Comparing anisotropic displacement parameters in protein structures

Ethan A. Merritt

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

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

The increasingly widespread use of synchrotron-radiation sources and cryo-preparation of samples in macromolecular crystallography has led to a dramatic increase in the number of macromolecular structures determined at atomic or near-atomic resolution. This permits expansion of the structural model to include anisotropic displacement parameters U^{ij} for individual atoms. In order to explore the physical significance of these parameters in protein structures, it is useful to be able to compare quantitatively the electron-density distribution described by the refined U^{ij} values associated with corresponding crystallographically independent atoms. This paper presents the derivation of an easily calculated correlation coefficient in real space between two atoms modeled with anisotropic displacement parameters. This measure is used to investigate the degree of similarity between chemically equivalent but crystallographically independent atoms in the set of protein structural models currently available from the Protein Data Bank.

1. Introduction

Because diffraction from crystals of macromolecules is typically poorer than that from crystals of small molecules, it is usually necessary to limit the number of parameters in the structural model. In particular, individual atoms in the structural model are generally assigned only four parameters: three positional parameters, x, y and z, and one thermal parameter B. However, routine cryo-preparation of protein sample crystals and the use of extremely intense synchrotron-radiation sources have led to an increasing number of cases in which diffraction from protein crystals can be measured to near-atomic resolution. In these cases, it is possible to expand the structural model to include individual anisotropic displacement parameters (ADPs) for each atom. This leads to a model containing nine parameters per atom, with six anisotropic displacement parameters U^{ij} replacing the single parameter B. This class of structural model is very common for small molecules and the physical significance of the refined U^{ij} parameters as indicators of atomic vibrational modes is well understood (Dunitz *et al.*, 1989). Crystals of proteins and other macromolecules differ from typical small-molecule crystals in a number of ways, however, and thus it is worth examining the reliability and physical significance of the U^{ij} parameters of protein models.

Several characteristic features of macromolecular crystals suggest that they may exhibit qualititatively different internal vibration modes and static disorder to those commonly found in small-molecule crystals. A protein consists of a polymer of amino-acid residues folded into secondary structural elements

such as α-helices and β-sheets. These in turn assemble into tertiary and quaternary structural elements. Each level of structural element may have a characteristic set of vibrational modes. For example, while amino acids other than proline and glycine exhibit torsional flexibility about the $C^\alpha - C^\beta$ bond, the additional modes of torsional vibration beyond C^β differ from one residue type to another. Furthermore, even in protein crystals which diffract to very high resolution, 50% or more of the unit-cell volume may contain unordered solvent. This relatively loose packing means that large domains within the ordered portion of the cell may exhibit vibrational modes or stochastic displacement that are not constrained by crystal lattice contacts. For example, domains or subdomains of the protein may flex relative to each other *via* hinge motions. All of these various vibrational modes contribute anisotropically to X-ray scattering. Unless these higher level modes are separately modeled, *e.g.* by a translation–libration–screw (TLS) treatment (Schomaker & Trueblood, 1968), their net effects are taken up by the refined values of the anisotropic displacement parameters U^{ij} associated with their individual constituent atoms. For this reason, we expect the U^{ij} parameters refined for typical protein atoms to reflect a greater degree of displacement than those typical for atoms in a small-molecule structure. This is already true in the case of isotropic models of protein structure, of course, and is one reason that the isotropic thermal parameters, B, in refined protein models are as large as they are.

There are several reasons why it would be useful to compare the values of U^{ij} between independent protein structures. On concern is the reliability of the refined U^{ij} parameters of a protein structural model. Even the highest resolution protein refinements to date do not approach the resolution possible for small-molecule structures. Consequently, the number of observations per parameter in a fully anisotropic protein model is less than for a typical small-molecule structure and the parameters are accordingly less well determined. The problem is made worse by the large size of the protein model, which makes it computationally impractical to invert the full matrix of normal equations to yield estimates for the standard uncertainty of the refined parameters. This is particularly true when U^{ij} parameters are included in the structural model, as the mixture of positional and displacement parameters produces a poorly conditioned matrix which requires greater numerical precision to invert (Kevin Cowtan, personal communication). In the absence of standard uncertainty values from matrix inversion, it would be useful to investigate the reproducibility of the refined U^{ij} values by comparing the values obtained from independent refinement of isomorphous structures.

The relatively low observation-to-parameter ratio, particularly at resolutions worse than 1 Å, can be mitigated by introducing restraints on the allowed values of the U^{ij} parameters during refinement. These are exactly analogous to geometric restraints on bond lengths and angles imposed during refinements at lower resolution. Unlike geometric restraints, however, we have as yet limited experience by which to choose optimal targets for ADP restraints (Merritt,

1999). The refined U^{ij} parameters in a protein model contain contributions from very local vibrational modes, as in small-molecule crystal structures, and also from larger scale modes of inter-domain flexibility and molecular libration. We do not know the relative magnitude of these contributions and, indeed, this may vary from one protein structure to another. This becomes important when we consider how best to refine protein structures containing non-crystallographic symmetry (NCS). Since data resolution is a major limiting factor in protein structure refinements and since the use of parameter restraints is, therefore, very helpful, it would be useful to know whether NCS restraints on atomic U^{ij} parameters are justified. If the net atomic displacement of individual atoms is dominated by local vibrational modes intrinsic to the protein, then it would make sense to consider applying NCS restraints to the U^{ij} parameters as well as to the positional parameters. If, on the other hand, the net atomic displacement is dominated by large-scale libration constrained by lattice packing, then one would not expect the U^{ij} parameters to exhibit the same degree of non-crystallographic symmetry as the positional parameters of those same atoms. Ideally, one would test this by refining a suitable structure both with and without NCS restraints on the U^{ij} parameters and comparing the R and R_{free} residuals resulting from the parallel refinements. This test is not possible using existing refinement programs, which do not support NCS restraint of U^{ij} parameters. We can, however, examine the U^{ij} parameters in protein structures refined with no such restraints to judge the degree to which the U^{ij} values of NCS related atoms are nevertheless correlated with each other.

