Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

On the molecular-replacement problem in the presence of merohedral twinning: structure of the N-terminal half-molecule of human lactoferrin

Wendy A. Breyer,† Richard L. Kingston,‡ Bryan F. Anderson and Edward N. Baker*§

Department of Biochemistry, Massey University, New Zealand

† Current address: Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, USA.
‡ Current address: Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA.
§ Current address: School of Biological Sciences, University of Auckland, Auckland, New Zealand.

Correspondence e-mail: ted.baker@auckland.ac.nz

Received 6 April 1998
Accepted 14 July 1998

The structure of a hemihedrally twinned protein crystal with two molecules in the asymmetric unit was solved by molecular replacement. The protein, a site-specific mutant of the N-terminal half-molecule of human lactoferrin, is able to undergo an internal rigid-body domain motion. Therefore, determining the structure required the independent positioning of four protein domains. The molecular-replacement solutions were obtained using a conventional real-space rotation function, and a translation function based on the linear correlation coefficient. Once the molecules were positioned, it was necessary to assign them to the appropriate twin domain. Several methods for doing this are described, one of which leads to a determination of the volume of each twin domain. In the appendix to the paper we discuss the interpretation of the self-rotation function in the presence of merohedral twinning.

1. Introduction

A twin is a composite crystal, containing consistently misoriented regions (twin domains) which are related by a symmetry operation (twinning operation). In the case of merohedral twinning, the diffraction patterns of the twin domains are exactly superimposed. This kind of twinning occurs when the twinning operations are members of the lattice point group (holohedry) but not the crystallographic point group (merohedry). Hemihedral twinning is a special case of merohedral twinning in which there are only two twin domains. Issues relating to crystal twinning in protein crystallography have been recently reviewed by Yeates (1997).

Diffraction data collected from a hemihedrally twinned crystal do not represent the true crystallographic intensities. The observed intensities are a linear combination of the true intensities (assuming that the size of the twin domains is large relative to the coherence length of the X-ray beam, so that interference effects can be neglected). Thus,

$$I_{\rm obs}(\mathbf{h}_1) = (1-\alpha)I(\mathbf{h}_1) + \alpha I(\mathbf{h}_2) \quad (1a)$$

$$I_{\rm obs}(\mathbf{h}_2) = \alpha I(\mathbf{h}_1) + (1-\alpha)I(\mathbf{h}_2), \quad (1b)$$

where $\mathbf{h}_1$ and $\mathbf{h}_2$ are reciprocal-lattice vectors that are related by the twinning operation, $I(\mathbf{h}_1)$ and $I(\mathbf{h}_2)$ are the true intensities associated with these lattice vectors, $I_{\rm obs}(\mathbf{h}_1)$ and $I_{\rm obs}(\mathbf{h}_2)$ are the observed intensities and α is the twin fraction (by convention, the fractional volume of the smaller twin domain with respect to the whole crystal, hence $0 < \alpha < 0.5$). When α is 0.50, the twin domains are of equal size and the twinning is said to be perfect. Perfect hemihedral twinning gives rise to higher order symmetry in the diffraction pattern than would result from the crystal space group alone.

When α is small, (1*a*) and (1*b*) represent a system of linear equations that can be solved for the true crystallographic intensities (see, for example, Britton, 1972; Igarashi *et al.*, 1997). However as α approaches 0.5, the system of equations becomes degenerate and it is impossible to obtain reliable estimates of the true intensities.

The method of molecular replacement has allowed the determination of a number of protein and viral structures using hemihedrally twinned crystals from which it was not possible to recover the true intensities (see, for example, Redinbo & Yeates, 1993; Lea & Stuart, 1995; Gomis-Rüth *et al.*, 1995). The interpretation of the cross-rotation function calculated from a hemihedrally twinned crystal is generally straightforward. It follows from (1*a*) and (1*b*) that the Patterson function calculated from the observed intensities of a hemihedrally twinned crystal is the superposition of the Patterson functions of the two twin domains (weighted by α, the twin fraction). Hemihedral twinning doubles the number of apparent molecular orientations and, therefore, doubles the number of solutions in the cross-rotation function. Assuming that the oriented molecules can be positioned using a conventional translation function, the problem that remains is the correct assignment of the molecular-replacement solutions to each of the twin domains in the crystal. We will discuss methods for doing this when the twinning is both perfect and imperfect.

In this paper, we describe crystals of a site-specific mutant of the N-terminal half-molecule of human lactoferrin which are hemihedrally twinned. Lactoferrin is an iron-binding protein and a member of the transferrin family (Baker, 1994). The molecule is divided into N-terminal and C-terminal lobes, which have clear sequence and structural homology. Each lobe (or half-molecule) is further sub-divided into two similarly folded domains comprising ~160 amino acids each (Fig. 1). The iron-binding site resides in the interdomain cleft, and opening and closing of this cleft by relative movement of the two domains is associated with iron binding and release (Anderson *et al.*, 1990; Gerstein *et al.*, 1993). Iron binding to lactoferrin depends on the synergistic binding of a carbonate anion in close proximity to the bound metal. In order to facilitate the study of lactoferrin, a recombinant N-terminal half-molecule has been constructed and characterized (Day *et al.*, 1992, 1993). The site-directed mutant of the N-terminal half-molecule discussed here has Arg121, which is involved in anion binding, mutated to Asp. We will refer to this mutant as R121D.

2. Protein expression, crystallization and X-ray data collection

The construction, expression and purification of the R121D site-specific mutant was accomplished by previously described methods (Faber *et al.*, 1996). Crystals were grown by micro-dialysis of protein solution against a solution containing 0.01 *M* Tris–HCl buffer pH 8.0 and either 12% 2-propanol or 12% ethanol. Although the lactoferrin in the crystallization trials contained iron, the crystals grew as colourless needles, consistent with loss of bound iron by the protein.

Diffraction data were collected at room temperature from a single-crystal of R121D conventionally mounted in a thin-walled glass capillary. Data were collected on an R-AXIS II-C system, utilizing Cu $K\alpha$ radiation from a Rigaku rotating-anode generator and using a Fuji imaging plate as a detector. Profile-fitted intensities were obtained from the oscillation images with the program *DENZO* (Otwinowski & Minor, 1997), and the data was scaled and merged using programs from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994).

X-ray diffraction data were indexed on a primitive lattice with cell parameters $a = b = 151.3$, $c = 48.6$ Å, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$. The diffraction pattern exhibits 6/*m* symmetry. The systematic absences ($l = 3n$) along the reciprocal-space vector 00*l* indicated space group $P6_2$ or its enantiomorph $P6_4$. Consequently, the initial data reduction assumed the presence of hexagonal symmetry. However, the crystals are truly trigonal (space group $P3_1$) with near-perfect hemihedral twinning, giving rise to the observed 6/*m* symmetry in the diffraction pattern. When this was discovered, the data were reprocessed imposing only trigonal symmetry. Statistics relating to the data processing are presented in Table 1. Diffraction from crystals of the R121D mutant extends to approximately 3 Å resolution.

