Acta Cryst. (1998). D**54**, 324–333

Molecular Replacement Using DNA Helical Symmetry

IGOR BAIKALOV AND RICHARD E. DICKERSON *

Molecular Biology Institute, University of California, Los Angeles, CA 90095–1570, USA. E-mail: red@mbi.ucla.edu

(Received 19 September 1996; accepted 23 July 1997)

Abstract

The efficiency of molecular-replacement methods in the structure analysis of B-DNA is markedly increased if a knowledge of the structural properties and helical symmetry of B-DNA is incorporated into molecular-replacement procedures. The separation of the most significant or most robust parameters, such as the location of helices in the unit cell, from the less well defined parameters, such as rotation around the helix axis, further improves the reliability of molecular replacement and avoids frameshift errors in the positioning of the model. This approach has been applied successfully to solve novel structures of four B-DNA decamers in various space groups.

1. Introduction

Molecular replacement is a powerful alternative to multiple isomorphous replacement for solving macromolecular structures, when heavy-atom derivatives are not available but when an appreciable portion of the desired structure is known in advance. Conformational flexibility of double-helical DNA oligomers, when even minor modification of the chemical structure often results in a change of crystal packing, makes it difficult to obtain isomorphous heavy-atom derivatives. On the other hand, knowledge of the general helix type allows one to use an idealized helix as a search model in molecular replacement, greatly advancing the process of structure determination. A trivial case of molecular replacement is represented by isomorphous structures, where the placement of molecules in the unit cell is the same. Refining the structure to account for local deviations from the model then concludes the solution. Thus, more than 50 B-DNA dodecamer crystal structures in orthorhombic space group $P2_12_12_1$, deposited to date with the Brookhaven Protein Data Bank, are isomorphous with the first one, C-G-C-G-A-A-T-T-C-G-C-G, which was solved by multiple isomorphous replacement (Drew *et al.*, 1981).

Application of molecular replacement methods in cases where the position and orientation of the helix in the unit cell are unknown is less straightforward. The conventional approach in protein crystallography separates rotation and translation of the molecule, so as to reduce a six-dimensional search (three rotation angles and three translations) to two consecutive three-dimensional searches. Both searches, rotation and translation, can utilize either real space (Huber, 1965; Crowther & Blow, 1967) or reciprocal space (Rossmann & Blow, 1962; Fujinaga & Read, 1987). The success of the translation search depends on the accuracy of the prior rotational solution, while the latter is primarily determined by the quality of the search model.

DNA helical symmetry presents a major obstacle for conventional methods of molecular replacement. At low resolution, typically used because of the low quality of the search model, base pairs of the double helix are practically indistinguishable. Therefore, the rotation function of B-DNA with ten base pairs per turn exhibits an almost perfect tenfold symmetry in the direction of the helix axis. In addition, these multiple solutions are duplicated by the approximate twofold symmetry between two strands of the DNA. The large number of possible rotational solutions makes the subsequent translation search extremely laborious and unreliable.

In a crystal structure built from B-DNA decamers stacked so as to approximate a continuous helix, two of the three rotational variables define the orientation of helix rods through the crystal, and the third describes rotation around the rod axis. The helix axis direction and hence the first two rotational variables usually can be established at the outset by observing a cluster of strong reflections at a spacing of 3.4 Å in the diffraction pattern, arising from base-pair stacking. The third rotational variable, in contrast, is weakly determined, and a search for it is better left until after a translation search. In a similar manner, two of the translation variables are strongly determined; those that define the lateral packing of helical rods in the crystal. These variables can be established by analytical search functions as described below.

The third rotation and third translation are best considered together, after the semi-continuous helical rods have been oriented and positioned correctly. The third rotation and translation are related by helical symmetry. If the helix was truly continuous, instead of being built from stacked duplexes, then a rotation of 36° would be equivalent to a translation of 3.4 Å (idealized values). Indeed, it is more efficient not to consider them as separate variables, but rather to consider first the orientation of the helix backbone around the rod axis (a

ISSN 0907-4449

series of discontinuous rotation/translation peaks in a search function), and lastly the breaks between one helix and the next, expressed as absent phosphate groups. This latter may be termed the frameshift problem, and can be the most difficult decision of all.

Local deviation of a structure from an idealized helix, *e.g.* the 50° twist seen at some positions in the monoclinic family of decamers (Privé *et al.*, 1991) compared with the average 36° for B-DNA, can overwhelm the frameshift issue. Preference given to one of the highest scored solutions in a full search, without considering other possible solutions, can lead easily to a frameshift error, or a mistake in locating the gaps between helices. This type of difficulty has been reported for almost every new B-DNA decamer packing mode (Heinemann & Alings, 1991; Heinemann *et al.*, 1992; Lipanov *et al.*, 1993; Goodsell *et al.*, 1993).

The most powerful clue in solving the structures of B-DNA oligomers in various space groups has been the occurrence in the diffraction pattern of clusters of especially strong reflections at a distance corresponding to 3.34 Å. Such clusters are a consequence of the stacking of base pairs atop one another along the B-DNA helix and their position reveals the direction of the helix axis. This observation significantly simplifies the molecular replacement problem, since one-dimensional rotation search around the helix axis is easily interpretable and helical redundancies can be accounted for (Grzeskowiak *et al.*, 1991). Thus, the same helical symmetry, which complicates a conventional approach to molecular replacement, can be very helpful when exploited properly.

Using all available information to reduce the number of unknowns is certainly the major step to success in molecular replacement. But even more important is the order in which remaining unknowns are searched. Zooming into the structure, going from coarse positioning to fine details, helps to avoid both false minima and ambiguity in the solution. The applications of the molecular-replacement method discussed in this paper have been developed to solve novel structures of the B-DNA decamers, pursuing as a general goal an optimal strategy to exploit helical symmetry in the DNA structure.

2. The trigonal decamers

The decamer C-G-A-T-C-G-A-T-C-G, denoted for brevity as KK after its primary investigators Kazunori Yanagi and Kazimierz Grzeskowiak, adopts a snugly packed orthorhombic cell, space group $P2_12_12_1$, with rods of stacked helices parallel to the 33.3 Å *c* axis (Grzeskowiak *et al.*, 1991). But methylation of the 6-amine position of adenine 7 causes C-G-A-T-C-G-^{6me}A-T-C-G (the KKMe decamer) to adopt an entirely different packing mode: trigonal space group $P3_221$, with cell dimensions 33.4 × 33.4 × 98.3 Å, γ = 120° (Baikalov *et al.*, 1993). The decamer C-C-A-A-C-I-T-T-G-G when crystallized in the presence of calcium ions (the CICa decamer) occupies monoclinic space group $C2$, again with helical rods parallel to the 34.2 Å *c* axis. But the identical base sequence, if crystallized in the presence of magnesium ions (the CIMg structure) adopts trigonal space group $P3_221$, with cell dimensions 33.2 × 33.2 × 94.8 Å, γ = 120° (Lipanov *et al.*, 1993). For both KKMe and CIMg, strong clusters of 3.4 Å reflections halfway between the a^* and b^* reciprocal axes indicate that columns of helices run parallel to the real-space axes a, b, and $-(a + b)$. Hence, the two rotational variables that define the orientation of the helix axis can be eliminated from the search at the outset.

