IMMUNOMETHODS 3, 180–190 (1993)

Structure Determination of Antibodies and Antibody–Antigen Complexes by Molecular Replacement

Axel T. Brünger

Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511

Three-dimensional structures of antibody and antibody–antigen complexes that have been solved to date exhibit highly conserved folding motifs, and the main structural differences between antibodies are restricted to the interdomain angles and the antigen-binding site. This knowledge has been successfully used to solve a large number of crystal structures of antibody fragments by molecular replacement. This paper reviews approaches that have been used in solving these crystal structures. © 1993 Academic Press, Inc.

In macromolecular crystallography, the determination of initial phases is the major obstacle in the determination of the structure of the crystallized molecule, because the observable diffraction information from a single crystal comprises only the amplitudes but not the phases of the reflections. This is commonly referred to as the phase problem. Although the phase problem has been solved in the case of small molecules (up to a few hundred atoms in the unit cell) through the so-called direct methods (1), the application of these methods to macromolecules has not yet succeeded, and one must resort to time-consuming and sometimes unsuccessful experimental methods.

Progress in obtaining approximate three-dimensional models of macromolecules from information other than crystallography suggests a knowledge-based approach to crystallographic phasing. The data base of known protein sequences is growing rapidly. The Human Genome Project (2) will likely increase the growth rate of this data base. Techniques for aligning sequences such as consensus templates (3) have been developed in order to recognize very distantly related proteins or protein domains and to carry out model building on the basis of the several hundred known protein structures (4). Many proteins are composed of domains that are similar in structure to a finite number of folding motifs (5). This suggests an approach to crystallographic phasing that consists of solving the problem on a case by case basis for each particular folding motif.

The method of choice for such a knowledge-based approach to crystallographic phasing is molecular replacement (6, 7), which is also referred to as Patterson search. It is a two-step procedure, first rotating ("rotation function") and then translating ("translation function") the known structure of a homologous search model in the unit cell of the target crystal in order to obtain a maximum correlation between observed and calculated diffraction data. Unfortunately, molecular replacement often fails if the search model is too inaccurate (8), that is, if the differences in atomic positions between the search model and the crystal structure are more than about 1 Å. The underlying cause may be found in the observation (9) that when the degree of sequence identity between two structures is less than 50%, the overall fold may be maintained but the relative orientations and positions of the secondary structural elements or domains can change substantially. Despite three decades of experience with molecular replacement there is only a partial understanding

1058-6687/93 $5.00

of the reasons for success or failure of the method. Antibodies represent an exception here; in recent years considerable experience has been gained in solving antibody structures by molecular replacement.

This review focuses on the structure determination by molecular replacement of Fab fragments of IgG antibodies and their complexes with antigens. Fab fragments consist of two polypeptide chains (heavy and light chain), each composed of two domains connected by flexible polypeptide segments (Fig. 1). The domains are composed of the immunoglobulin fold, a two-layer β-sheet sandwich (10–13). Apart from the antigen CDR (complementarity-determining region) loops, the individual domains of different Fabs are structurally very similar (14). The most important differences between Fabs are described by the variation in the interdomain angles and positions. The elbow angle, relating the directions of the two pseudo-twofold axes of the constant and variable domains (cf. Fig. 1), can range from about 130 to 180°. In addition, the variation of the heavy and light chains in the variable and constant domains can be up to 12° around and 3 Å along the pseudo-twofold axes (12). The differences in interdomain angles appear to be a consequence of the intrinsic flexibility of the linker regions and are probably not functionally related. In fact, the same Fab molecules can have different elbow angles in different crystal forms (13).

In view of the large variations between domains, molecular replacement has been extended by introducing additional parameters, such as the interdomain angles and positions, and by refining these parameters between the rotation and the translation stage of molecular replacement (15, 16). This approach, which is also referred to as generalized molecular replacement, has improved the discrimination between correct and incorrect orientations of the whole Fab and it has also increased the success rate of the translation function. In this paper we review the principles of molecular replacement and discuss its application to the solution of Fab antibody fragments and their complexes with antigens.

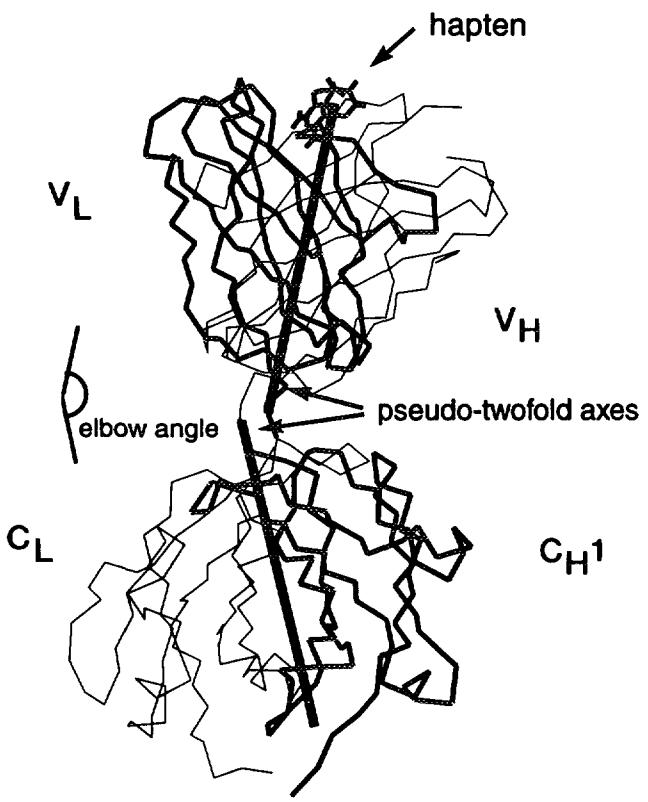

FIG. 1. Cα backbone plot showing the four domains of the AN02 structure (34, 35) (thin and medium lines), the hapten molecule (medium lines), and the pseudo-twofold axes (thick lines). The elbow angle is defined as the angle between the two directions given by the pseudo-twofold axes. The constant and variable domains of the light chain are denoted C_L and V_L, respectively. The constant and variable domains of the heavy chain are denoted C_H1 and V_H, respectively.

DESCRIPTION OF METHODS

In general, there are six parameters, three translations and three angles, required to place a search model in the unit cell of a crystal with one molecule per asymmetric unit. Additional parameters are required when more than one molecule is present in the asymmetric unit. Low-dimensional space groups can reduce the dimensionality of the problem because some of the translations may be arbitrary. For example, in space group $P1$, the translation of the search model is completely arbitrary.