3. Detection of twinning

Preliminary attempts at molecular replacement in space groups $P6_2$ and $P6_4$ suggested the presence of twinning. Clear solutions were obtained in rotation- and translation-function calculations, but the positioned molecules significantly overlapped (results not shown). A subsequent examination of intensity statistics indicated that the R121D crystals were hemihedrally twinned. Stanley has discussed the distribution of intensity measurements in the presence of perfect hemihedral twinning (Stanley, 1972). The expected value of the Wilson ratio,

$$\langle|\mathbf{E}|\rangle^2/\langle|\mathbf{E}|^2\rangle, \qquad (2)$$

calculated from the acentric data of single and hemihedrally twinned crystals is 0.785 and 0.885, respectively (here **E** represents the normalized structure factor). For the R121D data, the value of the Wilson ratio (evaluated over thin resolution shells to avoid the need for normalization) was 0.88. With a hemihedrally twinned protein crystal, the apparent Laue symmetry and systematic absences could only result from space groups *P*31 or *P*32. The twinning operation is a 180° rotation around an axis parallel to the crystallographic screw rotation (Koch, 1992). The apparent 6/*m* Laue symmetry and the observed value of the Wilson ratio both indicated that the hemihedral twinning in R121D was near-perfect ($\alpha \equiv 0.5$).

To precisely evaluate the twin fraction α, two statistical methods were employed. Firstly, inspection of the cumulative distribution of the statistic *H*, where *H* is defined as

$$H = |[I_{obs}(\mathbf{h}_2) - I_{obs}(\mathbf{h}_1)]|/[I_{obs}(\mathbf{h}_2) + I_{obs}(\mathbf{h}_1)], \qquad (3)$$

allows the determination of the twin fraction of a hemihedrally twinned crystal (Yeates, 1988). A plot of the cumulative distribution of H for the R121D diffraction data is shown in Fig. 2, along with the expected distributions for varying values of α. Inspection of this figure suggests a twin fraction for the crystals of ~0.43.

Another statistic for evaluating the twin fraction was suggested by Rees (1980), and is based on analysis of cumulative intensity distributions. In both test calculations and application to the actual data, this statistic failed to provide quantitative information (results not shown). However, the presence of hemihedral twinning was clearly indicated. This result is in agreement with the observations of others (Gomis-Rüth *et al.*, 1995).

Table 1
Data-processing statistics for R121D.

66 oscillation images collected from a single crystal were used in processing.

Upper resolution limit (Å)	Total unique observations	R_{merge}	$I/\sigma(I)$	Completeness (%)	Multiplicity
11.17	468	0.034	9.0	97	3.1
7.92	846	0.041	17.2	98	3.1
6.47	1135	0.057	12.4	100	3.1
5.61	1305	0.071	10.1	98	3.1
5.02	1470	0.073	10.0	98	3.1
4.58	1641	0.077	9.3	99	3.1
4.24	1784	0.087	8.3	99	3.1
3.97	1852	0.100	7.3	95	3.0
3.74	2052	0.119	6.2	99	2.8
3.55	2121	0.150	5.0	97	2.7
3.39	2155	0.179	4.3	94	2.6
3.24	2306	0.229	3.3	96	2.6
3.11	2319	0.315	2.4	93	2.4
3.00	2428	0.418	1.8	93	2.4
2.90	2469	0.522	1.4	92	2.2
Overall	26351	0.115	6.1	96	2.8

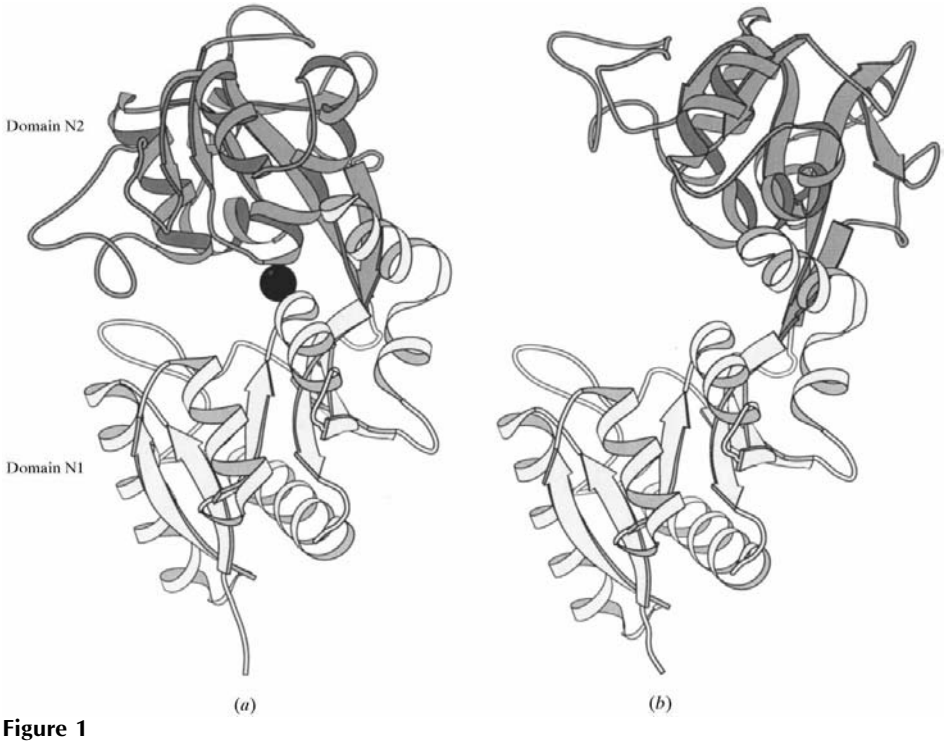

Figure 1
The rigid-body domain motion observed in lactoferrin. The closed (*a*) and open (*b*) states of the N-terminal half-molecule are illustrated. The two domains that comprise the half-molecule are shown in light and dark gray. Also shown is the position of the Fe atom in the inter-domain cleft in the closed structure. The diagram was prepared using the program *MOLSCRIPT* (Kraulis, 1991), and the atomic coordinates for the iron-bound and iron-free forms of human lactoferrin (Protein Data Bank codes 1LFG and 1LFH, respectively, Anderson *et al.*, 1989, 1990).