For the CIMg structure, the remaining unknown rotation about the helix axis and the three positional coordinates of the helix in the unit cell were first sought in a traditional manner, using standard rotation/translation techniques. A direct rotational search yielded five peaks separated by 36°, reflecting the tenfold symmetry of B-DNA decamer (see Fig. 1 of Lipanov *et al.*, 1993). The other five peaks are related by a twofold crystallographic dyad parallel to the helix axis. Although the best rotation solution yielded a strong peak in a three-dimensional translation search (Fig. 2 of Lipanov *et al.*, 1993), subsequent refinement failed in both space groups $P3_221$ and its enantiomorph $P3_121$ because of unacceptable intermolecular contacts. Furthermore, attempts to repeat the translation search using other rotation solutions led to a morass of multiple solutions. A different approach to solving these trigonal structures was needed.

2.1. *Zooming in: position the rods first*

Given that the decamer helices are stacked into semi-continuous rods of known orientation through the crystal (fixing two rotational variables), the most important variables to determine first are the two translations that describe the lateral packing of rods within the crystal. The final rotation and translation describe the motion of one rod about its own axis and the locations of breaks between stacked helices, and are less relevant at low resolution. Hence, it seems strategically desirable to solve for the packing of rods first, and to worry about axial rotation and translation of the rods later.

Space-group symmetry demands a packing of rods like that in Fig. 1. Six layers of rods are present at spacing of roughly 1/6 of the *c* axis (perpendicular to the plane of the page). The top layer 7 is related by a unit-cell translation along *c* to a bottom layer 1 (not shown). Twofold axes run along a at c = 1/6 (relating rods 2 and 3) and 4/6, along $-(a + b)$ at c = 0 and 3/6 (relating rods 4 and 5), and along b at c = 2/6 and 5/6 (relating rods 6 and 7). Helical rods 6 and 7′ (not 7) are

related by a twofold screw axis parallel to b at (a, c) = (1/2, 5/6). Operation of the threefold screw axis along c turns layer 2 into layer 4 and then 6, and layer 3 into 5 and then 7. In contrast, layers 1 and 2 cross at 120° angles, as do layers 3 and 4, and layers 5 and 6. For rod 6 in Fig. 1 as an example, the two translational variables that must be established are: (a) the displacement of the rod from the b axis in the direction of the dotted line, and (b) its exact height along the c axis perpendicular to the plane of the figure.

These are just the coordinates of rod 6 in a plane perpendicular to the b axis through the dotted line in Fig. 1, and can be obtained from that section of the three-dimensional Patterson map utilizing helical symmetry. Approximate tenfold-screw symmetry, exhibited by the helical rods at low resolution, can be reduced to a twofold screw symmetry for our purposes. A combination of the non-crystallographic and the crystallographic twofold screw axis (relating rods 6 and 7′) gives rise to a peak in the Patterson map, similar to a combination of twofold diads observed in protein crystallography (Epp *et al.*, 1971). Since two screw translations cancel each other, this peak lies in a plane perpendicular to the symmetry axis, and its position is determined by the pseudo-translation between rods 6 and 7′.

Fig. 2 shows a section of the Patterson map of KKMe perpendicular to the b axis. A simple self-Patterson peak at the origin is overlaid and obscured by a set of fringes that arises from sampling the Patterson map of helices 2, 3, 4 and 5 at other than right angles to their axes. The dotted line in Fig. 1 crosses helices 2–5 at an angle of 30° to their axes, and the horizontal spacing between fringes in Fig. 2, as expected, is the projection onto the Patterson plane of the spacing between base pairs, 3.34 Å/cos(30°) = 3.86 Å. Such fringes have been

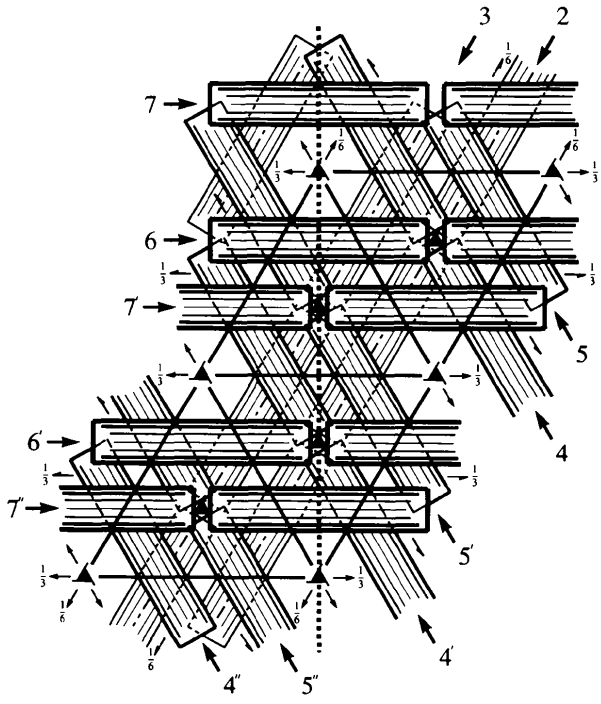

Fig. 1. Packing of helices in the trigonal $P3_221$ cell used by C-G-A-T-C-G-^{6me}A-T-C-G, and by C-C-A-A-C-I-T-T-G-G when crystallized with Mg^{2+} counterions. The origin is at the upper left threefold-screw symmetry axis sign, crossed by the vertical dotted line. Axis a extends diagonally down and to the left from the origin; axis b is horizontal to the right; axis c rises perpendicular to the plane of the page. Two unit cells are shown along the a axis. Semi-continuous rods of stacked decamer helices run in three directions parallel to the ab plane of the unit cell: along a, along b, and along $-(a + b)$ (diagonally upward to the left). Layers 2 and 3 are parallel although at different c levels (roughly 1/12 and 3/12), as are layers 4 and 5 (5/12 and 7/12), and layers 6 and 7 (9/12 and 11/12). For clarity, helices are drawn with only one third their true diameter.

