Acta Cryst. (1972). A28, 134

An Application of Vector Space Search Methods to the Patterson Function of Myoglobin

BY C. E. NORDMAN

Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48104, *U.S.A.*

(*Received* 7 *June* 1971)

By searching the 1·5 Å Patterson function of myoglobin (space group $P2_1$) with intragroup vector sets characteristic of the heme group and of the main chain of a 4- or 5-turn α-helix, it is shown that it is possible to determine the orientation of the heme, and the axial directions of the prominent helices *A*, *B*, *E*, *G*, and *H*. Additional searches performed with helices *A*, *G*, and *H* correctly reveal the true orientations about the helical axes. Translation searches with intergroup vectors, similarly yield the correct positions of the heme and helices *A*, *G*, and *H* relative to a 2_1 axis. The relative *y* coordinates (along 2_1) of the heme group and a helix can be unambiguously found by a search using vectors from the heme Fe atoms to the correctly oriented helix. Several image-seeking functions are explored by varying the selection and weighting of the search vectors. The best results are obtained with an image-seeking function defined as a weighted average over a selected subset of search vectors. The selection is made for each setting of the vector set, and is based on the value of the Patterson function at each search vector in relation to the distribution of values in the entire Patterson function.

Introduction

Computer programs for interpreting stored Patterson functions of structures containing known rigid groups have been developed in several laboratories in recent years (Sparks, 1961; Huber, 1965, 1970; Braun, Hornstra & Leenhouts, 1969; Hornstra, 1970; Nordman & Nakatsu, 1963; Schilling, 1970). Such programs have been used to determine the structures of several dozen small (*i.e.* non-protein) organic molecules.

There is a close relationship between these Patterson-space methods and the corresponding reciprocal-space methods, such as those based on the rotation functions of Rossmann & Blow (1962) or Tollin & Cochran (1964), and several reciprocal-space translation functions (Rossmann, Blow, Harding & Coller, 1964; Tollin, 1966; Crowther & Blow, 1967). The main difference between the Patterson space and the reciprocal space methods is the seemingly trivial one of computational layout. Beyond this, however, the two approaches offer entirely different possibilities in selective sampling of their respective spaces, carried out at execution time.

Successful applications of the reciprocal-space approach include several interesting results obtained with biological macromolecules. Tollin (1969) has used the known structure of the myoglobin molecule from sperm whale myoglobin (SWMb, space group $P2_1$) to determine the complete crystal structure of seal myoglobin (space group *A*2). Recently Lattman & Love (1970) have used the Rossmann–Blow rotation function to detect the similarity between the SWMb molecule and the molecules of single-chain lamprey hemoglobin (space group $P2_12_12_1$), and to determine the orientation of the molecules in the crystals of lamprey hemoglobin.

Zwick (1969) has used reciprocal-space methods to search for the heme group and α-helices in myoglobin. He has found the approximate orientation of the heme group, the directions of three helices, and the location of one of them. These results are at least qualitatively comparable to ours, as reported here.

In this paper, we present results obtained by Patterson-space searching applied to sperm whale myoglobin using the α-helix and the heme group as known rigid substructures. The procedures and programs used were essentially those described by Schilling (1970) and Nordman & Schilling (1970), modified to some degree as explained below.

The crystal data for SWMb are: space group $P2_1$, $a=64{\cdot}5$, $b=30{\cdot}86$, $c=34{\cdot}7$ Å, $\beta=106°$. The X-ray data for the native myoglobin to 1·5 Å resolution were kindly supplied to us in 1966 by Dr H. C. Watson.

Patterson functions and vector sets.

Three distinct versions of the Patterson function were used. All were sharpened and calculated with their origin peaks removed, the scaling having been done by an isotropic Wilson procedure. The first two Patterson functions were computed on a grid having $60\times60\times60$ points per unit cell. This choice, which was necessitated by program limitations on the IBM 7090, corresponds to the somewhat unfavorable grid intervals of $64{\cdot}5/60=1{\cdot}08$, $30{\cdot}86/60=0{\cdot}51$, and $34{\cdot}7/60=0{\cdot}58$ Å in the *x*, *y*, and *z* directions. These two Patterson functions were computed to 1·5 and 2·0 Å resolution respectively, each with a damping parameter $D=8$ Å^2 in the damping factor $\exp[-D(\sin\theta/\lambda)^2]$ applied to the 'point atom' (*i.e.* no-falloff) Patterson coefficients (Nordman & Schilling, 1970).

The third Patterson function was calculated as a 'point atom' (*i.e.* $D=0$) function to 1·5 Å resolution. It was evaluated on a grid having $120\times60\times60$ points per unit cell, which corresponds to nearly uniform grid steps of 0·54, 0·51, and 0·58 Å in the three directions.

The Fourier synthesis program of the *X-RAY* system (Stewart, Kundell & Baldwin, 1970) was used, suitably modified so as to produce the Patterson function in a form acceptable as input to the search programs. The majority of the searches described below were performed with this third Patterson function.

The myoglobin Patterson function differs markedly from its small-structure counterparts. The main distinction is the deep layer of unresolved vector density which constitutes the 'background' of the function. At 1·5 Å resolution and zero damping, the median of the computed values of the function represents a vector density of about 100 times the height of a single carbon–carbon peak, and 95% of the function falls in the range 80 to 120 carbon–carbon vectors.

The search vector components and overlap weights were computed by means of the *VEC* program, which has been described in some detail elsewhere (Nordman & Schilling, 1970). A considerable speed-up of the calculation was achieved by internally sorting the vectors according to the length of one component prior to the calculation of overlap weights.

The largest search group used was an 18-residue α-helix containing 90 atoms and yielding over 4000 distinct interatomic vectors. At 1·5 Å resolution and $D=0$, any two vectors that differ by a vector of magnitude less than 1·07 Å are expected to overlap to some degree. As a result of multiple overlap of incompletely resolved peaks, the effective vector density within the cluster of vectors typically reaches values of 10–20 times the peak height of a single carbon–carbon vector.

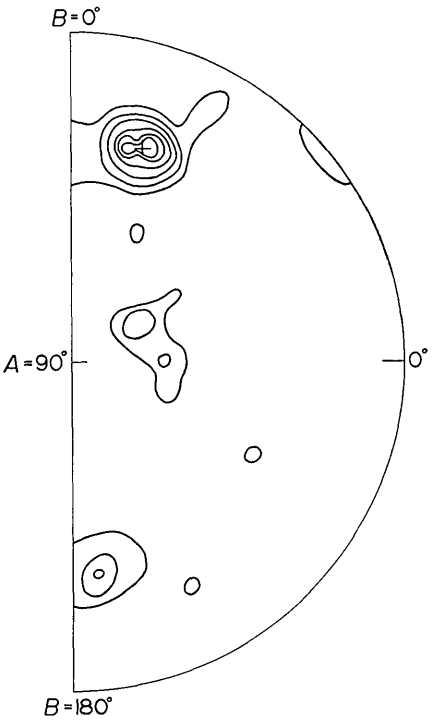

Fig. 1. Rotation search for the direction of the normal to the plane of the heme group. The plotted function is rmin(28,52) averaged over 5 settings, 72° apart, about the heme normal. The contours are at function values which exceed the mean by $n\sigma$, where $n=1,2,\ldots,6$.

