

# Expanding the model: anisotropic displacement parameters in protein structure refinement

**Ethan A. Merritt**

Department of Biological Structure, University of Washington, Seattle, WA 98195-7742, USA

Correspondence e-mail:  
merritt@u.washington.edu

Ethan Merritt first encountered crystallography as a graduate student in computer science, and was delighted to have found a research field offering a wide scope for developing and applying computational techniques in pursuit of biological goals. He earned a PhD in the group of M. Sundaralingam in Madison, dividing his efforts between macromolecular graphics, conformational modeling of nucleotides and the use of anomalous scattering to determine the stereospecificity of metal/nucleotide coordination in ATP-binding enzymes. He then joined the research staff of the Stanford Synchrotron Radiation Laboratory, where he worked to implement the X-ray optics, beamline-control systems and data-processing software needed to turn MAD phasing from a promising idea into a routine technique for structure determination. This effort culminated in a series of collaborations, notably with Hans Freeman and Wayne Hendrickson, which yielded the first protein structure determinations entirely from MAD phasing. In 1989, he moved to the University of Washington Medical School in Seattle. There his research interests have included structure-based drug design, development of the *Raster3D* graphics package and the improved exploitation of synchrotron radiation in protein crystallography. The review presented here grew from a convergence of these interests.

Recent technological improvements in crystallographic data collection have led to a surge in the number of protein structures being determined at atomic or near-atomic resolution. At this resolution, structural models can be expanded to include anisotropic displacement parameters (ADPs) for individual atoms. New protocols and new tools are needed to refine, analyze and validate such models optimally. One such tool, *PARVATI*, has been used to examine all protein structures (peptide chains >50 residues) for which expanded models including ADPs are available from the Protein Data Bank. The distribution of anisotropy within each of these refined models is broadly similar across the entire set of structures, with a mean anisotropy  $A$  in the range 0.4–0.5. This is a significant departure from a purely isotropic model and explains why the inclusion of ADPs yields a substantial improvement in the crystallographic residuals  $R$  and  $R_{\text{free}}$ . The observed distribution of anisotropy may prove useful in the validation of very high resolution structures. A more complete understanding of this distribution may also allow the development of improved protein structural models, even at lower resolution.

Received 25 November 1998

Accepted 30 March 1999

## 1. Introduction

Biologists and biochemists seeking to understand protein structure and function are always hungry for the highest resolution views possible of their favorite subject molecules. This appetite ensures an interested audience for each new protein crystal structure or NMR determination. The term 'high resolution', as applied to crystal structures of proteins, has for many years meant that X-ray diffraction data were measured to roughly 2 Å or better. This was a tacit acknowledgment that only in rare cases was it possible or practical to measure complete crystallographic data sets beyond 2 Å resolution from protein crystals. However, recent technological advances and the increased use of synchrotron radiation X-ray sources have largely removed this practical limit on data collection. Of course, not all protein crystals diffract beyond (or even as far as!) 2 Å resolution, but complete data collection is now possible from the surprisingly large number which do. One result is a dramatic increase in the number of protein crystal structures refined at atomic resolution (better than 1.2 Å) or at near-atomic resolution. A qualitative difference of crystallographic analysis at this resolution is that the large number of experimental observations, the measured Bragg intensities, allows expansion of the crystallographic model to include an anisotropic description of the position of each atom in the structure. This is in distinct contrast to refinement at lower resolution, where the smaller number of

observations does not allow this more than doubling of the number of parameters in the model. As of December 1997, only ten protein structures in the Protein Data Bank included anisotropic displacement parameters (ADPs) with the deposited model coordinates. This number doubled in the first six months of 1998, and it is abundantly clear from preliminary reports at meetings, from publications in advance of structure deposition and from data-collection logs at the various synchrotron radiation laboratories that many other such structures are in the pipeline.

Fortunately, increases in computational power have kept pace with the advances in data collection, so that analysis and interpretation of this flood of structures is computationally tractable. Nevertheless, new tasks can require new tools. The refinement protocols which have served protein crystallographers well in the past, implemented in programs such as *PROLSQ*, *TNT* and *X-PLOR* (Konnert & Hendrickson, 1980; Tronrud *et al.*, 1987; Brünger, 1992a), need to be updated to handle refinement at atomic resolution optimally. Virtually all of the atomic resolution protein refinements to date have depended on the timely modification of the program package *SHELX*, a mainstay of small-molecule crystallographic refinement, to also handle protein structures (Sheldrick & Schneider, 1997). The current burgeoning of very high resolution data collection from protein crystals is likely to inspire a next generation of refinement programs specifically designed to handle increasingly high resolution data and correspondingly complex models of protein structure.

In a similar fashion, the validation tools developed to monitor the quality of protein structural models must be expanded to handle the more complex atomic resolution models, with their increased number and variety of parameters. The total number of protein structures modeled with anisotropic thermal parameters is still small, but is growing rapidly. Several distinct benefits may be expected from systematic analysis and comparison of these structures as they become available. At a minimum, such analysis can provide a baseline for validation and guidance during the refinement of subsequent structures. We may also hope that analysis will lead to improvements in the specific protocols used for atomic resolution refinement of protein models and indeed to optimization of the basic parameterization of the models. In the long run, atomic resolution structures will provide a higher powered lens with which to re-inspect certain basic features of protein structure. One example of this is a re-evaluation of the rigidity of the peptide bond, which is observed in some instances to vary substantially from the ideally planar conformation normally assumed as a basic unit of protein structure (Stec *et al.*, 1995; MacArthur & Thornton, 1996; Merritt *et al.*, 1998). Another example is the possible insight into the local vibrational modes implied by corresponding atomic anisotropic displacement parameters (Stec *et al.*, 1995; Dunitz *et al.*, 1989; Anderson *et al.*, 1997; Harata *et al.*, 1998).

The database of protein structures determined by crystallographic analysis has become an important resource for researchers in many areas. For this reason, there is both an increased interest in optimization of the protocols used for

structure refinement and an increased emphasis on validation of the correctness of the resulting structural models (Dauter, Lamzin *et al.*, 1997). Several reviews of structural validation in general have appeared recently (Hooft *et al.*, 1996; EU 3-D Validation Network, 1998). I focus here on a set of issues specific to the optimal refinement and validation of protein models which include anisotropic displacement parameters. I also present a set of tools designed to assist in the analysis of protein models containing anisotropic displacement parameters. These tools were applied to all proteins of at least 50 residues for which ANISOU records were available from the Protein Data Bank at the time of writing.