4. Molecular replacement

Molecular-replacement calculations used the refined structure of the N-terminal half-molecule of lactoferrin as the search model (Day *et al.*, 1993). The R121D crystals were colourless, indicating that iron was not bound. Therefore, it was believed that the molecules present in the crystal might be in the open state (Fig. 1). Analysis of the domain closure in lactoferrin suggests stabilization of the open and closed states by close-packed interfaces (Gerstein *et al.*, 1993). However, in other proteins that undergo hinged rigid-body domain motions, multiple conformational states have been seen within a single crystal (Faber & Matthews, 1990). Independent positioning of the domains was therefore carried out as a precaution against obtaining a partially correct molecular-replacement solution. Domain N1 consisted of residues 4–90 and 251–320, and domain N2 consisted of residues 91–250 (Fig. 1).

4.1. The cross-rotation function

A real-space cross-rotation function was evaluated over the asymmetric unit (Rao *et al.*, 1980) for each domain using the program *X-PLOR* (Brünger, 1993). The highest peaks of the rotation function were used in Patterson correlation refinement (Brünger, 1990). Solutions were readily found for the N2 domains; however, solutions for the N1 domains were less clear. As expected, each cross-rotation-function solution had a complementary solution related by the twinning operation. However, for each domain there were no more than two unique solutions.

4.2. The translation function

All translation-function calculations employed the program *BRUTE* (Fujinaga & Read, 1987). In this program the linear correlation coefficient between observed and calculated structure-factor amplitudes is evaluated as the oriented search model is translated through the cell. The program was modified so that hemihedral twinning could be explicitly accounted for (*i.e.*, the calculated structure-factor amplitudes were twinned before evaluation of the correlation coefficient). For brevity, we refer to this as a 'twinned' translation function and the correlation coefficient on which it is based as the 'twinned' correlation coefficent. Additional input required for the program was the twin fraction α and the twinning operation.

Since the N2 domains gave the clearest signal in the cross-rotation function, these were positioned first. Parallel calculations were carried out in both $P3_1$ and $P3_2$ in order to resolve the space-group ambiguity. For comparison, both a regular translation function and a twinned translation function were computed. The translation function gave clearly contrasted solutions for all four rotations in space group $P3_1$ (Table 2). For space group $P3_2$, the maximum correlation coefficients were much lower for each domain, indicating that this is not the correct space group (results not shown).

For the twinned translation function, the maximum correlation coefficient is consistently higher for each positioned domain than in the regular translation function; however, so is the mean of the function. In both the twinned and conventional translation functions, the correct solution is equivalently contrasted from the mean. The higher mean in the twinned translation function is readily explained by the difference in the expected intensity distributions from single and hemihedrally twinned crystals. Thus, in this case, a translation function that explicitly accounts for hemihedral twinning (as suggested by Redinbo & Yeates, 1993; Yeates, 1997) did not assist in discriminating the correct molecular-replacement solution.

4.3. Assignment of the molecular-replacement solutions to the correct twin domains

As discussed, the presence of hemihedral twinning complicates the molecular-replacement problem. It is necessary to correctly assign the molecular-replacement solutions to the twin domains in the crystal. In the case of perfect hemihedral twinning, with a single molecule in the asymmetric unit, this problem does not arise (see, for example, Redinbo & Yeates, 1993). In this case the assignment is arbitrary.

One assignment method involves examining the relative magnitude of the correlation coefficient in the regular translation function In our case, this immediately suggests which of the solutions might derive from the smaller twin domain in the crystal, and which from the larger (Table 2). This was confirmed by examining the behaviour of the twinned correlation coefficient as the twin fraction was varied, for each of the four solutions (Fig. 3). Clearly, two of the solutions, which show maxima in the correlation coefficient at $a = 0.43$, belong in the larger twin domain. Conversely, the other two solutions, with maxima at $a = 0.57$, result from the smaller twin domain. Not only does this clearly assign the molecular-replacement solutions to the twin domains of the crystal, but it also provides an estimate of the twin fraction of the crystal (for further discussion, see Gomis-Rüth *et al.*, 1995).

After the molecular-replacement solutions were assigned to the correct twin domain, the two appropriate N2 domains were positioned with respect to a common origin. Patterson subtraction techniques (Zhang & Matthews, 1994), in which Patterson vectors arising from the positioned molecules are subtracted from the observed Patterson function, were then used to clarify the rotation-function results for the N1 domain. Again, two pairs of cross-rotation-function solutions were obtained, with the members of each pair related by the twinning operation. The translation function was unambiguous. The positioned N1 domains were assigned to the twin domains of the crystal as described for the N2 domains. Finally, the position of the N1 domains was determined with respect to the common origin.

Two molecules in the asymmetric unit corresponds to an unusually high Matthews coefficient (5.62 Å^3 Da^{-1}) or, equivalently, a solvent content of approximately 78% (Matthews, 1968). Therefore, we expected that there would be three (V_m of 3.75 Å^3 Da^{-1}, solvent content of 67%) or perhaps four (V_m of 2.81 Å^3 Da^{-1}, solvent content of 56%) molecules

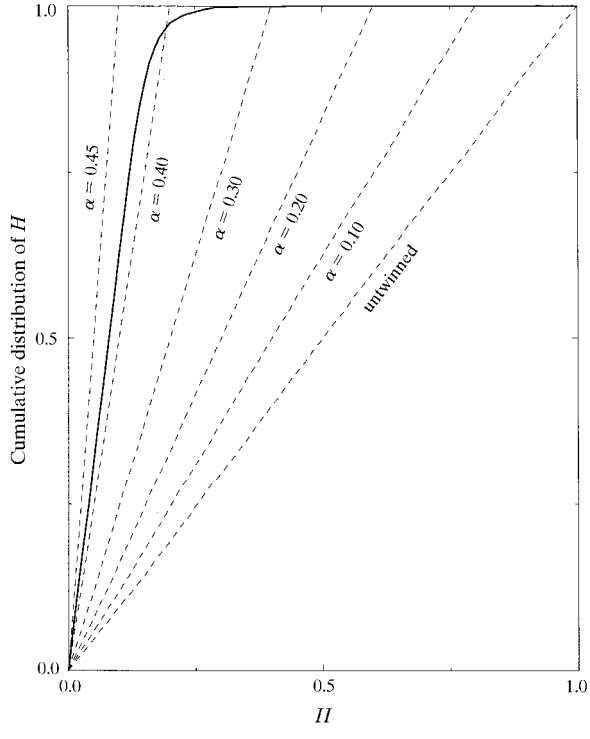

Figure 2
Estimation of the twin fraction (α) by comparison of twin-related acentric observations. The parameter H (Yeates, 1988) is a function of the intensity measurements related by the twinning operation (equation 3); its cumulative distribution for the R121D diffraction data is shown (solid line). Also shown are the expected cumulative distributions for varying values of the twin fraction (dashed lines). The strongest 44% of paired acentric observations were used to calculate the parameter H.