In searching the 1·5 Å 'point atom' Patterson function for α-helices it was noticed that the best results were obtained when the vector set was assumed to be slightly 'sharper' than the Patterson function. This observation can be rationalized by the reasonable assumption that the effective thermal motion of the helical main-chain atoms is less than that of the structure as a whole. Accordingly, the majority of the vector sets used in searches of the 1·5 Å 'point-atom' Patterson function were calculated with an assumed 1·2 Å resolution, and a D value of zero. These parameters give a maximum calculated overlap radius of 0·85 Å.

It is usually not practical to carry out a Patterson search using *all* – say, 4000 – interatomic vectors in a moderately large search group. Neither would it be correct to do so, since the overlapping of two nearly coincident vectors results in the near doubling of the weight of each; inclusion of *both* vectors in the search would amount to an overweighting of the pair. Stated differently, two nearly coincident search vectors will, in effect, convey the same information to the image-seeking function, and the use of both vectors is redundant. A procedure for arriving at the best set of search vectors is to select the highest-weight vectors, subject to the secondary condition that no two selected vectors differ by a vector whose magnitude is less than a specified minimum separation parameter. This selection of search vectors was carried out with a simple program, *VSEL* (Nordman & Schilling, 1970), using minimum separation parameters of 0·5 or 0·6 Å. The number of vectors used as search vectors ranged from 50 to 400 in the various searches described below.

Image-seeking functions

Because of the high overall level of the myoglobin Patterson function, many of the commonly used image-seeking functions are unsuitable. The value of the Patterson function typically exceeds the vector density of the search vector set everywhere, and consequently the search vector set can be accommodated in the Patterson function in any position.

The image-seeking function which we have found most useful in small-molecule problems is the 'minimum average' function, min(M,N) (Schilling, 1970). This function is defined as

$$\min(M,N)=\sum P_j/\sum w_j, \quad j=1,\ldots,M,$$

where P_j and w_j are, respectively, the Patterson value and weight of the search vector having the jth lowest value of P/w. The sums include the M search vectors having the lowest values of P/w, where M is a number typically chosen as 10 to 50% of the total number of search vectors, N. In the limit of $M=1$, the function becomes a weighted minimum function. At the other extreme of $M=N$ the minimum average function is

equivalent to the ordinary (unweighted) sum function, and $\sum w_j$ is a constant.

The rationale behind the 'minimum average' function is similar to that behind the minimum function: if the lowest P/w values in general represent search vector points at which the Patterson function is too low to accommodate the vector density of the search vector, then a high value of the $\min(M,N)$ function should sensitively reveal the correct orientation (translation) of the search vectors, by selectively displaying a good fit of even the worst-fitting vectors. Since, as mentioned above, the Patterson function of myoglobin is high enough to accommodate the search vectors everywhere, it is not surprising that the ordinary $\min(M,N)$ function is a poor image-seeking function in this application.

The manner in which it fails is, essentially, as follows: since the background is high, the *relative* fluctuations of the Patterson function are modest. Consequently, the search vectors enter the ranks of those having the lowest P_j/w_j, more by reason of a high value of w_j than a low value of P_j. Since w_j is a characteristic of the search vector itself and not of the Patterson point it scans, it follows that only those vectors with the highest w_j tend to determine the value of $\min(M,N)$. The result of the search then tends to depend only on the M search vectors with the highest values of w, and much of the structural detail contained in the larger set of N search vectors is lost.

A considerable improvement is achieved by using, what might be called, a weighted sum function:

$$\text{wsum}(N) = \sum_j w_j P_j / \sum_j w_j \quad j=1, \ldots, N.$$

Here, the summations include all N search vectors. The sum of the weights appearing in the denominator is a constant in this case. If we denote, by P_C, the calculated height of a single carbon–carbon vector peak, then

$$P_j^{sg} = w_j P_C$$

represents the value of the convolution molecule, *i.e.* of the Patterson function of the search group, evaluated at the vector point j. Consequently,

$$\text{wsum}(N) = (\sum_j P_j P_j^{sg})/(P_C \sum_j w_j) \quad j=1, \ldots, N$$

where the denominator is a constant. If the sum in the numerator is replaced by the corresponding integral, it can, alternatively, be evaluated in reciprocal space, and becomes equivalent to the criterion of fit used in rotation and translation functions as pointed out above.

An alternative to the weighted sum function can be developed by the following argument: as stated above, if the Patterson function everywhere exceeds every value of $w_j P_C$, it will not be possible to categorically rule out any rotational or translational positioning of the search vector set. However, it is possible, at least qualitatively, to estimate the likelihood that a search vector set of given w_j values is correctly placed in the Patterson function. For example, placing a high-weight vector at a high Patterson value and a lower-weight vector at a lower Patterson value represents a more likely solution than the reverse choice, even though either choice is formally compatible with the Patterson function.

Let us think of the Patterson function as the sum of two parts: the search group Patterson function, or