## 2. Isotropic and anisotropic displacement parameters

The X-ray scattering power of an atom decreases as the scattering angle increases, owing to the finite size of the electron cloud around the nucleus. For a given number of electrons, the larger this cloud is, the more rapidly the atom's scattering power falls off with scattering angle. To account for this scattering behaviour of real atoms in real crystals, one may add an angle-dependent term to the individual atomic scattering contributions  $f$  which are summed to yield the scattered intensities  $F_{\text{calc}}^2$ . This correction to  $f$  for a given atom has the form

$$f = f_0 \exp(-2\pi^2 \langle \mathbf{u}^2 \rangle \mathbf{h}^T \mathbf{h}) = f_0 \exp[-8\pi^2 \langle \mathbf{u}^2 \rangle (\sin^2 \theta / \lambda^2)], \quad (1)$$

where  $\langle \mathbf{u}^2 \rangle$  is the mean-square amplitude of vibration of that atom,  $\mathbf{h}$  is a reciprocal-lattice vector,  $\theta$  is the corresponding scattering angle and  $\lambda$  is the X-ray wavelength. The electron cloud of a vibrating atom, averaged over time, is larger than that of a similar atom at rest. Because the magnitude of vibration correlates with temperature, the parameter  $\mathbf{u}$  is often called a thermal parameter or temperature factor. This is misleading, however, since the smearing of the electron cloud at an atomic site in the crystal is a consequence not only of thermal vibration, but also of stochastic variation in the true location of the atomic center from one unit cell to the next. The preferred term is therefore 'displacement parameter' (Trueblood *et al.*, 1996). In macromolecular crystallography, it is more common to report the related parameter  $B = 8\pi^2 \langle \mathbf{u}^2 \rangle$ .

(1) describes an electron cloud which is uniformly smeared in all directions, and the parameter  $\mathbf{u}$  is thus an isotropic displacement parameter associated with the atom in question. The vibrational modes of bonded atoms are not isotropic, however, so (1) is at best an approximation to the actual scattering behavior of real protein atoms in a crystal lattice. To relax the implied insensitivity of scattering to direction while retaining computational simplicity, we can expand the single parameter  $\mathbf{u}$  into a  $3 \times 3$  symmetric tensor

$$\begin{pmatrix} U^{11} & U^{12} & U^{13} \\ U^{12} & U^{22} & U^{23} \\ U^{13} & U^{23} & U^{33} \end{pmatrix}.$$

The corresponding correction term, now anisotropic, becomes

$$f = f_0 \exp(-2\pi^2 \mathbf{h}^T \mathbf{U} \mathbf{h}). \quad (2)$$

The six independent components of the tensor,  $U^{ij}$ , are the anisotropic displacement parameters for this atom. They describe a probability distribution for the electron density which is a three-dimensional Gaussian. When this distribution is contoured at a fixed probability value, it yields an ellipsoid (Fig. 1). This representation was made familiar by the program *ORTEP* (Johnson, 1965).

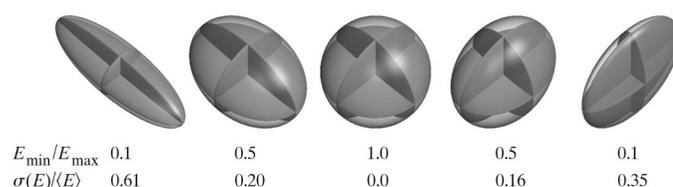
### 3. Open questions

The complexity of a refinable model is limited by the amount of available data or, more precisely, by the ratio of the number of observations (the data) to the number of model parameters (the complexity). Any increase in the ratio of observations to parameters can improve the quality of the refined structural model, yielding better estimates for the individual model parameters. In crystallography, the maximum number of observations is usually set by crystal quality and by the corresponding limited resolution of X-ray diffraction from the crystal. For protein crystals, the measurable number of observations is usually insufficient to justify free refinement of a full atomic model for the structure. To overcome this problem, the structural model can often be simplified, reducing the total number of model parameters. This is accomplished by introducing outside knowledge, for example, the existence of multiple stereochemically equivalent molecules in the asymmetric unit, to supplement the information contained in the measured data. The benefits of an increased observation-to-parameter ratio from a simplified structural model underlie a number of important crystallographic techniques, including rigid-body refinement, non-crystallographic symmetry constraints (Kleywegt, 1996), density averaging (Kleywegt & Read, 1997) and torsion-space refinement (Rice & Brünger, 1994). Alternatively, the effective number of model parameters can be reduced by imposing restraints on their allowed values. This approach was introduced by Konnert (1976) and remains fundamental to the success of most protein structure refinements. Of course, the resulting accuracy of a simplified or restrained model depends critically on the validity of the simplification or of the restraints that are imposed.

Conversely, in fortunate cases, the observation-to-parameter ratio may instead be improved by a larger number of observations. Doubling the achievable resolution of a diffraction experiment, say from 2 to 1 Å, corresponds to an eightfold increase in the number of Bragg intensities which can be measured. The increased number of observations can be used to support qualitative improvement of the structural model, by expanding it to include additional parameters which describe more subtle or more complex features of the structure. In particular, it allows expansion of the description of each atom to include a set of six anisotropic displacement parameters (ADPs), rather than a single isotropic thermal parameter as described above. This increases the total number of model parameters from four per atom (three positional parameters plus a single thermal parameter) to nine per atom (three positional parameters plus six thermal parameters).

Inclusion of ADPs yields global improvement of the model, as seen in lower  $R$  and  $R_{\text{free}}$  residuals, cleaner residual-density maps and lower uncertainties in atomic positions. Inclusion of atomic anisotropy in the model also makes possible the quantitative analysis of local features of the structure, such as torsional vibrations of individual protein side chains. Of course, expansion of the model to include ADPs once again lowers the observation-to-parameter ratio. Doubling the resolution from 2 to 1 Å and expanding the model to include ADPs is clearly a net gain, as the observation-to-parameter ratio has increased by more than a factor of three, even with the expanded model. However, there is a 'grey zone' covering the resolution range from  $\sim 1.6$  to  $\sim 1.2$  Å, in which there is some question whether the increase in the number of observations is sufficient to justify the increase in the number of model parameters.

