PROTEINS: Structure, Function, and Genetics 2:236–259 (1987)

Anisotropy and Anharmonicity of Atomic Fluctuations in Proteins: Analysis of a Molecular Dynamics Simulation

Toshiko Ichiye and Martin Karplus
Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

ABSTRACT **Positional probability density functions (pdf) for the atomic fluctuations are determined from a molecular dynamics simulation for hen egg-white lysozyme. Most atoms are found to have motions that are highly anisotropic but only slightly anharmonic. The largest deviations from harmonic motion are in the direction of the largest rms fluctuations in the local principal axis frame. Backbone atoms tend to be more nearly harmonic than sidechain atoms. The atoms with the largest anharmonicities tend to have pdfs with multiple peaks, each of which is close to harmonic. Several model pdfs are evaluated on the basis of how well they fit probability densities from the dynamics simulations when parameterized in terms of the moments of the distribution. Gram-Charlier and Edgeworth perturbation expansions, which have been successful in describing the motions of small molecules in crystals, are shown to be inadequate for the distributions found in the dynamics of proteins. Multipeaked distribution functions are found to be more appropriate.**

Key words: atomic probability distributions of proteins, anisotropy and anharmonicity of motions in proteins, multiple occupancy of atomic positions in proteins, molecular dynamics simulation, lysozyme

INTRODUCTION

A knowledge of the magnitudes, functional forms, and time scales of the atomic fluctuations in proteins is essential for a characterization of the internal motions that are thought to play an important role in their biological activity.[1] A variety of experimental[2–5] and theoretical methods[1,5,6] are currently being used to examine the dynamical properties of proteins. Although the magnitude and time course of the subnanosecond motions have been determined by experimental and theoretical studies, the functional form of the distribution of atomic fluctuations is less well characterized. It is, however, important to have a knowledge of the form of the distribution functions, in particular because of their role in the interpretation of certain experiments. Of greatest interest are the structural analyses of protein crystals by X-ray diffraction,[7] which in most cases have assumed that the atomic motions are harmonic and isotropic. Correspondingly, inelastic neutron-scattering analyses of macromolecules are based on the harmonic assumption, although isotropic motion is not required.[8] Moreover, certain theoretical models for protein dynamics (e.g., those based on a normal mode analysis of the protein treated as an isomorphous solid[9,10] or in terms of detailed potentials[11,12]) assume harmonic, but not isotropic, motions.

The simplest motional model assumes that each atom moves in a local potential of mean force which is isotropic and harmonic. The atomic fluctuations then obey a three-dimensional isotropic Gaussian distribution. Some work has been done to determine the deviations of the atomic fluctuations in proteins from the isotropic harmonic limit. The anisotropy has been analyzed in a number of theoretical studies[13–16] and in a few X-ray studies.[17,18] Theoretical studies have shown that particularly for sidechain atoms, the distributions tend to be highly anisotropic. Further, it has been demonstrated that the orientation of the anisotropy tensors found in molecular dynamics simulations do not have a simple relation to the static structure.[19] A comparison of molecular dynamics and high-resolution X-ray results has been made for the avian pancreatic polypeptide.[20]

Based on a pioneering temperature-dependent crystallographic analysis of myoglobin, Frauenfelder et al.[21] suggested that most of the local potentials are highly anharmonic and attempted to relate the degree of anharmonicity with the fluid- or solidlike behavior of different portions of the protein.[22] In a theoretical study, Levy et al.[11] found significant anharmonic effects in the analysis of the temperature dependence of the atomic fluctuations of an α-helix obtained from a molecular dynamics simulation. It was suggested that two types of limiting models could explain the results: the first ascribes the calculated effects to a temperature-independent anharmonic potential, in accord with the analysis of Frauenfelder et al.,[21] and the second to a harmonic potential whose force constant depends on the temperature, (i.e., a quasi-harmonic model).[23,24] The latter model has been suggested for the evaluation of the configurational

Received May 14, 1986; accepted May 13, 1987.

Address reprint requests to Dr. Martin Karplus, Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, MA 02138.

Toshiko Ichiye's present address is Department of Chemistry, University of California, Berkeley, CA 94720.

entropy contributed by atomic fluctuations in peptides and proteins.[25] Molecular dynamics simulations of proteins indicate that some anharmonicity is present in the atomic fluctuations, as measured by the third and fourth moments of the distributions.[15,16,26]

The results that we have outlined make clear that it would be of great interest to have a detailed study of the distribution functions associated with the atomic fluctuations in proteins. In this paper we determine the nature of the deviations from isotropic, harmonic probability distributions by use of a molecular dynamics simulation. The molecular dynamics technique, although subject to statistical errors, allows a direct comparison of model probability density functions with histograms of coordinates from a trajectory, whereas crystallographic studies must rely on a model to interpret the data. The simulation used is one for lysozyme.[27] We examine the distributions of fluctuations of individual atoms as well as measures of anisotropy and anharmonicity based on ratios of the second through fourth moments. An important element in the analysis is the distinction between deviations from harmonicity due to anharmonicity of a single well and that due to the occupancy of several harmonic wells. The effects of neglect of anisotropic and anharmonic contributions on X-ray crystallographic refinement of atomic positions and thermal parameters in proteins are discussed in another paper.[28]

In this paper measures of anisotropy and anharmonicity are defined and several model probability density functions are introduced. The methods used in the simulation and analysis are described briefly. The degree and type of anisotropy and anharmonicity present in distributions from molecular dynamics simulations of proteins are examined. It is determined how well model probability density functions describe these distributions when parameterized in terms of their lower-order moments. A concluding discussion is followed by an appendix that considers what values of the third and fourth moments are associated with significantly anharmonic motion.

