Acta Cryst. (1985). B41, 191-201

Restraints in Temperature-Factor Refinement for Macromolecules: an Evaluation by Molecular Dynamics*

BY HSIANG-AI YU AND MARTIN KARPLUS

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138, *USA*

AND WAYNE A. HENDRICKSON

Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, DC 20375, *USA*

(*Received* 28 *June* 1984; *accepted* 12 *December* 1984)

Abstract

Molecular-dynamics simulations are employed to evaluate the validity of the restraints on the relative magnitudes of isotropic and anisotropic temperature factors incorporated into refinement programs for macromolecules. A comparison of refinement and simulation results is made for the bovine pancreatic trypsin inhibitor. The isotropic refinement restraints are found to be in qualitative accord with the calculated fluctuations. However, the simulation yields mean values for the difference in temperature factors for bonded-neighbor and next-neighbor atom pairs that are twice as large as those assumed in the restraints. Significant differences between the values for side-chain and main-chain atoms are found in the simulation. In the anisotropic refinement, the weights used to restrain neighboring-atom temperature factors are much more restrictive than the variation in the temperature-factor values obtained from the simulation results. The orientations of the anisotropy tensors found in the simulation do not have a simple relation to the static structure. Thus, use of stereochemical assumptions in the refinement yields incorrectly oriented anisotropy tensors and reduced values for the anisotropies. The riding-motion model is shown to be invalid in general for the description of correlations in the motions of bonded- and next-neighbor atom pairs in proteins. This means that restraints on the difference in temperature factors of such atom pairs cannot be identified with the variance of their interatomic-distance distributions.

A variety of experimental and theoretical techniques have been employed recently to determine the dynamics of proteins (Karplus & McCammon, 1981, 1983). The picture of fluctuating proteins that has emerged supplements the static average structures obtained from X-ray diffraction studies of crystals. Information is now available concerning both the magnitudes and the time scales of the motions that occur at ordinary temperatures. Root-mean-square atomic displacements fall in the range 0·2 to 1·5 Å; their time scale is 0·1 to 50 ps, with the longer times generally associated with the larger amplitudes. The fluctuations have been shown to be composed of contributions from individual atom oscillations and from collective motions involving a few atoms, entire residues, and whole regions of a protein (Swaminathan, Ichiye, van Gunsteren & Karplus, 1982; Morgan & McCammon, 1983). More global fluctuations with a longer time scale also are expected to occur (Karplus & McCammon, 1983), but much less information concerning them is available.

Of particular importance for providing experimental information concerning the magnitudes of the fluctuations are the Debye-Waller (temperature) factors extracted from X-ray crystallographic refinements of protein structures (Frauenfelder, Petsko & Tsernoglou, 1979; Artymiuk, Blake, Grace, Oatley, Phillips & Sternberg, 1979; Northrup, Pear, McCammon, Karplus & Takano, 1980). This is based on the identification of the temperature factors with the mean-square atomic displacements. It is assumed that the fluctuations are harmonic so that the probability distribution for the displacement of an atom from its mean position is Gaussian. Since the motions can be anisotropic, the anisotropy tensor for the fluctuation of each atom would, in general, be required. A simplification that is often introduced assumes isotropic atomic motion and thereby reduces the required motional parameters by a factor of six; that is, only one isotropic temperature factor per atom, rather than six anisotropic temperature tensor components need to be determined. Once the temperature factors (isotropic or anisotropic) have been evaluated from the X-ray refinement, it is necessary to know what fraction of the observed temperature factor arises from atomic motions. Other contributions come from crystal disorder and errors in the refinement process itself.

* Supported in part by grants from the National Science Foundation and the National Institutes of Health.

0108-7681/85/030191-11$01.50 © 1985 International Union of Crystallography

Some information for interpreting the results of temperature-factor measurements for proteins have been provided by molecular-dynamics simulations. In particular, studies of a number of molecules (cytochrome *c*, bovine pancreatic trypsin inhibitor, and lysozyme) indicate that for most of the atoms the motions are nearly harmonic (Mao, Pear, McCammon & Northrup, 1982; van Gunsteren & Karplus, 1982*a*; Ichiye, Olafson & Karplus, 1985). There are, however, clear exceptions where the motions are highly anharmonic; many of these arise from multiple minima for the position of a given atom. As to the assumption of isotropic motion, the simulation results indicate that this is a very poor approximation (Northrup, Pear, Morgan, McCammon & Karplus, 1981; van Gunsteren & Karplus, 1982*a*); most of the atomic fluctuations are highly anisotropic so that use of an isotropic temperature factor necessarily yields an average result. Finally, the question of non-motional contributions to the temperature factor is more difficult to resolve and probably is different for each protein crystal and each refinement procedure. A simulation of cytochrome *c* and its comparison with the temperature factors determined by Takano & Dickerson (Northrup *et al.*, 1980; Takano & Dickerson, 1980) suggests that for this case about half of the observed temperature factors arose from atomic fluctuations. For the heme iron in myoglobin, comparison of room-temperature Mössbauer and X-ray analyses indicated that the motional contribution is about 60% of the temperature factor (Frauenfelder *et al.*, 1979).

Because the resolution of protein crystals limits the number of reflections that can be observed, additional assumptions beyond those described above are generally introduced into the refinement procedure to increase the likelihood of obtaining meaningful temperature factors. This is particularly true if a refinement including anisotropic atomic motion is attempted since the five additional required parameters per atom make the uncertainties due to the limited number of measurable reflections even more severe. The general approach that has been used is to introduce restraints that reduce the effective parameter space. For the determination of the average structure, it has long been customary to restrain the local geometry (Diamond, 1966; Konnert, 1976; Konnert & Hendrickson, 1978). More recently, analogous restraints on the relative temperature-factor values have been introduced into the refinement procedure for atom neighbors along the polypeptide chain (Konnert & Hendrickson, 1980). For refinements including anisotropic displacements, it has been suggested that the orientation of the anisotropy tensor can be determined from the stereochemistry (Konnert & Hendrickson, 1980); this reduces six independent parameters per atom to three.

Since the simplifying assumptions described above have been incorporated into effective refinement programs that are being widely applied to proteins and other macromolecules (*e.g.* DNA), it is important to have available tests of the validity of the procedure. This is difficult to obtain from the experiment at the present stage. It is the object of the present paper to examine the assumptions described above by comparing them with the results of a molecular-dynamics simulation. We shall focus on the bovine pancreatic trypsin inhibitor (BPTI) for which restrained isotropic and anisotropic temperature-factor refinements (Hendrickson & Konnert, 1980*a*) and molecular-dynamics simulations (van Gunsteren & Karplus, 1982*a*) are available. First, we compare the experimental and calculated variances of the temperature factors for pairs of bonded atoms and for pairs of atoms each bonded to the same atom. Then, we consider the calculated anisotropy tensor and determine whether it is simply related to the stereochemistry of a given atom. Finally, we analyze the effect of the stereochemical orientation assumption on the values of the tensor components determined in the refinement procedure.

Method

We summarize the nature of the restraints introduced into the X-ray refinement procedure and outline the approach used to test these assumptions with the results of a molecular-dynamics simulation.