The relatively modest gain in observation-to-parameter ratio in this grey zone may be strengthened by applying restraints to the ADPs during refinement. The more observations there are to determine the true values of the ADPs, the less need there is to hold them at or near idealized target values. Thus, at one end of the grey zone, near 1.6 Å, the observation-to-parameter ratio for an anisotropic model with nine parameters per atom is the same as it would have been at 2 Å for an isotropic model with only four parameters per atom. However, the isotropic model is well behaved during standard crystallographic refinement at 2 Å only because strong geometric restraints are imposed on the allowed values of the three positional parameters ( $xyz$ ). To obtain equivalently good behaviour during an anisotropic model refinement at 1.6 Å would require an equivalently powerful set of restraints on the six thermal parameters  $U^{ij}$ . At the other end of the grey zone, as one approaches true atomic



**Figure 1**

The five atoms shown above illustrate the physical significance of anisotropic displacement parameters (ADPs). They represent a range of shapes, from the prolate 'cigar' at the left to the oblate 'pancake' at the right, with a perfectly isotropic (spherical) example in the center. The corresponding ADPs have been chosen such that all five examples represent atoms with the same isotropic thermal parameter  $B_{\text{eq}}$ . Only the central atom, however, is adequately described by a single isotropic parameter. One can quantify the degree to which the other ellipsoids are non-spherical by defining anisotropy,  $A = E_{\min}/E_{\max}$ , as the ratio of the smallest and largest eigenvalues of the ADP matrix (Trueblood *et al.*, 1996). Thus, for a perfect sphere  $A = 1$ , while for extremely non-spherical ellipsoids  $A$  approaches 0. Note that small values of  $A$  may indicate either a cigar-like or pancake-like atom. Some authors have proposed using  $\sigma(E)/\bar{E}$ , *i.e.* the square root of the variance over the mean of the three principal axes, instead of  $A$  (Harata *et al.*, 1998; Longhi *et al.*, 1997). This quantity has the disadvantage of being more sensitive to cigar-shaped outliers than to pancake-shaped outliers, however. The two atoms flanking the central sphere, both having anisotropy  $A = 0.5$ , are typical of the atoms in protein models refined at atomic resolution.

**Table 1**

Accessible PDB entries as of December 1998 for protein structures with deposited anisotropic displacement parameters.

The analysis reported here necessarily omits consideration of structures for which no anisotropic displacement parameters were deposited with the PDB, even if the model was in fact refined anisotropically. Peptides shorter than 50 residues are not included.

PDB code	Protein	Residues in a.s.u.	Resolution (Å)	$R_1$	$\overline{A}(\sigma)$ protein	$\overline{A}(\sigma)$ solvent	Date†	Reference
1gci	<i>Bacillus lentus</i> subtilisin	269	0.78	0.101	0.65 (0.15)	0.42 (0.17)	1998	Khan <i>et al.</i> (1998)
2pvb	Pike parvalbumin	107	0.91	0.110	0.49 (0.16)	0.35 (0.13)	1998	DeClercq <i>et al.</i> , unpublished work
3lzt	Lysozyme	129	0.93	0.093	0.45 (0.15)	0.32 (0.14)	1998	Walsh <i>et al.</i> (1998)
1b0y	<i>Chromatium vinosum</i> HIPIP	86	0.93	0.155	0.37 (0.14)	0.37 (0.14)	1998	Parisinicapozzi <i>et al.</i> , unpublished work
1nls	Concanavalin A	237	0.94	0.127	0.43 (0.17)	0.39 (0.13)	1997	Deacon <i>et al.</i> (1997)
2fdn	<i>Clostridium acidurici</i> ferredoxin	55	0.94	0.100	0.49 (0.14)	0.36 (0.15)	1998	Dauter, Wilson <i>et al.</i> (1997)
1bxo	<i>Penicillium janthinellum</i> penicillopepsin	323	0.95	0.100	0.61 (0.16)	0.43 (0.16)	1998	Khan <i>et al.</i> (1998)
4lzt	Lysozyme	129	0.95	0.108	0.44 (0.16)	0.35 (0.13)	1998	Walsh <i>et al.</i> (1998)
1ixh	Phosphate-binding protein	321	0.98	0.117	0.48 (0.13)	Isotropic	1998	Wang <i>et al.</i> (1997)
1cex	<i>Fusarium solani</i> cutinase	214	1.00	0.094	0.49 (0.16)	0.38 (0.13)	1997	Longhi <i>et al.</i> (1997)
1lkk	Human P56 <sup>lck</sup> kinase SH2 domain	110	1.00	0.133	0.51 (0.13)	0.38 (0.15)	1996	Tong <i>et al.</i> (1996)
1ixg	Phosphate-binding protein	321	1.05	0.112	0.45 (0.14)	Isotropic	1998	Wang <i>et al.</i> (1997)
3sil	<i>Salmonella typhimurium</i> sialidase	379	1.05	0.116	0.50 (0.15)	0.35 (0.15)	1998	Garman <i>et al.</i> , unpublished work
1bkr	CH domain from $\beta$ -spectrin	109	1.10	0.141	0.45 (0.17)	0.41 (0.15)	1998	Banuelos <i>et al.</i> (1998)
1ctj	<i>Monoraphidium braunii</i> cytochrome $c_6$	89	1.10	0.138	0.38 (0.13)	0.42 (0.15)	1996	Frazzoli <i>et al.</i> (1995)
1iro	<i>Cl. pasteurianum</i> rubredoxin	53	1.10	0.083	0.46 (0.14)	0.40 (0.14)	1996	Dauter <i>et al.</i> (1996)
1lks	Hen egg-white lysozyme	129	1.10	?	0.35 (0.11)	0.32 (0.12)	1998	Steinrauf (1998)
1a6g	<i>Physeter catadon</i> myoglobin	151	1.15	0.128	0.45 (0.14)	0.40 (0.13)	1998	Vojtechovsky <i>et al.</i> , unpublished work
1rge	<i>Streptomyces aureofaciens</i> ribonuclease	192	1.15	0.109	0.43 (0.11)	0.48 (0.17)	1996	Sevcik <i>et al.</i> (1996)
1rfg	<i>S. aureofaciens</i> ribonuclease	192	1.20	0.106	0.48 (0.13)	0.45 (0.14)	1996	Sevcik <i>et al.</i> (1996)
1irn	<i>Cl. pasteurianum</i> rubredoxin	53	1.20	0.107	0.42 (0.14)	0.38 (0.13)	1996	Dauter <i>et al.</i> (1996)
3chb	Cholera toxin B-pentamer	515	1.25	0.130	0.46 (0.13)	0.44 (0.13)	1998	Merritt <i>et al.</i> (1998)
6fd1	<i>Azotobacter vinelandii</i> ferredoxin	106	1.35	0.153	0.47 (0.12)	0.42 (0.13)	1997	Stout <i>et al.</i> (1998)
1awd	<i>Chlorella fusca</i> ferredoxin	94	1.40	0.147	0.50 (0.14)	0.46 (0.13)	1998	Bes <i>et al.</i> , unpublished work
	<i>Escherichia coli</i> LT-I B-pentamer	1030	1.40	0.138	0.44 (0.13)	0.46 (0.12)		Merritt <i>et al.</i> , unpublished work
1a2p	<i>B. amyloliquefaciens</i> barnase	330	1.50	0.115	0.34 (0.13)	0.45 (0.14)	1998	Martin <i>et al.</i> , (1999)
1a8d	<i>Cl. tetani</i> tetanus toxin C fragment	452	1.57	0.188	0.36 (0.11)	0.54 (0.15)	1998	Knapp <i>et al.</i> unpublished work
1lkr	Hen egg-white lysozyme	258	1.60	0.106	0.56 (0.11)	0.42 (0.16)	1998	Steinrauf (1998)

† The date given is that of the most recent revision to the deposited PDB file.

resolution, these restraints may be applied loosely or removed altogether.