PROBABILITY DENSITY FUNCTIONS

The probability density function (pdf) as a function of the coordinate **r** at temperature T for an atom in a many-body system such as a protein is referred to as p(**r**,T). It is determined by the effective potential energy (potential of mean force), w(**r**,T), for that atom from the relation[29]

$$p(\mathbf{r},T) = \frac{e^{-\beta w(\mathbf{r},T)}}{\int e^{-\beta w(\mathbf{r},T)} d\mathbf{r}} \qquad (1)$$

in which β is equal to $1/k_BT$ and k_B is the Boltzmann constant. The potential of mean force is, as indicated, a function of the temperature; to simplify notation, we drop the reference to T in what follows, although simulation studies have indicated that the temperature dependence is likely to be important for proteins.[11]

Although the potential energy for a given atom is expressed as a sum of terms,[30] some of which are anharmonic, a common assumption is that the potential, and therefore the pdf, can be treated as harmonic in the neighborhood of the minimum. From a simulation, we can examine this assumption by calculating the pdf for each atom. Alternatively, since the pdf is determined from a knowledge of its moments,

$$\langle x^{n_1} y^{n_2} z^{n_3} \rangle = \int_{n_1,n_2,n_3 = 0,1,\ldots,\infty} x^{n_1} y^{n_2} z^{n_3} \, p(\mathbf{r}) \, d\mathbf{r}, \qquad (2)$$

where x,y,z are Cartesian components of **r**, the moments can be used to describe the probability distribution and therefore the shape of the local conformational potential by means of Eq. 1.

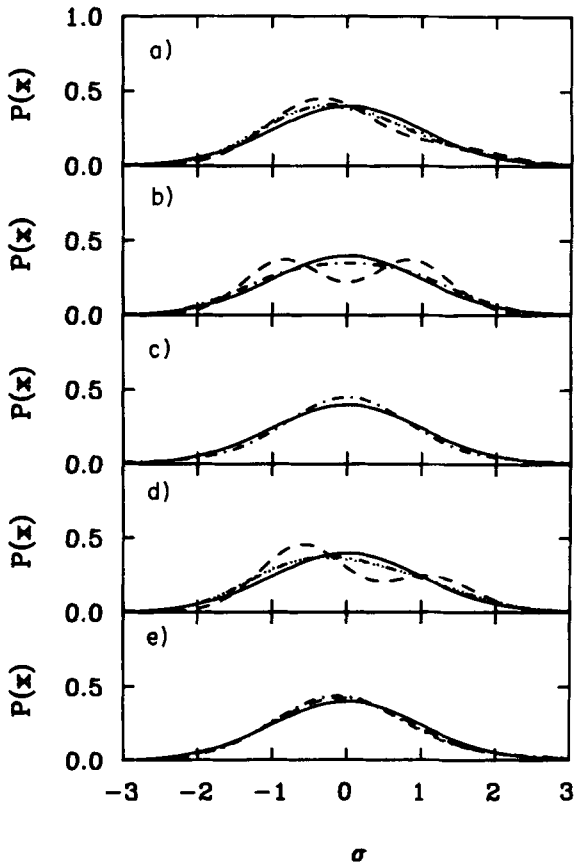

Fig. 1. Examples of anharmonic distributions with various values of skewness, α_3, and kurtosis, α_4. **a**: $\alpha_3 = 1/2$, $\alpha_4 = 0$. **b**: $\alpha_3 = 0$, $\alpha_4 = -1$. **c**: $\alpha_3 = 0$, $\alpha_4 = 1$. **d**: $\alpha_3 = 2^{-3/2}$, $\alpha_4 = -2^{-1/2}$. **e**: $\alpha_3 = 2^{-3/2}$, $\alpha_4 = 2^{-1/2}$. The distributions are given by the Gram-Charlier (—·—), Edgeworth (···), and double peak (– –); the Gram-Charlier and Edgeworth distributions are essentially identical in all cases shown. For reference, the Gaussian with the same value of σ is given by the solid line.

Anharmonicity

The major features of a pdf may be simply characterized by the second through fourth moments of the fluctuations about the mean position; they are denoted by $\mathbf{u} = (u_x,u_y,u_z)$, where $\mathbf{u} = \mathbf{r}-\langle\mathbf{r}\rangle$. For analysis of the dynamics results, the fluctuations are rotated to a local principal axis system, with coordinates $\mathbf{U} = (U_x,U_y,U_z)$ with $|\mathbf{U}| = |\mathbf{u}| = u_r$, in which U_x is defined as being along the direction of the largest mean-square displacement of the atom and U_z along the direction of the smallest. In this system, the covariances $\langle U_iU_j\rangle$, $i\neq j$ (but not higher-order cross-terms) equal zero, so the pdf is characterized by the magnitude of the fluctuations, measured by the second moments

$$\sigma_i^2 = \langle U_i^2\rangle, \tag{3a}$$

the asymmetry, measured by the coefficients of skewness,

$$\alpha_{3i} = \frac{\langle U_i^3\rangle}{\langle U_i^2\rangle^{3/2}} = \frac{\kappa_{iii}}{\sigma_i^3}, \tag{3b}$$

and the peakedness, measured by the coefficients of excess kurtosis,

$$\alpha_{4i} = \frac{\langle U_i^4\rangle}{\langle U_i^2\rangle^2} - 3 = \frac{\kappa_{iiii}}{\sigma_i^4} \tag{3c}$$

