DL/SCI/R23

MOLECULAR REPLACEMENT

Proceedings of the Daresbury Study Weekend,
15-16 February, 1985

Compiled by P. A. Machin

Science & Engineering Research Council
DARESBURY LABORATORY
Daresbury, Warrington WA4 4AD

ISSN 0144-5677

MOLECULAR REPLACEMENT

Proceedings of the Daresbury Study Weekend
15-16 February, 1985

Compiled by
P. A. Machin
Daresbury Laboratory

Science & Engineering Research Council
DARESBURY LABORATORY
1985

FOREWORD

The use of molecular replacement, for the determination of unknown protein structures which are homologous, or related to proteins whose structure is established, is an increasingly important component in x-ray analysis. It is an obvious tool to use for the elucidation of the structure of the modified proteins, which can be expected from the protein engineering initiative. Many research groups have tangled with this computationally simple but practically difficult technique, and one of the aims of this meeting was to analyse its unpredictable successes and failures. A theoretical introduction to the method and notation proved an essential preliminary to the subsequent discussion of results.

The meeting was organised and supported financially by the SERC Collaborative Computing Project in Protein Crystallography (CCP4) Daresbury Laboratory. The success that the meeting achieved was undoubtedly due to the considerable efforts of the invited speakers and our thanks are due to them for their talks and co-operation in the preparation of these proceedings. Particular thanks go to Eleanor Dodson and the Working Group of CCP4 for the time and effort they invested in planning the meeting.

We thank the Daresbury Laboratory and its Director, Professor L.L. Green, for the provision of organisational help and support, both for the meeting itself and for the publication of the proceedings. In particular we thank Mrs. Shirley Lowndes and Miss Karen Maunders and the Technical and Scientific Information Services staff, for their considerable assistance with the organisation of the meeting.

PELLA MACHIN

CONTENTS

INTRODUCTION

by

G. Dodson
Department of Chemistry, University of York, Heslington, York YO1 5DD

The achievements of crystallography in determining the structures of proteins rests on the remarkable success of the isomorphous substitution by heavy atoms in defining phases. There has always been, however, a persistent interest in exploiting the local symmetry that occurs fairly often in protein crystals (e.g., α-chymotrypsin and 2Zn insulin, both worked on by David Blow and Michael Rossmann). It is from these studies that many of the techniques and ideas of molecular replacement originated. Before the three-dimensional structures were known, these approaches were principally directed towards defining subunit organisation and symmetry. There was also the more distant prospect of using the structural redundancy for determining the phase or reducing the phase uncertainties. Gerard Bricogne has developed and applied very successfully these ideas in order to improve the electron density in an asymmetric unit with multiple copies of the molecule.

The character of the molecular replacement method for structure determination is now quite different; it usually involves relating a model structure, derived from an already established crystal structure, to the molecule (or molecules) in an unknown crystal. This has been stimulated first by the increasing and proper interest in related structures which, through altered crystal contacts or sequence, can reveal better such properties as ligand binding and structural behaviour. Secondly, there prove to be striking structural homologies between proteins of unrelated functions - and sometimes of unrelated sequences. The widespread appearance of the so-called nucleotide binding domain, first described by Michael Rossmann, the large families of trypsin-like enzymes, globins, insulins and immunoglobulins, all illustrate the variety and breadth of protein structural homologies. There has been some exploitation of these homologies and there will clearly be more, in spite of the complexities generated in these calculations by structural variation. The use of fragments or domains as successful starting points in structure determinations is as yet only a possibility; in favourable circumstances it does seem a possible tactic.

There are two major developments in molecular biology which are very relevant to protein crystallography and to molecular replacement methods in particular. The first of these is the technique of site-directed mutagenesis through which amino acid side chains can be specifically substituted in a protein; this has already led to x-ray studies on mutated proteins. There is an obvious partnership between the genetic experiments, with their specific variation of protein sequences, and protein crystallography, which will provide the essential framework for relating sequence, structure and function.

The second development is the monoclonal antibody. Already the crystal structure of lysozyme:-monoclonal antibody Fab complex has been reported at 6Å resolution. In addition to the obvious importance in getting detailed descriptions of antigen-antibody interactions, monoclonal antibody:protein complexes offer a possible route for structural determination through molecular replacement calculation on the known immunoglobulin moiety. More speculatively, this kind of complex may enable previously uncrystallisable proteins to be crystallised - with a built-in and efficient pathway to the determination of the structure. Hence this technique also could add greatly to the importance of molecular replacement methods in protein structural research.

These developments are likely to present protein crystallographers, therefore, with the challenge of solving quickly a selected series of structurally altered proteins: thus, molecular replacement methods now have new and immediate importance.

It is clear that a workshop concerned with the techniques (and notation!) used in molecular replacement and the problems and successes encountered in their application, is excellently timed, and we can expect new ideas and wider confidence will be generated by the papers and discussions.

INTRODUCTION TO ROTATION AND TRANSLATION FUNCTIONS

by

D.M. Blow
Blackett Laboratory, Imperial College, London SW7 2BZ

1. NON-CRYSTALLOGRAPHIC SYMMETRY

Suppose we have two identical objects in different positions. To superimpose one on the other we must rotate it and translate it. We can specify a rotation about the origin by three variables, which can be two to specify the direction of the rotation axis by its latitude and longitude, and one for the amount of rotation. Then we need another three variables to specify the translation - for example the amount of motion in the x, y and z directions.

There are many other ways the operation can be done, but they will always require six variables to be specified. For example, it is possible to choose the position of the rotation axis in such a way that the only subsequent translation which needs to be done is parallel to the axis. We still need three variables to specify the translation: now there are two to give the position of the axis (in a plane perpendicular to it) and one for the amount of translation parallel to the axis. There are also different ways of specifying the rotation.

This symposium relates to a range of crystallographic problems which arise when such a situation exists in crystals. I shall use the word subunit to represent a diffracting unit which is identical to sufficient accuracy to another: whether it be a different molecule, a different peptide chain in the same molecule, or two parts of the same peptide chain. The accuracy of the identity will define for the crystallographer the resolution of data for which he can assume the relationship will hold.

If there is more than one subunit in the crystallographic asymmetric unit, the first problem is to find the operation which superimposes one on the other - to determine the six variables which specify the rotation and translation (Case 1). If, on the other hand, the identical subunits are in different crystals, the problem is to find the transformation of one coordinate system which converts the coordinates of the subunit in that crystal to the coordinates of the other subunit in the other crystal (Case 2).

In either case, the problem is intricately entwined with crystallographic symmetry. In Case 1, the rotations and translations which are determined will depend on an arbitrary choice of which particular subunits are superimposed, out of the whole crystal lattice. In Case 2, the same type of arbitrary choice is involved, and in addition the choice of the origin of coordinates is very often (not always) dictated by crystallographic conventions.

There is a lattice symmetry about the possible solutions to the problem. If translation vectors $\underline{x}$, $\underline{y}$, $\underline{z}$ give a solution in Case 1, then the addition of any number of lattice translations to $\underline{x}$, $\underline{y}$, $\underline{z}$ generates another possible solution. Similarly, if the angles defining the rotation operation are α, β, γ an arbitrary number of 2π can be added to any of them without affecting the solution. So there are an infinite number of equivalent operations

$$\alpha + 2\pi A$$
$$\beta + 2\pi B$$
$$\gamma + 2\pi C$$
$$\underline{x} + U\underline{a}$$
$$\underline{y} + V\underline{b}$$
$$\underline{z} + W\underline{c}$$

where A, B, C, U, V, W can be any integer and $\underline{a}$, $\underline{b}$, $\underline{c}$ are crystal lattice translations parallel to $\underline{x}$, $\underline{y}$, $\underline{z}$.

The possible solutions thus form a six-dimensional lattice, and the existence of crystallographic symmetry means that other solutions exist so that this six-dimensional lattice is in general non-primitive. I know nothing about the six-dimensional space groups, but Rossmann and I considered the three-dimensional lattices of the rotation operations in our original paper[1]. Tollin, Main and

Rossmann[2] extended this work, and D. Moss[3] has recently pointed out a further extension to it, which seems previously to have been overlooked.

2. EULERIAN ANGLES

In practice it is convenient to compute the rotation function in terms of Eulerian angles, now referred to as (α, β, γ). I describe here the angular conventions used in the program currently available through CCP4[4], which is based on the mathematical methods of Crowther[5].

We define an orthogonal co-ordinate system ($\underline{X}$, $\underline{Y}$, $\underline{Z}$) and arrange the crystal cell so that the axis of highest symmetry is along $\underline{Z}$. In all symmetries higher than monoclinic, the crystallographic $\underline{c}$ direction coincides with $\underline{Z}$, and the crystallographic $\underline{a}$ direction is along $\underline{X}$. In the monoclinic system, the axes are permuted so that the crystallographic $\underline{b}$ is along $\underline{Z}$, and the crystallographic $\underline{c}$ coincides with $\underline{X}$.

The Eulerian rotation operations are defined with respect to the orthogonal system ($\underline{X}$, $\underline{Y}$, $\underline{Z}$)(Fig. 1).

(i) A rotation α of the coordinate system about $\underline{Z}$;
(ii) A rotation β about the new direction of $\underline{Y}$;
(iii) A rotation γ about the new direction of $\underline{Z}$.

This system is computationally convenient: in particular it fits well with the symmetry properties of the system. However, it has some disadvantages, as the three rotations are often far from orthogonal.

1. If β is zero the rotations α and γ have similar effects: all rotations with the same value of $(\alpha+\gamma)$ are identical.

2. If β is π, the rotations α and γ have opposite effects: all rotations with the same value of $(\alpha-\gamma)$are identical.

3. All possible rotations are included in the range $\alpha = 0,2\pi$; $\beta = 0,\pi$; $\gamma = 0,2\pi$.

Another well-known system is spherical polar rotation (χ, ω, ϕ) (Fig. 2). Here $\underline{Z}$ is the polar axis and ω and ϕ define the longitude and latitude respectively of the rotation axis. The angle of rotation about this axis is χ. In this system, all possible rotations are included in the range $\chi = 0,2\pi$; $\omega = 0,\pi$; $\phi = 0,\pi$.

3. SEARCH TECHNIQUES

The problem as presented so far has been a purely geometric one. If we have a set of coordinates $\underline{x}_i^A$, $\underline{y}_i^A$, $\underline{z}_i^A$ and $\underline{x}_i^B$, $\underline{z}_i^B$, $\underline{z}_i^B$ for the two molecules,

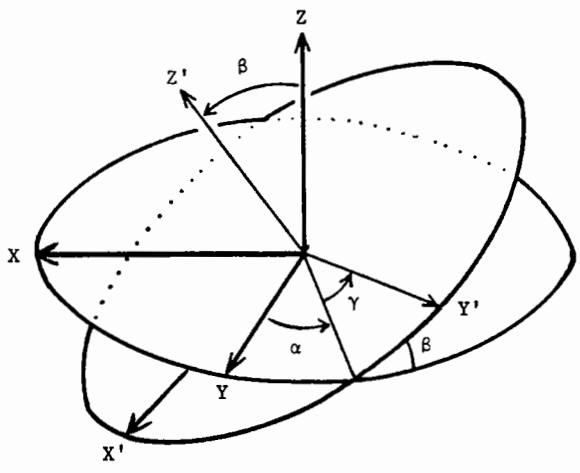


Fig. 1 Eulerian angles α,β,γ.

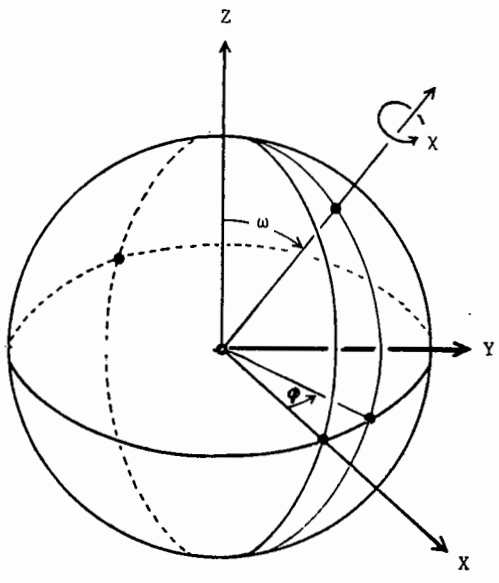


Fig. 2 Polar angles χ,ω,ϕ.

the solution is straightforward and can be solved as a standard eigenvalue problem (e.g. ref. 6).

The practical problem arises in experimental crystallography when the structure is unknown (or, in Case 2, one or both of the structures are unknown). The questions are then:

1. How can one detect that non-crystallographic symmetry exists?
2. How can one determine rotational and translational parameters which define the non-crystallographic symmetry operations?
3. Is this knowledge any help in solving structures?

The usual constraints on the practical problem are:

(a) The amplitudes of the structure factors are measurable by X-ray diffraction, but their phases are not accessible experimentally ("the phase problem").

(b) At the outset, no useful model is available for the structure or any part of it. (The usual situation in protein crystallography, but not in small molecule crystallography.)

In the absence of a direct solution to the practical problem, all methods will involve some kind of search. Searches in six variables easily become very large. It is hard to define how many values of each variable need to be tried in order to locate the solution, but in practice it cannot be less than 10 or so, implying a million search points, increasing to a billion if there are 30 search points in each variable. Such searches are clearly uneconomic, even with the latest computational equipment.

Small molecule crystallographers know well that the angular orientation of a molecule or ring system can often be determined long before the total structure is known. One of very few significant structures solved by a search procedure was the structure of hydroxyproline.[7] The molecule was treated as having a rigid five-memberedring with two dihedral angles as variables. So the search was eight dimensional, and it was carried out by a search procedure called a "method of non-local search". The search rapidly discovered the angular orientation of the five-membered ring, later the position of the molecule in the unit cell, and finally the proper value of the dihedral angles.

The key to the efficient solution of the problem is to find a way of separating the rotational and translational variables. If this is done, the search over millions of points can be reduced to two consecutive searches, each covering only a few thousand points.

The Patterson function[8] has certain characteristics which are well suited to this problem. Patterson pointed out that it contains <u>all</u> the information from the intensities (it has none from the phases): thus it includes all the experimental information, without any interpretive data. As a consequence of this, it is not dependent on any choice of origin: as we all know, it represents the collection of vectors between the elements of scattering density.

4. THE ROTATION FUNCTION

The basic idea of the rotation function is that the Patterson function of a molecule, or a subunit, will have a characteristic distribution of densities. In crystals, we are always dealing with arrays of units. If we think of the Patterson function as a collection of vectors, there will be some vectors between scattering density in the same subunit, and some vectors between one subunit and another. We can call these two types of vectors self vectors (S) and cross vectors (X). If we have two identical subunits in different crystals the array S for each will be the same, but the array X will be different. In case 1, with two identical subunits in the same crystal we will have two arrays S_A and S_B which are identical in shape, but in different orientations. A rotated coordinate system in Patterson space can be described, in which S_B is just like S_A in the original system. Multiplying the rotated Patterson function by the original will give a positive contribution from the product of S_A and its identical image S_B. If (following the usual habit) $F^2(000)$ has been omitted from the Patterson function, the other terms in the product will be negative as often as positive.

This positive tendency will exist all over the region containing S_A, so we can enhance the positive effect by looking simultaneously over the whole volume of vector space which it fills. This may be done by an integration which defines the rotation function. If we calculate

$$\mathcal{R}(R) = \int_A P(\underline{u}) \cdot P(\underline{u}_R) du$$

as a function of all pure rotation operations R, we expect the function to have significantly positive values when R represents the rotation of subunit B to subunit A. In case 1, there will also be a peak for the inverse rotation which rotates subunit A to subunit B.

4.1 WHAT ARE THE LIMITS OF SUMMATION?

It is important to point out that there are two types of summation (or integration) in the rotation function. The Patterson function itself is a summation over all reflexions, and the rotation function is a summation over a range of Patterson space. As we are in charge of the calculation, we can manipulate the limits of both these sums to our best advantage.

In crystallography there is always a resolution limit, beyond which it is impractical to extend our summation over reflexions. If our two subunits were precisely identical, it would be advantageous to extend our summation to the highest resolution available. In practice, subunits are not precisely identical and this would lead us to terminate the summation at lower resolution: or to be sophisticated, to weigh down high resolution terms. The resolution limit should be at least two or three times the expected mean co-ordinate difference between the two structures. In practice, with existing programs, computational limitations usually enforce a much restricted resolution.

It is also essential to omit very low resolution terms. The reason for this can be understood by considering what a low resolution structure actually looks like. Especially in crystals grown from high salt, the dominant features in a low resolution map are the solvent boundaries. Since the solvent regions are generally smaller in dimension than the molecules, the shapes which you see in maps at 10A - 8A resolution are the shapes of the solvent regions. Experience suggests that it is best to omit all terms with Bragg spacings greater than 7 or 8A.

Turning to the integration in vector space, it is obvious that no self-vector can exceed the largest diameter of the molecule. If the molecule is spherical, the density of its self-vectors forms an almost conical shape falling to zero at the diameter. If the molecule has a very elongated shape, the longest vectors will be equal to the length of the molecule, but there will be many more short vectors, especially shorter than the cross-sectional diameter.

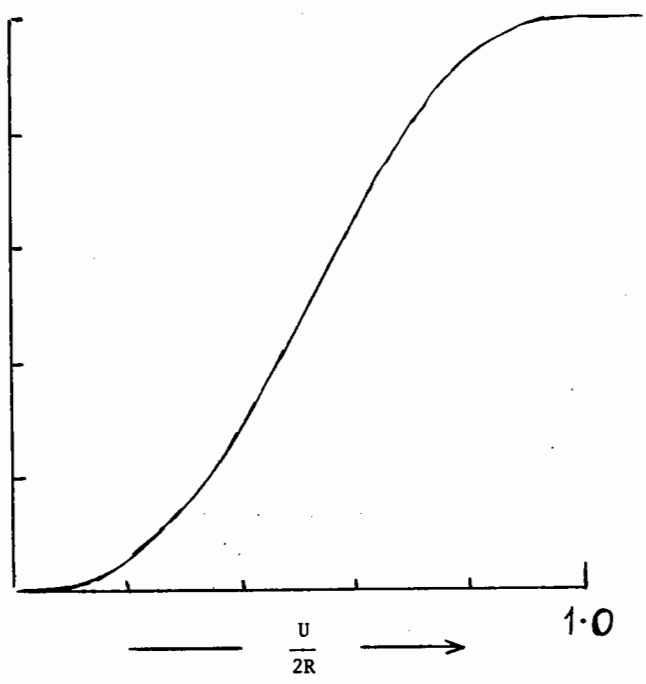


Fig.3 The Patterson function density for a uniform solid sphere of radius R, when integrated over a sphere of radius U about the origin, contains an integrated density given by

$$\frac{\pi^2 R^6}{36}\left(\frac{U}{2R}\right)^3\left[8-9\left(\frac{U}{2R}\right)+2\left(\frac{U}{2R}\right)^3\right]$$

for $0 \le U \le 2R$.

In our summation we obviously want to include regions of vector space where the self-vector density is high, and to omit those where it is zero, but where cross-vector density is high. Just where to make the cut-off is a matter which requires detailed study, but for an approximately spherical molecule I would suggest 75 - 80% of the diameter, which would include about 90% of the integrated Patterson density (Fig.3).

It is not necessary to integrate over a sphere, but simply a matter of convenience. In principle, it should be possible to obtain direct information about molecular shape by finding how the value of the peak in the rotation function varies as the volume of integration changes. In the early days,

I tried to discover the shape of the insulin hexamer (represented as a cylinder) by such a method. In those days the computations were too unwieldy and no useful results were obtained.

An important feature of the Patterson function is the origin peak, which is always the maximum value of the function. It expresses the fact that every bit of scattering density has an equal bit of scattering density exactly at a vector zero from it (namely, itself). Because the Patterson function is computed by a Fourier summation, the origin peak has a width characteristic of the resolution of the data, and it is normally spherical in shape. In the rotation function the origin peak is being compared to itself. The two spherical peaks simply add a constant positive term to the rotation function, and cause no particular trouble. They can be subtracted if desired.

There is one situation where origin peaks cause trouble - this is when there is a very flattened or elongated unit cell, and one of the lattice vectors is comparable to the molecular diameter. An integration in Patterson space over the average diameter of the molecule could possibly include an origin peak one lattice translation from the origin. It is essential to avoid including such an origin peak in the calculation. Either the radius of integration must be less than (the smallest lattice translation minus twice the resolution limit) or steps must be taken to subtract the origin peaks[9].

4.2 HOW TO DETECT NON-CRYSTALLOGRAPHIC SYMMETRY

When calculated, a rotation function usually looks very disappointing at first glance. There are a number of reasons for this. A self-rotation function (Case 1) has its own massive origin peak. Zero rotation obviously superimposes the whole Patterson function, and some other rotation which only superimposes the self-vectors will be miniscule by comparision. The Eulerian angles introduce an inconvenience. As already mentioned, if β is zero, the rotations α and γ have the same effect, so that all rotations with the same $(\alpha + \gamma)$ are identical. On the first page of your rotation function with $\beta = 0$, there is a massive ridge representing $(\alpha, 0, -\alpha)$, which are a whole set of equivalent zero rotations, and a series of other parallel ridges. Suggestions for replotting the rotation function in a different way to remove these problems have been made by Burdina[17] and Lattman[18].

After we have realised we must ignore the origin peak (in Case 1), it is still often necessary to do some statistics to find whether another peak is significant. I recommend calculating the mean and standard deviation of all the computed values after the origin peak has been removed: then find how many standard deviations from the mean the highest peak represents. In order to know how significant the peak is, one other fact is needed: that is, how many independent values the rotation function can have. Let me stress, this number only needs to be known very approximately. It is certainly no greater than the number of reflexions included in the Patterson function, to the appropriate resolution, so will often be a few thousand. In practice, if a peak exceeds 5 standard deviations it means something real; if the value is less than this it must be treated with scepticism.

If one is dealing with a symmetrical molecule, such as a dehydrogenase tetramer with 222 symmetry, there must be a series of peaks which bear a particular relationship to each other. Rossmann *et al.*[10] devised a "locked rotation function" which compares the values of the rotation function at different positions. This method can be used to enhance the significance level of a set of doubtful peaks.

4.3 FITTING A KNOWN STRUCTURE TO AN UNKNOWN ONE

An important use of the rotation function is to find the orientation of a molecule of known structure in a new crystal form. The simplest way to deal with this is to compute structure factors for the isolated molecule in a defined orientation[9]. The structure factors are those for a "model" crystal in which isolated molecules are spaced out on a lattice (usually on orthogonal lattice, but one in which the symmetry is P1). In the model crystal, one can arrange that the molecules are so far apart that none of the cross-vectors is as short as the chosen radius of integration. Then there are no cross-vectors in the model at this resolution. Avoid making the model crystal larger than necessary, as the number of structure factors to be computed increases rapidly.

5. TRANSLATION FUNCTIONS

When the rotational relationship of two subunits has been established, the other three variables to determine their relative positions can be found. A number of methods exist. When the two subunits are in the same crystal structure, and are related by a proper rotation (i.e. a rotation of $2\pi/n$) Rossmann et al.[11] discovered a relationship between the two sets of cross-vectors which produce peaks in a correlation function based on Patterson functions. This method is, however, difficult to use, and few examples of its use have been published. It remains the only available method for unknown structures.

When one of the structures is known (Case 2), a method due to Tollin[12] can be cast into the form of a Patterson-like convolution function[13].

Nowadays, the computer power available allows the use of a simpler method, which is to compute the crystallographic R factor as a function of the position of the subunit in the unit cell. This method was used for small molecules in 1964[14]. The molecular transform is computed for a single molecule, and values are added appropriately as the molecule and its symmetry-related copies range through the cell. When coupled with the FFT-based algorithm for structure factor calculation introduced by Agarwal[15], this method provides a computationally efficient solution to the translation problem for a molecule of known structure. It is available to the CCP through the program SEARCH, written by E. Dodson[16].

REFERENCES

1. M.G. Rossmann and D.M. Blow, Acta Cryst. 15 (1962) 24-31.

2. P. Tollin, P. Main and M.G. Rossmann, Acta Cryst. 20 (1966) 404-407.

3. D. Moss, Acta Cryst., in press.

4. CCP4 Program Suite Rotation Function-ALMN. User Documentation (Daresbury Laboratory, 1984).

5. R.A. Crowther in The Molecular Replacement Method (ed. M.G. Rossmann) (New York: Gordon & Breach, 1972) 173-178.

6. A.D. Maclachlan, Acta Cryst. A 32 (1972) 922-923.

7. R.L. Kayushina and B.K. Vainshtein, Kristallografiya 10 (1965) 833-844.

8. A.L. Patterson, Phys. Rev. 46 (1934) 372-376.

9. E.E. Lattman and W.E. Love, Acta Cryst. B 26 (1970) 1854-1857

10. M.G. Rossmann, G.C. Ford, H.C. Watson and L.J. Banaszak, J. Mol. Biol. 64 (1972) 237-249.

11. M.G. Rossmann, D.M. Blow, M.M. Harding and E. Coller, Acta Cryst. 17 (1964) 338-342.

12. P. Tollin, Acta Cryst. 21 (1966) 613-614.

13. R.A. Crowther and D.M. Blow, Acta Cryst. 23 (1967) 544-548.

14. A.K. Bhuiya and E. Stanley, Acta Cryst. 17 (1964) 746-748.

15. R.C. Agarwal, Acta Cryst. A34 (1978) 791.

16. CCP4 Program Suite. Search. User Documentation (Daresbury Laboratory, 1984).

17. V.I. Burdina, Kristallografiya 15 (1970) 623-630.

18. E.E. Lattman, Acta Cryst. B28 (1972) 1065-1068.

SYMMETRY AND THE ROTATION FUNCTION

by

D.S. Moss
Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College,
Malet Street, London WC1E 7HX.

1. ROTATION CONVENTIONS

The rotation function of Rossmann and Blow (1962) expresses the rotational correlation of two functions $P_0(x)$ and $P_1(x)$. Crowther's program (1972) parameterises the rotation function in terms of Eulerian angles α, β and γ. The rotation function may be written

$$F(\alpha,\beta,\gamma) = \int P_0(x) P_1(Rx)\, dV$$

where we may arbitrarily consider P_0 to be a stationary function and P_1 to be a function rotated by a matrix R. The volume of integration is a shell which may be chosen by the user.

There are two ways in which the two functions can be associated with the two relevant input channels of the program. In the Daresbury convention (Crowther and Dodson, 1981) the channel associated with the rotating function is such that P_1 is rotated by

γ about $\underline{z}$
β about $\underline{y}$
α about $\underline{z}$

in that order, clockwise looking away from the origin. In the other convention which is widely used, P_1 is rotated by

α about $\underline{z}$
β about $\underline{y}$
γ about $\underline{z}$

in that order, anticlockwise looking away from the origin. We shall follow the Darebury convention.

The Cartesian axes $\underline{x}$, $\underline{y}$ and $\underline{z}$ must not be confused with the crystal axes which may have different orientations.

It should be noted that the above conventions for the choice of rotation axes do not correspond to the original Rossmann and Blow (1962) convention ($\underline{z}$, $\underline{x}$, $\underline{y}$) used by CORELS (Sussman, Holbrook, Church & Kim 1977). The rotation of the co-ordinates can be accomplished by premultiplication of a column matrix of co-ordinates x by a rotation matrix R.

$$x' = Rx$$

In the Daresbury convention the matrix R is given in Table 1.

This operation may be necessary when wishing to compare rotations by expressing them in Eulerian angles. Assuming that R premultiplies the co-ordinates then in terms of Fortran functions the Eulerian angles are given by

$$\alpha = \text{atan2}(R_{23}, R_{13}) \qquad -\pi < \alpha \leq \pi$$
$$\beta = \text{acos}(R_{33}) \qquad 0 \leq \beta \leq \pi$$
$$\gamma = \text{atan2}(R_{32}, -R_{31}) \qquad -\pi < \gamma \leq \pi$$

2. DETERMINATION OF ROTATION AXIS AND ANGLE

Every rotation (α,β,γ) may be described as a rotation about a single axis by an angle χ. χ may be determined from the rotation matrix by

$$\cos\chi = \tfrac{1}{2}(R_{11}+R_{22}+R_{33}-1) \qquad 0 \leq \chi \leq \pi$$

The rotation axis may be expressed as a vector with direction cosines l_1, l_2 and l_3 where

$$l_1 = (R_{32}-R_{23})/2\sin\chi$$
$$l_2 = (R_{13}-R_{31})/2\sin\chi$$
$$l_3 = (R_{21}-R_{12})/2\sin\chi$$

For plotting stereograms it is useful to express the vector in terms of a spherical co-ordinates ω and ϕ where

$$\omega = \text{acos}(l_3) \qquad 0 \leq \omega \leq \pi$$
$$\phi = \text{atan2}(l_2/l_1) \qquad -\pi < \phi \leq \pi$$

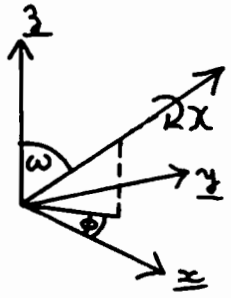


Fig. 1 Illustration of spherical co-ordinates

3. DEGENERATE SECTIONS

The sections of the rotation function at $\beta=0$ and $\beta=\pi$ are shown in figures 2 and 3 below. These sections are effectively one-dimensional and lines parallel to the diagonals illustrated are contours of constant density. The point $(\alpha,0,\gamma)$ represents a rotation of $\gamma+\alpha$ clockwise about z. The point (α,π,γ) represents a rotation of π about an axis in the xy plane at $\phi=\frac{1}{2}(\pi-\alpha+\gamma)$.

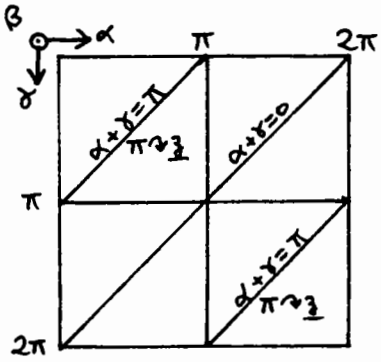


Fig. 2 Section at $\beta=0$

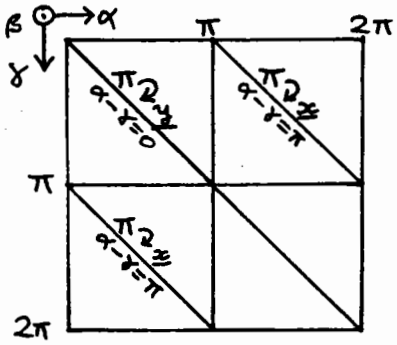


Fig. 3 Section at $\beta=\pi$

Sections close to these degenerate sections are highly distorted representations of rotation space.

4. DIADS IN THE SELF-ROTATION FUNCTION

Surfaces of constant χ in the Eulerian rotation function are given by the equation

$$\cos(\alpha+\gamma) = (2\cos\chi + 1 - \cos\beta)/(1+\cos\beta)$$

For diads ($\chi=\pi$) these surfaces are planes with equations given by

$$\beta = \pi, 3\pi, \ldots (2n+1)\pi \ldots$$

$$\alpha = \gamma = \pi, 3\pi, \ldots (2n+1)\pi \ldots$$

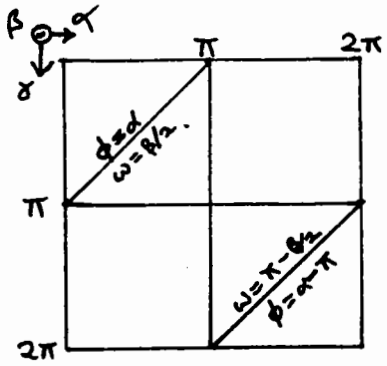


Fig. 4 General β section showing diagonal lines of intersection of planes corresponding to diad axes.

5. SYMMETRY OF THE ROTATION FUNCTION

In the cross-rotation of two Patterson functions P_0 and P_1 the group of operations which leaves the rotation function invariant is given by the direct product of the rotation groups (Q_0 and Q_1) of the functions.

$$D = Q_0 * Q_1$$

The elements of this direct product group D are pairs of symmetry operations, each operation being performed on its respective Patterson function.

In self-rotation, permutation of P_0 and P_1 leaves the rotation function invariant hence in this case the direct product group is

$$D = Q * Q * \pi_2$$

where Q is the relevant rotation group and π_2 is the permutation group of order two. Elements of D thus consist of three operations. It should be noted that the group D is independent of both the relative initial origin of the two functions and of the angular system used to express the rotations.

The parameterisation of proper rotations in terms of Eulerian angles gives rise a rotation function which is left invariant by space group symmetry operations performed on the co-ordinates (α,β,γ). This Eulerian space group G is related to D by a homomorphic mapping. The identity operation in D always maps into G, giving rise to general equivalent positions (α,β,γ) and $(\pi+\alpha,\beta,\pi+\gamma)$ which correspond to the same rotation.

The ability of the other elements of D to map into G depends on the orientation of the rotation axes in P_0 and P_1. In order that symmetry axes of P_0 and P_1 shall give rise to symmetry elements of G these axes must by parallel to z except in the case of diads which also give rise to symmetry when perpendicular to z. All axes of even order contain diads and give rise to the same symmetry when orientated perpendicular to z. It can be shown that the space group G depends only on the parity of the axes parallel or perpendicular to z (Moss, 1985). The order of the axes parallel to z will be denoted by p_0 and p_1 and those perpendicular to z by q_0 and q_1. With this notation the Eulerian rotation function space groups are shown in Tables 3 and 4.

6. SYMMETRY AND INTERPRETATION

We may summarise the relationship between the Patterson function rotation symmetry and the rotation function space group symmetry elements in the following statements where p_0 and p_1 are the orders of the axes parallel to z in the stationary and rotated functions respectively.

1) The dimensions of the primitive Eulerian cell are $a=2\pi/p_0 b=2\pi$ and $c=2\pi/p_1$.

2) A plane of symmetry always exists perpendicular to b and passes through the origin. The nature of this plane of symmetry depends on the parity of p_0 and p_1 in a way shown in Table 2.

3) Axes of even order perpendicular to z in the stationary or rotated functions give rise to b glide planes perpendicular to a or c respectively. Such a glide only passes through the origin if the corresponding Patterson axis is parallel to y.

4) Self rotation introduces a diagonal mirror plane into the rotation function which passes through the origin and is perpendicular to [101].

The practical implication of these statements is that a Patterson rotation axis produces the most helpful rotation space group symmetry when orientated parallel to z. In these circumstances it produces pure translational symmetry. When a Patterson symmetry axis is oriented perpendicular to z, space group symmetry only results when the axis is of even order and in this case if the order is greater than two then only the two-fold component produces space group symmetry. The b glide plane produced by this orientation is less convenient than the translational symmetry produced by orientation of the symmetry axis along z. It must be clearly understood that the x, y and z axes in this discussion are the Cartesian axes about which Eulerian rotation takes place and may not be parallel to the crystallographic axes having the same labels. This is especially true in the case of monoclinic Patterson functions where the unique crystallographic axis is usually chosen to be parallel to y.

The above discussion is also applicable when a Patterson function has axes of pseudosymmetry which may be either intermolecular or intramolecular. Such a pseudo-axis when oriented parallel to z will enable rotation function peaks corresponding to the solution and the pseudo-solution to be viewed on the same β section where they will have the same Eulerian distortion.

Considerations other than symmetry may sometimes be relevant when deciding how to orient a Patterson function with respect to Eulerian axes. Rotation function space is considerable distorted near the sections $\beta=0$ and $\beta=\pi$ where the angles α and γ are degenerate. The occurrence of peaks close to these sections may make interpretation inconvenient and in such cases reorienting a Patterson function may aid interpretation even if some Patterson symmetry no longer produces such useful effects.

When the rotation symmetry of a Patterson function belongs to a cyclic or dihedral point group then alignment parallel to z of the principal symmetry axis ensures that each axis gives rise to space group symmetry in the rotation function. However in the case of the cubic rotation groups 23 and 432, it is not possible to align the axes so that they are simultaneously all parallel or perpendicular to z. Hence not all the symmetry axes will produce space group effects in any given rotation function. For example, if a Patterson function of rotation group 432 is aligned with a 4-fold axis parallel to z then only the 422 sub-group will be effective in producing rotation function space group symmetry.

Each 3-fold axis will give rise to sets of three point within the asymmetric unit of the rotation function where the function is equal valued. This situation is analogous to the occurrence of non-crystallographic 3-fold axes in the crystal unit cell. In order that one of the 3-fold axes shall produce rotation function symmetry the Patterson function must be oriented with a 3-fold parallel to $\underline{z}$ in which case symmetry appropriate to point group 32 will result.

A similar problem exists when the Patterson function exhibits icosahedral symmetry 532 within the radius of integration. In this case the sub-groups 52 or 3 produce rotation function space group effects according to whether a 5-fold or 3-fold axis is parallel to $\underline{z}$.

Table 1

Rotation matrix which premultiplies a column matrix of co-ordinates.

$$R = \begin{bmatrix} \cos\alpha\cos\beta\cos\gamma-\sin\alpha\sin\gamma & -\cos\alpha\cos\beta\sin\gamma-\sin\alpha\cos\gamma & \cos\alpha\sin\beta \\ \sin\alpha\cos\beta\cos\gamma+\cos\alpha\sin\gamma & -\sin\alpha\cos\beta\sin\gamma+\cos\alpha\cos\gamma & \sin\alpha\sin\beta \\ -\sin\beta\cos\gamma & \sin\beta\sin\gamma & \cos\beta \end{bmatrix}$$

Table 2

Rotation function symmetry elements as a function of axial parity.

	$\underline{p}_0=2\underline{n}+1$	$\underline{p}_0=2\underline{n}$
$\underline{p}_1=2\underline{n}+1$	$\underline{n}$	$\underline{c}$
$\underline{p}_1=2\underline{n}$	$\underline{a}$	$\underline{m}$

$\underline{p}_0$ and p_1 are the orders of the rotation axes parallel to $\underline{z}$ in the stationary and rotated functions respectively. The Table indicates how the nature of the glide or mirror plane produced in the rotation function perpendicular to $\underline{b}$ depends on the parities of $\underline{p}_0$ and $\underline{p}_1$.

Table 3

Matrix of cross-rotation space groups as a function of axial parity

$\underline{P}_0$ is the stationary function and $\underline{P}_1$ is the rotated function. In the respective functions $\underline{p}_0$ and $\underline{p}_1$ are the orders of the axis parallel to $\underline{z}$ and $\underline{q}_0$ and $\underline{q}_1$ are the orders of the axes perpendicular to $\underline{z}$. Cell dimensions are $\underline{a} = 2\pi/\underline{p}_0$, $\underline{b} = 2\pi$, $\underline{c} = 2\pi/\underline{p}_1$. Numbers in parentheses are the space group numbers in International Tables for Crystallography (1983).

$\underline{P}_1$	$\underline{P}_0$	$\underline{q}_0=2\underline{n}+1$		$\underline{q}_0=2\underline{n}$	
		$\underline{p}_0=2\underline{n}+1$	$\underline{p}_0=2\underline{n}$	$\underline{p}_0=2\underline{n}+1$	$\underline{p}_0=2\underline{n}$
$\underline{p}_1=2\underline{n}+1$		$\underline{P}n$ (7)	$\underline{P}c$ (7)	$\underline{P}bn2_1$ (33)	$\underline{P}bc2_1$ (29)
$\underline{q}_1=2\underline{n}+1$	$\underline{p}_0=2\underline{n}$	$\underline{P}a$ (7)	$\underline{P}m$ (6)	$\underline{P}ba2$ (32)	$\underline{P}bm2$ (28)
$\underline{p}_1=2\underline{n}+1$		$\underline{P}2_1nb$ (33)	$\underline{P}2cb$ (32)	$\underline{P}bnb$ (56)	$\underline{P}bcb$ (54)
$\underline{q}_1=2\underline{n}$	$\underline{p}_1=2\underline{n}$	$\underline{P}2_1ab$ (29)	$\underline{P}2mb$ (28)	$\underline{P}bab$ (54)	$\underline{P}bmb$ (49)

Table 4

Matrix of self-rotation space groups as a function of axial parity

Rotation axes parallel and perpendicular to $\underline{z}$ have orders $\underline{P}$ and $\underline{q}$ respectively. The cell dimensions of the primitive cell are $\underline{a}=\underline{c}=2\pi/\underline{p}$, $\underline{b}=2\pi$. The centred cells for the two space groups not based on a primitive lattice have unit cell vectors $\underset{\sim}{a}'=\underset{\sim}{a}-\underset{\sim}{c}$, $\underset{\sim}{b}'=\underset{\sim}{b}$ and $\underset{\sim}{c}'=\underset{\sim}{a}+\underset{\sim}{c}$. Numbers in parentheses are the space group numbers in International Tables for Crystallography (1983).

	$\underline{p}=2\underline{n}+1$	$\underline{p}=2\underline{n}$
$\underline{q}=2\underline{n}+1$	$\underline{B}'\underline{ma}2$ (39)	$\underline{B}'\underline{mm}2$ (38)
$\underline{q}=2\underline{n}$	$\underline{P}4_2/\underline{nbm}$ (138)	$\underline{P}4_2/\underline{mbm}$ (132)

References

1. R.A. Crowther and E.J. Dodson, Science and Engineering Research Council Programme Suite for Protein Crystallography, Daresbury Laboratory, England, (1981).