There is an additional subtlety in that a restraint may be applied either to an individual parameter value or to some quantity derived jointly from several parameters. In the present case, the most obvious candidate for such a derived quantity is the net anisotropy of an atom (Fig. 1). This single value  $A$ , derived from the six anisotropic displacement parameters, describes the degree to which the model for a scattering atom is non-spherical. A restraint on  $A$  during refinement of the ADPs is exactly analogous to the use of geometric restraints, for example on bond length, to guide refinement of atomic coordinates. However, in order to apply a restraint at all, one must have some *a priori* expectation for the distribution of the corresponding parameter values. That is, one must have some set of target values, or at least a target variance, to which the parameters are restrained. But just what should be our expectations for the degree or distribution of anisotropy in the atoms of a protein structure? We may break this uncertainty down into a series of questions open for investigation.

(i) Is there a distribution of anisotropy among the atoms of a protein which is typical of protein structures in general, or is it an intrinsic property of individual proteins?

(ii) Is the degree of anisotropy within a given structure predictable from some more global (and more easily measurable) property of that structure?

(iii) If we can establish expectations for the mean and variance of the anisotropy within a protein structure, how can these expectations best be used in structure refinement?

(iv) Is it possible to find a simplified representation of local anisotropy, say at the level of whole residue side chains rather than individual atoms, which captures the essential features of a protein structure in far fewer parameters?

If we can answer these questions even partially through analysis of atomic resolution structures, the resulting knowledge may improve our ability to model protein structure, even at lower resolution.

#### 4. Analysis of structures available to date

An initial analysis of anisotropy was made by examining all applicable protein structures available as of December 1998. This included all structures in the Protein Data Bank containing at least 50 peptide residues in the asymmetric unit and for which anisotropic displacement parameters were either deposited or made available by the depositors on request. Also included were two structures currently being refined in the author's laboratory. The complete set of structures analyzed is listed in Table 1.

In order to perform this analysis conveniently, the *rastep* utility program of the *Raster3D* visualization package (Merritt & Bacon, 1997) was modified to tabulate various statistical properties of the ANISOU records from a standard PDB-

format input file. Use of *rastep* as a validation/analysis tool is simplified by providing the user with a higher level graphical interface named *PARVATI* (Protein Anisotropic Refinement Validation and Analysis Tool). This interface is implemented as a server-side computational resource (a cgi script) which is accessible using any standard WWW browser. *PARVATI* partitions the records of an uploaded PDB file into protein, solvent, ligand or other categories for separate analysis. Graphical and tabular representations of the distribution of anisotropy in the various structural components are returned to the originating WWW browser (Fig. 2).

#### 4.1. Is there a typical distribution of anisotropy common to protein structures?

The anisotropy of an individual atom is defined by  $A = E_{\min}/E_{\max}$  (Trueblood *et al.*, 1996), where  $E_{\min}$  and  $E_{\max}$  are the minimum and maximum eigenvectors of the ADP matrix and are also related to the lengths of the shortest and longest principal axes of the corresponding ellipsoid (Fig. 1). The mean anisotropy,  $\bar{A}$ , of the individual protein structures deposited with the PDB ranges from 0.34 to 0.65, as shown in Table 1 and Fig. 3. One must approach this set of values cautiously, as any restraint holding the atoms to be approximately isotropic during refinement will both shift the resultant value of  $\bar{A}$  upward and distort the shape of the distribution. Unfortunately, information on the restraints applied during refinement is not available for many of the structures tabulated.

The majority of the models show substantial similarity in the distribution of anisotropy, with a consensus value of  $\bar{A}$  somewhere in the region of 0.4–0.5 and an approximately symmetric distribution about this mean, with  $\sigma_A$  values of the order of 0.1–0.2. This is a significant departure from a purely isotropic model for atoms in a protein structure. It is, therefore, hardly surprising that the inclusion of ADPs in the model in general leads to a dramatic improvement in the power of

the model to predict scattering amplitudes ( $F_{\text{calc}}$ ) and also to a corresponding dramatic decrease in the standard crystallographic residuals  $R$  and  $R_{\text{free}}$ . Three structures of the present set (1a2p, 1a8d, 1lkr; Table 1) are in the upper portion of the grey zone of resolution, between 1.4 and 1.6 Å, where strong restraints on anisotropy are called for. It is interesting that all three are slight outliers from the consensus distribution of  $A$  (dotted lines in Fig. 3), which may indicate that the restraints used in refinement were not optimal. Two other outliers to the consensus distribution are substantially more isotropic than the remaining structural models in the current set. One of these, the 0.78 Å refinement of *Bacillus lentus* subtilisin (1gci; Table 1), is the highest resolution model in the current set. It also exhibits a very low mean value for protein thermal parameters ( $\langle B_{\text{eq}} \rangle = 7.9 \text{ \AA}^2$ ). It will be very interesting to see whether the low anisotropy exhibited by this model is replicated in future extremely high resolution (<0.8 Å) refinements.

The set of structures in Table 1 includes a reasonable range of protein sizes, protein tertiary structures and crystallographic properties such as solvent content. Therefore, although the total number of structures is not large (and is somewhat biased toward lysozyme!), the general agreement in mean anisotropy found from these refinements may indicate that it is representative of a wide range of proteins.

#### 4.2. Is the degree of anisotropy in a structure predictable from more global properties?

Next we may ask if there is evidence from this set of structures that the expected value of  $\bar{A}$  or  $\sigma_A$  should be modified based on other characteristics of the structure. It would be plausible, for instance, if structures determined at liquid-nitrogen temperature exhibited less anisotropy than those collected at room temperature. For an example of a model which would make this prediction, consider a case where the refined net anisotropy of atomic centers in protein structures at room temperature is a consequence primarily of thermal vibration of rigid groups (domains or subdomains) made up of isotropic atoms. This is conceptually analogous to a swinging string of pearls. At sufficiently low temperature, the large-scale rigid-body motion would be damped and would no longer mask the local isotropy of the individual constituent atoms. A more general hypothesis would be that the mean anisotropy, and perhaps also the variance  $\sigma_A^2$ , should depend on the overall mean  $B$  value of the structure. This hypothesis is tested in Fig. 4(a), which seems to indicate that neither the distinction between cryo- and room-temperature structures nor the generalization to a dependence of  $\bar{A}$  on  $\bar{B}_{\text{eq}}$  is justified. This is no way invalidates models of domain motion in individual proteins, of course; it

### 3CHB Summary

	Anisotropy		# atoms
	Mean	Sigma	
Protein	0.459	0.130	4103
Solvent	0.441	0.127	758
Hetatms	0.469	0.146	342
<b>Total</b>	<b>0.457</b>	<b>0.131</b>	<b>5203</b>
C	0.466	0.127	2756
O	0.453	0.132	923
N	0.444	0.141	729
S	0.502	0.116	37