Restrained X-ray refinement

The restrained refinement procedure developed by Konnert & Hendrickson (1980) minimizes a function of the form

$$\Phi = \sum_i w_i(|F_o|_i - |F_c|_i)^2 + \sum_l w_l V_l^2 + \text{other restraints}, \quad (1)$$

where the w's are weights assigned to the observations, the V's are the variances of the interatomic-distance distributions and the F_c's and F_o's are the calculated and observed structure factors. The form of equation (1) is such that ideally the weights w_l correspond to the inverse of the variance V_l of a set of observations; thus, w_l can be chosen to obtain V_l values having the desired distributions. In the actual refinement procedure, an average weight w is used for a given type of restrained internal coordinate (Konnert & Hendrickson, 1980).

Isotropic restraints

In applications of equation (1) to proteins, the sum over l has included both bonded atoms and atoms bonded to a common atom; that is, it is assumed that as a result of the bonding there is a correlation between the motions of such pairs of

atoms which limits their relative fluctuations. Konnert & Hendrickson (1980) suggested that the 'riding-motion' model of Busing & Levy (1964) could be used as a guide for restraining the variance V_{ab} of the internuclear distance between atoms a and b. For the isotropic case with atom b 'riding' on atom a, V_{ab} is given by

$$V_{ab} = \Delta^2 - \frac{\Delta^4}{d_0^2} + \cdots, \qquad (2)$$

where d_0 is the distance between the mean positions of the atoms and $\Delta^2 = \langle u_b^2 \rangle - \langle u_a^2 \rangle$. Here $\langle u^2 \rangle$ is the mean-square displacement of an atom along a given direction of movement. This is a component of the mean-square value, $\langle r^2 \rangle$, of the general instantaneous displacements, r, from the mean atomic position. In the isotropic case, $\langle u^2 \rangle = \frac{1}{3}\langle r^2 \rangle$. For the correlations represented by the 'riding-motion' approximation to be physically meaningful, Δ^2 must be positive. As the distinctions between atoms a and b are usually obscure, the absolute value is taken. With the first-order approximation to equation (2), the square of the variance used in isotropic refinement is

$$V'^2_{ab} = [\langle u_b^2 \rangle - \langle u_a^2 \rangle]^2 = [B_b - B_a]^2/(8\pi^2)^2, \qquad (3)$$

where B is the isotropic temperature factor.

With the variance restrained to have the form given in equation (3) the weighting factor w_l in equation (1) is chosen to obtain what is regarded as a suitable magnitude for V'_{ab}. In the application to BPTI, values chosen for average weighting factors are $(0{\cdot}013\ \text{Å}^2)^{-2}$ for directly bonded main-chain atoms (bond-length restraint), $(0{\cdot}019\ \text{Å}^2)^{-2}$ for both directly bonded side-chain atoms and next-nearest-neighbor main-chain atoms and $(0{\cdot}025\ \text{Å}^2)^{-2}$ for side-chain atoms bonded to the same atom (bond-angle restraint). This choice is in the range used in many applications of the refinement program.

Anisotropic restraints

To determine the anisotropy of the thermal fluctuations, the components of the anisotropy tensor are included as additional parameters in the least-squares refinement procedures. Without the isotropy assumption, six parameters are needed to specify the orientation of the thermal ellipsoid and the magnitudes of mean-squared atomic fluctuations along the principal axes. Based on a knowledge of the stereochemistry of a macromolecule, the orientation of the thermal ellipsoids can be chosen to be consistent with certain directions for maximum and minimum displacements. For proteins, a dictionary for determining the principal axes of the anisotropy tensor based on the amino acid sequence has been developed by Konnert & Hendrickson (1980). In general, the first axis direction, $\hat{c}_1$, is chosen to point along the bond from the preceding atom in the sequence. A second vector, $\hat{\mathbf{s}}$, is then introduced. For the main chain, $\hat{\mathbf{s}}$ points in some other direction along another bond involving the atom under consideration (see Fig. 1*a*) and the principal axis coordinate system is then assumed to be given by $\hat{\mathbf{c}}_3 = \hat{\mathbf{c}}_1 \times \hat{\mathbf{s}}$, $\hat{\mathbf{c}}_2 = \hat{\mathbf{c}}_3 \times \hat{\mathbf{c}}_1$. For the side chains, similar choices are made where possible; special designations are required for terminal and ring atoms, and some examples are given in Fig. 1(*b*) and (*c*).

Equations for calculating anisotropic variances with the riding-motion assumption have been given by Konnert & Hendrickson (1980). Following their notation, we approximate the joint one-dimensional variance component for atoms a and b along a direction v by the expression

$$V_{v,ab} = \Delta^2_{v,ab}\cos^2\theta + \Delta^4_{v,ab}[\sin^4(\theta/2) - 3\cos^2\theta\sin^2\theta]/d_0^2, \qquad (4)$$

where θ is the angle between the unit vector v and $\langle \mathbf{R} \rangle$, the vector from atom a to b in the average structure; d_0 is the magnitude of $\langle \mathbf{R} \rangle$ and

$$\Delta^2_{v,ab} = \langle u^2_{v,b} \rangle - \langle u^2_{v,a} \rangle \qquad (5)$$

if b is riding on a in direction v. Here $\langle u^2_{v,a} \rangle$ is the mean-squared displacement of atom a from its mean position along the direction v. This can be calculated from a knowledge of its anisotropic temperature factors $\beta_i^a (i = 1, 2, 3)$ expressed in terms of the local principal axis system that diagonalizes the anisotropy tensor (*i.e.*, $\beta_i^a = 2\pi^2 \langle u^2_{i,a} \rangle$). If the $\mathbf{d}_i^a$ are the unit vectors in this coordinate frame,

$$\langle u^2_{v,a} \rangle = \sum_{i=1}^{3} \beta_i^a (\mathbf{v} \cdot \mathbf{d}_i^a)^2/(2\pi^2). \qquad (6)$$

The variance used in the refinement program is then approximated by a sum over six such one-

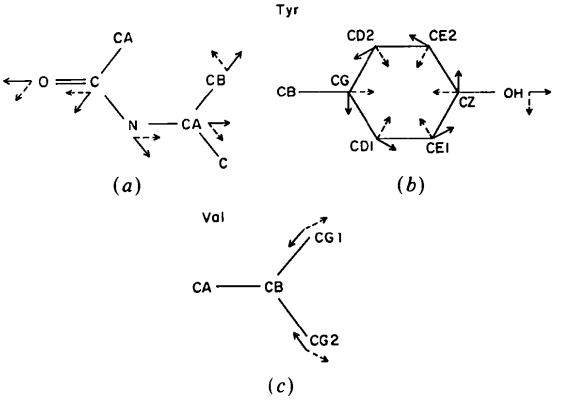


Fig. 1. Examples for choice of $\hat{c}_1$ (——→), $\hat{s}$ (----→) in the stereochemical dictionary for proteins. (*a*) Main-chain $\hat{s}$: O(CA----→C); side-chain $\hat{s}$: CB(C----→CA). (*b*) Tyrosine $\hat{c}_1$: CG(CD2——→CD1), CD1(CG——→CE1), CE1-(CD1——→CZ), CZ(CE1——→CE2), OH(CZ——→OH); $\hat{s}$: ring atoms (from the atom to the center of ring), OH-(CE2——→CE1). (*c*) Valine $\hat{s}$: CG1(CA----→CG1).

dimensional variances; that is

$$V'_{ab} = \sum_{v=1}^{6} V_{v,ab}/2, \qquad (7)$$

where the six directions v correspond to the three $\mathbf{d}_i^a (i = 1, 2, 3)$ and the three $\mathbf{d}_i^b (i = 1, 2, 3)$.