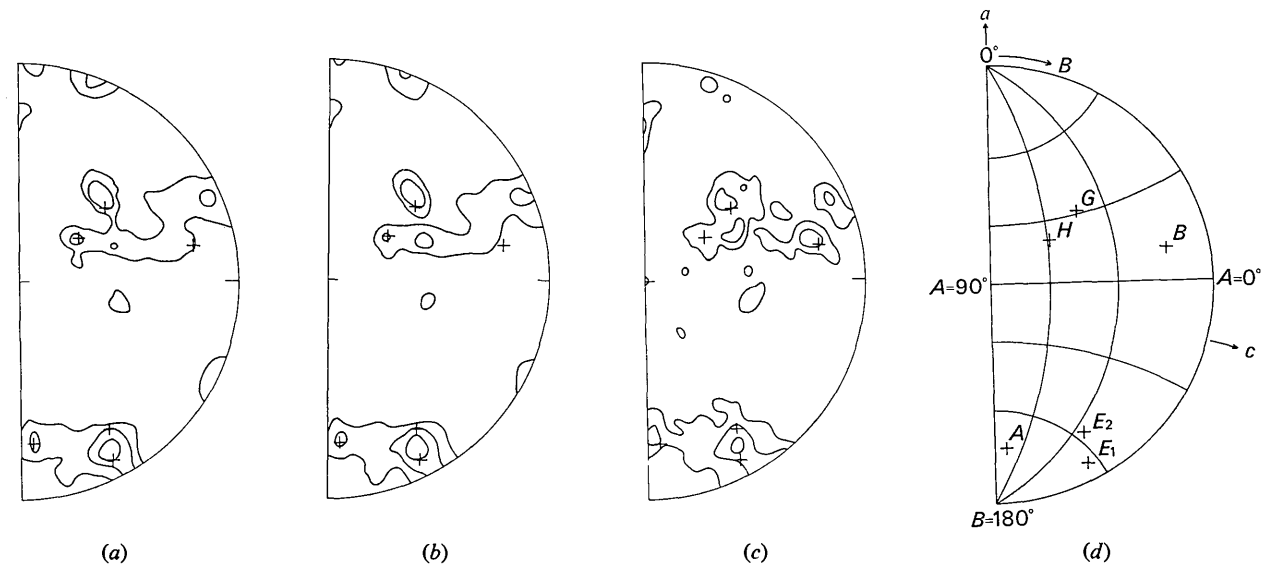

Fig. 2. Rotation (A,B) searches for the directions of the α-helices, calculated with a modified wsum function and averaged over C as explained in the text. Contour levels are at $n\sigma$ above the mean of the function, where $n=1,2,3$. Patterson function resolution, number of search vectors, and assumed helical parameters are: (*a*) 1·5 Å, 66 vectors, $n=3{\cdot}60$, $d=1{\cdot}50$ Å; (*b*) 2·0 Å, 66 vectors, $n=3{\cdot}60$, $d=1{\cdot}50$ Å; (*c*) 1·5 Å, 72 vectors, $n=3{\cdot}72$, $d=1{\cdot}47$ Å. Fig. 2(*d*) is a key to the coordinate system and to the identity of the helices, whose true directions are marked in (*a*)–(*c*).

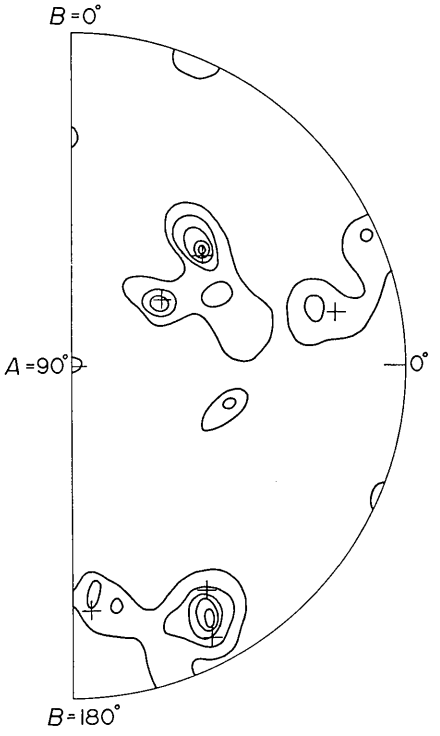

Fig. 3. Rotation search with vectors of a 5-turn α-helical model. The rmin(40,70) function is contoured at values which exceed the mean by $n\sigma$, where $n=1,\ldots,5$.

convolution molecule, and the 'background' Patterson function. If the distribution of 'background' Patterson values were known, the probability that a search vector j is correctly placed at a Patterson function value P could be judged by whether or not $P-w_jP_C$ is a likely 'background' value. The distribution of 'background' Patterson values cannot be known, of course, but if the search group is small compared to the entire structure, the distribution of values at the grid points of the Patterson function itself is a reasonable (although slightly high) approximation. We introduce this distribution by assigning to each value of $P-w_jP_C$ a number proportional to the *rank* of this value among the grid point values of the entire Patterson function. On a scale normalized to a maximum value of 1000, the number 700, for example, is assigned to a $(P-w_jP_C)$ value that exceeds 70% of the values of the Patterson function. When used in this sense, a high rank corresponds to a good fit of the search vector to the Patterson function, a lower rank to a worse one. We now introduce the 'minimum average' feature by defining an image-seeking function rmin(M,N) as the average of the above-defined ranks of the M lowest-ranking $(P-w_jP_C)$ values among those of the N search vectors.

The rmin(M,N) and wsum(N) functions are the two principal image-seeking functions used in this study. Several examples and comparisons of the two are given in the following sections.

Rotation searches

The structure of the search group is specified in terms of a Cartesian coordinate system (X,Y,Z). The Euler angles (A,B,C) represent a rotation of the search group coordinate system from an initial orientation in which, for a monoclinic crystal, $X||\mathbf{a}$, $Y||\mathbf{b}$, and $Z||(\mathbf{a}\times\mathbf{b})$. For our purposes, the Euler angles are clockwise rotations, viewed toward the origin, and carried out in sequence as follows (Schilling, 1970):

A is a turn about the original X axis,
B is a turn about the *rotated* Y axis, and
C is a turn about the *rotated* X axis.

A and B can be visualized as the longitude and latitude representing the direction of the rotated X axis on the surface of a sphere. The (unrotated) monoclinic **a** direction is the 'north' pole ($A=0$, $B=0$), **b** is on the equator ($A=90$, $B=90°$), and **c** is at $A=0$, $B=\beta$.

To a first approximation, the intragroup vector set of the heme group is a circular disk, and that of the α-helix a cylindrical rod. If the approximate symmetry axis of the group is specified to lie along X, the search in C will be relatively insensitive, since it corresponds to rotation about the axis of the disk or cylinder. It is then possible to separate the search for the direction (A,B) of the approximate symmetry axis from the subsequent search for the proper azimuth (C) about this axis. To further reduce the C-dependence in the (A,B) searches, the image-seeking function was roughly averaged over C for each value of A and B, by calculating the mean of five settings separated by 72° in C.

In rotation searches it was generally found that the angular resolution could be increased by giving *long* search vectors increased weight, or otherwise emphasizing their role in the image-seeking function. In the wsum(N) function this added emphasis can be introduced by replacing the overlap weight w with a modified weight w', which includes a dependence on the vector length L. In conjunction with the older α-helix rotation searches, several expressions for w' were qualitatively compared, and the form $w'=(w-2{\cdot}0)(L-5{\cdot}0)$ Å selected. The choice of constants in this expression was not a critical one, however. In the rmin(M,N) function, the overlap weight enters into the expression for the search-group Patterson function, w_jP_C, and a replacement of w with w' would mis-scale w_jP_C relative to P. The emphasis on vector length, radius, *etc.*, in rmin(M,N) rotation searches was therefore introduced in the *selection* of search vectors, rather than in their weighting.