Retrieve EPS  
version of plot

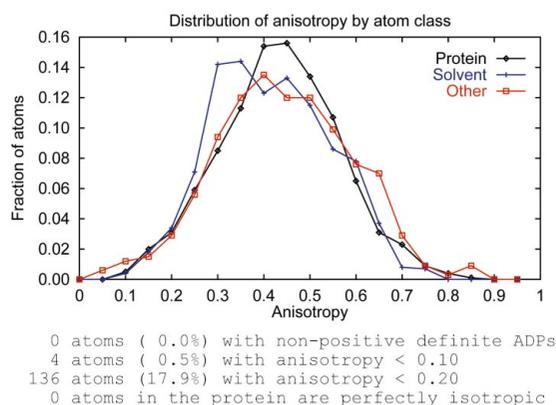


Figure 2

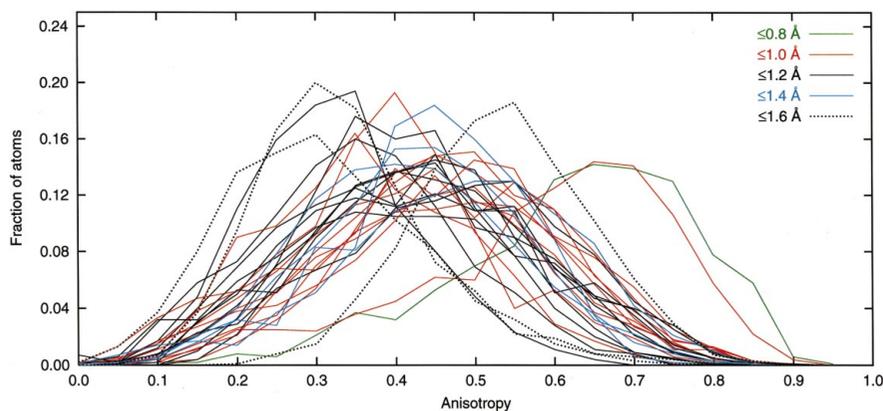
Example of a WWW page returned by the *PARVATI* WWW server, containing an overall summary of the distribution of anisotropy within an uploaded PDB file. The server also returns a listing of atoms which are outliers, as well as suggestions for revised restraint weights and plots of the spatial distribution of anisotropy within the protein (not shown).

simply indicates that temperature is not a predictor of mean anisotropy for the current ensemble of refined structures.

It is still possible that anisotropy and  $B$  values are correlated within a structure, however, and this modified hypothesis is tested in Fig. 4(b). It is clear that the models refined at a resolution of 1 Å or better exhibit a stronger correlation of  $A$  with  $B_{\text{eq}}$ . These highest resolution structures all also have a low overall mean  $\overline{B}_{\text{eq}}$ , but this by itself does not distinguish them from many of the structures refined at somewhat lower resolution. Based on the set of structures currently available, we may tentatively conclude that one should expect a correlation of anisotropy with  $B_{\text{eq}}$  for atoms within a structure. This correlation is less striking at lower resolution, however, perhaps because of the increased importance of restraints applied during refinement or to the fact that the lower observation-to-parameter ratio inherently decreases the accuracy of the refined ADP parameters.

#### 4.3. Choice of restraint weights applied to the ADPs during refinement

The structures analyzed here were all refined using versions of the program *SHELXL* (Sheldrick & Schneider, 1997), so discussion of restraints will focus specifically on the two types of restraints which *SHELXL97* applies to anisotropic displacement parameters during refinement. The first of these is controlled by the ISOR command; it restrains the ADPs of each atom to be approximately isotropic. The strength of this restraint may be set independently for arbitrary sets of atoms. The ISOR restraint was originally intended for application to discrete water molecules in structures refined at high resolution. It may also be used to restrain protein atoms during refinement of a structure in the grey zone of resolution or to restrain specific poorly behaved atoms (*e.g.* atoms whose ADPs become non-positive definite during refinement). The choice whether ISOR restraints are needed at all, or what set of target deviations is optimal, can be made by tracking the



**Figure 3**

Distribution of anisotropy exhibited in structures deposited with the Protein Data Bank. Each curve on the plot represents the distribution of anisotropy seen for protein atoms belonging to a single deposited structure from among those listed in Table 1. The curves are color coded to indicate the resolution of the corresponding structure refinement, as listed in the key. The three structures with the lowest resolution (between 1.4 and 1.6 Å) are shown in dotted lines.

residuals  $R$  and  $R_{\text{free}}$  (Brünger, 1992b) resulting from repeated refinement runs (Fig. 5). This has the major drawback of being extremely expensive computationally, as each run may require anywhere from several hours to several days of CPU time, depending on the size of the structure and the number of observations. In order to make the optimization of restraints more tractable, it would therefore be useful to find a set of heuristic guidelines. Figs. 3 and 4(a) suggest that well behaved protein refinements yield a roughly symmetric distribution of anisotropy within the protein, with a mean in the range 0.4–0.5 and a  $\sigma$  of  $\sim 0.15$ . Fig. 5 further suggests that the asymmetry introduced by a non-optimal choice of restraint weights is fairly obvious, as is the attribution of the asymmetry to either too strong or too weak restraints. Note, however, that the asymmetry induced by applying a strong ISOR restraint is to some extent an artifact of the mode of implementation in *SHELXL*. Because the restraint target is specifically  $A = 1$ , perfect isotropy, the distribution tends to crowd up against this side of the mean with a long skewed tail on the side toward  $A = 0$ . Yet the analysis in Fig. 3 does not support the notion that our default assumption should be perfect isotropy. It would be interesting to investigate whether a different implementation, which allowed specification of a target anisotropy other than  $A = 1$ , would be more optimal in the specific case of including strongly restrained ADPs for the protein atoms in a structure at marginal resolution.

A second category of restraints applied to the anisotropic displacement parameters is controlled by the SIMU command in *SHELXL*. This restraint limits the extent to which the corresponding ADPs of nearby atoms can differ from each other. The primary effect of this restraint is that the axes of deformation (that is, the direction rather than the magnitude of anisotropy) of adjoining atoms are roughly parallel. Because the SIMU restraint is applied to the absolute, rather than relative, values of the individual ADPs, it has a secondary effect of limiting the variation in  $B_{\text{eq}}$  between adjacent atoms. In this respect, its function is analogous to restraints on r.m.s.  $\Delta B$  applied by *X-PLOR* and other programs commonly used

to refine isotropic protein models. It is worth noting that the *SHELXL* implementation distinguishes ‘terminal’ atoms, those with only one bonded neighbor, from all other atoms. If the SIMU target applied to terminal atoms is too stringent, this will manifest itself in  $(F_{\text{obs}} - F_{\text{calc}})$  difference-density maps as a preponderance of negative density (indicating that  $B_{\text{eq}}$  is too low) at the atomic centers corresponding to terminal atoms. The overall stringency of SIMU restraints may also be judged by evaluating the reasonableness of the overall r.m.s.  $\Delta B_{\text{eq}}$  for bonded atoms in the protein structure, just as one would for a structure refined isotropically. Changes to the strength of SIMU restraints can be validated by monitoring the effect on  $R_{\text{free}}$ .