While a full six-dimensional search is no longer beyond the scope of available computing resources, a much less computing-intensive two-step procedure often produces the same results as the full search (6, 7). First, the search model is rotated without considering its crystallographic symmetry mates and the best agreement with the observed diffraction data is obtained (rows 1 and 2 in Fig. 2). Second, the correctly oriented search model and all its symmetry mates are then translated in order to find the best agreement with the diffraction data (row 4 in Fig. 2). This procedure is justified by the fact that the agreement between the observed and the computed diffraction data is dependent on the orientation, but not on the position,

of the search model when the symmetry mates are excluded. The drawback to this two-step procedure consists of a low correlation between a single copy of the search model and the diffraction data especially in high-symmetry space groups. This problem becomes even more pronounced in the case of poor or incomplete search models. This may produce an erroneous solution of the rotation function, which in turn will cause a failure of the translation function.

Rotation Function

Rotation functions have been formulated in real space (orthogonal Å coordinate space) as well as in reciprocal space. While the two formulations are equivalent in principle, there are differences in the numerical approximations employed and in the way certain weighting and cutoff schemes can be applied. All rotation functions involve the computation of the Patterson function or its reciprocal space analog for the search model and the observed diffraction data. A Patterson function is a three-dimensional scalar field that is obtained by Fourier transformation of the diffraction intensities. Patterson functions can be interpreted as the collection of all interatomic vectors in the unit cell of the crystal.

For all rotation functions except those of the "direct" type described below (Eq. [3]), the search model is placed in an artificial orthogonal unit cell. The dimensions of this unit cell are chosen to be greater than the extent of the molecule plus the maximum length of the Patterson vector used in the rotation function. Therefore, intermolecular Patterson vectors produced by the artificial unit cell are removed from the rotation function.

In the real space formulation, the Patterson function of the crystal P_x and that of the search model P_m are computed. P_x and P_m are then rotated with respect to each other and a product correlation is computed,

$$\mathrm{Rot}(\Omega) = \int_U P_x(r)P_m(\Omega r)dV, \qquad [1]$$

where the 3×3 rotation matrix Ω is described by three angles (e.g., Eulerian angles), r is the integration variable, and U is the volume of integration, usually spherical, centered at the origin. To reduce computing requirements, only the strongest peaks of the model Patterson function P_m are used in the integration and the value of the crystal Patterson function P_x at these peaks is computed by interpolation (17, 18). The integration is typically carried out over a shell that eliminates the trivial origin peak and all Patterson vectors beyond a certain radius.

By Fourier transformation of Eq. [1] and application of Parseval's theorem, one obtains a reciprocal space formulation of the rotation function,

$$\mathrm{Rot}'(\Omega) = \sum_{\mathbf{h}} F_x^2(\mathbf{h})F_m^2({}^t\Omega\mathbf{h}), \qquad [2]$$

where $\mathbf{h} = hkl$ are the Miller indices of the crystal, ${}^t\Omega$ denotes the transpose of the rotation matrix Ω, and F_x and F_m are the structure factors of the crystal and the search model, respectively. F_x^2 is obtained from the observed diffraction intensities, whereas F_m is obtained by the structure factor expression that describes the particular search model. Crowther (19) developed an expansion of $\mathrm{Rot}'(\Omega)$ in terms of spherical harmonics functions which allows computation of the rotation function by Fourier transformation. This procedure greatly speeds up the rotation function but care must

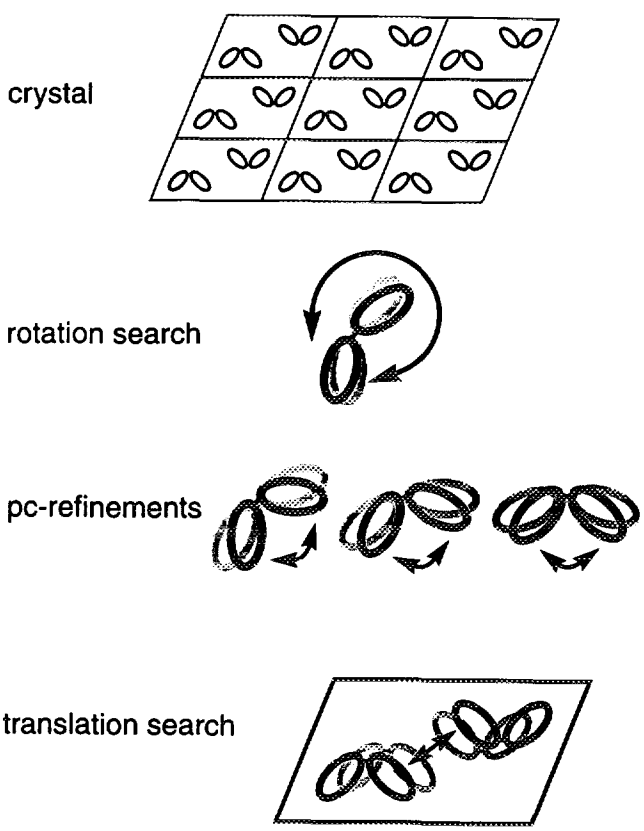

FIG. 2. Illustration of the generalized molecular replacement strategy: rotation function, PC refinement, and translation function.

be taken to include a sufficient number of spherical harmonics terms to ensure proper convergence.

High-symmetry space groups or poor models often produce fairly weak signals for the rotation function. In this case, the noise produced by the numerical approximations (interpolation or spherical harmonic expansion) used to compute Eqs. [1] and [2] can make it difficult to detect the correct solution. These difficulties can be reduced when computing the Patterson function for each sampled orientation of the atomic model. In X-PLOR [21] this approach is used for the computation of a correlation coefficient (referred to as PC, Patterson correlation),

$$\mathrm{PC}(\Omega) = \frac{\langle |E_x|^2 |E_m(\Omega)|^2 \rangle - \langle |E_x|^2 \rangle \langle |E_m(\Omega)|^2 \rangle}{\sqrt{\langle |E_x|^4 - \langle |E_x|^2 \rangle^2 \rangle \langle |E_m(\Omega)|^4 - \langle |E_m(\Omega)|^2 \rangle^2 \rangle}}, \quad [3]$$

where E_x denotes the normalized observed structure factors, and $E_m(\Omega)$ denotes the normalized structure factors of the search model placed in the unit cell of the crystal in an arbitrary position and oriented according to Ω *without considering the crystallographic symmetry mates.* The angle brackets denote an averaging over the set of observed reflections expanded to $P1$. In a different context, Hauptman showed that PC is a measure of the phase accuracy of an atomic model (20). At present, the main use of Eq. [3] is for the so-called PC refinement described below, where only a subset of orientations is evaluated. However, the direct rotation function could become the method of choice on massively parallel computers as it is highly parallelizable.