2. International Tables for Crystallography, Vol A; (Dordrecht, Holland: Reidel Publishing Company, 1983).

3. D.S. Moss, Acta Cryst., (1985) in press.

4. M.G. Rossmann and D.M. Blow, Acta Cryst. 15, (1962) 24-31.

5. J.L. Sussman, S.R. Holbrook, G.M. Church and S. Kim, Acta Cryst. A33, (1977) 800-804.

ROTATION ANGLES

by

P. R. Evans
MRC Laboratory for Molecular Biology, Hills Road, Cambridge CB2 2QH

This paper considers the various ways of representing a rotation matrix by three angles, and in particular the relationship between Eulerian angles and polar angles for rotation functions. These alternatives, and the different conventions for the directions, names, signs and origins for these angles can be confusing, so it always best to consider the rotation matrix itself as the basic definition of the rotation, and to regard the angles as derived from the matrix. Even for the matrix, it is necessary to distinguish between the matrix which rotates the coordinates of an object, and its inverse (=transpose) which rotates the axes in the opposite sense.

Three types of angular representation of a rotation matrix may be distinguished. Within each type, various conventions of axes have been used.

(i) Eulerian angles where the first rotation axis is the same as the last, eg successive rotations of γ about z, β about the new y, and α about the new z. This is the usual convention for rotation functions. Note that when the second rotation β is zero, the first and last rotations are equivalent, so only $\alpha + \gamma$ can be determined. Similarly, when $\beta = 180°$, only $\alpha - \gamma$ can be determined.

$$R = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$$= \begin{bmatrix} \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\gamma & -\cos\alpha\cos\beta\sin\gamma - \sin\alpha\cos\gamma & \cos\alpha\sin\beta \\ \sin\alpha\cos\beta\cos\gamma + \cos\alpha\sin\gamma & -\sin\alpha\cos\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\sin\beta \\ -\sin\beta\cos\gamma & \sin\beta\sin\gamma & \cos\beta \end{bmatrix} \quad (2)$$

(ii) Eulerian angles where the first rotation axis is different from the last, eg succesive rotations of ϕ about z, θ about the new x, and ρ about the new y. Such conventions are often used for rigid body refinements, since the position of equivalence of the first and last rotations ϕ and ρ occurs when the second rotation $\theta = +-90°$, rather than when $\theta = 0$. In refinement, it is common to start with a zero rotation, ie an identity matrix, in which case all angles are zero, and the three parameters $\phi\theta\rho$ are orthogonal rotations away from this position. (iii) Polar angle conventions, eg a rotation of an angle χ about an axis whose direction is defined by the two angles ϕ and ω. These conventions are convenient for self-rotation functions, which can then be examined for n-fold rotation axes.

$$R = \begin{bmatrix} l^2+(m^2+n^2)\cos\chi & lm(1-\cos\chi)-n\sin\chi & nl(1-\cos\chi)+m\sin\chi \\ lm(1-\cos\chi)+n\sin\chi & m^2+(l^2+n^2)\cos\chi & mn(1-\cos\chi)-l\sin\chi \\ nl(1-\cos\chi)-m\sin\chi & mn(1-\cos\chi)+l\sin\chi & n^2+(l^2+m^2)\cos\chi \end{bmatrix} \quad (3)$$

where

$$\begin{bmatrix} l \\ m \\ n \end{bmatrix} = \begin{bmatrix} \sin\omega\cos\phi \\ \sin\omega\sin\phi \\ \cos\omega \end{bmatrix}$$

Self-rotation functions in polar angle space

The usual fast rotation function program of Crowther[1] calculates a map on an Eulerian angle grid. The position of any peak in Eulerian angles may be converted to polar angles, using relations derived from equating elements of matrices (2) and (3). However, it is often useful to look at whole sections of the self-rotation function at constant polar angle χ, for example to look for 222 symmetry. In general, there is no simple correspondence between the Eulerian grid and the polar angle grid, as can be seen by comparing the traces of matrices (2) and (3).

$$\text{Trace}(R) = \cos(\alpha + \gamma)[\,1 + \cos\beta\,] + \cos\beta$$

$$= 1 + 2\cos\chi$$

A dyad axis, $\chi = 180°$, is particularly simple: this can also be seen from the fact that for a twofold axis $R = R^{-1} = \tilde{R}$, ie R is symmetric, so equating matrix elements:

$$\cos\alpha = -\cos\gamma$$

$$\sin\alpha = \sin\gamma$$

$$\text{ie} \quad \alpha + \gamma = 180°$$

Thus it is easy to extract the $\chi = 180°$ section from the Eulerian angle rotation function, but other rotation angles are more difficult.

An alternative approach has been adopted by Tanaka [2], who has adapted the fast rotation function program to polar angles. Instead of rotating the second Patterson P_1 relative to the fixed first one P_0, he rotates the first by $(\alpha, \beta, 0)$ and the second by (α, β, γ) so that their relative rotation is γ about an axis defined by α $(=90° + \phi)$ and β $(= \omega)$.

$$F(\phi,\omega,\chi) = \int R(\alpha,\beta,0)\, P_0(\underline{x}) .\, R(\alpha,\beta,\gamma)\, P_1(\underline{x})\, dV$$

He shows how this modification can be put into the Crowther program. This modified version runs at least twice as slowly as the Euler angle version, and produces ω $(= \beta)$ sections: this map may then be resectioned to produce χ sections (Paula Fitzgerald, personal communication).

REFERENCES

1. R.A.Crowther in The Molecular Replacement Method, edited by M.G.Rossmann, pp. 173-183. (New York: Gordon & Breach, 1972)
2. Nobuo Tanaka, Acta Cryst. (1977) **A33**, 191-193

INVESTIGATIONS INTO THE LIMITATIONS OF A ROTATION AND A TRANSLATION FUNCTION

by

A.J. Schierbeek, R. Renetseder, B.W. Dijkstra and W.G.J. Hol
Laboratory of Chemical Physics, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands

1. INTRODUCTION

Between 1962 and 1967 the molecular replacement method has been developed at the MRC Laboratory in Cambridge, UK, which allows the elucidation of an unknown protein structure starting from the structure of a related molecule[1]. This approach has been used in the past with great success in many laboratories. With an increasing number of protein structures being available, the molecular replacement method will become of even greater importance in the future.

In our laboratory, the molecular replacement method has been used with succes in the structure determination of a number of phospholipases A_2[2,3]. As we envisage the use of the method for solving the structure of Triose phosphate isomerase from Trypanosoma brucei[4], starting from the known structure of the chicken enzyme[5], and of lipoamide dehydrogenase from A. vinelandii[6], starting from the known structure of human erythrocyte glutathione reductase[7], we looked into the limits of the molecular replacement methods for solving structures. As both T. brucei TIM and lipoamide dehydrogenase crystallize in space group $P2_12_12_1$, tests calculations were carried out using actinidin[8] as a model for solving the structure of the closely related enzyme papain[9,10]. All calculations were carried out using Crowther's fast rotation function[11] and Crowther and Blow's translation function[12].

2. FAST ROTATION FUNCTION STUDIES

A flow diagram of the fast rotation function[11], as programmed by Crowther is given in Figure 1. The program used can handle a maximum of 30 orders for the spherical Bessel functions. It is important to choose with care the triclinic cell used for calculating the model structure factors. It is advisable to use cell dimensions which are equal to the maximum dimensions of the model molecule in the a, b and c directions increased by the radius of the

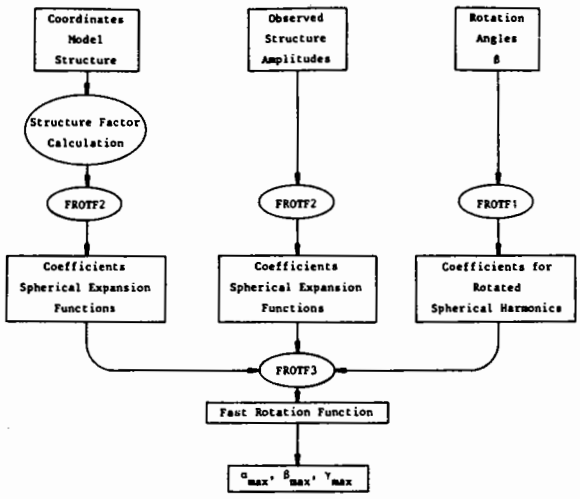


Fig. 1 Flow diagram of the fast rotation function calculations.

sphere in Patterson space to be employed in the rotation function calculations. In this manner no cross vectors between model molecules enter into the rotation function calculations.

2.1 Results with Phospholipases A_2

The structures of bovine native phospholipase A_2 and of transaminated bovine phospholipase A_2 have been determined at high resolution by Dijkstra et al.[13,14]. Starting with these models, the structures of bovine prophospholipase A_2 and of porcine native phospholipase A_2 could be determined by molecular replacement methods[2,3]. Some characteristics of the enzymes and crystals involved are listed in Table 1.

The fast rotation function gave only the correct solution in the case of the bovine prophospholipase structure when the appropriate resolution limits were chosen. As indicated in Figure 2, a narrow resolution range of 4.3 - 3.2 Å did not yield the correct answer while the range 6.0 - 3.2 Å worked perfectly well. This was a first indication that in the fast rotation function, the resolution range should not be taken too narrow. This point was also evident from rotation function studies on porcine

Table 1

PHOSPHOLIPASE A_2

(MW ~ 14.000)

	Bovine Native	Bovine Pro	Porcine Native
Space Group	$P2_12_12_1$	$P3_121$	$P3_121$
a (Å)	47.0	46.9	69.8
b (Å)	64.5	46.9	69.8
c (Å)	38.2	102.0	67.7
Medium	50% MPD	50% MPD	20% MeOH
Amino acid sequence identity	100%	100% + 7 extra at N-terminal	85%

phospholipase A_2 as summarized in Table 2.

2.2 Results with actinidin and papain

Highly refined structures of the sulphydryl proteases papain[9] and actinidin[8] are available. Some characteristics of these enzymes are given in Table 3. This pair of molecules forms a more stringent test of power of the molecular replacement method than the phospholipases as there is only ~ 50% amino acid sequence identity (Table 1). It should be pointed out, however, that papain and actinidin have a highly similar folding pattern with ~ 600 atoms having an r.m.s. difference of only ~ 0.4 Å[10].

As model structure was taken actinidin which was

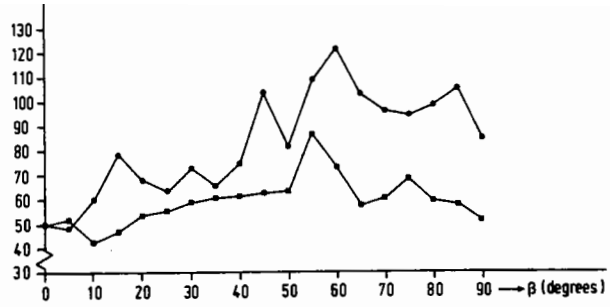


Fig. 2 Fast rotation function studies on bovine prophospholipase A_2 using transaminated phospholipase A_2 as model structure[2]. The highest values in each section with constant β are plotted in the vertical direction on an arbitrary scale based upon the highest value in the section β = 0° which is given a value of 50. Squares indicate results with data between 6 - 3.2 Å. In the latter case the maximum is at (α = 10°, β = 60°, γ = 85°) while the correct solution is (α = 25°, β = 55°, γ = 87.5°) which was obtained with the data between 6 - 3.2 Å (circles). The radius of integration was in both cases 19.2 Å. The model structure comprised 956 non-hydrogen atoms and was placed in a rectangular unit cell with dimensions 74 x 63 x 78 Å. An overall temperature factor of 15 $Å^2$ was used for calculating the model structure factors.

placed in a rectangular unit cell of 70 x 66 x 62 Å such that no intermolecular vectors shorter than 22 Å occurred. An overall temperature factor of 15 $Å^2$ was used. A number of initial tests varying the

Table 2

PORCINE PHOSPHOLIPASE A_2

Summary of results obtained with fast rotation function

	Resolution Range (Å)					
	3-5	3-6	3-7	3-8	3-9	3-10
No. of refls (F_{obs})	5042	5703	6041	6222	6349	6432
No. of refls (F_{calc})	4115	5141	5620	5907	6139	6326
Rotation function*	1.1	1.2	1.3	1.3	1.7	1.7

* ratio (peak height of correct solution) / (height of highest noise peak)

Table 3

PAPAIN and ACTINIDIN

(MW ~ 22.000)

	Papain	Actinidin
Space Group	$P2_12_12_1$	$P2_12_12_1$
a (Å)	45.2	78.2
b (Å)	104.6	81.9
c (Å)	50.9	33.0
Medium	67% MeOH	20% sat. Amm. Sulphate
pH	9.2	6.0
Amino acid sequence identity	100%	48%
No. of atoms	1654	1666

radius of integration from 6 to 17 Å showed that radii of 6 - 11 Å gave no useful results. The differences between the results for radii from 12 to 17 Å were only marginal. The resolution range was kept constant, 7 - 3 Å. The conclusion seems warranted that even for these protein molecules with dimensions of 45 x 44 x 36 $Å^3$ a radius of integration of > 11 Å appears to be sufficient.

The fast rotation function appeared to be insensitive to the omission of weak reflections, although it seems that leaving out the weakest reflections does not harm the results at all. A test for the sensitivity to changes in the resolution limits gave a similar result as in the case of the phospholipases: narrow resolution ranges of ~ 1 Å gave poor results, particularly at lower resolution. This is distinctly different from the self-rotation function in the case of haemocyanin[15,16] where resolution ranges of 6 - 5 Å gave excellent results. This may be related to the much larger size of the haemocyanin molecule and the larger radius of integration employed.

The effect of the completeness of the model structure on the rotation function results were also investigated. It appeared that removal of one quarter of the model molecule does usually not lead to a poor result, although omitting the last quarter of the model molecule, and incorporating all atoms, leads to a result where the highest peak is no longer the correct solution of the rotating problem. It also appeared that removing side chains beyond C^β does give a somewhat better result than incorporating all atoms. Quite surprisingly, using only the 218 C^α carbons gave also the correct solution, but leaving out one quarter of the C^α atoms is disastrous. It can therefore be concluded that the model can be surprisingly incomplete and still reveal the correct answer of the rotation problem.

3. TRANSLATION FUNCTION STUDIES

Crowther & Blow[12] derived a translation function for positioning a known molecule relative to a symmetry axis in an unknown structure. The translation function is calculated as $\underline{t}$, the distance between the centroid of two molecules related by the rotational symmetry operation under consideration. These authors also suggested that the self-vectors be removed from the observed Patterson leading to the following Fourier summation:

$$T(\underline{t}) = \sum_{\underline{h}} \{ |F_{OBS}(\underline{h})|^2 - \sum_{i=0}^{n-1} |F_M(\underline{h}[A_i])|^2 \} F_M(\underline{h}) F_M^*(\underline{h}[A]) \exp(-2\pi i \underline{h}.\underline{t}),$$

where $[A_i]$, i=0, n-1 are all crystallographic rotational symmetry operations, [A] the crystallographic symmetry operation for which $T(\underline{t})$ is calculated and F_M the structure factors of the model molecule. The simple flow diagram for the calculation of the translation function is shown in Figure 3. It is obviously important to have a proper scaling of the

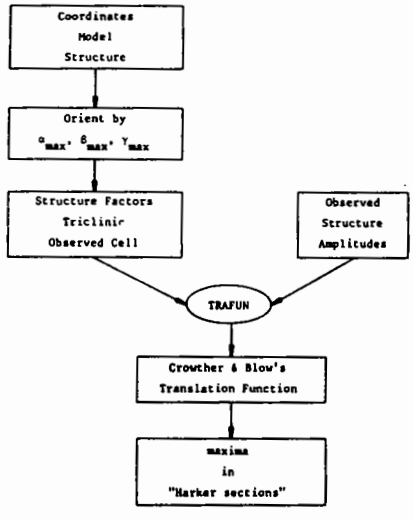


Fig. 3

F_M with respect to the F_{obs} when subtracting self-vectors. This can simply be done by equating

$\sum_{\underline{h}} |F_{obs}(\underline{h})|^2$ to $\sum_{\underline{h}} \sum_{i=0}^{n-1} |F_M(\underline{h}[A_i])|^2$ in case the

model structure contains roughly the same number of atoms as the structure to be solved. If the model structure contains only a fraction of the atoms of the unknown structure then the sum of the $|F_M|^2$ could be set equal to that fraction times the sum of the $|F_{obs}|^2$, assuming that Wilson statistics hold.

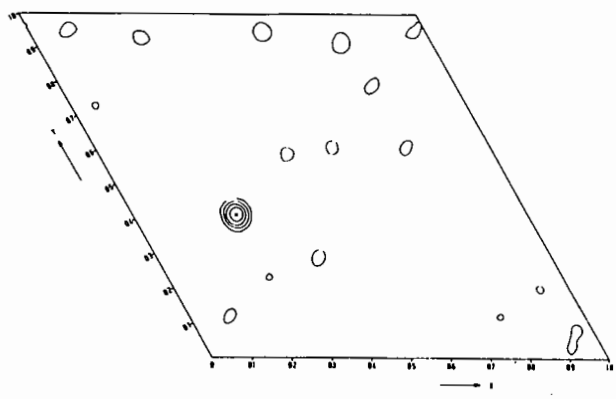


Fig. 4 Section with z_{tr} = 1/3 in the translation function, calculating cross vectors between model molecules related by a 3_1 axis parallel to the a-axis. The occurrence of a major peak in this section and not in the section z_{tr} = 2/3 shows that the space group of prophospholipase is $P3_121$ and not its enantiomorph $P3_221$. Contours are drawn at intervals of 20 arbitrary units, starting at level 20. The favourable peak-to-noise ratio is typical for the translation function.

3.1 Phospholipases A_2

As in many applications in other laboratories for other problems, Crowther & Blow's translation function gave excellent results in the case of bovine prophospholipase A_2[2] using the transaminated bovine phospholipase A_2 as an initial structure (Figure 4). In this case data between 3.2 and 5.0 Å were used whereas the orientational errors in α, β as well as γ were $\sim 2°$. The highest and correct peak had a value of 115 arbitrary units while the second highest peak had a value of 38.

In solving the structure of porcine phospholipase A_2 starting from bovine phospholipase A_2[3] similarly good results were obtained. As Table 4 shows, the results of the translation function did not critically depend on the resolution limits although the range 3 - 10 Å gave considerably worse results than the range 3 - 5 Å. This may be due to crystal packing effects as discussed by Crowther and Blow in their original paper[12]. Comparison of Tables 4 and 2 indicates that the rotation and translation functions seem to have opposite optimal resolution ranges.

3.2 Papain and actinidin

The sensitivity of the translation function for a number of parameters was investigated by using the actinidin structure as a model to determine the papain structure. In Table 5 is shown the very best

Table 4

PORCINE PHOSPHOLIPASE A_2

Summary of results obtained with Crowther & Blow's translation function

	Resolution Range (Å)					
	3-5	3-6	3-7	3-8	3-9	3-10
No. of refls (F_{obs})	5042	5703	6041	6222	6349	6432
No. of refls (F_{calc})	4115	5141	5620	5907	6139	6326
Translation function*	2.1	2.5	2.5	2.6	2.5	1.4

* ratio $\frac{\text{(peak height of correct solution)}}{\text{(height of highest noise peak)}}$

Table 5

ACTINIDIN and PAPAIN

Removal of self vectors in translation function

Model Structure	Self-vectors Removed?	TRF_{max} [1] x=0	TRF_{sol} [2] x=½	TRF_{max} [1] y=0	TRF_{sol} [2] y=½	TRF_{max} [1] z=0	TRF_{sol} [2] z=½
Papain	yes	21	99	24	99	24	99
Actinidin	yes	61	99	47	99	51	99
Actinidin	no	65	99	89	99	82	99

The translation function was calculated with resolution limits 3-7 Å and with F_{min} = 50. Actinidin was oriented according to the superposition of the coordinates.

1) TRF_{max} is the maximum translation function value in this section. As no peaks are expected in these sections, TRF_{max} can be considered as an indicator of the maximum noise level.

2) TRF_{sol} is the translation function value of the peak at the position of the solution of the translation problem. The maximum value in the two translation function sections (x=0 & x=½) or (y=0 & y=½) or (z=0 & z=½) is set to 99.

result one could expect, namely when using the complete and refined papain molecule as a starting model. It appears that refined actinidin as starting point gives a distinctly less good peak-to-noise ratios but the correct answer is nevertheless obtained without any problem whatsoever. Table 5 also shows that removal of self-vectors in the observed Pattersons gives much better peak-to-noise ratios in two of the three "Harker sections" of the translation function.

The weak reflections appeared to be important for an optimal translation function result. On the other hand, the translation function results were very insensitive to the choice of resolution limits. For instance the range 5 - 4 Å with 977 F_{obs} gave almost as good results as the range 10 - 3 Å with 4735 F_{obs} values. This is distinctly different from the results presented in Table 4 for the case of porcine phospholipase A_2.

Insight into the sensitivity of errors in orientation is quite important as it appears that the rotation function often gives peaks deviating somewhat from the exact solution. For the case of actinidin and papain, the correct solution obtained from a superposition of coordinates, has a value of β = 110° which means that α and γ are quite orthogonal to each other. So far, our calculations showed that errors of in one of the angles of up to ~ 5° can be tolerated.

Finally, the effects of the completeness of the model was investigated. With the actinidin model in exactly the correct orientation it appeared that using only C^{α} atoms, i.e. 13% of the total number of actinidin non-hydrogens, the highest peaks occurred at the correct positions in the sections x = ½, y = ½ as well as z = ½. Preliminary results do indicate, however, that with such incomplete model structures the translation function becomes extremely sensitive to orientational errors.

4. CONCLUSIONS

The investigations described above may be summarized as follows:

Fast rotation function

* Leaving out weak reflections does not harm
* Do not use too narrow a resolution range ($\Delta \sim 2$ Å)
* Do not use too small a radius (> 12 Å)
* Removing side chains from start model may help
* Start model can be surprisingly incomplete.

Crowther & Blow translation function

* Remove self-vectors, with appropriate weight
* Resolution range quite unimportant
* Do NOT remove weak reflections
* Orientational errors of ~ 5° allowed
* Start model can be quite incomplete.

Further studies are undoubtedly necessary to establish how generally valid these conclusions are.

ACKNOWLEDGEMENTS

We like to thank Dr. R.A. Crowther for the fast rotation function program and Dr. E. Lattman for an initial version of the translation function program. The actinidin coordinates were kindly provided by Dr. E.N. Baker. We are indebted to Prof. J. Drenth for his continuous encouragement and interest.

REFERENCES

1. M.G. Rossmann, The Molecular Replacement Method, Gordon and Breach, New York (1972).

2. B.W. Dijkstra, G.J.H. van Nes, K.H. Kalk, N.P. Brandenburg, W.G.J. Hol and J. Drenth, Acta Crystallogr. B38, (1982) 793.

3. B.W. Dijkstra, R. Renetseder, K.H. Kalk, W.G.J. Hol and J. Drenth, J. Mol. Biol. 168, (1983) 163.

4. R.K. Wierenga, W.G.J. Hol, O. Misset and F.R. Opperdoes, J. Mol. Biol. 178, (1984) 487.

5. D.W. Banner, A.C. Bloomer, G.A. Petsko, D.C. Phillips, C.I. Pogsen, I.A. Wilson, P.H. Corran, A.J. Furth, J.D. Milman, R.E. Offord, J.D. Priddle and S.G. Waley, Nature 255, (1975) 609.

6. A.J. Schierbeek, J.M. van der Laan, H. Groendijk, R.K. Wierenga and J. Drenth, J. Mol. Biol. 165, (1983) 563.

7. R. Thieme, E.F. Pai, R.H. Schirmer and G.E. Schulz, J. Mol. Biol. 152, (1981) 763.

8. E.N. Baker, J. Mol. Biol. 141, (1980) 441.

9. I.G. Kamphuis, K.H. Kalk, M.B.A. Swarte and J. Drenth, J. Mol. Biol. 179, (1984) 233.

10. I.G. Kamphuis, J. Drenth and E.N. Baker, J. Mol. Biol. 182, (1985) in press.

11. R.A. Crowther in "The Molecular Replacement Method", (M.G. Rossmann, editor) pp. 173-178. Gordon and Breach, New York, 1972.

12. R.A. Crowther and D.M. Blow, Acta Crystallogr. 23, (1967) 544.

13. B.W. Dijkstra, K.H. Kalk, W.G.J. Hol and J. Drenth, J. Mol. Biol. 147, (1981) 93.

14. B.W. Dijkstra, K.H. Kalk, J. Drenth, G.H. de Haas, M.R. Egmond and A.J. Slotboom, Biochemistry 23, (1984) 2759.

15. E.J.M. van Schaick, W.G. Schutter, W.P.J. Gaykema, A.M.H. Schepman and W.G.J. Hol, J. Mol. Biol. 158, (1982) 457.

16. W.G.J. Hol, A. Volbeda and W.P.J. Gaykema, see this issue.

Note added in proof: Dr. R.K. Wierenga informed us that the Trypanosoma brucei triose phosphate isomerase structure has been solved using procedures described in this paper, starting from the chicken triose phosphate isomerase structure. We like to thank the Oxford group for making the refined chicken TIM coordinates available.

REVIEW OF SPACE GROUP GENERAL TRANSLATION FUNCTIONS THAT MAKE USE OF KNOWN STRUCTURE INFORMATION AND CAN BE EXPANDED AS FOURIER SERIES

by

I.J. Tickle
Department of Crystallography, Birkbeck College, London WC1E 7HX.
for CCP4 meeting on "Molecular Replacement"

1. ABSTRACT

The purpose of this paper is to review the current status of the various techniques used to solve the translation problem in the molecular replacement method. I intend to deal only with the case where the trial structure is known, that is, the coordinates relative to an arbitrary origin, and optionally the thermal parameters, of a homologous structure are known, and the orientation has been accurately determined by means of the Rotation Function. The latter is of course being dealt with separately in this meeting. Then, some results for two structures will be discussed; the first, γ-crystallin IV, where the translation had previously been determined by R-factor search, and the second, chymosin, where the result was not known beforehand.

2. INTRODUCTION

In the papers of Tollin on the Q-function[1] and Crowther & Blow on the T-function[2], emphasis was placed on the separation of the general 3-dimensional translation function into several 2-dimensional functions. This was clearly important before the advent of the crystallographic Fast Fourier Transform[3]. Considering only space groups with only n-fold rotation and/or screw axes (face and body-centring symmetry can be ignored), and taking only the asymmetric units related by one n-fold axis at a time, the translation function varies only in a plane perpendicular to the axis. Thus one 2-dimensional function is obtained for each rotation/screw axis in the space group i.e. 1 for all polar point groups (2,3,4,6), 3 for orthorhombic (222), 4 for point group 32, 5 for 422, 7 for 622, 7 for 23 and 13 for 432.

The disadvantages of this approach are firstly that noise is introduced into the translation function by ignoring vector peaks in the Patterson of the unknown structure arising between asymmetric units other than those under consideration, and secondly in high symmetry apolar space groups one is still faced with the task of extracting the required 3-D vector from up to 13 number field sections, many of which, due to noise, may well not have the correct solution peak as the highest in the section.

The 3-dimensional functions proposed by Crowther & Blow[2] and by Harada, Lifchitz & Berthou[4] overcome these difficulties by computing a single translation function in which the solution peak will contain the combined information from all the 2-dimensional functions, thereby improving the signal-to-noise ratio. In fact Crowther & Blow incorrectly formulated their T_2 function, making an unnecessary approximation, and subsequently Bott & Sarma[5] reported failure of the T_2 function. It should be mentioned that even though a 3-dimensional function is computed, advantage can still be taken of the fact that the problem breaks down into 2-dimensional functions in producing the Fourier coefficients of the translation function; the final computation is then a 3-dimensional Fourier transform using FFT.

The earlier papers on the translation problem[1,6] indicated the importance of 'sharpening' (i.e. normalising) the structure amplitudes, but this seems not to have been taken up in more recent papers, with the exception of Harada et al.[4]. The advantages of this procedure are: i) the datasets are automatically scaled together by taking the data from the known and unknown structures separately and making $\overline{E^2}=1$ in shells of constant volume in reciprocal space; and ii) data is sampled evenly in reciprocal space. The 'sharpening' is actually to some extent undesirable since it leads to series termination ripples in the Pattersons, but this has not been found to be a serious problem. Incidentally the use of

normalised amplitudes also appears to be advantageous in the Rotation Function (I.J. Tickle, unpublished).

It should be pointed out that the current practice of applying a rather severe low resolution cut-off (e.g. 6Å) in rotation and translation functions is nothing more than a crude sharpening technique. A 20Å cut-off has been found to be necessary because $\overline{E^2}$ can change rather rapidly at this resolution and there is usually an insufficient number of reflections to give a statistically meaningful average. The omission of any data is always questionable, as it inevitably introduces random and systematic errors.

Another important factor often overlooked is the noise introduced by failing to eliminate from the Patterson, vector peaks between atoms within the same asymmetric unit (self-vectors). If one makes the approximation that the known and unknown structures have the same self-vector set, then it is a simple matter to substract its Fourier transform from the transform of the unknown Patterson, i.e. the intensities.

These apparent difficulties encountered with the scalar-product type of translation function (i.e. the Q and T functions) have led various authors to propose alternatives. These fall into two major classes: R-factor and overlap. The principal disadvantage of the R-factor function[7,8] is that it cannot be computed by a Fourier transform in order to take advantage of the FFT algorithm, and therefore unless one omits systematically (e.g. the small intensities) a large proportion of the data, or works at low resolution or computes on a coarse grid, or in fact all of these, the compute time can become prohibitive, typically 100 times slower than the T-function under the same conditions. In addition the conventional crystallographic R-factor is actually a less efficient statistic than the correlation coefficient, which is closely related to the scalar product. This means that the correlation coefficient should have a higher discriminatory power, provided it is used correctly.

R-factor (on F): $\dfrac{\sum |F_o - F_c|}{\sum F_o}$

Correlation coefficient (on E^2)

$$\frac{\sum n E_o^2 E_c^2 - n\overline{E_o^2}.\overline{E_c^2}}{[\sum (E_o^2 - \overline{E_o^2})^2 . \sum (E_c^2 - \overline{E_c^2})^2]^{\frac{1}{2}}}$$

(note $\overline{E_o^2}=1$ by definition; $\overline{E_c^2}$ depends on overlap.)

Scalar product (on E^2): $\sum E_o^2 E_c^2$

Overlap function: $\sum F_c^2$

Different forms of the overlap function have been proposed[4,5,9]; probably the easiest to compute is that of Harada et al.[4] who defined it simply as the origin peak of the Patterson of the trial structure; this will be smallest when no atomic densities overlap.

3. RESULTS: γ-crystallin IV

γ-crystallin IV crystallises in space group $C222_1$, and has 85% sequence homology with γ-crystallin II whose structure has been refined to 1.6Å.[10] The rotation function at 3Å resolution has been determined, structure factors for one molecule of γ-crystallin II in the unit cell of γ-crystallin IV computed, and the translation vector determined by an R-factor search program.[11] A program T2ORTH was written to test various T functions; the formulae used in the program are derived in the Appendix; this is essentially a synthesis of the work of Crowther & Blow[2], and of Harada et al.[4] Table 1 shows the results; the column labelled S/N is the signal-to-noise ratio of the correct peak relative to the highest incorrect one. The best S/N relative to the mode of the distribution of the function was 12.3.

4. CONCLUSIONS: γ-crystallin IV

i) Using F's as coefficients the S/N increases on omitting data with d>6Å, and further improves when scaled in shells. Omitting data with d<3.5Å reduces the S/N. Including the self-vectors greatly reduces the S/N. In contrast, using E's as coefficients and omitting data with d>6Å reduces the S/N; in fact omitting data in any way, with d<3.5Å or leaving out centric zones, has the same effect. The

conclusion is that the 'sharpening' effect produced with F's by omitting data with d>6Å can be achieved with E's by using all the available data (at least in the range 20-3Å). Line 12 in the table is the translation/overlap function of Harada et al.[4]; rather surprisingly the inclusion of self-vectors in this has a positive effect on the S/N. The last four lines show a fairer comparison between the translation and translation/overlap functions since this was developed for the case of the high-temperature form of lysozyme at 6Å resolution, and where the sequence homology is 100%. The conclusion is that the translation/overlap function may be advantageous if only low resolution data is available and the homology is high.

5. RESULTS and CONCLUSIONS: chymosin

Chymosin crystallises in space group I222 or $I2_12_12_1$ and has ~ 20% sequence homology with endothia pepsin whose structure has been refined to 2.1Å[12]. Computer graphics model building was used to modify the sequence, build in the side-chains and re-build the loops at the periphery.[13] The sequence homology is greatest in the central core of the enzyme near the active site but the loops have low homology. The isotropic thermal parameters of the endothia pepsin were retained as far as possible in the chymosin model, but new atoms were given twice the last known thermal parameter for the purposes of the structure factor calculation. A rotation function at 3Å resolution using normalised amplitudes and a 2.5° sampling interval confirmed a previous result.[14] Table 2 shows the results using the T2ORTH program, assuming space group I222. The best peak was 7.9σ above the mode. No significant peak was found for space group $I2_12_12_1$.

The conclusions are similar to those for γ-crystallin, except that the translation/overlap function is no longer the best, although there is some indication that it improves relatively when high resolution (3.5-3Å) data is omitted. The most consistent results, however, are obtained when the self vectors are subtracted.

Table 1. Gamma-crystallin IV (C2221) / Gamma-crystallin II

Function	Coefficient	Resolution	Centrics?	Scaling	Self-vectors?	S/N
1) Translation	F	20 - 3	Y	Overall	Subtracted	1.7
2) Translation	F	6 - 3	Y	Overall	Subtracted	3.2
3) Translation	F	6 - 3	Y	Shells	Subtracted	4.2
4) Translation	F	6 - 3.5	Y	Shells	Subtracted	3.6
5) Translation	F	6 - 3	Y	Shells	Included	1.1
6) Translation	E	20 - 3	Y	Shells	Subtracted	5.7
7) Translation	E	6 - 3	Y	Shells	Subtracted	4.5
8) Translation	E	20 - 3.5	Y	Shells	Subtracted	4.2
9) Translation	E	20 - 3	N	Shells	Subtracted	4.8
10) Translation	E	20 - 3	Y	Shells	Included	3.4
11) Transl./Overlap	E/F	20 - 3	Y	Shells	Subtracted	5.4
12) Transl./Overlap	E/F	20 - 3	Y	Shells	Included	5.9
13) Translation	E	20 - 6	Y	Shells	Subtracted	0.8
14) Translation	E	20 - 6	Y	Shells	Included	FAILS
15) Transl./Overlap	E/F	20 - 6	Y	Shells	Subtracted	1.3
16) Transl./Overlap	E/F	20 - 6	Y	Shells	Included	1.9

Table 2. Chymosin (I222) / Chymosin model.

Function	Coefficient	Resolution	Self-vectors?	S/N
1) Translation	F	6 - 3	Subtracted	1.5
2) Translation	E	20 - 3	Subtracted	2.1
3) Translation	E	20 - 3	Included	-1.4
4) Transl./overlap	E/F	20 - 3	Subtracted	2.2
5) Transl./overlap	E/F	20 - 3	Included	0.8
6) Translation	F	6 - 3.5	Subtracted	1.1
7) Translation	E	20 - 3.5	Subtracted	1.6
8) Translation	E	20 - 3.5	Included	-1.8
9) Transl./Overlap	E/F	20 - 3.5	Subtracted	1.6
10) Transl./Overlap	E/F	20 - 3.5	Included	1.3

6. APPENDIX

Crowther & Blow T_2 functions[2] (with self-vector subtraction):

$$T_2(\underline{t})=\int_V (P_o(\underline{u})-P_m(\underline{u})).P_c(\underline{u},\underline{t}).d\underline{u} \quad (1)$$

Harada et al. TO/O function[4]:

$$\frac{TO(\underline{t})}{O(\underline{t})} = \frac{\int_V P_o'(\underline{u}).P_c'(\underline{u},\underline{t}).d\underline{u}}{P_c(\underline{O},\underline{t})} \quad (2)$$

Form of translation function found best for γ-crystallin IV and chymosin:

$$Q(\underline{t})=\int_V (P_o'(\underline{u})-P'_m(\underline{u})).(P_c'(\underline{u},\underline{t})-P_m'(\underline{u})).d\underline{u} \quad (3)$$

where:

$$P_o(\underline{u})=\sum_h I_o(\underline{h}).\exp -2\pi i\underline{h}.\underline{u} \quad (4)$$

(Patterson of unknown structure, from observed intensities $I_o(\underline{h})$.)

$$P_m(\underline{u})=\sum_h I_m(\underline{h}).\exp -2\pi i\underline{h}.\underline{u} \quad (5)$$

(Sum of separate Pattersons of n asymmetric units of the known structure in their correct orientations, from the calculated intensities of the model, $I_m(\underline{h})$.)

$$P_c(\underline{u},\underline{t})=\sum_h I_c(\underline{h},\underline{t}).\exp -2\pi i\underline{h}.\underline{u} \quad (6)$$

(Patterson of complete unit cell of trial structure made up of n asymmetric units of known structure in their correct orientations, but in unknown positions determined by $\underline{t}$.)

$$P_c(\underline{O},\underline{t})=\sum_h I_c(\underline{h},\underline{t}) \quad (7)$$

(Origin peak height of $P_c(\underline{u},\underline{t})$.)

$P_o'(\underline{u})$, $P'_m(\underline{u})$ and $P'_c(\underline{u},\underline{t})$ are the corresponding sharpened Pattersons computed with normalised intensities $E^2_{\underline{h}}$ in place of $I_{\underline{h}}$.

$$I_o(\underline{h})=|F_o(\underline{h})|^2 \quad (8)$$

$$I_m(\underline{h})=\sum_k n|F_c(\underline{h},\underline{t})_k|^2 \quad (9)$$

$$I_c(\underline{h},\underline{t})=(\sum_k {}^nF_c(\underline{h},\underline{t})_k).(\sum_l {}^nF_c^*(\underline{h},\underline{t})_l)$$
$$=\sum_k {}^n\sum_l {}^nF_c(\underline{h},\underline{t})_k.F_c^*(\underline{h},\underline{t})_l \quad (10)$$

$I'_o(\underline{h})$, $I'_m(\underline{h})$ and $I'_c(\underline{h},\underline{t})$ are obtained by using E in place of F in (8), (9) and (10)

$$F_c(\underline{h},\underline{t})_k=\sum_j f_j \exp 2\pi i\underline{h}.\underline{s}_{jk}(\underline{t}) \quad (11)$$

This is the structure factor for the kth asymmetric unit, where f_j is the scattering factor for the jth atom in the model structure, and

$$\underline{s}_{jk}(\underline{t})=A_k.(\underline{r}_j+\underline{t})+\underline{d}_k \quad (12)$$

is the coordinate vector of the jth atom in the kth asymmetric unit of the trial structure, where $A_k,\underline{d}_k$ defines the kth general equivalent position of the space group and $\underline{r}_j$ is the coordinate vector of the jth atom of the model structure relative to its local origin before application of the translation $\underline{t}$. Note that this is not the same as Crowther & Blow's definition of $\underline{t}$. The corresponding structure factor is:

$$|F_m(\underline{h})|.\exp i\phi_m(\underline{h})=\sum_j f_j \exp 2\pi i\underline{h}.\underline{r}_j \quad (13)$$

Substituting (12) and (13) in (11)

$$F_c(\underline{h},\underline{t})_k=|F_m(\underline{h}A_k)|.\exp i(\phi_m(\underline{h}A_k)+2\pi\underline{h}.\underline{d}_k).\exp 2\pi i\underline{h}A_k\underline{t} \quad (14)$$

Therefore $|F_c(\underline{h},\underline{t})_k| = |F_m(\underline{h},\underline{t})_k|$ (15)

and from (9) $I_m(\underline{h})=\sum_k |F_m(\underline{h}A_k)|^2$ (16)

From (10) and (14):

$$I_c(\underline{h},\underline{t})=\sum_k {}^n\sum_l {}^n|F_m(\underline{h}A_k)|.|F_m(\underline{h}A_l)|.\exp i[\phi_m(\underline{h}A_k)-\phi_m(\underline{h}A_l)+2\pi\underline{h}.(\underline{d}_k-\underline{d}_l)].\exp 2\pi i\underline{h}.(A_k-A_l).\underline{t} \quad (17)$$

We now have everything required to expand (1), (2) or (3) in terms of known amplitudes and phases. For example, substituting (4), (5) and (6) in (3):

$$Q(t)=\sum_{\underline{h}}\sum_{\underline{h}'}(I'_o(\underline{h})-I'_m(\underline{h})).(I'_c(\underline{h}',\underline{t})-I'_m(\underline{h}')).$$
$$\int_v \exp(-2\pi i(\underline{h}+\underline{h}').\underline{u}).d\underline{u}$$
$$=\sum_{\underline{h}}(I'_o(\underline{h})-I'_m(\underline{h})).(I_c'(\underline{h},\underline{t})-I'_m(\underline{h})) \qquad (18)$$

Substituting (8), (16) and (17) in (18)

$$Q(\underline{t})=\sum_{\underline{h}}(|E_o(h)|^2-\sum_k{}^n|E_m(\underline{h}A_k)|^2).$$
$$\sum_k{}^n\sum_l{}^n E_m(\underline{h}A_k).F_m(\underline{h}A_l).$$
$$k\neq l$$
$$\exp i[\phi_m(\underline{h}A_k)-\phi_m(\underline{h}A_l)+2\pi\underline{h}.(\underline{d}_k-\underline{d}_l)].$$
$$\exp 2\pi i\underline{h}.(A_k-A_l).\underline{t} \qquad (19)$$

Crowther & Blow[2] stated incorrectly that this expression could not be cast as a single Fourier series because of the $\sum_k\sum_l$ summations. However Bricogne (unpublished) and more recently Harada et al.[4] showed that although it cannot be expressed as a Fourier series in $\underline{h}$, it can be so expressed in terms of the modified index $\underline{h}'=-\underline{h}.(A_k-A_l)$. In fact

$$Q(\underline{t})=\sum_{\underline{h}'}|F_{\underline{h}'}|\exp i\phi_{\underline{h}'}.\exp -2\pi i\underline{h}'.\underline{t} \qquad (20)$$

where $\sum_{\underline{h}'}=\sum_{\underline{h}}\sum_k\sum_l$
$$k\neq l$$
$$|F_{\underline{h}'}|=(|E_o(\underline{h})|^2-\sum_{k'}{}^n|E_m(\underline{h}A_{k'})|^2).$$
$$|E_m(\underline{h}A_k)|.|E_m(\underline{h}A_l)| \qquad (21)$$

and $\phi_{\underline{h}'}=\phi_m(\underline{h}A_k)-\phi_m(\underline{h}A_l)+2\pi\underline{h}.(\underline{d}_k-\underline{d}_l)$.
The other forms of the translation function can be expanded similarly. E.g. in (2):
for $P_c(\underline{0},\underline{t})$ the Fourier coefficient is:

$$|F_{\underline{h}'}|=|F_m(\underline{h}A_k)|.|F_m(\underline{h}A_l)|$$

7. REFERENCES

1. P. Tollin, Acta Cryst, 21, (1968) 613-614.
2. R.A. Crowther & D.M. Blow, Acta Cryst. 23, (1967) 544-548.
3. L.F. ten Eyck, Acta Cryst. A29, (1973) 183-191.
4. Y. Harada, A. Lifchitz & J. Berthou, Acta Cryst. A37, (1981) 398-406.
5. R. Bott & R. Sarma, J. Mol. Biol. 106, (1976) 1037-1046.
6. P. Tollin & W. Cochran, Acta Cryst. 17, (1964) 1322-1324.
7. P.E. Nixon & A.C.T. North, Acta. Cryst. A32, (1976) 320-325.
8. A.D. Rae, Acta Cryst. A33, (1977) 423-425.
9. W.A. Hendrickson & K.B. Ward, Acta Cryst. A32, (1976) 778-780.
10. L. Summers, G. Wistow, M. Narebor, D.S. Moss, P.F. Lindley, C. Slingsby, T.L. Blundell, H. Bartunik & K. Bartels in Peptide and Protein Reviews, Vol. 3 ed. M. Hearn; (New York & Basel: Marcel Dekker, 1984) 147.
11. H.P.C. Driessen & H. White, following paper.
12. L.H. Pearl & T.L. Blundell, FEBS Letters 174, (1984) 96-101.
13. T.L. Blundell & L. Sibanda, personal communication.
14. N. Andreeva & M. Saffro, personal communcation.