2. Anisotropic displacement parameters

The scattering power of an atom falls off with scattering angle because it does not act as a point scatterer, but rather as an electron cloud of finite extent. To account for this scattering behaviour, one may add an angle-dependent term to the individual atomic scattering factor for each atom. If this correction is limited to a single parameter, it is usually applied in the form of a spherical Gaussian,

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

where $\langle u^2 \rangle$ is the mean-square amplitude of displacement, h is a reciprocal-lattice vector, λ is the X-ray wavelength and θ is the corresponding scattering angle. In macromolecular crystallography, it is more common to express this in terms of an isotropic displacement parameter $B = 8\pi^2 \langle u^2 \rangle$. If sufficient data are available, this correction can be expanded by replacing the single isotropic displacement parameter with a set of six parameters U^{ij} which constitute the unique elements of a symmetric 3×3 tensor U describing a three-dimensional Gaussian model for the anisotropic fall-off in scattering power with scattering angle for that atom. The correction then takes the form

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

This corresponds in real space to a three-dimensional Gaussian probability distribution for the electron density of that atom. This distribution is commonly visualized by contouring an isolevel of probability, which yields the surface of an ellipsoid.

3. Comparing the anisotropy of equivalent atoms

There are at least three features of the 3×3 symmetric tensor U that may be described easily, each of which might be used in comparing the anisotropic displacement parameters of corresponding atoms. The simplest of these is the magnitude of the equivalent isotropic displacement parameter $U_{\mathrm{eq}} = 1/3(U^{11} + U^{22} + U^{33})$. This is related to the B factor usually reported for an atom in an isotropic protein model by the relationship $B = 8\pi^2 \langle U_{\mathrm{eq}}^2 \rangle$. Any given isolevel contour of the spherical Gaussian probability distribution determined by U_{eq} encloses a sphere containing the same integrated probability enclosed by the ellipsoidal isolevel contour of the original distribution U. As we shall see below, in comparing the anisotropic model for a pair of atoms it may be desirable to normalize them first so that they have the same U_{eq}.

The second obvious feature of the anisotropic model for an atom is the degree to which it deviates from being spherical. This is called the anisotropy, A, and is formally defined as the ratio of the smallest to the largest eigenvalue of the 3×3 matrix U (Trueblood *et al.*, 1996). In real space, A corresponds to the axial ratio of the ellipsoidal isosurface for that atom. In a perfectly spherical distribution all axes are equal, so $A = 1$ describes a spherical atom. As a distribution becomes increasing non-spherical, its axial ratio becomes more extreme and thus A drops toward 0. Note that a given value of anisotropy $0 < A < 1$ may describe either a prolate 'cigar-like' atom or an oblate 'pancake-like' atom, depending on whether the length of the third axis is more like that of the shortest or the longest axis.

Both U_{eq} and A capture some sense of the 'shape' of the anisotropic atom, but they do not convey any information about the orientation of the ellipsoid. Two atoms with identical values of U_{eq} and A may nevertheless be quite unlike with respect to their direction of vibration. Of course, the U^{ij} parameters of the 3×3 tensor U themselves fully describe the anisotropic model, but it would be convenient to have some scalar metric that describes how similar or dissimilar two tensors U and V are to each other. To do this, we will turn our attention to the electron distributions $\rho_u(x)$ and $\rho_v(x)$ in real space that correspond to the anisotropic displacement tensors U and V.

4. The correlation coefficient in real space

It is common in macromolecular crystallography to express the similarity of two electron-density maps by calculating the correlation coefficient between them. If the two maps are calculated on the same three-dimensional grid, their correlation coefficient is given by

$$\mathrm{cc} = \frac{\sum_i (\rho_{u_i} - \overline{\rho_u})(\rho_{v_i} - \overline{\rho_v})}{[\sum_i (\rho_{u_i} - \overline{\rho_u})^2 \sum_i (\rho_{v_i} - \overline{\rho_v})^2]^{1/2}},$$

where the index i covers all relevant grid points. In the limit of an infinitely fine grid extending over all space, both $\overline{\rho_u}$ and $\overline{\rho_v}$ approach 0 and the corresponding continuous summation becomes

$$\mathrm{cc}_{uij} = \frac{\int \rho_u(x)\rho_v(x)}{[\int \rho_u(x)\rho_u(x) \int \rho_v(x)\rho_v(x)]^{1/2}}, \tag{1}$$

where the integrals are taken over all space.