The goal of the rotation function is to find the global maximum of Eq. [1], [2], or [3]. This is usually accomplished through a grid search of the angular parameters that describe the orientation of the search model. Parameters that influence the rotation function are the completeness and accuracy of the model, the weighting functions and cutoffs used for the rotation function, and the resolution range and selection of diffraction data.

Generalized Molecular Replacement: Patterson Correlation or Intensity Refinement

Recently, we have generalized molecular replacement by introducing an intervening step between the rotation function and the translation function whose purpose it is (i) to improve the discrimination between correct and incorrect orientations for the rotation function and (ii) to improve the accuracy of the search model prior to the translation function (15). Independently, a similar approach was developed by Yeates and Rini (16). Our approach considers a set of likely orientations found by the rotation function, whereas Yeates and Rini's approach considers only the highest rotation function peak.

Since the details of the procedure are published elsewhere (15), we summarize only the main points here (Fig. 2). First, a rotation function is evaluated. A large number of orientations corresponding to the highest peaks of the rotation function are selected. Here one makes the ad hoc assumption that the correct orientation is among this selected subset. This assumption appears to be reasonable in many cases. One then introduces a small number of parameters p that describe the most dominant differences that are expected to occur between the crystal structure and the search model. So far, the method has been most successful in choosing as parameters p the orientations and positions of large rigid groups of atoms (secondary structural elements or protein domains). For n rigid bodies this requires $6n - 3$ parameters, where three parameters specifying the arbitrary overall position have been subtracted. Next is the most important step, which consists of refinements of p against $-$PC (Eq. [3]) in order to maximize PC. Refinement is carried out against $-$PC since minimization algorithms locate minima as opposed to maxima; a minimum of $-$PC corresponds to a maximum of PC. The major difference from crystallographic R factor refinement (22) is that PC refinement is carried out in a triclinic $P1$ cell, that is, without crystallographic symmetry. In fact, the inclusion of crystallographic symmetry operators appears to be redundant (see discussion below).

It should be noted that a large number of the PC refinements are actually carried out for incorrect orientations of the search model; only the PC refinements starting close to the correct orientation are expected to yield a relatively large correlation coefficient after PC refinement. Furthermore, one expects the search model to become more accurate in the latter cases. A necessary but insufficient condition for the correct solution of the crystal structure is that PC assumes a maximum. Thus, one selects another subset of orientations that have produced the largest correlation coefficients after PC refinement. The final step consists of translation functions using the PC-refined search models for this subset of orientations. The goal of the generalized molecular replacement procedure is to reduce the number of possible orientations to be checked

by subsequent translation functions and to improve the accuracy of the search model *prior to* the translation function.

Instead of PC, Yeates and Rini (16) use a residual error function,

$$Y = \sum_{\mathbf{h}} [F_x^2(\mathbf{h}) - s \sum_j F_m^2(S_j\mathbf{h})], \qquad [4]$$

where S_j comprises the symmetry operators of the crystallographic space group, s is a scale function, and the summation is carried out over a set of unique, observed reflections. As is the case for PC refinement, the Y function is refined against the parameters p of the search model. In contrast to PC in Eq. [3], the Y function includes all symmetry mates of the search model. Note that the *intensities* of the symmetry mates (F_m^2) are added; the summation of the complex structure factors F_m is not possible because the position of the search model is unknown. A weight function that deemphasizes the unknown intermolecular vectors in Patterson space can be applied (16). The effect of this weight function is probably similar to that of using normalized structure factors for PC (Eq. [3]).

Translation Function

Once the orientation of the search model is known, the position of the molecule must be found. There are a large number of approaches to this problem. Translation functions can use the Patterson function (23, 24), the correlation coefficient between the observed and the calculated structure factor amplitudes (25), or phased structure amplitudes, as in the phased translation function (26–30). The reader is also referred to the review article by Beurskens and co-workers (31).

Translation functions that make use of the Patterson function are derived from the analog to Eq. [2] with crystallographic symmetry,

$$\text{Trans}'(\delta) = \sum_{\mathbf{h}} F_x^2(\mathbf{h})[\sum_j F_m(S_j\mathbf{h})\exp(2\pi i S_j\mathbf{h}\cdot\delta)]^2, \qquad [5]$$

where the first sum is carried out over a unique set of the observed reflections and the second sum over all symmetry operators S_j of the space group. The simple functional form of Eq. [5] makes it possible to use Fourier transformation for the efficient evaluation of the whole function. Various modifications and simplifications are usually employed to increase sensitivity and performance. For example, the terms on the right hand side of Eq. [5] can be reorganized into intermolecular and intramolecular contributions. As the translation function is independent of the intramolecular contributions of the search model, the corresponding terms are usually omitted. Furthermore, contributions of self-vectors in the observed Patterson function can be removed from the translation function by replacing F_x^2 in Eq. [5] by

$$F_x^2 - k \sum_j F_m^2(S_j\mathbf{h}). \qquad [6]$$

Information about molecular packing has also been incorporated in the translation function (24).

The translation function based on the correlation coefficient (25) makes use of the target

$$\text{corr} = \frac{\langle |E_x|^2 |E_{calc}|^2 \rangle - \langle |E_x|^2 \rangle \langle |E_{calc}|^2 \rangle}{\sqrt{\langle |E_x|^4 \rangle - \langle |E_x|^2 \rangle^2 \rangle \langle |E_{calc}|^4 \rangle - \langle |E_{calc}|^2 \rangle^2 \rangle}}. \qquad [7]$$

This is the same definition as that for PC (Eq. [3]), except that the averaging operations are performed over an asymmetric unit of diffraction data and that E_m is now replaced by E_{calc}, which denotes the normalized structure factors of the search model and all its symmetry mates. No modifications of the normalized structure factors are required. The correlation coefficient includes information about molecular packing (25). In general, this type of translation function appears to be more sensitive than the translation functions based on Eq. [5]. However, this comes at a substantial computational expense because the correlation coefficient does not allow for a simple Fourier expansion of the whole translation function. As is the case for the direct rotation function (Eq. [3]), this type of translation function will probably become the method of choice on massively parallel computers.