MOLECULAR REPLACEMENT AND THE CRYSTALLINS

by

H.P.C. Driessen* and H. White
Laboratory of Molecular Biology, Department of Crystallography,
Birkbeck College, Malet Street, London WC1E 7HX, UK.

1. INTRODUCTION

The three-dimensional structure of bovine lens γ-II crystallin has been solved to 1.6Å resolution in the Department of Crystallography of Birkbeck College, London[1-2]. This protein consists of two similar globular domains, each comprising two similar Greek key motifs. The two domains pack together with a single connection and are related by a pseudo 2-fold axis (Fig. 1). In each domain the two Greek key motifs form a pair of four-stranded antiparallel β-pleated sheets, each sheet composed of 3 strands from one motif and 1 from the other. The two motifs in their turn are related by a pseudo 2-fold axis. This structure therefore has a very high internal symmetry.

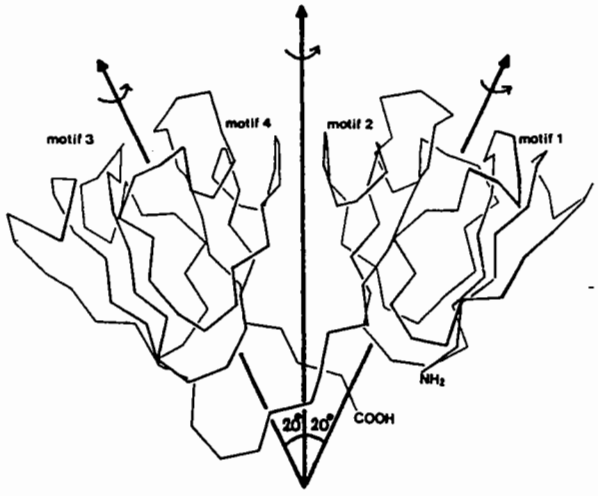


Fig.1 The internal symmetry of γ-II.

γ-II crystallin is a monomeric member of the β,γ-superfamily of crystallins, which comprises at least 13 different polypeptides having a protein sequence homology ranging from 30 to 98%[3-4]. All these proteins have the invariant and conservatively varied residues, necessary to form a similar structure to that of γ-II. Bovine γ-IV is a monomeric protein in solution and crystallizes with 1 molecule in the asymmetric unit in space group $C222_1$[2]. The homology with γ-II is 80%.It is therefore clear that γ-IV is a good candidate for molecular replacement studies, and as such one of the few examples for β-sheet proteins.

2. γ-II CROSS-ROTATED AGAINST ITSELF

2.1 The Orientation of the Search Molecule

In order to get a feeling for the problem an initial study was started with a rotated γ-II molecule against itself. Since the crystallins have a pseudo 2-fold axis, which for the γ-II backbone is present to an rms residual of 1.5Å with a screw component of 0.44Å, there are expected to be two peaks in the rotation map. One will have the correct orientation, while the pseudo-peak will represent the molecule rotated around the central pseudo 2-fold axis. Therefore this pseudo 2-fold axis was aligned parallel to z. This has two advantages: 1) the solution and pseudo-solution will be present on approximately the same β-section with a difference in γ of about 180° (the first rotation in Crowther's version of the rotation function[5] is γ about z showing symmetry in the search model), which makes the distinction between them easier, 2) the solution and pseudo-peaks will stand out among the spurious peaks, facilitating the interpretation.

2.2 The Choice of the Unit Cell

The size of the γ-II molecule is approximately 58 x 27 x 26Å. This molecule was placed in a cell as small as possible in order to keep computing costs low and contrast in maps high[6]. The geometrical mean of the ellipsoid semi-axes was chosen as the radius of integration and the cell dimensions were made thrice as large, although it was realised that this would cause the presence of some intermolecular vectors in rotation studies. The rotated molecule was put into an orthogonal cell of P1 symmetry with a=b=c=60Å, and structure factors were calculated using the atomic temperature factors of γ-II.

* Paper presented by H.P.C. Driessen

2.3 The Pseudo-Solution and the Solution for γ-II

As expected, the pseudo-solution and solution were immediately conspicuous on rotation maps. However, the solution-peak was about twice as high as the pseudo-peak over all resolution shells and radii of integration used, and the position of the pseudo-peak was subject to large shifts. An artificial temperature factor of 20Å² was used for both the model and the 'unknown' in all further work. For precise determinations a 60 Bessel function with stepsize 2.5° was used[5].

2.4 The Effect of the Resolution and Radius for γ-II

Although the pseudo-solution is not a problem in this 'ideal' case, the spurious peaks are, as can be inferred from a signal-to-noise plot (Fig. 2). For a Patterson radius of 15.0Å the map is extremely noisy with data of 4.0-5.0Å, although the spurious peaks are not significantly higher than the solution. The situation improves at 3.0-4.0Å, where the solution becomes the highest peak. The noisiness increases again at 2.1-3.0Å. At a radius of 15.0Å there appears no linear relationship between noise and resolution of the data used. The degree of noise possibly reflects errors both in the rotation function and in the intensity data.

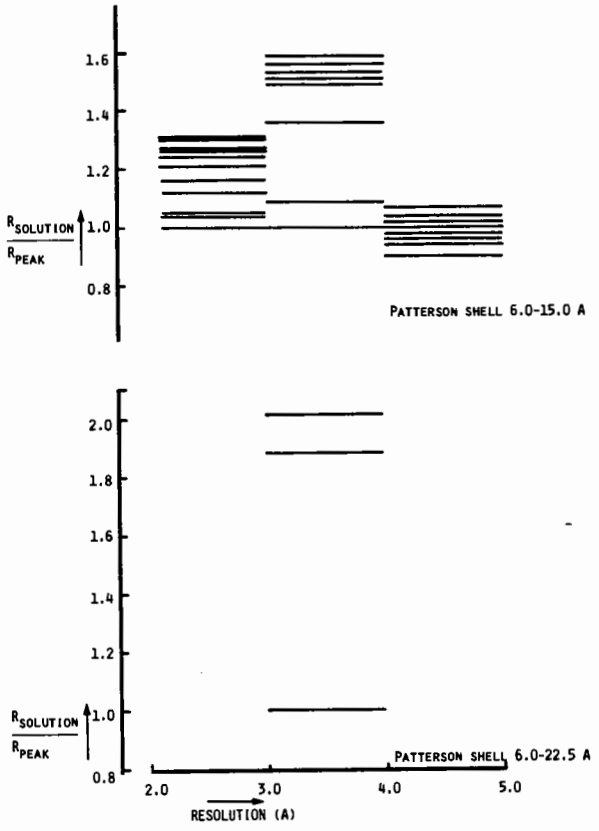


Fig.2 The effect of the resolution and the radius of integration on the signal-to-noise ratio for rotated γ-II against itself.

The effect of altering the radius of integration is quite dramatic. When this is increased to 22.5Å the signal-to-noise ratio at 3.0-4.0Å becomes 1.9 for the next highest peak. Clearly 22.5Å is a good value to find the best fit in cross-rotations for the crystallins.

2.5 The Accuracy of the Rotation Solution for γ-II

How accurate must the solution be? It is clear that any errors here will be carried into the next step, the determination of the translation parameters. The 60 Bessel version of the rotation function limits the stepsize in α and γ to 2.5°, while in β it is possible to go smaller than 0.5°[5]. More accuracy can be obtained by interpolation if the peak is very consistent in its position both as a function of the resolution and the radius of integration.

For γ-II the solution is as close as 1° in α, 2° in β and 0° in γ, when no interpolation is used. This result was used for a 0.7Å stepsize R-factor search (version of E. Dodson [7]). The correct solution could be clearly distinguished and was 0.88Å rms apart from the actual coordinates, yielding an R-factor of 39% at 5.0-10.0Å. To refine the six rigid body parameters the constrained-restrained least-squares program CORELS was used[8]. With three cycles of rigid body refinement the R-factor dropped to 32% and the structure finished in the correct position. This clearly indicates that even a good rotation result with acceptable translation parameters does need correction with rigid body refinement.

A more 'realistic' situation was chosen, where the three Eulerian angles in the rotation function were 2.5° in error. A smaller stepsize was used in the R-factor search leading to a solution with an R-factor of 47.5% at 5.0-10.0Å, which is 1.56Å rms out. 7 Cycles of CORELS at 5.0-10.0Å were required to reach the correct parameters.

What are the limits for rigid body refinement to be able to correct the molecular replacement errors? For γ-II against itself it appears that rotational errors have a worse effect. When the error was 2.3Å rms, or about 1½ bond distances, it took 10 cycles at 7.0-10.0Å to correct; at 5.0-10.0Å refinement was unsuccessful. For translational errors only, 2.5Å could be covered in 15 cycles at 5.0-10.0Å. The lesson for a real case appears to be that the rotation solution should be determined as accurately as possible, preferably with interpolation. After the molecular replacement the rms error should not be much larger than one bond distance.

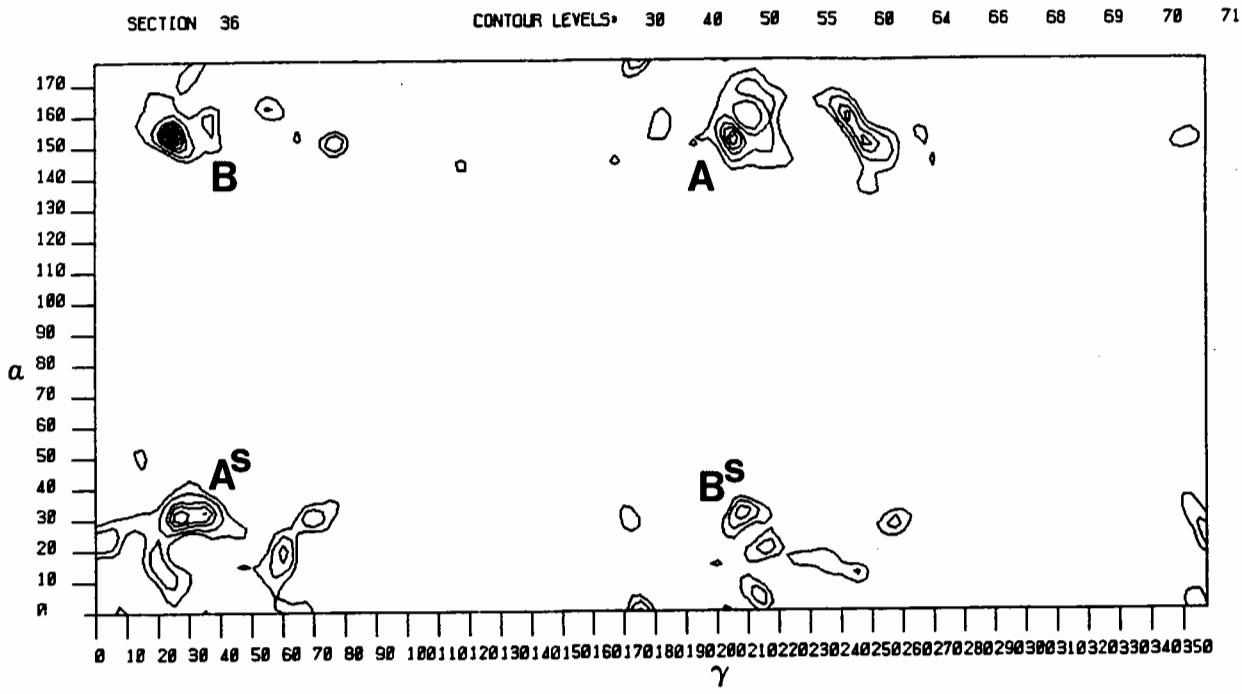


Fig.3 Cross-rotation map of Υ-II against Υ-IV: β=87.5°, resolution 3.0-4.0Å, Patterson shell 6.0-22.5Å. Peaks A^S and B^S are tails from the symmetry-equivalents of peaks A and B above β=90.0°.

3. Υ-II CROSS-ROTATED AGAINST Υ-IV

3.1 The Rotation Map of Υ-II against Υ-IV

For the rotation studies on Υ-IV the same rotated Υ-II molecule as in the Υ-II case study was used. One of the first features of the cross-rotation map to emerge was that there is no distinct difference between solution and pseudo-solution. On the β=87.5° section (Fig. 3) peak B, which is assumed to be the solution, and peak A are approximately 180° apart in Υ. Peak A is somewhat higher at β=85.0°. Since both peaks A and B are very close to β=90.0°, where a 2_1-screw axis is present, the β=87.5° section also shows peaks tailing from the symmetry-equivalents at β=92.5° and 95.0°. This gives rise to an apparent pseudo 2_1-screw axis at β=87.5°.

3.2 The Effect of the Resolution and Radius for Υ-IV

The signal-to-noise plot for Υ-II against Υ-IV (Fig. 4) at a radius of 22.5Å is distinctly worse than for the 'ideal' case discussed above. The noisiness of the maps is much higher and there are many peaks higher than the solution-peak in resolution shells 3.0-4.0Å and 4.0-5.0Å. They are, however, not significantly higher, and their position is not very consistent. At the highest resolution, 2.3-3.0Å, the solution B is the highest peak as expected. This resolution is needed to resolve solution B and pseudo-solution A. Clearly the difference between the two domains of Υ-IV is not as outstanding as in the 'ideal' case of Υ-II against itself. This may be due to the difference between the model Υ-II and the unknown Υ-IV or the lack of resolving power in the cross-rotation for proteins as distant as 80% in sequence homology.

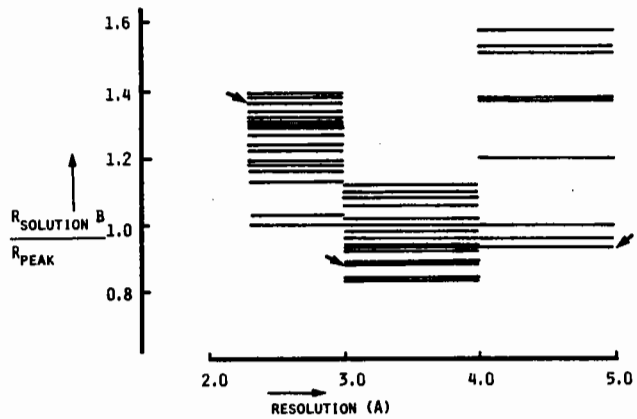


Fig.4 The effect of the resolution on the signal-to-noise ratio for Υ-II against Υ-IV. The Patterson shell is 6.0-22.5Å. The arrow denotes the pseudo-solution A.

3.3 The Cross-Rotation Solution for γ-IV

It is clear from sections 3.1 and 3.2 that the pseudo-solution A is not easily distinguishable from the solution B. On the other hand, there are indications for this interpretation. First, peak B is the highest at 2.3-3.0Å, where the differences between side chains should be most apparent. Second, peak B is very sharp, while A is very broad. Third, and linked to the last argument, for peak B the maximum difference in the Eulerian angles α, β and γ as a function of the resolution and radius is small, of the order of 1 to 2°. For peak A the order is 3 to 5°. This then lead to the choice of two different results for peak A, one for a radius of 14.0Å (A1) and one for 22.5Å (A2), while for B only one at 22.5Å was aken. The value of 22.5Å was used because of the results in the γ-II case study.

For peaks A1, A2 and B the position was optimised by using small β-steps and interpolation, although the value of the latter was arguable for the more variable A-peaks.

3.4 The R-Factor Search for γ-IV

In an R-factor search at 5.0-10.0Å with a stepsize of 0.5Å, it immediately became apparent that the A-peaks do not represent the correct rotation solution. R-factors for peak A2 (α=26.5, β=95.2°, γ=25.5°) were hardly above the background. Furthermore, they did not yield a structure with a reasonable packing in the unit cell with space group symmetry elements going straight through molecules. For peak A1 (α=32.0, β=92.0, γ=30.8°) two peaks with an acceptable packing were found, but their R-factors were rather large (53.8 and 54.8% at 5.0-10.0Å). These peaks disappeared in the background at 4.0-10Å. Nevertheless, they were subjected to rigid body refinement at 7.0-10.0 Å, but this gave no improvement in the statistics. At this resolution their correlation coefficients were of the order of 20% compared with the correct solution having a value of 54%. The A-peaks from the cross-rotation were not considered for further study.

The B-peak (α=155.0, β=86.8, γ=24.5°) on the other hand produced 12 peaks with acceptable molecular packing and with R-factors ranging from 45.7 to 52.8% at 5.0-10.0Å, much lower values than for the A-peaks. The translation peaks were more accurately determined with 0.1Å micro-grid searches.

3.5 The Determination of the R-Factor Search Solution

On the X-section, where the solution was present, a pseudo-translation solution was also visible (Fig. 5). This type of pseudo-peak was abundant, particularly at low resolution and with centric reflections only. It was therefore necessary to extend the initial low resolution search to high resolution and to use all the reflections. Since the R-factor search, even with microgrids around peaks only, is computationally expensive, the statistics at higher resolution for all peaks were determined with CORELS. At the same time a rigid body refinement at 7.0-10.0Å would give all peaks the best possible result, and would ensure that no peak was dismissed without optimal checking. R-factors and correlation coefficients were then calculated at 3.0-10.0Å (Table 1). The correlation coefficient turns out to be a particularly sensitive marker, indicated by a peak which is about twice as high as the next peak. This peak, with the lowest R-factor, shows a good stereochemical packing in the unit cell, even when side chains are taken into account. It has thus been accepted as the solution.

The shifts found in the rigid body refinement were quite

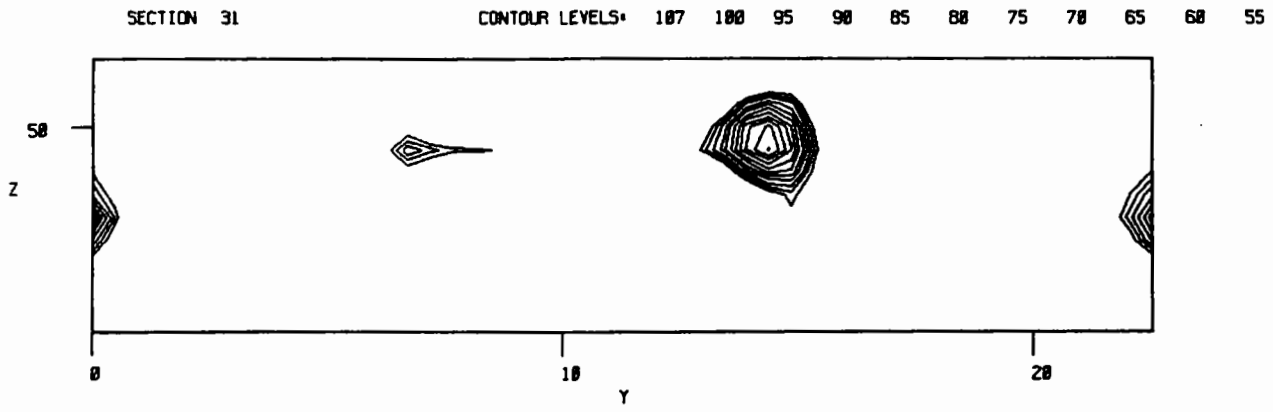


Fig.5 R-factor search map for γ-IV: x=13.6Å, y=0.0-22.5Å, z=45.5-51.5Å; resolution 7.0-10.0Å; stepsize 0.5Å. The map shown has been obtained for rotation solution B.

PEAK	R (%)	C (%)	RMS SHIFT (A)
SOLUTION	48.1	41.5	0.66
A	54.7	22.6	0.74
B	54.7	22.6	0.81
C	53.8	21.8	0.65
D	53.7	22.3	0.64
E	56.1	15.5	1.36
F	53.6	25.4	0.62
G	54.5	21.6	0.58
H	53.9	22.7	0.76
I	55.9	17.2	0.96
J	55.9	16.9	1.13
K	55.9	17.2	0.73

Table 1. R-factor and Correlation coefficient at 3.0-10.0A and rms shift at 7.0-10.0A after rigid body refinement of all possible R-factor search peaks for rotation solution B.

large, up to one bond distance, confirming that the radius of convergence of CORELS at 7.0-10.0A is large.

4. REFINEMENT OF γ-IV

4.1 Rigid Body Refinement of γ-IV

For the solution of the cross-rotation of γ-II against γ-IV the six rigid body parameters were now refined at 5.0-10.0A, because the actual position had been chosen at this resolution (Fig 6). The molecule was then refined at 3.0-10.0A. The rms movement with respect to the molecular replacement solution was a minimal 0.26A, indicating that the solution was quite accurate. Not surprisingly the rotational parameters showed the largest changes.

Subsequently the two domains of this molecule were refined as two independent rigid bodies without restraints. This was necessary, because in the connecting peptide between the two domains of γ-IV there is a deletion with respect to γ-II. The refinement was completely stable, and the R-factor dropped to about 47% at 2.3-10.0A. With respect to the molecular replacement solution the total rms movement was 0.88A, a considerable change. It appears that the domains have rotated around the central pseudo 2-fold axis giving a closer interdomain contact. The screw component has been reduced from 0.44 to 0.23A. This shows that the domains in γ-IV are related by a 2-fold axis which has less pseudo character than in γ-II. This could explain partly why the solution and the pseudo-solution were difficult to resolve in the rotation study.

Since the R-factor was as yet calculated for a model having an overall temperature factor of 20A², the refinement proceeded with two temperature factors per residue, one for the main chain and one for the side chain. The differences between γ-II and γ-IV now became apparent, and were mainly at the expected positions, i.e. in most of the side chains where a replacement has taken place, in the connecting peptide and in the C-terminal tail. The R-factor fell to below 40%.

4.2 Model Building and Restrained Refinement

The resulting map at 2.3A quite clearly showed the distinction between correctly and incorrectly positioned atoms. New density was visible, as were some new lattice contacts. The γ-IV residues were built in, and 9% of the atoms were omitted. The new model was subjected to restrained refinement at 2.3-5.0A using atomic temperature factors with the program RESTRAIN [9]. Currently the R-factor is 39% and the correlation coefficient nearly 70%. Further work is in progress.

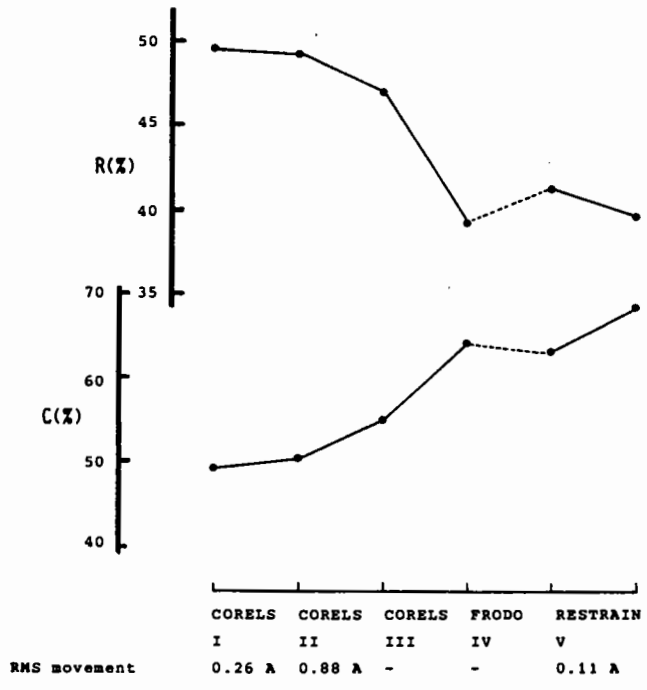


Fig 6. Refinement of the molecular replacement solution for γ-IV (statistics at 2.3-10.0A)

I rigid body:
5 cycles at 5.0-10.A F>3σ
5 cycles at 3.0-10.A F>3σ

II 2-domain rigid body:
6 cycles at 3.0-10.0A F>3σ

III 2 temperature factors per residue:
6 cycles at 2.3-10.0A F>3σ

IV model building of the γ-IV sequence; 9% of atoms omitted

V restrained refinement with atomic temperature factors:
6 cycles at 2.3-5.0A all F

5. CONCLUSIONS

The following conclusions can be drawn:

1) it was advantageous to do model studies with the search molecule
2) the rotation solution had to be as accurately determined as possible.
3) for this type of β-sheet protein high resolution data were a necessary requirement.
4) molecular replacement results could be improved by rigid body refinement with CORELS;
5) rigid body refinement of distinct structural domains gave a large improvement.

6. ACKNOWLEDGEMENTS

We thank Miss L. Summers for kindly providing us with the 1.6Å coordinates of γ-II. H.D. is the recipient of an EMBO Fellowship.

7. REFERENCES

1. G. Wistow, B. Turnell, L. Summers, C. Slingsby, D. Moss, L. Miller, P. Lindley and T. Blundell, J. Mol. Biol. 170, (1983) 175.

2. L. Summers, G. Wistow, M. Narebor, D. Moss, P. Lindley, C. Slingsby, T. Blundell, H. Bartunik and K. Bartels in Peptide and Protein Reviews, Vol. 3, ed. M. Hearn; (New York and Basel: Marcel Dekker, 1984), 147.

3. G.A.M. Berbers, W.A. Hoekman, H. Bloemendal, W.W. de Jong, T. Kleinschmidt and G. Braunitzer, Eur. J. Biochem, 139, (1984) 467.

4. J.G.G. Schoenmakers, J.T. den Dunnen, R.J.M. Moormann, R. Jongbloed, R.W. van Leen and N.H. Lubsen in Human Cataract Formation, CIBA Foundation Symposium 106, (1984) 208.

5. R.A. Crowther and E.J. Dodson, (1981), Science and Engineering Council Programme Suite for Protein Crystallography, Daresbury Laboratory, UK.

6. A. Lifchitz, Acta Cryst. A39, (1983) 130.

7. E.J. Dodson, (1982), Science and Engineering Council Programme Suite for Protein Crystallography, Daresbury Laboratory, UK.

8. J.L. Sussman, S.R. Holbrook, G.M. Church and S.-H. Kim, Acta Cryst. A33, (1977) 800.

9. I. Haneef, D.S. Moss, M.J. Stanford and N. Borkakotti, Acta Cryst., in press.

MOLECULAR REPLACEMENT: THE METHOD AND ITS PROBLEMS

by

E.J. DODSON

Department of Chemistry, University of York, Heslington, York YO1 5DD

Similar and homologous molecules will generally crystallise in different forms; molecular replacement techniques give us a handle for relating their structures within the crystals. The technique assumes that relative to some orthonormal axial system [I,J,K] we can generate the new coordinate set x1 from the known coordinate set x0.

To write this in full, the vector

$$[I\ J\ K] \begin{vmatrix} x_1 \\ y_1 \\ z_1 \end{vmatrix} = [I\ J\ K]\ [\Omega] \begin{vmatrix} x_0 \\ y_0 \\ z_0 \end{vmatrix} + \begin{vmatrix} tx \\ ty \\ tz \end{vmatrix} \qquad (1)$$

where [I J K] are the orthonormal axial system

$\begin{vmatrix} x \\ y \\ z \end{vmatrix}$ the coordinates of individual atoms

$[\Omega]$ a rotation matrix,

$\begin{vmatrix} tx \\ ty \\ tz \end{vmatrix}$ a translation vector.

In practice we assume the axial system for both sides of eq.(1) are the same,

and simply write $\underline{x}1 = [\Omega]\ \underline{x}0 + \underline{t}$.

But that assumption introduces the first possibility for confusion. We are dealing with crystals, which all too often do not have orthonormal axes, so we must be clear how we have defined the orthonormal axes relative to the crystal axes. The program ALMNFR, based on Tony Crowther's fast fourier summation for calculating the Rossmann/Blow rotation overlap allows three ways for defining these.

They are 1. I along a , K along c*.
2. I along b , K along a*
3. I along c , K along b*

The atomic coordinates $\underline{x}0$ relative to these axes are derived from the fractional $x0 = [Q]^{-1}\ xf$ and the elements of [Q] for each system are given in Appendix 1.

It reduces the amount of calculation to have the highest symmetry along K, so option 3 is used for monoclinic spacegroups, option 1 for all high symmetry spacegroups with 3,4 or 6 fold axes along c. For orthorhombic spacegroups the choice is arbitrary.

The second area where confusion is rife is in the definition of $[\Omega]$. Any matrix with real elements and determinant equal to unity represents a rotation in real space, and any rotation can be expressed in terms of three independent angles. There are many ways of doing this, but as long as the user knows how his/her matrix has been defined, the matrix elements should be the same for each and all methods. It is simply necessary to make sure all the programs used have the same convention built into them!

One useful form is given in Int. Tab. Vol.2 p.63.

If $[\Omega]$ describes a rotation of ψ about a vector with direction cosines $l_1\ l_2\ l_3$, then

$$[\Omega] = \begin{vmatrix} \cos\psi+l_1^2(1-\cos\psi) & l_1l_2(1-\cos\psi)-l_3\sin\psi & l_3l_1(1-\cos\psi)+l_2\sin\psi \\ l_1l_2(1-\cos\psi)+l_3\sin\psi & \cos\psi+l_2^2(1-\cos\psi) & l_2l_3(1-\cos\psi)-l_1\sin\psi \\ l_3l_1(1-\cos\psi)-l_2\sin\psi & l_3l_2(1-\cos\psi)+l_1\sin\psi & \cos\psi+l_3^2(1-\cos\psi) \end{vmatrix}$$

If the rotation axis makes an angle of ω with the K_0 axis, and its projection in the I_0J_0 plane makes an angle of ϕ to the I_0 axis, (measured as positive towards the J_0 axis), the direction cosines are

$$\text{Sin } \omega \text{ Cos } \phi \qquad \sin\omega \sin\phi \qquad \cos\omega$$

and the substitution of these values gives $[\Omega]$ in

its spherical polar form. It is often useful to use this form when searching for non-crystallographic symmetry within the asymmetric unit, where we expect to find a two-fold or a three-fold or even a seventeen-fold axis, and therefore know the expected value of ω.

Tony Crowther defines $[\Omega]$ using the system given below.

If the two sets of coordinates are defined relative to an orthonormal axial system I, J, K, the matrix $[\Omega]$ can be written as a function of three Eulerian angles, α, β, γ.

If α is a rotation about the initial direction of K_0, β is a rotation about the subsequent direction of J_1, and γ is a rotation about the final direction of K_2, then

$$[\Omega] = \begin{vmatrix} \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\gamma & -\cos\alpha\cos\beta\sin\gamma - \sin\alpha\cos\gamma & \cos\alpha\sin\beta \\ \sin\alpha\cos\beta\cos\gamma + \cos\alpha\sin\gamma & -\sin\alpha\cos\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\sin\beta \\ -\sin\beta\cos\gamma & \sin\beta\sin\gamma & \cos\beta \end{vmatrix}.$$

Appendix 2 shows why this rotation of axes is equivalent to the rotating of the coordinates relative to fixed axes by α, then β, then γ.

The actual multiplications by $[\Omega]$ are sometimes done in real space, sometimes in reciprocal space, and it is important to be clear whether $[\Omega]$ is being used to pre-multiply the column vector of coordinates, or post-multiply a row vector of indices.

THE ROTATION OVERLAP FUNCTION

So assuming we now know what we are doing, let us consider methods for finding the three angles which define the rotation matrix $[\Omega]$.

Rossmann and Blow (1962) defined a Patterson overlap function:

$$R[\Omega] = \int_V P(u')\, P(u)\, dv$$

where P(u') and P(u) are the values of the Patterson functions at the points u' and u where $u' = [\Omega]\, u$ and V is the volume over which matching is expected. When $[\Omega]$ is correct parts of the Patterson functions will overlap and $R[\Omega]$ will be large.

The idea is simple, but it would take a very long time to check the values of $R[\Omega]$ for many different $[\Omega]$ by direct integration.

Rossmann and Blow show that if P(u) is expanded in the usual way as a Fourier series:

$$P(u) = \sum_{hkl} F^2_{hkl} \exp(-2\pi i\, \underline{h}.\underline{u})$$

$R[\Omega]$ simplifies to

$$R[\Omega] = U/v^3 \sum_{\underline{h}} \sum_{\underline{p}} F^2_{\underline{h}}\, F^2_{\underline{p}}\, G_{\underline{h}.\underline{p}}$$

where $\underline{h}$ and $\underline{p}$ range over reciprocal space and $\underline{h} = [\underline{p}]\, [\Omega]$.

$G_{h.p}$ is a shape function which can be defined for simple volumes of integration. In fact $G_{h.p}$ is virtually zero unless $\underline{h}$ is close to $\underline{p}$.

The first program for calculating $R[\Omega]$ used a lot of computing resources.

Tony Crowther showed that by expanding the Patterson density within a spherical volume in terms of spherical harmonics it was possible to use fast fourier transforms to calculate values of the overlap function. I have collated some of his mathematical treatment of this problem in Appendix 3, since it has never been completely written up for publication.

Consideration of this overlap function suggests two further pitfalls which can cause difficulties.

1. The overlap search is done by matching spherical shells of Patterson space. This shell should ideally contain most of the intra-molecular vectors, but few of the inter-molecular ones resulting from crystal packing. This is possible if the molecule is roughly globular, and in a large volume of solvent, like haemoglobin. The choice of radius for the sphere of integration is determined by the dimensions of molecule, its shape and its packing in the crystal. If the molecule is markedly ellipsoidal as

is 6PGDH vectors will be omitted if the radius is limited to the molecules smallest dimension; on the other hand if the radius is set to the largest dimension, the spherical volume will be grotesquely larger than the molecular volume. (Paul Carr; see below).

When there is close crystal packing with small solvent volume and an irregular shaped molecule as with despentapeptide insulin, any change of radius can alter the appearance of the overlap map considerably. Figure 2 illustrates how changing the radius from 11 Å to 13 Å altered the map. Increasing the radius to 15 Å also increased the noise level.

Obviously if the sphere radius is greater than half a cell edge the same Patterson density will be included twice, and if the radius were set greater than a cell edge the origin peak would be included again.

The inner radius is used in a function to modify the F terms to remove the origin peak. It must be at least equal to the resolution limit.

This problem of non-spherical molecules is an insoluble one when using the Crowther-Blow method and I feel it is the main reason why some overlap maps look like the haemoglobin one (fig.1) and some like the DPI ones (fig.2). (Derewenda et al (1981), Bi Ru-Chang et al (1983)).

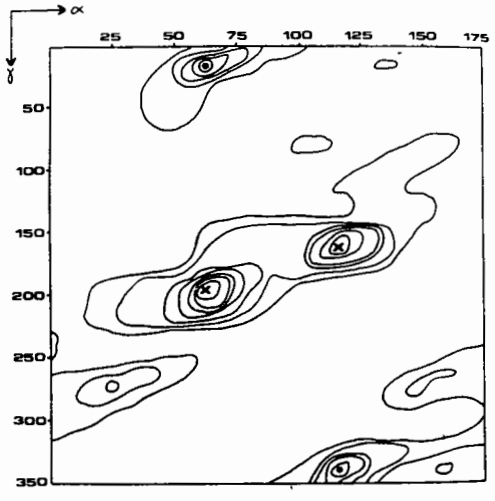


Fig.1. A section through a rotation function map of native haemoglobin, form II. 10-4 Å data has been used; a single αβ dimer was used as the search unit. The value of β on this section is 55°. × denotes the true maxima, • denotes the pseudosolutions.

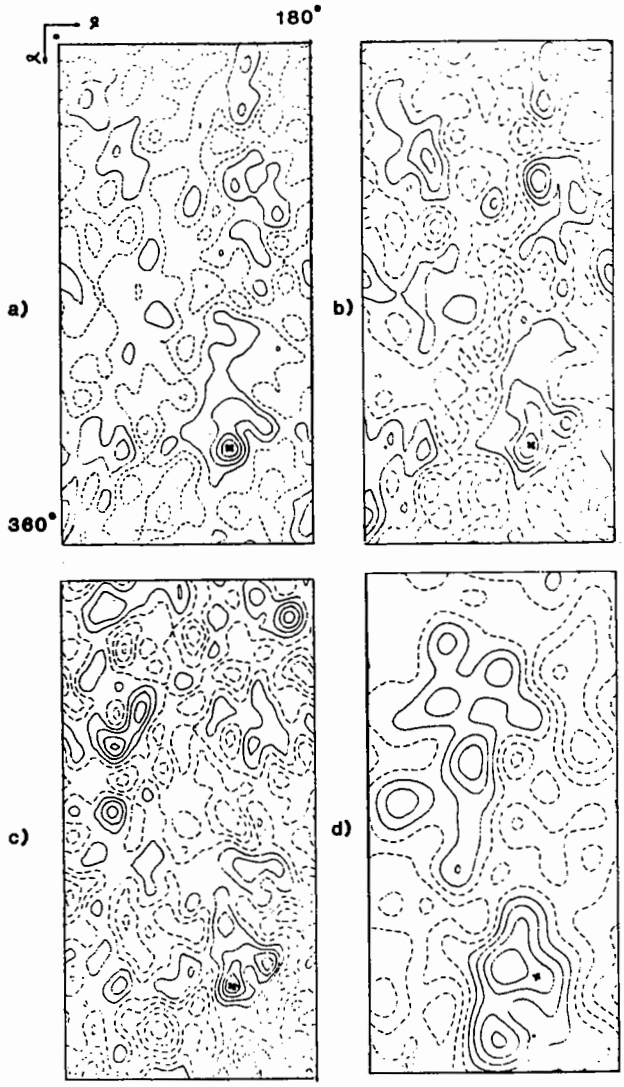


Fig.2. The β = 85° section for various beef DPI C2/model rotation functions. (a) The model contains 70% of the expected atoms; the sphere of integration extends to 13 Å; sharpening factor B = 20 Å; data range 8-3 Å. This map gave the clearest maxima. (b) as (a) but with the sphere of integration limited to 11 Å. Reducing the volume of integration has reduced the height of the true peak relative to spurious ones. (c) as (a) but with the model limited to main-chain atoms only. Again, spurious peaks are now the same height as the true one. (d) as (a) but with data limited to 12-5 Å. The maximum peak is now seriously misplaced.