4.1. Derivation of cc$_{uij}$

We would like to calculate the quantity in (1) directly from the two associated tensors U and V. We know that the probability distribution in real space $\rho_u(x)$ corresponding to an ADP tensor U is given by

$$\rho_u(x) = \left(\frac{\det U^{-1}}{8\pi^3}\right)^{1/2} \exp(-\tfrac{1}{2}x^T U^{-1} x),$$

so the integral of the product of two such distributions described by U and V in the numerator of (1) is given by

$$\int \rho_u(x)\rho_v(x) = \left(\frac{\det U^{-1}}{8\pi^3}\right)^{1/2} \left(\frac{\det V^{-1}}{8\pi^3}\right)^{1/2} \\ \times \int \exp\{-\tfrac{1}{2}[x^T(U^{-1} + V^{-1})x]\}. \tag{2}$$

Now let us suppose that we had an ADP matrix W such that $W^{-1} = (U^{-1} + V^{-1})$. It would correspond to a hypothetical real-space probability distribution $\rho_w(x)$, where

$$\rho_w(x) = \left(\frac{\det W^{-1}}{8\pi^3}\right)^{1/2} \exp(-\tfrac{1}{2}x^T W^{-1} x),$$

and since this is a probability distribution, its integral over space $\int \rho_w(x)$ must equal 1. Therefore, we have

$$1 = \int \rho_w(x) = \left(\frac{\det W^{-1}}{8\pi^3}\right)^{1/2} \int \exp(-\tfrac{1}{2}x^T W^{-1} x),$$

$$\int \exp(-\tfrac{1}{2}x^T W^{-1} x) = \left(\frac{\det W^{-1}}{8\pi^3}\right)^{-1/2}. \tag{3}$$

Substituting (3) into (2), we find that

$$\int \rho_u(x)\rho_v(x) = \left(\frac{\det U^{-1}}{8\pi^3}\right)^{1/2} \left(\frac{\det V^{-1}}{8\pi^3}\right)^{1/2} \\ \times \int \exp\{-\tfrac{1}{2}[x^T(U^{-1} + V^{-1})x]\} \\ = \left(\frac{\det U^{-1}}{8\pi^3}\right)^{1/2} \left(\frac{\det V^{-1}}{8\pi^3}\right)^{1/2} \left(\frac{\det W^{-1}}{8\pi^3}\right)^{-1/2} \\ = \left[\frac{\det U^{-1} \det V^{-1}}{8\pi^3 \det(U^{-1} + V^{-1})}\right]^{1/2}. \tag{4}$$

The corresponding self-product is

$$\int \rho_u(x)\rho_u(x) = \left(\frac{\det U^{-1}}{64\pi^3}\right)^{1/2}. \tag{5}$$

Substituting (4) and (5) into the original expression for the correlation coefficient gives

$$\mathrm{cc}_{uij}(U, V) = \frac{\int \rho_u(x)\rho_v(x)}{\left[\int \rho_u(x)\rho_u(x) \int \rho_v(x)\rho_v(x)\right]^{1/2}}$$

$$= \frac{\left[\dfrac{\det U^{-1} \det V^{-1}}{8\pi^3 \det(U^{-1} + V^{-1})}\right]^{1/2}}{\left[\left(\dfrac{\det U^{-1}}{64\pi^3}\right)^{1/2}\left(\dfrac{\det V^{-1}}{64\pi^3}\right)^{1/2}\right]^{1/2}}$$

$$= \frac{\left(\det U^{-1} \det V^{-1}\right)^{1/4}}{\left[(1/8)\det(U^{-1} + V^{-1})\right]^{1/2}}, \quad (6)$$
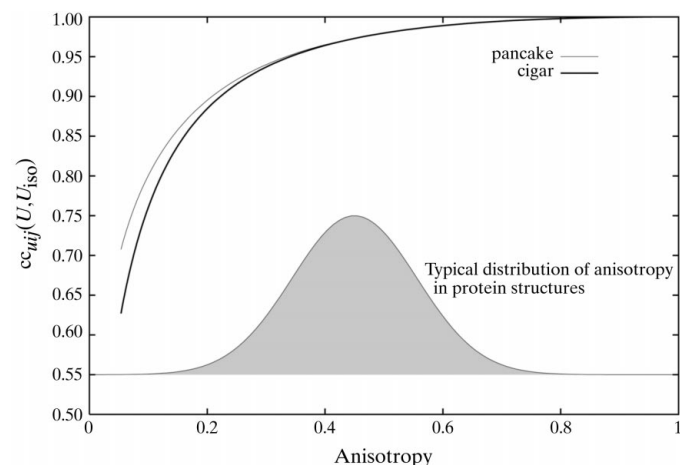
which, as we wanted, is easily calculated directly from U and V.

4.2. Properties of the unnormalized and normalized cc_{uij} correlation coefficient

Let us see what values cc_{uij} takes on when applied to a few simple cases. If U_{iso} and V_{iso} describe a pair of isotropic atoms, with $U_{\mathrm{iso}}^{11} = U_{\mathrm{iso}}^{22} = U_{\mathrm{iso}}^{33} = U_{\mathrm{eq}}$ and similarly for V_{iso}, then we have

$$\mathrm{cc}_{uij}(U_{\mathrm{iso}}, V_{\mathrm{iso}}) = (U_{\mathrm{eq}}V_{\mathrm{eq}})^{3/4}\left[\tfrac{1}{2}(U_{\mathrm{eq}} + V_{\mathrm{eq}})\right]^{-3/2}, \quad (7)$$