Equation (4) is derived by truncating a Maclaurin series expansion (Konnert & Hendrickson, 1980) and keeping only terms up to second order in $\Delta^2_{v,ab}$. This procedure is valid if the series converge rapidly. However, failure of the expansion can occur if convergence is slow and negative variances can result for certain values of θ. For the limiting case of $\Delta^2_{v,ab}$ equal to d_0^2, the series is ill behaved; negative values of $V_{v,ab}$ result for θ in the ranges $0 \le \theta \le 24°$, $81° \le \theta \le 99°$, and $156° \le \theta \le 180°$.

Comparisons of the results obtained with the riding-motion assumption [equation (5)] with those expected for uncorrelated motion can be made by using

$$\Delta^2_{v,ab} = \langle u^2_{v,a} \rangle + \langle u^2_{v,b} \rangle \qquad (8)$$

in equation (4).

Restrained refinement application

Stereochemically restrained refinement of the high-pH crystal form of BPTI was carried out as a test in the development of refinement methods. The diffraction data and starting model for this refinement were those from Deisenhofer & Steigemann (1975), as supplied by J. Deisenhofer. The unrestricted atomic radii parameters from the real-space refinement (Diamond, 1971) were converted to isotropic B values for the refinement in reciprocal space.

After 12 cycles of restrained refinement of this model, the distance variances as measured by equation (3) were substantially reduced while preserving stereochemical ideality (*e.g.*, r.m.s. deviations from ideal bond lengths went from 0·019 to 0·018 Å) and reducing the crystallographic R value from 0·208 to 0·188 for the 7976 data in the range of spacings between 7 and 1·5 Å. Reduction in average variances ranged from a decrease from 0·038 to 0·012 Å^2 for main-chain atoms to a decrease from 0·16 to 0·031 Å^2 for the angle-related (1-3) distances in side chains. Two rounds of manual revision and 16 further cycles of isotropic refinement produced the model used to initiate anisotropic refinement. The revisions were based on inspection of Fourier syntheses without the use of computer graphics. They involved rather minor changes to ten residues and the net addition of 34 water molecules and 12 protein atoms. The final model then comprises 454 non-hydrogen protein atoms and 81 water O atoms. This model does not allow for conformational heterogeneity (disorder).

The anisotropic refinement included only the 6894 data in the 7-1·5 Å range that had $|F| > 10{\cdot}75e$ ($\sim 2\sigma$, absolute scale). This refinement used procedures described above for thermal-parameter restraints (Konnert & Hendrickson, 1980) and included various other stereochemical restraints described by Hendrickson & Konnert (1980*b*). Thirteen refinement cycles reduced R from 0·159 to 0·143 while conformity with ideal geometry was kept essentially unchanged. The weighting factor used for each $V^2_{v,ab}$ term was $(0{\cdot}0033\ \text{Å}^2)^{-2}$ in the bond-length restraints and $(0{\cdot}0075\ \text{Å}^2)^{-2}$ in the bond-angle restraints. With these weights, the mean of variances [equation (7)] was reduced from 0·0094 to 0·0035 Å^2 for main-chain bonds and by a similar factor to 0·0049 Å^2 for side-chain bonds. The corresponding variances for (1-3) angle distances were not appreciably changed and the resulting means were 0·014 and 0·020 Å^2 for main-chain and side-chain atoms, respectively. The refinement statistics for the resulting anisotropic model are given in Table 1. This model was used for comparison with the molecular-dynamics simulation.

For comparison with the isotropic analysis of simulation results, the anisotropic thermal parameters of the model described above were converted to their isotropic equivalents and three additional cycles of isotropic refinement were performed. The R value was reduced from 0·155 to 0·152 in this refinement which led to the stereochemical statistics cited in Table 1. The isotropic thermal-restraint weightings were as described in an earlier section.

Molecular-dynamics simulation

The molecular-dynamics simulation of the bovine pancreatic trypsin inhibitor used in this work has been described previously (van Gunsteren & Karplus, 1982*a*). It treated the molecule as composed of 454 extended atoms plus four interior water molecules and included all the degrees of freedom of the extended atoms (*i.e.*, no bond lengths or angles were constrained); H atoms appear only implicitly as part of the extended atoms. The simulation had an equilibration period of 30 ps and then was continued for an analysis period of 25 ps; the mean temperature of the simulation was 300 K. A subset of 500 coordinates spaced at 0·05 ps were used for the present analysis. The r.m.s. deviations of the dynamics average structure from the X-ray structure are 2·28 Å for main-chain atoms, 3·14 Å for side-chain atoms and 2·73 Å overall.

Given the atomic positions as a function of time, it is straightforward to calculate the quantities that enter into the thermal-parameter refinement; the program *CHARMM* (Brooks, Bruccoleri, Olafson, States, Swaminathan & Karplus, 1983) developed at Harvard was used for this purpose. Specifically, it is possible to calculate the actual distance fluctuations between any pair of atoms from the trajectory. We have determined these for bonded atoms (*i.e.*, the

Table 1. *X-ray refinement statistics for BPTI*

Number of atoms	
Protein	454
Water	81
Number of reflections	
Total ($\infty - 1{\cdot}5$ Å)	8079
Resolution range ($7 - 1{\cdot}5$ Å)	6989
Significant ($7 - 1{\cdot}5$ Å, $\|F\| > \sim 2\sigma$)	6894

Deviations from ideality

	Isotropic		Anisotropic	
Feature	Weighting sigma*	R.m.s. deviation	Weighting sigma†	R.m.s. deviation
Bonding distances (Å)				
Bond length (1-2)	0·02	0·021		0·020
Angle-related distance (1-3)	0·03	0·035		0·034
Intraplanar distance (1-4)	0·04	0·038		0·036
Planar groups (Å)				
Deviation from plane	0·02	0·010		0·010
Chiral centers (Å^3)				
Chiral volume	0·15	0·181		0·181
Non-bonded contacts (Å)				
Single torsion	0·50	0·180		0·179
Multiple torsion	0·50	0·180		0·181
Possible hydrogen bond	0·50	0·246		0·242
Torsion angles (°)				
Planar (*e.g.* peptide ω)	2	3·9		3·9
Staggered (*e.g.* aliphatic χ_1)	15	15·9		16·6
Transverse (*e.g.* aromatic χ_2)	20	18·1		18·2
Thermal parameters (Å)				
Main-chain‡ bond fluctuation	0·113	0·104	0·058	0·059
Side-chain‡ bond fluctuation	0·138	0·143	0·058	0·070
Main-chain (1-3) fluctuation	0·138	0·129	0·144	0·121
Side-chain (1-3) fluctuation	0·159	0·176	0·144	0·142
R value:		0·152		0·143

* See Hendrickson & Konnert (1980*a*, *b*) for definitions.
† If blank, the anisotropic weight is as for the isotropic case.
‡ Main-chain atoms are N, C^α, C, O and side-chain atoms include all the rest of the atoms.

fluctuations in the bond lengths) and for the next-neighbor atoms. This result can be compared directly with the variances of these distances obtained by use of the riding-motion model in both isotropic and anisotropic refinements and the simulation. For the isotropic case, we consider V'_{ab} as defined in equation (3) with the assumption that the displacements are isotropic so that $\langle u^2 \rangle = \frac{1}{3}\langle r^2 \rangle$ for each atom. For comparison with the anisotropic refinement, V'_{ab} as defined by equation (7) is used.