2. The overlap function is a Patterson function using squared and sharpened values of F^2, and is therefore likely to be distorted by a few large terms. I find it is always necessary to sharpen the input intensities to give an even distribution of large F^2s against resolution; the program lists the number of terms in each $\sin\theta/\lambda$ shell after sharpening thus allowing you to check whether the chosen sharpening factor is sensible. Another problem arises with values of F calculated from a model

structure; obviously differences between the two structures will affect the degree of overlap. Some differences are inevitable; usually the known structure will have a different sequence or crystal form, so it may seem safer to omit many sidechains and external regions. However the DPI example (fig.2) shows that caution can be overdone - a better result was obtained when more atoms were included even though many were not correctly placed in the new form. Presumably their presence gave a more realistic distribution of F_{calc}.

The best way round these problems is to repeat the overlap search several times using different shells of intensities. Any true solution should be at least positive in all such searches, even if it is not the local maxima. There seems no way of predicting which shell will be the best for your problem; if there has been a good deal of movement between different protein domains the best result may come from low resolution matching; but if there are packing difficulties then the solution may only show up when high resolution data is used allowing matching of more precise features. The work reported today shows how unwise it is to use parameters simply because they work for someone else!

3. Crystal symmetry can make interpretation of results very complex. There are formulae for predicting symmetry equivalents of Eulerian angles, but it is essential to remember that the formulae depend both on the crystal symmetry AND the orthogonalising convention. I now use a version of the program which tabulates all equivlaent sets of Eulerian angles for any maxima, with the spherical polar equivalents and the direction cosines of the axis of rotation. Appendix 4 gives the derivation of the equivalents.

It is worth illustrating this with the results of the self-rotation search of $P2_1$ insulin. This form has a hexamer in the asymmetric unit, so we expect to find 4 maxima; one corresponding to a three-fold rotation, and three others corresponding to two-fold rotations, all perpendicular to the three-fold, and at 60° or 120° to each other. Here are extracts of the program output giving the symmetry equivalents of each maximum, first calculated in Eulerian angles, and then transformed to spherical polar angles.

Peak	α	β	γ	ω	φ	ψ	l_1	l_2	l_3	
20.1	170	20	160	85	20	180	.94	.34	.09	(1)
	170	20	-20	90	110	170				
	-170	20	-20	90	-70	170	(2-fold)			
	-170	20	160	95	20	180	.94	.34	-.09	
14.2	120	75	110	120	-108	178	-.26	-.83	-.50	
	120	75	-70	89	163	120				
	-120	75	-70	89	-17	120	(2-fold)			
	-120	75	110	120	72	178	.26	.83	-.50	(2)
14.5	65	150	100	142	-65	122	.26	-.56	-.79	(3)
	65	150	-80	48	-155	93				
	-65	150	-80	48	25	93	(3-fold)			
	-65	150	100	142	115	122				
19.0	110	145	35	125	-35	180	.67	-.47	-.57	
	110	145	-145	90	-125	110				
	-110	145	-145	90	55	110	(2-fold)			
	-110	145	35	55	-35	180	.67	-.47	.57	(4)

Check 1

2-fold axes must be at 60° or 120° to each other.

(1)*(2)

$l_{11}*l_{12} + l_{21}*l_{22} + l_{31}*l_{32} = .48 = \cos(60)$

(1)*(4)

$l_{11}*l_{14} + l_{21}*l_{24} + l_{31}*l_{34} = .52 = \cos(60)$

(2)*(4)

$l_{12}*l_{14} + l_{22}*l_{24} + l_{32}*l_{34} = .50 = \cos(60)$

Check 2

2-fold axes must be at 90° to 3-fold axis.

(1)*(3)

$l_{11}*l_{13} + l_{21}*l_{23} + l_{31}*l_{33} = .0022 = \cos(90)$

(2)*(3)

$l_{12}*l_{13} + l_{22}*l_{23} + l_{32}*l_{33} = -.0171 = \cos(90)$

(4)*(3)

$l_{14}*l_{13} + l_{24}*l_{23} + l_{34}*l_{33} = -.0129 = \cos(90)$

In the DPI study the difficulty was not that the complexity of the solution was disguised by the crystal symmetry, but that there were high noise levels and ambiguous peaks. We were however able to cross-check all the results from the four crystal forms by using the suitable symmetry equivalents of the maxima to find equivalent matrices matching A to B, B to C and A to C, etc.

4. There is another insoluble problem which can arise from crystal symmetry which cannot be avoided. If the non-crystallographic symmetry axis is approximately aligned with a crystal symmetry axis, any maxima of the search function will tend to smear into its symmetry equivalent, and it is often not possible to be sure where the true peak is. This hazard is obviously more likely to arise with higher symmetry space groups.

POSITIONING THE MOLECULE IN ITS CELL

In most of the work done at York we have found the translation parameters by searching for a minimum R value between the observed intensities and the different sets of calculated amplitudes generated as the molecule was moved through the unit cell. This method was used by Cutfield et al (1974) and described fully by Nixon and North (1976). The program used (called SEARCH) uses the fact that if the partial F_c's, F_{c1}, F_{c2}, ..., are calculated for all the symmetrically equivalent model molecules, orientated according to the appropriate rotation matrix, then the value of F_c for any set of translations, t_1, t_2,..., where the ti are symmetrically equivalent, will be given by

$$F_c(hkl) = F_{c1}(hkl)\exp(-2i\pi\underline{h}.\underline{t}_1) + F_{c2}(hkl)\exp(-2i\pi\underline{h}.\underline{t}_2) + \ldots$$

So, once the F_ci are calculated it is only necessary to sum them together with appropriate phase modifications to generate the different sets of F_c's. It is possible to use the inverse fast Fourier transform for space group P1 using the cell dimensions of the observed intensity set to calculate all the required F_ci relatively inexpensively. The scale factor between F_o and F_c varies by as much as 20% depending on the position of the molecule in the unit cell, and the amount of overlap in the symmetry-related positions. The scale factor was determined by comparing $\langle F_o \rangle$ and $\langle F_c \rangle$, with some allowance for the incompleteness of the model. Earlier calculations had shown that the positions of correct minima were not very sensitive to the changes in the scale.

In Beef DPI the space group was C2, where the position along the y axis is arbitrary, and the search was relatively cheap to run. For haemoglobin space group P21212 it was necessary to search one quarter of the unit cell, which took a good deal of computing resources. However the answer was extremely sharp. It would be interesting to test the translation function on this problem, and see if there is any gain in accuracy when using the R factor search.

In any space group with only two symmetry operations it is important to exclude reflections where either F_{c1} or F_{c2} is weak from the R-factor calculation. For any such reflection $|F_c|$ is almost independent of the values of t_1 and t_2. In general, such space groups have an R value for a correctly oriented molecule in any position in the unit cell considerably below that of a randomly wrong structure.

In our experience the characteristics of the translation parameter search gives some verification to the quality of the rotational parameters. If these are inaccurate, or if there is a poor match between the model and the new structure, it is impossible to find consistent clear solutions for the translation parameters.

VERIFYING THE RESULTS

We use two procedures to check the quality of our fitting. The first is to run a distance/angles calculation to see that the suggested fit is not causing different symmetry models to collide. The program, PRJANG, does this very quickly. (It has been modified to step through the unit cell with a reduced set of atoms, usually the CAs, to chart possible and impossible locations within the cell; a very poor poor-man's graphics substitute.)

The best check is to generate phases from the model in the new cell and see if these phases are any use. If there is heavy atom derivative data, then these phases should show up the heavy atom sites. In DPI they were sufficiently accurate to show sulphur atoms which had been excluded from the phasing (fig.3).

The final test of the model is: is it good enough to allow us to refine it to an accurate structure. Obviously there will be some gross errors where the model structure and the new form are different.

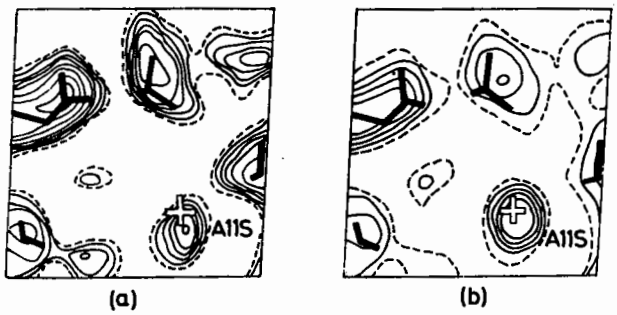


Fig.3. Part of the F_{obs} Fourier map showing the electron density at the sulphur atom (A 11) excluded from the phasing. Map (a) is unweighted and map (b) was calculated using modified Sim weights (Bricogne, 1976).

Nevertheless in our experience if the refinement of some parts of the new form can proceed smoothly and generate phases sufficiently accurate to allow rebuilding of the different fragments, then the search for isomorphous derivatives is bypassed, and a useful accurate crystal structure can quickly be obtained. We have found initial R factors of about 60% for 2A data fall to 30%-35% automatically when the initial rotation and translation parameters have been substantially correct. At that point rebuilding and extension of the model has been fairly straightforward. A word of warning; incorrect structures have "refined" to about 40%, then stuck. Refinement procedures will always lower the R factor whether the model is nonsense or not, and other biochemical criteria are necessary in deciding on the correctness of any solution.

REFERENCES

1. Bi Ru-Chang, S.M. Cutfield, E.J. Dodson, G.G. Dodson, F. Giordano, C.D. Reynolds and S.P. Tolley, Acta Cryst. B39, (1983) 90.

2. R.A. Crowther, in: The Molecular Replacement Method. A collection of Papers on the Use of Non-crystallographic Symmetry, edited by M.G. Rossmann, (New York: Gordon & Breach, 1972) pp.173.

3. Z.S. Derewenda, E.J. Dodson, G.G. Dodson and A.M. Brzozowski, Acta. Cryst. A37, (1981) 407.

4. P.E. Nixon and A.C.T. North, Acta Cryst. A32, (1976) 320.

5. M.G. Rossmann and D.M. Blow, Acta Cryst. 15, (1962) 24.

APPENDIX 1

The Orthonormalisation Matrix

These simple expressions for rotation matrices are only valid when the co-ordinates are given relative to an orthonormal set of axes. It is necessary to define a matrix to convert the crystal lattice ($\underline{a}$,$\underline{b}$ $\underline{c}$), to the chosen orthonormal frame (I_o,J_o,K_o). Different crystal space groups are more conveniently converted in different ways so there are three options available in the rotation and matching programmes.

The real space co-ordinates in the orthonormal frame are in Ångströms, and the reciprocal space units are $Å^{-1}$.

Option 1 I_o is parallel to $\underline{a}$, K_o is parallel to $\underline{c}^*$
Option 2 I_o is parallel to $\underline{b}$, K_o is parallel to $\underline{a}^*$
Option 3 I_o is parallel to $\underline{c}$, K_o is parallel to $\underline{b}^*$

Let us summarise the mathematics. The co-ordinates of the unique molecule with respect to the crystal axes are

$$\begin{vmatrix} \xi_i \\ \eta_i \\ \zeta_i \end{vmatrix} \quad i = 1, \ldots \eta$$ where η is the number of atoms in the molecule

From knowledge of the crystal axial system we can define a (3 × 3) orthonormalisation matrix [Q] to convert the fractional co-ordinates and the reciprocal indices (h k l) to an orthonormal form.

Define
$$\begin{vmatrix} \xi \\ \eta \\ \zeta \end{vmatrix} = [Q] \begin{vmatrix} x \\ y \\ z \end{vmatrix}$$
where ξ is defined relative to the crystal axes, and x is defined relative to the orthonormal axial system. ξ units range from 0 to 1; x is measured in As.

Then
$$\begin{vmatrix} x \\ y \\ z \end{vmatrix} = [Q]^{-1} \begin{vmatrix} \xi \\ \eta \\ \zeta \end{vmatrix}$$

Also, defining orthonormal indices [H K L], with respect to the same axial system as the orthogonalised coordinators such that

$$[h\ k\ l] \begin{vmatrix} \xi \\ \eta \\ \zeta \end{vmatrix} = [H\ K\ L] \begin{vmatrix} x \\ y \\ z \end{vmatrix}$$

then $[H\ K\ L] = [Q][h\ k\ l]$.

For option 1

$$[Q]^{-1} = \begin{vmatrix} a & b\cos\gamma & c\cos\beta \\ 0 & b\sin\gamma & -c\sin\beta\cos\alpha^* \\ 0 & 0 & c\sin\beta\sin\alpha^* \end{vmatrix}$$

For option 2

$$[Q]^{-1} = \begin{vmatrix} a\cos\gamma & b & c\cos\alpha \\ -a\sin\gamma\cos\beta^* & 0 & c\sin\alpha \\ a\sin\gamma\sin\beta^* & 0 & 0 \end{vmatrix}$$

For option 3

$$[Q]^{-1} = \begin{vmatrix} a\cos\beta & b\cos\alpha & c \\ a\sin\beta & -b\sin\alpha\cos\gamma^* & 0 \\ 0 & b\sin\alpha\sin\gamma^* & 0 \end{vmatrix}$$

APPENDIX 2

Rotation of Axes v. Rotation of Co-ordinates

The interpretation of rotation function maxima often leads to confusion over whether axes or co-ordinates are moving. (The Patterson density can be described as a set of co-ordinates).

The definition of Eulerian angles are usually given as rotation of axes. Crowther's system first rotates the axes about K_0 by α, then rotates by β about the new axis, J_1, followed by a rotation of γ about the new axis, K_2.

In practice however we want to match one set of density to another, relative to a <u>fixed</u> axial system.

Using matix notation it is easy to visualise how this definition of $[\Omega]$ affects the co-ordinates. A co-ordinate can be written as

$$\underset{\sim}{I}_0 x + \underset{\sim}{J}_0 y + \underset{\sim}{K}_0 z \equiv \underset{\sim}{I}_0\ \underset{\sim}{J}_0\ \underset{\sim}{K}_0 \begin{vmatrix} x \\ y \\ z \end{vmatrix}$$

If we rotate the axes by $[\Omega_\alpha] = \begin{vmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{vmatrix}$

the new vector relative to these axes is

$$\underset{\sim}{I}_1 x + \underset{\sim}{J}_1 y + \underset{\sim}{K}_1 z \equiv [\underset{\sim}{I}_0\ \underset{\sim}{J}_0\ \underset{\sim}{K}_0]\ [\Omega_\alpha] \begin{vmatrix} x \\ y \\ z \end{vmatrix}$$

Rotating the axes

$$\underset{\sim}{I}_1,\ \underset{\sim}{J}_1,\ \underset{\sim}{K}_1 \text{ by } [\Omega_\beta] = \begin{vmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{vmatrix}$$

gives a new vector

$$\underset{\sim}{I}_2 x + \underset{\sim}{J}_2 y + \underset{\sim}{K}_2 z = [\underset{\sim}{I}_0\ \underset{\sim}{J}_0\ \underset{\sim}{K}_0]\ [\Omega_\alpha]\ [\Omega_\beta] \begin{vmatrix} x \\ y \\ z \end{vmatrix}$$

and finally

$$\underset{\sim}{I}_3 x + \underset{\sim}{J}_3 y + \underset{\sim}{K}_3 z = [\underset{\sim}{I}_0\ \underset{\sim}{J}_0\ \underset{\sim}{K}_0]\ [\Omega_\alpha]\ [\Omega_\beta]\ [\Omega_\gamma] \begin{vmatrix} x \\ y \\ z \end{vmatrix}$$

where

$$[\Omega_\gamma] = \begin{vmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

and

$$[\Omega] = [\Omega_\alpha]\ [\Omega_\beta]\ [\Omega_\gamma]$$

Obviously the final co-ordinates relative to $[I_0\ J_0\ K_0]$ are

$$[\Omega_\alpha]\ [\Omega_\beta]\ [\Omega_\gamma] \begin{vmatrix} x \\ y \\ z \end{vmatrix}$$

implying rotation of the co-ordinates, by first γ, then β, then α.

APPENDIX 3

This is based on Tony Crowther's description of the Fast Rotation Function in the book "The Molecular Replacement Method" and his program description. The conventions used here are those used in the program which are different from those in the book.

The rotation function is generally used to correlate a spherical volume of a given Patterson density with rotated versions of either itself or another Patterson density. Since we are dealing with rotations of spherical volumes, it is likely that a more natural form for the rotation function than that given by Rossmann and Blow[10] should be derivable if, instead of working with the Cartesian Fourier components $|F_{\underset{\sim}{h}}|^2$ of the crystal, we expand the Patterson density within the sphere in terms of more appropriate functions, namely spherical harmonics.

Using spherical polar co-ordinates (r,θ,ϕ) for the vector $\underset{\sim}{u}$ we can expand

$$P(r,\theta,\phi) = \sum_{\ell mn} a_{\ell mn} S_{\ell mn}(r,\theta,\phi) .$$

We define the normalised expansion functions:

$$S_{\ell mn}(r,\theta,\phi) = \hat{j}_\ell(k_{\ell n} r)\, \hat{Y}_\ell^m(\theta,\phi)$$

Here $\hat{J}_\ell$ is a normalised spherical Bessel function of order ℓ and $k_{\ell n}$ is such that $\hat{j}_{\ell n}(k_{\ell n} a)=0$, $(n=1,2\ldots)$ where a is the radius of the chosen sphere of Patterson density.

$$\hat{j}_\ell(k_{\ell n} r) = \frac{\sqrt{2}}{a^{3/2}\, j_{\ell-1}(k_{\ell n} a)}\, j_\ell(k_{\ell n} r)$$

The normalised spherical harmonics $\hat{Y}_\ell^m(\theta,\phi)$ are given in terms of the associated Legendre polynomials P_ℓ^m by

$$\hat{Y}_\ell^m(\theta,\phi) = i^{|m|-m} \sqrt{\frac{(2\ell+1)}{4\pi}\, \frac{(\ell-|m|)!}{(\ell+|m|)!}}\, P_\ell^m(\cos\theta)\, e^{im\phi}$$

When integrated throughout the sphere of radius a, the $S_{\ell mn}$ then satisfy the orthogonality relation:

$$\int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{r=0}^{a} S^*_{\ell mn}\, S_{\ell' m' n'}\, r^2 \sin\theta\, dr\, d\theta\, d\phi = \delta_{\ell\ell'}\delta_{mm'}\delta_{nn'} \qquad (1)$$

Evaluation of almn

$$a_{\ell mn} = \int_{r<a} P(r,\theta,\phi)\, \hat{j}_\ell(k_{\ell n} r)\, \hat{Y}_\ell^{m*}(\theta,\phi)\, r^2 \sin\theta\, dr\, d\theta\, d\phi$$

The Patterson density can also be represented as

$$P(\underline{u}) = P(r,\theta,\phi) = \frac{1}{V} \sum_{\underset{\sim}{h}} |F_{\underset{\sim}{h}}|^2 \exp(2\pi i\, \underline{u}.\underline{h})$$

where the vector $\underline{u}$ can be represented in spherical polar co-ordinates (r,θ,ϕ) and the reciprocal lattice vector $\underline{h}$ is represented in spherical polars as (R_h,Θ_h,Φ_h).

So

$$a_{\ell mn}$$

$$= \frac{1}{V} \sum_{\underset{\sim}{h}} |F_{\underset{\sim}{h}}|^2 \int_{r<a} \exp(2\pi i \underline{u}.\underline{h}) \hat{j}_\ell(k_{\ell n} r) \hat{Y}_\ell^{m*} r^2 \sin\theta\, dr d\theta d\phi$$

$$= \frac{1}{V} \sum_{\underset{\sim}{h}} |F_h|^2\, T^*_{\ell mn}(R_h,\Theta_h,\Phi_h)$$

where $T_{\ell mn}$ is the Fourier transform of a normalised expansion function, $S_{\ell mn}$, sampled at the reciprocal lattice points $\underset{\sim}{h}$ of the crystal. But (Morse and Feshbach, p.1467):

$$\int_0^\pi d\theta \int_0^\pi \sin\theta\, d\theta\, e^{2\pi i \underline{u}.\underline{h}}\, \hat{Y}_\ell^{m*}(\theta,\phi)$$

$$= 4\pi\, i^\ell\, \hat{Y}_\ell^{m*}(\Theta_h,\Phi_h)\, j_\ell(2\pi r R_h)$$

so

$$T^*_{\ell mn} = 4\pi\, i^\ell\, \hat{Y}_\ell^{m*}(\Theta_h,\Phi_h) \int_0^a \hat{j}_\ell(k_{\ell n} r)\, j_\ell(2\pi r R_h) r^2\, dr.$$

$$\int_0^a \hat{j}_\ell(k_{\ell n} r) j_\ell(2\pi r R_h) r^2\, dr =$$

$$\frac{\sqrt{2}}{a^{3/2} j_{\ell-1}(k_{\ell n} a)}\, \frac{a^2}{(k_{\ell n}^2 - 4\pi^2 R_h^2)} \left| -k_{\ell n} j_{\ell-1}(j_{\ell n} a) j_\ell(2\pi a R_h) \right|$$

$$= \sqrt{2a}\, \frac{k_{\ell n} j_\ell(2\pi a R_h)}{\left[4\pi^2 R_h^2 - k_{\ell n}^2\right]}$$

$$T^*_{\ell mn} = 4\pi\, i^{\ell}\, \sqrt{2a}\, \frac{k_{\ell n} j_{\ell}(2\pi a R_h)}{\left[4\pi^2 R_h^2 - k_{\ell n}^2\right]}\, \hat{Y}^{m*}_{\ell}(\Theta_h, \Phi_h)$$

$$= 4\pi\, \sqrt{2a}\, \frac{k_{\ell n} j_{\ell}(2\pi a R_h)}{\left[4\pi^2 R_h^2 - k_{\ell n}^2\right]}\, \hat{P}^{m}_{\ell}(\cos\,\Theta_h)\, e^{i(\ell\frac{\pi}{2} - m\Phi_h)}$$

so $a_{\ell mn} =$

$$\frac{4\pi\,\sqrt{2a}}{V} \sum_{\underset{\sim}{h}} \left|F_{\underset{\sim}{h}}\right| 2 \frac{k_{\ell n} j_{\ell}(2\pi a R_h)}{\left[4\pi^2 R_h^2 - k_{\ell n}^2\right]}\, \hat{P}^{m}_{\ell}(\cos\,\Theta_h)\, e^{i(\ell\frac{\pi}{2} - m\Phi_h)}$$

Note that the transforms $T_{\ell mn}$ of the expansion functions $S_{\ell mn}$ are orthogonal on the infinite domain in the transform. Since the summation over $\underset{\sim}{h}$ used in generating the $a_{\ell mn}$ will be truncated after a finite number of terms, the estimated coefficients $a_{\ell mn}$ will be in error. Better estimates could be found by expanding the Patterson density $P(r,\theta,\phi)$ directly in terms of the $S_{\ell mn}$, which are orthogonal on the finite domain $r<a$. This would require prior calculation of the Patterson density from the intensities.

If we perform a rotation $[\Omega]$ specified by Eulerian angles (α,β,γ), a rotated harmonic $\Omega\{Y^m_\ell\}$ of degree $\underline{\ell}$ may be expressed as a weighted sum of the $(2\ell+1)$ unrotated harmonics Y^m_ℓ of degree $\underline{\ell}$ in the form (Hamermesh, 1962)[7]:

$$\Omega\{\hat{Y}^m_\ell(\theta,\phi)\} = \sum_{q=-\ell}^{\ell} D^{\ell}_{mq}(\Omega)\, \hat{Y}^q_\ell(\theta,\phi) \qquad (2)$$

The $D^{\ell}_{mq}(\Omega)$ have the form:

$$D^{\ell}_{mq}(\Omega) = e^{-im\alpha}\, d^{\ell}_{mq}(\beta)\, e^{-iq\gamma}$$

The matrix elements $d^{\ell}_{mq}(\beta)$ of the rotation group can be conveniently calculated by recurrence relations (Altmann and Bradley, 1963)[1]. They refer to rotations of spherical harmonics and need therefore to be calculated only once for all rotation functions.

If we now have two Patterson densities $P_1(r,\theta,\phi)$ and $P_2(r,\theta,\phi)$, we may expand them within the spherical volume $r<a$ in the form:

$$P_1(r,\theta,\phi) = \sum_{\ell mn} a^*_{\ell mn}\, \hat{j}_\ell(k_{\ell n} r)\, \hat{Y}^{m*}_\ell(\theta,\phi) \qquad (3)$$

$$P_2(r,\theta,\phi) = \sum_{\ell' m' n'} b_{\ell' m' n'}\, \hat{j}_{\ell'}(k_{\ell' n'} r)\, \hat{Y}^{m'}_{\ell'}(\theta,\phi) \qquad (4)$$

The rotation overlap is then defined as:

$$R(\Omega) = \int_{\text{sphere}} P_1(r,\theta,\phi)\, \Omega\{P_2(r,\theta,\phi)\}\, r^2 \sin\theta\, dr\, d\theta\, d\phi$$

where $\Omega\{P_2\}$ is the rotated version of P_2 resulting from the rotation $\Omega(\alpha,\beta,\gamma)$. Substituting from (3) and (4) gives:

$$R(\Omega) = \int_{\text{sphere}} \sum_{\ell mn} a^*_{\ell mn}\, \hat{j}_\ell(k_{\ell n} r)\, \hat{Y}^{m*}_\ell(\theta,\phi)$$

$$\sum_{\ell' m' n'} b_{\ell' m' n'}\, \hat{j}_{\ell'}(k_{\ell' n'} r)\, \Omega\{\hat{Y}^{m'}_{\ell'}(\theta,\phi)\}\, dV$$

Substituting for the rotational harmonics from (2) we obtain:

$$R(\Omega) = \int_{\text{sphere}} \sum_{\ell mn} a^*_{\ell mn}\, \hat{j}_\ell(k_{\ell n} r)\, \hat{Y}^{m*}_\ell(\theta,\phi)$$

$$\sum_{\ell' m' n'} b_{\ell' m' n'}\, \hat{j}_{\ell'}(k_{\ell' n'} r) \sum_{q} D^{\ell'}_{m' q}\, \hat{Y}^{q}_{\ell'}(\theta,\phi)\, dV$$

Using the orthogonality relations (1) for the expansion functions this now reduces to:

$$R(\Omega) = \sum_{\ell m m' n} a^*_{\ell mn}\, b_{\ell m' n}\, D^{\ell}_{m' m}(\Omega)$$

We may perform the radial summation $\underline{n}$ independently of the rotation Ω and so writing

$$C_{\ell mm'} = \sum_{n} a^*_{\ell mn}\, b_{\ell m' n} \qquad (5)$$

we get finally:

$$R(\Omega) = \sum_{\ell mn'} c_{\ell mm'}\, D^{\ell}_{m' m}(\Omega)$$

or

$$R(\beta,\alpha,\gamma) = \sum_{mm'} \left(\sum_{\ell} c_{\ell mm'}\, d^{\ell}_{m' m}(\beta)\right) e^{-im'\gamma - im\alpha} \qquad (6)$$

where $[\Omega]$ is defined in terms of the Eulerian angles α, β and γ.

Note:

The rotation matrices for spherical harmonics are calculated from recurrence relations (27) and (28) of Altmann and Bradley (1963). Notice however that because we are considering a rotated density rather than a rotated axial system, the $D^{\ell}_{mq}(\Omega)$ that are used in the Fast Rotation Function are the Hermitian conjugates (transposed complex conjugates) of those used by Altmann and Bradley.

It is now apparent that by using expansion functions appropriate to the rotation group (rather than the Fourier series appropriate to the translation group of the crystal), the rotation function has been split into two parts. The coefficients $c_{\ell mm'}$ refer to a particular pair of Patterson densities and are independent of the rotation Ω. The coefficients $D^{\ell}_{m',m}(\Omega)$, which contain the whole rotational part of the problem, refer to rotations of spherical harmonics and are independent of the particular Patterson densities. This new form for rotational correlations is comparable with that derived previously for translational correlations (Blow, Rossmann, Harding and Coller, 1964; Crowther and Blow, 1967), where the same type of separation occurs.

Furthermore, because of the form of the rotation coefficients $D^{\ell}_{m',m}$ two of the three summations in expression (6), those over $\underline{m}$ and $\underline{m}'$, occur as Fourier series. They may therefore be performed very efficiently by using fast Fourier summing techniques (Cooley and Tukey, 1965).

Computation of the fast rotation function therefore proceeds as follows. We set up and store the elements $d^{\ell}_{m',m}(\beta)$ of the rotation matrices for each β. This need only be done once.

The Bessel function $\hat{j}_{\ell}(2\pi a R_h)$ are calculated from the standard recurrence relation, taking care to set $j_{\ell}(x) = 0$ for small values of the argument since the recursion otherwise diverges.

The normalised associated Legendre polynomials $\hat{P}^{m}_{\ell}(\cos\theta_h)$ are computed by <u>backwards</u> recursion using Wiggins and Saito (1971) Algorithm 2, modified to generate the normalised functions. Note that forward recursion using the standard formulae is highly unstable.

From the two sets of expansion coefficients we calculate the coefficients $c_{\ell mn}$, using equation (5). Then for each β we perform the $\underline{\ell}$ summation in expression (6). The resulting coefficients form the input to a fast Fourier transform program which produces values of the rotation function for all α and γ for each β.

Preliminary studies indicate that this new way of evaluating the rotation function is at least 100 times faster than previous methods. Furthermore, the approximations introduced in previous methods by truncating the interpolation function G and by using only large values of $|F_{\underset{\sim}{h}}|^2$ (Tollin and Rossmann, 1966) need no longer be made. The fast rotation function is therefore potentially more accurate, though whether it will be significantly so in practice remains to be investigated. A more detailed description of the fast rotation function and its applications is in preparation.

These are some further notes from the program description.

5. SYMMETRIES OF THE COEFFICIENTS $a_{\ell mn}$

(a) Because the Patterson density $P(r,\theta,\phi)$, is real the $a_{\ell mn}$ satisfy a condition analogous to the Friedel relation, namely:

$$a_{\ell mn} = (-1)^{m} a^{*}_{\ell mn}$$

Hence the coefficients $c_{\ell mm'}$ in the rotation function summation satisfy

$$c_{\ell -m -m'} = (-1)^{m+m'} c^{*}_{\ell mm'}$$

But the rotation matrices $d^{\ell}_{m',m}(\beta)$ are real and satisfy

$$d^{\ell}_{m',m} = (-1)^{m+m'} d^{\ell}_{-m',-m}(\beta)$$

so finally the coefficients $f_{mm'}$ satisfy $f_{mm'} = f^{*}_{-m-m'}$ and the final Fourier series defining the rotation function has a real sum as expected.

(b) The Patterson density is centrosymmetric so only those coefficients with $\underline{\ell}$ even are non-zero, because

of the symmetry of the spherical harmonics $Y_\ell^m(\theta,\phi)$.

(c) The spherically symmetrical terms $\ell=0$ may be omitted as they contribute only a constant to the rotation function.

(d) If Patterson P_1 has a p_1-fold rotation axis along K_o, only those $a_{\ell mn}$ for which m is a multiple of p_1 are non-zero. If Patterson P_2 has a p_2-fold rotation axis along K_o, only those $b_{\ell'm'n'}$ for which m' is a multiple of p_2 are non-zero. The unit cell of the rotation function is then reduced to $0 \leq \alpha \leq 2\pi/p_1$ and $0 \leq \gamma \leq 2\pi/p_2$. We can take advantage of this when computing the final Fourier series $\Sigma f_{mm'} e^{-imr-im'r}$ and compress the coefficients $f_{mm'}$: the Fourier summation then computes only a single unit cell, which is sampled $1/p_3$ as finely in α and $1/p_2$ as finely in γ as it would be if there were no symmetry.

6. NUMBER OF TERMS TO BE INCLUDED FOR A GIVEN FOURIER CUT-OFF

If intensities are to be included out to a Fourier cut-off of RESMAX Å for a cut-off sphere of radius a Å in the Patterson density, the maximum value of m and hence of ℓ that should be included is approximately

$$\ell_{max} = \frac{2\pi a}{RESMAX}$$

The peak of the transform $T_{\ell mn}(R,\Theta,\Phi)$ of an expansion function $S_{\ell mn}(r,\theta,\phi)$ occurs at a radius of approximately

$$R_{\ell n} \doteq \frac{k_{\ell n}}{2\pi}$$

Hence for a given ℓ, radial terms in n for which

$$R_{\ell n} = \frac{k_{\ell n}}{2\pi} < \frac{1}{RESMAX}$$

should be included. This may be written as

$$\lambda_{\ell n} < \frac{2\pi a}{RESMAX} \text{ where } j_\ell(\lambda_{\ell n}) = 0$$

As presently implemented the program allows a largest value for ℓ of 60.

REFERENCES

1. S.L. Altmann and C.J.Bradley, Phil. Trans. Roy. Soc. Lond. A255, (1963) 193.

2. D.M. Blow, M.G. Rossmann, M.M. Harding and E. Coller, Acta Cryst. 17, (1964) 338.

3. J.W. Cooley and J.W. Tukey, Maths. of Computation 19, (1965) 297.

4. R.A. Crowther and D.M. Blow, Acta Cryst. 23, (1967) 544.

5. R.A. Crowther, The Fast Rotation Function in "The Molecular Replacement Method" ed. M.G. Rossmann.

6. R.A. Crowther, Program description, private communication.

7. M. Hamermesh, "Group Theory and its Application to Physical Problems" (Reading, Mass.; Addison-Wesley Publishing Co. Inc., 1962).

8. E.E. Lattman and W.E. Love, Acta Cryst. B26, (1967) 1854.

9. Morse and Feshbach, "Methods of Mathematical Physics" Part II, (New York; McGraw-Hill, 1953) p.1467.

10. M.G. Rossmann and D.M. Blow, Acta Cryst. 15, (1962) 24.

11. P. Tollin and M.G. Rossmann, ACta Cryst. 21, (1966) 872.

12. R.A. Wiggins and M. Saito, "Evaluation of Computational Algorithms for the Associated Legendre Polynormals by Interval Analysis, Bull. Seismological Soc. Amer. 61 (1971) 375.

APPENDIX 4

We know [Ω], a rotation matrix and the symmetry matrices for the crystal space groups $[S_{1j}]j = 1$, $nsym_1$ for crystal 1, and $[S_{2i}]i = 1, \ldots nsym_2$ for crystal 2.

Now

$$[\Omega_{11}] \begin{vmatrix} x_2 \\ y_2 \\ z_2 \end{vmatrix} = \begin{vmatrix} x_1 \\ y_1 \\ z_1 \end{vmatrix} \quad [\underset{\sim}{x}] \text{ referred to orthonormal axes}$$

$$\text{and} \quad \begin{vmatrix} x \\ y \\ z \end{vmatrix} = [Q]^{-1} \begin{vmatrix} \xi \\ \eta \\ \zeta \end{vmatrix} \quad [\xi] \text{ referred to crystal axes}$$

$$\text{i.e.} \quad [\Omega_{11}] \; [Q_2]^{-1} \begin{vmatrix} \xi_2 \\ \eta_2 \\ \zeta_2 \end{vmatrix} = [Q_1]^{-1} \begin{vmatrix} \xi_1 \\ \eta_1 \\ \zeta_1 \end{vmatrix} \qquad (1)$$

If $[\Omega_{ij}]$ is the rotation matrix to rotate

$$[S_{2i}] \begin{vmatrix} \xi_2 \\ \eta_2 \\ \zeta_2 \end{vmatrix} \text{ to } [S_{1j}] \begin{vmatrix} \xi_1 \\ \eta_1 \\ \zeta_1 \end{vmatrix}$$

$$[\Omega_{ij}] \; [Q_2]^{-1} \; [S_{2i}] \begin{vmatrix} \xi_2 \\ \eta_2 \\ \zeta_2 \end{vmatrix} = [Q_1]^{-1} \; [S_{1j}] \begin{vmatrix} \xi_1 \\ \eta_1 \\ \zeta_1 \end{vmatrix} \qquad (2)$$

Matrix manipulation equating $\begin{vmatrix} \xi_1 \\ \eta_1 \\ \zeta_1 \end{vmatrix}$ from equations (1) and (2) gives

$$[Q_1] \; [\Omega_{11}] \; [Q_2]^{-1} = [S_{1j}]^{-1} \; [Q_1] \; [\Omega_{ij}] \; [Q_2]^{-1} \; [S_{2i}]$$

i.e.

$$[\Omega_{ij}] = [Q_1]^{-1} \; [S_{1j}] \; [Q_1] \; [\Omega_{11}] \; [Q_2]^{-1} \; [S_{2i}]^{-1} \; [Q_2]$$

The generation of the $[\Omega_{ij}]$ is easy once $[\Omega_{11}]$ is known and the programme now tabulates for each (Ω_{11}) which gives a maximum overlay all the $[\Omega_{ij}]$ and the sets of (α,β,γ) and (ω,ϕ,ψ) which they define.

MOLECULAR REPLACEMENT STUDIES OF 6PGDH FROM BACILLUS STEAROTHERMOPHILUS

BY

P.D. Carr[1,3,*] E.J. Dodson[2], T.J. Greenhough[1,4],

and

J.R. Helliwell[1,5]

1 Department of Physics, University of Keele, Keele, Staffordshire ST5 5BG, UK
2 Department of Chemistry, University of York, Heslington, York YO1 5DD, UK
3 Present Address: Department of Pure and Applied Physics, UMIST, P.O. Box 88, Manchester M60 1QD, UK
4 Present Address: Department of Biochemistry, University of Alabama, University Station, Birmingham, Alabama 35294, USA
5 Present Address: SERC Daresbury Laboratory, Warrington WA4 4AD, Cheshire, UK

1. INTRODUCTION

Small crystals (≲0.15mm) of 6-phosphogluconate dehydrogenase (6PGDH) extracted from Bacillus stero-thermophilus[(1)] have been used to collect synchrotron data to 4.5Å[(2)]. Using this data the space group and unit cell dimensions have been determined [(2)]. The space group is $P3_221$ with cell dimensions $a = b = 123.1$Å, $c = 147.7$Å, $\alpha = \beta = 90°$, $\gamma = 120°$.

The structure of 6PGDH extracted from sheep's liver has been solved to a resolution of 2.6Å by the Oxford group[(3)]. They kindly sent us a set of atomic coordinates which have been used for the molecular replacement calculations discussed in this paper 6PGDH is a dimer of molecular weight 100,000 daltons. The dimer axis is the sheep liver crystals (space group $C222_1$) is crystallographic at $z = 1/4$, the two monomers being related by the symmetry operator $-x, y, \frac{1}{2} - z$. The dimer axis is non crystallographic in the bacterial cell. The 6PGDH monomer is a roughly ellipsoidal molecule with major axes of 90x60x60Å[(4)].

2. SELF ROTATIONS

The program ALMN written by Tony Crowther and extensively modified by Eleanor Dodson was used to calculate the Rotation Function for the sheep liver F_{calcs} rotated onto the bacterial F_{obs}. Two separate datashells were used. These were 4.5-6Å and 6-10Å. The outer radius of the Patterson sphere of integration was varied between 10 and 40Å. The best results were obtained using a sphere of 3-30Å.

Two peaks were observed on the maps produced using the 6-10Å datashell. The first was at $\alpha\beta\gamma = 90,30,90$ with a peak height of 16.8 (relative to an origin peak of 50). The second was a smaller peak at $\alpha\beta\gamma = 30,20,30$, with a peak height of 13.2. Both of these peaks were fairly broad extending over a large range in Eulerian space (see fig. 1). The rest of the Eulerian maps were very flat.

When the 4.5-6Å datashell was used the second of these two features was not observed, the original peak at $\alpha\beta\gamma = 90,30,90$ being the only major feature of the output Eulerian map. The peak was sharper for this datashell although it still extended over a large area of Eulerian space (see fig. 2).

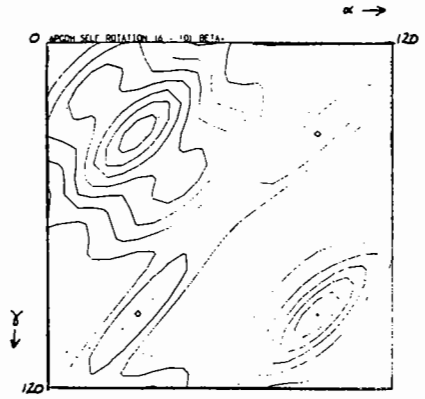
Changing the size of the integration sphere altered the peak heights but did not alter their position on the output maps.

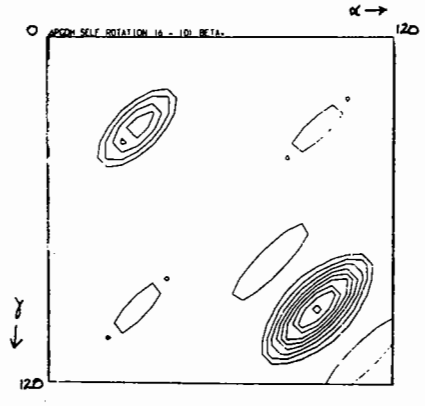
From the Eulerian coordinates of the major feature in the self rotation maps one is able to deduce the two spherical polars (ω, ϕ) defining the direction of the dimer axis in the bacterial cell. However in this particular case $\phi = 90°$ (i e perpendicular to a crystallographic two fold) which leads to an ambiguity in the ω value. Therefore from the self rotation calculations we were able to deduce that the dimer axis in the bacterial cell was at an angle of either 15° or 105° to c, 90° to a and 30° to b.