Table 2
Translation function for the oriented N2 domains in space group $P3_1$.

The translation function was calculated using terms from 8–4 Å resolution. The cross-rotation-function solutions A and A' are related by the twinning operation, as are the solutions B and B'. Because $P3_1$ is a polar space group, only a two-dimensional search was needed. The translation function was computed over the region $\mathbf{a} \times \mathbf{b}$, which is larger than the region 1/3 $(\mathbf{a} - \mathbf{b}) \times 1/3$ $(\mathbf{a} + 2\mathbf{b})$ actually required (Hirshfeld, 1968). For each cross-rotation-function solution there was only a single significant peak in the translation function (taking into account the space-group symmetry operations and alternative origin choices). Note that the same translation vector gives rise to the maximum correlation coefficient in both functions. For calculation of the twinned translation function, a twin fraction $\alpha = 0.40$ was assumed.

	Regular translation function			Twinned translation function		
Rotation-function solution	Maximum CC	Mean of function	Standard deviation of function	Maximum CC	Mean of function	Standard deviation of function
A	0.150	0.059	0.010	0.217	0.087	0.012
A'	0.191	0.075	0.010	0.233	0.091	0.013
B	0.151	0.065	0.009	0.226	0.096	0.013
B'	0.194	0.083	0.010	0.235	0.101	0.013

in the asymmetric unit. However, using Patterson subtraction techniques, no further cross-rotation-function solutions could be identified for either the N1 or N2 domains. The final correlation coefficient between observed and calculated structure-factor amplitudes is 0.65 (including all four positioned domains, and data between 8 and 3 Å resolution). If the calculated structure-factor amplitudes are first twinned (with $\alpha = 0.43$), the final correlation coefficient is 0.78.

The methods described here to assign each of the positioned molecules to the twin domains in the crystal rely on α being different from 0.5. However, even in the case of perfect hemihedral twinning it should still be possible to solve complicated molecular-replacement problems.

The total structure factor for each twin domain results from the vector addition of the structure factors due to each positioned molecule in that domain. If a positioned molecule is assigned to the incorrect twin domain, this vector addition will be inappropriate and the correlation with the observed structure-factor amplitudes should be reduced relative to the correct assignment. Calculations with the R121D data, where one protein domain was purposely placed in the wrong twin domain, have confirmed this idea (results not shown). This suggests that even in the case of perfect hemihedral twinning, complicated molecular-replacement problems can be resolved. In addition, many of the possibilities are likely to be eliminated on the basis of simple packing considerations.

5. Verification of the model

The four positioned domains provided a physically reasonable model for the crystal. On inspection, the positioned N1 and N2 domains were seen to form two half-molecules of lactoferrin. Both half-molecules were in an essentially equivalent 'open' conformation, just as in the structure of intact iron-free lactoferrin (Fig. 1). There was no overlap between symmetry-related molecules. As noted before, the calculated solvent content of the crystals is very high. However, the positioned molecules form a plausible crystal lattice, with a large solvent channel in the direction of the unique axis (Fig. 4). In the *Appendix* to this paper, a general method for predicting all peaks of the self-rotation function in the presence of merohedral twinning is derived. The model accounts for all peaks in the self-rotation function computed from the R121D diffraction data. Hence, despite the anomalously high solvent content of the crystals, the model appeared correct.

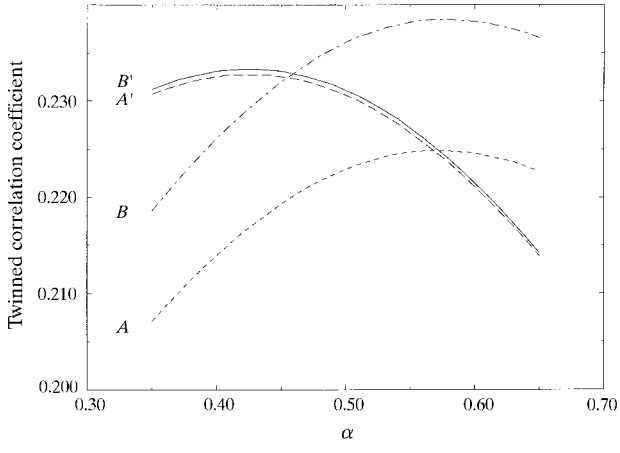

Figure 3
Behaviour of the correlation coefficient as a function of the twin fraction α for the positioned N2 domains of lactoferrin. The four molecular-replacement solutions can be divided into two pairs (A and B). The solutions A and A' are related by the twinning operation as are the solutions B and B'. For each solution, the twinned correlation coefficient was evaluated for values of α between 0.30 and 0.70.

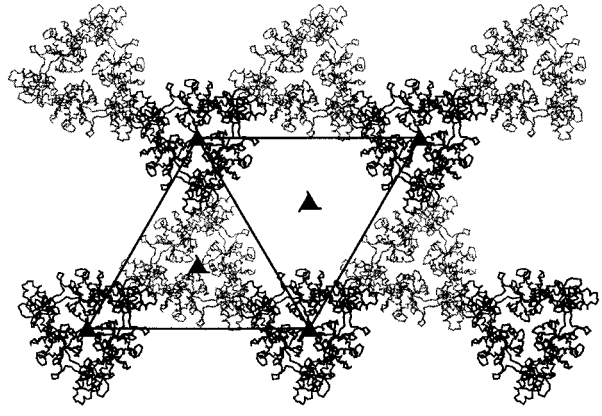

Figure 4
A diagram illustrating the molecular packing in the R121D crystals. A view down the unique axis of the trigonal crystals is shown. The two N-terminal half-molecules of lactoferrin found in the crystallographic asymmetric unit are represented in light and dark gray, respectively. They are related by a 37° rotation almost exactly about z. Also shown are the symmetry elements of space group $P3_1$.

Table 3
Refinement statistics.

Resolution (Å)	20–3.0
Refinement program	*SHELXL*-97
Refinement protocol	Conjugate-gradient least-squares minimization
R	0.150
R_{free}†	0.191
Twin fraction	0.442
Number of parameters	20321
Number of restraints	27728
RMSD bond distances (target) (Å)‡	0.005 (0.010)
RMSD 1–3 distances (target) (Å)‡ (bond-angle restraint)	0.016 (0.020)
NCS restraint targets‡	
RMSD B values (Å^2)	4
RMSD 1–4 distances (Å) (torsion-angle restraint)	0.02
Ramachandran plot§	
Most allowed (%)	82.0
Allowed (%)	16.9
Generously allowed (%)	0.5
Disallowed (%)	0.4

† The data set aside for calculation of R_{free} includes, for each randomly selected reflection, its twin-related reflection. ‡ The target root-mean-square deviations (RMSD) from ideal values are used in the weighting of restraints in least-squares calculations. Restraints on bond and torsion angles are imposed by restraining non-bonded distances. Ideal values are derived from the geometry library of Engh & Huber (1991). Non-crystallographic symmetry (NCS) restraints were applied to the torsion angles of NCS-related molecules and to the B values of NCS-related atoms. Within a molecule, each atom has an individual isotropic B value. Tight restraints were imposed on the differences in the B values of bonded atoms. § The allowed regions of the Ramachandran plot were defined by the program *PROCHECK* (Laskowski *et al.*, 1993).