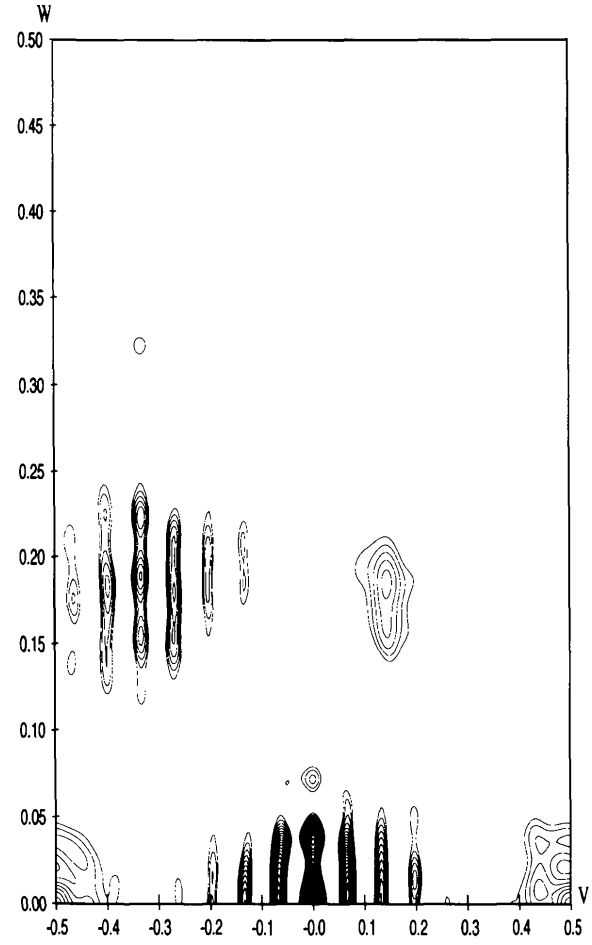

Fig. 2. Patterson map section through the origin, perpendicular to the b axis, for C-G-A-T-C-G-^{6me}A-T-C-G (the KKMe helix). The vertical axis is parallel to c. The horizontal axis extends along the dotted line in across two unit cells, from the uppermost threefold symbol to the lower, a distance of: $a(3)^{1/2}$ = 57.8 Å. This axis is plotted from −0.5 at the left, through the origin at center, to +0.5 at the right. Resolution ranges from 8.0 to 3.0 Å; map is contoured at $\sigma/3$.

encountered only in trigonal space groups, and are absent when this Patterson section method is used with monoclinic or orthorhombic space groups.

To right and left at ±0.50 along the horizontal axis are the compound peaks representing vectors between helices 6 and 6′, or 7 and 7′. These pairs of helices are inverted relative to one another because of the half-cell displacement along *b* that results from the 120° angle at γ. Wherever the major groove faces up (toward the viewer in Fig. 1) along helix 6, it faces down (into the page) in helix 6′ at the same location measured relative to the dotted axis. The two rods cannot be mapped onto one another by a simple lateral translation parallel to the dotted line. Hence, the collection of Patterson vectors between these inverted helices is the complex constellation seen at the lower right and left of Fig. 2. These features are not overlaid by fringes because no sets of helices in layers 2–5 are separated by half the total horizontal displacement.

The single peak at (0.145, 0.185) establishes the separation of rods 6 and 7′ (not 7) in Fig. 1. Rods 7′ and 7 have inverted orientation because of the half-cell displacement just mentioned. Rods 7 and 6 also are inverted by the action of the twofold axis along *b* that relates them. As a consequence, for rods 7′ and 6 the two effects cancel, and they exhibit the same orientation. If the absence of two phosphates at each break between decamer helices is ignored, then rod 6 can be mapped onto rod 7′ by a simple lateral translation along the dotted line of Fig. 1, followed by a rise along *c*. Indeed, these are precisely the two translations whose values are being sought from the Patterson map. The identity of orientation of helices 6 and 7′ means that Patterson vectors pile up to produce one single peak, visible at (0.145, 0.185). To the left of this single peak is yet another set of fringes arising from Patterson peaks between rods 2 *versus* 3, and 4 *versus* 5, inclined to the Patterson plane. These fringes mask the compound peak arising from vectors between inverted rods 6 and 7.

This interpretation of the perpendicular Patterson section has been tested by calculations using model helices positioned properly in the trigonal cell: (*a*) rods positioned at the same *c* level like rods 6 and 6′, (*b*) rods at levels 6 and 7 only, and (*c*) rods at all levels 2–7 as in the true structure. In addition, the interpretation of the fringes and their spacing has been validated by observing sections through the Patterson map of an ideal B-DNA helix as a function of the angle between helix axis and the Patterson section plane.

A separation of rods 6 and 7′ by 0.145 along the two-cell dotted line corresponds to a fractional displacement across a single cell of 2 × 0.145 = 0.290. Hence, the *X* coordinate of rod 6 is (1 − 0.290)/2 = 0.355. With a vertical separation of 0.185, rods 6 and 7 sit below and above level *c* = 5/6 by 0.093, or at 0.740 and 0.926, respectively.

2.2. *Helical rod rotation and translation*

With the determination of the packing of rods, two rotational and two translational parameters are established. The final two coordinates to be determined are the rotation of the stacked helical rod about its axis and translation along that axis. They are decidedly of secondary importance. If the B-DNA helix were a solid rod (approximated at low resolution), these two parameters would not exist at all. If the rod were a continuous B-DNA helix, without the deletion of phosphates every ten steps (approximated at intermediate resolution), then they would merge into a single variable: rotation by 36° would be identical to a translation of −3.34 Å. Conversely, rotation of 36° accompanied by translation of +3.34 Å would map an ideal, continuous helix onto itself one step farther along the helix. Clearly, it is inefficient to search for values for these two parameters early in the molecular-replacement analysis process. But when everything else has been determined correctly, it is time to turn to these remaining quantities.

To this end, a simultaneous rotation/translation search was carried out using the correlation coefficient between observed and calculated intensities (Fig. 3). For the calculation of I_c, an infinite continuous helix of alternating C and G bases was employed, constructed with ideal Arnott fiber-derived coordinates. Only data between 8.0 and 4.0 Å were included, in order that the difference between G·C and A·T base pairs might be ignored. A recurrent pattern of solutions related by rotations of 36° and translations of Δb = 0.10, or 3.34 Å, exhibited a correlation coefficient (peak height) of 0.654. This establishes the location of the two sugar–phosphate backbones along the endless rod, but does not yet indicate where one helix ends and the next begins.

The breaks between decamer helices were located by successive refinement of all ten possible screw solutions at higher resolution, 8.0 to 3.0 Å, using the correct base sequence. As seen from the progress of refinement in Fig. 4, one solution displayed a significantly lower *R* factor throughout refinement. This solution employed space group $P3_221$; the enantiomorph $P3_121$ was rejected because of strong forbidden intermolecular contacts. The *R* factor was 23.5% before addition of any ions or solvent molecules. Hence, the KKMe structure could be considered as having been solved, subject only to continued refinement and solvent searches. For further details, see Baikalov *et al.* (1993).