3. CROSS ROTATIONS

Using the sheep liver atomic cordinates placed in a P1 cell, structure factors were calculated for

*Paper presented by P.D. Carr





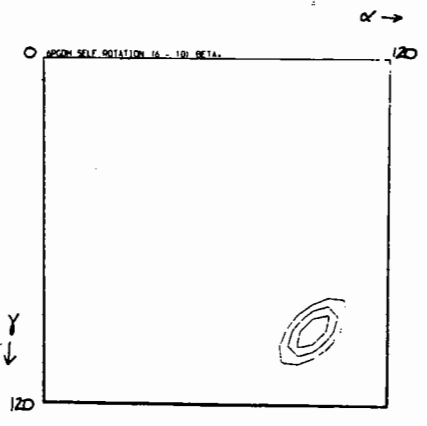
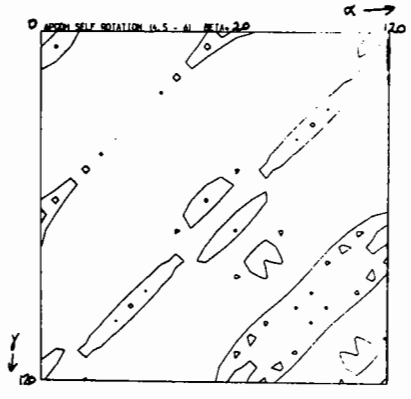
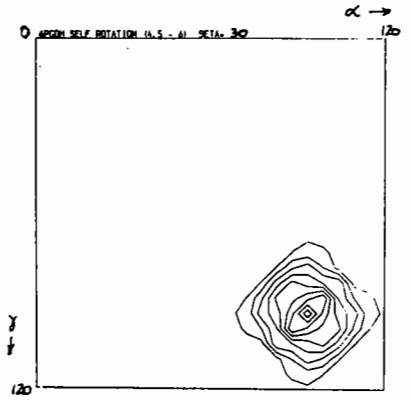


Fig. 1 Contour maps of the Self Rotation Function Using the 6-10Å datashell. Contours every 1.0 above 7.0.





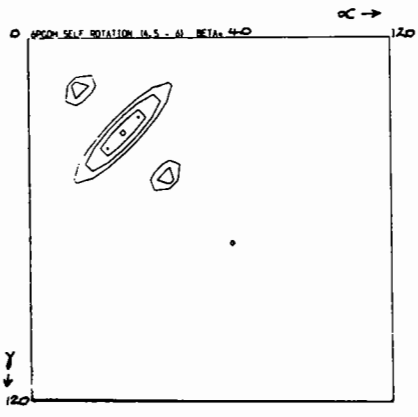


Fig. 2 Contour maps of the Self Rotation Function Using the 4.5-6Å datashell. Contours every 1.0 above 2.0.

Table 1 The three 'best' cross rotation peaks, their symmetry equivalents and directional cosiner. Patterson sphere. 3-30Å, datashells, 4.5Å and 6-10Å.

α	β	γ	Equivalent	ω	φ	χ
35	45	-15	α,β,γ	67.3	115	49.0
155	45	-15	120+α, β, γ	23.8	175	143.2
95	-45	165	60+α,-β,180+γ	151.6	55	107.1
85	135	165	120-α,180-β,180+γ	108.7	-130	154.6
25	-135	-15	60-α,-180+β,γ	87.9	-70	135.2
145	-135	-15	180-α,β-180,γ	69.4	-10	161.4
55	60	30	α,β,γ	40.5	102.5	100.6
175	60	30	120+α, β, γ	249.41	-17.5	158.4
115	-60	-150	60+α,-β,γ-180	117.5	42.5	68.6
65	120	-150	120-α,180-β,γ-180	111.3	-162.5	136.7
5	-120	30	60-α,β-180,γ	80.2	-102.5	123.0
125	-120	30	180-α,β-180,γ	60.6	-42.5	167.6
110	80	125	α,β,γ	136.6	-97.5	138.6
50	-80	-55	α-60,-β,γ-180	93.0	-37.5	80.1
170	-80	-55	α +60, -β,γ -180	44.9	22.5	131.4
10	100	-55	120-α,180-β,γ-180	107.8	122.5	107.1
130	100	-55	240-α,180-β,γ-180	62.9	-177.5	118.7
70	-100	125	180-α,-180+β,γ	129.8	62.5	170.4

Table 2 Value of the Rotation Function (RF) obtained at points corresponding to (i) ω, ∅, χ = 105,90,180 (or Lane group equivalents) and (ii) ω, ∅, χ = 15,90,180 (or Lane group equivalents).

	α	β	γ	Monomer $RF_{4.5-6}$	Monomer RF_{6-10}	Dimer $RF_{4.5-6}$	Dimer $RF_{6.10}$	
(i)	90	15	270	-35	14	- 4	19	
	45	20	310	-26	- 5	- 4	- 6	
	15	20	230	-30	- 8	7	-71	
	15	20	50	44	36	- 4	- 6	
	45	20	130	3	-28	7	-71	
	35	25	320	-19	-12	14	1	
	25	25	220	-32	-27	-13	-95	
	25	25	40	-14	41	14	1	
	35	25	140	- 5	-53	-13	-95	
	30	30	330	- 8	-22	16	1	
	30	30	210	4	-35	23	-56	
	30	30	30	3	41	16	1	
	30	30	150	-19	-21	23	-56	
	20	35	335	5	-38	-22	-35	
	35	35	205	19	- 3	21	-24	
	20	35	155	- 6	10	29	-24	
	40	35	25	41	27	-22	-35	
	20	40	335	13	-26	0	-28	
	40	40	205	6	8	1	-31	
	40	40	25	30	30	0	-28	
	20	40	155	-39	6	1	-31	
	15	45	340	26	- 5	46	16	
	45	45	200	5	-22	- 9	-58	
	45	45	20	38	47	46	16	*
	15	45	160	- 31	-15	- 9	-58	
	15	50	340	3	-16	39	49	
	45	50	200	7	-46	-16	-61	
	45	50	20	47(54)	63	39	49	*
	15	50	160	-52	-20	-16	-61	
	10	55	340	9	- 6	- 1	67	
	50	55	200	5	-36	-12	-35	
	50	55	20	8(67)	51	- 1	67	
	10	55	160	-38	3	-12	-35	
	10	60	345	9	- 1	5(44)	100	†
	50	60	195	-24	-14	-11	-35	
	50	60	15	44	104	5(44)	100	*
	10	60	165	5	-14	-11	-35	
	5	65	345	40	7	- 7	96	
	55	65	195	-10	9	6	-33	
	55	65	15	- 1	99	- 7	96	
	5	65	165	24	-15	6	-33	
	5	70	345	46	15	- 4	80	
	55	70	195	22	5	- 8	-10	
	55	70	15	- 8	71	- 4	80	
	5	70	165	- 9	5	- 8	-10	
	5	75	345	19	23	-27	85	
	55	75	195	11	12	- 1	8	
	55	75	15	- 7	62	-27	85	
	5	75	165	0	0	- 1	8	
	5	80	345	- 9	23	-30	110	
	55	80	195	-23	25	- 4	32	
	55	80	15	-10	73	-30	101	
	5	80	165	19	- 9	-4	32	
	0	85	345	4	13	-12	123	
	60	85	195	-11	31	37	53	
	60	85	15	-16	101	-12	123	
	0	85	165	16	- 2	37	53	

† dimer related to (50,60,10)

(ii)	30	75	90	- 8	29	21	-24
	30	75	270	36(82)	-72	21	-24
	80	80	80	2	12	41	-30
	100	80	100	-28	12	-46	-23
	100	80	280	3	-55	41	-30
	80	80	260	- 7	- 8	-46	-23
	100	85	75	8	-26	0	-19
	80	85	105	5	-47	11	-100
	80	85	285	43	-20	0	-19
	100	85	255	2	-48	11	-100

both a 6PGDH monomer and dimer. These F_{calcs} were used to perform cross rotation functions using the same datashells, sphere of integration and temperature factors as were used for the self rotations. Problems were encountered in trying to decide on an appropriate integration sphere for the cross rotations because of the size of the 6PGDH molecule. Any sphere chosen will necessarily either exclude large numbers of intramolecular vectors or include large numbers of intermolecular vectors for this elliptical molecule. The limitation of the maximum order of Bessel functions used being 30 also restricted the range of integration radius, a, and maximum resolution, RESMAX, to $\frac{2\pi a}{RESMAX} \leqslant 30$. Several combinations were tried and the results presented in this paper are those obtained using a spherical shell with an inner radius of 3Å and an outer radius of 30Å.

The resulting maps had many peaks and bands of high density on them. Unlike the self-rotation maps the peak heights and positions were vulnerable to the choice of calculating parameters. The peaks were all listed and checked for consistency between datashells and monomer/dimer runs. Only three peaks showed any real signs of consistency and for these and their symmetry related positions the spherical polars were calculated. None of these were very close to the $\omega\phi$ = 105,90 or 15,90 that we were expecting (see table 1).

As simple inspection of the major peaks did not reveal what we were expecting we decided to predict where the cross rotation peaks would occur, using our knowledge of ω and ϕ from the self rotations and the fact that the dimer axis is parallel to b in the sheep liver cell. Using the expression;

$$[\Omega]_{C222_1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} DC1 \\ DC2 \\ DC3 \end{bmatrix}_{P3_221}$$

expressing the directional cosines DC1,DC2 and DC3 in terms of spherical polars and the $[\Omega_{12}]$, $[\Omega_{22}]$ and $[\Omega_{32}]$ elements in terms of Eulerian angles $\alpha\beta\gamma$ then by comparing matrix elements we obtained the following expressions:

$$-\cos\alpha.\cos\beta.\sin\gamma - \sin\alpha.\cos\gamma = \sin\omega.\cos\phi$$
$$-\sin\alpha.\cos\beta.\cos\gamma + \cos\alpha.\cos\gamma = \sin\omega.\sin\phi$$
$$\sin\beta.\sin\gamma = \cos\omega$$

using these expressions, values for α and γ for each β section were calculated for both ω,ϕ = 15,90 and 105,90. Table 2 shows values of the cross rotation function obtained for the monomer, dimer and both datashell runs at these calculated positions on the output Eulerian maps.

Only one position gave high values for all four runs this was at $\alpha\beta\gamma$ = 50,60,15. This was in the same area as the $\alpha\beta\gamma$ = 55,60,30 peak which was one of the three hopefuls from simple inspection. A similar procedure was carried out for the smaller self rotation feature but none of the predicted peaks gave consistency over all four runs. The spherical polars for $\alpha\beta\gamma$ = 50,60,15 were $\omega\phi\chi$ = 105,90, - 62.4.

4. TRANSLATION SEARCH

An R-factor search using the program SEARCH was conducted on a 1Å grid for both a $P3_221$ cell and a $P3_121$ cell, as we did not know which of this enanteomorphic pair was the correct group. A finer 0.2Å grid was used around the translations with lowest R-factors. The minimum R values obtained were 52.75% for the $P3_121$ cell and 50.30% for the $P3_221$ cell. Neither of these were strikingly low compared to the minima on other sections nor were they very close to 45% which one would expect for a good fit of the model structure.

The programs LSQKAB and PRJANG were used to apply the rotation matrix, translation vector and symmetry operations to the sheep liver atomic coordinates and also to check for overlapping of the C_α atoms in the bacterial cell. The results show that fewer overlaps occurred for the translation containg symmetry operators when the $P3_221$ cell was used.

REFERENCES

1. B.M.F. Pearse and J.I. Harris, Febs Letters 38, (1973) 49.

2. P.D. Carr, Ph.D thesis, University of Keele, (1984).

3. M.J. Adams, I.G. Archibald, C.E. Bugg, A. Carne, S. Gover, J.R. Helliwell, R.W. Pickersgill and S.W. White, Embo Journal 2, (1983) 1009.

4. M.J. Adams, J.R. Helliwell and C.E. Bugg, J. Mol. Biol. 112, (1977) 183.

THE MOLECULAR REPLACEMENT METHOD AND GRAMICIDIN S

by

M.M. Harding

IPI Chemistry Department, Liverpool University, Liverpool L69 3BX

1. INTRODUCTION

The molecular replacement method was the key to solving the structure of the hexagonal form of gramicidin S, so I wish to take a look back at the history of this.

Gramicidin S is a cyclic decapeptide antibiotic which was characterised by Synge and others in 1948[1,2]

L-pro - L-val - L-orn - L-leu - D-phe
↑ ↓
D-phe - L-leu - L-orn - L-val - L-pro

Very soon Dorothy Hodgkin, Gerhardt Schmidt and Beryl Oughton (Rimmer) were looking at a variety of crystalline forms, contemplating a crystal structure determination, and building models.[3] Late in the 1950's the hexagonal N-acetyl form, with chloro-, bromo-, and iodo-derivatives was chosen (see table) and data collection went ahead. (I became involved in the work at this stage). No solution for the structure was forthcoming despite all these heavy atoms, probably because the isomorphism was not very good and the pseudosymmetry confused the issue. From 1965 onward I made a series of attempts to solve it by what is now known as the molecular replacement method.

As I see it the molecular replacement method involves three stages:
(i) finding a structure model for all or a sufficiently large part of the molecule in the unknown structure,
(ii) finding the orientation and position in the cell of the model structure, usually by rotation and translation functions, sometimes by structure factor calculation,
(iii) calculating phases from the model and an electron density map which allow the remainder of the structure to be found and refined. It is only in stage (iii) that you can know you have been successful.

None of the early attempts with gramicidin S were successful, but many times during the period 1965-78 new models for the gramicidin S molecule appeared and were tried. I shall describe the series of models, and then the methods used to try to fit them.

2. THE MODELS

The main models were
a) one of Hodgkin and Oughton's models, based on a β-sheet[3]
b) a model put forward by Sheraga et al[4], resulting from energy calculations, and similar to a β-sheet,
c) a model put forward by Liquori et al[5], again based on energy considerations, but more α in type,
d) a model based on torsion angles deduced from n.m.r. data by Ovchinnikov[6] and more like a) than b) or c)
e) the molecule of gramicidin S as found in 1978 in York[7] by the crystal structure determination of its urea complex.

In every case the backbone polypeptide chain and β-carbon atoms were used, but side chains were not, on the grounds that the side chain might adopt different conformations.

3. METHODS OF SEARCHING

At an early stage, a rotation function was used, and indicated a β-sheet orientation roughly normal to the c axis; any more precise results from rotation and translation functions seemed to be precluded by the multiplicity of orientations present as a result of the crystal symmetry, $P6_522$. The main search was a stepwise one, looking for regions of structure factor agreement, preceded by searches for regions where van der Waals contacts were acceptable . These search

methods were very much limited by the computing power available at the time, but assisted by the symmetry and pseudosymmetry; the latter was assumed to hold exactly.

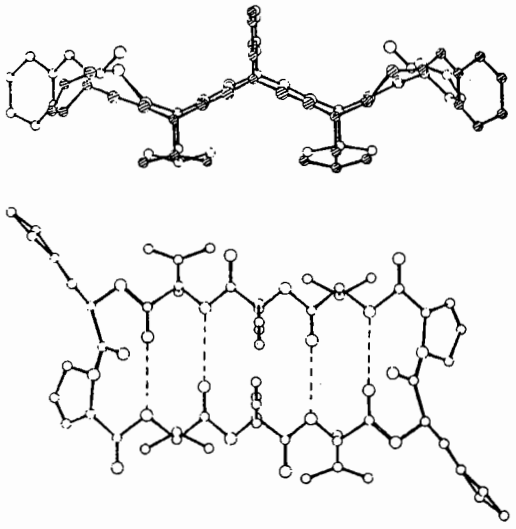


Fig. 1(a) Hodgkin and Oughton model of gramicidin S. Hydrogen atoms are omitted. Oxygen atoms are largest, then nitrogen, then carbon: (below) view along the molecular twofold axis and (above) view nearly normal to this axis. In the first view the atoms in the upper part of the molecular are shaded (Redrawn from Schmidt et al. (1957).

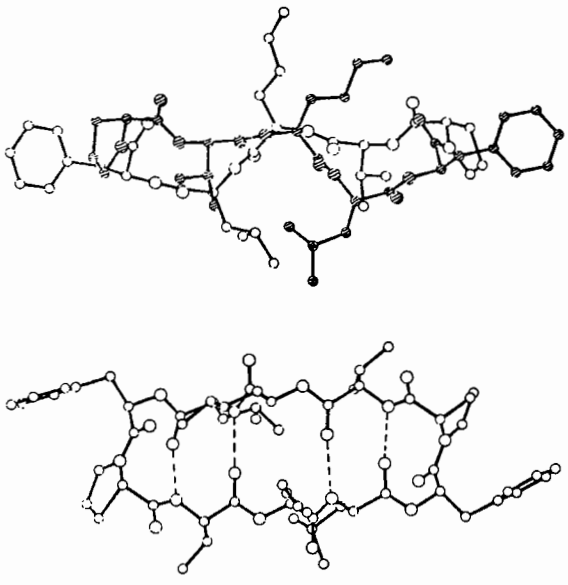


Fig. 1(b) Experimentally determined conformation of gramicidin S in trigonal crystals (Hull et al, 1978); (below) view along the approximate molecular 2-fold axis, and (above) view normal to this axis. Shading as in fig. 1(a).

The model molecule must be placed on a crystallographic diad where it has two degrees of freedom, rotation about the diad (ϕ) and translation along it (D); the twofold axis may be parallel to $[10\bar{1}0]$ or $[11\bar{2}0]$, and the space group may be $P6_122$ or $P6_522$. The structure factor calculcations were done for 30-90 reflections; these reflections were chosen carefully so that the largest and the smallest observed reflections from each parity group were included. The structure factor agreement was assessed by a correlation coefficient; an example for one model is shown in Fig 2.

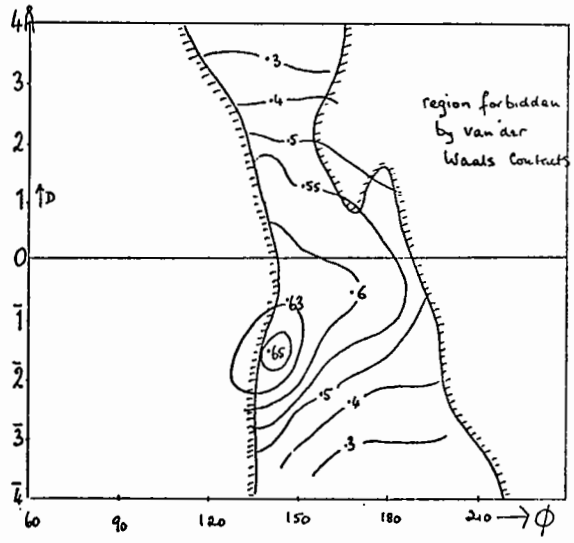


Fig. 2 Structure factor agreement as shown by the correlation coefficient between F_{obs} and F_{calc} evaluated in the region of space where van der Waals contacts are acceptable. The model was Liquori's in 1967, with 2-fold axis parallel to $[10\bar{1}0]$ in $P6_122$.

4. DISCUSSION

Success was apparent when the phenyl rings and parts of other side chains turned up clearly in the map phased on the York model. The structure which emerged is shown in Fig 3. The molecules are tilted in relation to the sixfold screw axis, and their repetition up it makes a twisted β-sheet similar to that found in proteins.

Subsequent work on this form of gramicidin S has been assisted by a new data set, from Yves Mauguin in Paris, and by much computational help from John

Campbell. Two of the three half molecules in the asymmetric unit have refined nicely and the third is in a mess and it is still not entirely clear what is wrong. The two good molecules are not significantly different from the York model in their backbone conformation, but some of the side chains are different.

But it would be useful to look backward through the models tried and see why we did not get a solution earlier. The penultimate model, Ovchinnikov's, was fitted in the cell with the correct tilt and with the centre of gravity in the right position, but the mirror image of the correct peptide chain orientation appeared to fit best; the model does not have the pronounced distortion from a simple β-sheet that the true structure does. Compare fig 1(a) and 1(b). Looking further back the tilt and position in the cell were approximately right in all cases where the model was of a β- type.

So, perhaps, not surprisingly, the lesson we should learn from this is that a model with stereochemistry close to that of the true structure is essential if the molecular replacement method is to work.

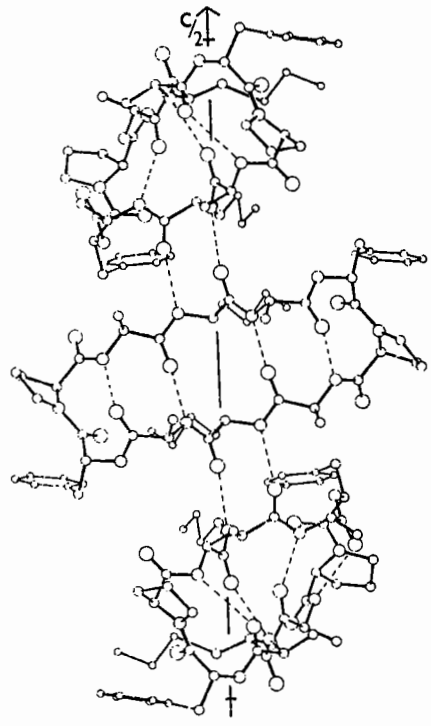


Fig. 3 Structure found for hexagonal N-acetyl gramicidin S. One molecule is viewed along its two-fold axis, and two more are related by the six-fold screw axis along c. Val. ine and leucine side chains are omitted for clarity.

TABLE

N-acetyl gramicidin S

$(C_{32}\ H_{48}\ N_6\ O_6)_2$, MW 1140
Hexagonal
a = 27.5 c = 55.4 A
Spacegroup $P6_522$ (or $P6_122$)
Z = 18 implies 3 pentapeptides
per asymmetric unit, each decapeptide molecule must lie on a 2-fold symmetry axis.
Oscillation photographs, CuK_α radiation,
dmin = 1.9Å along a^*, 1.6Å along c^*.

pseudosymmetry relates the three chemically identical peptides in the asymmetric unit approximately: x y z

$\frac{1}{3}+x, \frac{2}{3}+y, \frac{1}{2}+z$
$\frac{2}{3}+x, \frac{1}{3}+y, \frac{1}{2}+z$

further details in reference (8)

REFERENCES

1. Consden, R.J., Gordon, A.H., Martin, A.J.P., and Synge, R.L.M. (1947). Biochem. J. 41, 596.
2. Synge, R.L.M. (1948). Biochem. J. 42, 99.
3. Schmidt. G.M.J., Hodgkin, D.C., and Oughton B.M. (1957). Biochem. J. 65, 744.
4. Vanderkooi, G., Leach, S.J., Nemethy, G., Scott, R.A., and Scheraga, H.A. (1966). Biochemistry 5, 2991.
5. Liquori, A.M., Santis, P.de, Kovacs, A., and Mazzarella, L (1966). Nature, Lond. 211, 1039.
6. Ovchinnikov, Yu. A., Ivanov, V.T., Bystrov, V.F., Miroshnikov, A.I., Shepel, E.N., Abudullaev, N.D., Efremov, E.S., and Senyavina, L.B., (1970). Biochem. biophys. Res. Commun. 39, 217.
7. Hull, S.E., Karlsson, R., Main, R., Woolfson, M.M., and Dodson, E.J. (1978). Nature Lond. 275, 206.
8. Harding M.M., in 'Structural Studies on Molecules of Biological Interest', ed G. Dodson, J.P. Glusker and D. Sayre, Clarendon Press, Oxford, 1981, p198-206.

LOW RESOLUTION STRUCTURES OF TWO FORMS OF PHOSPHOFRUCTOKINASE

by

P. R. Evans, G.W.Farrants*, M.C.Lawrence+ & Y.Shirakihara
MRC Laboratory for Molecular Biology, Hills Road, Cambridge CB2 2QH

The T-state conformation of B.stearothermophilus phosphofructokinase (PFK), and the R-state conformation of E.coli PFK, have been solved at low resolution from the atomic model derived earlier from the crystal structure of B.stearothermophilus PFK in the active R-state conformation[1]. This paper briefly describes the solutions to illustrate how structures corresponding to known models can be located in low resolution diffraction data, and tells a cautionary tale about using heavy atom derivatives to check a molecular replacement solution.

1. T-STATE B.STEAROTHERMOPHILUS PFK

(i) Location of the model in the cell

B.stearothermophilus PFK crystallizes in space group $P2_12_12_1$, cell dimensions 131.5, 114.7, 96.2 Åfrom 2M potassium tartrate with 10mM 2-phosphoglycollate (an analogue of the allosteric inhibitor phosphoenolpyruvate). These crystals have one tetramer in the asymmetric unit. They diffract rather weakly to about 2.9Å resolution, but at present, data are available only to 7Å resolution. All the calculations described here have used data from 10 to 7Å resolution. The first evidence of the location of the molecular 222 axes came from the native Patterson function. This showed a large peak on one of the Harker sections at uvw = 0.23, 1/2, 0.44: this can be interpreted as arising from a molecular dyad axis approximately parallel to the crystallographic y axis, passing through x=0.12, z=0.47 in the x-z plane. Figure 1 shows the 180° rotation section of the self-rotation function, projected along the crystallographic y axis. This shows the three orthogonal crystallographic dyads as large peaks, the central one obscuring the parallel molecular q axis. The other two molecular axes are on the periphery (marked p and r in figure 1) rotated by about 20° from the crystallographic axes: these three molecular axes p, q & r form a 222 set.

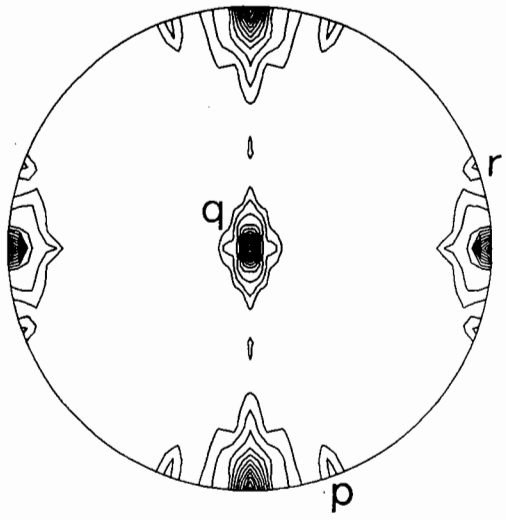


Figure 1. A stereographic projection down the y axis of the T-state B.stearothermophilus self-rotation function. The three molecular dyad axes are marked p, q & r.

The identification of which of these axes corresponds to each molecular axis in the model from the R-state conformation was done from a cross-rotation function between the T-state data and structure factors calculated from a R-state tetramer placed in a large P1 unit cell (208 x 208 x 208 Å). The four peaks expected from superimposing two tetramers are reduced to two because of the molecular dyad lying parallel to the crystallographic screw dyad. The largest two peaks in the cross-rotation function were found in positions consistent with the orientation of the molecular axes from the self-rotation function.

An R-factor search was used to locate the position of the centre of the tetramer along the y axis (using the position on x and z from the native Patterson), and to improve the rotation parameters. The top curve in figure 2 shows the search along y with the molecular axis parallel to the y axis. This orientation gives only a small dip in the R-factor. The other curves show the search with the axis tilted by + or -4°

Current addresses:

* Department of Biochemistry, University of Sheffield, Sheffield S10 2TN

+ Institute of Electron Microscopy, South African Medical Research Council, PO Box 70, Tygerberg 7505, South Africa

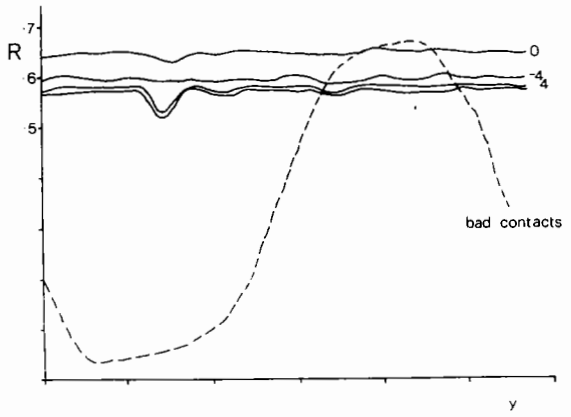


Figure 2. R-factor searches for the B.stearothermophilus T-state molecule (solid lines) along the y-axis. The four lines correspond to orientations of the model (defined by Eulerian angles ρ θ φ) of -20°,0,0; -20°,0,-4°; -20°,0,+4°; and -19.8°,-2.8°,+4.7° in descending order. The dashed line is a crystal packing search, the number of α-carbons less than 6Å apart for each position of the tetramer.

around the z axis, and with the axes in the final orientation from the rigid body refinement (see below). This last curve and the one at +4° tilt show a good minimum. This example shows how the R-factor search may be used to explore a part of the 6-dimensional rotation-translation space of the solution, once an approximate solution is found. In this way the rotation parameters may be improved over those obtained from the rotation function, which is relatively inaccurate because it is based on Patterson functions. Also shown in figure 2 (dashed line) is a crude estimate of the number of bad contacts between crystallographically related molecules (the number of α-carbons closer than 6Å to each other). For these crystals, the packing search is not very discriminating between alternative positions, but it does limit the possible y positions to about 1/3 of the cell. These searches give a clear indication of the placing of the R-state model in the T-state cell.

(ii) Refinement

Starting from the model of the R-state tetramer rotated and translated into the T-state cell as described above, three methods of refinement were used to try to improve the model, using data from 10-7Å resolution.

(a) Rigid-body refinement taking the subunit as the basic unit, and maintaining the 222 symmetry of the tetramer. This refinement included the 6 parameters defining the orientation and position of the subunit relative to the molecular axes, 6 parameters defining the tetramer in the cell, plus a scale factor. During the course of this refinement, the subunits rotated by about 3.5° about the molecular p axis.

(b) A smoothed individual atom refinement, in which individual atom shifts for main chain atoms were taken from the gradient of the difference map[2], and were then smoothed out by taking a running average along the main chain. This was first done with the four subunits treated independently: it was found that they tended to move in directions related by the molecular symmetry, which suggested that the refinement was improving the model. Later refinement was done for one subunit using the averaged difference map.

(c) Rigid-body refinement using CORELS[3], breaking each subunit into 2, 6 then 8 rigid groups. The four subunits were kept equivalent by averaging the transformations for each group over the four subunits between each cycle[4].

The results from these refinements were not conclusive because of the limited resolution of the data. The change in quaternary structure is clear, a rotation of the p-axis dimer relative to the other dimer by nearly 8° about the p axis. Within each subunit there are structural changes, but they are difficult to characterize from refinement at this resolution. Also the refinements did not account for all the changes in structure, since the difference maps after both refinements (b) and (c) still showed significant features which could not be interpreted at this stage.

(iii) A cautionary tale

It is sometimes difficult to know if a proposed molecular replacement solution is correct. At one stage during the analysis of the PFK T-state structure we had problems with the translation search (which later went away), and during this period we did a good deal of work with the tetramer 7Å away from its correct position along

y (at y=21Å instead of 14Å), but in the correct orientation. Since we were unsure of the truth of the solution, we tried to use a poor mercury derivative (the best we had) as a check on the calculated molecular replacement phases, by calculating a difference map. This map had peaks at many of the predicted mercury positions, near the three thiol groups on each subunit, giving the misleading impression that this solution was correct. It is true that when we had the correct solution, the mercury peaks were higher and less sensitive to the exact orientation of the model, but with 5 of the 6 rotation-translation parameters correct, the phases were sufficiently correct for the presence of heavy-atom peaks to give us a false indication that we had the right solution. The presence of peaks in a heavy-atom difference map is clearly not sufficient evidence for the complete truth of a proposed solution.

2. R-STATE E.COLI PFK

The R-state B.stearothermophilus model was located in the E.coli crystal cell in a similar way to that used for the B.stearothermophilus T-state crystals, using rotation functions and R-factor searches at 6Å resolution. E.coli PFK crystallizes from 2-methyl-2,5-pentanediol containing fructose-6-phosphate and Mg/ADP, in space group $P2_12_12$, cell dimensions 112.0, 85.4, 96.4 Å with half a tetramer in the asymmetric unit. The molecule must therefore lie somewhere on the crystallographic dyad axis along z. Figure 3 shows the 180° section of the self-rotation function looking along z. As before, two molecular dyad axes can be seen on the periphery of the section (marked p & r), forming a 222 set with the crystallographic dyad. The orientation of the model was found from a cross-rotation function which showed that the crystallographic z dyad coincides with the molecular q axis. Figure 4 shows the R-factor search used to explore the translation of the centre of the tetramer along z and its rotation around the z-axis. The best value for this rotation of 28.4° was found from rigid-body refinement (type (a) above). The data is now being extended to 2.5Å resolution, and will be phased by molecular replacement and two-fold averaging.

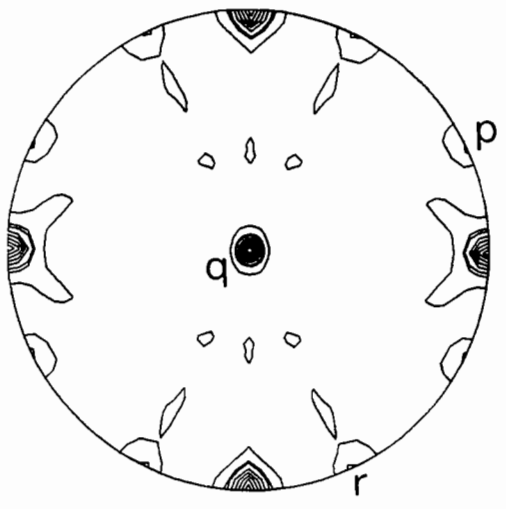


Figure 3. A stereographic projection down the z axis of the R-state E.coli self-rotation function. The three molecular dyad axes are marked p, q & r.

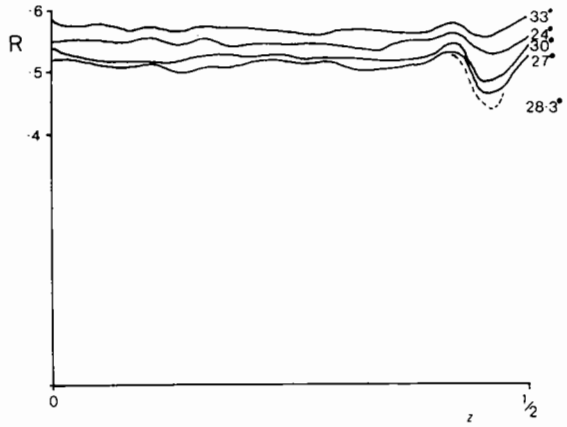


Figure 4. R-factor searches for the E.coli R-state molecule along the z-axis, for different rotations of the molecular p-axis from the crystallographic x-axis.

REFERENCES

1. P.R.Evans, G.W.Farrants, and P.J.Hudson, Phil.Trans.Roy.Soc.Lond., **B293**, (1981) 53-62
2. A.Jack & M.Levitt, Acta Cryst. (1978), **A34**, 931-935
3. J.L.Sussmann, S.R.Holbrook, G.M.Church & S.H.Kim, Acta Cryst. (1977), **A33**, 800-804
4. A.G.W.Leslie, Acta Cryst. (1984), **A40**, 451-459

USE OF MULTI-DIMENSIONAL SEARCH METHODS FOR STRUCTURE DETERMINATION OF LARGE MOLECULES

by

Dov Rabinovich

Department of Structural Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel

Search methods are perhaps the most obvious and direct implementations of molecular replacements principles. However, direct and naive applications of these methods result in prohibitive computing times and poor discrimination criteria. On the other hand, we have shown[1] that small molecular structures can directly and efficiently be solved by a systematic search procedure using geometrical criteria such as global molecular packing considerations and hard sphere atom-atom contacts combined with a set of multi-level sieving operatons, by means of which only a limited number of accepted trial structures are left for further examination by R-factor calculations.

This procedure proved, however, to be inefficient for large molecular structures since it was time consuming as a result of the large number of atoms. It also yielded too many acceptable trial structures owing to the relatively small number of intermolecular atom-atom contacts resulting from the fact that a large fraction of the unit cell volume consists of solvent molecules. Consequently, we have developed a new multi-dimensional search approach[2] incorporating global packing criteria with structure-factor calculations for very-low-resolution x-ray data, which proved to be efficient for the solution of large-molecule structures.

The approach owes its efficiency to four distinct features. The first, global packing considerations, can reduce the number of the search grid points in the six dimensional parameter-space by orders of magnitude. Secondly, the use of very-low-resolution x-ray data renders the approach reasonably insensitive to deviations of the model from the actual structure and permits the use of a coarse parameter-space grid on the one hand, and reduces considerably the structure-factor computation time, on the other. Thirdly, the use of the transforms of the model and the fringe functions of its origin allows extremely fast and efficient algorithms for the calculations of the structure-factors. Lastly, a great improvement in the performance of the approach is achieved by using group scatterers in lieu of the individual atomic scatterers, a procedure justified by the use of low-resolution data.

Several criteria for testing the agreement between the observed and calculated structure factors were considered, e.g. product functions, correlation coefficients and the discrepancy factor R. The latter, which is more indicative for further refinement, was also found to be more discriminative. Accordingly, the set of trial structures (typically 20) with the lowest R was subjected to a further rigid-body least-squares refinement procedure with increasingly higher resolution to yield the correct structure.

Using a computer program, ULTIMA, based on this approach, we have solved the structures of three A-DNA octamers, an A-DNA decamer, a B-DNA dodecamer and a Z-DNA hexamer. We have also reproduced the solutions of another B-DNA dodecamer[2], of $tRNA^{phe}$[3] and of two proteins, lysozime[4] and concanavalin A[5] (see Table 1).

The models for the oligonucleotides were taken from Arnott and Hukins[6]; for the $tRNA^{phe}$ from the refined structure of the monoclinic form[7]; for the triclinic lysozyme from the tetragonal form[8] and for cocanavalin A from the demetallised form[9].

Group scatterers for the oligonucleotides were the phosphates, sugars and base moieties. Most of the amino acids were divided into two groups, where one group consisted of the backbone atoms, the other of the side chain ones.

Table 1

Compound	S.G	NA[a]	NG[b]	ΔO[c]	ΔU[d]	G.R[e]	Res.	R	R_{rb}[f]
GGTATACC	$P6_1$	322	46	30.0	3	0.03	25-10	42	23
GGGGCTCC[g]	$P6_1$	324	46	30.0	3	0.03	25-10	45	28
GGGGTCCC[g]	$P6_1$	324	46	30.0	3	0.03	25-10	46	28
ACCGGCCGGT	$P6_1 22$	404	56	22.5	3	0.33	25-10	41	22
CGCGAATTTGCG[g]	$P2_1 2_1 2_1$	488	70	22.5	3	0.016	25-10	47	31
TGCGCG[g]	$P2_1 2_1 2_1$	242	34	30.0	3	0.01	25-10	32	24
CGCGAATTCGCG[(2)]	$P2_1 2_1 2_1$	486	70	22.5	3	0.016	25-10	42	28
$tRNA^{phe}$[(3)]	$P2_1 22_1$	1652	228	15.0	4	0.15	50-17	40	20
Lysozyme[(4)]	P1	1000	209	22.5	-	0.70	25-10	43	29
Concanavalin A[(5)]	I222	1807	362	15.0	4	0.01	25-15	36	28

[a] Number of atoms

[b] Number of group scatterers

[c] Rotational grid step (^{0})

[d] Translational grid step (A)

[e] Fraction of grid point searched; indicates the efficiency of global packing considerations.

[f] R-factor after 4 cycles of rigid-body refinement

[g] Mutant DNA

REFERENCES

1. D. Rabinovich and Z. Shakked, Acta Cryst. A40 (1984) 195.

2. R.M. Wing, H.R. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura and R.E. Dickerson, Nature 278 (1980).

3. S.-H. Kim, G.J. Quigley, F.L. Suddath, A. McPherson, D. Sneden, J.J. Kim, J. Weinzierl and A. Rich, Science 179 (1973) 285.

4. J. Moult, A. Yonath, W. Traub, A. Smilansky, A. Podjarny, D. Rabinovich and A. Saya, J. Mol. Biol. 100 (1976) 179.

5. G.M. Edelman, B.A. Cunningham, G.N. Reeke Jnr., J.W. Becker, M.J. Waxdal and J.L. Wang, Proc. Nat. Acad. Sci., U.S.A., 69 (1972) 2580.