The phased translation function has the potential to combine high sensitivity with computational efficiency (29, 30). This translation function follows the practice sometimes used in small-molecule crystallography, when a correctly oriented but misplaced partial structure can be found (32). Briefly, the space-group symmetry of the crystal is artificially reduced to $P1$, and the electron density map is calculated with phases derived from a model that is correctly oriented but arbitrarily positioned within the unit cell. Recognition of the symmetry-related molecules, however distorted, by a translational search allows for the positioning of the symmetry elements.

The goal of the translation function is to find the global maximum of Eq. [5] or [7]. This is usually ac-

complished through a grid search of the translational parameters that describe the position of the search model. Parameters that influence the translation function are the completeness and accuracy of the search model and the resolution range and selection of diffraction data.

RESULTS AND DISCUSSION

Rotation Functions

Cygler and Anderson (33) performed extensive studies of the performance of various rotation functions using intact Fab models as well as single domains. Parameters that were studied included the resolution range of the data, the outer integration radius of the Patterson function, the number of strongest reflections, and the included model atoms or domains. When the search was conducted with a complete Fab model, success depended critically on the resolution range employed and on whether side-chain atoms were included or excluded. However, there were no clear guidelines for complete Fab molecules. When the search was conducted with either the constant or the variable domain of the Fab, data in the resolution range 10–4 Å proved to be the most important. Models consisting of only Cα atoms produced the correct solution, albeit at a substantially lower signal-to-noise ratio. Interestingly, it is better to include the CDR loops despite the dissimilarity of their conformations in the search model and in the crystal. Generally, the positions and peak heights of the rotation function were much more dependent on the various parameters (resolution range, integration radii, completeness of the search model) in the case of a poor or incomplete model than in the case of a good model. A large outer integration radius (larger than 24 Å) was necessary. It was also shown that including more reflections in the rotation function might be beneficial for a poorer model. It should be pointed out that these tests were carried out for a low-symmetry crystal structure of the HED10 monoclonal antibody Fab fragment (space group $P2_1$).

Attempts to solve the structure of an anti-dinitrophenyl spin-label monoclonal antibody Fab fragment AN02 (34, 35) and the structure of a Fab Digoxin complex (36, 37) failed using conventional rotation functions (using MERLOT (38) and program PROTEIN (18)) of either the intact Fab or its individual constant or variable domains. The former Fab crystallized in space group $P6_522$, while the latter Fab crystallized in space group $P2_1$ with two molecules in the asymmetric unit.

Patterson Correlation or Intensity Refinement

We subsequently succeeded in obtaining phases for the AN02 and the Fab/digoxin structures by using generalized molecular replacement (37, 35). The correct rotation function solutions were clearly identified and the significance of the translation function was greatly enhanced through PC refinement. Other structures that were solved by this strategy include a complex of an anti-angiotension II Fab with bound peptide (39, 40), which crystallized in space group $P4_1$, and the V_L-related fragment of the human κI Bence Jones protein Wat, which crystallized in space group $P6_4$ (41; unpublished results). In the case of the Fabs, PC refinements of 21 parameters specifying the orientations and relative positions of the constant and variable domains of the heavy and light chains were carried out; in the case of the Wat V_L dimer, PC refinement of the 9 parameters specifying the orientations and relative position of the two monomers were carried out.

Brady and co-workers (42) solved a chimeric Fab fragment of a tumor cell binding antibody after employing PC refinement of a large number of the most likely orientations of the search model. Rini and coworkers (43) solved the crystal structure of an HIV-1 neutralizing antibody (50.1) in complex with its V3 loop peptide antigen by generalized molecular replacement using intensity refinement (16). In retrospect, a feature common to all these cases is that the inaccuracy of the model in combination with the presence of noncrystallographic symmetry or the high crystal symmetry of the space group made it difficult to identify the correct orientation of the molecule(s) or single domains by conventional rotation functions.

If successful, PC refinement can discriminate between correct and incorrect orientations of the search model. Indeed, the two highest peaks of the rotation function for the 26-10 Fab were incorrect (37) despite the fact that they were related to each other by the noncrystallographic twofold symmetry operation of the crystal. PC refinement was carried out at two different resolution ranges for 150 of the highest grid points of the rotation function. Two of the 300 PC refinements produced significantly higher PC values which corresponded to the correct orientations of the two molecules in the asymmetric unit. In both cases, PC refinement modified certain interdomain angles between 10 and 20°; the elbow angle between the variable and constant domains increased by about 10°. The success of PC refinement was resolution-dependent: one of the ori-

entations was obtained at 15–4 Å resolution, the other at 15–3.5 Å. This can be understood by the "landscape" of PC as a function of the refined rigid body parameters. The correct solution emerges as the global maximum and its location is resolution-independent. However, there are a number of local maxima in which PC refinement can get trapped. The location of these local maxima is resolution-dependent.

It would seem that the multiple maxima problem for PC refinement could be overcome by using predominantly low-resolution reflections in the refinement process. Indeed, when PC is plotted as a function of the elbow angle and resolution range (shown for AN02 (35) in Fig. 3) the width of the global maximum of PC increases with decreasing resolution range. The elbow angle of the AN02 crystal structure was modified between 130 and 180° by rotating the C_L, C_H1 dimer around an axis that intersects the two linker regions while keeping the V_H, V_L dimer stationary (cf. Fig. 1). However, this analysis is misleading since it addresses only a single degree of freedom. Observed differences between Fabs include not only the elbow angle but also other interdomain angles and positions between the domains of the Fab. Thus, in PC refinement these additional degrees of freedom also must be optimized, giving rise to local maxima in multidimensional space that persist and become even more pronounced at low resolution (Krukowski and Brünger, unpublished results). These local maxima probably correspond to out-of-register superpositions of the β-sheet pattern of the Fab domains.