The structure has now been successfully refined (G. B. Jameson, manuscript in preparation) using the program *SHELXL* (Sheldrick & Schneider, 1997). This program implements a method to refine twinned crystal structures (Pratt *et al.*, 1971; Jameson, 1982). Interactive model building was carried out with the program *TURBO-FRODO* (A. Roussel, A. G. Inisan and C. Cambillau). The initial model contained residues 4–320. Refinement began with values for R and R_{free} of 0.33, which decreased to 0.25 on introduction of a parameter describing the twin fraction. The first electron-density map revealed the entire C-terminal helix (residues 321–332), which was missing in the structure of the iron-bound half-molecule (Day *et al.*, 1993). The current state of the refinement is summarized in Table 3.

6. Summary

We have discussed the molecular-replacement solution of a hemihedrally twinned crystal, in which multiple protein domains were positioned in the asymmetric unit. The principal problem encountered was the assignment of the positioned molecules to the twin domains in the crystal.

Firstly, we consider the case where the twinning is imperfect. The relative height of peaks in the cross-rotation function cannot be used to reliably assign these peaks to the twin domains in the crystal (the overlap of self- and cross-vectors in the Patterson function prevents this). If translation-function solutions can be obtained, then they can be assigned to the larger or smaller twin domains in the crystal using the magnitude of the correlation coefficient between the observed and calculated structure factors. Additionally, the behaviour of the twinned correlation coefficient as a function of the twin fraction will clearly assign the molecular-replacement solutions to the appropriate twin domain. This also serves to give an estimate of the twin fraction of the crystal consistent with the method of Yeates (1988).

Finally, even in the case of perfect hemihedral twinning, it should still be possible to solve complicated molecular-replacement problems. This can be performed by examining the behaviour of the correlation coefficient as the positioned fragments are systematically permuted among the twin domains.

APPENDIX *A*
Interpretation of the self-rotation function in the presence of merohedral twinning

The self-rotation function can be used to detect non-crystallographic rotational symmetry (Rossmann & Blow, 1962). It is defined as the angular auto-correlation function of the Patterson synthesis. Here we consider the implications of merohedral twinning for the interpretation of the self-rotation function.

There is a direct correspondence between the symmetry and orientation of the molecules in a crystal and the intramolecular Patterson vectors (self-vectors). Hence, the problem of predicting all peaks of the rotation function reduces to finding the set of all rotations that leave a molecule invariant or rotate it into the orientation of another molecule in the crystal. For ordinary crystals, a general method to predict all peaks of the self-rotation function has been described by Litvin (1975, 1987). The peaks of the self-rotation function are expressed in terms of the rotational operations of the crystal space group and the point-group symmetry and orientation of the molecules in the asymmetric unit of the crystal.

We extend this method to crystals exhibiting merohedral twinning. Subsequently, we use this result to account for the observed peaks in the self-rotation function calculated from the R121D diffraction data.

A1. The self-rotation function in the absence of twinning

First, we restate Litvin's result, adopting a slightly differing notation for clarity. We consider a crystal made up of identical protein motifs. In this sense, a motif may refer to a single polypeptide or an assembly of distinct polypeptides (*e.g.* an oligomeric protein or a viral capsid). Let P denote the point-group symmetry of the protein motif (noting that the motif may possess no symmetry other than the identity operation). Let $\mathbf{r}$ be a vector corresponding to the centre of mass of a motif (for protein motifs possessing dihedral or cubic point-group symmetry, the centre of mass of the motif will correspond to the point-group centre).

In the asymmetric unit of the crystal, we have vectors corresponding to the centre of mass of q protein motifs. We can denote these vectors as $\mathbf{r}_{11}$, $\mathbf{r}_{12}$, ..., $\mathbf{r}_{1q}$ (the first subscript

indicates that the vectors are found in the selected asymmetric unit of the crystal, the second subscript discriminates between the motifs within the asymmetric unit). For each motif, we denote its corresponding orientation as M_{11}, M_{12}, ..., M_{1q}. We denote the elements of the point group of each motif as $\{P_{11}\}$, $\{P_{12}\}$, ..., $\{P_{1q}\}$.

Let R denote the point group of the crystal space group. The members of this group, the n rotational operators of the space group, are the matrices $[R_1]$, $[R_2]$, ..., $[R_n]$.

We consider the case where $\mathbf{r}_{11}$, $\mathbf{r}_{12}$, ..., $\mathbf{r}_{1q}$ are general position vectors of the space group. We can generate all the possible orientations of the protein motif in the crystal using $[R_1]$, $[R_2]$, ..., $[R_n]$, the rotational operators of the space group.

We have, for ($j = 1, ..., n$), ($\alpha = 1, ..., q$),

$$[R_j]M_{1\alpha} = M_{j\alpha}.$$

Note that these ($q \times n$) orientations are not necessarily distinct.

In addition, we define for all $\alpha = 1, ..., q$ a rotation $[N_\alpha]$ such that

$$[N_\alpha]M_{11} = M_{1\alpha}.$$

These rotations map the first motif in the asymmetric unit into the orientation of the q other equivalent motifs.

It was shown by Litvin (1975) that the peaks of the self-rotation function correspond to the distinct rotations contained in the set of rotations

$$\{S(j\alpha, k\beta)\} = [R_k][N_\beta]\{P_{11}\}([R_j][N_\alpha])^{-1}$$

for ($j, k = 1, ..., n$), ($\alpha, \beta = 1, ..., q$) and every element of the point group $\{P_{11}\}$.