The Cartesian atomic coordinates in α-chlorohemin (Koenig, 1965) provided the geometry of the heme group. Slight adjustments were made to impose strict 4*mm* symmetry and to make the group planar except for the iron atom, which was assumed to be 0·25 Å out of the plane. The assumed group contained 33 atoms, including the 8 carbon atoms α to the heme group. Intraheme vectors were computed to 1·2 Å resolution,

and search vectors selected using a separation parameter of 0·5 Å.

Fig. 1 shows the result of a heme (A,B) rotation search carried out with the 52 vectors for which $(w-3{\cdot}0)L>18$ Å. The rmin$(M,52)$ function was computed at 487 points in A–B space, and consequently involved $52\times487\times5$ evaluations of the Patterson function, done by linear interpolation between the stored values. The CPU time required on an IBM 360/67 computer was 179 sec, or 1·4 ms per vector setting. Several versions of the function, for different choices of M, were computed simultaneously. A subsequent fine search in A and B served to pinpoint the dominant maximum, which unambiguously shows the correct orientation of the heme group normal. In the function shown in Fig. 1, the peak height at the correct (A,B) angles exceeds the mean value of the function by 6·9 times the standard deviation of the function from its mean.

One-dimensional C searches for the correct orientation of the heme group about its correctly oriented normal were considerably less striking. Nevertheless, the highest maximum in both wsum and rmin calculations was found to lie at, or within a few degrees of, the correct C value.

A considerable variety of α-helix searches was carried out, both with the early coarse-grid ($60\times60\times60$ points per cell) Patterson functions, and with the later ($120\times60\times60$ points per cell) version. For reasons of economy, several variations explored with the older Patterson functions and the IBM 7090 computer were not repeated with the later Patterson function and the IBM 360/67. The results presented in the following paragraphs illustrate the main points to be made.

The vector sets for the older α-helix (A,B) searches were calculated for a 4-turn α-helix containing 74 atoms, or 14·8 5-atom residues. The 'standard' α-helical parameters were assumed, *i.e.* $n=3{\cdot}60$ residues per turn and $d=1{\cdot}50$ Å per residue. The values $n=3{\cdot}72$ residues per turn and $d=1{\cdot}47$ Å per residue were also used.

The (A,B) searches based on these vector sets used the wsum image-seeking function averaged over five values in C, 360/5 degrees apart, at each setting in A and B. The above mentioned compromise between vector length, L, and overlap weight, w, was introduced by evaluating the w sum function as $\sum w'P/\sum w'$, where the modified weight w' was chosen as the quantity $(w-2{\cdot}0)$ $(L-5{\cdot}0)$ Å.

Three such (A,B) searches performed with the older, coarse-grid Patterson function are shown in Fig. 2. The result in Fig. 2(*a*) was obtained with the Patterson function computed to 1·5 Å resolution using $D=8{\cdot}0$ Å^2. 66 vectors with $w'>20{\cdot}0$ Å were used. Fig. 2(*b*) shows a similar search with the X-ray data limited to 2·0 Å resolution. The 'standard' α-helix was assumed in both cases. Fig. 2(*c*) shows a search of the 1·5 Å Patterson function with 72 vectors ($w'>20{\cdot}0$ Å) based on the 'non-standard' α-helix with $n=3{\cdot}72$ and $d=1{\cdot}47$. The true helical directions are indicated with crosses in the contour maps and are identified in Fig. 2(*d*). Helix E has a 7° bend in the middle; the directions of the two halves are shown separately.

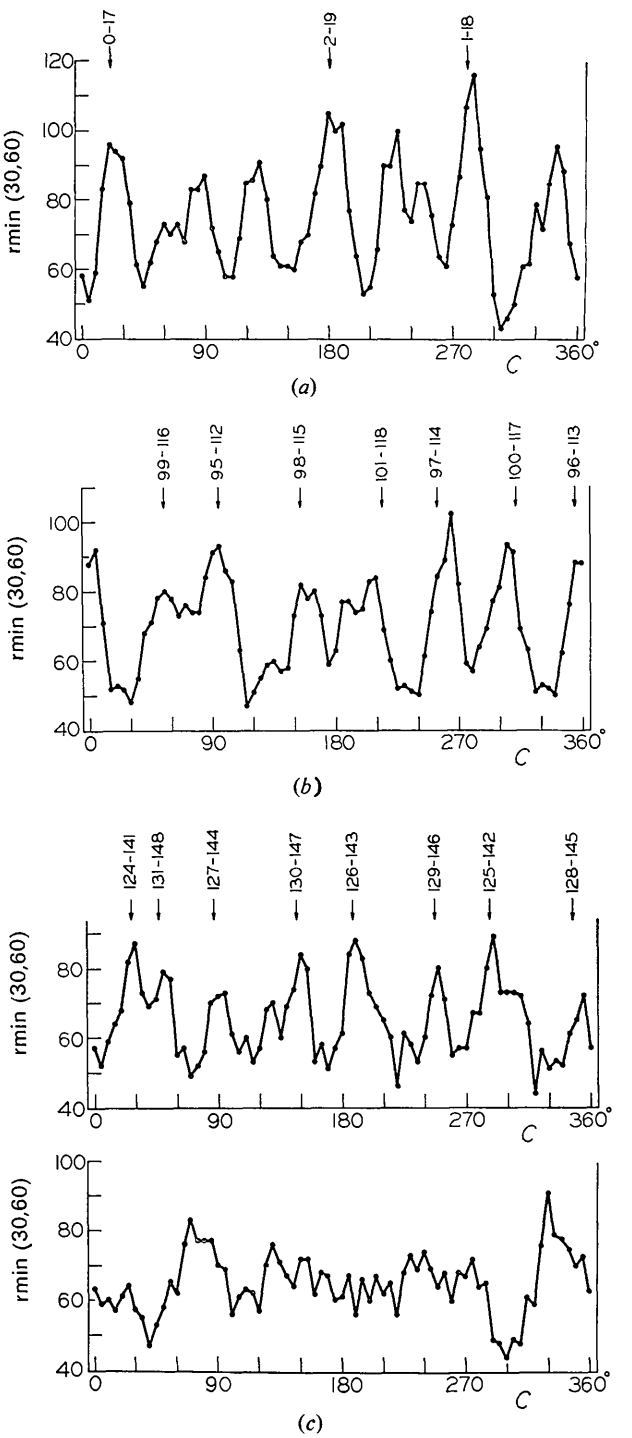

Fig. 4. Results of searches in the angular coordinate (C) about the correctly oriented axis of the α-helix. (*a*) and (*b*) represent helices A and G pointed in the true direction, (*c*) represents helix H with the model helix correctly pointed (top), and turned end for end (bottom). Arrows indicate the true positions of best fit calculated from the atom coordinates.