Here κ_{iii} and κ_{iiii} are third and fourth cumulants.[31] A pdf with a long "tail" on the positive side has positive skewness (Fig. 1a). A pdf with a flat peak has negative kurtosis (Fig. 1b) while one with a sharp peak and long "tails" has positive kurtosis (Fig. 1c). If a given Cartesian component i of the potential is harmonic, $\alpha_{3i} = \alpha_{4i} = 0$. Higher-order measures that are zero for the harmonic case can be defined analogously in terms of higher cumulants; i.e.,

$$\alpha_n = \kappa_n/\sigma^n \tag{4}$$

In the three-dimensional case, measures of anharmonicity independent of anisotropy require examination of the individual components U_x,U_y,U_z rather than u_r because only the inequality

$$\frac{\langle u_r^4\rangle}{\langle u_r^2\rangle^2} > \frac{5}{3}$$

holds in general for a distribution that is harmonic but anisotropic[32]; in the isotropic case,

$$\frac{\langle u_r^4\rangle}{\langle u_r^2\rangle^2} = \frac{5}{3}.$$

Anisotropy

The anisotropy of the distribution can be parameterized in several ways. We consider two quantities. The first, A_1,[19]

$$A_1 = \left[\frac{\langle U_x^2\rangle}{\frac{1}{2}(\langle U_y^2\rangle + \langle U_z^2\rangle)}\right]^{1/2} - 1, \tag{5a}$$

gives the fractional amount by which $\langle U_x^2\rangle$ is larger than the fluctuations along the other two directions; A_1 is 0 for isotropic motion. The second measure, A_2,

$$A_2 = \left[\frac{\langle U_y^2\rangle}{\frac{1}{2}(\langle U_y^2\rangle + \langle U_z^2\rangle)}\right]^{1/2} - 1, \tag{5b}$$

is defined similarly. It describes to what extent the fluctuations may be considered equal in magnitude along the two smaller principal axis directions, U_y and U_z.

The measures of anisotropy used here are related to those of Northrup et al.,[14] $f_2 = \langle U_y^2\rangle/\langle U_x^2\rangle$ and $f_3 = \langle U_z^2\rangle/\langle U_x^2\rangle$, by $A_1 = [\frac{1}{2}(f_2+f_3)]^{-1/2} - 1$ and $A_2 = [f_2/\frac{1}{2}(f_2+f_3)]^{1/2} - 1$. In van Gunsteren and Karplus,[15,16] the average value of $\langle U_x^2\rangle^{1/2}$ for various groups of atoms was divided by that of $\langle U_z^2\rangle^{1/2}$.

Model Probability Density Functions

Model pdfs parameterized by the moments of the distribution can be used to describe the functional form of the actual distribution. We consider several model pdfs. All are normalized so that the total integrated probability is equal to one. The argument of each pdf includes all of the parameters needed to determine the distribution; therefore, the mean position, $m = \langle x\rangle$, is included even though the pdfs do not depend explicitly on it. We list the one-dimensional distributions in what follows; their generalization to three dimensions and associated potential functions are given elsewhere.[28,32] Derivations and further details are presented in reference 32. Examples of the distributions are shown in Figure 1a–e.

Gaussian distribution

The Gaussian pdf, corresponding to a harmonic potential in one dimension, is given by

$$p_G(u;m,\sigma) = \frac{1}{\sigma(2\pi)^{1/2}} e^{-\frac{1}{2}u^2/\sigma^2} \tag{6}$$

where $\sigma = \langle u^2\rangle^{1/2}$.

Gram-Charlier and Edgeworth expansions of the distribution function

An anharmonic distribution can be described by an expansion in terms of Hermite polynomials about a

Gaussian pdf. One such expansion, the Gram-Charlier expansion, is given by

$$p_{GC}(u;m,\sigma,\alpha_3,\alpha_4,\cdots) = p_G(u;m,\sigma)\left[1 + \frac{\alpha_3}{3!}H_3\left(\frac{u}{\sigma}\right) + \frac{\alpha_4}{4!}H_4\left(\frac{u}{\sigma}\right) + \cdots\right] \quad (7)$$

where $H_i(x)e^{-\frac{1}{2}x^2} = (-1)^i(d^i/dx^i)e^{-\frac{1}{2}x^2}$.[33]

The Edgeworth expansion has the form

$$p_E(u;m,\sigma,\alpha_3,\alpha_4,\ldots) = p_G(u;m,\sigma)\left[1 + \frac{\alpha_3}{3!}H_3\left(\frac{u}{\sigma}\right) + \frac{\alpha_4}{4!}H_4\left(\frac{u}{\sigma}\right) + \frac{10\alpha_3^2}{6!}H_6\left(\frac{u}{\sigma}\right) + \ldots\right] \quad (8)$$

It is equivalent to the Gram-Charlier expansion in the limit but combines terms for which the coefficients of the Hermite polynomials are of the same order to magnitude.[33] In Eq. 8, each term of the Edgeworth expansion is written on a separate line. Unless otherwise specified, we include terms through the fourth moment.

The Gram-Charlier and Edgeworth expansions have been used to describe anharmonic motion in small-molecule crystals for the purpose of refining X-ray data.[34-36] Some caution is required in their use, however, because both expansions may have regions where the probability has negative values when the third and/or fourth moments are large since the higher-order Hermite polynomials have negative regions.