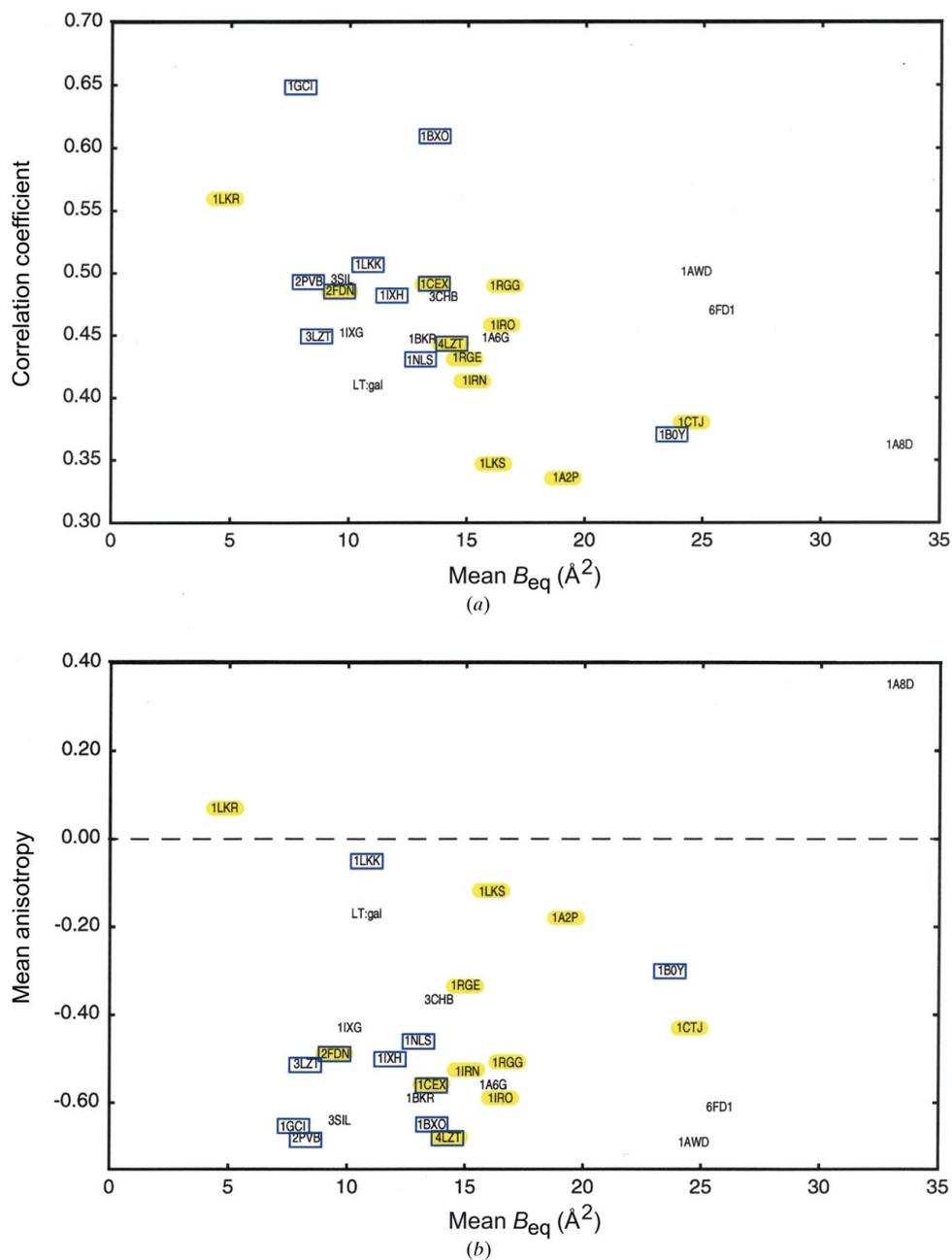
The underlying rationale for the SIMU restraint is an expectation that atoms belonging to a relatively rigid group will undergo thermal vibration in unison. Analysis of the validity of this rationale is beyond the scope of the current

discussion, although it should be noted that torsional vibrations, such as that about the  $C^\alpha - C^\beta$  bond of rigid side chains such as threonine and valine, clearly violate this expectation of parallel vibration. Re-formulation of this restraint for application to specific side-chain types during refinement is another interesting area open for exploration.

#### 4.4. Anisotropy of water molecules associated with the protein structure

Most of the questions regarding expectations for anisotropy and the choice of optimal restraints may be asked all over again with respect to the treatment of discrete water molecules in atomic resolution structural models. A fraction of the protein crystals which diffract to atomic resolution contain only a very small region of disordered solvent. Two examples of this are crambin (Stec *et al.*, 1995; Teeter, 1992) and  $\gamma$ B-crystallin (Kumaraswamy *et al.*, 1996). In these structures, virtually all the water molecules making up the solvent volume of the crystal are well ordered and are therefore explicitly present in the structural model. This high degree of solvent ordering may, in fact, contribute to the extremely good scattering properties of these crystals. They are not typical of protein crystals in general, however, and it is not clear that the behaviour of water molecules in the refined models of such structures could be used to derive heuristic expectations for the behaviour of water in more typical protein structure refinements.

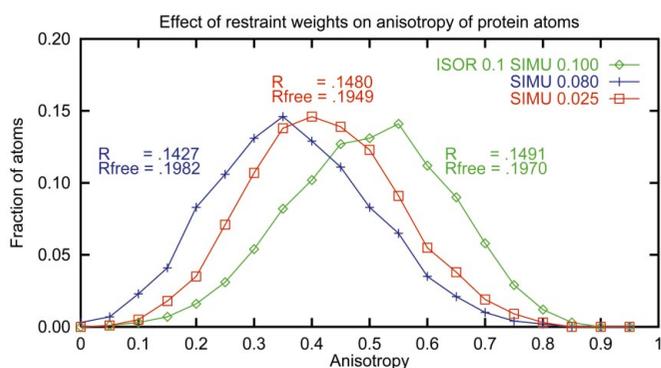
Nevertheless, many of the crystal structures in Table 1 are not particularly exceptional with regard to their total water content or the fractional volume of disordered solvent. The mean anisotropy for water molecules in these deposited structures ranges from less than that of the corresponding protein (*e.g.* barnase) to somewhat greater (*e.g.* P56<sup>1ck</sup>).