which is equal to 1 only if $U_{\mathrm{eq}} = V_{\mathrm{eq}}$. This means that if we want to compare the shape of two atoms, as distinct from their relative size, then we must first scale one to the other. This can be performed by shrinking or expanding it uniformly so that $U_{\mathrm{eq}} = V_{\mathrm{eq}}$. This correction is needed to correctly identify the similarity of two sets of atoms which differ only in their overall magnitude of thermal vibration, as might be expected, fo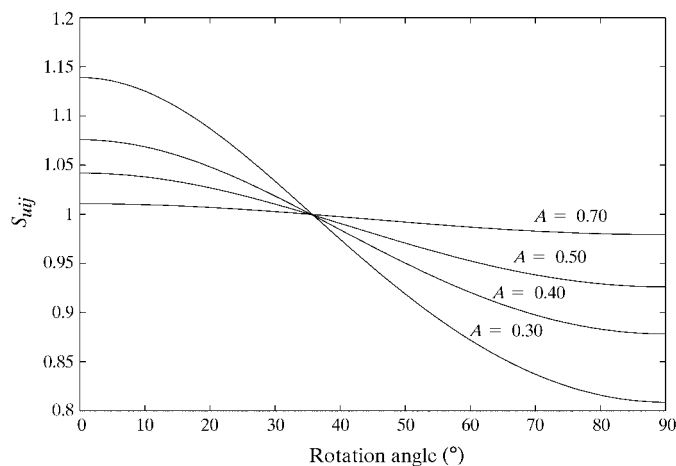r example, in paired determinations of a structure at room temperature and at 120 K. Another useful check is the correlation of the anisotropic electron distribution described by a general tensor U with its own isotropic approximation U_{iso}. This value, $\mathrm{cc}_{uij}(U, U_{\mathrm{iso}})$, is plotted in Fig. 1 as a function of the anisotropy of U. The value of $\mathrm{cc}_{uij}(U, U_{\mathrm{iso}})$ is 1 for an isotropic atom and drops as the atom described by U becomes more anisotropic. When the anisotropy becomes extreme, the value of cc_{uij} drops asymptotically to zero. For atoms in a protein structure refined at near-atomic resolution, the mean anisotropy is typically in the range 0.4–0.5 (Merritt, 1999), and for atoms whose anisotropy is near this mean value $\mathrm{cc}_{uij}(U, U_{\mathrm{iso}})$ is above 0.95 (Fig. 1).

With these results in hand, we are ready to formulate the precise test we will use for a pairwise comparison of atoms, each with a fully general set of anisotropic displacement parameters. The question we will ask is: if these two atoms had the same size, would they look more like each other than like an isotropic atom of that same size. Having the 'same size' here means having the same U_{eq} value and this can be achieved by multiplying the tensor V by $U_{\mathrm{eq}}/V_{\mathrm{eq}}$. Our comparison then becomes

$$S_{uij}(U, V) = \frac{\mathrm{cc}_{uij}[U, (U_{\mathrm{eq}}/V_{\mathrm{eq}})V]}{\mathrm{cc}_{uij}(U, U_{\mathrm{iso}})\mathrm{cc}_{uij}(V, V_{\mathrm{iso}})}. \quad (8)$$

This normalized coefficient will be greater than 1 whenever the two atoms described by U and V are more similar to each other than to an isotropic atom, and will be less than or equal to 1 otherwise. The value of S_{uij} is specifically sensitive to the orientation of the atoms being compared. This may be shown by generating two identical ADP tensors which differ only by a pure rotation and following the value of S_{uij} as a function of the rotation angle (Fig. 2). For prolate atoms with a typical anisotropy of about 0.45, this value varies from 1.06 when the atoms are perfectly aligned to 0.90 when their long axes are maximally misaligned. Note that if either atom being

Figure 1
The behaviour of the coefficient $\mathrm{cc}_{uij}(U, U_{\mathrm{iso}})$ as a function of the anisotropy A of the atom described by U. Two curves are shown. One corresponds to a prolate (cigar-shaped) atom described by an ADP tensor U whose eigenvalues are $(1, A, A)$. The second corresponds to an oblate (pancake-shaped) atom whose ADP eigenvalues are $(1, 1, A)$. The inset Gaussian distribution indicates a typical distribution of anisotropy among the atoms of a protein structure refined at near-atomic resolution. The distribution shown has mean anisotropy $A = 0.45$ and $\sigma = 0.15$, as found from a survey of structures in the PDB (Merritt, 1999).
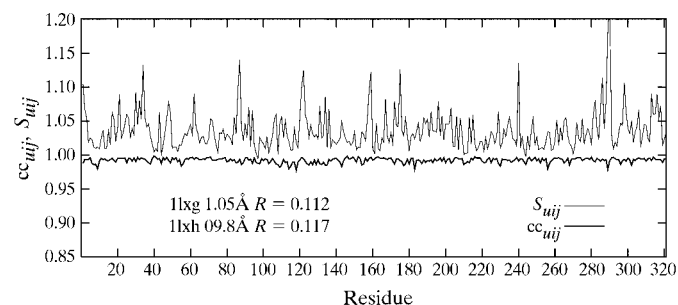

Figure 2
The behaviour of the similarity measure S_{uij} between two atoms differing only in orientation. Each curve represents the value of S_{uij} between two prolate atoms with anisotropy A and eigenvalues $(1, A, A)$ for their ADP tensors. The two atoms differ from each other by rotation about an axis perpendicular to the long axis of the corresponding ellipsoid.

compared is approximately isotropic, then the value of S_{uij} will be near 1. Thus, atom pairs which are very similar but also nearly isotropic will not score as highly as atoms which are similar and are also distinctly anisotropic. This is not likely to be a problem in comparing protein structures, as few atoms refined with ADPs remain isotropic unless strongly restrained (Merritt, 1999). Therefore, we will proceed to use this normalized coefficient S_{uij} to compare the similarity of the anisotropic models for equivalent atoms in refined protein models.