Power law potential

Anharmonic effects may be included by parameterizing the potential rather than the pdf. Frauenfelder et al.[21] assumed the functional form of the three-dimensional potential to be an isotropic power law independent of temperature,

$$w_\nu(\mathbf{r}) = au_r^{1/\nu} \quad (9)$$

to determine anharmonic effects in the protein myoglobin from the temperature dependence of the temperature factors (B-factors). For a harmonic potential, $\nu = \frac{1}{2}$. We can obtain the pdf corresponding to the potential given in Eq. 9 for a given value of ν by using Eq. 1; the result is

$$p_{PL}(\mathbf{u};\mathbf{m},\nu) = \frac{(\beta a)^{3\nu}}{4\pi\nu\Gamma(3\nu)}e^{-\beta a u_r^{1/\nu}} \quad (10a)$$

which depends only on $u_r = |\mathbf{u}|$ and not on the components. The one-dimensional pdf corresponding to such a potential is

$$p_{PL}(u;m,\nu) = \frac{(\beta a)^\nu\Gamma(2\nu)}{2\Gamma(3\nu)}e^{-\beta a u^{1/\nu}}\sum_{k=0}^{2\nu-1}\frac{(\beta a u^{1/\nu})^k}{k!} \quad (10b)$$

for $2\nu \geq 1$, integer. The quantities σ^2 and α_{4i} are given by

$$\sigma^2 = \frac{\Gamma(5\nu)}{3\,\Gamma(3\nu)(\beta a)^{2\nu}} \approx \frac{1}{15^{1/2}}\left|\frac{5^5}{3^3(e\beta a)^2}\right|^\nu \quad (11a)$$

$$\alpha_{4i} = \frac{9\,\Gamma(7\nu)\,\Gamma(3\nu)}{5\,\Gamma(5\nu)^2} - 3 \approx \frac{9}{21^{1/2}}\left|\frac{7^7 3^3}{5^{10}}\right|^\nu - 3 \quad (11b)$$

where e is the base of the natural logarithm. We note that α_4 is independent of both a and β, as is true for all of the higher moments normalized with respect to σ. The value of α_4 varies slowly with ν for $\nu < \frac{1}{2}$ and rapidly with ν for $\nu > 1$. Since the potential is symmetric about the origin, α_{3i} is zero.

Expansions of the potential

The potential function can be expanded in a power series. This type of expansion has been used in many spectroscopic, X-ray crystallographic,[37] solid state,[38] and theoretical[39] applications. If we truncate after the 4th power, the potential is

$$w(x) = cu'^2 - gu'^3 - fu'^4 \quad (12)$$

where u' is the displacement measured from a minimum in the potential corresponding to the mean position at absolute zero. The actual mean position for this potential depends on the temperature; i.e., $\langle u'\rangle = (3g/4c^2)k_BT$ If $g \neq 0$ and/or $f > 0$, this potential form is useful only for small oscillations since the minimum at $u' = 0$ is only a local minimum.

Although the pdf and the moments corresponding to the potential in Eq. 12 cannot be obtained in closed form, approximate expressions for the moments can be derived when $a_3^2, a_4 \ll 1$. Defining $s^2 = (2!\beta c)^{-1}$, $a_3 = 3!s^3\beta g$, and $a_4 = 4!s^4\beta f$, the quantities σ, α_3, and α_4 through terms of order 2 in a_3 and order 1 in a_4 are given by

$$\sigma^2 \approx s^2\left(1 + \frac{1}{2}a_4 + a_3^2\right) \quad (13a)$$

$$\alpha_3 \approx a_3\left(1 + \frac{11}{4}a_4\right) \quad (13b)$$

$$\alpha_4 \approx a_4 + 3a_3^2 \quad (13c)$$

Multiple Gaussians

For cases in which the expansion about a single Gaussian is not adequate (e.g., a distribution corresponding to multiple minima), we introduce a sum over Gaussian peaks, properly normalized. This could represent multiple occupancy in a crystallographic refinement. If an atom can be in either of two harmonic potential wells with different probabilities, we can represent the pdf as

$$p_M(u; m, u_\pm, \sigma_\pm) = \frac{1}{(w_-\sigma_- + w_+\sigma_+)(2\pi)^{1/2}}\left[w_-\exp\left(-\frac{(u+u_-)^2}{2\sigma_-^2}\right) + w_+\exp\left(-\frac{(u-u_+)^2}{2\sigma_+^2}\right)\right] \quad (14)$$

where $u_\pm$, which are defined to be positive, are the positions of the maxima relative to m, $w_\pm$ are the relative probabilities of being in each well ($w_+ + w_- = 1$), and $\sigma_\pm$ is the standard deviation of each peak, each of which is assumed isotropic. The parameters $u_\pm$, $w_\pm$, and $\sigma_\pm$ in principle, can all depend on temperature. In the limits when one of the peaks becomes infinitely small (either $w_+ \to 0$ or $w_- \to 0$) or when the separation between the peaks approaches zero ($u_+ + u_- \to 0$), the pdf approaches a Gaussian.

We assume that the two peaks have the same half-width to reduce the number of parameters so $\sigma_+ = \sigma_- = \sigma_0$, which is equivalent to assuming that the force constants are the same for the two potential wells. Since $u_\pm$ are measured from the mean, $w_-u_- = w_+u_+$. The higher moments are

$$\sigma^2 = \sigma_0^2 + u_+u_- \quad (15a)$$

$$\alpha_3 = u_+u_-(u_+ - u_-)/\sigma^3 \quad (15b)$$

$$\alpha_4 = u_+u_-(u_+^2 - 4u_+u_- + u_-^2)/\sigma^4 = -2u_+^2u_-^2/\sigma^4 + \alpha_3^2\sigma^2/(u_+u_-) \quad (15c)$$

For $u_+ = u_-$, $\alpha_3 = 0$ and $\alpha_4 = -2u_\pm^4/\sigma^4$, i.e., the pdf is symmetric and the kurtosis is negative with the magnitude increasing with increasing peak separation. If $u_+ > u_-$, there is positive skewness ($\alpha_3 > 0$). Furthermore, any skewness, positive or negative, increases α_4, but the contribution of the skewness decreases with increasing peak separation. The double peak pdf is illustrated in Figure 1a–e. Note that not all values of α_3 and α_4 correspond to real values for u_+ and u_- and that the most negative value possible for α_4 is -2. It is easy to generalize Eq. 14 to more peaks,[32] although it is necessary to have more information than the first four moments from a simulation to obtain the parameters m, u_i, and σ_i if there are more than two peaks.