Fig. 2(*a*) clearly shows the correct orientations of the axes of four of the five prominent helices in the molecule, namely *A*, *E*, *G*, and *H*. These helices contain respectively 16, 20, 19, and 26 residues, and all conform closely to the 'standard' α-helical values of *n* and *d* (Watson, 1969). The similarity between Fig. 2(*a*) and the only slightly more ambiguous Fig. 2(*b*) suggests that the directions of 4-turn or longer α-helices should be recoverable from somewhat less well-resolved Patterson functions than those used here.

The fifth prominent helix in the myoglobin molecule is *B*, with 16 residues. The true values of *n* and *d* for this helix are 3·72 and 1·47 respectively (Kendrew, 1962; Watson, 1969). It is interesting to note that the peak indicating helix *B* is sharply enhanced in Fig. 2(*c*), while the peaks of the four 'standard' helices are reduced in height.

For the searches of the more recent, $120 \times 60 \times 60$ point per cell Patterson function, a 'standard' 5-turn α-helical segment of 90 atoms, or 18 residues, was used. A separation parameter of 0·6 Å was used to eliminate close overlaps among the initial set of 4005 intragroup vectors. For the (A,B) rotation searches, the search vectors were chosen as those having the highest values of $(w-3{\cdot}0)\ X$, where X is the component of the vector along the axis of the helix.

Searches for the directions of the helices were made using the rmin function, and averaging over *C*, as before, by a 5-step rotation. An example is shown in Fig. 3. This search was made with 70 vectors for which $(w-3{\cdot}0)X > 25$ Å. The weights, *w*, of these vectors ranged from 17·5 to 5·1, and their lengths from 4·8 to 18·4 Å. The steps in *A* and *B* were taken as 5° over most of the angular range; near the poles the steps in *A* were larger. Regions where the function exceeded its mean value by 1·5 σ were recalculated on a finer grid.

Fig. 3 is a clear improvement over Fig. 2(*a*). One reason for this is the denser Patterson grid; however, it is also likely that the rmin (40,70) image-seeking function is inherently better than the modified wsum(66) function. This point is discussed more fully below.

There is no obvious interpretation of any of the

Table 1. *Comparison of the image-seeking functions* wsum(N) *and* rmin(N,N) *for translation searches*

Search vectors	N	Weight cutoff	(True peak-mean)/σ wsum(N)	(True peak-mean)/σ rmin(N,N)	False peaks higher than true* wsum(N)	False peaks higher than true* rmin(N,N)
Heme–heme (2_1)	50	3·10	4·58	4·29	0	0
Heme–heme (2_1)	75	2·52	4·45	3·51		
A–*A* (2_1)	300	3·56	2·58	3·48	3	0
A–*A* (2_1)	400	3·23		3·76		
G–*G* (2_1)	200	3·21	3·61	3·88	0	0
G–*G* (2_1)	300	2·91	3·60	3·66	0	
H–*H* (2_1)	200	3·21	3·04	3·71	2	2
H–*H* (2_1)	300	2·99	2·58	2·77	7	
(2 heme Fe)–*H*	180		5·66	4·94	0	0

* Missing entries signify that an exhaustive peak search was not made.

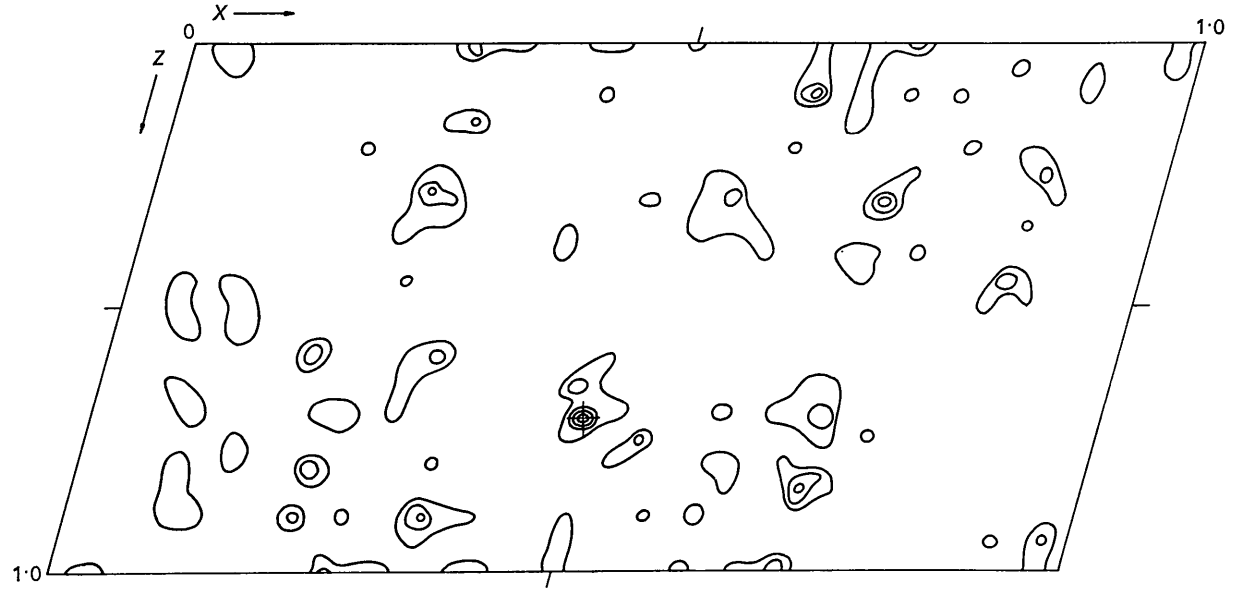

Fig. 5. Translation search for the *x* and *z* coordinates of the heme group. Intergroup vectors between two screw-axis related heme groups were used. The wsum(50) image-seeking function is contoured at values exceeding the mean by $n\sigma$, where $n = 1, \ldots, 4$. The true position is indicated at (0·49, 0·71) and corresponds to the highest peak in the map.

lower peaks, some of which reach values of mean+2σ in the image-seeking functions. It is possible that the mean+2·2σ peak below the center of Fig. 3 represents helix *D* (7 residues), whose true direction ($A=38$, $B=93°$, rmin = mean+0·5σ) is only 8° from the peak maximum. On the other hand, the two other minor helices, *C* (7 residues, $A=80$, $B=78°$) and *F* (10 residues, $A=7$, $B=108°$) lie at rmin values of only mean−0·3σ and mean+0·0σ respectively, and are far removed from any significant maxima in rmin.

It should be pointed out that the quadrant (A,B) of the sphere, shown in Figs. 2 and 3, and its 'antipode' $(180°+A, 180°-B)$ do not represent identical searches when the search group is totally asymmetric. They represent an end-for-end rotation of the helix. In the 'cylindrical' approximation to the vector set, which is implied by the averaging over *C*, no significant differences were found between corresponding peak heights in the two quadrants. Accordingly, only one of the two quadrants is shown.