4.3. Implicit assumptions

Calculation of cc_{uij} and S_{uij} purely from the U^{ij} parameters implicitly assumes that the atoms being compared are superimposed such that their centers are coincident. Thus, the measure is insensitive to small conformational differences in their respective residues or to small inaccuracies in the NCS operation used to superimpose them. One could alternatively compare the electron-density distributions in real space, including both positional and thermal parameters. This approach was taken by Peters-Libeu & Adman (1997) to compare the immediate environment of metal centers among multiple protein structures, although in that case the atoms were modeled using only a single isotropic displacement parameter.

The normalization of one atom by U_{eq}/V_{eq} in the numerator of (8) to yield S_{uij} (U, V) is a post hoc correction to the refined model parameters V^{ij}. It implicitly assumes that the net ADP tensor from the refinement may be partitioned into a component arising from purely local vibrational modes and a more global, in this case isotropic, component. If such a partition can actually be made, it would more properly be performed by adding parameters to the refined model. So one might imagine refining a set of TLS parameters to describe large-scale motions by a hinged domain, while retaining the full six-parameter ADP description for the constituent atoms within the domain. Since the TLS parameters and the individual ADP parameters would be highly correlated, however, it is not clear that refinement of such a model would be stable.

5. Application to structural models in the Protein Data Bank

5.1. Isomorphous structures and the overall reliability of U^{ij} values

To date, only a relatively small number of protein structures have been refined with individual anisotropic displacement parameters and deposited in the Protein Data Bank. Among those that are present in the data bank, however, there are several sets of independently refined isomorphous structures. These include *Escherichia coli* phosphate-binding protein (Wang *et al.*, 1997; PDB codes 1ixh, 1ixg), hen egg-white lysozyme (Walsh *et al.*, 1998; PDB codes 3lzt, 4lzt) and *Streptomyces aureofaciens* ribonuclease (Sevcik *et al.*, 1996; PDB codes 1rgg, 1rge). The phosphate-binding protein structural pair consists of the wild-type protein determined to 0.98 Å and the Thr141Asp single-site mutant determined to 1.05 Å. The lysozyme pair consists the same structure determined to 0.95 Å at room-temperature and to 0.93 Å at liquid-nitrogen temperature. The pair of ribonuclease structures consists of the apoenzyme refined at 1.20 Å and a complex with 2′-GMP refined at 1.15 Å. This pair of structures is of particular interest because there are two crystallographically independent ribonuclease molecules in the cell, allowing us to compare the similarity of isomorphous copies across the two refined models and also the similarity of the NCS-related copies within each refined model.

The mean value of S_{uij} per residue for each of these structural pairs is shown in Figs. 3, 4 and 5. In calculating the mean value per residue, paired atoms from the two structures were included only if (i) they had identical atom names except for the chain identifier, (ii) no alternate conformation was indicated by the full set of atom-identifier fields in the PDB file and (iii) their superimposed positions agreed within 1.5 Å. Since no positional parameters enter into the calculation of S_{uij}, it is not sensitive to slight deviations from perfect superimposition of the two isomorphous structures. The 1.5 Å threshold was set to prevent comparison of inconsistently
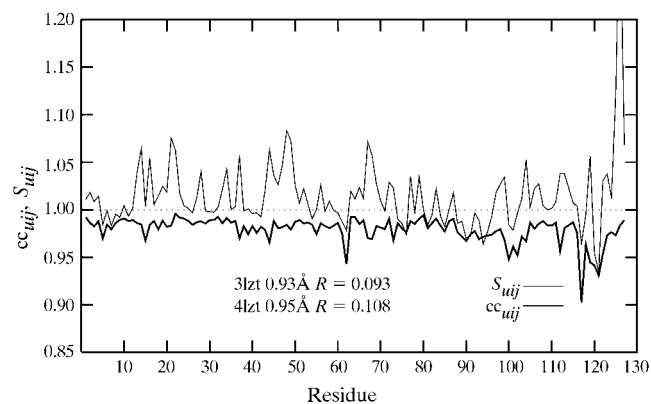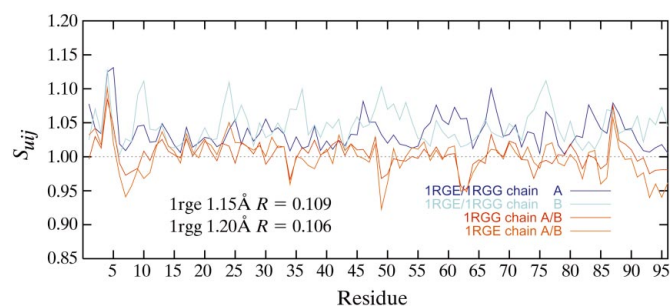