The rotational operators of the space group are defined in terms of the crystallographic basis vectors, which are not in general orthogonal. In practice, all the rotations are most conveniently defined with respect to some orthonormal basis. Thus, if [OM] is the transformation from the crystallographic to the orthonormal frame, we can write

$$\{S(j\alpha, k\beta)\} = [\mathrm{OM}][R_k][\mathrm{OM}]^{-1}[N_\beta]\{P_{11}\} \times ([\mathrm{OM}][R_j][\mathrm{OM}]^{-1}[N_\alpha])^{-1}. \quad (4)$$

A2. The self-rotation function in the presence of merohedral twinning

It is straightforward to extend this result to account for the presence of merohedral twinning. For a merohedrally twinned crystal with m independent twin domains, we denote the rotational operations relating the twin domains as $[T_1]$, $[T_2]$, ..., $[T_m]$. The first matrix, $[T_1]$, is the identity operation. To denote the orientation of the motifs in the crystal we add a third subscript, which indicates the twin domain. Hence, we have for all $j = (1, ..., n)$, $\alpha = (1, ..., q)$, $u = (1, ..., m)$,

$$[N_\alpha]M_{111} = M_{1\alpha 1}$$
$$[R_j]M_{1\alpha 1} = M_{j\alpha 1}$$
$$[T_u]M_{j\alpha 1} = M_{j\alpha u}.$$

It follows that

$$M_{111} = ([T_u][R_j][N_\alpha])^{-1}M_{j\alpha u}.$$

We need to determine the set of rotations which will transform the orientation $M_{j\alpha u}$ into the orientation $M_{k\beta v}$. The unique rotations in this set of rotations comprise the predicted peaks of the self-rotation function.

The set of rotations $[T_v][R_k][N_\beta]\{P_{111}\}([T_u][R_j][N_\alpha])^{-1}$ clearly satisfies this condition, since

$$\begin{aligned}[T_v][R_K][N_\beta]\{P_{111}\}([T_u][R_j][N_\alpha])^{-1}M_{j\alpha u} &= [T_v][R_k][N_\beta]\{P_{111}\}M_{111} \\ &= [T_v][R_k][N_\beta]M_{111} \\ &= M_{k\beta v}.\end{aligned}$$

Now we show that all rotations that satisfy this condition are members of the set $\{[T_v][R_k][N_\beta]\{P_{111}\}\ ([T_u][R_j][N_\alpha])^{-1}\}$.

Let $[S(j\alpha u, k\beta v)]$ denote the rotation which will transform the orientation $M_{j\alpha u}$ into the orientation $M_{k\beta v}$.

$$[S(j\alpha u, k\beta v)]M_{j\alpha u} = M_{k\beta v}$$
$$[S(j\alpha u, k\beta v)][T_u][R_j][N_\alpha]M_{111} = [T_v][R_k][N_\beta]M_{111}$$
$$([T_v][R_k][N_\beta])^{-1}[S(j\alpha u, k\beta v)][T_u][R_j][N_\alpha]M_{111} = M_{111}.$$

$([T_v][R_k][N_\beta])^{-1}[S(j\alpha u, k\beta v)][T_u][R_j][N_\alpha]$ is an element of the point group $\{P_{111}\}$, since it leaves M_{111} invariant. Therefore, for some element $[P_{111}]$ of $\{P_{111}\}$,

$$([T_v][R_k][N_\beta])^{-1}[S(j\alpha u, k\beta v)][T_u][R_j][N_\alpha] = [P_{111}]$$
$$[S(j\alpha u, k\beta v)] = [T_v][R_k][N_\beta][P_{111}]([T_u][R_j][N_\alpha])^{-1}.$$

Consequently, all the peaks of the self-rotation function are members of the set

$$\{S(j\alpha u, k\beta v)\} = [T_v][R_k][N_\beta]\{P_{111}\}([T_u][R_j][N_\alpha])^{-1}$$

for ($j, k = 1, ..., n$), ($\alpha, \beta = 1, ..., q$), ($u, v = 1, ..., m$) and every element of the point group $\{P_{111}\}$.

Again, considering that the rotational operators of the space group and the operations which relate the twin domains need to be applied in the orthonormal frame, we can write this expression as

$$\{S(j\alpha u, k\beta v)\} = [\mathrm{OM}][T_v][R_k][\mathrm{OM}]^{-1}[N_\beta]\{P_{111}\} \times ([\mathrm{OM}][T_u][R_j][\mathrm{OM}]^{-1}[N_\alpha])^{-1}. \quad (5)$$

A3. Application to R121D

We now apply the expression derived above to interpret the self-rotation function calculated from the R121D diffraction data. There are two molecules in the asymmetric unit of the crystal. The molecules are in an essentially identical conformation (despite the potential for rigid-body domain motion within the molecule) and thus we can treat them as equivalent.

The orthonormal frame in which the rotations are described has the Cartesian X axis coincident with **a**, and the Cartesian Z axis coincident with (**a** $\times$ **b**) (or **c***, in the convention employed by the Protein Data Bank). The matrices that

convert between the crystallographic and orthonormal frame (see Giacovazzo, 1992) are

$$\mathrm{OM} = \begin{pmatrix} 151.300000 & -75.650000 & 0.000000 \\ 0.000000 & 131.029643 & 0.000000 \\ 0.000000 & 0.000000 & 48.600000 \end{pmatrix},$$

$$\mathrm{OM}^{-1} = \begin{pmatrix} 0.006609 & 0.003816 & 0.000000 \\ 0.000000 & 0.007632 & 0.000000 \\ 0.000000 & 0.000000 & 0.020576 \end{pmatrix}.$$