Searches in *C* at the peak values of both (A,B) and $(180°+A, 180°-B)$ of helices *A*, *G*, and *H* were carried out in order to establish the correct orientation about the helix axis, and the end-for-end direction of the

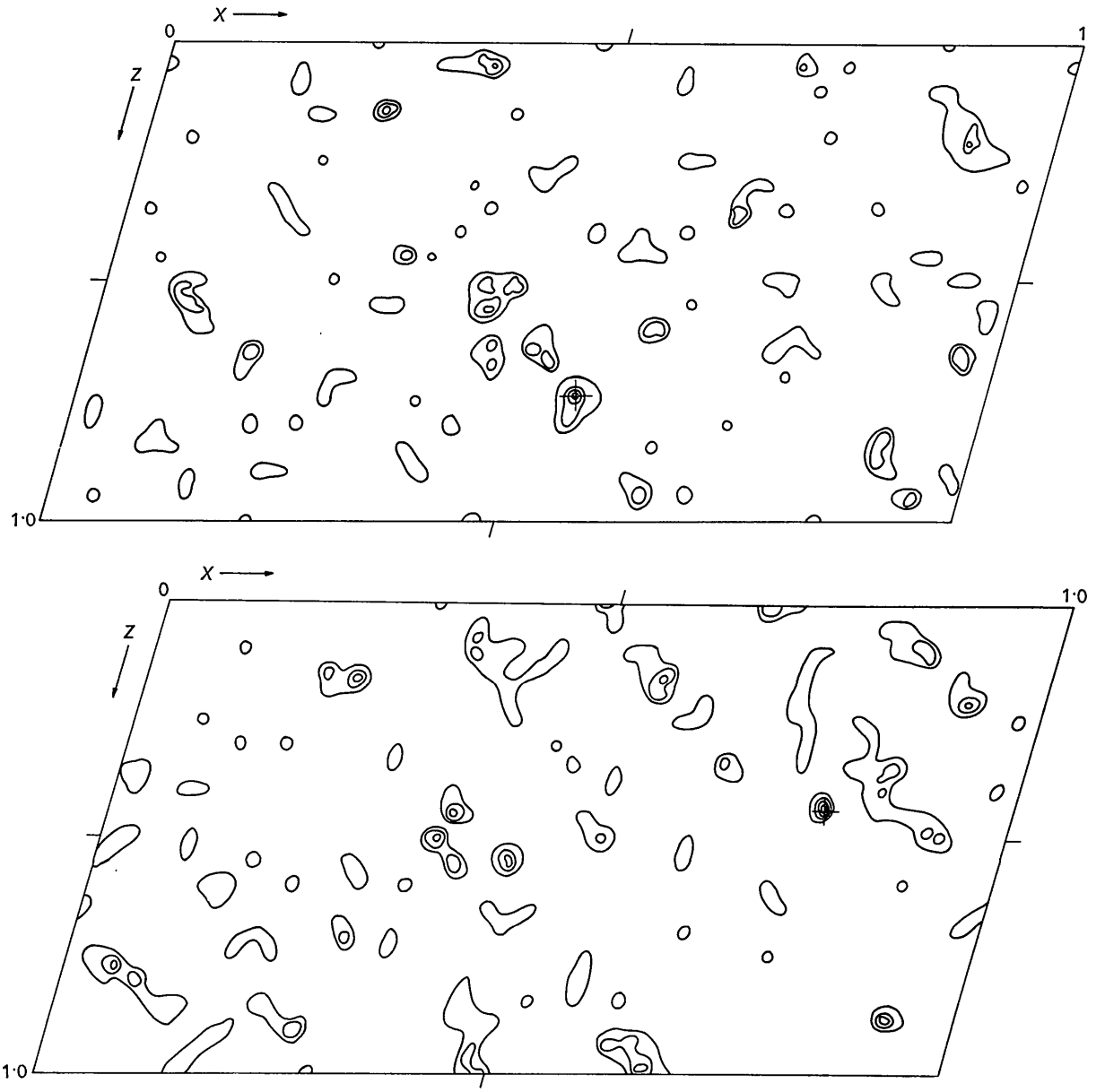

Fig. 6. Translation searches for the *G* helix (top) and the *H* helix (bottom), each performed with sets of vectors between helices related by the 2_1 axis. The rmin(150,200) function is contoured at values exceeding the mean by $n\sigma$, where $n=1, \ldots, 4$. The true positions are indicated by crosses, and correspond to the highest peak in each case.

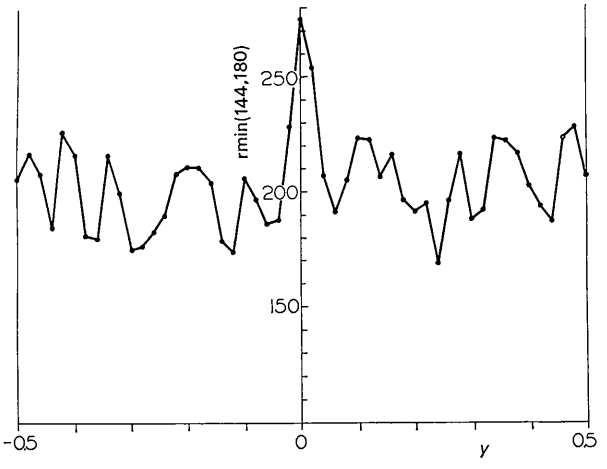

Fig. 7. One-dimensional search for the relative y coordinate of the heme group and helix H, both correctly placed in x and z. The 180 vectors from the two Fe atoms to the 90-atom helical model were used. The vector set was arbitrarily generated so that the true y value is zero.

helix. Vectors for these searches were selected so as to maximize the quantity $(w-3{\cdot}0)XR$, where R is the radial and X, as before, is the axial component of the vector relative to the helix axis.

Fig. 4 shows the results for the three helices correctly 'aimed' and for helix H turned end-for-end. The A helix consists of 16 residues (3–18), the G helix of 19 residues (100–118), and the H helix of 24 residues (125–148) (Kendrew, 1962; Watson, 1969). There can be no unique way to fit the 18-residue model helix to these units. Instead, we expect to find a sequence of several good coincidences separated by $360°/n \simeq 100°$. The true C values corresponding to these coincidences were, for each helix, determined by least-squares fitting of the model helix to the atomic coordinates of the helix in the myoglobin structure. These calculated C values are marked in the upper margins of the plots. Other acceptable positions, 100° apart, are also indicated. The accompanying numbers designate the residues in the myoglobin chain that 'match' the 18-residue model helix at the particular setting. It is understood that the model helix overhangs the end of the myoglobin helix, whenever the residue sequences in Fig. 4 extend outside the helical regions given above. It can be seen from the plots that the maxima of the rmin(30,60) function match the true positions very well.