Figure 3
Pairwise comparison of refined anisotropic displacement parameters in independently refined isomorphous structures of *E. coli* phosphate-binding protein (Wang *et al.*, 1997; PDB codes 1ixh, 1ixg). Both structures were determined at 100 K. The mean value of the correlation coefficient cc_{uij} and of the similarity index S_{uij} is calculated for each residue.
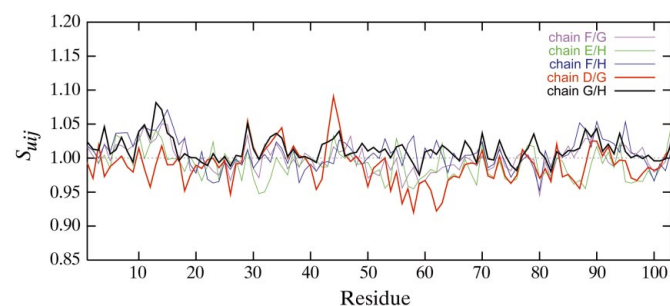

Figure 4
Pairwise comparison of refined anisotropic displacement parameters in independently refined isomorphous structures of hen egg-white lysozyme at room temperature and at 120 K (Walsh *et al.*, 1998; PDB codes 3lzt, 4lzt). The mean value of the correlation coefficient cc_{uij} and of the similarity index S_{uij} is calculated for each residue.

assigned atoms (for example, the atoms $O^{\delta 1}$ and $N^{\delta 2}$ of asparagine side chains) and of atoms belonging to side chains with substantially different conformations in the two structural models.

In the paired structures of the phosphate-binding protein and the ribonuclease, the mean value of S_{uij} is greater than 1 for all residues (Figs. 3 and 5). The overall mean value of S_{uij} for the entire phosphate-binding protein in the isomorphous pair is 1.04, with a standard deviation of 0.04. For the ribonuclease structures, the isomorphous chains A and B both have an overall mean S_{uij} of 1.04, with a standard deviation of 0.03. In the pair of lysozyme structures, the mean value of S_{uij} drops slightly below 1 for several residues, indicating that the similarity of the ADPs for these residues is no better than random. The overall mean S_{uij} for the lysozyme structures is 1.01, with a standard deviation of 0.05.

Figure 5
Pairwise comparison of refined anisotropic displacement parameters in independently refined isomorphous structures of *S. aureofaciens* ribonuclease (Sevcik *et al.*, 1996; PDB codes 1rgg, 1rge). Both structures were determined at room temperature. The similarity index S_{uij} is calculated for each residue. There are two crystallographically independent copies of the ribonuclease molecule in the cell, so comparisons can be made within each structure or between the two structures. One pair of plots (light and dark blue lines) shows the similarity between the same chain (A or B) as seen in each of two isomorphous structures; the other pair of plots (red lines) shows the similarity between the NCS-related chains within each of the two structures.
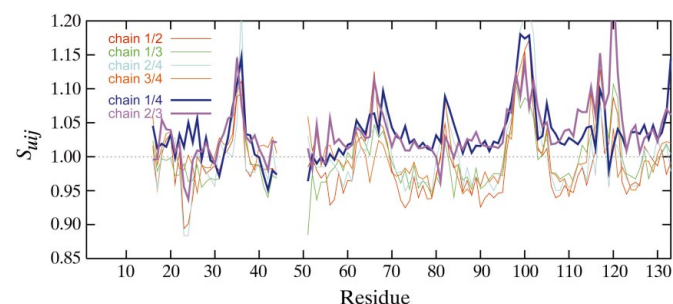

Figure 6
Mean similarity in anisotropy S_{uij} per residue in pairwise comparisons of monomers in the cholera toxin B-pentamer (Merritt *et al.*, 1998; PDB code 3chb). The five monomer chains, identified D through H, are related by fivefold non-crystallographic symmetry; the r.m.s. deviation of C^{α} positions from perfect non-crystallographic symmetry for chains compared pairwise ranges from 0.25 to 0.38 Å. The ADPs of chains G and H (black line) are the most similar; those of chains D and G (red line) the least similar. Three of the ten other possible pairwise comparisons are also shown. Only in the regions of residues 43–45 and 88–91 are the anisotropic models for all of the chains more similar to each other than to an isotropic model.

These three structural pairs support the view that the refined values of the U^{ij} parameters are not artifactual, since there is substantial agreement in the values assumed by equivalent atoms in independently refined isomorphous structures. However, this comparison of isomorphous structures is not informative as to whether the similarity in refined ADPs is a consequence of stereochemical equivalence within the protein structure or to equivalent packing environments within the crystal lattice.

6. Non-crystallographic symmetry

In order to compare the anisotropic displacement parameters U and V of two NCS-related atoms, the 3×3 tensor V must be suitably rotated in accord with the non-crystallographic symmetry operation. This is performed by taking $V' = R^{T}VR$, where R is the rotational component of the NCS operation relating the two atoms. Three sets of protein models from the Protein Data Bank were chosen to investigate the adherence of anisotropic displacement parameters to non-crystallographic symmetry. The first of these is the pair of ribonuclease structures, 1rge and 1rgg, already used as an isomorphous pair (Fig. 5). The second is the cholera toxin B-pentamer refined at 1.25 Å (Merritt *et al.*, 1998; PDB code 3chb), which contains five crystallographically independent copies of the toxin's 103-residue B subunit. The third case is the 1.14 Å refinement of a streptavidin mutant (Freitag *et al.*, 1999; PDB code 1swu), which contains four crystallographically independent subunits per asymmetric unit. In all cases, the NCS operation relating the two peptide chains to be compared was determined by least-squares superposition of C^{α} atoms.