The effective local potential represented by the distribution is given by the logarithm of Eq. 14; i.e.,

$$\beta w(u) = \frac{(u+u_-)^2 + (u-u_+)^2}{4\sigma_0^2} - \log[\cosh f(u) + (w_+ - w_-)\sinh f(u)] \quad (16)$$

where $f(u) = [2u(u_+ + u_-) - u_+^2 + u_-^2]/(4\sigma_0^2)$. Thus, there are two harmonic wells centered at u_+ and u_-; the cosh term makes transitions possible and the sinh term leads to asymmetry. Thus, unlike the potential expansion, there are exact expressions for both the pdf and the potential and the parameters $u_\pm$ and σ_0 can be solved for exactly in terms of σ, α_3, and α_4 (i.e., quantities easily calculated from a dynamics simulation) from Eq. 15. Furthermore, the pdf is always positive.

In many applications, it is of interest to reduce the number of parameters. One reasonable way is to assume that there are only two isotropic Gaussian peaks which lie along the direction of the largest principal axis and the fluctuations in the other axis directions are harmonic; i.e., the distributions for the U_y and U_z directions are Gaussian and that for the U_x-direction is given by Eq. 14.

A multiminimum potential has been used by Nadler and Schulten[40] to describe Mössbauer spectra of proteins fluctuating between different conformational substates. Their potential, given by

$$w(x) = b\exp[-\sin 2(n\pi x/sx_0)] \quad (17)$$

where $b(e-1)/e$ is the height of the barriers between adjacent substrates and e is the base of the natural logarithm, has n minima if motions are confined to the interval $0 \leqslant x \leqslant 2x_0$. This potential is not harmonic when $n=1$ and gives a more complicated pdf than the multiple Gaussians.

SIMULATION METHOD

For the purpose of analyzing the characteristics of the atomic motions, we have used the probability densities and their moments obtained from a molecular dynamics trajectory for the protein hen egg-white lysozyme (129 amino acids, 1,001 heavy atoms); details of simulation for lysozyme are presented elsewhere.[27] After an equilibration period of 26 picoseconds (ps), the simulation was continued for an additional 34 ps of which the first 30 ps were used for this analysis. The time step was 0.001 ps and the mean temperature was equal to 304°K. The moments and the probability densities were calculated from coordinates at intervals of 0.05 ps and 0.01 ps, respectively.

The probability densities were estimated by calculating normalized histograms of the coordinates with an interval size of 0.1 Å. The moments of the distribution were obtained from molecular dynamics simulations by equating the time averages over the trajectory, indicated by overbars,

$$\overline{u_i^n} = \frac{1}{T}\int_0^T [u_i(t)]^n\,dt \approx \frac{1}{N}\sum_{i=1}^{N}[u_i(t_i)]^n \quad (18)$$

with phase space averages (see Eq. 2); i.e., we equate $\langle u_i^n \rangle$ and $\overline{u_i^n}$.

The question of how well the pdfs obtained from the dynamics simulations are represented by the various model pdfs given in the last section is examined by using a goodness-of-fit criterion based on the χ^2-test[31]

$$\chi^2 = \sum_{j=1}^{k} \frac{[Np_o(u_j) - Np_c(u_j;\Theta_n)]^2}{Np_c(u_j;\Theta_n)} \qquad (19)$$

where p_o and p_c are the observed and model distributions, respectively. The observed values [in this case, $U_x(t_i)$, $U_y(t_i)$, or $U_z(t_i)$] have been divided into k intervals with u_j the value at the center of the jth interval, N the total number of independent observations, and $Np_o(u_j)$ the number of observations with value u_j. If the motion of the atom is Brownian, then N may be estimated as T/τ where τ is the correlation time and T is length of the time series.[32] The correlation times of local motions of protein atoms are estimated as $\tau \approx 0.2$ ps[41] so that for the 30 ps simulation of lysozyme, N is ~150.

The probability of the value u_j for the model distribution being tested is $p_c(u_j;\Theta_n)$, where $\Theta_n = (\theta_1,...,\theta_n)$ are the n parameters determining the model. If the observed values are exactly equal to the predicted values $p_o(u_j) = p_c(u_j)$, then $\chi^2 = 0$. However, for a real situation where the observed frequencies fluctuate