To analyze the thermal-ellipsoid orientation of the anisotropic displacements, we compute the six components of the mean-square atomic displacement tensor for each atom in a molecule fixed coordinate system. Each atom anisotropy tensor can be diagonalized to obtain the principal components and their orientation in a local coordinate system. It is the relation of this local coordinate system to that assumed in the stereochemical dictionary which is examined. For this purpose, we define three coordinate systems and the rotation matrices to go from one to another. The anisotropy tensor for the atoms is determined from the simulation in the laboratory frame defined by the unit vectors, $\mathsf{L} = (\hat{\mathbf{l}}_1, \hat{\mathbf{l}}_2, \hat{\mathbf{l}}_3)$. For each atom a there are two other coordinate frames to consider; the first, $\mathsf{d}^a = (\mathbf{d}_1^a, \mathbf{d}_2^a, \mathbf{d}_3^a)$, is the local principal axis frame that diagonalizes the calculated anisotropy tensor (introduced above) and the second, $\mathsf{c}^a = (\mathbf{c}_1^a, \mathbf{c}_2^a, \mathbf{c}_3^a)$, is determined by the stereochemical dictionary. We introduce the transformation matrices D^a and C^a,

$$\mathsf{d}^a = \mathsf{D}^a\mathsf{L} \qquad \mathsf{c}^a = \mathsf{C}^a\mathsf{L} \tag{9}$$

with components given by the direction cosines $(\mathsf{D}^a)_{ij} = (\mathbf{d}_i^a \,.\, \hat{\mathbf{l}}_j)$ and $(\mathsf{C}^a)_{ij} = (\mathbf{c}_i^a \,.\, \hat{\mathbf{l}}_j)$. Finally we need the relation between c^a and d^a given by the matrix R^a, where

$$\mathsf{d}^a = \mathsf{R}^a\mathsf{c}^a \tag{10a}$$

and

$$\mathsf{R}^a = \mathsf{D}^a(\mathsf{C}^a)^{-1} \quad \text{or} \quad (\mathsf{R})_{ij}^a = (\mathbf{d}_i^a \,.\, \mathbf{c}_j^a) \equiv \cos\theta_{ij}^a. \tag{10b}$$

The deviation between the stereochemical frame c^a and the dynamic frame d^a for atom a is assessed by comparing the minimum, α^a, of the angles

$$\alpha^a = \text{minimum} \times |(\theta_{11}^a, \theta_{21}^a, \theta_{31}^a, \pi - \theta_{11}^a, \pi - \theta_{21}^a, \pi - \theta_{31}^a)|, \tag{11a}$$

where θ_{ij}^a is defined in equation (10*b*); see Fig. 2. Equation (11) takes into account the assumption made in the X-ray refinement that thermal motion is centrosymmetric. When $\mathbf{c}_1^a$ is chosen to be a bond vector (see Fig. 1), α^a is the value of the smallest angle that any of the dynamical principal axes makes with the bond vector. The angular deviation found in this way can be compared with the angle of 31·9° that corresponds to random relative orientation within the allowed range, $0 \le \alpha^a \le 54°44'$; the numerical value was obtained by evaluating an expression involving elliptic integrals (see Appendix).*

Alternatively, it is possible to examine whether the largest component, $\mathbf{d}_1^a$, of the anisotropy tensor is oriented along the vector $\mathbf{c}_3^a$, which where possible is defined to be perpendicular to the plane corresponding to two bonds in which atom a is involved (see

* The Appendix has been deposited with the British Library Lending Division as Supplementary Publication No. SUP 42088 (3pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

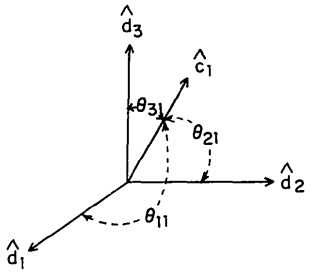


Fig. 2. Definition of angles determining the angles α and γ [see equations (11) in text].

above). For this case the angle, γ^a, is given by (see Fig. 2)

$$\gamma^a = \text{minimum}\,|\theta_{13}, \pi - \theta_{13}| \qquad (11b)$$

and a random orientation corresponds to 57·3°.

Results and discussion

In this section we present the results obtained and discuss their significance for evaluating the assumptions made in implementing the various refinement procedures.

Isotropic restraints

The results for the isotropic temperature-factor analysis of the bovine pancreatic trypsin inhibitor obtained from the simulation and the X-ray data are given in Table 2; part (*a*) is concerned with bonded atom pairs (1-2 pairs) and part (*b*) with atom pairs that are bonded to a common atom (1-3 pairs). The mean value of the variances and the standard deviation about the mean of the variance distributions are listed. Since differences between temperature factors are being considered, we neglect any disorder contributions in the experimental results on the assumption that the disorder contribution is approximately constant (Northrup *et al.*, 1980; Frauenfelder *et al.*, 1979). We consider the averages over all atoms and then separately those for the main-chain atoms (N, C^α, C, O) and side-chain atoms C^β, C^γ *etc.*). For both (1-2) and (1-3) pairs, it is clear that the distance fluctuations calculated from the trajectory are much smaller than those corresponding to the riding motion [equation (3)]; that is, the actual motions of (1-2) and (1-3) atom pairs are highly correlated due to the harmonic force constants associated with the bond lengths and the bond angles, as well as other interactions included in the empirical energy function (van Gunsteren & Karplus, 1982*a*). The riding-motion values are about a factor of 100 times larger than the 1-2 distance fluctuations and a factor of 10 times larger than the 1-3 distance fluctuations. Clearly, the riding-motion model is not valid in general for these atomic fluctuations in a protein. This is in accord with the analysis of the dynamics by Swaminathan *et al.* (1982), which have demonstrated the collective character of the motions. A riding-motion approximation might be applicable for an atom relative to the center of mass of the collective 'particle' to which it is attached, but not directly for atom pairs. As pointed out by Busing & Levy (1964), the riding-motion assumption is expected to be valid for a light particle 'riding' on a heavy atom (*e.g.*, an H attached to an O atom). For proteins, the atoms or extended atoms being studied have similar masses so that it is not surprising that the riding-motion assumption fails.