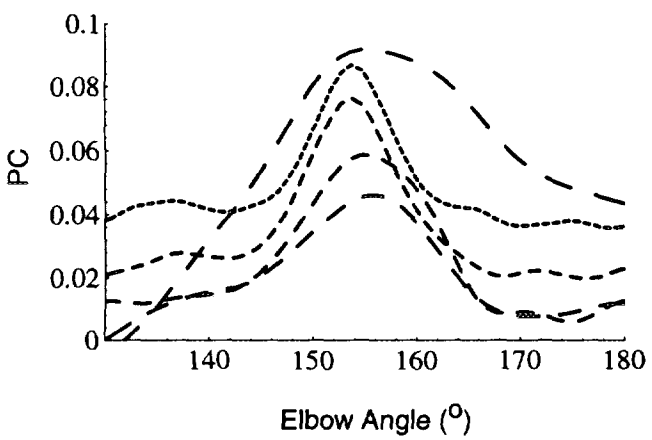

FIG. 3. PC as a function of the elbow angle between an artificially modified AN02 structure and the AN02 crystal structure (34, 35) (with elbow angle 153.6°). The elbow angle is defined as the angle between the directions of the pseudo-twofold axes of symmetry of the V_L–V_H and C_L–C_H1 domain pairs of the Fab (see also Fig. 1). The elbow angle was modified by rotating the C_L, C_H1 dimer around an axis that intersects the Cα positions of particular residues in the two linker regions while keeping the V_H, V_L dimer stationary. PC was evaluated at 15–2.5 (narrow dashes), 15–4, 15–6, 15–8, 15–11 Å (wide dashes) resolution. At low resolution (15–8 and 15–11 Å), direct summation of atomic structure factors was used, whereas at higher resolution a fast Fourier transformation method (21, 62) was used to compute E_m (cf. Eq. [3]).

As a consequence of the local maxima, the radius of convergence of PC refinement is resolution-dependent. Figure 4a shows the rms (root-mean-square) differences of the modified AN02 structure before and after PC refinement at 15–4 Å resolution; Fig. 4b shows the values of PC after PC refinement. It appears that PC refinement is capable of undoing the artificial elbow angle change of AN02 up to about 13°, in agreement with previous results (35). It should be noted that this result is model- and data-sensitive. For example, PC refinement of a model for the chimeric Fab fragment of a tumor cell binding antibody accomplished a modification of the elbow angle by about 18° (42). However, this large radius of convergence was accomplished only after substantial improvement of the search model and by using a complete data set to 3.1-Å resolution data.

Figure 4c shows the rms differences for PC refinement of the modified AN02 structure starting at 15–11 Å resolution followed by gradual extension to 15–4 Å resolution. Figure 4d shows the values of PC at 15–11 Å resolution after the PC refinements. PC refinement does not converge to the correct solution even when the initial elbow angle is close to the correct one (153.6°). Furthermore, incorrect configurations of the four Fab domains can be found with PC values at 15–11 Å resolution larger than that of the correct solution (compare Figs. 3 and 4d). It should be pointed out that all PC refinements included a packing function that prevented the domains from penetrating each other; thus, the incorrect configurations obtained in Fig. 4d are chemically reasonable. It appears that PC refinement at lower resolution (15–11 Å) has a much poorer radius of convergence than at higher resolution (15–4 Å).

Yeates and Rini argued that the inclusion of crystallographically related symmetry mates in the intensity refinement (Eq. [4]) improves convergence behavior (16). The crystallographic symmetry operations can also be included in PC refinement by replacing $E_m(\mathbf{h})$ in Eq. [3] by

$$\sum_j |E_m(S_j\mathbf{h})|^2, \qquad [8]$$

where the sum extends over all crystallographic symmetry operators S_j. Note that the complex structure factors of the symmetry-related molecules cannot be added because the positions of the symmetry elements are not yet known. Thus, the contribution of symmetry-related molecules can be included only by summation of amplitudes or intensities (16). The data in Fig. 5 suggest that the inclusion of crystallographic symmetry does not improve the radius of convergence of PC refinement. In fact, the inclusion of the symmetry mates simply scales PC by a factor of about 3.25 (Fig. 5a). Upon division by this factor, the PC values using crystallographic symmetry are almost identical to the PC values computed without considering symmetry. Therefore, the inclusion of crystallographic symmetry in the computation of PC appears to be redundant. However, it is conceivable that the incorporation of molecules related by noncrystallographic symmetry might be useful to enhance the signal-to-noise ratio in analogy to the "locked" rotation function (44).

PC refinement improves the significance of the translation function by making the search model more accurate (15, 16). For example, in the case of the 26-10 Fab fragment, the PC-refined model produced the