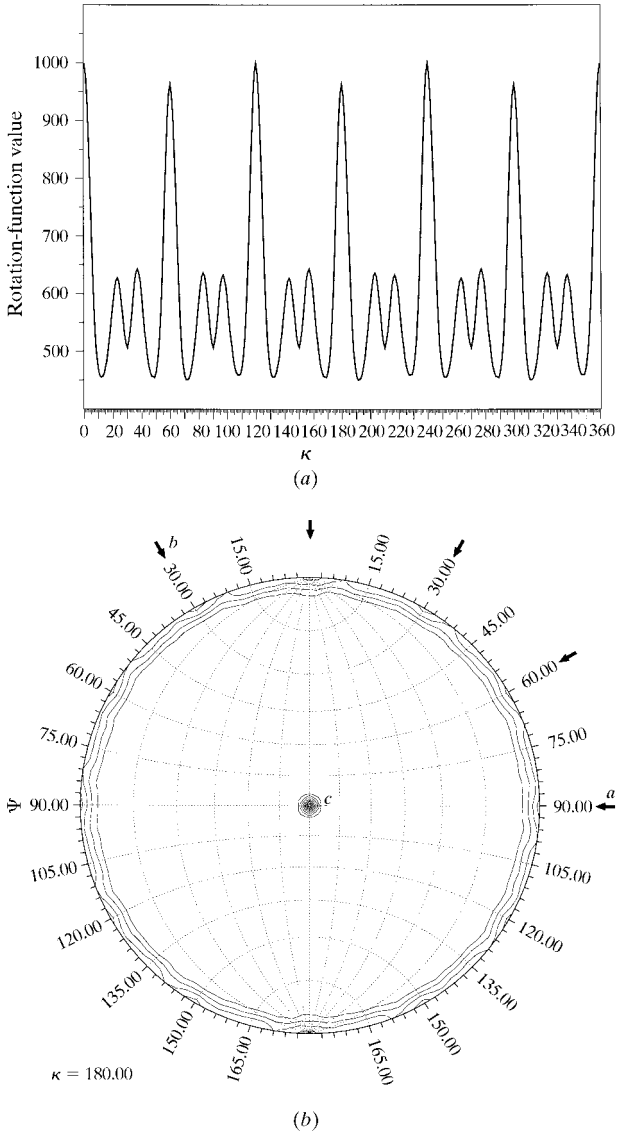

Figure 5
(*a*) A one-dimensional plot of the R121D self-rotation function for all rotations about the direction of the unique crystal axis ($\varphi,\psi = 90^\circ$). (*b*) A stereographic projection of the R121D self-rotation function for all rotations with $\kappa = 180^\circ$. Positions of the apparent twofold rotations perpendicular to the principal axis are indicated by the solid arrows. In both cases, the self-rotation function was calculated by the reciprocal-space method of Rossmann & Blow (1962) using the program *GLRF* (Tong & Rossmann, 1990). A spherical radius of integration of 25 Å was employed, and diffraction data between 10 and 3 Å resolution were used in the calculation.

The molecule does not possess any internal symmetry. Thus, the point group of the motif has only a single element, the identity operation:

$$\{P_{111}\} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The three rotational operators of the space group $P3_1$ (in the crystallographic frame) are

$$R_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$R_2 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$R_3 = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

There are two twin domains in the crystal, and the matrices which define the rotational relationship between them (in the crystallographic frame) are

$$T_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$T_2 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Finally, the rotations which map the distinct copies of the motif in the asymmetric unit into one another are

$$N_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$N_2 = \begin{pmatrix} 0.80272 & -0.59617 & 0.01510 \\ 0.59627 & 0.80191 & -0.03727 \\ 0.01011 & 0.03892 & 0.99919 \end{pmatrix}.$$

The peaks of the self-rotation function predicted using (4) are presented in Table 4. Over all rotation space, a total of 18 peaks are predicted, and all of these correspond to rotations approximately about the direction *Z*. A one-dimensional plot of the R121D self-rotation function for rotations about this axis is presented in Fig. 5(*a*). The peaks of the self-rotation function predicted on the basis of the model are clearly seen in this plot.

In order to fully validate the model, it remains to show that there are no other significant peaks in the self-rotation function. In fact this is not the case, and there are several strong peaks corresponding to apparent twofold rotations perpendicular to the principal axis of the crystal (Fig. 5*b*). In terms of absolute magnitude, these peaks are larger than those

Table 4
Predicted peaks of the self-rotation function for R121D and their assignment.

These should be compared with the actual self-rotation function shown in Fig. 5(*a*). The relative heights of the rotation-function peaks are explained by this assignment. For brevity, only the assignment of the crystallographically unique region of rotation space ($0^\circ \leq \kappa \leq 120^\circ$) is shown. Rotations are described in spherical polar angles, where φ is the angle from the Cartesian X axis and ψ is the angle from the Cartesian Y axis (with $0^\circ \leq \varphi,\psi \leq 180^\circ$ and $0^\circ \leq \kappa \leq 360^\circ$). Their assignment is given in terms of the transformation of the orientation $M_{j\alpha u}$ into the orientation $M_{k\beta v}$. The first subscript indexes the orientations related by space-group symmetry, the second the distinct orientations found in the asymmetric unit of the crystal and the third the orientations due to the merohedral twinning of the crystal (see text for details).

Rotation (φ, ψ, κ) ($^\circ$)	Assignment
0, 0, 0	$M_{111} \rightarrow M_{111}$, $M_{211} \rightarrow M_{211}$, $M_{311} \rightarrow M_{311}$, $M_{112} \rightarrow M_{112}$, $M_{212} \rightarrow M_{212}$, $M_{312} \rightarrow M_{312}$, $M_{121} \rightarrow M_{121}$, $M_{221} \rightarrow M_{221}$, $M_{321} \rightarrow M_{321}$, $M_{122} \rightarrow M_{122}$, $M_{222} \rightarrow M_{222}$, $M_{322} \rightarrow M_{322}$
85, 93, 23	$M_{111} \rightarrow M_{222}$, $M_{211} \rightarrow M_{322}$, $M_{311} \rightarrow M_{122}$, $M_{112} \rightarrow M_{221}$, $M_{212} \rightarrow M_{321}$, $M_{312} \rightarrow M_{121}$
94, 90, 37	$M_{121} \rightarrow M_{111}$, $M_{221} \rightarrow M_{211}$, $M_{321} \rightarrow M_{311}$, $M_{122} \rightarrow M_{112}$, $M_{222} \rightarrow M_{212}$, $M_{322} \rightarrow M_{312}$
90, 90, 60	$M_{111} \rightarrow M_{212}$, $M_{211} \rightarrow M_{312}$, $M_{311} \rightarrow M_{112}$, $M_{112} \rightarrow M_{211}$, $M_{212} \rightarrow M_{311}$, $M_{312} \rightarrow M_{111}$, $M_{121} \rightarrow M_{222}$, $M_{221} \rightarrow M_{322}$, $M_{321} \rightarrow M_{122}$, $M_{122} \rightarrow M_{221}$, $M_{222} \rightarrow M_{321}$, $M_{322} \rightarrow M_{121}$
89, 91, 83	$M_{111} \rightarrow M_{321}$, $M_{211} \rightarrow M_{121}$, $M_{311} \rightarrow M_{221}$, $M_{112} \rightarrow M_{322}$, $M_{212} \rightarrow M_{122}$, $M_{312} \rightarrow M_{222}$
91, 89, 97	$M_{121} \rightarrow M_{212}$, $M_{221} \rightarrow M_{312}$, $M_{321} \rightarrow M_{112}$, $M_{122} \rightarrow M_{211}$, $M_{222} \rightarrow M_{311}$, $M_{322} \rightarrow M_{111}$
90, 90, 120	$M_{111} \rightarrow M_{311}$, $M_{211} \rightarrow M_{111}$, $M_{311} \rightarrow M_{211}$, $M_{112} \rightarrow M_{312}$, $M_{212} \rightarrow M_{112}$, $M_{312} \rightarrow M_{212}$, $M_{121} \rightarrow M_{321}$, $M_{221} \rightarrow M_{121}$, $M_{321} \rightarrow M_{221}$, $M_{122} \rightarrow M_{322}$, $M_{222} \rightarrow M_{122}$, $M_{322} \rightarrow M_{222}$

predicted on the basis of the model. While these peaks cannot be predicted by the expression developed above, they are reproduced in a self-rotation function computed from the model structure factors (whether or not they are twinned).