The corresponding analysis for anisotropic variance distributions is given in Table 3. The calculations for the riding-motion model use equations (4)-(7); for the uncorrelated motion distribution, equation (5) is replaced by equation (8). In the analysis of the dynamics results, the local principal axis systems d^a and d^b calculated from the dynamics were used in equation (6). In comparison with Table 2, there are 10 to 30% reductions in the dynamics values for the mean variances based on the riding-motion assumption in going from the isotropic to the anisotropic

Table 2. *Isotropic variance distribution for distances*

(*a*) Directly bonded pairs (1-2)

	Atom pairs	Mean (Å^2)	Standard deviation (Å^2)
Dynamics trajectory			
Exact	All	$6{\cdot}7\times10^{-4}$	$1{\cdot}5\times10^{-4}$
	Main chain	$6{\cdot}4\times10^{-4}$	$1{\cdot}3\times10^{-4}$
	Side chain	$6{\cdot}8\times10^{-4}$	$1{\cdot}8\times10^{-4}$
Riding motion [equation (3)]	All	0·040	0·067
	Main chain	0·020	0·028
	Side chain	0·067	0·095
X-ray refinement			
Riding motion [equation (3)]	All	0·015	0·014
	Main chain	0·011	0·009 (0·013)*
	Side chain	0·020	0·018 (0·019)*

(*b*) Next-nearest-neighbor pairs (1-3)

	Atom pairs	Mean (Å^2)	Standard deviation (Å^2)
Dynamics trajectory			
Exact	All	$3{\cdot}4\times10^{-3}$	$1{\cdot}8\times10^{-3}$
	Main chain	$2{\cdot}6\times10^{-3}$	$0{\cdot}6\times10^{-3}$
	Side chain	$3{\cdot}6\times10^{-3}$	$2{\cdot}6\times10^{-3}$
Riding motion	All	0·054	0·084
	Main chain	0·027	0·036
	Side chain	0·094	0·134
X-ray refinement			
Riding motion	All	0·024	0·024
	Main chain	0·017	0·014 (0·019)*
	Side chain	0·031	0·030 (0·025)*

* Numbers in parentheses give the square-root of the weighting factors $[(1/w_l)^{1/2}]$ used in equation (1) for the X-ray refinement.

Table 3. *Anisotropic variance distribution for interatomic distances*

(*a*) Directly bonded pairs (1-2)

	Atom pairs	Mean (Å^2)	Standard deviation (Å^2)
Dynamics trajectory			
Uncorrelated motion	All	0·186	0·316
[equation (5)]	Main chain	0·142	0·083
	Side chain	0·246	0·496
Riding motion	All	0·031	0·079
[equation (7)]	Main chain	0·014	0·016
	Side chain	0·056	0·122
X-ray refinement			
Riding motion	All	0·0044	0·0031
[equation (7)]	Main chain	0·0035	0·0023 (0·0033)*
	Side chain	0·0049	0·0034 (0·0033)*

(*b*) Next-nearest-neighbor pairs (1-3)

	Atom pairs	Mean (Å^2)	Standard deviation (Å^2)
Dynamics trajectory			
Uncorrelated motion	All	0·230	0·230
	Main chain	0·159	0·082
	Side chain	0·382	0·364
Riding motion	All	0·042	0·065
	Main chain	0·019	0·021
	Side chain	0·087	0·105
X-ray refinement			
Riding motion	All	0·018	0·016
	Main chain	0·014	0·011 (0·0075)*
	Side chain	0·020	0·023 (0·0075)*

* *See footnote in Table 2.*

analysis. The relative magnitudes in the various categories of dynamics results are similar to those found from the isotropic treatment.

The analysis of the dynamics shows unequivocally that the difference between the temperature factors of (1-2) or (1-3) atom pairs cannot be identified with the variance of their interatomic-distance distributions. In fact, a corresponding analysis of a trajectory of BPTI with *fixed* bond lengths (van Gunsteren & Karplus, 1982*b*) shows essentially the same distribution of temperature factors for both (1-2) and (1-3) pairs. This confirms that, with the conformational flexibility available to protein, it is not meaningful to extract information on bond-length fluctuations from the relative X-ray temperature factors of (1-2) atom pairs. The existence of small bond-length fluctuations does not arise from equal-magnitude fluctuations for the two atoms involved; in fact, the two are not related in any simple way. Instead what is important is that the motions of the atoms be highly correlated.

The above results do *not* invalidate the temperature-factor refinement procedure of Konnert & Hendrickson (1980). For the analysis of the X-ray data, it is not the correctness of the riding-motion model that is at issue, but rather whether or not the restraints applied in the refinement process are appropriate. To focus on this point, we compare the values for the variances, V'_{ab}, obtained from the dynamics and the X-ray analysis of BPTI. As shown in Table 2 for the isotropic case, the results found from the dynamics are similar to those from the X-ray refinement for both (1-2) and (1-3) neighbor pairs. For the (1-2) and (1-3) pairs, the calculated mean variance for all atoms is about twice the X-ray estimate with the chosen restraints. Examining the difference between main-chain and side-chain atom pairs, we see that the dynamics yields a difference in standard deviations of a factor of three for both (1-2) and (1-3) neighbors, while the refinement employs restraints that differ by a factor of only 1·5; the simulation results for pairs where one is a main-chain atom and the other a side-chain atom are comparable to those where both atoms are part of the main chain.

Figs. 3 and 4 show a comparison of the distributions for V'_{ab} of (1-2) and (1-3) pairs obtained from the restrained isotropic refinement (Figs. 3*a* and 4*a*) and the simulation (Figs. 3*b* and 4*b*). The overall distributions are rather similar, but there are larger differences, as expected from the results in Table 2, for the separate main-chain and side-chain distributions. As is evident from the figures and the standard deviations for the variances listed in Table 2, the dynamics distribution is somewhat broader than that obtained from the refinement. For the (1-3) pairs, the largest variances from the simulation correspond primarily to charged side chains (Lys 26, Glu 49, Arg 17, 39 and

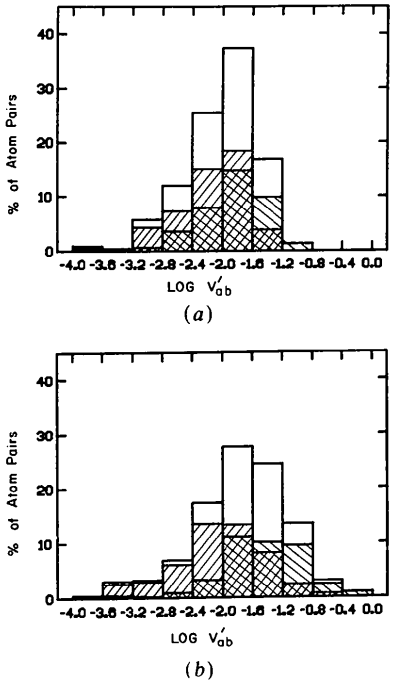


Fig. 3. Isotropic variance distribution, V'_{ab}, for interatomic distances of directly bonded atom pairs (1-2) in BPTI [equation (3)] (there are 468 such pairs); the logarithm of V'_{ab} is plotted because of the wide range of values. (*a*) X-ray; (*b*) dynamics. The all-atom results correspond to the height of the white bars, the main chain to the height of the portion marked (▨), and the side chain to the portion marked (▧).

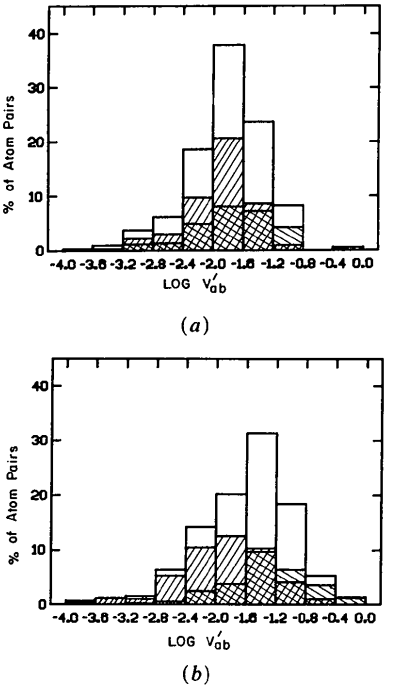


Fig. 4. Isotropic variance distribution, V'_{ab}, of interatomic distances of next-nearest-neighbor pairs (1-3) in BPTI (there are 630 such pairs). (*a*) X-ray; (*b*) dynamics; for description see legend of Fig. 3.