3. Tests of the method with other trigonal data sets

The same Patterson section strategy was applied to the trigonal CIMg helix, after a general rotation/translation search had failed. The best solution was found, which differed from the earlier solution in displacing the inter-

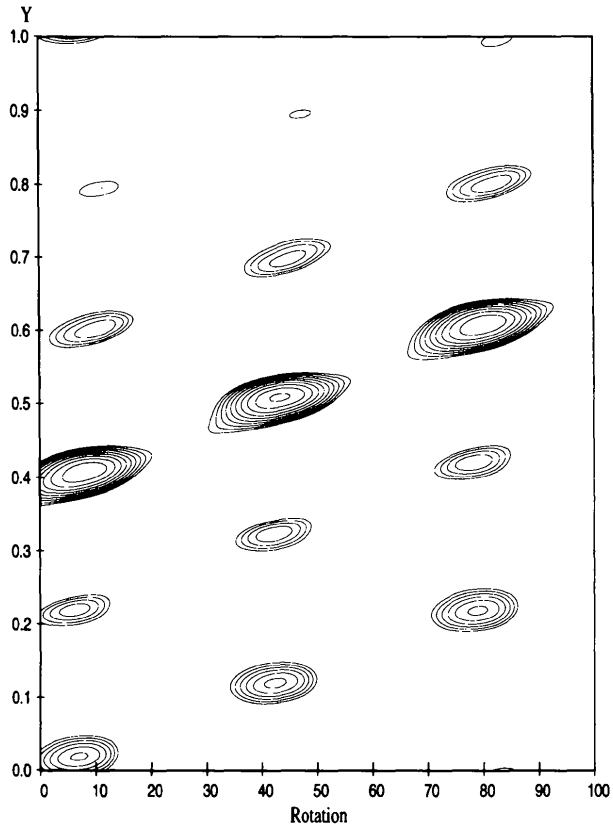

Fig. 3. Plot of correlation function between I_o and I_c for the KKMe helix as one varies the rotation about a helical rod (horizontal) and displacement along the rod (vertical), after rods have been properly positioned. Ten equal contour levels are drawn for correlation coefficient levels between 0.45 and 0.65. The highest peaks in the middle, at level 0.65, differ by rotations of 36° and translations of 3.34 Å (0.10 in c), and represent the same physical solution for an ideal continuous helix.

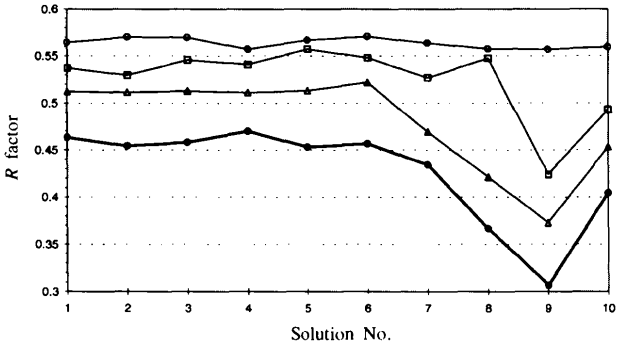

Fig. 4. Rigid-body refinement of ten choices for the break between stacked KKMe helices, at 3.0 Å resolution. From top to bottom, the four curves represent: circles = rigid helix positioned for the best R factor, with no other energy terms present; squares = ten rigid base pairs, with soft 'repel' potential for van der Waals interactions to avoid unfavorable intermolecular contacts at early stages of refinement; triangles = 20 rigid nucleotides, Lennard–Jones potential; filled circles = 58 rigid bases, sugars, and phosphates. Observe the clear and consistent preference for model 9, once the refinement of components of the ideal helix began.

helix junctions by one base pair. This solution in fact was isomorphous with that found for KKMe.

The approach outlined here was tested next by using it with a trigonal cell in which the helices are packed differently than with KKMe and CIMg. Kopka *et al.* (1994) have reacted the anti-tumor drug anthramycin with the decamer of sequence C-C-A-A-C-G-T-T-G-G, separated and purified two fractions with covalently bound drug, and crystallized both fractions. One fraction, called Ant-1, adopts either trigonal space group $P3_121$ or $P3_221$, with cell dimensions 33 × 33 × 108 Å, $\gamma = 120°$. Strong clusters of 3.4 Å reflections once again indicate stacked rods parallel to the a, b and $-(a + b)$ axes. But a Patterson map section through the origin perpendicular to the b axis, Fig. 5, is quite different from that of Fig. 2. Peaks between layers of parallel rods now occur at $w = 0.50$ instead of $w = 0.185$. Layers analogous to 6 and 7 in Fig. 1 lie a half-cell apart rather

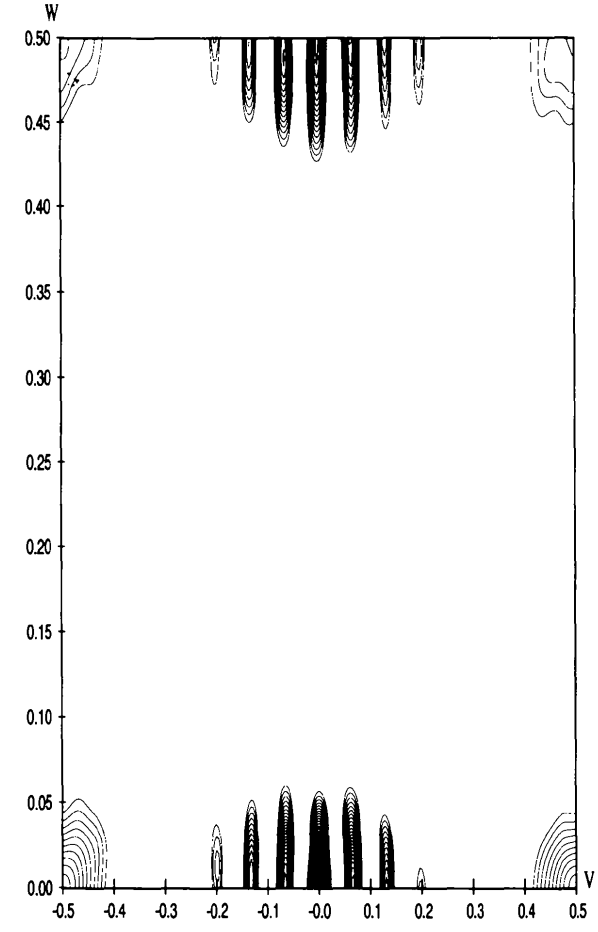

Fig. 5. Patterson map section through the origin, perpendicular to the b axis, for the Ant-1 covalent complex of anthramycin with C-C-A-A-C-G-T-T-G-G in space group $P3_121$. The Patterson peak between rods 6 and 7 of Fig. 1 occurs at $(v,w) = (0.0,0.5)$, indicating that layers of parallel rods are half a c cell axis apart rather than one-sixth as with KKMe. Resolution ranges from 8.0 to 3.0 Å; map is contoured at $\sigma/3$.

than being adjacent. Hence the orientation of rods in successive layers up the c axis is: a, $-(a + b)$, b, a, $-(a + b)$, b, whereas that for the KKMe helix was: a, a, $-(a + b)$, $-(a + b)$, b, b.