6. S. Arnott and D.W.L. Hukins, Biochem. Biophys. Res. Commun. 47 (1972) 1504.

7. B. Hingerty, R.S. Brown and A. Jack, J. Mol. Biol. 124 (1978) 523.

8. D.C. Phillips, Proc. Nat. Acad. Sci., U.S.A., 57 (1967) 484.

9. M. Shoham, A. Yonath, J.L. Sussman, J. Moult, W. Traub and A.J. Kalb (Gilboa), J. Mol. Biol. 131 (1979) 137.

EXPERIENCE WITH THE APPLICATION OF PATTERSON SEARCH TECHNIQUES

by

Robert Huber
Max-Planck-Institut fuer Biochemie, D-8033 Martinsried bei Muenchen

Hoppe discovered that the Patterson function of a crystal composed of discrete molecules can be decomposed into molecular functions, which he called Faltmolekuele[1,2]. There are two different types of Faltmolekuele, those with equal indices and those with mixed indices. The first type represents the intra-molecular vector set. It is centered at the origin of the unit cell and has a structure and orientation which is uniquely derived from the molecular structure and orientation. Consequently it may serve to determine the orientation of the molecule in the crystal cell if the molecular structure is known. The second type represents the inter-molecular vector sets between molecules related by the crystal symmetry. The structure of these Faltmolekuele depends only on the molecular structure and orientation. They are centered at the difference vectors of the two molecules they are derived from. Consequently, they can be constructed if the molecular structures and orientations are known and may then serve to determine the translation relative to the crystal symmetry elements.

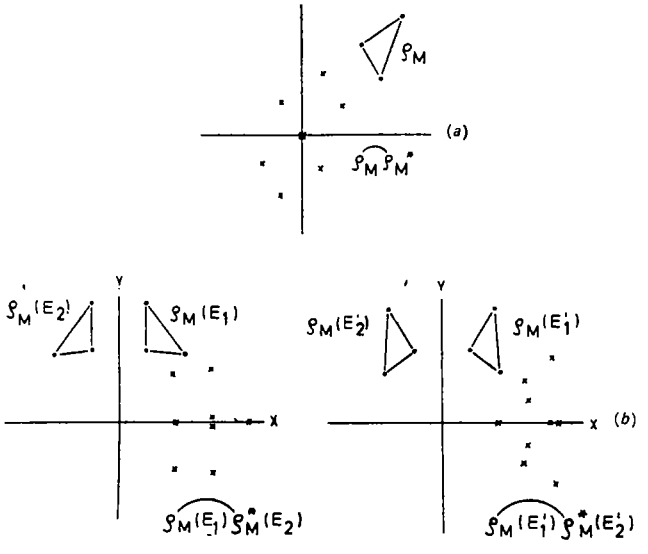


Fig.1 Faltmolekuele with equal indices derived from a triangular structure (ρ_M). A number of properties are immediately obvious: the internal structure of the Faltmolekuele $\widehat{\rho_M\rho_M}^*$ (the intra-molecular vector set) is evidently determined by the molecular structure. It is centered at the origin of the coordinate system and is independent of translation. Its orientation follows from the orientation of ρ_M.

Faltmolekuele were constructed graphically before computer programmes were available and rotated or slid over Patterson projections to observe the optimal fit. A number of crystal structures of organic molecules were solved in this way[3-9].

In the lower half of fig.1 ρ_M in orientation E_1 is repeated by a diad along Y and the mixed indexed Faltmolekuele $\widehat{\rho_M(E_1)\rho_M}^*(E_2)$ constructed. E_2 follows from E_1 by the diad symmetry operation. It is obvious that the Faltmolekuele is located on the Harker line Y = 0 at a distance 2X from the diad. Its internal structure depends on both, the molecular structure and orientation, as can be seen when $\widehat{\rho_M(E_1)\rho_M}^*(E_2)$ and $\widehat{\rho_M(E_1')\rho_M}^*(E_2')$ are compared. Equation (1) formulates the relations:

$$\rho = \sum_{k}^{K} \rho_M(E_k, r_k);$$

$$P = \widehat{\rho\rho}^* = \sum_{k=1}^{K} \widehat{\rho_M\rho_M}^*(E_k, 0) +$$

$$\sum_{k\neq k'}^{K} \sum_{k'=1}^{K} \widehat{\rho_M(E_k)\rho_M}^*(E_{k'})[r_k - r_{k'}]$$

ρ_M is the electron density of an individual molecule in arbitrary orientation and translation. Its orientation and translation in the crystal cell is defined by E_k and r_k respectively. The summation over the K molecules in general equivalent positions of the unit cell yields ρ. * denotes the operation of inversion.

The necessary sequence of operations to position a known molecule in a crystal cell is therefore to determine the orientation by correlating the Patterson function with the intra-molecular vector set. Thereafter the cross-vector sets can be constructed and the translation parameters relative to the symmetry elements determined by correlating the Patterson function and cross-vector sets. The Faltmolekuele method was programmed by Huber[10,11] and Nordman and Nakatsu[12] and different criteria for the correlation between the Patterson function and vector set were used.

Rossmann and Blow[13] discovered that the product correlation between the Patterson function and self-vector sets can be expressed in reciprocal space. They also recognized that internal symmetry in the asymmetric unit ρ_M is reflected in the Patterson map from which it can be derived by a rotational self-correlation function of the Patterson function. Figure 2 shows a motif with internal symmetry. ρ_M

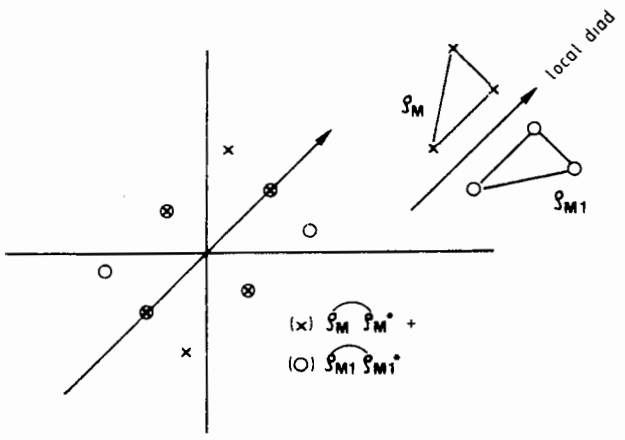


Fig. 2

is related to ρ_{M1} by a diad axis of symmetry. The self-vector sets of ρ_M and ρ_{M1} are related by the rotational part of the symmetry operator relating ρ_M and ρ_{M1}. It can be analysed with the Rotation function by calculating the self-correlation of the Patterson function with respect to three angular variables describing the rotational symmetry operator. The self-correlation calculation must be restricted to a certain radius around the origin of the Patterson map to make the number of inter-molecular vectors small as they do not obey these symmetry relations. The truncation requires a convolution with the Fourier transform of the shape function[13,14].

The translation function expresses the correlation between the Patterson function and cross-vector sets in reciprocal space. It calculates the product correlation and was formulated by Crowther and Blow[15] in an elegant way as a Fourier series.

In correlation calculations between the Patterson function and Faltmolekuele in real space a variety of criteria of fit can be used[10,12] but do not seem to be superior to the product correlation. The product correlation has the advantage of being insensitive to the relative scaling of the Patterson function and Faltmolekuele. The exclusion of parts of the Patterson map and Faltmolekuele is very simple in real space, an advantage over reciprocal space methods.

The angular correlation calculation may be performed in various systems. The Eulerian angular system allows the symmetry properties of the Patterson map and Faltmolekuele to be taken into account by restricting the required minimum scan range. The space group dependent symmetry properties have been analysed in a lucid way by Rossmann and Blow[13] and Tollin, Main and Rossmann[16]. The polar angular system is convenient for the search of given rotation axes by two-dimensional scans.

The following series of references documents our experiences with Patterson search techniques: the programmes employed were almost always the real space orientation correlation in the Protein programme system[17] and the translation function programme of Lattman modified by Deisenhofer and Huber. The examples analysed were proteins and protein protein complexes of various sizes ranging from 5 to 100 kD. The models used in the searches stemmed from crystal structures of the same molecules in different crystal forms or from closely related, homologous variants or from components of protein complexes. They comprised various proportions of the asymmetric unit from 100 to 30 percent and were of various degrees of refinement. In the following sections I shall not repeat details described in the original publications but try to summarize and generalize various aspects.

A. Calculation Procedures

The model Patterson map was usually calculated from the atomic model in a cubic box with cell edges twice the longest molecular dimension, to isolate the intra-molecular vector set. For the orientation correlation calculation a variable number of maxima were picked from the model Patterson function to be compared with the crystal Patterson function. Inner and outer radii of 5 Å and 15 to 30 Å respectively, were applied. The outer radius was chosen small for the initial rough scan and increased later in the fine scans. The number of maxima extracted from the model Patterson map was also varied from about 2000 to 8000 depending on the problem. The signal to noise ratio increases with the number of peaks used. In the citrate synthase oxaloacetate complex the use of a very large number of peaks was a prerequisite

for the correct solution[18]. The resolution applied was usually to the limit of the available data set but not beyond 2.5 Å.

This orientation correlation function did not always result in an outstandingly high signal of the correct solution. In doubtful cases the highest peaks in the correlation function were tested by the translation function.

The translational functions were usually calculated at the resolution allowed by the experimental data set. The molecular Fourier transform was generated from the molecular coordinates placed in a cubic unit cell with dimensions twice the largest diameter of the molecule. The translation function programme allows orientation of the Fourier transform according to the orientational angles determined in the previous step. Translation functions corresponding to all inter-molecular vectors generated by the space group symmetry were calculated. The multiple determination of translation components in space groups of higher symmetry is most welcome. Discrimination of enantiomorphous space groups is possible by calculating the appropriate translation function. As the orientational symmetry operator is identical in the two possible space groups the translation vector occurs in different sections of the same three-dimensional translation function map[18-21].

The orientational and translational parameters obtained by Patterson search techniques were refined before refinement of the atomic model was initiated. Appropriate procedures are rigid body least squares refinement procedures in reciprocal space[22,23].

B. Properties of the Search Model

It is obvious that the signal of the correlation functions increases with the similarity of search and target molecules. A search model from a refined crystal structure is advantageous. The crystal structure analyses of chymotrypsinogen using models of various refinement stages demonstrates this aspect[24]. We note however that identical molecules crystallizing in different crystal lattices may show deviations of about 0.4 Å in C^{α} atom positions as revealed by the very high resolution analyses of two forms of PTI[25,26]. These deformations are due to the crystal lattice packing forces.

Most of the successful examples referred to in Refs.(18-21, 26, 27-29) had r.m.s. deviations between search models and final refined crystal structure models of about 0.5 to 1 Å. Models with insufficient similarity, yet clearly recognizeable amino acid sequence homology, did not yield a significant solution. The trypsin model was insufficient to find kallikrein, a related protease[30].

C. Size of the Search Model

It is clear that the signal of the correlation calculations increases with the proportion of the search model in the asymmetric unit. The search model may be part of a complex of different proteins or may be identically repeated in the asymmetric unit. In the latter case analysis of the self rotation function is useful as it places constraints on the possible orientations of the independent molecules[31]. In our experience a search with one half of the asymmetric unit in the structure analysis of chymotrypsinogen proceeded without problems[24]. Orientation and translation searches were performed independently for the two molecules. A cross-translation function solved the origin problem. In crystals of C-phycocyanin of Agmenellum quadruplicatum[32] however three $\alpha\beta$ units of known structure[33] build up the asymmetric unit. The search model comprises one third of the asymmetric unit. A Patterson search analysis was successful only because certain plausible assumptions about orientation and position of the molecules could be made[23].

REFERENCES

1. W. Hoppe, Acta Cryst. 10 (1957) 750.
2. W. Hoppe, Z. Elektrochem. 61 (1957) 1076.
3. W. Hoppe, Z. Krist. 117 (1962) 4.
4. W. Hoppe, Z. Krist. 114 (1960) 393.
5. W. Hoppe and G. Will, Z. Krist. 113 (1960) 104.
6. W. Hoppe and R. Rauch, Z. Krist. 115 (1961) 141.

7. W. Hoppe, Z. Krist. 117 (1962) 4.

8. G. Will, Z. Krist. 119 (1963) 1.

9. R. Huber and W. Hoppe, Chemische Berichte 98 (1965) 2403.

10. R. Huber, Acta Cryst. 19 (1965) 353.

11. R. Huber in Crystallographic Computing Proceedings of the 1969 International Summer School on Crystallographic Computing, ed. by F.R. Ahmed (Munksgaard, 1969).

12 C.E. Nordman and K. Nakatsu, J. Amer. Chem. Soc. 85 (1963) 353.

13. M.G. Rossmann and D.M. Blow, Acta Cryst. 15 (1962) 24.

14. E.E. Lattman and W.E. Love, Acta Cryst. B26 (1970) 1854.

15. R.A. Crowther and D.M. Blow, Acta Cryst. 23 (1967) 544.

16. P. Tollin, P. Main and M.G. Rossmann, Acta Cryst. 20 (1966) 404.

17. W. Steigemann, Ph.D. Thesis, Technische Universitaet, Muenchen (1974).

18. G. Wiegand, S. Remington, J. Deisenhofer and R. Huber, J. Mol. Biol. 174 (1984) 205.

19. W. Bode, H. Fehlhammer and R. Huber, J. Mol. Biol. 106 (1976) 325.

20. H. Fehlhammer, W. Bode and R. Huber, J. Mol. Biol. 111 (1977) 414.

21. Z. Chen and W. Bode, J. Mol. Biol. 164 (1983) 283.

22. J.L. Sussman, S.R. Holbrook, G.M. Church and S. Kim, Acta Cryst. A33 (1977) 800.

23. R. Huber and M. Schneider, J. Appl. Cryst. (1985), in press.

24. D. Wang, W. Bode and R. Huber, J. Mol. Biol. (1985), in press.

25. J. Walter and R. Huber, J. Mol. Biol. 167 (1983) 911.

26. A. Wlodawer, J. Walter, R. Huber and L. Sjoelin, J. Mol. Biol. 180 (1984) 301.

27. W. Bode, O. Epp, R. Huber, M. Laskowski, Jr. and W. Ardelt, Eur. J. Biochem. 147 (1985) 387.

28. M. Bolognesi, G. Gatti, E. Menegatti, M. Guarneri, M. Marquart, E. Papamokos and R. Huber, J. Mol. Biol. 162 (1982) 839.

29. M. Marquart, J. Walter, J. Deisenhofer, W. Bode and R. Huber, Acta Cryst. B39 (1983) 480.

30. W. Bode, Z. Chen, K. Bartels, C. Kutzbach, G. Schmidt-Kastner and H. Bartunik, J. Mol. Biol. 164 (1983) 237.

31. M. Buehner, H.-J. Hecht, R. Hensel and U. Mayr, J. Mol. Biol. 162 (1982) 819.

32. M.L. Hackert, C. Abad-Zapatero, S.E. Stevens Jr. and J.L. Fox, J. Mol. Biol. 111 (1977) 365.

33. T. Schirmer, W. Bode, R. Huber, W. Sidler and H. Zuber, J. Mol. Biol. (1985), in press.

REAL SPACE VS. RECIPROCAL

by

M. Buehner* and H.J. Hecht
A.G. Roentgenstrukturanalyse, Universitaet Wuerzburg, Am Hubland, D-8700 Wuerzburg, Fed. Rep. Germany

1. INTRODUCTION

Most components of the complex procedure called structure analysis are rigidly associated with a given space, either real or reciprocal, the transition between both spaces being afforded by the Fourier transformation. In the subset of molecular replacement procedures, however, there are some components which can be adapted to execution in either space. This symmetry of spaces in theory, and the limitations of either space for practically applicable algorithms, have been clearly spelt out i.a. by G. Bricogne[1]. Although the theoretical equivalence of spaces stands undisputed, real world conditions such as running time and memory availability of computers (or their limitation by computer centre management and/or fees) require shortcuts, simplifications, limitation of series etc., which in practical application lead to advantages and disadvantages of one space over the other for a given set of experimental conditions. Therefore, in real life, the question is often not whether to use real space or reciprocal, but rather real space and reciprocal, whichever is best suited.

In Wuerzburg we have not done a systematic comparison of spaces for molecular replacement algorithms, but we have encountered a few cases where a comparison could be done during some protein structure projects. Both projects described here are being done by molecular replacement exclusively, i.e. without any independent phase information from isomorphous replacement, so that phase combination is not an issue of this paper.

2. THE PROGRAMMES

DENS expands atomic co-ordinates into a density map for structure factor calculation using FFT.
ROTFUN is a modification of M.G. Rossmann's reciprocal space rotation function programme[2].
TRAFUN is a short translation function programme[3] which simply combines model structure factors with F-obs into translation function coefficients. These coefficients are subsequently transformed into the translation function map by FFT.
TRALS is a translations-only least squares procedure. Translations are applied as phase shift factors and thus no Fourier transformation is required. Therefore this programme is very fast, allowing extensive translational searches with "zeroing in" by least squares at each point.
RIGID is a complete rigid body least squares procedure, refining all 6 degrees of freedom. It is very time consuming so that in our case its application is limited by the availability of computing funds (or lack thereof).
DISPRF (DIrect SPace Rotation Function) was written using the angle/matrix framework of ROTFUN. It works on Patterson maps (3 rotational parameters) as well as on electron density maps (3 rotational + 3 translational parameters). In its present stage it can deal with proper single rotations and 222-systems ("locked" rotation function). Improper rotations can be simulated by using the cross-rotation mode between identical data sets and employing the translation parameters to do the shifting along the axis.
AVERAG is a programme which allows averaging[4,5] the density of subunits grouped about 1 or more origins. Either it uses a "labelled" mask (where each legal density point is marked by the number of its origin), or it works within spheres of a given radius. Where spheres overlap the maximum function is used for selecting the output density, following the experience that density parts belonging to the currently used origin come out strong whereas parts belonging to neighbouring origins are chopped into noise, because the local symmetry operator used does not hold for them.
MASK uses monomeric model density provided by DENS and puts it into the natural unit cell to make a labelled mask which describes the molecular volume about each non-crystallographic origin.
All other programmes used are adapted standard products from various sources.

*Paper presented by M. Buehner.

The programmes mentioned above were designed for our university's present central computer (Telefunken TR 440) which is a slow (~VAX/750) non-virtual machine with limited memory, lacking a good and fast sort/merge package, but equipped with very good random access file handling. The real space programmes DISP, AVERAG and MASK do not use peak lists or the double sorting technique, but work on scratch density maps. A scratch map consists of a sphere (or spheres) containing the coherent density around an origin, so that no symmetry operations or cell translations need be applied. Scratch density is held in paginated sections in random access files. The look-up scratch map is mirrored into a large array by a procedure that mimicks virtual addressing.

All programmes are available for distribution, please contact M.B.

3. APPLICATION

Table 1 attempts to survey the application of the different methods in different spaces, as to our present knowledge. It might seem odd that "Films" (i.e. visual inspection of precession photographs) is assigned a column of its own in the table, but the power of immediate film analysis should by no means be underestimated. Its strength is not the accurate quantitative evaluation of rotational or translational parameters, but besides giving an overall picture of the situation it can impose crucial constraints on the range of validity and on the interpretation of the results of other methods. Pseudo systematic absences e.g. can be the most reliable translation information at all since they are independent of any model.

Both structures mentioned below have non-crystallographic symmetry, but they differ in complexity and size of the asymmetric unit. For both projects model co-ordinates were available. Uteroglobin had been solved in a different crystal form (monomer in the asymmetric unit) by J.P. Mornon in Paris, with whom we are co-operating, and LDH co-ordinates were kindly provided by M.G. Rossmann (pig heart LDH ternary complex) or could be obtained from the protein data bank (dogfish muscle apo LDH).

3.1 Uteroglobin from Rabbit

Uteroglobin is a small steroid binding protein consisting of 2 identical polypeptide chains of 70 amino acids. It crystallises in many different crystal forms what might be an indication of flexibility or conformational variability. We have been looking at a crystal form in space group $P\,2_1$ ($a = 43.3$ Å, $b = 38.1$ Å, $c = 34.5$ Å, $\beta = 90.7^{\circ}$) with a dimer of molecular weight $2 \times 7{,}900$ in the asymmetric unit. The crystals contain 67 % protein by volume[6].

This crystal structure is an example for the power of looking at films: a) The monoclinic angle ß is close to a right angle, b) an upper layer precession photograph (1kℓ) at 2.5 Å resolution showed only a handful of deviations from exact mm-symmetry[6] although only 1 mirror plane is required by crystallographic symmetry, c) an hOℓ precession photograph revealed good systematic absences on the hOO-line out to about 4 - 5 Å resolution whereas the OOℓ-line did not display anything special[6]. The general intensity distribution on all layers recorded exhibi-

Table 1

Molecular replacement methods as used in reciprocal and real space. Capitals indicate names of programmes.

Method	Reciprocal Space			Real Space	
	Films	Fourier	Bessel	Patterson	Density
Rotation search	spikes etc.	ROTFUN	+	DISPRF	-
Translation search	ps. syst. abs.	TRAFUN	-	+	packing
l. sq.	-	TRALS	-	-	-
Rot. + Tra. search	-	+	-	-	DISPRF
l. sq.	-	RIGID	-	-	+
Phase Improvement	-	?	-	-	AVERAG

ted an mm-symmetry of quite reasonable quality.

These observations from films suggest that there are non-crystallographic 2-fold axes roughly in the direction of the a- and c-axes, and that the axis parallel c is most likely the molecular dimer axis since it appears to be a proper rotation. Furthermore, this molecular dyad and the centre of the molecule must be close to x = 1/4 and x = 3/4, to produce a translation by a/2 required by the pseudo systematic absences on h00. Thus the situation is close to P 2_1's supergroup P $2_1 2_1 2$, and only the angle of rotation of the molecule about its axis and the position of its centre of gravity along this axis (z) are really unknown.

A general search for dyads using programme ROTFUN indeed revealed a high peak (79 % of origin) at the polar angles $\psi = 90.0^\circ$ and $\varphi = 0.4^\circ$, i.e. midway between a^* and a. Due to its special position on the a/c-plane this peak is mirrored at $\psi = 90.0^\circ$ and $\varphi = 90.4^\circ$ (i.e. between $-c^*$ and -c) by symmetry imposed by the crystallographic b-axis. Model co-ordinates from Mornon's structure determination were then used to calculate the cross-rotation function ir reciprocal space (ROTFUN). The whole asymmetric unit of Eulerian space was searched although it would have sufficed just to superimpose the model's dimer axis on the vector at $\psi = 90.0^\circ$ and $\varphi = 90.4^\circ$. The map was essentially flat except for the molecular axes' superposition.

When searching for a dimer using a monomeric model one would expect 2 solutions 180° apart. What we did find, however, were 4 peaks of about equal height roughly 90° apart. So we found 2 equally possible dimers instead of one. One of these solutions must be spurious, created by packing. Inter-molecular vectors must be expected to play a major role in this tightly packed unit cell. To resolve the ambiguity we resorted to cross-rotation in Patterson space. Unfortunately, the symmetry of spaces was proved beyond doubt by the equal failure of both methods to discriminate between the 2 solutions.

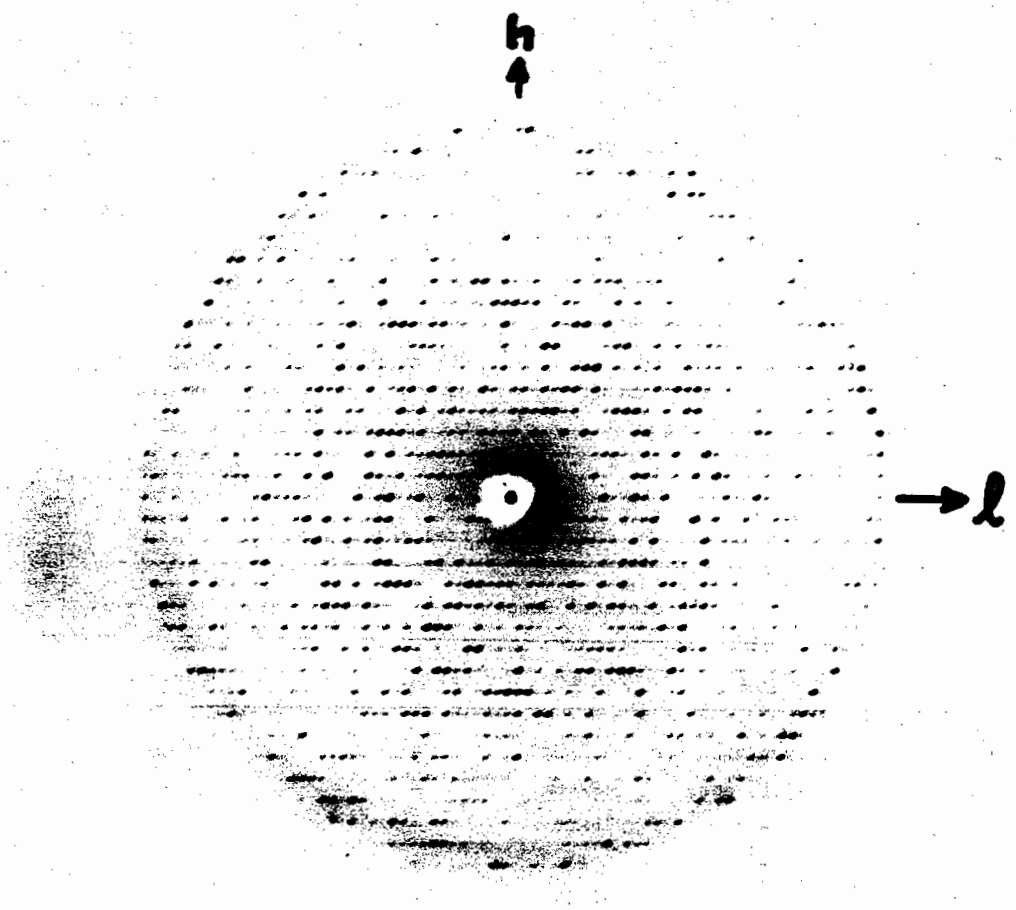


Fig.1 Precession photograph of the h0ℓ reciprocal lattice plane of L. casei LDH. Cu K_α, prec. angle = 9.0°.

Obviously the model is not good enough to successfully tackle the problem. The conformation of the protein must be quite different in both crystal forms. Sure enough, under those conditions the translation function did not work properly either, and rigid body calculations were a mere sink for computing money. Packing calculations narrowed down the possible range of the position of the molecule to within a couple of Ångstroems, but the result came out the same for both alternatives.

The best way out of the dilemma seemed to be to use the redundancy of non-crystallographic symmetry. The more likely looking solution[7] was applied after rigid body refinement. A $2F_o-F_c$ electron density map was subjected to cyclic averaging: After each cycle, structure factors were calculated from the averaged density, a $2F_o-F_c$ map was calculated, and the non-cryst. symmetry operators were readjusted using the new map. After 6 cycles convergence was about reached at R = 19.3 % (10.0 - 2.5 Å res.) and the map was inspected by model building. The density was very clean, but no significant modification of the model was apparently needed, much to our surprise. It turned out that the whole averaging procedure had done nothing more than ridding the faulty model of its noise without removing its systematic errors. This indicates clearly that averaging under the prevailing conditions (only 2 copies, parameters very close to a crystallographic supergroup) did not exert sufficient power to improve the model.

The lesson to be learned here is that being close to a supergroup, i.e. non-crystallographic symmetry mimicking crystallographic symmetry, diminishes, or even destroys, the power of the averaging process. The reason is that a data set sampling the molecular transform repeatedly at about the same grid points does not contain really independent redundant information. Only if a non-crystallographic symmetry operator is sufficiently distant from systematic crystallographic directions to sample the molecular transform at truly different points is the additionally gained information independent and the averaging process can exert its power by using true redundancy.

3.2 Lactate Dehydrogenase from Lactobacillus casei

This bacterial LDH is regulated by allosteric effec-

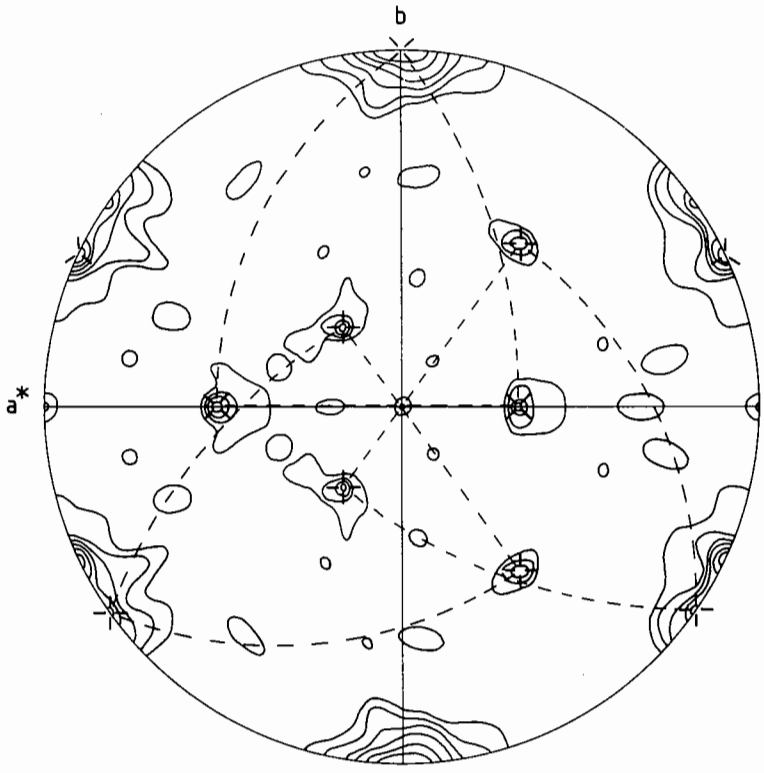


Fig.2 Stereographic projection of the self-rotation function of L. casei LDH (search for dyads).

tors, Fru-1,6-P_2 and Mn^{2+}, in contrast to vertebrate isoenzymes. The molecule is a tetramer consisting of 4 identical polypeptide chains of 325 amino acids. It crystallises in space group C 2 (a = 169.20 Å, b = 85.35 Å, c = 180.18 Å, β = 91.3°). The unit cell contains 6 tetrameric molecules, i.e. there are 1 1/2 tetramers (mol. weight 6 x 35.440) in the asymmetric unit[8], which is not quite expected for tetramers in a monoclinic space group. The protein content of the crystals is 43 % by volume.

At low resolution 00ℓ-reflexions are absent or very weak for ℓ ≠ 3n (fig.1). This shows clearly that a prominent translation vector of size c/3 exists and the cell content therefore must be a multiple of 3. A general search for dyads in reciprocal space (ROTFUN) solved the riddle by revealing (fig.2) one 222-system of axes in a special position, including the crystallographic b-axis, and an unrestricted second 222-system plus its symmetry equivalent. This suggests a central tetramer on a crystallographic dyad contributing a dimer to the asymmetric unit and a second, peripheral, tetramer in a general position at z = ± 1/3, making the asymmetric unit 6 subunits = 1 1/2 tetramers. One dyad of the peripheral tetramer is close to the a/b-plane at ψ = 126°. The total arrangement of dyads together with the systematic absences makes c an approximate 3_1-axis. Therefore, the arrangement of molecules in the unit cell resembles strongly the space goup P $3_1$21 which is a supergroup to C 2.

All axes were identified in terms of the molecular system P/Q/R by cross-rotation function with the model of dogfish apo-LDH, again in reciprocal space. The answers were unambiguous in all cases, showing that the model was appropriate. The dyads were then checked in Patterson space, with deviations within the limits of error (± 2°).

The position of the central tetramer can be set to 0/0/0 and only the origin of the peripheral tetramer need be determined. First, the translation function was used in the classical way, to determine the distance between symmetry related components. Thus we found 2x and 2z for the peripheral tetramer, but not y. The translation function was then modified to accept 2 independent models instead of a pair of symmetry related components and was used to determine the vector between the central and peripheral tetramers (fig.3). The answer was unambiguous at .270/.145/.672. However, only the z-range between .13 and .91 of the translation maps could be used. The part close to the origin was plagued by high noise, obviously produced by self vectors. The noisy part of the map could safely be ignored, however, because the systematic absences permitted only small bands of z-values about 1/3 and 2/3. Packing calculations confirmed the solution, but, as in the case of uteroglobin, packing alone could not have resolved the ambiguity between 1/3 and 2/3. The translation parameters were then least squares refined, and finally rotation and translation parameters were further im-

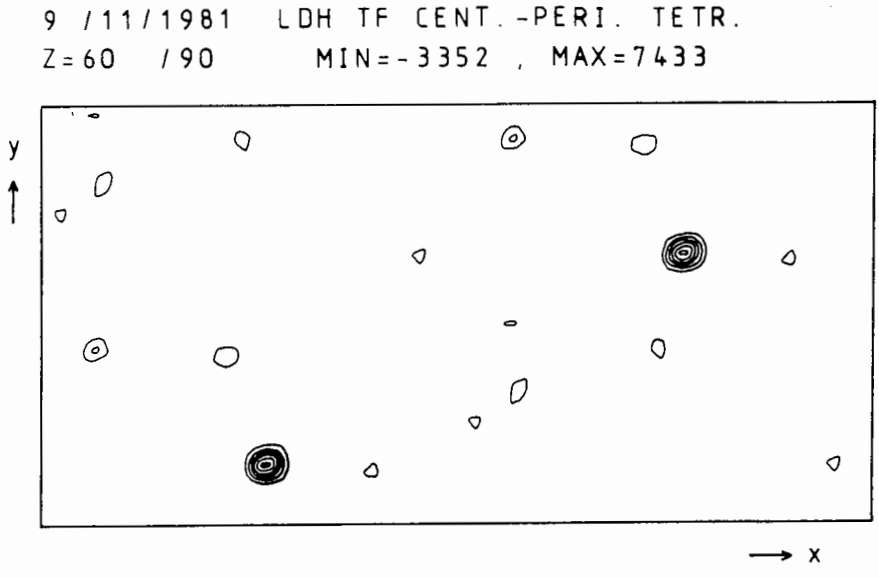


Fig.3 Translation function of L. casei LDH, section at z=2/3. Peaks appear twice due to the C-centred cell.

Table 2

Statistics of calculated structure factors during averaging of LDH.

	R (%)	Correlation coefficient	1. Model	Ave 1	Ave 2	Ave 3
1. Model	48.06	.1839	-			
Ave 1	41.97	.4604	49.6	-		
Ave 2	34.66	.5372	54.8	20.2	-	
Ave 3	33.22	.5912	57.9	25.7	11.6	-
Ave 4	29.95	.6426	60.3	33.4	21.0	15.2

R and the correlation coefficient were calculated in the resolution range of 10.0 - 3.0 Å.

$$R = \frac{\Sigma \left| |F_o| - |F_c| \right|}{\Sigma |F_o|} \cdot 100, \qquad \text{Correl. coeff.} = \frac{n\Sigma(|F_o| \cdot |F_c|) - \Sigma|F_o| \cdot \Sigma|F_c|}{\{[n\Sigma(|F_o|^2) - (\Sigma|F_o|)^2] \cdot [n\Sigma(|F_c|^2) - (\Sigma|F_c|)^2]\}^{1/2}}$$

proved by rigid body least squares. A $2F_o-F_c$ map was interpreted by model building and the amino acid sequence was changed from the model (pig heart LDH) to the L. casei LDH sequence. Then we could proceed by cyclic averaging:

All 6 independent copies of the LDH subunit were averaged. In cycles 1 - 3 straight $2F_o-F_c$ maps were used without weighting. Reflexions were excluded by resolution limits (10.0 - 3.0 Å) only, the maximum molecular radius used was 42 Å. After cycle 3 the procedure was changed towards molecular radius 45 Å, using Sim-weighted $2F_o-F_c$ maps and rejecting reflexions if $2\left||F_o|-|F_c|\right| / |F_o|+|F_c| > 1.2$. The results are outlined in table 2.

After model building and after each cycle of averaging the non-crystallographic symmetry operators were optimised in the $2F_o-F_c$ density using DISPRF. In table 3 the orientation of the molecular dyads is compared for all steps in the molecular replacement procedure, and in table 4 the same is done for the position of the peripheral tetramer (the central tetramer was always held at 0./0./0.).

The average-4 map could not be checked yet at the interactive display, but the average-3 map still contains spots of bad density and shows traces of the model. We have been able to identify and tentatively orient the allosteric effector fructose-1,6-bis-phosphate in its binding site, but a Co^{2+} ion (Co was used instead of Mn for reasons of crystal stability and scattering power) could not yet be found. Obviously, even with the power of 6-fold averaging it is not trivial to get rid of all influences of the primary model.

Table 3

Orientation of all molecular 2-fold axes of LDH (in spherical polar angles Ψ / φ).

		Central Tetramer			Peripheral Tetramer		
		1 (R)	2 (Q)	3 (-P)	4 (R)	5 (Q)	6 (-P)
Rec. Space	6 Å	0./0.	90./36.0	90./126.0	126.6/-1.0	133.8/133.2	65.0/69.0
Patt. Space	3 Å	0./0.	90./34.70	90./124.70	126.42/-1.72	133.47/132.65	65.33/68.47
Rigid Body	6 Å	1.1/-98.8	89.36/33.08	89.22/123.09	126.63/-.67	131.17/129.90	62.68/66.76
1. Model	3 Å	0./0.	90./33.55	90./123.55	126.70/-.35	131.19/130.36	62.78/67.11
Ave 1		0./0.	90./33.58	90./123.58	126.68/-.47	131.29/130.38	62.88/67.10
Ave 2		0./0.	90./33.64	90./123.64	126.64/-.49	131.34/130.38	62.89/67.14
Ave 3		0./0.	90./33.65	90./123.65	126.66/-.52	131.37/130.44	62.96/67.15
Ave 4		0./0.	90./33.70	90./123.70	126.68/-.52	131.39/130.50	63.00/67.17

Table 4

Position of the peripheral LDH tetramer (in cell fractions).

		x	y	z
Translation Function	6 Å	.270	.145	.672
Translation L. Sq.	6 Å	.2698	.1437	.6732
Rigid Body L. Sq.	6 Å	.26776	.14172	.67543
1. Model	3 Å	.2675	.1386	.6740
Ave 1		.2677	.1387	.6741
Ave 2		.2678	.1387	.6742
Ave 3		.26782	.13862	.67424
Ave 4		.26794	.13860	.67420

4. CONCLUSIONS

As far as the reliability of rotation results is concerned, the data in table 3 seem to imply that the Patterson method is doing slightly better than reciprocal space, but the difference is not really significant. In our view there remain 2 differences between real and reciprocal space: 1) Reciprocal space programmes work faster than Patterson space programmes (Rossmann's by a factor of 3-5, Crowther's much faster still), 2) Patterson rotation peaks are smoother than reciprocal space peaks. Therefore, reciprocal space offers itself for general surveys whereas Patterson seems more suited for fine tuning and for compromises if peaks appear split or rugged in reciprocal space.

For translation purposes, real space (i.e. packing) is no match for reciprocal space, because it uses shape information only and not internal structure. It is, however, very valuable for excluding large parts of (often quite noisy) translation function maps. Therefore, translation information from both real and reciprocal space should always be combined. Pseudo systematic absences on precession films are often the most reliable translation information at all.

Once parameters are roughly established they should first be refined by rigid body procedures. For further refinement direct space rotation/translation search seems superior to R-search, because it allows to work in genuine electron density, while R-search and rigid body least squares depend on model co-ordinates with all their shortcomings. Density, on the other hand, can be shifted in the direction of the correct structure by either the $nF_o+m(F_o-F_c)$ - technique or, even more, by averaging and recycling the averaged density.

For phase improvement by use of the redundant information from multiple molecular copies there is as yet no alternative to real space (density averaging [5]). Although reciprocal space methods are still being discussed[9] the stage of a really applicable algorithm or even working computer code does not seem to have been reached yet. Whatever the approach, one caveat seems advised: The power of the method depends strongly on the redundant information being really independent through non-identical sampling of the molecular transform.

ACKNOWLEDGEMENTS

All model building was done at the graphics facilities of the Max-Planck-Institut fuer Biochemie at Martinsried, made available to us by R. Huber. We thank Prof. Huber for his continuing generous support.

REFERENCES

1. G. Bricogne in "Computational Crystallography" (D. Sayre ed.), (Oxford: Clarendon Press, 1982), p. 258.

2. M.G. Rossmann and D.M. Blow, Acta Cryst. 15, (1962) 24.

3. R.A. Crowther and D.M. Blow, Acta Cryst. 23, (1967) 544.

4. M. Buehner, G.C. Ford, D. Moras, K.W. Olsen and M.G. Rossmann, J. Mol. Biol. 82, (1974) 563.

5. G. Bricogne, Acta Cryst. A 30, (1974) 395.

6. M. Buehner and M. Beato, J. Mol. Biol. 120, (1978) 337.

7. M. Buehner, A. Lifchitz, R. Bally and J.P. Mornon, J. Mol. Biol. 159, (1982) 353.

8. M. Buehner, H.J. Hecht, R. Hensel and U. Mayr, J. Mol. Biol. 162, (1982) 819.

9. M.G. Rossmann in "Computational Crystallography" (D. Sayre ed.), (Oxford: Clarendon Press, 1982), p. 159.

ROTATION FUNCTIONS, DENSITY AVERAGING AND PHASE EXTENSION: PANULIRUS INTERRUPTUS HAEMOCYANIN

by

Wim G.J. Hol, Anne Volbeda and Wil P.J. Gaykema
Laboratory of Chemical Physics, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands

1. INTRODUCTION

Haemocyanins are the copper-containing, non-heme, oxygen-transport molecules occurring in a large number of invertebrates. The cylindrical molluscan haemocyanins differ completely in architecture from the hexameric or multihexameric arthropodan haemocyanins[1,2]. The structure of the single-hexameric haemocyanin from the spiny lobster Panulirus interruptus, an arthropod, has been solved by isomorphous and molecular replacement methods[3-8]. A number of characteristics of this haemocyanin are summarized in Table 1, whereas Table 2 lists some of the details of the structure determination. Here we will focus on the role which molecular replacement methods played in the course of the determination of a structure where the entire hexamer with MW ~ 470.000 occurred in the asymmetric unit.

