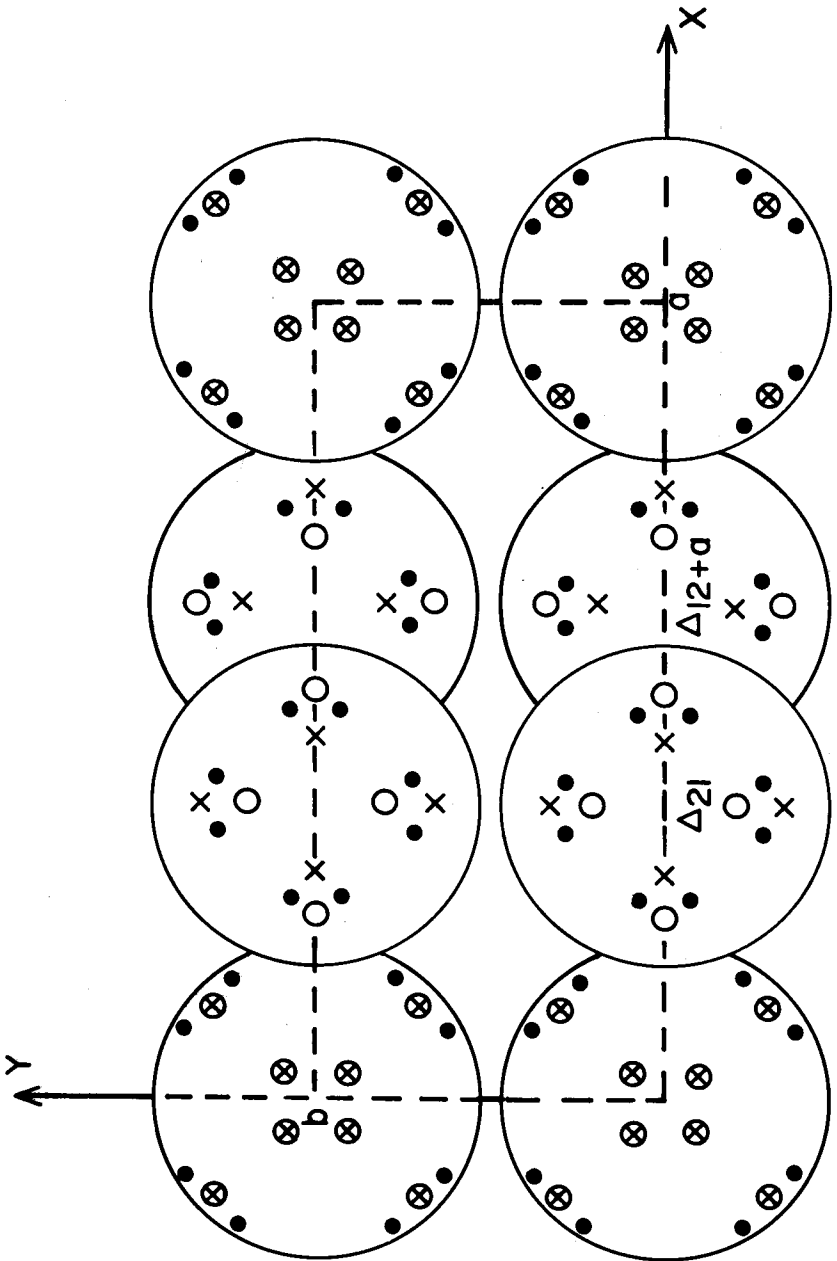
TABLE 2

A	x
2_z	$\Delta_{21},\ \Delta_{21} + t_x,\ \frac{1}{2}t_y,\ \Delta_{21} + \frac{1}{2}t_y,\ \Delta_{12} + t_x + \frac{1}{2}t_y$
2_{xy}	$\Delta_{21},\ \Delta_{12} + t_x$
$2_{\bar{x}y}$	$\Delta_{21},\ \Delta_{12} + t_x$
2_x	$\frac{1}{2}t_y$
4_z	$\frac{b}{2}\,(1,1,0),\ \frac{b}{2}\,(-1,1,0) + t_x$
4_z^3	$\frac{b}{2}\,(1,1,0),\ \frac{b}{2}\,(-1,1,0) + t_x$
m_x	$\Delta_{21},\ \Delta_{12} + t_x$
m_y	$\Delta_{21},\ \Delta_{12} + t_x,\ \frac{1}{2}t_y,\ \Delta_{21} + \frac{1}{2}t_y,\ \Delta_{12} + t_x + \frac{1}{2}t_y$
$\bar{1}$	$\frac{1}{2}t_x,\ \frac{1}{2}(t_x+t_y)$
$\bar{4}_z$	$\Delta_{21},\ \Delta_{12} + t_x,\ \frac{b}{2}\,(1,1,0),\ \frac{b}{2}\,(-1,1,0) + t_x$
	$\Delta_{21} + \frac{b}{2}\,(1,1,0),\ \Delta_{21} + \frac{b}{2}\,(-1,1,0)$
	$\Delta_{12} + t_x + \frac{b}{2}\,(1,1,0),\ \Delta_{12} + t_x + \frac{b}{2}\,(-1,1,0)$
$\bar{4}_z^3$	$\Delta_{21},\ \Delta_{12} + t_x,\ \frac{b}{2}\,(1,1,0),\ \frac{b}{2}\,(11,1,0) + t_x$
	$\Delta_{21} + \frac{b}{2}\,(1,1,0),\ \Delta_{21} + \frac{b}{2}\,(-1,1,0)$
	$\Delta_{12} + t_x + \frac{b}{2}\,(1,1,0),\ \Delta_{12} + t_x + \frac{b}{2}\,(-1,1,0)$

1) the rotation A is such that

$$P_{jkt}(r_k - r_j + t + A^{-1}u) = P_{j'k't'}(r_{k'} - r_{j'} + t' + u);$$

2) $A(r_k - r_j + t - x) = r_{k'} - r_{j'} + t' - x;$

3) the vector $\frac{1}{2}y$, where $-y = r_k - r_j + t - x$ is within the volume U.

The first two conditions demand that $P_{jkt}(r)$ is transformed into $P_{j'k't'}$ by a rotation A about the point x, and the third condition, that the volume U of the Patterson function centered at x intersects both $P_{jkt}(r)$ and $P_{j'k't'}(r)$. Using group theoretical arguments like those used for the rotation function one can show that all rotations which satisfy the first condition are those denoted by $\{A(jkj'k')\}$