53), though Leu 6 and 29 and Ile 19 also have large variances. Many of these are due to dihedral-angle transitions (van Gunsteren & Karplus, 1982*a*), which are not expected to be adequately sampled in the present trajectory. In the isotropic X-ray refinement, very large side-chain variances do not occur, presumably due, at least in part, to the use of average restraints with narrow widths.

Anisotropic restraints

Similar conclusions can be drawn from Table 3 and Figs. 5 and 6 where the anisotropic refinement is compared to dynamics simulation data. The need for different weightings for main-chain and side-chain atom pairs is very clear in this case. Moreover, comparing the X-ray with the dynamics variances, it is apparent that the great reduction in the values assumed for the refinement variances (relative to the isotropic case) is not justified. This is in accord with the results of the X-ray refinement where the (1-3) pairs have variances two to three times the assumed restraints (*i.e.*, 0·014 and 0·020 Å^2 compared with the restraint 0·0075 Å^2).

The above analyses suggest that restrained refinement procedures for temperature factors can yield meaningful results for proteins and other macromolecules. Obviously it is necessary to choose reasonable values for the assumed variances; these could be estimated from dynamical simulations. It would be interesting to have the results of an unrestrained refinement for comparison. This is particularly true for directly bonded atoms (1-2 pairs), where the simulation indicates that it is appropriate to introduce different weights for main-chain and side-chain atom pairs, while for (1-3) main-chain pairs, restraints may also be useful. For the (1-3) side-chain pairs the simulation results suggest that the restraints are of doubtful significance. In fact, the isotropic distribution for (1-3) side-chain pairs is similar to that found for arbitrary atom pairs in the protein separated by 9 to 12 Å.

Orientation and anisotropy

For an analysis of the orientational aspects of the anisotropic refinement of BPTI, we compare the local principal axis system for the anisotropy tensor of a given atom with that from the stereochemical dictionary results. Figs. 7(*a*) and 7(*b*) show the average value of α [equation (11*a*)] and γ [equation (11*b*)] as a function of the atom type. It can be seen that for all the atoms and for most of the atom types the values of α and γ are close to, but somewhat smaller than, those expected for a random relative orientation of the two coordinate systems ($\alpha_{random} = 31{\cdot}9°$; $\gamma_{random} = 57{\cdot}3°$). Only for the main-chain carbonyl oxygen (O) does the local principal axis system orientation appear to be significantly correlated with the bonding. For the carbonyl O it has been noted previously (van

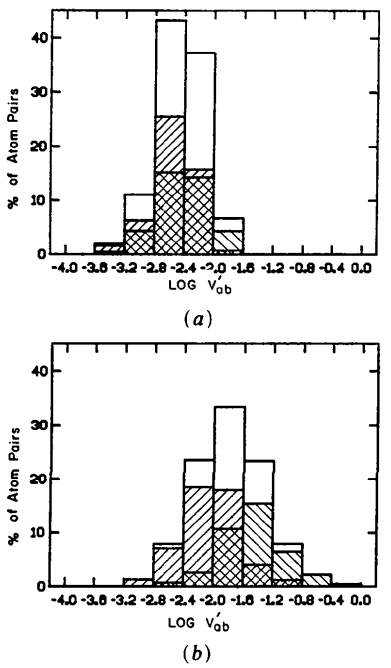


Fig. 5. Anisotropic variance distribution, V'_{ab}, for interatomic distance of directly bonded atom pairs (1-2) in BPTI [equation (7)]; there are 467 such pairs excluding C^{γ}-$C^{\delta 1}$ of Leu 29 which has a negative variance calculated by equation (4). (*a*) X-ray; (*b*) dynamics; for description see legend of Fig. 3.

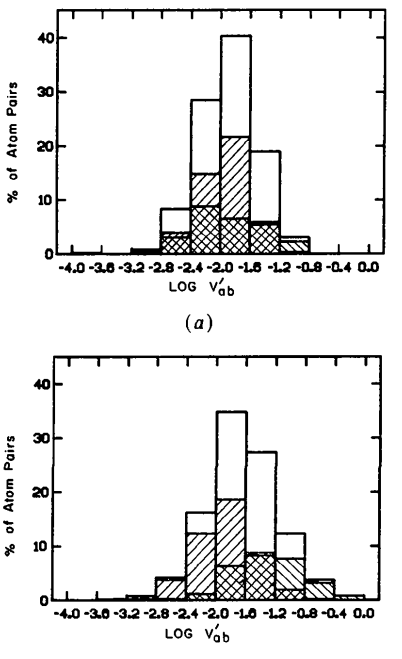


Fig. 6. Anisotropic variance distribution, V'_{ab}, for interatomic distances of next-nearest-neighbor pairs (1-3) in BPTI (there are 630 such pairs). (*a*) X-ray; (*b*) dynamics; for description see legend of Fig. 3.

Gunsteren & Karplus, 1982*a*; Mao *et al.*, 1982) that the dominant motion is perpendicular to the C=O bond.

If the anisotropy were simply related to the bonding, the principal component of the anisotropy tensor directed along a bond to the atom of interest would have the smallest value. To examine this possibility we show in Fig. 8 the percentage of the atoms of a given type that have α as the angle between $\hat{\mathbf{c}}_1$ (in general, the bond vector) and $\hat{\mathbf{d}}_3$, the principal axis with the smallest eigenvalue. A value of 33·3% corresponds to atomic displacements that are, on the average, random with respect to the chemical bonds. Main-chain N and carbonyl C atoms have values close to random, while the results for other atoms do suggest that the fluctuations are somewhat larger in directions perpendicular to chemical bonds, in accord with the values of γ. The carbonyl O has the highest percentage value, 75%, a result consistent with the out-of-plane motion mentioned above.

A simple measure of the anisotropy of the atomic fluctuations is provided by the quantity A defined as

$$A = (\langle \Delta d_1^2 \rangle)^{1/2} / [(\langle \Delta d_2^2 \rangle + \langle \Delta d_3^2 \rangle)/2]^{1/2} - 1, \quad (12)$$

where Δd_1, Δd_2, and Δd_3 are the eigenvalues of the anisotropy tensor in order of decreasing magnitude; a value of zero corresponds to isotropic motion. In the d frame, the anisotropic X-ray refinement gives A in the range 0·004 to 0·63 with mean 0·12, whereas the dynamics simulation gives A in the range from 0·08 to 2·62 with mean 0·73. Use of the stereochemical c frame to analyze the dynamics reduces the mean value to 0·41.