**Figure 4**

(a) The mean anisotropy  $\bar{A}$  of the protein atoms in each structure is plotted against the mean temperature factor  $B_{eq}$  for that structure. Room-temperature structure determinations are highlighted in yellow. Refinements performed at 1  $\text{\AA}$  resolution or better are in blue boxes. While the largest anisotropy mean (lowest  $\bar{A}$ ) is indeed observed in a subset of the room-temperature structures, there does not appear to be any general correlation of  $\bar{A}$  with  $B_{eq}$ . (b) The correlation of anisotropy with  $B_{eq}$ . For each structure, the correlation between anisotropy and  $B_{eq}$  was calculated for all protein atoms. This correlation is in general negative, indicating that atoms with higher  $B_{eq}$  are also more anisotropic (lower value of  $\bar{A}$ ). The horizontal axis of the plot indicates the mean temperature factor  $B_{eq}$  of the structure. The vertical axis indicates the correlation coefficient between  $\bar{A}$  and  $B_{eq}$  for individual atoms within the structure. Room-temperature structure determinations are highlighted in yellow. Refinements performed at 1  $\text{\AA}$  resolution or better are in blue boxes. It is clear that the higher resolution structures tend to exhibit a stronger correlation of  $\bar{A}$  with  $B_{eq}$ .

However, in most cases it is not possible to tell whether the restraints applied to the water molecules were, in fact, optimal. As part of the 1.25 Å refinement of the cholera toxin B-pentamer, a systematic evaluation explored the effect that the ISOR restraint weight applied to water molecules has on the resulting values of  $A$ ,  $R$  and  $R_{\text{free}}$  (Merritt *et al.*, 1998). This analysis indicated that  $R_{\text{free}}$  was lowest and the distribution of anisotropy plausible when ISOR was chosen to yield a value of  $A$  for waters which was slightly less than that obtained for the protein atoms (Fig. 2). A similar, though less extensive, evaluation has been made for the ongoing 1.4 Å resolution refinement of *E. coli* LTB (unpublished results). Even in this case, falling as it does in the middle of the resolution grey zone, treating the water molecules anisotropically seems to be justifiable: relative to an isotropic water model, the residual  $R$  drops from 0.146 to 0.138,  $R_{\text{free}}$  drops from 0.195 to 0.191 and the resulting distribution of anisotropy for water molecules ends up similar to that of the protein atoms. A tentative heuristic can therefore be proposed for the treatment of ordered water molecules during refinement of structures lying in the resolution grey zone between 1.6 and 1.2 Å. At the end of refinement with isotropic water molecules, the default weighting (ISOR 0.1 in *SHELXL97*) can be used to restrain an initial anisotropic treatment of waters. A *PARVATI* plot such as that given in Fig. 2 can then be used to compare the resulting distribution of anisotropy for protein atoms and for water molecules. If the two distributions are substantially different, it is probably worth re-running the refinement with



**Figure 5**  
Effect of varying restraint strength on resultant distribution of anisotropy. This figure shows the distribution of anisotropy among the 9697 protein atoms of the LTB–galactose complex refined against 1.4 Å data (unpublished results; Table 1). Three refinement runs were made from the same starting point and differed only in the restraints applied to the ADPs of protein atoms. Each run required 27 h of CPU time on a 400 MHz DEC Alphaserber 4100. One run (green line) restrained each atom to be approximately isotropic (ISOR 0.1) and applied a relatively weak restraint to the similarity of ADPs belonging to adjoining atoms (SIMU 0.100). The second (blue line) and third (red line) runs dropped the restraint toward isotropy and applied increasingly stringent restraints on ADP similarity (SIMU 0.080 and 0.025). The three runs yielded a mean anisotropy  $\bar{A}$  of 0.524, 0.393 and 0.446, respectively. The second run, with the weakest restraints, yielded the lowest standard crystallographic residual  $R$  as expected, but resulted in ADPs skewed toward extreme anisotropy. The third run, employing the strongest SIMU restraint, yielded the lowest  $R_{\text{free}}$  value as well as a more symmetric distribution of anisotropy than either of the first two runs.

the restraint weight adjusted so as to shift the water distribution toward that of the protein atoms. The validity of including anisotropic displacement parameters for the water molecules and the choice of restraint weights may be judged by following the  $R_{\text{free}}$  residual.

## 5. Inclusion of ADPs in lower resolution structures

In certain cases, a model for anisotropy may be introduced into protein structure refinement even at low-to-moderate resolution. One is the somewhat special case of metal centers. Here, one can limit expansion to a full anisotropic model (six thermal parameters per atom) to a specific small set of atoms, for example the Fe and S atoms in an Fe/S cluster. These electron-dense and well ordered atoms are likely to be robustly determined by the diffraction data, and in general will contribute substantial peaks to a residual-density map if they are inadequately described by a purely isotropic treatment. For example, Libeu *et al.* (1997) used *SHELXL* to refine a series of structures containing reduced or oxidized forms of pseudoazurin point mutants at resolutions near 2 Å. Only a Cu atom and two of its four liganding atoms (a methionine  $S^{\delta}$  and a cysteine  $S^{\gamma}$ ) were treated anisotropically. The addition of these 15 new parameters to the model did not noticeably lower the crystallographic  $R$  factor, but did improve residual-density maps. In this case, the largest component of anisotropy was in the direction of shifts between oxidation centers, which could reflect either true displacement along those vectors or the presence of an admixture of oxidation states.

### 5.1. Other representations of anisotropy

Although anisotropic displacement parameters are assigned to individual atoms, this may obscure the physical basis for the distribution of electron density which they describe. The electron density at atomic centers is non-spherical not only because of oscillatory motions of the individual atoms, but also because of both dynamic and static disorder at the level of larger structural units or indeed of the entire protein molecule. A second path to modeling anisotropy, which requires a relatively small number of added parameters, is to describe the anisotropic behavior of groups of atoms rather than of individual protein atoms. For example, a set of up to 20 parameters can be used to describe the TLS (translation/libration/screw) motion of any rigid body (Schomaker & Trueblood, 1968). This approach, as implemented in the *RESTRAIN* program (Driessen *et al.*, 1989), has been used to model domain motion in protein structures refined even at modest resolution (Moss *et al.*, 1996; Papiz & Prince, 1996). At intermediate resolution, a TLS model can be refined to describe the motion of individual side chains (Holbrook *et al.*, 1985; Howlin *et al.*, 1989), thereby reducing the number of refined parameters relative to a model with six ADPs per atom. For example, in the simple case of no overall domain or backbone motion, the  $U^{ij}$ s of the rigidly linked atoms in a Trp side chain are highly correlated by torsion-angle rotations about the  $C^{\beta}-C^{\gamma}$  bond. The anisotropy of the atoms in the indole group may, therefore, be

described as well by a single set of  $\leq 20$  TLS parameters as by 54 ADPs (six for each of nine atoms).