$$\{A(jk,j'k')\} = [\{R(jj')\} \cap \{R(kk')\}] + \bar{I}[\{R(jk')\} \cap \{R(kj')\}].$$

If this set of rotations is empty, then $P_{jkt}(r)$ and $P_{j'k't'}(r)$ are not congruent. For rotations A contained in $\{A(jkj'k')\}$, the positions x are calculated from the second condition.

The concept of the translation function $T(x,A)$ was introduced by Rossmann, Blow, Harding and Coller in 1964[8]. However they considered only the case where $A = m$, and only in this case has the translation function been put to actual use with some success. Why these people use only this special case is unclear; it is probably because no general method was developed to predict the peaks of the translation function. One now has such a general method, which needs only very elementary group theoretical arguments to derive.

REFERENCES

1. Progress in Physics, Biophysics, a reprint series. W.A. Benjamin, New York, 1969 and the reviews therein by L. BRAGG and A.C.T. NORTH.

2. The molecular replacement method, M.G. Rossmann, ed., New York, Gordon & Breach, 1972.

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

4. D.B. LITVIN, Acta Cryst. A31, 407 (1975).

5. K. AKERVALL, B. STRANDBERG, M.G. ROSSMANN, V. BENGTSSON, K. FRIDBORG, H. JOHANNISON, K. KANNAN, S. LOVGREN, G. PETEF, B. OBERG, D. EAKER, S. HJERTEN, L. RYDEN and I. MORING, Cold Spring Harbour Symp. Quant. Biol. 36, 469, 487 (1971).

6. A. KLUG, Cold Spring Harbour Symp. Quant. Biol. 36, 483 (1971).

7. D.B. LITVIN, Acta Cryst. A33 (1977).

8. M.G. ROSSMANN, D.M. BLOW, M.M. HARDING and E. COLLER, Acta Cryst. 17, 338 (1964).

9. A. JACK, Acta Cryst. A29, 545 (1973).

10. J.N. CHAMPNESS, A.C. BLOOMER, G. BRICOGNE, P.J.G. BUTLER and A. KLUG, Nature 259, 20 (1976).