However, it is not immediately clear from Fig. 5 whether the single peak representing the pileup from aligned helices 6 and 7′ is to be found at $(v, w) = (0, \frac{1}{2})$ or at $(\frac{1}{2}, \frac{1}{2})$ that is, whether the single peak occurs at the upper left and right corners of the map, or whether it is obscured by the set of fringes in the center. If, in Fig. 1, helical rods 6 and 7 were to be shifted so that they are superimposed along the b axis, or at $a = 0$, then the peak representing vectors between same-orientation rods 6 and 7′ would occur at $(v, w) = (\frac{1}{2}, \frac{1}{2})$. Conversely, if helical rods 6 and 7′ were to be superimposed at $a = \frac{1}{2}$, then their vectors would create a peak at $(v, w) = (0, \frac{1}{2})$. Simulations of both arrangements with model helices yield Patterson sections that are nearly identical, and do not decide the issue. But the question can be settled by means of rotation/translation plots like that for KKMe in Fig. 3. Such a plot establishes the multiple rotation/translation solutions, differing by 36° and 3.34 Å, that define the same correct orientation of the double helix along the rods shown in Fig. 1. It positions the major and minor grooves of the helix correctly, without defining the precise location of the gap between one stacked decamer helix and the next.

Plots for both enantiomers of the space group are shown in Fig. 6, for the model in which the helices of layers 6 and 7 occur at $a = \frac{1}{2}$. The correlation function maximum is PC = 0.424 for space group $P3_121$ (Fig. 6*a*), and 0.381 for $P3_221$ (Fig. 6*b*). The model with helices at $a = 0$ can be ruled out since the correlation function in the equivalent plots never rises above PC = 0.07. Hence,

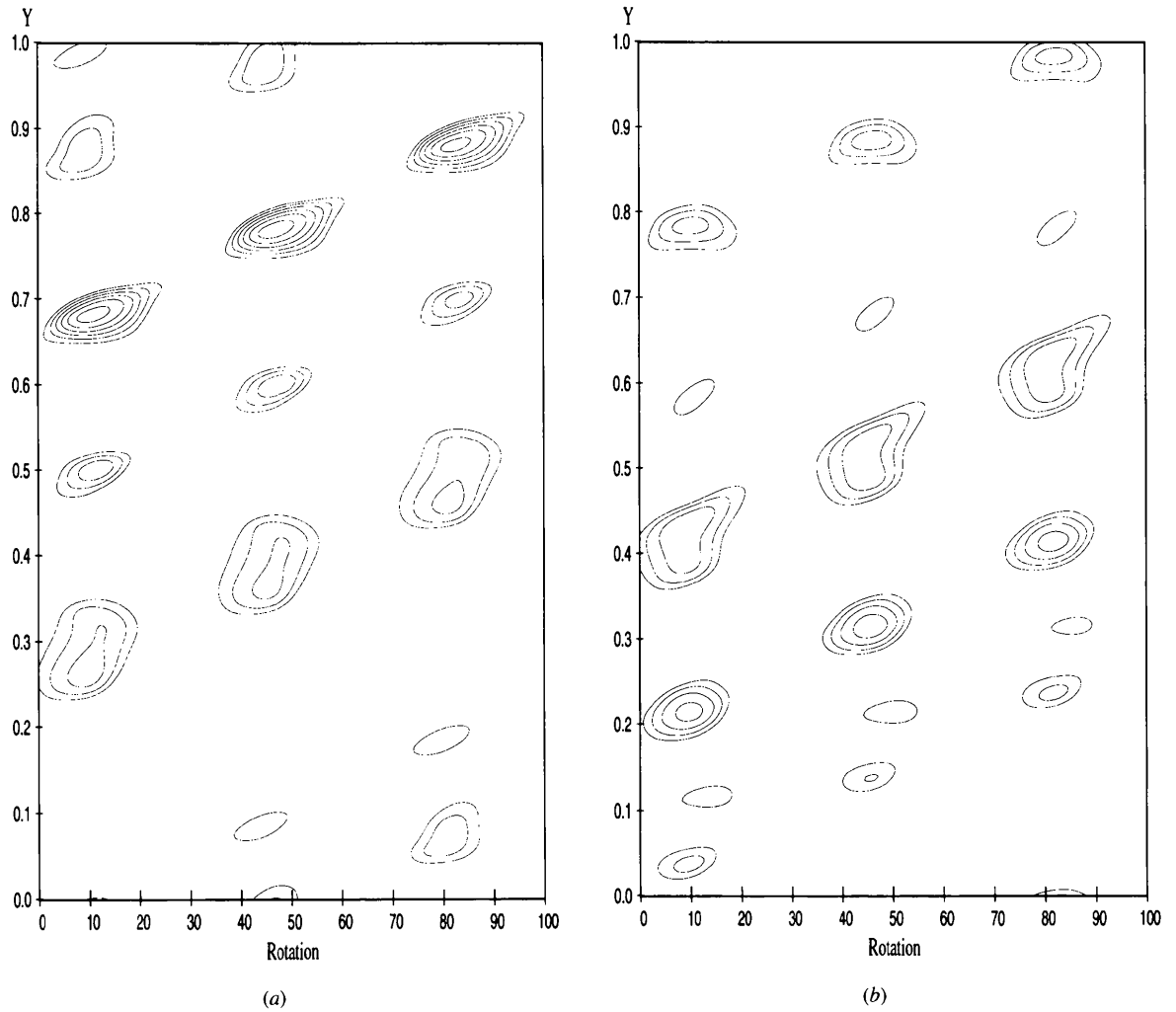

Fig. 6. Two-dimensional rotation/translation searches around and along the helix axis, as Fig. 3, but for the Ant-1 DNA–drug complex. (*a*) Space group $P3_121$, rods 6 and 7′ at $a = \frac{1}{2}$. Maximum value of correlation coefficient is 0.424. (*b*) Space group $P3_221$, rods 6 and 7′ at $a = \frac{1}{2}$. Maximum peak height 0.381. Ten equal contour levels are drawn for correlation coefficients between 0.3 and 0.5. Similar maps for both space group enantiomers with rods 6 and 7 at $a = 0$ gave no correlation coefficients above 0.07.

the obscured peak in Fig. 5 should be sought at $v = 0$, or beneath the fringes at the center of the top plot border.