The inter-chain similarity of U^{ij} parameters for atoms in chains A and B of the *S. aureofaciens* ribonuclease structures is low overall, although there are regions of good agreement (Fig. 5). Whereas the overall similarity for the ribonuclease molecules paired by isomorphism between structures was

Figure 7
Mean similarity in anisotropy S_{uij} per residue in pairwise comparisons of the four crystallographically independent subunits of the tetramer of the streptavidin point mutant Tyr43Phe (Freitag *et al.*, 1999; PDB code 1swu). Residues 45–52 belong to a flexible loop which is not completely modeled in all of the subunits (Freitag *et al.*, 1997) and hence are omitted from this comparison. The displacement parameters for atoms in chains 1 and 4 are similar to each other (blue line), as are those for atoms in chains 2 and 3 (magenta line). The similarity is distinctly less when comparing either chain 1 or 4 with chain 2 or 3.

greater than 1 [mean S_{uij} = 1.04 (0.03)], the overall similarity between the two NCS-related molecules internal to each structure was in each case less than 1 [mean S_{uij} = 0.98 (0.01) for 1rgg and 0.97 (0.02) for 1rge]. This is consistent with the published analysis of the structure (Sevcik *et al.*, 1996), in which it is noted that the r.m.s. deviation of main-chain atoms between chains *A* and *B* is significantly larger than the expected error in the refined coordinates. In particular, conformational differences at residues 48, 61–64, 74–77 and 94 are attributed to differences in the lattice environment. Fig. 5 shows that the calculated S_{uij} index for these regions is also worse than random and confirms that this similarity index can highlight residues whose ADPs are substantially influenced by the local packing environment within the crystal lattice.

The cholera toxin B-pentamer consists of five peptide chains, labeled *D*, *E*, *F*, *G* and *H*, arranged with regular five-fold symmetry about a central pore. The pentamer contains five receptor-binding sites, which in this structure are occupied by a receptor-derived pentasaccharide (Merritt *et al.*, 1998). Pairwise comparisons of S_{uij} between the chains are shown in Fig. 6. As with the ribonuclease comparison, the degree of similarity between NCS-related residues in these pairwise comparisons is near random overall. The mean overall S_{uij} for the entire peptide chain in the ten possible pairwise comparison ranges from 0.99 (0.04) for chains *G/D* to 1.01 (0.04) for chains *G/H*. However, when the plots are examined in finer detail several features emerge. Inclusion of chains *D* and *F* in the pairwise comparisons with the remaining three chains reduces the overall similarity. This is very likely to be a reflection of the fact that chains *D* and *F* are involved in a lattice-packing interaction which constrains their vibrational freedom and which prevents the receptor-binding sites in these chains from being simultaneously occupied by bound saccharide. This, in turn, frees the loop consisting of residues 51–60 in these two chains to flex within the local packing environment in at least half of the copies present in the cell, as this loop is in general only tightly ordered when sugars are bound (Merritt *et al.*, 1998). The inter-chain similarity, as indicated by S_{uij} > 1, is greatest in two distinct regions of the structure. The first of these is the turn consisting of residues 43–45. These residues form a turn projecting from the surface of the assembled pentamer and exhibit substantial anisotropy (Merritt *et al.*, 1998). These residues are part of a strongly antigenic epitope, as shown by response to ten-residue challenge peptides after exposure to intact toxin pentamer (Takahashi *et al.*, 1996). The similarity in ADPs probably indicates a characteristic vibrational mode of this three-residue turn that is found in all five crystallographically independent chains. The second region of high similarity between chains is the receptor-binding site, which is comprised of residues 10–14, 33, 56, 61 and 88–91. Residues 11–14 of chain *D* are a specific exception owing to the packing interaction mentioned above.

Streptavidin forms a homotetramer with D_2 (222) symmetry. A biotin-free structure of an engineered streptavidin core tetramer, containing residues 1–133 of the wild-type protein with the single substitution Tyr43Phe, has been fully refined at atomic resolution (Freitag *et al.*, 1999; PDB code 1swu). Residues in all four subunits of this structure are very similar with respect to their overall magnitude of displacement B_{iso} and to their magnitude of anisotropy A (data not shown). The ADPs of the four peptide chains in this structure are compared pairwise in Fig. 7 using the similarity index S_{uij}. It is strikingly evident that the ADPs for atoms in the four subunits do not follow the full non-crystallographic D_2 symmetry of the streptavidin homotetramer. Instead, subunits 1 and 4 show similar ADPs overall, as do subunits 2 and 3, but the two pairs of subunits are unlike each other. This is consistent with the previous observation (Hendrickson *et al.*, 1989) that the streptavidin tetramer can be considered to be a dimer of dimers owing to the extensive inter-subunit contacts between the pairs of subunits 1 + 2, and 3 + 4. From this single structure, therefore, we cannot determine whether the asymmetry of ADPs within the tetramer is primarily a consequence of the lattice environment or is a fundamental property of the streptavidin tetramer. Wild-type streptavidin and engineered variants have been crystallized in a number of crystal forms, several of which diffract to very high resolution (Freitag *et al.*, 1997; R. E. Stenkamp, personal communication). This set of structures may eventually provide an excellent test case for exploring the extent to which the modes of atomic displacement within the structure are affected by the lattice packing environment.