Table 1

Characteristics of
Panulirus interruptus haemocyanin

MW	6 x 77.000
Residues per subunit	657
Copper ions per subunit	2
Carbohydrate moieties per subunit	1
No. of disulfide bridges per subunit	3
Subunit types native Hc	*a*, *b* & *c*
Amino acid sequence difference *a* vs *b*	~ 3%

2. STUDIES AT 10 Å RESOLUTION: THE SYMMETRY OF THE MOLECULE

Using 32 precession films a set of 1440 structure factors out to 10 Å resolution was obtained[4]. Both the conventional Rossmann & Blow rotation function[9] and Crowther's fast rotation function[10] were applied and gave results which were in excellent agreement with each other[4]. Some results of the conventional rotation function are shown in Figure 1. In these calculations data between 10 - 25 Å and

Table 2

Structure Determination of
Panulirus interruptus haemocyanin

pH	4.5
Buffer	0.01 M acetate
Temperature	4°C
Space group	$P2_1$
Cell dimensions	119.8 x 193.1 x 122.2 Å; β=118.1°
Daltons per asymmetric unit	~ 470.000
Subunit types in crystal	*a* & *b*
No. of sites of Pt-derivative	36
No. of sites of Hg-derivative	70
No. of refls with MIR phases	32721
No. of refls with only MR phases	31121
Total no. of phased reflections	63842

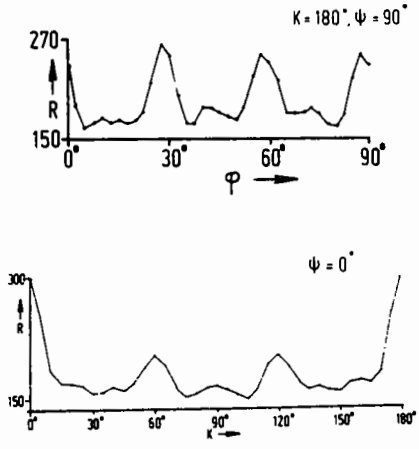


Fig. 1 Rotation function studies with 10 Å resolution data. Upper Figure: search for two-fold non-crystallographic axes perpendicular to the b-axis. Lower Figure: search for rotation axes parallel to the b-axis. Angular parameters as defined in Rossmann and Blow[9].

a radius of 45 Å were used. Reducing the radius to 25 Å had hardly any effect on the calculations. Changing the resolution range to 10 - 12.5 Å did not give much different results. Varying the number of large terms between 109 and 427 had little effect on the appearance of the rotation function.

As the polar rotation axes were defined as in the paper by Rossmann & Blow[9], the results of Figure 1 indicate that the three-fold axis of the hexamer runs approximately parallel to the b-axis whereas the molecular two-fold axes run nearly parallel to the a- and c-axis and to the bisectrice of these two axes. The point group symmetry of the molecule is 32.

3. RESULTS AT 5 Å RESOLUTION: THE SHAPE OF THE HEXAMER AND THE ACCURACY OF ROTATION FUNCTION RESULTS

With a 5 Å resolution data set, obtained by the oscillation method[4], the conventional rotation function was used to establish the direction of the molecular symmetry elements with greater precision. As depicted in Figure 2, each two-fold rotation (2_{intra}) relating vector sets of subunits within one hexamer, is accompanied by a pseudo two-fold rotation (2_{inter}) which relates vector sets of subunits within two different hexamers. These two non-crystallographic dyads are perpendicular to each other and are related as follows:

$$\begin{aligned} 2_{intra}&: \kappa = 180° \quad , \phi \quad , \psi = 90-\Delta \\ 2_{inter}&: \kappa = 180-2\Delta, \ \phi+90, \ \psi = 90 \end{aligned} \qquad (1)$$

neglecting the 2_{inter} rotation with $\kappa = 180+2\Delta$ for simplicity. Calculations with small angular increments showed that at $\phi = -1°$, 50° and 120° maxima occur for $\kappa \neq 180°$ (Fig. 3). Consequently, the first set of rotation operations are pseudo-twofolds relating different hexamers and the second set comprises the true dyads relating subunits within one hexamer. Inspection of the positions of the peaks in Figure 3 shows that the angular relationships of equation (1) hold well.

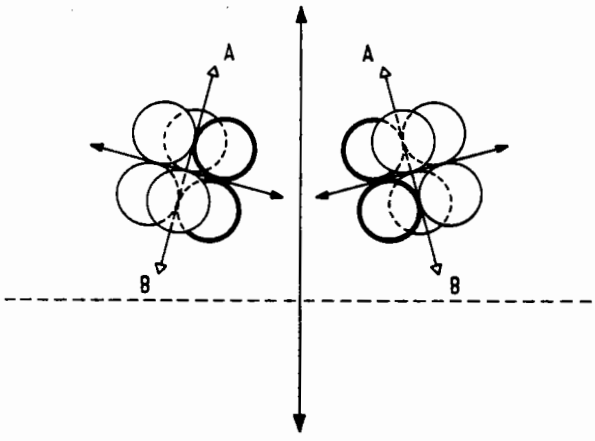


Fig. 2 Schematic drawing of 2 hexameric antiprisms related by a crystallographic two-fold axis (for the rotation function the translation parallel to the two-fold in space group $P2_1$ is irrelevant and has been omitted for clarity). The local three-fold and one local two-fold has been drawn for each hexamer. The associated pseudo two-fold is perpendicular to the plane of the paper with a rotation unequal to 180°.

The accuracy of the rotation function results could be tested because the two heavy atom derivatives used for phase determination at 5 Å resolution contained a large number of sites (Table 2 and ref. 4). The sites allowed an independent assessment of the direction of the non-crystallographic symmetry elements because in the heavy atom parameter refinement all positions were treated independently. As Table 3 shows, the angular parameters derived from the rotation function and from the heavy atom sites agree very well with each other: - only two angles differ by more than 0.5°, one of which is a $\Delta\phi$ of 2.1° when ψ is close to 0° and thus a change in ϕ is not greatly affecting the direction of the axis.

The 5 Å resolution studies[4] did thus not only reveal the shape of the haemocyanin hexamer but also showed that the direction of local symmetry elements can be established with considerable accuracy using rotation function procedures.

4. PHASE IMPROVEMENT AT 4 Å RESOLUTION: THE LOCATION OF THE COPPER IONS

Reprocessing the oscillation films of the two derivatives resulted in quite complete sets of structure factors to 4 Å resolution, while a 3.2 Å native data set was already available. After a few cycles of heavy atom parameter refinement at 4 Å, a new MIR electron density map was calculated.

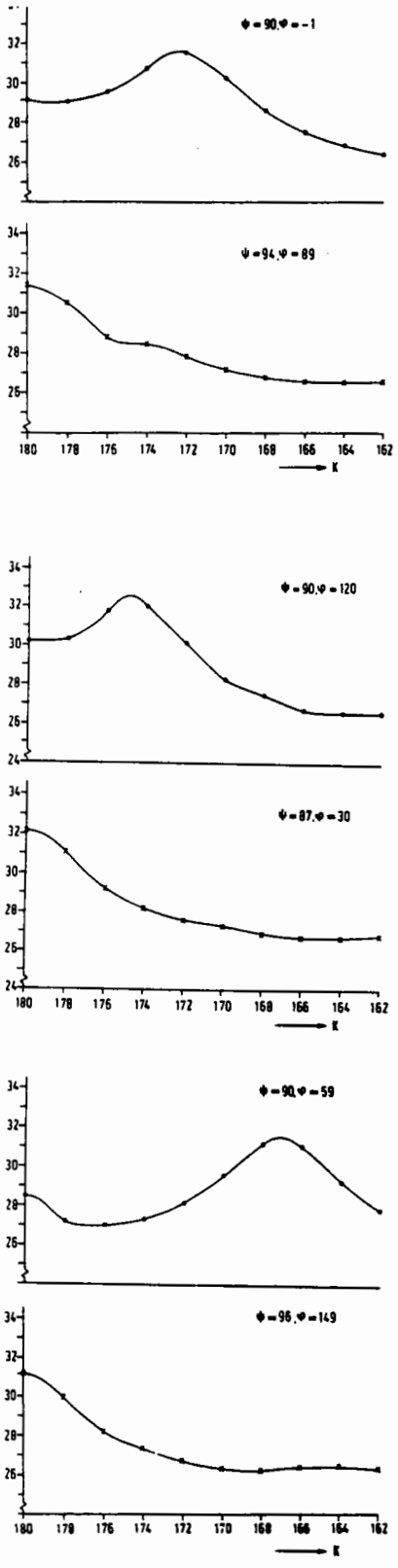


Fig. 3 Pairs of local "true" dyads (κ = 180°) and associated pseudo-dyads (κ ≠ 180°) are revealed by 5 Å rotation function studies. The radius of integration was taken as 50 Å. The resolution range was 6 - 5 Å.

Table 3

Angular parameters of local symmetry elements

A. Rotation Function with 6-5 Å data

	2_A	2_C	2_D	3
ϕ	29.6	89.2	149.4	142.0
ψ	87.4	94.2	96.4	6.5
κ	180.0	180.0	180.0	120.0

B. Heavy-atom Superposition (5 Å data)

ϕ	29.7	89.3	149.6	144.1
ψ	87.3	93.4	96.2	6.4
κ	180.3	180.3	180.5	119.9

From the heavy atom parameters, the centre of the hexamer could be obtained with great accuracy[5]. Six-fold averaging of the density lead to a much improved map which immediately revealed the dinuclear copper sites: two maxima, of equal height, at a mutual distance of ~ 3.8 Å and much higher than any other feature in the map, occurred in the centre of each subunit[5].

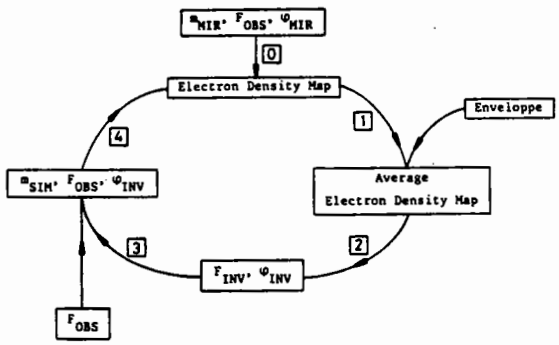


Fig. 4 The density averaging and phase improvement cycle. Instep "2" phases can be calculated for reflections which did not contribute to the calculation of the electron density maps in steps "0" or "4".

Hereafter, Bricogne's program system for density averaging and phase improvement[11] was employed to enhance the quality of the map (Figure 4). Great care was taken in order to ensure that the molecular enveloppes of neighbouring hexamers did not interpenetrate. Six cycles were carried out at 4 Å resolution. The MIR phases were immediately replaced by the phases obtained by Fourier inversion of the six-fold averaged map. Sim weights were applied to calculate new electron density maps and relative Wilson

plots were used in each cycle to obtain scale and temperature factors which were applied to F_{inv}. The overall difference between the final MR phases after six cycles and the initial MIR phases was 60.4°. The overall R-factor, defined as $\Sigma||F_{OBS}| - |F_{INV}||/\Sigma|F_{OBS}|$, was 20.3% after six cycles, whereas the mean Sim weight, or figure of merit, increased from 0.67 after the first cycle to 0.77 after the sixth cycle. Figure 5 shows the phase changes, as function of resolution, in successive cycles. As observed in several other applications of this procedure, convergence was obtained after a few cycles. The 4.0 Å electron density map obtained in this way by using

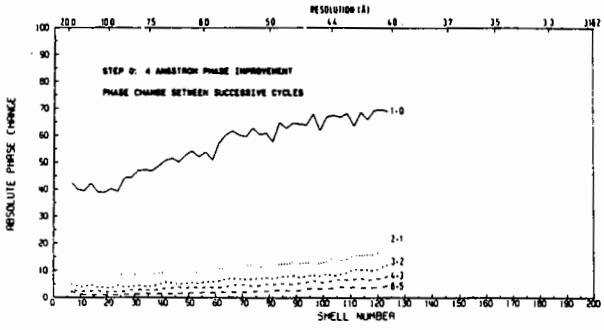


Fig. 5 Phase changes (degrees) between successive cycles in the phase improvement step at 4 Å resolution.

32721 reflections formed the starting point for calculating *de novo* phases for an additional 31121 reflections between 4.0 and 3.2 Å for which no MIR phase information was available.

5. PHASE EXTENSION TO 3.2 A RESOLUTION: THE ARCHITECTURE OF PANULIRUS INTERRUPTUS HAEMOCYANIN

The Fourier inversion (step "2" in the cycle depicted in Figure 4) allows the calculation of phases for reflections with higher resolution than used for the calculation of the electron density map in step "1". As several authors have shown[12-15] that non-crystallographic symmetry leads to phase relationships between reflections with similar resolution, it was decided to expand the phases from 4 to 3.2 Å in ten "Steps" of increasing resolution. Each "Step" consisted of one phase expansion cycle followed by three cycles of phase improvement at constant resolution. Only in the last "Step" a total of eleven phase improvement cycles were carried out as a "finishing touch".

The results obtained by this procedure were better than we had dared to hope for. The entire polypeptide chain could be traced, disulphide bridges and carbohydrate moiety could clearly be located and an "X-ray sequence" could be established which showed good correspondence with the amino acid sequence recently completed by Vereijken, Soeter, Bak and Beintema.

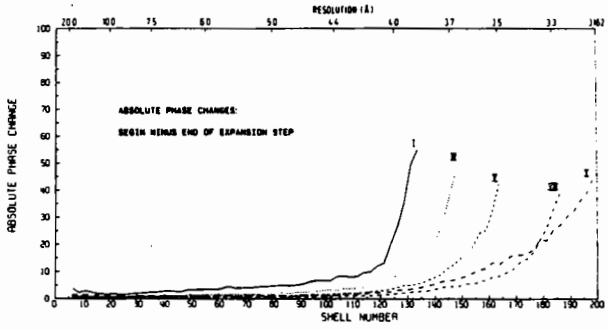


Fig. 6 Phase changes, in degrees, between the beginning and end of the phase expansion Steps I, III, V, VIII and X. In Steps I, III, V and VIII 3 phase improvement cycles at constant resolution were carried out, while after the final expansion Step X 11 phase improvement cycles were carried out.

The phase changes in a number of "Steps" are shown in Figure 6. It appears that the newly added reflections are subject to major changes in phases while the previously calculated phases change considerably less. The effect of the procedure on the electron density map is shown by Figures 7 and 8 which are quite representative for the tremendous improvement of the electron density map occurring at many places of the molecule during the phase extension procedure.

ACKNOWLEDGEMENTS

We like to thank Drs. M.G. Rossmann, R.A. Crowther and G. Bricogne for making computer programs available to us.

REFERENCES

1. E.F.J. van Bruggen, W.G. Schutter, J.F.L. van Breemen, M.M.C. Bijlholt and T. Wicherthes, in "Electron Microscopy of Proteins" (Edited by M. Harris), pp. 1-37. Academic Press, New York (1981).

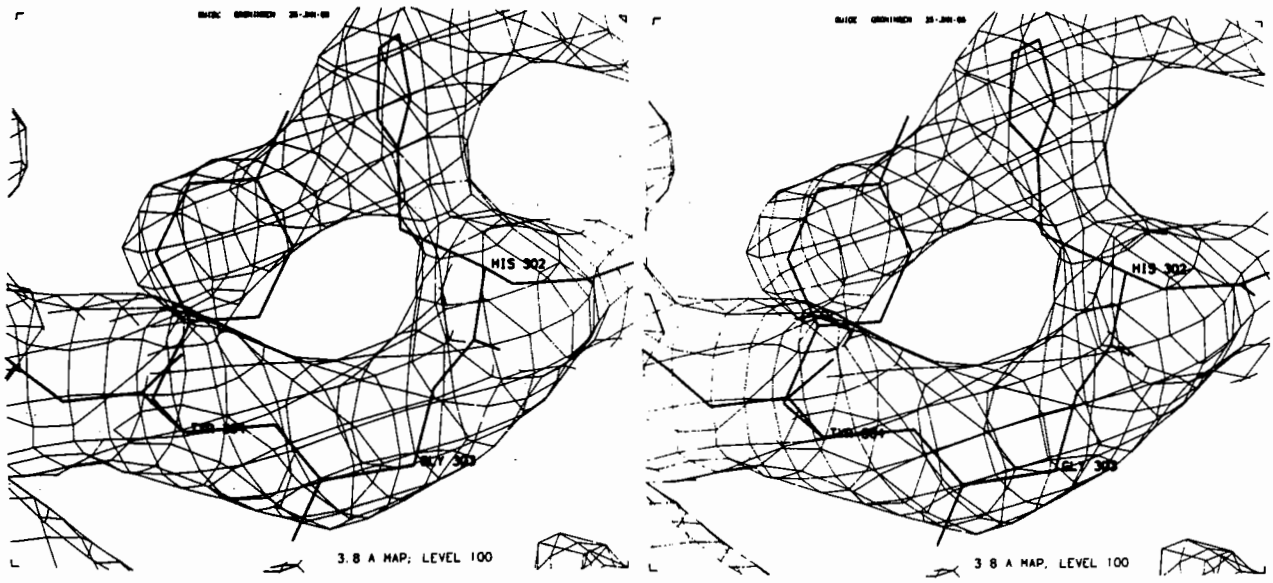


Fig.7

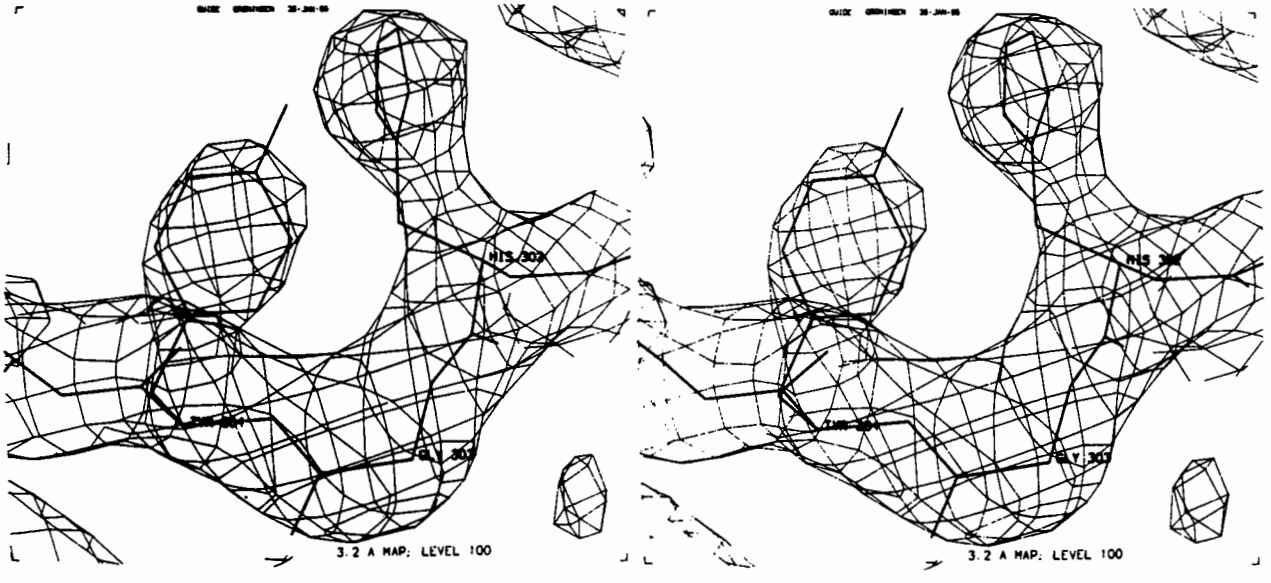


Fig.8

Fig. 7 and 8 Stereofigures of the electron density map at the 3.8 and 3.2 Å resolution stages of the phase determination procedures. Both the tyrosine and histidine side chains are much better defined at 3.2 Å than at 3.8 Å resolution.

2. K.E. van Holde and K.I. Miller, Q. Rev. Biophysics 15, (1982) 1.

3. H.A. Kuiper, W. Gaastra, J.J. Beintema, E.F.J. van Bruggen, A.M.H. Schepman and J. Drenth, J. Mol. Biol. 99, (1975) 619.

4. E.J.M. van Schaick, W.G. Schutter, W.P.J. Gaykema, A.M.H. Schepman and W.G.J. Hol, J. Mol. Biol. 158, (1982) 457.

5. W.P.J. Gaykema, E.J.M. van Schaick, W.G. Schutter and W.G.J. Hol, Chemica Scripta 21, (1983) 19.

6. W.P.J. Gaykema, W.G.J. Hol, J.M. Vereijken, N.M. Soeter, H.J. Bak and J.J. Beintema, Nature 309, (1984) 23.

7. B. Linzen, N.M. Soeter, A.F. Riggs, H.-J. Schneider, W. Schartau, M.D. Moore, E. Yokota, P.Q. Behrens, H. Nakashima, T. Takagi, T. Nemoto, J.M. Vereijken, H.J. Bak, J.J. Beintema, A. Volbeda, W.P.J. Gaykema and W.G.J. Hol, Submitted to Science.

8. W.P.J. Gaykema, A. Volbeda and W.G.J. Hol, in preparation.

9. M.G. Rossmann and D.M. Blow, Acta Crystallogr. 15, (1962) 24.

10. R.A. Crowther in "The Molecular Replacement Method" (Edited by M.G. Rossmann) pp. 174-178. Gordon and Breach, New York (1972).

11. G. Bricogne, Acta Crystallogr. A32, (1976) 832.

12. M.G. Rossmann and D.M. Blow, Acta Crystallogr. 16, (1963) 39.

13. P. Main and M.G. Rossmann, Acta Crystallogr. 21, (1966) 67.

14. P. Main, Acta Crystallogr. 23, (1967) 50.

A NEW SAMPLING METHOD FOR MOLECULAR REPLACEMENT AND ANGULAR CORRELATION

by

Rolf Karlsson
Department of Structural Biology, Biocentre, CH-4056 Basle, Switzerland

1. INTRODUCTION

It is sometimes necessary to calculate the structure factors of a molecular density (or an approximate vector set), which is part or the whole of a known structure, for a new crystallographic environment. This amounts to the isolation of a molecular density ρ_A from the crystal structure ρ and its replication in order to form a new crystal structure.

Isolation of ρ_A is achieved by multiplying the periodic structure ρ with a shape function (e.g. a molecular mask) by which all density except ρ_A is deleted. The continuous Fourier transform of ρ_A is thus evaluated by the convolution of the structure-factors of ρ and the continuous transform G of the shape function, and subsequently sampled onto the new reciprocal lattice.

This principle forms the basis of the well-known structure factor relationships derived by Main and Rossmann[1] and Crowther[2] for the purpose of utilising non-crystallographic redundancies for phase refinement (local symmetry averaging), which were also implicit in the rotation function of Rossmann and Blow[3].

However, their procedure requires the cumbersome evaluation of G at general positions in reciprocal space. In our work, this problem is circumvented by obtaining the continuous transform ρ_A in two steps rather than one.

First the structure factors of ρ_A in the old unit cell are obtained swiftly by standard discrete convolution, or Fourier transform techniques, in which the structure factors of G are used. In a second step the continuous transform of ρ_A is obtained by modified Lagrange interpolation in reciprocal space. The interpolation is a convolution and thus corresponds to a multiplication of the periodic structure of ρ_A with a single unit cell with smoothed boundaries. This requires that the nonperiodic density ρ_A be entirely enclosed within the unit cell.

Although the method is of general use, only the rotation function problem or angular correlation is considered here.

2. ANGULAR CORRELATION

In 1970 Lattman and Love[4] presented a fast version of the rotation function of Rossmann and Blow to determine the orientation of a known molecule in an unknown crystal structure. They calculated the structure factors of the self vector set or molecular Patterson function from atomic coordinates. The transform at the rotated reciprocal lattice points were obtained by linear interpolation.

In the present method the structure factors of an approximate self vector set are obtained by the convolution of the Patterson coefficient F_o^2 based on observed intensities and the structure factors of a spherically symmetric shape function. The present method can thus be considered as a further development of their method. In the following, only self rotation will be considered but extension to cross rotation is easily done.

Criteria for rotational relationships. The overlap criterion or rotation function is usually of the form

$$R = \sum_h F_o^2(h)\, F(u),$$

where $u = hC$, C is the transformation matrix in real space, h is a reciprocal lattice vector and F(hC) are the structure factors of the rotated approximate self vector set.

If F(h) is used instead of F_o^2 (as in Crowther's fast rotation function[5]), a normalised criterion for rotational relationships is obtained:

$$R' = \sum_h F(h)F(u) \;/\; \sum_h F^2(h).$$

It is also possible to calculate a normalised difference criterion:

$$D = \sum_h \{F(h) - \langle F(u)\rangle\}^2 \; / \; \sum_h F^2(h),$$

where the averaged F(u) is taken over the Laue symmetry.

Shape functions. Spherically symmetric shape functions such as a Gaussian function and a sphere convoluted with itself or with a Gaussian function are convenient. These functions are similar to the 'shaded' rotation function of Prothero and Rossmann[6], which uses the product of a Gaussian function and a sphere as shape function. However a pure spherical shape function is discontinuous at radial distance from its centre; this gives rise to a high noise level in the structure factors F at the highest resolution. As a consequence, truncation of the transform of the sphere at different resolutions can give quite different results especially for difficult cases.

Interpolation. Linear (2 by 2 by 2 point) interpolation seems to be as good as higher interpolations. Even the very fast nearest neighbour or one point interpolation, although not giving very smooth functions of the R' and D criteria, seems to give very similar results. Possible influence from neighbouring origin peaks are however reduced with increasing number of interpolation points.

REFERENCES

1. M.G. Rossmann and D.M. Blow, Acta Cryst. 15, (1962) 24.

2. P. Main and M.G. Rossmann, Acta Cryst. 27, (1966) 67.

3. R.A. Crowther, Acta Cryst. 22, (1967) 758.

4. E.E. Lattman and W.E. Love, Acta Cryst. B26, (1970) 1854.

5. R.A. Crowther, in The Molecular Replacement Method. Ed. M.G. Rossmann (London; Gordon and Breach, 1972) p.174.

6. J.W. Prothero and M.G. Rossmann, Acta Cryst. 17, (1964) 768.

CORELS - HOW RIGID IS YOUR MOLECULE?

by

A.G.W. Leslie
Blackett Laboratory, Imperial College, London SW7 2BZ, U.K.

1. INTRODUCTION

The determination of an unknown structure using the molecular replacement method can be subdivided into two quite distinct parts. The first involves finding an appropriate model structure and solving the rotation and translation functions, while the second is concerned with the refinement of the resulting model. The most common strategy is to use CORELS to optimise the position and orientation of the complete molecule in the unit cell, and then go straight into restrained least squares (RLS) refinement, alternating rounds of refinement with manual rebuilding using an interactive graphics display. In some cases it may be necessary initially to omit those parts of the model that are clearly incorrect and rebuild them as the refinement proceeds. Experience has shown that this approach is very often successful, and there are several examples of structures solved in this way in the literature. However, there are some cases in which the errors in the initial model are sufficiently large to prevent convergence to the correct solution, or in less extreme examples, to make the refinement process an extremely long and painful one.

I would like to describe an example where an alternative approach, using CORELS to improve the molecular model before going on to RLS refinement, has proved very successful. This approach was adopted in the refinement of the structures of the apo- and 1 NAD per tetramer species of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) from Bacillus stearothermophilus, using the refined holo-enzyme structure as a starting model[1,2]. Before considering these examples in detail, I would like to give a brief description of the CORELS program itself.

2. CORELS

CORELS is a COnstrained REstrained Least Squares refinement program originally written by Joel Sussman and colleagues[3,4] in the early 1970's to refine the structure of a tRNA molecule. It has been improved considerably in recent years, primarily by Joel Sussman, and in its present form it is applicable to both protein and nucleic acid structures. The feature that makes CORELS different to other RLS programs is that the program refines the parameters of constrained groups of atoms rather than individual atomic parameters. Within a constrained group, all bond lengths and angles are fixed at values determined from small molecule crystallography. The refined parameters are then the position and orientation of the group, and optionally specified dihedral angles within the group (thus a constrained group is not necessarily rigid). A constrained group may be as small as a single amino acid or as large as an entire molecule. This approach results in a considerable reduction in the number of refined parameters, and as a consequence, a significant improvement in both the speed and radius of convergence. It is the increase in the radius of convergence when compared with conventional RLS methods that makes CORELS particularly attractive when employing the molecular replacement technique. This is achieved principally by starting refinement using only low resolution data (e.g. 6Å or even lower resolution), which is possible because of the small number of parameters being refined. The function minimised in the refinement is:

$$Q = \omega_F DF + \omega_D DD + \omega_T DT \qquad (1)$$

where the first term is the usual structure factor contribution:

$$DF = \sum_h \omega_h \left(|F_{obs,h}| - |F_{calc,h}|\right)^2$$

Table 1

The rigid groups used in the 4Å CORELS refinement of GAPDH.

CORELS group	Secondary Structural units	Residue numbers
1	$\beta_A, \beta_B, \beta_C$	0-8,23-24,71-77
2	α_B	9-22
3	α_C	36-52
4	β-structure	53-70
5	α_D	78-88
6	α_E	101-113
7	$\beta_D, \beta_E, \beta_F$	89-100,114-120,142-147
8	β_E-β_F loop	121-141
9	catalytic domain	148-311
10	C-terminal helix	312-333

The second term is for stereochemical restraints, which are all expressed as distance restraints between pairs of atoms:

$$DD = \sum_d \omega_d \, (D_{obs,d} - D_{calc,d})^2$$

where $D_{obs,d}$ is the "ideal" distance between two atoms and $D_{calc,d}$ is the distance calculated from the model. (Planarity restraints are dealt with in this way by the introduction of dummy atoms.)
The final term restrains the model to a set of target coordinates, and is expressed as:

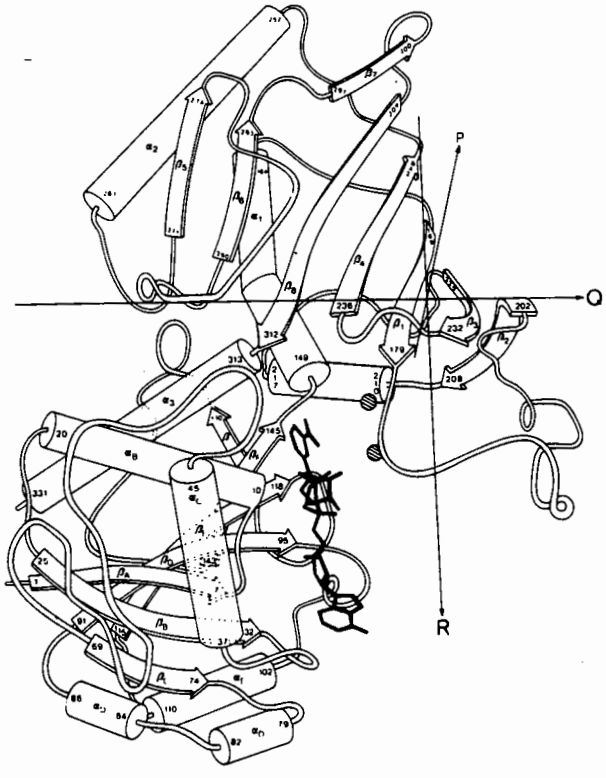
$$DT = \sum_i \omega_i \, (|\underline{X}_{targ,i} - \underline{X}_{calc,i}|)^2$$

$\underline{X}_{targ,i}$ is a vector defining the position of the $\underline{i}$th atom in the target coordinates and $\underline{X}_{calc,i}$ is the vector defining its position in the model.
The quantity Q in equation (1) is an explicit function of all the group positional and thermal parameters, and can be written as:

$$Q = f(t,R,\psi,B)$$

In this expression, t and R refer to the translation vector and rotation matrix to be applied to each constrained group, ψ describes the internal dihedral angles of each group and B represents the thermal parameters. Group derivatives are obtained by differentiation of (1) with respect to individual atomic parameters and application of the chain rule[3].

Fig.1 A schematic representation of the structure of one subunit of B. stearothermophilus holo-GAPDH.



3. THE APPLICATION TO GAPDH

GAPDH is a tetramer of four identical subunits, total MW 145000, that requires the cofactor nicotinamide adenine dinucleotide (NAD) for activity. The structure of the holo-enzyme, with four molecules of NAD per tetramer, has recently been refined to an R-factor of 22% at 2.4Å resolution (unpublished work of A.J. Wonacott and P.C.E. Moody). A schematic diagram of the structure of one subunit is shown in Figure 1. Each subunit is made up of two domains, the coenzyme binding domain (residues 1-147 and 312-333) and the catalytic domain (residues 148-311). The four subunitsare arranged with 222 molecular symmetry, with the catalytic domains forming the inner core of the tetramer. The refined holo-enzyme structure was used as a starting model for the refinement of the apo-enzyme (no bound NAD), which crystallises in the same space group ($P2_1$) but with changes in cell dimensions of up to 5Å. The crystallographic asymmetric unit contains the complete tetramer in both cases. Diffraction data were collected to 6Å resolution from a mercury and a platinum derivative, and the heavy atom coordinates were used to locate the molecular centre and the orientation of the 222 molecular symmetry axes, although this could have been done equally well using rotation and translation functions.

The initial model had an R factor of 44.5% for data to 6Å resolution. This model was then refined using CORELS. Initially only data to 6Å resolution were included, in order to get a reasonably large radius of convergence. Each subunit of the tetramer was divided into three rigid groups, corresponding to the coenzyme binding domain excluding the C-terminal helix, the catalytic domain and the C-terminal helix. Six cycles of CORELS refinement reduced the R factor to 39.4%, with a total shift of 5° in the orientation of the coenzyme binding domain of each subunit. To improve the model further it became necessary to divide the structure up into a larger number of rigid groups. The choice of rigid groups was made by examining a 6Å resolution ($F_o - F_c$), α_c electron density map and by looking at the secondary and tertiary structure of the enzyme. This resulted in each subunit being divided into 10 rigid groups (Table 1), with each group corresponding to an element of secondary structure or a group of elements. The resolution of the X-ray data was extended to 4Å, and refinement reduced the R factor at this resolution to 32.3%. When refined individual atomic temperature factors from the holo-enzyme structure were added to the refined coordinates of the apo-enzyme the resulting model gave an R factor of 30.2% to 3Å resolution. Further refinement was performed using the Hendrickson-Konnert RLS refinement program (unpublished work of T. Skarzynski and A.J. Wonacott). In 19 cycles all the serious errors in stereochemistry introduced during the CORELS refinement (these were primarily at the junctions between rigid groups) had been eliminated, and the R factor had dropped to 25.3% at 2.5Å resolution. At this stage a minor rebuild of the model was required.

The same strategy was adopted for the refinement of the 1 NAD per tetramer species. This form crystallises in an different space group ($P2_12_12$), but the molecular packing is in fact very similar to that in the crystalline apo- and holo-enzymes. The molecular orientation and the position of the molecular centre were again located by analysing the positions of the heavy atoms in mercury and platinum derivatives. The refined apo-enzyme coordinates were used as a starting model, giving an initial R factor of 40.3% at 6Å resolution. Eight cycles of CORELS refinement at 6Å resolution, followed by a further eight cycles at 4Å resolution, using the same rigid group definitions as in the apo-enzyme refinement, reduced the R factor to 29.5%. These coordinates were then combined with the holo-enzyme individual atomic temperature factors and the resulting model was refined for a further eleven cycles using the Hendrickson-Konnert program. Again, the refinement eliminated all serious errors in stereochemistry, and gave a final R factor of 23.9% for data to 3Å resolution.

4. CONCLUSIONS

The total r.m.s. shift in α-carbon positions for the coenzyme binding domain of GAPDH was 1.2Å for the apo-enzyme refinement, with a maximum shift of 2.3Å for the α-carbons in the helix α_D. It is unlikely that shifts of this magnitude would have been within the radius of convergence of a conventional RLS program. The use of CORELS was a much

Table 2

The structural stability of helical secondary structures in molecules undergoing a large conformational change on ligand binding. For any given enzyme, each helix in turn has been superimposed in the two structural forms of the enzyme, using W. Hendricksons superposition program[8]. The superposition was performed using only α-carbons, and the resulting rms deviation in α-carbon positions was noted for each helix. The table gives the overall rms deviation for all helices, and the maximum and minimum deviation for any one helix.

Enzyme	Domain rotation (degrees)	Number of helices	Total number of residues in helices	rms deviations in α-carbon positions after superposition		
				overall	maximum (Å)	minimum
GAPDH	5	4x7	4x86	0.16	0.24	0.09
LADH[6]	7	9	95	0.37	0.45	0.27
Citrate Synthase[7]	18	20	315	0.66	1.17	0.27

simpler alternative to solving the apo-enzyme structure either using multiple isomorphous replacement techniques or using the molecular replacement method with the initial omission and subsequent rebuilding of a large part of the coenzyme binding domain. The results of the work on GAPDH suggest that CORELS refinement is the ideal way of tackling those structural problems where the structure of an enzyme in one state of ligation is known, and where the enzyme is suspected of having undergone a substantial conformational change as the result of a change in the state of ligation by substrate or cofactors. There are of course examples where the conformational change is so large that it would be outside the radius of convergence of even a CORELS refinement. In such cases, however, it may still be possible to locate the larger domain and use this to calculate an electron density map from which it is possible to obtain an initial estimate of the position of the smaller domain. This would provide a greatly improved starting model for further CORELS refinement.

This approach depends on two factors for its success. Firstly, the known structure must be well determined, and preferably already refined at high resolution. Secondly, it must be possible to approximate the conformational change by rigid body movement, of complete domains at the crudest level, or of elements of secondary structure (α-helices and strands of β-sheet) at a more sophisticated level. The analysis of the conformational changes in hexokinase[5], liver alcohol dehydrogenase[6], citrate synthase[7] and GAPDH[1,2] all suggest that this requirement is usually satisfied, although to somewhat varying extents (see Table 2).

The use of CORELS as a precursor to restrained least squares refinement has been of considerable value in the case of GAPDH. As the number of structures solved using molecular replacement techniques increases, it seems likely that CORELS will have an increasingly important contribution to make to protein structure determination.

REFERENCES

1. A.G.W. Leslie and A.J. Wonacott, J. Mol. Biol. 178, (1984) 743.

2. A.G.W. Leslie, Acta Cryst. A40, (1984) 451.

3. J.L. Sussman, S.R. Holbrook, G.M. Church and S.H. Kim, Acta Cryst. A33, (1977) 800.

4. O. Herzberg and J.L. Sussman, J. Appl. Cryst. 16, (1983) 144.

5. W.S. Bennett Jr. and T.A. Steitz, Proc. Natl. Acad. Sci. U.S.A. 75, (1978) 4848; J. Mol. Biol. 140, (1980) 183; Ibid, 140, (1980) 211.

6. H. Eklund, J.P. Samama, L. Wallen, I.C. Branden, A. Akeson and T.A. Jones, J. Mol. Biol. 146 (1981) 561.

7. S. Remington, G. Wiegand and R. Huber, J. Mol. Biol. 158, (1982) 111.

8. W.A. Hendrickson, Acta Cryst. A35, (1979) 158.

SOME EXPERIENCES WITH HAEMOGLOBIN REFINEMENT

by

Zygmunt S. Derewenda*
Chemistry Department, University of York, Heslington, York YO1 5DD.

1. INTRODUCTION

The molecular replacement method (MR) has become an increasingly popular alternative to the multiple isomorphous replacement (MIR) technique of solving macromolecular crystal structures. The two methods yield model structures which may then be used to calculate sets of F_c and α_c and may subsequently be subject to least-squares refinement. The quality of these initial sets of F_c and α_c seems comparable for both methods. However, the nature of errors introduced with each of the models may be different.

The atomic coordinates derived from the MIR map will suffer mainly from random errors with more serious faults confined to less well defined and/or external residues with comparatively high temperature factors. During the refinement of actinidin[1], the r.m.s. shift between the starting and final sets of coordinates was 0.5 Å. The quality of this particular starting model was considered to be high, but the r.m.s. difference quoted is probably typical for many structures determined from MIR electron density maps at 2.0-2.8 Å resolution. Considering the large theoretical radius of convergence of least squares refinement, one might expect these model structures to refine automatically and rapidly.

Although models used in the MR calculations may be superior in terms of their stereochemistry, even small errors in the rotation and translation parameters may lead to systematic shifts of entire parts of the structure relative to true positions. There are also likely to be moves of 1-2 Å in elements of secondary structure which will further weaken the match. This situation may cause severe problems during least-squares refinement. In this paper I wish to report some experiences with the refinement of human semioxyhaemoglobin at 2.1 Å resolution from a set of atomic coordinates obtained by MR.