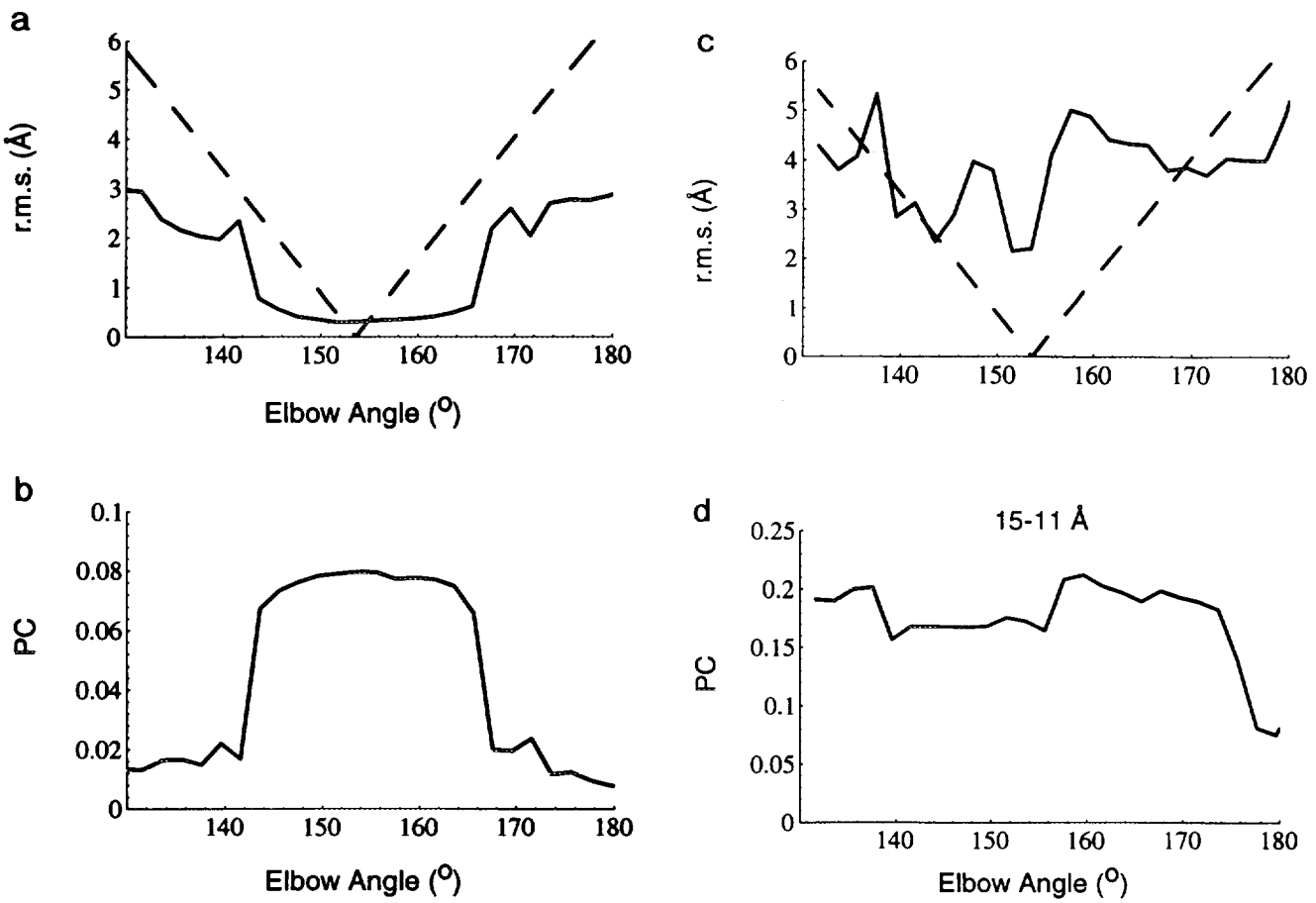

FIG. 4. Radius of convergence of PC refinement of AN02 at 15–4 and 15–11 Å resolution. The refinement proceeded against the sum of PC (Eq. [3]) and the van der Waals energy between the backbone atoms of the Fab domains. The latter restraint prevents the formation of overlapping configurations of the Fab domains. (a) The rms difference for Cα atoms of an artificially modified AN02 structure before (dashed lines) and after (solid lines) PC refinement at 15–4 Å resolution vs the initial elbow angle. The elbow angle of the AN02 structure was modified by rotating the C_L, C_H1 dimer around an axis that intersects the Cα positions of particular residues in the two linker regions while keeping the V_H, V_L dimer stationary. (b) The value of PC after PC refinement at 15–4 Å resolution vs the initial elbow angle. (c) Same as (a) for PC refinement starting at low resolution followed by extension to high resolution. PC refinements were first carried out at 15–11 Å resolution, followed by 15–8, 15–6, and, finally, 15–4 Å resolution. At low resolution (15–8 and 15–11 Å) direct summation of atomic structure factors was used, whereas at higher resolution a fast Fourier transformation method (21, 62) was used to compute E_m. (d) The value of PC after PC refinement at 15–11 Å resolution vs the initial elbow angle.

correct peak with seven standard deviations above the next highest peak, whereas the correctly oriented but unrefined search model produced a noisy translation function (37) that was very difficult to interpret. Similar improvements were reported when intensity-based refinement was used (16).

Translation Functions

Cygler and Anderson (45) compared translation functions for Fabs based on Eq. [5] using the Crowther and Blow method (23), the correlation coefficient (Eq. [7]), and the phased approach. They found that the performance of the Crowther and Blow translation function was rather poor when only a quarter of the unknown Fab fragment was used. The phased translation function appeared to work best, although the comparison with the correlation coefficient-based translation function was hampered by the use of a limited resolution range for the latter one.

Generally, the success of the translation function critically depends on the correct orientation of the molecule and the similarity between the search model and the crystal structure. PC refinement or intensity-based refinement can result in a substantial improvement of the search model. If the intervening PC refinement is successful, the signal-to-noise ratio is very large regardless of the type of translation function and the parameters used (resolution range and diffraction data included).

Choice of Model

The success of molecular replacement largely depends on how well the search model describes the unknown crystal structure. While generalized molecular replacement can reduce differences between the search model and the crystal structure due to different interdomain orientations and positions, the domains themselves must match well for a successful application of the method. Modeling the unknown portions of the molecule through energy minimization was successful in at least one case (42). It is a general but unexplained observation that well-refined search models produce more significant translation and rotation function solutions despite the fact that the atomic rms differences between the search model and the crystal structure can be larger than the rms differences between a highly refined model and a less well-refined model.

Ideally, knowledge about all known structures of immunoglobulin domains with a high degree of homology to the unknown structure should be used. Leahy and co-workers (46) used a "composite" search model consisting of 10 different immunoglobulin light-chain variable domains to solve the soluble form of the human T-cell co-receptor CD8 at 2.6-Å resolution. By giving greater weight to highly conserved features and down-weighting variable regions such as the CDR loops, this composite model was more representative of the particular crystal structure and permitted inclusion of more information during molecular replacement. In addition, the resulting electron maps showed less model bias toward the search model.

Determination of Antigen Structure

Once the structure of the antibody has been solved by molecular replacement, it is possible to obtain the