The choice of direction along the helix axis is somewhat less clear-cut, and some 100° periodicity was found in all three C searches performed with the model helix pointing in the wrong direction. The result for helix H is shown in Fig. 4(*c*). The pattern of maxima separated by 100° is less prevalent than in the correctly pointed search. It is quite possible that the distinction could be improved by selecting search vectors so as to preferentially include vectors that involve the oxygen and beta-carbon atoms. It is also quite possible that a mathematical analysis of data such as those in Fig. 4 could provide a more reliable criterion for judging the presence or absence of a cyclic repeat, and perhaps even yield a 'best' value for the corresponding ΔC. Neither of these possibilities was pursued further, however.

Translation searches

All translation searches described here were done with the 1·5 Å 'point-atom' ($D=0$) Patterson function computed on a grid of $120\times60\times60$ points per cell. Vector sets were computed to 1·2 Å resolution; the separation parameter used in selecting search vectors was 0·5 Å for heme–heme vectors, 0·6 Å for others.

A two-dimensional heme–heme translation search is shown in Fig. 5. The search was carried out with intergroup vectors from one correctly oriented 'reference' heme group to its screw-axis related mate. The coordinates of the search map refer to the displacement of the search vector set, that is, a peak at (x,z) implies that the 'reference' group is to be translated by $(-x/2,-z/2)$ from the position assigned to it in the vector calculation. In this case, the vector set was calculated with the center of the heme group located on the 2_1 axis; hence, the peak at (0·49, 0·71) in the search map correctly gives the location of the heme group center as $x=-0{\cdot}245$, $z=-0{\cdot}355$.

Similar two-dimensional translation searches were made using vectors between pairs of screw-axis related 18-residue segments of α-helix, properly oriented to match helices A, G, and H.

The interhelix vector set between two non-parallel α-helices is much more 'diffuse' than either an intrahelix or a heme–heme vector set. That is to say that extremely high overlap weights are less likely to occur in an interhelix vector set, and that the set occupies a

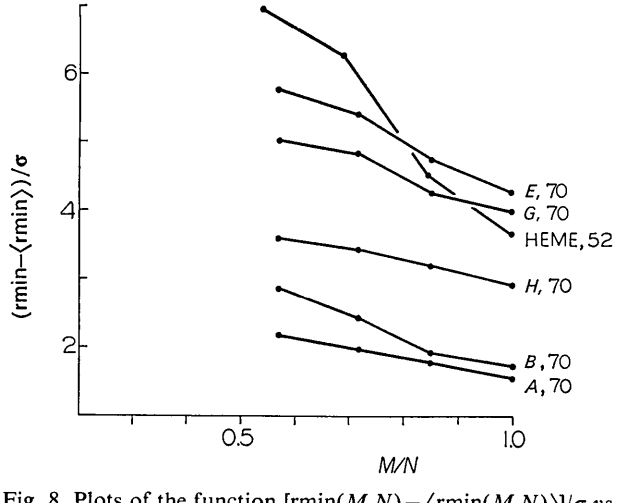

Fig. 8. Plots of the function $[\mathrm{rmin}(M,N)-\langle\mathrm{rmin}(M,N)\rangle]/\sigma$ *vs.* the ratio M/N for rotation searches. For each curve the value of $\mathrm{rmin}(M,N)$ is the peak value at the true (A,B) setting for the search group in question. The quantity $\langle\mathrm{rmin}(M,N)\rangle$ is the mean of $\mathrm{rmin}(M,N)$ taken over all directions (A,B), and σ is the standard deviation from this mean. The total number of vectors N used in each search is shown on the right.

larger volume of Patterson space. It is then reasonable to expect that a larger number of search vectors will be needed to adequately portray the cross-Patterson function, assuming that the minimum separation between the selected search vectors is constant.

Fig. 6 shows two of the α-helix translation searches. In both cases, the highest peak in the search function falls at, or very near, the true position. The third α-helix translation search was similarly unambiguous.

By means of a three-dimensional search it should be possible to find the positions, relative to one another, of two correctly oriented, symmetrically unrelated groups. A partial search of this kind was done with the 180 vectors from the two heme iron atoms to a 90-atom model helix oriented as helix *H*. The calculation was performed, not as a complete three-dimensional search, but as a set of five one-dimensional searches in y at discrete values of x and z. If the x and z components of the vector joining the origins of the heme group and the *H* helix are correctly assigned, (*i.e.*, taken as corresponding to the highest peaks in Figs. 5 and 6), then the y search of Fig. 7 is obtained. Here, the vector set was arbitrarily generated so that $y=0$ corresponds to the true position. The fact that a strong maximum occurs at $y=0$ shows that the relative y coordinates of the two groups can be found in this way.

Four more y searches were done using (x,z) values indicated by lower (false) peaks in Figs. 5 and 6. None of these exhibited a clear maximum, and none had a value of rmin(144, 180) higher than 235 on the scale of Fig. 7. We conclude, then, that such one-dimensional y searches could be used not only to find relative y coordinates, but also to resolve ambiguities that might exist in the (x,z) searches.

Discussion

In the course of this work a large number of searches were calculated, covering a variety of choices of parameters. Many of these searches were less successful than the examples shown above. In this section we discuss some criteria which seem to determine the likelihood of success.

It appears that for any given search group the rotation search tends to be safer than the translation searches, that is, less likely to give an ambiguous or false result. There are two reasons for this: differences between the group and the optimally-fitted model are amplified in the intergroup vector set, because the errors in the x and z components of vectors between screw-axis related atoms are twice the errors in the corresponding atomic coordinates. Secondly, as was pointed out above, the intragroup vector sets tend to contain more vectors of very high weight.

Table 1 and Figs. 8 and 9 present summaries of obtained results. As the basis for comparison we have used the number of standard deviations by which the peak representing the true solution exceeds the mean value of the search function.

Table 1 presents a comparison between the image-seeking functions wsum(N) and rmin(N,N). The results suggest that for a given set of N search vectors rmin(N,N) is a somewhat better image-seeking function than wsum(N), but the difference is not great. There is also a suggestion that nothing is gained by including search vectors having weights less than about 3·2 carbon–carbon vectors, and that inclusion of too many weak vectors might even be harmful.

Figs. 8 and 9 demonstrate the advantage of using rmin(M,N) with the 'minimum average' feature included, *i.e.*, with the execution-time selection of the M lowest values of $P_j - w_j P_C$ for inclusion in the image-seeking function. In every case, an improvement can be achieved by choosing M less than N. It appears, then, that the best choice of image-seeking function is rmin(M,N), where M is 50–80% of the value of N.

In summary, the results of this work suggest that Patterson-space search techniques could be useful in the crystallography of biological macromolecules, namely for purposes such as: (1) to detect the presence, to determine the orientation and, in favorable cases, the position relative to a symmetry element of helices and other known or assumed relatively large substructures; (2) in the presence of a heavy atom, to determine the position of any such oriented substructure relative to the heavy atom; (3) to determine the orientation and position of any heavy-atom containing rigid group, regardless of size.