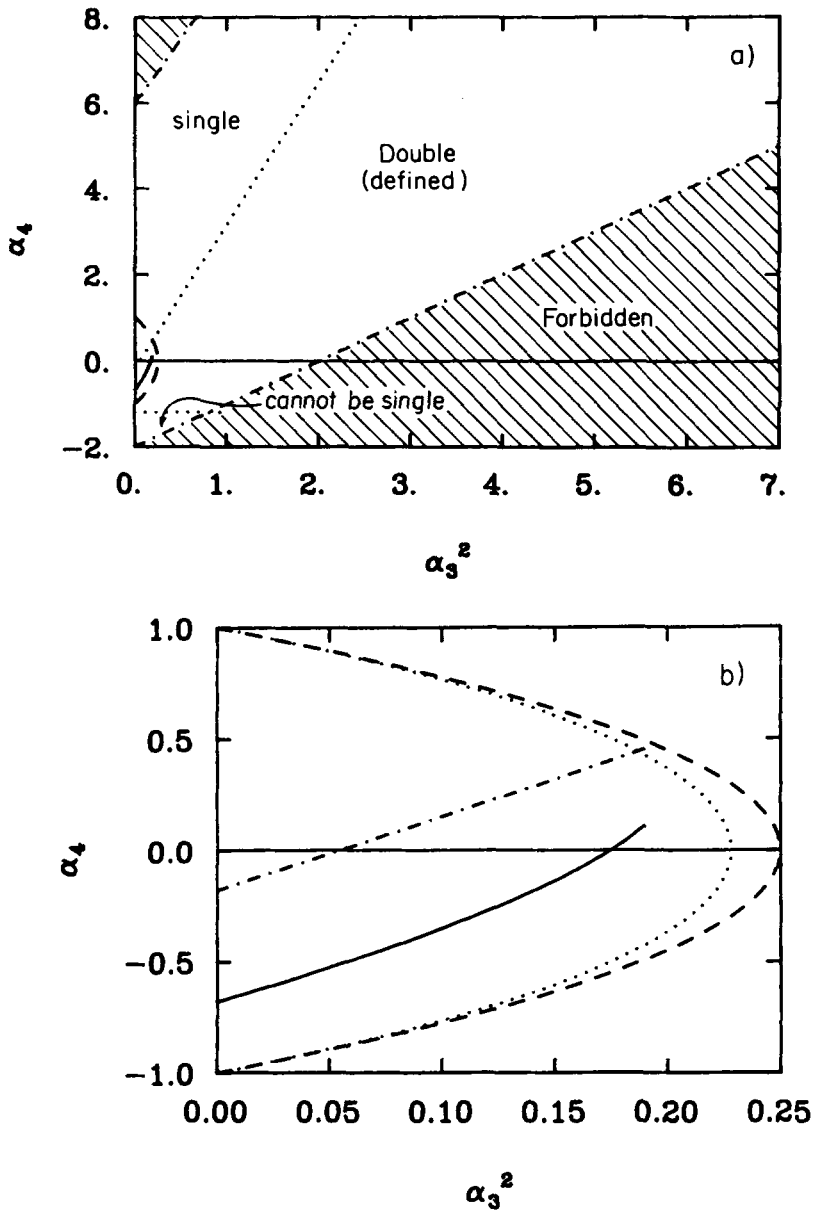

Fig. 2. Classification of values of α_3 and α_4. **a:** The region of allowed values of α_3 and α_4 ($-2 + \alpha_3^2 < \alpha_4 < 6 + 3\alpha_3^2$) is delineated by (—·—) and the forbidden range is shaded. The separation between the values which are defined as single peak and those defined as double peak is given by $\xi^2 = 0.3$ (···) and the separation between the values defined as double peak and those which must be at least double peaks is given by $\alpha_4 = -1.2$ (···). The harmonic region in the neighborhood of $\alpha_3 \equiv \alpha_4 \equiv 0$ is enlarged in **b** where $\Delta\chi^2$ for the Gram-Charlier (—·—), Edgeworth (···), and double peak (– – –) pdfs are equal to N/24; see Eq. 20a, A2, A3.

due to statistical errors, $\chi^2 \approx \nu$ where $\nu = k-n-1$ if the observations are a sample from the model distribution c.[31] The value of χ^2/ν should be approximately independent of k if there are enough points in each interval; equal size intervals (0.1, 0.2, 0.3, or 0.5 Å) were used so that the value of k was between 9 to 26. If $\chi^2 \gg \nu$, the observed frequencies probably come from a distribution which differs from the distribution c, either in the values of the parameters Θ_n or in the form of the distribution.

We define criteria for anharmonicity and presence of multiple peaks. A distribution is considered anharmonic if

$$\Delta\chi^2_{GC} = N\left[\frac{\alpha_3^2}{3!} + \frac{\alpha_4^2}{4!}\right] > \frac{1}{24} \qquad (20a)$$

and double peaked if the above condition holds plus

$$\xi^2 = u_+u_-/\sigma^2 > 0.3 \qquad (20b)$$

The basis for these criteria is given in Appendix A.

Figure 2 shows the values of α_3 and α_4 which have $\Delta\chi^2$ less than N/24, for three models of the parent pdf—namely, the Gram-Charlier, Edgeworth, and double peak. Regardless of the functional form of the pdf, $-2 + \alpha_3^2 < \alpha_4 < 6 + 3\alpha_3^2$.[31] Furthermore, for a unimodal distribution, $\alpha_4 \geq -1.2$.[31] The model pdfs evaluated at several values of α_3 and α_4 are illustrated in Figure 1a–e. For the Gram-Charlier and Edgeworth pdfs, the values shown all have $\Delta\chi^2_{GC}/N \approx 1/24$. For the double peak pdf, $\Delta\chi^2_{dp}/N$ are 0.066 (Fig. 1a), 0.13 (Fig. 1b), 0.12 (Fig. 1d), and 0.063 (Fig. 1e). The double peak pdf is undefined for $\alpha_3=0$, $\alpha_4=1$. The model pdfs evaluated at several values of α_3 and α_4 corresponding to $\xi^2=0.3$ are shown in Figure 3a–d. The value of $\Delta\chi^2_{GC}/N$ are 0.00135 (Fig. 3a), 0.059 (Fig. 3b), 0.58 (Fig. 3c), and 2.6 (Fig. 3d). Although the values of $\Delta\chi^2_{GC}$ for the pdfs in Figure 3c and d predict anharmonic distributions, the double well pdf is close to the harmonic pdf between $\pm 3\sigma$ for both cases.

ANISOTROPY AND ANHARMONICITY RESULTS

In this section, we calculate the anisotropy of the fluctuations to assess the limitations of the isotropic motion model and the range of values of α_3 and α_4 to evaluate the degree of anharmonicity.