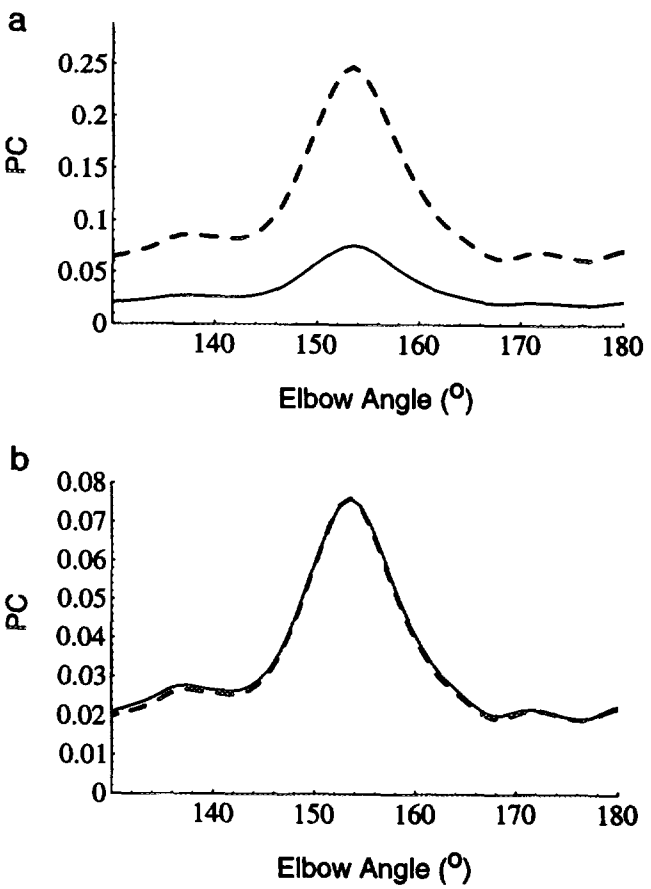

FIG. 5. (a) PC as a function of the elbow angle between an artificially modified AN02 structure and the AN02 crystal structure (34, 35). The elbow angle of the AN02 structure was modified by rotating the C_L, C_H1 dimer around an axis that intersects the Cα positions of residues in the two linker regions while keeping the V_H, V_L dimer stationary. Dashed lines incorporate the symmetry-related molecules using Eq. [8], whereas solid lines use Eq. [3]. (b) Same as (a) except that the dashed line has been scaled by a factor of 1/3.25.

conformation of an antigen bound to the active site. This is relatively straightforward by standard difference map techniques when the antigen is small (e.g., Ref. 35). Care must be taken to avoid artificial movement of CDR side chains into the antigen density during refinement without the antigen being present. This movement can easily obscure the location and conformation of the antigen molecule. The problem can be alleviated by applying harmonic restraints to the side-chain atoms near the antigen binding site prior to inclusion of the antigen (21).

The question arises of what percentage of the unit cell content can be successfully phased by this bootstrap procedure. The solution of a DNase–actin complex (47) starting with the known structure of DNase and boot-strapping the actin structure showed that about 50% of a model can be obtained from the known portion of the model. Test calculations on the HyHEL5–lysozyme complex (48) indicated that the lysozyme molecule could have been traced after generalized molecular replacement with the Fab fragment alone and without knowledge of the lysozyme structure (Bussiere and Brünger, unpublished results). Conversely, Sheriff and co-workers showed that they could locate lysozyme, which accounts for only 23% of the entire HyHEL5–lysozyme complex, by molecular replacement without knowledge of the Fab (49).

Recently Solved Structures

A large number of Fab fragments and complexes with antigens have been solved by various molecular replacement methods in the past 2 years. In addition to the structures already mentioned in this review, they include a specific antibody (Fab 17/9) to a peptide immunogen from influenza virus hemagglutinin (50), a FvD1.3 fragment and its complex with egg-white lysozyme (51), the complex between the monoclonal anti-lysozyme Fab D1.3 and the anti-idiotopic FabE225 (52), the Fab fragment of an anti-phenylarsonate monoclonal antibody (36–71) (53), the HyHEL-10 Fab–lysozyme complex (54), and an Fab specific for a *Salmonella O*-polysaccharide antigen (55). It should be noted that this is only a partial list.

CONCLUSION

Recent successes in solving antibody crystal structures suggest that the X-ray structure determination of Fab fragments of antibodies may now be carried out much more efficiently than previously possible. Studies of antibodies or their complexes with antigens play an increasing role in rational drug design using monoclonal technology (56, 57). Molecular replacement based on immunoglobulin folds is not limited to Fab fragments. For example, the CD2, CD4, and CD8 classes of cell adhesion molecules exhibit immunoglobulin folds (46, 58–60). Indeed, the soluble form of the human T-cell co-receptor CD8 was solved by molecular replacement (46).

The lessons learned from applying molecular replacement to Fabs might also be useful for other multidomain proteins with flexible linkers or hinges between domains.

A remaining challenge is the multiple maxima problem that is encountered in the generalized molecular replacement strategy: if the search model is too far from the crystal structure, PC refinement is likely to get trapped in local maxima. Modern optimization techniques such as simulated annealing (61) might be able to overcome this problem. Another challenge is the appropriate inclusion of the combined knowledge of all immunoglobulin structures known to date. Composite search models might address this problem (46). Thus, we are not too far from considering the crystallographic phase problem solved for immunoglobulin molecules.

ACKNOWLEDGMENTS

I am grateful to Drs. Jiangsheng Jiang, Temple Burling, and Paul Adams and to Mr. Luke Rice for critically reading the manuscript. The author thanks the Pittsburgh Supercomputer Center for support (Grant DMB890008P).

REFERENCES

1. Karle, J., and Hauptman, H. (1950) *Acta Crystallogr.* **3,** 181–187.
2. Watson, J. D. (1990) *Science* **248,** 44–49.
3. Taylor, W. R. (1988) *Protein Eng.* **2,** 77–86.
4. Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E., and Thornton, J. M. (1987) *Nature* **326,** 347–352.
5. Richardson, J. S. (1985) *Methods Enzymol.* **115,** 341–380.
6. Hoppe, W. (1957) *Acta Crystallogr.* **10,** 750–751.
7. Rossmann, M. G., and Blow, D. M. (1962) *Acta Crystallogr.* **15,** 24–31.
8. Huber, R. (1985) *in* Molecular Replacement: Proceedings of the Daresbury Study Weekend (Machin, P. A., Ed.), pp. 58–61, Sci-