7. Concluding remarks

The S_{uij} index proposed here is derived from the correlation coefficient between two anisotropic Gaussian density distributions in real space, but is easily calculated directly from the corresponding anisotropic displacement parameters U^{ij}. The mean value of S_{uij} for multiple pairs of atoms conveys in a single number the similarity or dissimilarity of two sets of anisotropic displacement parameters. Therefore it provides a convenient measure by which anisotropic displacement parameters may be compared quantitatively across an extended set of atoms, as in a residue-by-residue comparison of protein structures. Possible applications include: (i) quantitative comparison of ADP values from replicate or parallel structure determinations, as a check on the reliability of the refined values, (ii) quantitative comparison of ADP values associated with atoms related by non-crystallographic symmetry or between secondary structural elements found in multiple protein structures and (iii) quantitative comparison of the effect of different refinement protocols, programs or restraint schemes on the resulting ADP estimates. Such investigation of refinement protocols is beyond the scope of this report, but a set of near-atomic resolution models from the Protein Data Bank has been used in preliminary exploration of the first two applications.

Evaluation of per-residue values of S_{uij} between three pairs of independently refined isomorphous structures in the Protein Data Bank showed substantial agreement between equivalent ADPs. Although this is clearly a very small set of test cases, it is reassuring evidence that the ADP values refined

for protein structures determined at near-atomic resolution are not artifacts of the data measured from a particular crystal. Demonstration of similar ADPs in isomorphous protein structures does not, of course, address the issue of whether the similarity arises primarily from intrinsic vibrational modes of the protein or from rigid-body motions characteristic of the crystal lattice. Previous attempts to estimate the magnitude of overall libration of proteins relative to the crystal lattice have reached different conclusions. For example, Stec *et al.* (1995) compared a set of models for crambin at 0.83 Å resolution containing translation–libration–screw (TLS) parameters for one, two or three rigid groups and concluded that rigid-body libration of the entire crambin molecule contributed 60% of the overall mobility. Anderson *et al.* (1997), however, found little evidence for overall libration from a *post hoc* analysis of individually refined atomic ADPs in the 1 Å resolution structure of pheromone Er-1.

Evaluation of the similarity between the ADPs of equivalent atoms related by non-crystallographic symmetry within a structure is another possible way of examining intrinsic vibrational modes separately from lattice vibration and from librational modes arising as a consequence of packing constraints. In the three cases presented here, the mean S_{uij} for peptide chains related by non-crystallographic symmetry was lower than for paired atoms in isomorphous structures. This is particularly clear in the case of the ribonuclease structures, where both isomorphous and non-crystallographic symmetry-related copies of the protein may be compared within the same pair of structures. We may tentatively conclude that the crystal lattice environment in these structures has a large effect on the ADPs. This suggests that rigid-body librational modes contribute a substantial component of the net anisotropic displacement for each atom. Because the total number of TLS parameters required to describe these modes is small relative to the number of parameters in a model with individually refined atomic ADPs, it also supports the potential value of adding TLS models for rigid-body vibration of proteins or of protein domains into structure refinements at medium resolution.

The evaluation of S_{uij} between protein subunits related by non-crystallographic symmetry may nevertheless be informative. Several regions of the proteins examined here exhibit substantial similarity despite having different lattice environments, including the residues that form the five receptor-binding sites in the cholera toxin B-pentamer. Conversely, in the case of the streptavidin tetramer it is at least possible that the asymmetry of ADPs is itself an intrinsic characteristic of the homotetramer rather than being an artifact of the crystal environment. Both the possible dominance of rigid-body motions and the possible intrinsic asymmetry of vibrational modes for non-crystallographic symmetry-related monomers suggest that it would not in general be useful to apply NCS restraints to the ADPs during protein structure refinement.

References

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

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

Freitag, S., Le Trong, I., Klumb, L., Stayton, P. S. & Stenkamp, R. E. (1997). *Protein Sci.* **6**, 1157–1166.

Freitag, S., Le Trong, I., Klumb, L. A., Stayton, P. S. & Stenkamp, R. E. (1999). *Acta Cryst.* D**55**, 1118–1126.

Hendrickson, W. A., Pähler, A., Smith, J. L., Satow, Y., Merritt, E. A. & Phizackerley, R. P. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 2190–2194.

Merritt, E. A. (1999). *Acta Cryst.* D**55**, 1109–1117.

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

Peters-Libeu, C. & Adman, E. T. (1997). *Acta Cryst.* D**53**, 56–77.

Schomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* B**24**, 63–76.

Sevcik, J., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1996). *Acta Cryst.* D**52**, 327–344.

Stec, B., Zhou, R. & Teeter, M. M. (1995). *Acta Cryst.* D**51**, 663–681.

Takahashi, I., Kiyono, H., Jackson, R. J., Fujihashi, K., Staats, H. F., Hamada, S., Clements, J. D., Bost, K. L. & McGhee, J. R. (1996). *Infect. Immun.* **64**, 1290–1298.

Trueblood, K. N., Bürgi, H.-B., Burzlaff, H., Dunitz, J. D., Gramaccioli, C. M., Schulz, H. H., Shmueli, U. & Abrahams, S. C. (1996). *Acta Cryst.* A**52**, 770–781.

Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, A., Lamzin, V. S. & Wilson, K. S. (1998). *Acta Cryst.* D**54**, 522–546.

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