Anisotropies

The measures of anisotropy, A_1 and A_2 (Eq. 5a,b), of the atomic fluctuations in the lysozyme simulation were calculated for all of the atoms. Since the value of A_2 is fairly uniform, we discuss only A_1 in detail. Table I gives average values and Figure 4a shows the distribution for A_1. The average value of A_1 for all atoms is 0.85, which indicates that the fluctuations along the principal axis U_x with the largest value are almost twice as big as the fluctuations along the average of the U_y and U_z directions. However, not many have anisotropies large than this; i.e., 71% of the atoms have $A_1 < 1$ and only 5% have $A_1 > 2$. Since only 3% have $A_1 < 0.2$, relatively few have nearly isotropic motion. The sidechain atoms have, on the average, a higher anisotropy (0.93) than the backbone atoms (0.77) and the anisotropy increases with distance from the backbone (Table II). More sidechain atoms are anisotropic since only 59% have

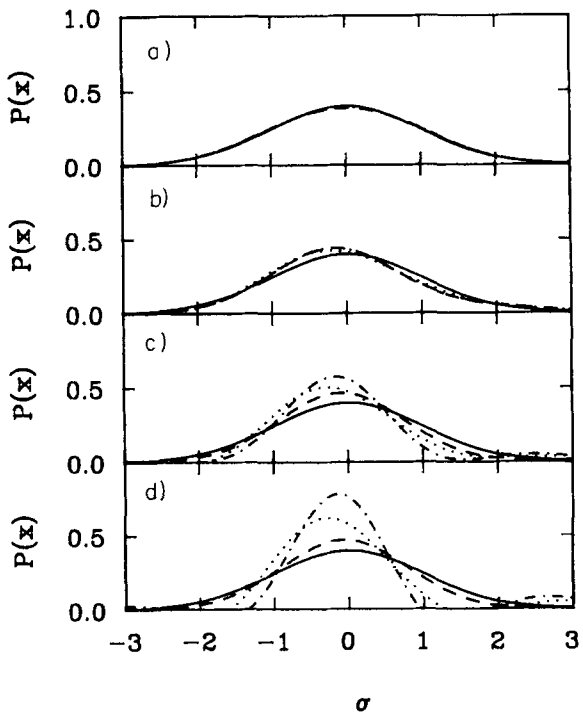

Fig. 3. Examples of pdfs with $\xi^2 = 0.3$. **a**: $\alpha_3=0$, $\alpha_4=-0.18$. **b**: $\alpha_3=0.5$, $\alpha_4=0.65$. **c**: $\alpha_3=1$, $\alpha_4=3.15$. **d**: $\alpha_3=1.5$, $\alpha_4=7.32$. The distributions are the Gram-Charlier (–·–), Edgeworth (···), and double peak (– –). For reference, the Gaussian with the same value of σ is given by the solid line.

TABLE I. Average Anisotropies*

Atom types	A_1	A_1'	A_2	$\langle U_x^2\rangle^{1/2}$	$\langle U_y^2\rangle^{1/2}$	$\langle U_z^2\rangle^{1/2}$
All	0.85 (0.55)	0.23 (0.28)	0.14 (0.07)	0.63	0.35	0.24
Backbone	0.77 (0.50)	0.21 (0.28)	0.13 (0.06)	0.50	0.29	0.21
Sidechain	0.93 (0.59)	0.25 (0.28)	0.15 (0.07)	0.75	0.41	0.27

*Average over atom types with the standard deviation in parentheses.