At atomic resolution, there may be sufficient information to partition anisotropy into local and large-scale components. In this case, TLS refinement of domain motions can be complementary to refinement of anisotropic displacement parameters for individual atoms (Stec *et al.*, 1995). Finally, as noted above in passing, it might be fruitful to explore the use of TLS models to generate restraints on individual atomic displacement parameters. As the database of atomic resolution structures refined with an expanded model for anisotropy grows, exploration may show that the ADPs within recurring structural elements are correlated. That is, there may be vibrational modes or static distributions of positional disorder that are characteristic of, say, certain side chains,  $\alpha$ -helices or other secondary-structural elements. At that point, not only will we have expanded our model for individual crystal structures, but we will also have expanded our understanding of protein structures in general.

## 6. Availability of the validation/analysis tools

The *rastep* and *PARVATI* programs are available from the author. Access to an automated version of these tools is available as an on-line validation/analysis resource *via* the WWW at URL <http://www.bmsc.washington.edu/parvati/>.

## References

Anderson, D. H., Weiss, M. S. & Eisenberg, D. (1997). *J. Mol. Biol.* **273**, 479–500.

Banuelos, S., Saraste, M. & Carugo, K. D. (1998). *Structure*, **6**, 1419–1431.

Brünger, A. T. (1992a). *X-PLOR. Version 3.1. A System for X-ray Crystallography and NMR*. Yale University, New Haven, CT, USA.

Brünger, A. T. (1992b). *Nature (London)*, **355**, 472–474.

Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1997). *Curr. Opin. Struct. Biol.* **7**, 681–688.

Dauter, Z., Wilson, K., Sieker, L. C., Meyer, J. & Moulis, J.-M. (1997). *Biochemistry*, **36**, 16065–16073.

Dauter, Z., Wilson, K., Sieker, L. C., Moulis, J.-M. & Meyer, J. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 8836–8840.

Deacon, A., Gleichmann, T., Kalb, A. J., Price, H., Raftery, J., Bradbrook, G., Yariv, J. & Helliwell, J. R. (1997). *J. Chem. Soc. Faraday Trans.* **93**, 4305–4317.

Driessen, H., Haneef, M. I. J., Harris, G. W., Howlin, B., Khan, G. & Moss, D. S. (1989). *J. Appl. Cryst.* **22**, 510–516.

Dunitz, J., Schomaker, V. & Trueblood, K. N. (1989). *J. Phys. Chem.* **92**, 856–867.

EU 3-D Validation Network (1998). *J. Mol. Biol.* **276**, 417–436.

Frazão, C., Soares, C. M., Carrondo, M. A., Pohl, E., Dauter, Z., Wilson, K. S., Hervás, M., Navarro, J. A., De la Rosa, M. A. & Sheldrick, G. M. (1995). *Structure*, **3**, 1159–1169.

Harata, K., Abe, Y. & Muraki, M. (1998). *Proteins Struct. Funct. Genet.* **30**, 232–243.

Holbrook, S. R., Dickerson, R. E. & Kim, S.-H. (1985). *Acta Cryst.* **B41**, 255–262.

Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **381**, 272.

Howlin, B., Moss, D. S. & Harris, G. W. (1989). *Acta Cryst.* **A45**, 851–861.

Johnson, C. K. (1965). Report ORNL-3794. Oak Ridge National Laboratory, Oak Ridge, TN, USA.

Khan, A. R., Parrish, J. C., Fraser, M. E., Smith, W. W., Bartlett, P. A. & James, M. N. G. (1998). *Biochemistry*, **37**, 16839–16845.

Kleywegt, G. J. (1996). *Acta Cryst.* **D52**, 842–857.

Kleywegt, G. J. & Read, R. J. (1997). *Structure*, **5**, 1557–1569.

Konnert, J. H. (1976). *Acta Cryst.* **A32**, 614–617.

Konnert, J. H. & Hendrickson, W. A. (1980). *Acta Cryst.* **A36**, 344–350.

Kumaraswamy, V. S., Lindley, P. F., Slingsby, C. & Glover, I. D. (1996). *Acta Cryst.* **D52**, 611–622.

Libeu, C. A. P., Kukimoto, M., Nishiyama, M., Horinouchi, S. & Adman, E. T. (1997). *Biochemistry*, **36**, 13160–13179.

Longhi, S., Czjzek, M., Lamzin, V., Nicolas, A. & Cambillau, C. (1997). *J. Mol. Biol.* **268**, 779–799.

MacArthur, M. W. & Thornton, J. M. (1996). *J. Mol. Biol.* **264**, 1180–1195.

Martin, C., Richard, V., Salem, M., Hartley, R. & Manguen, Y. (1999). *Acta Cryst.* **D55**, 386–398.

Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.

Merritt, E. A., Kuhn, P., Sarfaty, S., Erbe, J. L., Holmes, R. K. & Hol, W. G. J. (1998). *J. Mol. Biol.* **282**, 1043–1059.

Moss, D. S., Tickle, I. J., Theis, O. & Wostrack, A. (1996). *Proceedings of the CCP4 Study Weekend*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 105–113. Warrington: Daresbury Laboratory.

Papiz, M. Z. & Prince, S. M. (1996). *Proceedings of the CCP4 Study Weekend*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 115–123. Warrington: Daresbury Laboratory.

Rice, L. M. & Brünger, A. (1994). *Proteins Struct. Funct. Genet.* **19**, 277–290.

Schomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* **B24**, 63–76.

Sevcik, J., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1996). *Acta Cryst.* **D52**, 327–344.

Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.

Stec, B., Zhou, R. & Teeter, M. M. (1995). *Acta Cryst.* **D51**, 663–681.

Steinrauf, L. K. (1998). *Acta Cryst.* **D54**, 767–780.

Stout, C. D., Stura, E. A. & McRee, D. E. (1998). *J. Mol. Biol.* **278**, 629–639.

Teeter, M. M. (1992). *Dev. Biol. Stand.* **74**, 63–72.

Tong, L., Warren, T. C., King, J., Betageri, J. R. & Jakes, S. (1996). *J. Mol. Biol.* **256**, 601–610.

Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* **A43**, 489–501.

Trueblood, K. N., Bürgi, H.-B., Burzlaff, H., Dunitz, J. D., Gramaccioni, C. M., Schulz, H. H., Shmueli, U. & Abrahams, S. C. (1996). *Acta Cryst.* **A52**, 770–781.

Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, A., Lamzin, V. S. & Wilson, K. S. (1998). *Acta Cryst.* **D54**, 522–546.

Wang, Z., Luecke, H., Yao, N. & Quiocho, F. A. (1997). *Nature Struct. Biol.* **4**, 519–522.