The author is indebted to Dr H. C. Watson for the myoglobin X-ray data, to Dr C. M. Venkatachalam for the atom coordinates of the model helices, and to Drs J. W. Schilling and R. Hoge for many helpful discussions. Support of this investigation under grants HE-08612 and GM-15259 from the National Institutes of Health is gratefully acknowledged.

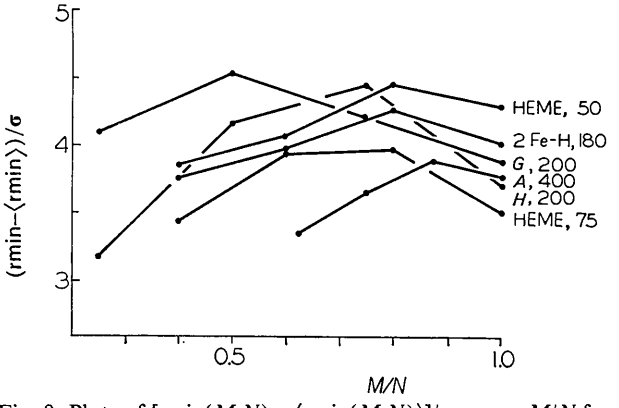

Fig. 9. Plots of $[\text{rmin}(M,N) - \langle\text{rmin}(M,N)\rangle]/\sigma$ versus M/N for translation searches. Here, rmin(M,N) is the peak value at the true (x,z) setting for each interhelix or interheme vector set (A, G, H, and HEME), or the peak value at the true y setting for the vector set 2Fe-helix *H*. The means of rmin (M,N) are taken over all (x,z) respective y values of the search in question, and σ is the corresponding standard deviation from the mean. The total number of vectors N used in each search is shown on the right.

References

BRAUN, P. B., HORNSTRA, J. & LEENHOUTS, J. I. (1969). *Philips Res. Rep.* **24**, 85.
CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**,544.
HORNSTRA, J. (1970). *Crystallographic Computing*, p. 103. Edited by F. R. AHMED. Copenhagen: Munksgaard.
HUBER, R. (1965). *Acta Cryst.* **19**, 353.
KENDREW, J. C. (1962). *Brookhaven Symp. Biol.* **15**, 216.
KOENIG, D. F. (1965). *Acta Cryst.* **18**, 663.
LATTMAN, E. E. & LOVE, W. E. (1970). *Acta Cryst.* B**26**, 1854.
NORDMAN, C. E. & NAKATSU, K. (1963). *J. Amer. Chem. Soc.* **85**, 353.
NORDMAN, C. E. & SCHILLING, J. W. (1970). *Crystallographic Computing*, p. 110. Edited by F. R. AHMED. Copenhagen: Munksgaard.
ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24.
ROSSMANN, M. G., BLOW, D. M., HARDING, M. M. & COLLER, E. (1964). *Acta Cryst.* **17**, 338.
SCHILLING, J. W. (1970). *Crystallographic Computing*, p. 115. Edited by F. R. AHMED. Copenhagen: Munksgaard.
SPARKS, R. A. (1961). *Abstr. Amer. Cryst. Assoc. Meeting*, p. 37.
STEWART, J. M., KUNDELL, F. A. & BALDWIN, J. C. (1970). *The X-ray System of Crystallographic Programs*. Univ. of Maryland, College Park, Maryland.
TOLLIN, P. (1966). *Acta Cryst.* **21**, 613.
TOLLIN, P. (1969). *J. Mol. Biol.* **45**, 481.
TOLLIN, P. & COCHRAN, W. (1964). *Acta Cryst.* **17**, 1322.
WATSON, H. C. (1969). *Prog. Stereochem.* **4**, 299.
ZWICK, M. (1969). *Abstr. Amer. Cryst. Assoc. Winter Meeting*, p. 74.

Acta Cryst. (1972). A**28**, 143

Choice of Scans in X-ray Diffraction

BY S. A. WERNER

Scientific Research Staff, Ford Motor Company, Dearborn, Michigan, U.S.A. and Department of Nuclear Engineering, University of Michigan, Ann Arbor, Michigan, U.S.A.

(*Received* 20 *September* 1971)

In a recent paper (Werner, S. A. *Acta Cryst.* (1971). A**27**, 665) it was pointed out that the choice of scans in a neutron diffraction experiment should be based on the criterion that the diffracted beam enters the detector on its centerline for each angular setting of the crystal. The same criterion should be applied in X-ray diffraction. Since the spectral distribution of a source of X-rays and neutrons is quite different, conclusions regarding the optimum coupling between the detector and crystal motions are different in these two cases. In this paper, formulas are derived (within the framework of certain gaussian approximations) for the optimum scanning ratio g in equatorial plane X-ray diffraction experiments on single crystals. For the case when a monochromator is not used, g is independent of scattering angle $2\theta_B$ for a large range of instrumental parameters and Bragg angles θ_B. It is found that a θ–2θ scan is essentially never advisable. An expression for g is derived for the case when a planar monochromator is used in symmetric Bragg reflection. The optimum scan is found to depend on the scattering angle, but not in such a marked way as in the neutron case. Coupling the detector and crystal motions in the manner suggested allows one to decrease the acceptance aperture to its minimum width, thus keeping the background due to thermal diffuse scattering (TDS) and incoherent scattering as low as possible.

I. Introduction

The problem of the selection of scans in single crystal neutron diffraction experiments was discussed in a recent paper (Werner, 1971). The choice of scans in X-ray diffraction requires special attention because the geometry and spectral distribution of a source of X-rays and neutrons are quite different. Equatorial plane X-ray diffraction (like neutron diffraction) experiments on single crystals are generally carried out using either an ω-scan (crystal rotating, detector fixed) or a θ-2θ scan (detector coupled 2:1 to the crystal). In view of the widespread use of instrumentation involving tape-controlled and computer-controlled diffractometers, restricting the scanning of Bragg reflections to these two modes is not a necessary constraint. The purpose of this paper is to examine the question of whether there is in general a better way to scan Bragg reflections in X-ray diffraction.

Over the years a number of papers have been published on the theory of measuring integrated intensities and on the various geometrical considerations necessary in X-ray diffraction experiments on single crystals [see for example Alexander & Smith (1962), Burbank (1964), Ladell & Spielberg (1966)]. A summary of the results of these papers regarding the necessary size of the receiving aperture and the range of scan is given in the book by Arndt & Willis (1966). However, the analysis given in these papers does not permit the experimentalist to readily make a decision on the optimum coupling ratio between the detector and crystal motions. The intent of this paper is to derive an