Comparison of Fig. 6(*a*) and 6(*b*) with one another and with Fig. 3 suggests strongly that $P3_121$ is the correct space group. This is reinforced by the subsequent refinements to search for the break between helices. Models for this space group consistently refined to lower R values than those for the other enantiomer. However, the break between helices could not be established with certainty, for two instructive reasons. The main problem is the low resolution of the data set: 2.8 Å, compared with 2.0 Å for the KKMe helix. For such fine structure as the presence or absence of individual P atoms, the signal at this low resolution has sunk below the noise level.

Another perturbing factor, however, is the presence of a still-unlocated drug molecule bound to the double helix. One anthramycin molecule, $C_{16}H_{17}N_3O_4$, contributes 166 electrons to the scattering of X-rays, whereas the two missing phosphates together amount to only 60 electrons. It is futile to attempt to locate inter-helix junctions until the drug molecule first has been positioned. And it is at this point that the low resolution of the data set brings the analysis to a close (see below for another crystal form of anthramycin/DNA).

This example illustrates particularly well the importance of determining the six rotation/translation parameters in the correct order. Some of the six parameters are more robust than others *i.e.* more strongly determined and hence less sensitive to problems with data-set quality or resolution. The most efficient procedure, and that most likely to end in success, is to establish these strongly determined parameters before even beginning to search for values for the more sensitive ones.

4. Tests of the method with other space groups

The Patterson section approach to finding relative positions of helix axes also has been tried on structures in non-trigonal space groups. Decamer C-A-T-G-G-C-C-A-T-G, called CAT, crystallizes in orthorhombic space group $P2_12_12_1$ with cell dimensions 36.6 × 42.5 × 34.7 Å (Goodsell *et al.*, 1993). Significant deviation of these cell parameters from the well known orthorhombic decamer standards (Grzeskowiak *et al.*, 1991; Quintana *et al.*, 1992; Yuan *et al.*, 1992) implied a new packing arrangement, providing an opportunity to test the method in a different space group.

Strong reflections at 3.4 Å resolution along the crystallographic axes c revealed the direction of the helix axes. Patterson map section through the origin in the *ab* plane (Fig. 7) contains two unique peaks in addition to the origin peak. The cross-peak between parallel rods of helices at (X, Y, Z) and $(\frac{1}{2} - X, -Y, \frac{1}{2} + Z)$ is at (0.4, 0.5), while the compound peaks at (0.0, 0.5) correspond to inverted pairs of helices at (X, Y, Z) and $(-X, \frac{1}{2} + Y, \frac{1}{2} - Z)$. Coordinates of the cross-vector $(2X - \frac{1}{2} = 0.4, 2Y = 0.5)$ give the position of a helical rod in the XY plane: $X = 0.45$, $Y = 0.25$. The orthorhombic space group has the added simplification that all stacked helical rods run parallel, and hence the fringes that obscure details in trigonal cases are absent.

A simultaneous rotation/translation search, shown in Fig. 8, exhibits familiar repeats separated by 36°, but for each rotation solution there are four equivalent peaks. These multiple peaks result from a combination of the different choice of origin and the molecular dyad relating two strands of the idealized B-DNA.

A rigid-body refinement of ten screw solutions starting at (0.45, 0.50, 0.14), 18° gives the best solution with R factor of 31% for data between 8 and 3 Å resolution, and is essentially the same as the solution described in Goodsell *et al.* (1993) after a frameshift mistake was corrected. Thus, the method works well with the orthorhombic space group, and helps to avoid erroneous solutions.

5. Positioning helices along twofold axes

It is not uncommon in crystals of DNA oligomers for a crystallographic twofold axis to pass through the center of the helix, perpendicular to the helix axis. Under these conditions the asymmetric unit is only a single helix strand, or alternatively half the base pairs of the helix.

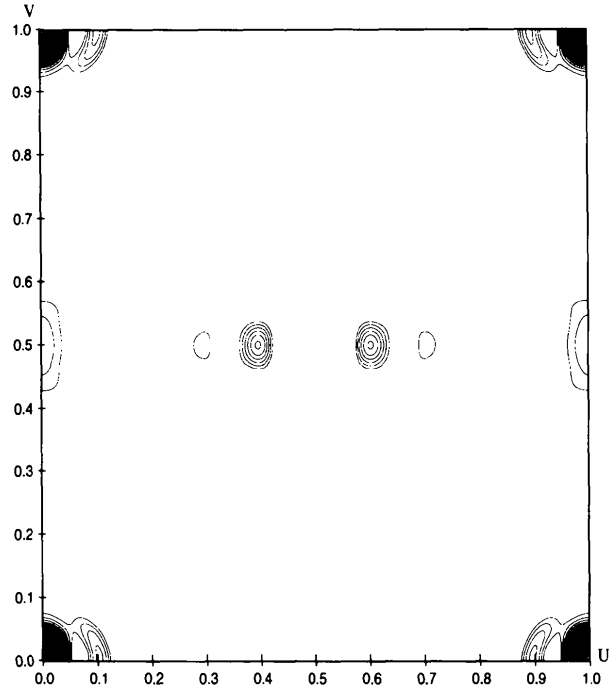

Fig. 7. Patterson map section through the origin, perpendicular to the c axis, for C-A-T-G-G-C-C-A-T-G (the CAT decamer). The horizontal and vertical axes are parallel to a and b, respectively. Resolution ranges from 8.0 to 3.0 Å; map is contoured at $\sigma/3$.

This is the situation, for example, with tetragonal A-DNA octamers and monoclinic B-DNA decamers (Dickerson, 1992). Under such conditions two translational variables are established by the location of the twofold axis in the cell, and the only remaining translation is the positioning of the helix along that axis. And the three rotational degrees of freedom are reduced to one unknown rotation and a two-valued choice: a propeller-like rotation of the helix about the twofold axis, and the choice of whether the major groove or the minor groove at the middle of the helix faces a given direction along the axis. Hence, only a two-dimensional search is called for: translation along the twofold and rotation about it, with two choices of groove orientation to be examined.