Fig. 9 gives the anisotropy for the C^α atoms and carbonyl O atoms as a function of residue number. For the C^α atoms, we see that the exact calculated anisotropy (dashed line) based on equation (12) is significantly larger than that (solid line) obtained by using the stereochemical values from the simulation; the latter are obtained by orienting the calculated anisotropy tensor along the ĉ coordinate system and then using the diagonal elements in decreasing order (Δc_1, Δc_2, and Δc_3) instead of Δd_1, Δd_2, Δd_3, respectively, in equation (12). Similar results are found for the carbonyl O, except that the calculated anisotropy values tend to be larger and that the Δd_i's and Δc_i's

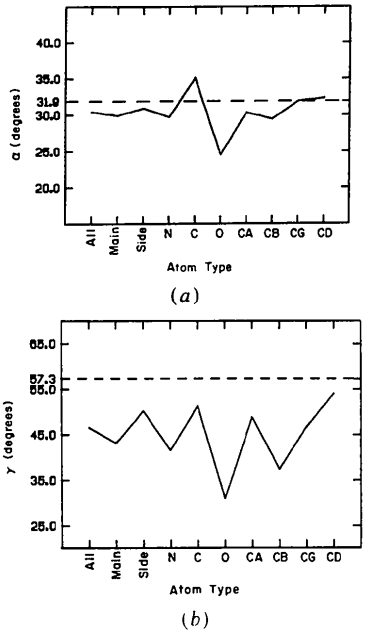


Fig. 7. Values of angle giving the orientation of the stereochemical frame with respect to the local principal axis frame from dynamics for the anisotropy tensor as a function of atom types; a random orientation corresponds to the dotted line (see text): (*a*) angle α [equation (11*a*)]; (*b*) angle γ [equation (11*b*)]. Atom types CG, CD include all atoms in γ and δ positions, respectively.

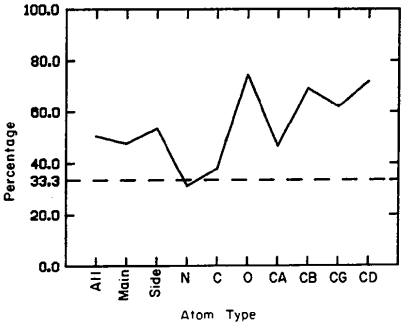


Fig. 8. Percentage of atoms for which the orientation of smallest dynamic displacement axis ($\hat{d}_3$) makes the smallest angle (α) with the bond vector ($\hat{c}_1$). Atom types CG, CD include all atoms in γ and δ positions, respectively.

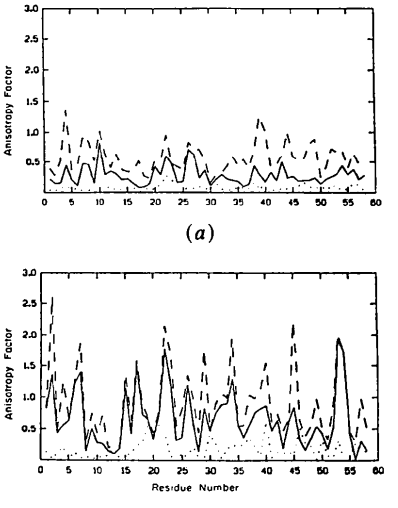
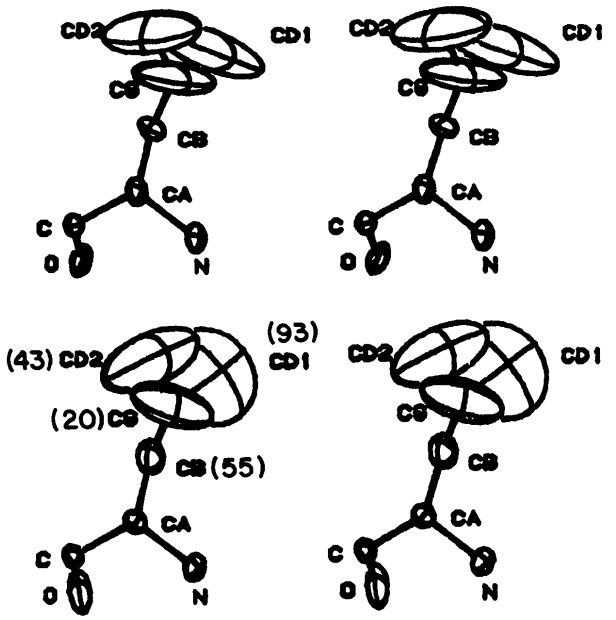


Fig. 9. Anisotropy of fluctuations of (*a*) C^α and (*b*) carbonyl O: (- - -) dynamics anisotropy calculated in local principal axis frame; (——) dynamics anisotropy calculated in stereochemical frame; (· · ·) X-ray anisotropy based on refinement in stereochemical frame.

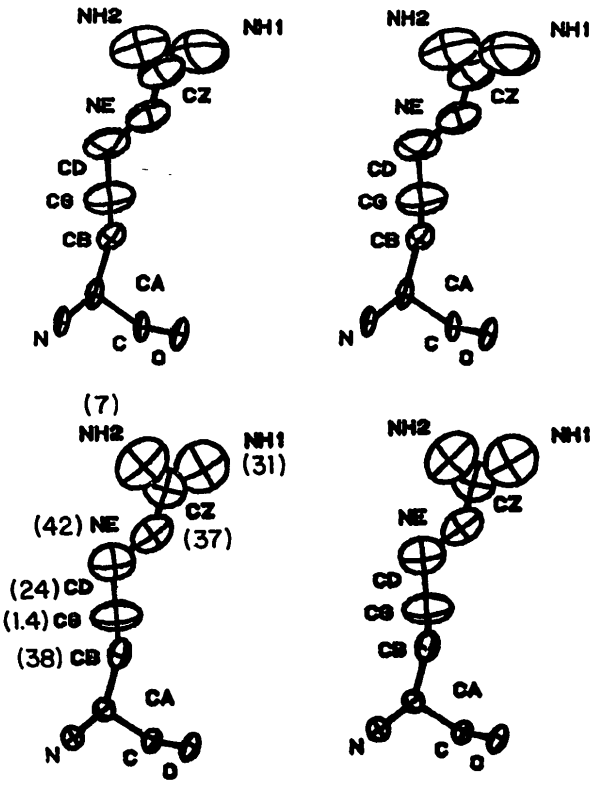
resulting from the simulation are more similar to each other than for the C^{α} atoms. This is because for the carbonyl O atoms the anisotropy tensor orientation obtained from the stereochemical model is relatively close to the true principal axis system. Fig. 9 also compares the calculated A values with the A values obtained (dotted line) from the refinement. We see that the latter are still smaller for both the C^{α} atoms and carbonyl O. At least part of this difference is likely to be due to the restraints, although approximations in the simulation may also be involved.

Fig. 10 shows *ORTEP* drawings of thermal ellipsoids from the simulation for selected residues with large side-chain motions in both the c and the d coordinate frames. Numbers in parentheses give the percentage reduction in the anisotropy [equation (12)] obtained in the c frame relative to that in the d frame. The significantly more isotropic behavior generally observed in the c frame is evident; a particularly striking case involves the $C^{\delta 1}$ atom of Leu 29. Also, the difference in orientation of many of the thermal ellipsoids in the two frames is clear. Atom C^{γ} of Arg 39 is an exception where the agreement is excellent with only 1·4% reduction. For C^{β} of Asn 44, the direction of the smallest eigenvector ($\hat{\mathbf{d}}_3$) coincides well with $\hat{\mathbf{c}}_1$ along the C^{α}–C^{β} bond; the value of α as given by equation (11) is 7·7°, comparable to $\alpha = 8°$ for C^{γ} of Arg 39. However, mixing of the other two components (*i.e.*, $\hat{\mathbf{d}}_1$ makes angles of 124 and 34·5° with $\hat{\mathbf{c}}_2$ and $\hat{\mathbf{c}}_3$, respectively, compared to 88 and 5·8° for C^{γ} of Arg 39) gives a 61% reduction in the anisotropy.