ence and Engineering Research Council, The Librarian, Daresbury Laboratory, Daresbury, United Kingdom, February 1985.
9. Chothia, C., and Lesk, A. M. (1986) *EMBO J.* **5,** 823–826.
10. Amzel, L. M., and Poljak, R. J. (1979) *Annu. Rev. Biochem.* **48,** 961–997.
11. Davies, D. R., Padlan, E. A., and Sheriff, S. (1990) *Annu. Rev. Biochem.* **59,** 439–473.
12. Wilson, I. A., Rini, J. M., Fremont, D. H., Fieser, G. G., and Stura, E. A. (1991) *Methods Enzymol.* **203,** 153–176.
13. Branden, C., and Tooze, J. (1991) Introduction to Protein Structure, p. 302, Garland Publishing, New York.
14. Chothia, C., and Lesk, A. (1987) *J. Mol. Biol.* **196,** 901–917.
15. Brünger, A. T. (1990) *Acta Crystallogr. A* **46,** 46–57.
16. Yeates, T. O., and Rini, J. M. (1990) *Acta Crystallogr. A* **46,** 352–359.
17. Huber, R. (1965) *Acta Crystallogr. A* **19,** 353–356.
18. Steigemann, W. (1974) Ph.D. thesis, Technische Universität München.
19. Crowther, R. A. (1972) *in* The Molecular Replacement Method, Int. Sci. Rev. No. 13 (Rossmann, M. G., Ed.), Gordon & Breach, New York.
20. Hauptman, H. (1982) *Acta Crystallogr. A* **38,** 289–294.
21. Brünger, A. T. (1992) X-PLOR: A System for X-Ray Crystallography and NMR, Yale Univ. Press, New Haven, CT.
22. Hendrickson, W. A. (1985) *Methods Enzymol.* **115,** 252–270.
23. Crowther, R. A., and Blow, D. M. (1967) *Acta Crystallogr.* **23,** 544–548.
24. Harada, Y., Lifchitz, A., Berthou, J., and Jolles, P. (1981) *Acta Crystallogr. A* **37,** 398–406.
25. Fujinaga, M., and Read, R. J. (1987) *J. Appl. Crystallogr.* **20,** 517–521.
26. Colman, P. M., Fehlhammer, H., and Bartels, K. (1976) *in* Crystallographic Computing Techniques (Ahmed, F. R., Huml, K., and Sedláček, B., Eds.), pp. 248–259, Munksgaard, Copenhagen.
27. Doesburg, H. M., and Beurskens, P. T. (1983) *Acta Crystallogr. A* **39,** 368–376.
28. Read, R. J., and Schierbeek, A. J. (1988) *J. Appl. Crystallogr.* **21,** 490–495.
29. Cygler, M., and Desrochers, M. (1989) *Acta Crystallogr. A* **45,** 563–572.
30. Bentley, G. A., and Houdusse, A. (1992) *Acta Crystallogr. A* **48,** 312–322.
31. Beurskens, P. T., Gould, R. O., Bruins Slot, H. J., and Bosman, W. P. (1987) *Z. Kristallogr.* **179,** 127–159.
32. Karle, I. L., and Karle, J. (1971) *Acta Crystallogr. B* **26,** 1891–1898.
33. Cygler, M., and Anderson, W. F. (1988) *Acta Crystallogr. A* **44,** 38–45.
34. Leahy, D. J., Hynes, T. R., McConnell, H. M., and Fox, R. O. (1988) *J. Mol. Biol.* **203,** 829–830.
35. Brünger, A. T., Leahy, D. J., Hynes, T. R., and Fox, R. O. (1991) *J. Mol. Biol.* **221,** 239–256.
36. Strong, R. K. (1990) Ph.D. thesis, Harvard University.
37. Bruñger, A. T. (1991) *Acta Crystallogr. A* **47,** 195–204.
38. Fitzgerald, P. (1988) *J. Appl. Crystallogr.* **21,** 273–278.
39. Garcia, K. C., Ronco, P., Verroust, P. J., and Amzel, L. M. (1989) *J. Biol. Chem.* **264,** 20463–20466.
40. Garcia, K. C., Ronco, P. M., Verroust, P. J., Brünger, A. T., and Amzel, L. M. (1992) *Science* **257,** 502–507.
41. Stevens, F. J., Westholm, F. A., Panagiotopoulos, N., Schiffer, M., Popp, R. A., and Solomon, A. (1981) *J. Mol. Biol.* **147,** 185–193.
42. Brady, R. L., Edwards, D. J., Hubbard, R. E., Jiang, J.-S., Lange, G., Roberts, S. M., Todd, R. J., Adair, J. R., Emtage, J. S., King, D. J., and Low, D. C. (1992) *J. Mol. Biol.* **227,** 253–264.
43. Rini, J. M., Stanfield, R. L., Takimoto-Kamimura, M., Stura, E. A., Profy, A. T., and Wilson, I. A. (1992). Preprint.
44. Tong, L., and Rossmann, M. G. (1990) *Acta Crystallogr. A* **46,** 783–792.
45. Cygler, M., and Anderson, W. F. (1988) *Acta Crystallogr. A* **44,** 300–308.
46. Leahy, D. J., Axel, R., and Hendickson, W. A. (1992) *Cell* **68,** 1145–1162.
47. Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F., and Holmes, K. C. (1990) *Nature* **347,** 37–44.
48. Sheriff, S., Silverton, E. W., Padlan, E. A., Cohen, G. H., Smith-Gill, S. J., Finzel, B., and Davies, D. R. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 8075–8079.
49. Sheriff, S., Padlan, E. A., Cohen, G. H., and Davies, D. R. (1990) *Acta Crystallogr. B* **46,** 418–425.
50. Rini, J. M., Schulze-Gahmen, U., and Wilson, I. A. (1992) *Science* **255,** 959–965.
51. Bhat, T. N., Bentley, G. A., Fischmann, T. O., Boulot, G., and Poljak, R. J. (1990) *Nature* **347,** 483–485.
52. Bentley, G. A., Boulot, G., Riottot, M. M., and Poljak, R. J. (1990) *Nature* **348,** 254–257.
53. Rose, D. R., Strong, R. K., Margolies, M. N., Gefter, M. L., and Petsko, G. A. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 338–342.
54. Padlan, E. A., Silverton, E. W., Sheriff, S., Cohen, G. H., Smith-Gill, S. J., and Davies, D. R. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5938–5942.
55. Rose, D. R., Cygler, M., To, R. J., Przybylska, M., Sinnott, B., and Bundle, D. R. (1990) *J. Mol. Biol.* **215,** 489–492.
56. Tramontano, A., Jonda, K. D., and Lerner, R. A. (1986) *Science* **234,** 1566–1570.
57. Wolff, M. E., and McPherson, A. (1990) *Science* **345,** 365–366.
58. Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K., and Stuart, D. I. (1992) *Nature* **360,** 232–239.
59. Wang, J., Yan, Y., Garrett, T. P. J., Liu, J., Rodgers, D. W., Garlick, R. L., Tarr, G. E., Husain, Y., Reinherz, E. L., and Harrison, S. C. (1990) *Nature* **348,** 411–418.
60. Ryu, S.-E., Kwong, P. D., Truneh, A., Porter, T. G., Arthos, J., Rosenberg, M., Dai, X., Xuong, N.-H., Axel, R., Sweet, R. W., and Hendrickson, W. A. (1990) *Nature* **348,** 419–426.
61. Kirkpatrick, S., Gelatt, C. D., Jr., and Vecchi, M. P. (1983) *Science* **220,** 671–680.
62. Brünger, A. T. (1989) *Acta Crystallogr. A* **45,** 42–50.