These ideas were tested with the other crystalline DNA–anthramycin complex, Ant-2. This form adopts the familiar trigonal space group $P3_221$, but with a completely different crystal packing (Kopka *et al.*, 1994). Double helices no longer are stacked in endless rods parallel to the *ab* plane. Instead, discrete helices spiral up the threefold screw axis along *c*, making a constant angle of roughly 20° with that axis as evidenced by clusters of strong 3.4 Å reflections in the diffraction pattern. The asymmetric unit is only a single strand or half of a decamer double helix. In this space group, twofold axes lie parallel to real-space axes *a*, *b* and $-(a + b)$. Hence, only two variables remain undetermined: the distance of a helix from the origin along the *a* axis, and the rotation of the helix axis away from the perpendicular or *c* direction.

Fig. 9 shows a two-dimensional rotation/translation search for Ant-2 using the correlation coefficient between I_o and I_c, the latter calculated from an ideal Arnott B-DNA starting model skewered on the crystallographic twofold axis parallel to *X*. The single large peak of height PC = 0.36 yields an unambiguous answer: rotation of helix axis 24° from the vertical (in agreement with the strong 3.4 Å reflections) and *X* coordinate of 0.68. Reversing the helix to face the major groove in the opposite direction at the middle of the helix, or choosing the opposite space-group enantiomer, $P3_121$, yielded maximum correlation peak heights in the maps of less than 0.20.

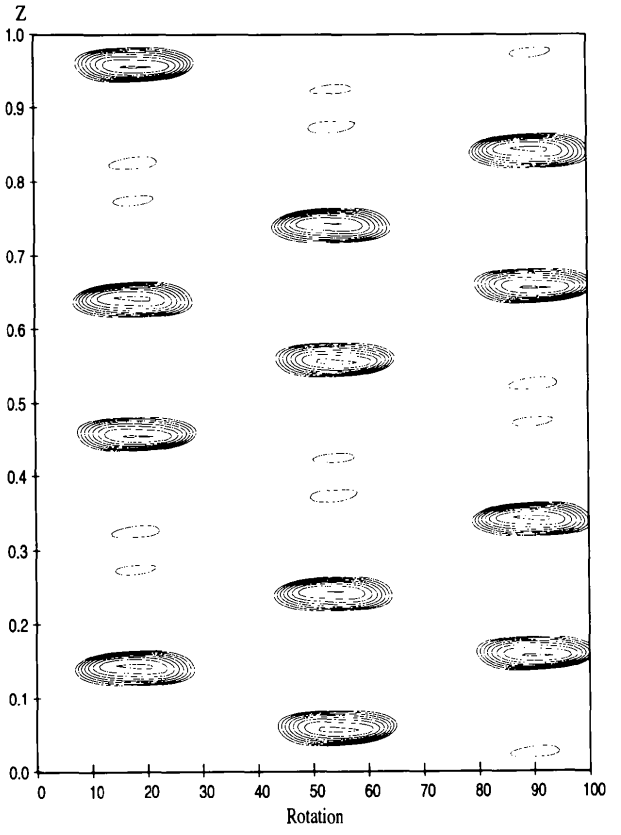

Fig. 8. Two-dimensional rotation/translation search around and along the helix axis for the CAT decamer. All peaks are approximately the same height, PC = 0.460. Ten equal contour levels are drawn for correlation coefficients between 0.3 and 0.5.

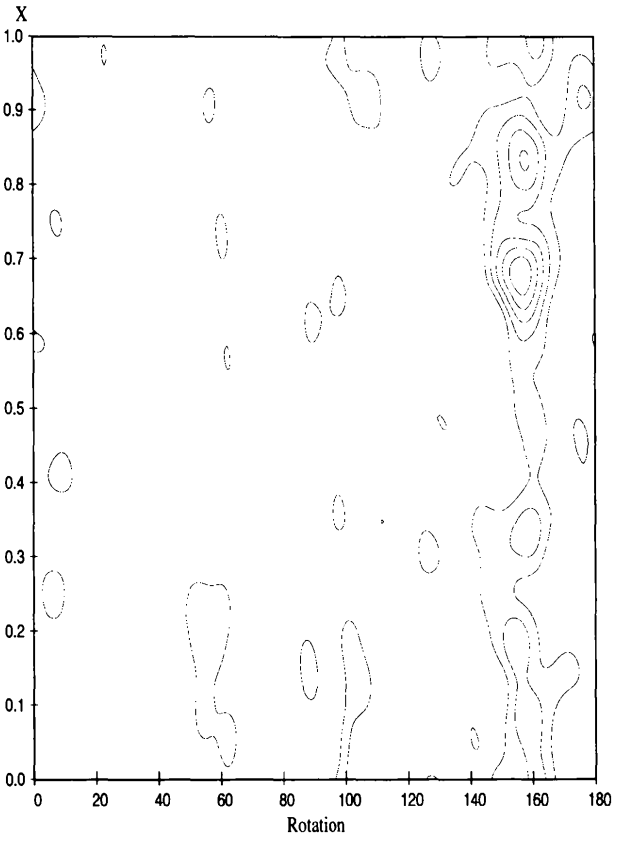

Fig. 9. Direct search of rotation of the helix about a diad through its center (horizontal) and displacement of the helix along the twofold axis (vertical) for the anthramycin/DNA complex Ant-2 in space group $P3_221$. Resolution ranges from 8.0 to 4.0 Å; map is contoured at 1σ. A single 6σ peak of height 0.36 corresponds to the solution (X = 0.68, κ = 24°).

6. A general solution of the problem of helix orientation

Knowledge of helix axis direction or of molecular diad position greatly simplifies a structure solution. However, helical symmetry alone can provide useful information. A self-rotation function $f(\psi,\varphi,\kappa)$ (Rossmann & Blow, 1962), where κ represents rotation about the (ψ,φ) axis, has a maximum whenever the (ψ,φ) axis is parallel to the true helix axis and κ is a multiple of the helical repeat. For a helical molecule with a 36° twist, the direction of the helix axis (ψ,φ) should be discernible *via* strong peaks on (ψ,φ) sections at κ = 36, 72, 108 or 144°. Fig. 10(*a*) shows the 36° section for the Ant-2 data. The correct peak is present in each section, if one knows where to look. But experimental noise and deviations of the helix from ideality can combine to contribute multiple peaks that make each individual section difficult to interpret by itself. Fig. 10(*b*) shows a *minimum function* prepared from the four individual plots, using at every point the lowest value from all four maps. This simple procedure accentuates common features present in all maps, and carries along the true answer, which appears as the highest peak in the minimum function map, at ψ = 66, φ = 90°. These angles correspond to a helix inclination of 24° relative to the z axis, which is in perfect agreement with the result of the direct search for Ant-2.