Problems in the interpretation of the self-rotation function have arisen from the confusion of the packing symmetry of particles with the symmetry of the particles themselves (see Åkerval *et al.*, 1971; Klug, 1971; Litvin, 1975, 1987) or the inability to resolve closely situated rotation-function peaks (see, for example, Muckelbauer *et al.*, 1995). In some cases, the peaks in the self-rotation function are poorly defined, despite the presence of non-crystallographic symmetry (Jones *et al.*, 1991). The unassigned peaks in the R121D self-rotation function present a problem of a different kind.

At this stage, we cannot identify the origin of these peaks. They are clearly not due to superposition of sets of self-vectors in the conventional sense, since they do not correspond to any simple rotation between the molecular orientations in the crystal. The most plausible hypothesis at this time is that they arise from a combination of the inversion symmetry of the Patterson function (which reduces to twofold rotational symmetry on planes passing through the origin) and the special packing arrangement found in the crystals. In support of this idea, the self-rotation functions of other molecules found in similar space groups with a high solvent content also display such peaks. A particularly good example is sperm-whale myoglobin crystallized in space group *P*6 (Phillips *et al.*, 1990). Myoglobin has no internal symmetry, and in this crystal form there is only one molecule in the asymmetric unit. However, in the self-rotation function there are a number of apparent twofold rotations perpendicular to the principal axis (results not shown). The solvent content of this crystal is 62% and there are large solvent channels in the crystal in the direction of the principal axis just as for R121D.

Support is gratefully acknowledged from the USA National Institutes of Health (grant HD-20859), the Howard Hughes Medical Institute (International Research Scholar award to ENB), Massey University (doctoral scholarship to RLK) and the Fulbright Foundation (graduate award to WAB). We also wish to thank Dr Catherine Day for constructing the mutant studied in this paper and Dr Geoff Jameson for valuable discussions and the communication of results in advance of publication.

References

Åkerval, K., Strandberg, B., Rossmann, M. G., Bengtsson, U., Fridborg, K., Johannisen, H., Kannan, K. K., Lövgren, S., Petef, G., Öberg, B., Eaker, D., Hjertén, S., Rydén, L. & Moring, I. (1971). *Cold Spring Harbour Symp. Quant. Biol.* **36**, 469–483.
Anderson, B. F., Baker, H. M., Norris, G. E., Rice, D. W. & Baker, E. N. (1989). *J. Mol. Biol.* **209** 711–734.
Anderson, B. F., Baker, H. M., Norris, G. E., Rumball, S. V. & Baker, E. N. (1990). *Nature (London)*, **344**, 784–787.
Baker, E. N. (1994). *Adv. Inorg. Chem.* **41**, 389–463.
Britton, D. (1972). *Acta Cryst.* A**28**, 296–297.
Brünger, A. T. (1990). *Acta Cryst.* A**46**, 46–57.
Brünger, A. T. (1993). *X-PLOR* Version 3.1. A System for X-ray Crystallography and NMR. Yale University, Conneticut, USA.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Day, C. L., Anderson, B. F., Tweedie, J. W. & Baker, E. N. (1993). *J. Mol. Biol.* **232**, 1084–1100.
Day, C. L., Stowell, K. M., Baker, E. N. & Tweedie, J. W. (1992). *J. Biol. Chem.* **267**, 13857–13862.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Faber, H. R., Baker, C. J., Day, C. L., Tweedie, J. W. & Baker, E. N. (1996). *Biochemistry*, **35**, 14473–14479.
Faber, H. R. & Matthews, B. W. (1990). *Nature (London)*, **348**, 263–266.
Fujinaga, M. & Read, R. J. (1987). *J. Appl. Cryst.* **20**, 517–521.
Gerstein, M., Anderson, B. F., Norris, G. E., Baker, E. N., Lesk, A. M. & Chothia, C. (1993). *J. Mol. Biol.* **234**, 357–372.

Giacovazzo, C. (1992). Editor. *Fundamentals of Crystallography*, pp. 61–140. Oxford University Press.
Gomis-Rüth, F. X., Fita, I., Kiefersauer, R., Huber, R., Avilés, F. X. & Navaza, J. (1995). *Acta Cryst.* D**51**, 819–823.
Hirshfeld, F. L. (1968). *Acta Cryst.* A**24**, 301–311.
Igarashi, N., Moriyama, H., Mikami, T. & Tanaka, N. (1997). *J. Appl. Cryst.* **30**, 362–367.
Jameson, G. B. (1982). *Acta Cryst.* A**38**, 817–820.
Jones, E. Y., Walker, N. P. C. & Stuart, D. I. (1991). *Acta Cryst.* A**47**, 753–770.
Klug, A. (1971). *Cold Spring Harbour Symp. Quant. Biol.* **36**, 483–488.
Koch, E. (1992). *International Tables for Crystallography*, Vol. C, edited by A. J. C. Wilson, pp. 10–14. Dordrecht: Kluwer.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lea, S. & Stuart, D. (1995). *Acta Cryst.* D**51**, 160–167.
Litvin, D. B. (1975). *Acta Cryst.* A**31**, 407–416.
Litvin, D. B. (1987). *Patterson and Pattersons*, edited by J. P. Glusker, B. K. Patterson & M. Rossi, pp. 515–523. New York: Oxford University Press.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Muckelbauer, J. K., Kramer, M., Minor, I., Tong, L., Zlotnick, A., Johnson, J. E. & Rossmann, M. G. (1995). *Acta Cryst.* D**51**, 871–887.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Phillips, G. N. Jr, Arduini, R. M., Springer, B. A. & Sligar, S. G. (1990). *Proteins*, **7**, 358–365.
Pratt, C. S., Coyle, B. A. & Ibers, J. A. (1971). *J. Chem. Soc. A*, pp. 2146–2151.
Rao, S. N., Jih, J.-H. & Hartsuck, J. A. (1980). *Acta Cryst.* A**36**, 878–884.
Redinbo, M. R. & Yeates, T. O. (1993). *Acta Cryst.* D**49**, 375–380.
Rees, D. C. (1980). *Acta Cryst.* A**36**, 578–581.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Stanley, E. (1972). *J. Appl. Cryst.* **5**, 191–194.
Tong, L. & Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 783–792.
Yeates, T. O. (1988). *Acta Cryst.* A**44**, 142–144.
Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.
Zhang, X.-J. & Matthews, B. W. (1994). *Acta Cryst.* D**50**, 675–686.