---

* On leave from the Department of Crystallography, University of Lodz, Poland.

2. THE HAEMOGLOBIN CASE

At the time of our initial calculations, the most accurate model of haemoglobin available was that of human deoxyhaemoglobin refined by G. Fermi at 2.5 Å resolution using real-space refinement techniques[2]. Non-crystallographic, molecular symmetry was assumed to be ideal, and accordingly the model was composed of one dimer. A complete molecule (a tetramer) was obtained by rotating the dimer 180° around the molecular dyad (i.e. the molecular Y axis).

However, in an attempt to investigate the quaternary structure of the molecule in the crystal in question, and in order to assess the power of the computing techniques, a single dimer was used as the search model. The final model of the entire molecule was composed of two dimers, each of which was independently oriented using the fast rotation function program[3]. The positions of the dimers within the unit cell were established simultaneously using a six-dimensional version of an R-factor minimization type translation function with incorporated Fast Fourier Transform. The details of this work have been published elsewhere[4].

The above calculations were carried out using 3.5 Å diffractometer data for four forms of haemoglobin. In order to refine and analyse the structure of semioxyhaemoglobin, a 2.1 Å photographic data set was collected using synchrotron radiation at LURE. To reduce computing costs, the positions of the two dimers established earlier were used to obtain the starting set of coordinates.

The structure has since been extensively refined. The initial stages of refinement were carried out using the unrestrained Fast Fourier Transform refinement method of Agarwal[5], with periodic restoration of proper stereochemistry using the MODELFIT method[6]. In later stages the energy refinement procedure of Jack and Levitt[7] was also

used, while the last stage was performed with the Konnert-Hendrickson restrained refinement procedure[8], modified by E.J. Dodson to make use of the FFT. During the course of the refinement $2F_o-F_c$ and F_o-F_c maps were plotted out at various points. The structure was reviewed and rebuilt in parts several times. Although certain details of static disorder and water structure are still being investigated the structure is essentially refined. The conventional crystallographic R-factor for a restrained coordinate set including 84 water molecules is 20.5%.

3. PROBLEMS ENCOUNTERED DURING THE REFINEMENT

Most of the problems encountered during the refinement of the semioxyhaemoglobin structure are representative of various difficulties resulting from the use of MR coordinates in general. Using somewhat liberal criteria they may be classified as follows:

(i) Errors in the rotation and translation parameters established by the MR technique.

(ii) Inaccuracies of the starting model; in the case of semioxyhaemoglobin the starting model was obtained from real-space refinement at 2.5 Å resolution. Both these factors could have resulted in errors, though one would expect them to be randomly distributed and in general not exceed 0.5 Å for well resolved groups of atoms.

(iii) Influence of crystal packing forces and protein solvent interaction. These are probably the most difficult differences to assess a priori. One may expect that a number of residues involved in intermolecular contacts may exhibit marked conformational changes; the different nature of protein-solvent interactions (e.g. high-salt vs. low-salt solvent) may also cause differences in the hydrogen bond network.

(iv) Lack of water structure in the initial model structure; although crystallographically identifiable water molecules rarely constitute more than 10% of the scattering matter (and thereby their direct effect on the course of protein refinement is nearly negligible), their absence in the initial model may lead to errors such as displacements of atoms hydrogen bonded to water molecules towards the latters' positions, or the refinement of some sidechains into the density of the water molecules and not that of the protein.

(v) Functional differences; normally the structure of the particular form of the protein molecule is not known, and the initial model represents only a related derivative. These structural differences are of major biochemical importance and every care should be taken to separate them from any of the previously described differences.

The above presented problems may be grouped into two categories: a) general errors affecting the entire structure (points 1 and 2), b) localized or highly localized errors affecting individual residues or atoms.

Since the functional differences are impossible to foresee they cannot be classified a priori. All localized errors, and particularly the large errors often cannot be corrected automatically and are normally detected and corrected during the inspection of the electron density map. It is therefore the first category of errors which has a particular effect on the course of automatic refinement, and especially the errors introduced by the rotation and translation parameters.

4. THE ACCURACY OF THE MR SOLUTION IN SEMIOXYHAEMOGLOBIN

Once the structure is fully refined, it is easy to check the accuracy of the initial rotation and translation parameters by simple least-squares matching of the starting model on the final structure. I shall henceforth refer to the model obtained by the MR method as the MR model, and to the model obtained a posteriori by least squares matching on the final structure as the LSQ model. Table 1 compares the rotation (Eulerian α, β and γ) and translation ($|\vec{t}_x|, |\vec{t}_y|, |\vec{t}_z|$) parameters of the two models.

It is clear that the MR method gave an accurate solution, with orientation errors mostly within 1°. The translation parameters are also fairly accurate with only one serious discrepancy (0.6 Å for $|\vec{t}_x|$ of dimer 1).

Fig.1 shows r.m.s. differences (averaged separately

Table 1

The comparison of rotational and translational parameters derived from the MR solution and obtained from a least-squares matching of the initial model onto the final set of coordinates*.

Model	DIMER 1						DIMER 2						R%
	Eulerian angles			Translation vector (Å)			Eulerian angles			Translation vector (Å)			
	α	β	γ	t_x	t_y	t_z	α	β	γ	t_x	t_y	t_z	
MR	66°	-124°	161°	4.12	1.25	15.22	68°	56°	-164°	3.59	1.39	15.6	47.2
L1	66.7°	-123.6°	162.6°	3.88	1.35	15.34	67.7°	56.4°	-164.4°	3.51	1.36	15.57	43.7
L2	67.1°	-123.6°	163.1°	3.68	1.39	15.40	67.6°	56.7°	-164.9°	3.48	1.35	15.5	41.2
LSQ	67.1°	-123.7°	163.1°	3.44	1.32	15.42	67.3°	56.4°	-165.3°	3.47	1.46	15.45	39.3

*Parameters for models denoted L1 and L2 have been obtained as described in section 5.

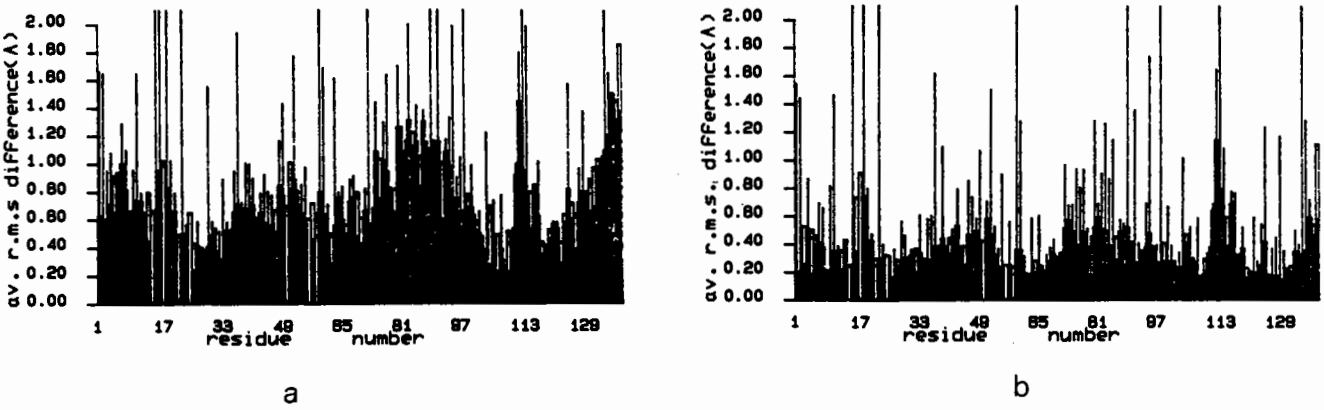


Fig.1 r.m.s. differences (averaged separately over main- and side-chain atoms) between each of the MR and LSQ models and the final set of coordinates of the α_1 chain of semioxyhaemoglobin:
a) MR model vs. final structure; b) LSQ model vs. final structure;
thick bars - mainchain atoms; thin bars - sidechain atoms.

over main-chain and side-chain atoms in each residue) between each of the MR and LSQ models and the final structure (only one chain is shown). It is easy to see that even small errors in the positioning of the model structure introduced serious displacements into the starting set of coordinates.

Figure 2 shows the exact nature of these errors. Figure 2a shows that the F helix was systematically displaced by over 1.0 Å along its axis. On the other hand the G helix shows a very good fit, except for few sidechains (fig.2b).

In order to assess the significance of these errors during the refinement I have simultaneously subjected the LSQ and MR models to refinement, using the Konnert-Hendrickson restrained refinement with FFT.

5. THE REFINEMENT OF LSQ AND MR MODELS

The efficiency of the refinement is highly dependent on the quality of the set of phases obtained with the initial model. If they are close to the correct values, the difference Fourier map should reveal well defined gradients, and accordingly convergence should be fast.

Figure 3 shows the quality of the $2F_o-F_c$ maps calculated using the LSQ and MR models to obtain the values of F_c and α_c. The region shown is that of

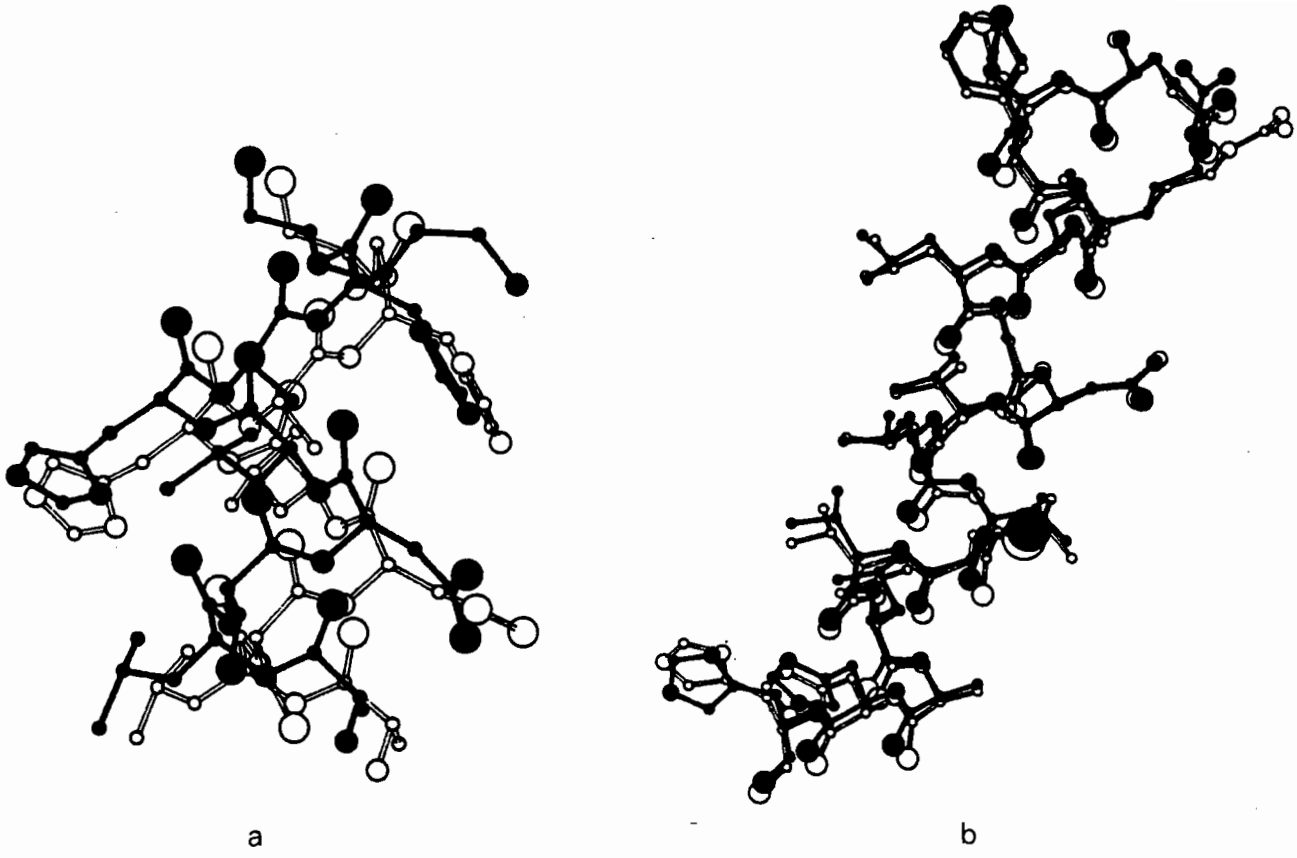


Fig.2 The agreement of the MR with the final structure:
a) region of poor fit - α_1 F helix
b) region of good fit - β_1 G helix
Filled bonds and atoms: final semioxyhaemoglobin structure
Empty bonds and atoms: MR model

Fig.3 $2F_o-F_c\alpha_c$ electron density maps obtained directly from MR and LSQ models without refinement.
a) MR model; b) LSQ model
The maps were calculated and contoured in an identical way; filled bonds - final semioxyhaemoglobin structure; empty bonds - trial model (MR or LSQ).

the α_1F helix, which was earlier shown to be considerably displaced in the MR model.

Ideally, the $2F_o-F_c$ map should reveal the true structure. It is the sum of $F_c\alpha_c$ and $2(F_o-F_c)\alpha_c$ Fourier transforms. The factor of 2 puts the features of the difference map on the absolute scale. The minima indicating wrongly placed atoms should therefore cancel out those features on the $F_c\alpha_c$ map, while new features should reveal the true structure on an absolute scale. This is based on the assumption that the phases are accurate; in other words that the model is fairly close to the correct structure.

Figure 3a shows that the MR model is sufficiently displaced to disturb the phases in such a way that the map reveals significant "ghosting", (i.e. reappearance of the initial erroneous model) and the resulting density spreads over the model and the true positions. The LSQ model yields a much more reliable set of phases, and the density is sharper and emphasises the true positions much more clearly.

Both models have been subjected to 2 rounds of least-squares refinement each consisting of 5 xyz and 3 unrestrained temperature factor refinement cycles. During the first round the distance restraints were set to the value of e.s.d. for a single bond to 0.01 A, somewhat stronger than usual, to keep proper geometry in spite of large shifts in the early stages of refinement. At the beginning of the second round this value was increased to that normally used (0.02 A), and a file with 84 water molecules was added.

Figure 4 shows the progress of both refinements in terms of the decrease of the conventional R factor. Normally one would be inclined to consider the progress of the MR model refinement as at least satisfactory. It is clear, however, that the LSQ model is much better, and the effort involved in rebuilding would probably be much less.

Figure 5 shows that while the LSQ model refines quite easily into the true structure, the MR model (in spite of the encouraging R factor value) is too much in error in some parts to refine adequately. The above calculations indicate that the success and

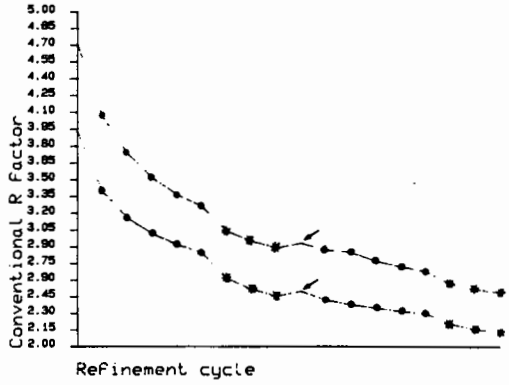


Fig.4 The progress of the refinement of the MR and LSQ models
- ● x,y,z refinement cycle
- * B refinement cycle
- → water file added; restraints relaxed

efficiency of the least-squares refinement of coordinates obtained by the MR method depends strongly on the accuracy of the parameters used to position the initial model.

Having refined the semioxyhaemoglobin structure it is trivial to check these parameters against the LSQ model. With hindsight it would have been much more valuable to assess the accuracy of the MR solution in the course of the refinement, and apply the necessary corrections. I will now show that this is indeed possible.

Following the first round of automatic refinement of the MR model (5 cycles of xyz and 3 of B factor refinement), 91 atoms located not further than 12 Å from the molecule's centre were selected. It was found that the temperature factors of these atoms were all close to or under 20 $Å^2$ (mean value for the structure) and were therefore assumed to have refined correctly. An identicial set of 91 atoms from the initial model used in the MR calculations was fitted by least-squares. The resulting values of orientation and translation parameters were used to reposition the initial model and the refinement was restarted. Another 5 cycles of xyz and 3 cycles of B factor refinement were performed and the procedure was repeated, this time with 254 atoms located within the sphere of 15 Å radius. Three such cycles of orientation refinement were enough to bring the model within 0.2° and 0.05 Å of the LSQ model (Table 1).

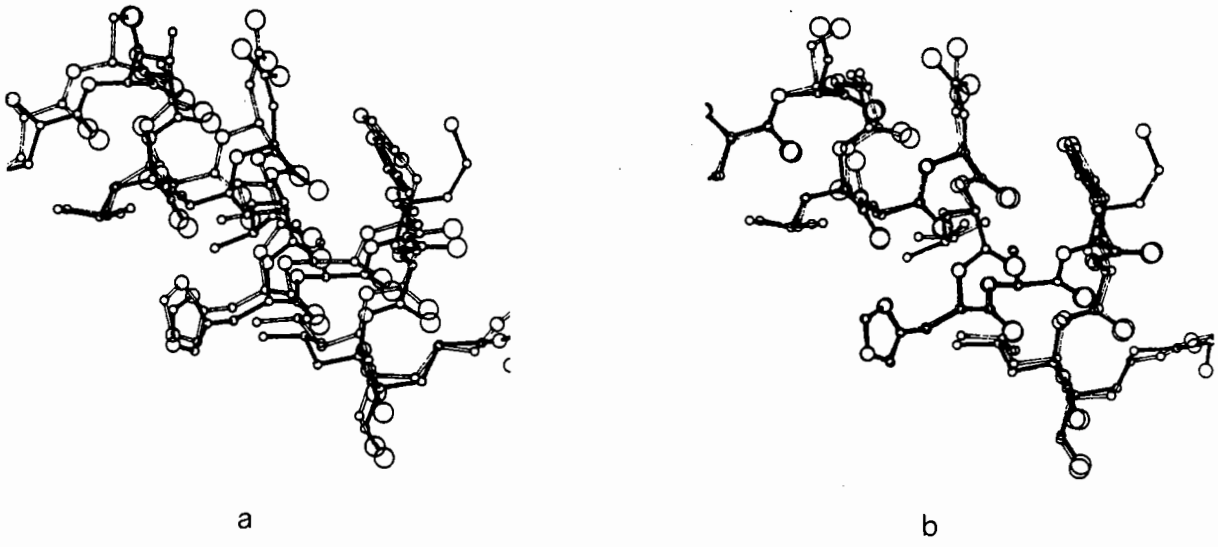


Fig.5 The refined MR and LSQ models (10 cycles of x,y,z and 6 of B refinement as shown on fig.4):
a) MR refined model vs. final semioxyhaemoglobin structure
b) LSQ refined model vs. final semioxyhaemoglobin structure
The part of the structure shown is the α_1F helix.

7. CONCLUSIONS

It is generally agreed that the accurate positioning of the starting model is vital for successful subsequent refinement, and there are a number of possible ways in which one can do it. One useful method is the CORELS[9] technique, i.e. rigid body refinement in the early stages. The calculations presented here indicate that, even though least squares refinement is unable to correct overall errors in orientation and translation parameters, it provides accurate information which can be readily used to improve the starting model. It seems that the least squares refinement approach is very powerful for rigid structures and will possibly cope with independent shifts of several subunits or domains. In our calculations, however, we have used only two independent parts of the haemoglobin molecule.

Since these calculations were carried out, the method was used to refine the structure of methaemoglobin in crystals obtained from poly(ethylene glycol). These crystals seemed to be isomorphous with those of semioxyhaemoglobin and accordingly semioxyhaemoglobin atomic coordinates were used as the initial model for refinement. The starting R factor value was 46% and it was brought down to 27% in the course of the refinement. An electron density map calculated at this stage revealed no serious errors, but little could be done about improving the agreement of the model with the diffraction data. At this point the stragegy described here was used, and it was found that the molecule in crystals of methaemoglobin is shifted by approximately 0.7 Å along the X axis. The initial model was then reconstructed by positioning the semioxyhaemoglobin dimers in the methaemoglobin crystal unit cell in agreement with these observations and the refinement was resumed. Two cycles of this procedure allowed to bypass the local minimum encountered earlier, and the current R factor value is now 21%.

ACKNOWLEDGEMENTS

The calculations described in this paper were carried out during 3 months on leave from the University of Lodz, Poland. I gratefully acknowledge the help of the University of York, in particular of the Chemistry Department and Computing Service for providing the facilities; as well as my home University for a travel grant and indulgence in respect to my frequent absences.

The work referred to has had a number of contributors, of which Dr. G.G. Dodson and Mrs. E.J. Dodson are most outstanding, and whose

constant help and encouragement I acknowledge most gratefully. I thank R.C. Liddington for supplying the data of methaemoglobin which were collected at Daresbury Laboratory. Various stages of the work were financed by the Medical Research Council, Science and Engineering Research Council, European Molecular Biology Organization and Polish Ministry of Science and Technology.

REFERENCES

1. E.N. Baker and E.J. Dodson, Acta Cryst. A36 (1980) 559.

2. G. Fermi, J. Molec. Biol. 97 (1975) 237.

3. R.A. Crowther in The Molecular Replacement Method. A Collection of Papers on the Use of Noncrystallographic Symmetry, ed. by M.G. Rossman (New York: Gordon and Breach, 1972) 173.

4. Z.S. Derewenda, E.J. Dodson, G.G. Dodson and A.M. Brzozowski, Acta Cryst. A37 (1981) 407.

5. R.C. Agarwal, Acta Cryst. A34 (1978) 791.

6. E.J. Dodson, N.W. Isaacs and J.R.S. Rollet, Acta Cryst. A32 (1976) 311.

7. A. Jack and M. Levitt, Acta Cryst. A34 (1978) 931.

8. J.M. Konnert and W.A. Hendrickson, Acta Cryst. A36 (1980) 344.

9. J.L. Sussman, S.R. Halbrook, G.M. Church and S.H. Kim, Acta. Cryst. A33 (1977) 800.

MOLECULAR REPLACEMENT:
A Discussion and Review of the Contributions

by

T.L. Blundell and I.J. Tickle

Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, London University, Malet Street, London WC1E 7HX

The method of molecular replacement depends on three of the most aesthetically pleasing and elegant aspects of crystallography; symmetry, molecular transforms and the convolution theorem. It is therefore an area where intuitive arguments based on concepts have a place alongside the more rigorous mathematical analyses.

Let us begin our review by reiterating the aspects of symmetry which are often ignored in practice as molecular replacement practitioners rush to their computers. An n-fold rotation axis - whether it be a proper rotation or a screw rotation axis - leads to n-fold symmetry in the diffraction pattern and its transform, the Patterson function, whether it be crystallographic or local symmetry. Thus a local two fold axis leads to pseudo 2/m symmetry which can be observed most easily in sections parallel to the local two fold axis. The Patterson has a pseudo-Harker plane through the origin and perpendicular to the local axis; the higher the order of the axis the more peaks on the Harker section. The Harker plane is thus located as a plane containing a higher concentration of vector peaks. Alternatively, the Harker plane transforms in reciprocal space to a line of extra intensity in the direction of the rotation axis - the well known "spike" which is evident in the diffraction patterns of viruses but which can be seen even for the pseudo 2-fold axes in the insulin hexamer.

The existence of more than one molecule or subunit in the asymmetric unit gives a diffraction pattern that is derived from the molecular transform of the molecule in each orientation. When two molecules in different crystal systems are compared the problem is still essentially the same; we must recognise the relative orientations of the two molecular transforms in reciprocal space or the two self vector sets in Patterson space. Much of the meeting addressed the problem of what model should be used to calculate the transform although this was not always explicitly discussed.

When two identical molecules are compared there is considerable advantage in using a refined protein structure at the highest resolution possible. Thermal parameters will be crucial - except for low resolution structure analyses - as these have the effect of emphasising the ordered and more rigid parts of the structure. Although the crystal packing will limit the dynamics of the protein at the points of lattice contact, recent work on comparison of the thermal parameters defined by x-ray analysis and those deduced from normal mode analysis or molecular dynamics shows that x-ray parameters do reflect the intrinsic internal dynamics of the system. They therefore are relevant to more than one crystal system.

When there is no identical structure which has been refined by x-ray analysis of another crystal system, there is the possibility of using proteins modelled by homology, for example see the work by Tickle (this volume) for chymosin. However, this offers little advantage unless thermal motion and disorder are modelled. We suggest that this might most economically be achieved using normal mode analysis, the results of which can be expressed straightforwardly as thermal ellipsoids for the calculation of structure factors.

When the homology between the search structure and the molecule in the unknown crystal structure is not high, the success seems to depend critically on the nature of the protein structure. For example an α-helical protein such as a globin can be successfully studied at low (~ 6 Å) resolution. This success may be related to the fact that helices have low resolution images which correspond to rods of density which are not too dependent upon the amino acid sequence. However, the low resolution image of a β-sheet structure is not so obviously related to the path of its mainchain and may be critically dependent on the sequence. In this case higher (~ 3 Å) resolution data may be appropriate even though the molecular structures are not identical

At very high (~ 1.5 Å) resolutions other problems may arise due to small differences in the relative positions of topologically equivalent secondary structures.

Much of the method of molecular replacement of Rossmann and Blow can be most easily understood in terms of the convolution theorem. This states that the transform of A convoluted with B can be calculated from the product of the transforms of A and B. Thus the rotation function involves the product of two self vector Patterson functions which can be calculated as the convolution of two intensity patterns in Eulerian space or:

$$R = \int P(x_2)P(x_1)dx_1 = (U/V^3) \sum_p \sum_h |F_p|^2 |F_h|^2 G_{h,h'}$$

in which each self Patterson is the convolution of the molecule with its inverse. On the other hand the translation function involves the product of the Patterson function with a Patterson function involving cross vector terms, which is the convolution of the translated and rotated molecule with the inverse.

With respect to rotation functions several questions arose during the discussions. First the success of the Munich group in using Patterson functions is significant. Could the apparently smaller success rate of many other groups be due to the limitations of the number of reflections in many of the reciprocal space programs?

A general question concerning rotation functions appears to be the precision of the rotation angles defined. These must depend on the resolution of the data used. In fact the consistency of the rotation angles with data shells of different resolutions appears to be a good guide to a correct solution. Clearly the higher the resolution the better. For large molecules with a radius of ~ 60 Å, the resolution should be reasonably high; an error of 1° in the rotational parameter will lead to a 1 Å error at the periphery of the molecule. If used in molecular replacement or symmetry averaging this will have a serious effect on the interpretation - see for example the work on the semioxygenated haemoglobin (this volume) where the structure near the rotation axis was clearer and more easily interpretable than that at the molecular periphery.

One aspect of translation functions that has been paid little attention is the use of overlap functions. Any pair of molecules in a crystal unit cell must occupy distinct volumes which do not overlap. The closest contact cannot be less than a van der Waals distance ~ 3 Å. This can be re-expressed by saying that in a set of cross-Patterson vectors between molcules related by a crystallographic operation ($\underline{Ax} + \underline{d}$) there should be no interatomic vectors in a sphere of radius 3 Å around the origin. The cross-Patterson is given by

$$P_{12}(\underline{u}) = \int_v \rho_1(\underline{x})\rho_2(\underline{x} + \underline{u})\, d\underline{x}$$

This is the convolution of one molecule with its inverse operated on by the symmetry. Thus if $\underline{t}$ is the translation between 1 and 2

$$P_{12}(u) = \Sigma\, F_M(\underline{h})\, F_M^*(\underline{h}.\underline{A}) \exp -2\pi i(\underline{h}.\underline{t} - \underline{h}.\underline{u})$$

This function should be zero within radius 3 Å of the origin.

$$\therefore OF = \int_0^{3\text{Å}} \Sigma\, F_M(\underline{h}) F_M^*(\underline{h}.\underline{A}) \exp -2\pi i\, \underline{h}.\underline{t} \exp 2\pi i\, \underline{h}.\underline{u}\, dV$$

This is part of the calculation implied by the translation function of Crowther and Blow and may be used to weight the terms in the function.

The weight should be dependent on $\frac{1}{OF}$ and the new function might be

$$T(t) = \frac{1}{OF} \Sigma |F_{obs}(\underline{h})|^2 F_M(\underline{h})\, F_M^*(h.A) \exp(-2\pi i\, \underline{h}.\underline{t})$$

This is similar to the translation overlap function of Lifschitz et al except that OF is there defined simply as the origin peak height of $P_{12}(u)$.

Finally we consider analysis of the rotation and translation functions. There is some confusion in the definition of the significance of peaks. To avoid this it would be sensible for both the number of standard deviations above the mean and above the next significant peak to be quoted for the peak of choice.

The Table shows the forms of crystallographic agreement values (R) and product moment correlation

coefficients (C) which may be of use in evaluating the results. Although the expected values for random and perfect solutions are well known, further work is necessary to define the expected values especially when the orientation is correct and only the translation is in error, and when only a few reflections are used. In general it seems that correlation functions are more discriminating than R-factors.

Table

Conventional crystallographic residual

$$R = \frac{\Sigma_h \left| F_{obs} - F_{calc} \right|}{\Sigma_h F_{obs}}$$

Product moment correlation coefficient

$$C = \frac{\Sigma_h [(X_{obs} - \bar{X}_{obs}) \cdot (X_{calc} - \bar{X}_{calc})]}{\left(\Sigma_h (X_{obs} - \bar{X}_{obs})^2 \cdot \Sigma_h (X_{calc} - \bar{X}_{calc})^2 \right)^{1/2}}$$

($X = F$, E, F^2 or E^2, note $\bar{E}^2 = 1$)

Random*: $R = 0.586 - (0.203/n)^{1/2}$ (acentric);
$R = 0.828 - (0.455/n)^{1/2}$ (centric)
$C = 0$

Perfect: $R = 0$
$C = 1$

*The authors thank Prof. A.J.C.Wilson for comments on the forms of these expressions.
n is the number of reflections.

LIST OF DELEGATES

ACHARYA, R. Dr.
Laboratory of Molecular Biophysics
Rex Richards Building
Department of Zoology
University of Oxford
South Parks Road
Oxford OX1 3QU

ADAMS, M.J. Dr.
Laboratory of Molecular Biophysics
Rex Richards Building
Department of Zoology
University of Oxford
South Parks Road
Oxford OX1 3QU

BAILEY, S. Ms.
Birkbeck College
University of London
Malet Street
London WC1E 7HX

BAKER, P. Mr.
Department of Biochemistry
University of Sheffield
Sheffield S10 2TN

BANNER, D.W. Dr.
EMBL
Meyerhofstrasse 1
Postfach 10.2209
D-6900 Heidelberg
German Fed. Rep.

BARFORD, D. Mr.
Laboratory of Molecular Biophysics
Department of Zoology
University of Oxford
South Parks Road
Oxford OX1 3QU

BHADBHADE, M. Dr.
Laboratory of Molecular Biophysics
Rex Richards Building
Department of Zoology
University of Oxford
South Parks Road
Oxford OX1 3QU

BLOOMER, A.C. Dr.
Medical Research Council
Laboratory of Molecular Biology
University Postgraduate Medical School
University of Cambridge
Hills Road
Cambridge CB2 2QH

BLOW, D.M. Prof.
Blackett Laboratory
Imperial College
Prince Consort Road
London SW7 2BZ

BLUNDELL, T.L. Prof.
Department of Crystallography
Birkbeck College
University of London
Malet Street
London WC1E 7HX

BRICK, P. Dr.
Biophysics Group
Blackett Laboratory
Imperial College
London SW7 2AZ

BROWN, R. Mr.
EMBL
Meyerhofstrasse 1
Postfach 10.2209
D-6900 Heidelberg
German Fed. Rep.

BUEHNER, M. Dr.
Forschergruppe Roentgenstrukfuranalyse
Biologischer Makromolekülе
Universität Würzburg
Am Hubland
D-8700 Würzburg
German Fed. Rep.

CAMPBELL, J. Dr.
SERC Daresbury Laboratory
Daresbury
Warrington WA4 4AD

CARR, P. Mr.
Department of Pure & Applied Physics
University of Manchester
Institute of Science & Technology (UMIST)
P.O. Box 88
Manchester M60 1QD

CHATTOPADHYAY, T. Mr.
Department of Crystallography
Birkbeck College
University of London
Malet Street
London WC1E 7HX

CLEASBY, A. Ms.
Department of Crystallography
Birkbeck College
University of London
Malet Street
London W1E 7HX

COOPER, S. Ms.
Birkbeck College
University of London
Malet Street
London WC1E 7HX

DEREWENDA, U. Dr.
Department of Chemistry
University of York
Heslington
York YO1 5DD

DEREWENDA, Z. Dr.
Department of Chemistry
University of York
Heslington
York YO1 5DD

DIDEBERG, O. Dr.
Chef du Travaux
Laboratoire de Cristallographie
Institut de Physique 85
Universite de Liege
B-4000 Liege
Belgium

DODSON, E.J. Mrs.
Department of Chemistry
University of York
Heslington
York YO1 5DD

DODSON, G. Dr.
Department of Chemistry
University of York
Heslington
York YO1 5DD

DRIESSEN, H. Dr.
Department of Crystallography
Birkbeck College
University of London
Malet Street
London WC1E 7HX

ELIOPOULOUS, E.E. Mr.
Astbury Department of Biophysics
University of Leeds
Leeds LS2 9JT

ELLIS, G.H. Mr.
Laboratory of Molecular Biophysics
Rex Richards Building
Department of Zoology
University of Oxford
South Parks Road
Oxford OX1 3QU

EVANS, P.R. Dr.
MRC Laboratory of Molecular Biology
University Medical School
Hills Road
Cambridge CB2 2QH

FARRANTS, G. Dr.
Department of Biochemistry
University of Sheffield
Sheffield S10 2TN

FORD, G.C. Dr.
Department of Biochemistry
University of Sheffield
Sheffield S10 2TN

GAMBLIN, S. Mr.
Department of Biochemistry
University of Bristol
Medical School
University Walk
Bristol BS8 1TD

GEDDES, A.J. Dr.
Astbury Department of Biophysics
University of Leeds
Leeds LS2 9JT

GOVER, S. Dr.
Laboratory of Molecular Biophysics
Rex Richards Building
Department of Zoology
University of Oxford
South Parks Road
Oxford OX1 3QU

HARDING, M.M. Dr.
Department of Inorganic, Physical
& Industrial Chemistry
Donnan Laboratories
University of Liverpool
Gove Street
P.O. Box 147
Liverpool L69 3BX

HARLOS, K. Dr.
Laboratory of Molecular Biophysics
Rex Richards Building
Department of Zoology
University of Oxford
South Parks Road
Oxford OX1 3QU

HARRISON, P.M. Prof.
Department of Biochemistry
University of Sheffield
Western Bank
Sheffield S10 2TN

HELLIWELL, J. Dr.
SERC, Daresbury Laboratory
Daresbury
Warrington WA4 4AD

HILL, C. Mr.
Department of Chemistry
University of York
Heslington
York YO1 5DD

HOL, W.G.J. Dr.
Laboratory of Chemical Physics
University of Groningen
Nijenborgh 16
9747 AG Groningen
The Netherlands

HUBER, R. Prof.
Max-Planck-Institut f. Biochemie
D-8033 Martinsreid bei Munchen
Munchen
German Fed. Rep.

HUNTER, W. Dr.
Chemical Laboratories
University of Cambridge
Lensfield Road
Cambridge CB2 1EW

KARLSSON, R. Dr.
C/o Professor J.N. Jansonius
Biozentrum der Universitat Basel
Klingelbergstrasse 70
CH-4056 Basel
Switzerland

KNIGHT, S. Mr.
SLU, Molecular Biology
BMC
Box 590
S-75724 Uppsala
Sweden

KOKKINIDIS, M. Dr.
EMBL
X-ray Division
Meyerhoffstrasse 1
D-6900 Heidelberg
German Fed Rep.

KORBER, F.C.F. Dr.
Astbury Department of Biophysics
University of Leeds
Leeds LS2 9JT
West Yorkshire

LESLIE, A.G.W. Dr.
Blackett Laboratory
Imperial College
London SW7 2BZ

LIDDINGTON, R. Mr.
Department of Chemistry
University of York
Heslington
York Y01 5DD

LILJAS, L. Dr.
SLU, Molecular Biology
BMC
Box 590
S-75724 Uppsala
Sweden

MACHIN, P. Miss
(Secretary)
SERC Daresbury Laboratory
Daresbury
Warrington WA4 4AD

MAHADEVAN, D. Mr.
Department of Crystallography
Birkbeck College
University of London
Malet Street
London WC1E 7HX

MCLAUGHLIN, P. Mr.
Laboratory of Molecular Biophysics
Rex Richards Building
Department of Zoology
University of Oxford
South Parks Road
Oxford OX1 3QU

MOSS, D.S. Dr.
Department of Crystallography
Birkbeck College
University of London
Malet Street
London WC1E 7HX

MUIRHEAD, H. Dr.
Department of Biochemistry
University of Bristol
University Walk
Bristol BS8 1TD

MURRAY-RUST, J. Dr.
Department of Crystallography
Birkbeck College
University of London
Malet Street
London WC1E 7HX

NAJMUDIN, S. Mr.
Department of Crystallography
Birkbeck College
University of London
Malet Street
London WC1E 7HX

NORTH, A.C.T. Prof.
Astbury Department of Biophysics
University of Leeds
Leeds LS2 9JT

OLIVA, G. Mr.
Department of Crystallography
Birkbeck College
University of London
Malet Street
London WC1E 7HX

PARKER, M. Mr.
Laboratory of Molecular Biophysics
Department of Zoology
Rex Richards Building
University of Oxford
South Parks Road
Oxford OX1 3QU

PHILLIPS, S.E.V. Dr.
Astbury Department of Biophysics
University of Leeds
Leeds LS2 9JT

PICKERSGILL, R. Dr.
Laboratory of Molecular Biophysics
Rex Richards Building
Department of Zoology
University of Oxford
South Parks Road
Oxford OX1 3QU

RABINOVICH, D. Prof.
Weizmann Institute of Sciences
C/o Chemical Laboratories
University of Cambridge
Lensfield Road
Cambridge CB2 1EW

RAWAS, A. Mr.
Department of Biochemistry
University of Bristol
Woodland Road
Bristol

REYNOLDS, C. Dr.
Department of Chemistry
University of York
Heslington
York YO1 5DD

RICE, D.W. Dr.
Department of Biochemistry
University of Sheffield
Sheffield S10 2TN

RIZKALLAH, P. Mr.
Astbury Department of Biophysics
University of Leeds
Leeds LS2 9JT

RULE, S. Mr.
SERC Daresbury Laboratory
Daresbury
Warrington WA4 4AD
and
Department of Chemistry
University of Liverpool
Donnan Laboratories
Gove Street
P.O. Box 147
Liverpool L69 3BX

RYPNIEWSKI, W.J. Mr.
MRC Laboratory of Molecular Biology
Hills Road
Cambridge

SAMRAOUI, B. Mr.
Laboratory of Molecular Biophysics
Rex Richards Building
Department of Zoology
University of Oxford
South Parks Road
Oxford OX1 3QU

SAUL, F. Dr.
Institut Pasteur
Immunologie Structurale
28 Rue du Docteur Roux
F-75724 Paris Cedex 15
France

SAWYER, L. Dr.
Department of Biochemistry
University of Edinburgh Medical School
Hugh Robson Building
George Square
Edinburgh EH8 9XD

SHIRAKIHARA, Y. Dr.
MRC Laboratory of Molecular Biology
Hills Road
Cambridge

SKARZYNSKI, T. Dr.
Blackett Laboratory
Imperial College of Science & Technology
Prince Consort Road
London SW7 2AZ

SMITH, J. Dr.
Department of Biochemistry
University of Sheffield
Sheffield S10 2TN

STUART, D. Dr.
Laboratory of Molecular Biophysics
Rex Richards Building
Department of Zoology
University of Oxford
South Parks Road
Oxford OX1 3QU

SUTTON, B. Dr.
Laboratory of Molecular Biophysics
Rex Richards Building
Department of Zoology
University of Oxford
South Parks Road
Oxford OX1 3QU

TICKLE, I.J. Dr.
Department of Crystallography
Birkbeck College
University of London
Malet Street
London WC1E 7HX

TICKLE, Jasmine Dr.
Department of Crystallography
Birkbeck College
University of London
Malet Street
London WC1E 7HX

TOUGARD, P. Dr.
Institut Pasteur
25 Rue du Docteur Roux
F-75724 Paris Cedex 15
France

TOWNS, E. Ms
SERC Daresbury Laboratory
Daresbury
Warrington WA4 4AD

TREHARNE, A.C. Ms.
Room EB14
Birkbeck College
University of London
Malet Street
London WC1E 7HX

WATSON, F. Dr.
Department of Crystallography
Birkbeck College
University of London
Malet Street
London WC1E 7HX

WATSON, H.C. Dr.
Department of Biochemistry
University of Bristol
Medical School
University Walk
Bristol BS8 1TD

WHITE, J. Dr.
Department of Biochemistry
University of Sheffield
Sheffield S10 2TN

WONACOTT, A. Dr.
Blackett Laboratory
Imperial College of Science & Technology
Prince Consort Road
London SW7 2AZ