Such a general search procedure depends upon the presence of a regular helix in the crystal, but neither on a particular value for helical twist, nor on base pairs being perpendicular to the helix axis. Hence, in principle it can be used with any class of DNA helix; indeed, it can be used to establish the helical twist itself and hence the class of helix. Amplification of the common peaks in the consecutive sections of the self-rotation function should appear only when an increment angle κ has a value close to that of the actual helical twist. Therefore, a series of minimum functions calculated at different κ values in the range of 30–36° should be sufficient to differentiate an A from a B helix.

7. Conclusions

The most important goal of the strategy advocated here has been to separate the search variables and focus on the most significant variables before turning to less important ones. Zooming into the structure not only shortens the process of solving the structure, it provides the most efficient way to obtain an *unambiguous* solution. In the case of DNA helices, the very first step is to find the direction of the helix axis in the unit cell, and the second step is to position the helices correctly. If helices are stacked into quasi-continuous rods, or if a crystallographic twofold axis relates the two strands of each helix, then these are powerful restraints that can be used to further the molecular replacement. It helps to consider the DNA helices first as endless helical rods at low resolution, while the location of the missing phosphates that separate one helix from another in the stack is very much a third-order consideration. If no other restraints are available, then the helical symmetry alone can reveal the direction of the helix axis, significantly reducing the number of unknowns in further

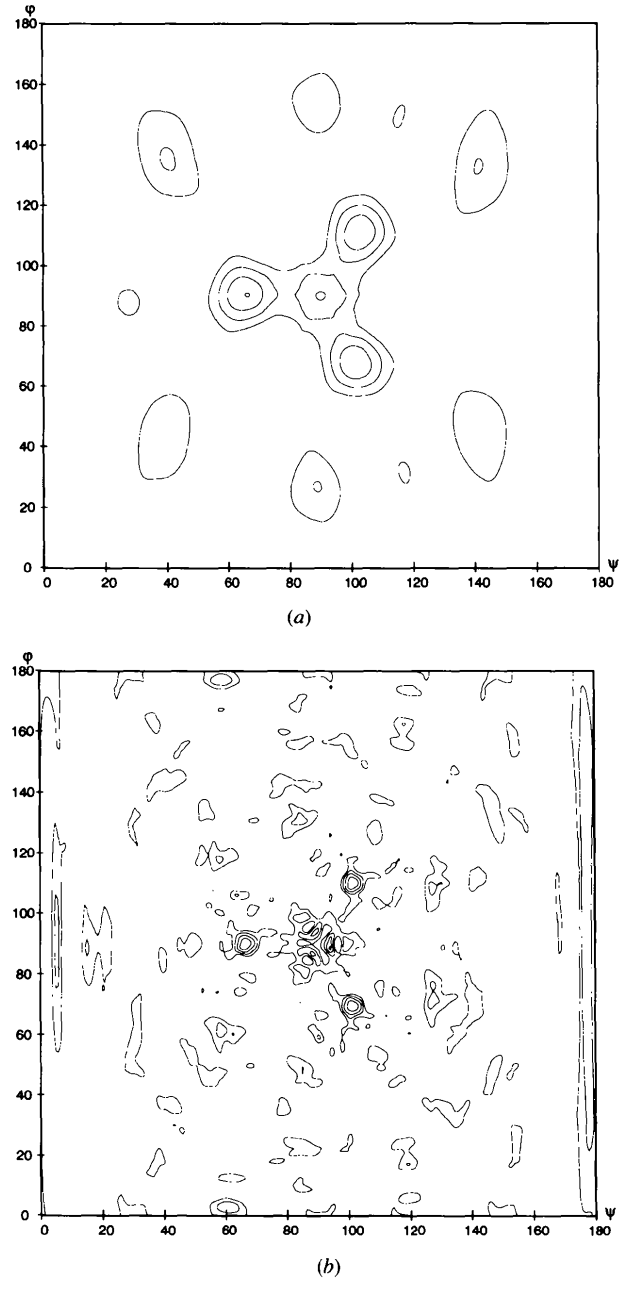

Fig. 10. (*a*) Section of κ = 36° through the self-rotation function $f(\psi,\varphi,\kappa)$ for Ant-2. Horizontal axis, ψ, is the inclination *versus* y axis, vertical axis, φ, is the azimuthal angle. Resolution ranges from 8.0 to 4.0 Å; map is contoured at 1σ. (*b*) The minimum function constructed by taking at each point the lowest value on the four individual sections of κ = 36, 72, 108 and 144°.

search. Although the approach outlined here is in no way an ultimate one, it proved to be effective with various space groups and provides a solid framework for investigating new ones.

We thank Mary Kopka, Andrei Lipanov and David Goodsell for the use of their data and for inspiring the development of the method. This work was conducted with the support of NIH program project GM-31299, NSF grant MCB-8916261, and American Cancer Society grant DHP-22.

References

Baikalov, I., Grzeskowiak, K., Yanagi, K., Quintana, J. & Dickerson, R. E. (1993). *J. Mol. Biol.* **231**, 768–784.

Crowther, R. A. & Blow, D. M. (1967). *Acta Cryst.* **23**, 544–548.

Dickerson, R. E. (1992). *Methods Enzymol.* **211**, 67–111.

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 2179–2183.

Epp, O., Steigemann, W., Formanek, H. & Huber, R. (1971). *Eur. J. Biochem.* **20**, 432–437.

Fujinaga, M. & Read, R. J. (1987). *J. Appl. Cryst.* **20**, 517–521.

Goodsell, D. S., Kopka, M. L., Cascio, D. & Dickerson, R. E. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 2930–2934.

Grzeskowiak, K., Yanagi, K., Privé, G. G. & Dickerson, R. E. (1991). *J. Biol. Chem.* **266**, 8861–8883.

Heinemann, U. & Alings, C. (1991). *EMBO J.* **10**, 35–43.

Heinemann, U., Alings, C. & Bansal, M. (1992). *EMBO J.* **11**, 1931–1939.

Huber, R. (1965). *Acta Cryst.* **19**, 353–356.

Kopka, M. L., Goodsell, D. S., Baikalov, I., Grzeskowiak, K., Cascio, D. & Dickerson, R. E. (1994). *Biochemistry*, **33**, 13593–13610.

Lipanov, A., Kopka, M. L., Kaczor-Grzeskowiak, M., Quintana, J. & Dickerson, R. E. (1993). *Biochemistry*, **32**, 1373–1389.

Privé, G. G., Yanagi, K. & Dickerson, R. E. (1991). *J. Mol. Biol.* **217**, 177–199.

Quintana, J. R., Grzeskowiak, K., Yanagi, K. & Dickerson, R. E. (1992). *J. Mol. Biol.* **225**, 379–395.

Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.

Yuan, H., Quintana, J. & Dickerson, R. E. (1992). *Biochemistry*, **31**, 8009–8021.