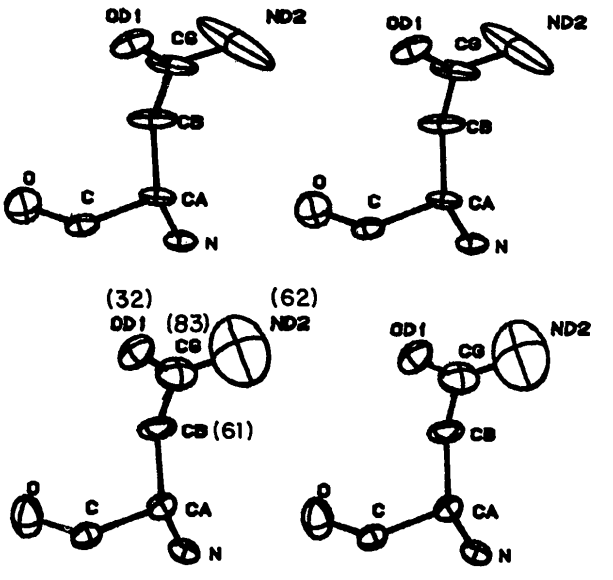


(*a*)

Fig. 10. *ORTEP* drawings (Johnson, 1965) for thermal ellipsoids of selected residues in the dynamic and stereochemical principal axis frames; only comparison within the same residue is meaningful since different scaling factors have been used for different residues: (*a*) Leu 29; (*b*) Arg 39; (*c*) Asn 44. In each case the upper drawing refers to the dynamic d frame and the lower to the stereochemical frame c. Numbers in parentheses give percentage reduction in anisotropy [equation (12)] obtained from frame c relative to frame d.



(*b*)



(*c*)

Fig. 10. (*cont.*)

Summary and cautions

The atomic fluctuations obtained from a dynamics simulation of the bovine pancreatic trypsin inhibitor have been used to test the application of temperature-factor restraints in the X-ray refinement of protein structures. Although the riding-motion model on which the restraints are based is found to be invalid, the general behavior of the difference in the fluctuations for bonded- and next-neighbor atom pairs obtained from the simulation is consistent in order of magnitude with the restraints used in the isotropic refinement procedure. Quantitatively, the calculated variances are somewhat larger than the X-ray values, suggesting that less restrictive temperature-factor restraints should be used. For the anisotropic refinement the variances assumed are clearly too small. The simulation results indicate that different weights are appropriate for restraining the temperature factors for main-chain and side-chain atoms. Restraint of (1-4) atom pairs, which has been suggested, appears unnecessary since even (1-3) side-chain pairs have variances that are comparable with those for pairs of distant atoms.

The use of local principal frames for anisotropic temperature factors based on knowledge of geometry alone does not provide thermal-ellipsoid orientations in accord with the dynamics simulation. The result of a refinement with incorrectly oriented ellipsoids is a reduction in the apparent anisotropy. An alternative would be to perform X-ray refinements with the local principal frames obtained from molecular-dynamics simulation. This could serve as a test of the dynamics and may simplify the introduction of anisotropic temperature factors into the refinement of proteins and other macromolecules.

The above conclusions are tempered by the fact that the molecular-dynamics simulation on which they are based involves a variety of approximations (*e.g.*, empirical potential function, neglect of solvent and crystal environment, finite time truncation). This means that some of the detailed quantitative results are likely to be in error, but the qualitative conclusions should be valid. It is hoped that the present analysis is just a first step in future interactions between dynamical simulations and X-ray studies of macromolecules.

We wish to thank B. Brooks, R. Bruccoleri, T. Ichiye, D. States and S. Swaminathan for helpful discussion and assistance with the *CHARMM* program. We also thank J. Deisenhofer for supplying the diffraction data and starting model for the X-ray refinement.

References

ARTYMIUK, P. J., BLAKE, C. C. F., GRACE, D. E. P., OATLEY, S. J., PHILLIPS, D. C. & STERNBERG, M. J. E. (1979). *Nature (London)*, **280**, 563-570.

BROOKS, B. R., BRUCCOLERI, R. E., OLAFSON, B. D., STATES, D. J., SWAMINATHAN, S. & KARPLUS, M. (1983). *J. Comput. Chem.* **4**, 187-217.

BUSING, W. R. & LEVY, H. A. (1964). *Acta Cryst.* **17**, 142-146.

DEISENHOFER, J. O. & STEIGEMANN, W. R. (1975). *Acta Cryst.* B**31**, 238-250.

DIAMOND, R. (1966). *Acta Cryst.* **21**, 253-266.

DIAMOND, R. (1971). *Acta Cryst.* A**27**, 436-452.

FRAUENFELDER, H., PETSKO, G. A. & TSERNOGLOU, D. (1979). *Nature (London)*, **280**, 558-563.

GUNSTEREN, W. VAN & KARPLUS, M. (1982*a*). *Macromolecules*, **15**, 1528-1544.

GUNSTEREN, W. VAN & KARPLUS, M. (1982*b*). *Biochemistry*, **21**, 2254-2274.

HENDRICKSON, W. A. & KONNERT, J. H. (1980*a*). *Biophys. J.* **32**, 645-647.

HENDRICKSON, W. A. & KONNERT, J. H. (1980*b*). *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 13.01-13.23. Bangalore: Indian Academy of Sciences.

ICHIYE, T., OLAFSON, B. & KARPLUS, M. (1985). To be published.

JOHNSON, C. K. (1965). *ORTEP.* Report ORNL-3794. Oak Ridge National Laboratory, Tennessee.

KARPLUS, M. & MCCAMMON, J. A. (1981). *Crit. Rev. Biochem.* **9**, 293-349.

KARPLUS, M. & MCCAMMON, J. A. (1983). *Annu. Rev. Biochem.* **53**, 263-300.

KONNERT, J. H. (1976). *Acta Cryst.* A**32**, 614-617.

KONNERT, J. H. & HENDRICKSON, W. A. (1978). *Acta Cryst.* A**34**, S47.

KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A**36**, 344-350.

MAO, B., PEAR, M. R., MCCAMMON, J. A. & NORTHRUP, S. H. (1982). *Biopolymers*, **21**, 1979-1989.

MORGAN, J. D. & MCCAMMON, J. A. (1983). *Biopolymers*, **22**, 1579-1593.

NORTHRUP, S. H., PEAR, M. R., MCCAMMON, J. A., KARPLUS, M. & TAKANO, T. (1980). *Nature (London)*, **287**, 659-660.

NORTHRUP, S. H., PEAR, M. R., MORGAN, J. D., MCCAMMON, J. A. & KARPLUS, M. (1981). *J. Mol. Biol.* **153**, 1087-1109.

SWAMINATHAN, S., ICHIYE, T., VAN GUNSTEREN, W. & KARPLUS, M. (1982). *Biochemistry*, **21**, 5230-5240.

TAKANO, T. & DICKERSON, R. E. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 6371-6375.