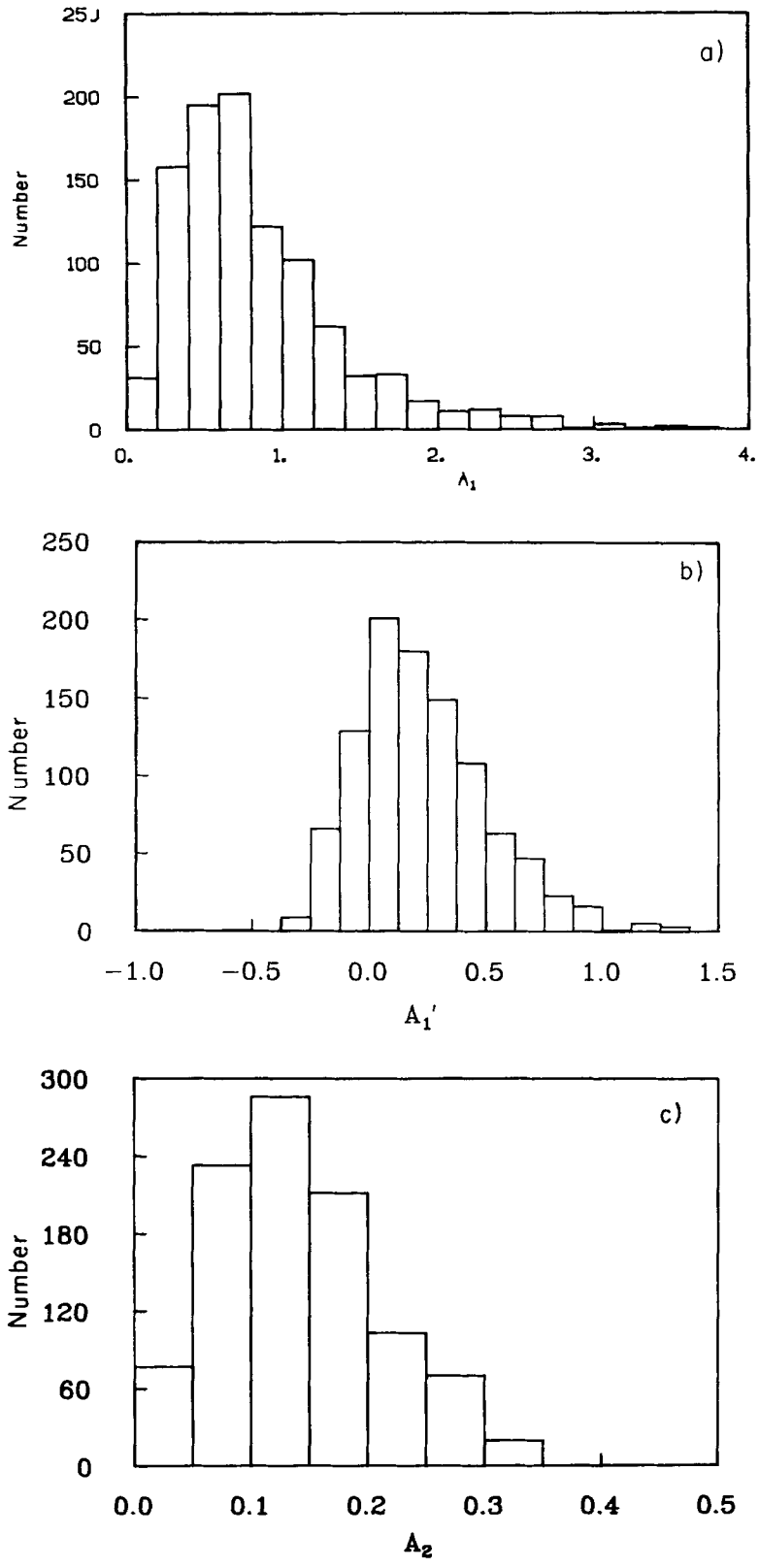

Fig. 4. Distributions of anisotropy from the trajectory for the 1,001 atoms of lysozyme. The number of atoms with a given anisotropy is shown: (**a**) A_1; (**b**) A_1'; (**c**) A_2.

$A_1 < 1$ and 7% have $A_1 > 2$ vs. 77% and 3%, respectively, for backbone atoms. The average values of the fluctuations in each principal axis direction (Table I) show the greatest difference between backbone and sidechain in the U_x direction and the smallest in the U_z direction.

The anisotropies averaged in shells about the geometric center are given in Table III; the anisotropies of sidechains atoms (and the range of anisotropies) tend to increase with distance from the geometric center whereas the effect for backbone atoms is very small. It has been shown[27,32] that $\langle u_r^2 \rangle^{1/2}$, the rms atomic fluctuation, increases with distance from the center as well.

A_1 is highly correlated with large fluctuations; the correlation coefficient of A_1 with $\langle U_x^2 \rangle^{1/2}$ is 0.75 and with $\langle u_r^2 \rangle^{1/2}$ is 0.67. Figure 5 shows a scatterplot of A_1 vs. $\langle u_r^2 \rangle^{1/2}$. This indicates that the use of isotropic temperature factors for X-ray refinement is a poorer approximation for atoms with large fluctuations. Of the backbone atoms, the carbonyl oxygens have the largest anisotropies; the direction of the largest fluctuations is about 70° to the C–O bond. The anisotropies of the backbone atoms in three of the α-helices, 5–15, 24–34, and 88–96, are rather small (0.51–0.57), whereas they are rather high in the β sheet (~1).

Earlier in this paper, it was proposed that anisotropic and anharmonic effects can be included by assuming two isotropic Gaussian peaks with the same width. The anisotropy of either peak can be estimated by a quantity A_1' where $\langle U_x^2 \rangle$ in Eq. 5a is replaced by σ_0^2, the width of either peak along the U_x direction (Eq. 15). From Table I, A_1' has an average value of 0.23 ± 0.28, which indicates that the width of either peak in the direction of the largest principal axis is on the average only 20% larger than the average of the widths in the two smallest directions (see Fig. 4b). The rms value of σ_0 over all atoms is 0.38 Å, which is close to that of $\langle U_y^2 \rangle^{1/2}$ (0.35 Å) and of $\langle U_z^2 \rangle^{1/2}$ (0.24 Å) (Table I). These results indicate that if the double peak pdf is a good representation of the pdf in the U_x direction, an anisotropic pdf can be described as a double peak pdf with two isotropic peaks lying along

TABLE II. Average Anisotropies (A_1) as a Function of Atom Type

Atom type	A_1*
N	0.68 (0.43)
C	0.74 (0.45)
O	0.93 (0.60)
C_α†	0.73 (0.47)
C_β†	0.74 (0.45)
γ†	0.90 (0.55)
δ†	0.95 (0.56)
ϵ	1.03 (0.67)

*Average over atom types with the standard deviation in parentheses.
†Not including prolines.

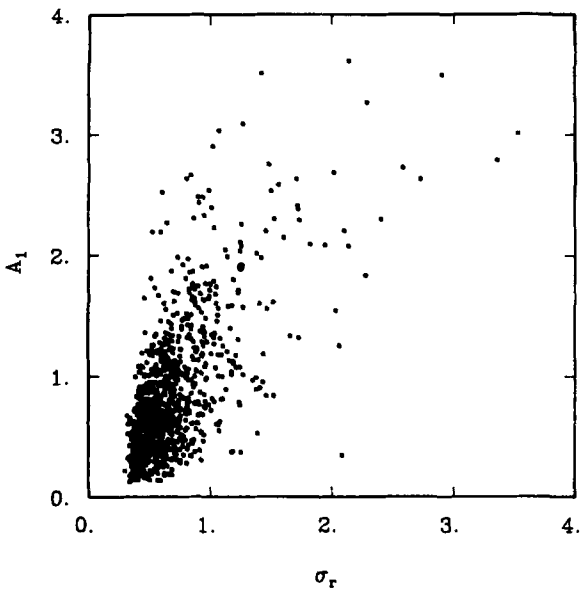

Fig. 5. Correlation of A_1 and $\sigma_r = \langle u_r^2 \rangle^{1/2}$.

TABLE III. Anisotropies in Shells

Shell (Å)	Number of atoms in shell*	Anisotropies, A_1†	
		Backbone	Sidechain
0–6	57	0.83 (0.54)	0.78 (0.35)
6–9	143	0.68 (0.46)	0.89 (0.65)
9–12	269	0.74 (0.47)	0.91 (0.50)
12–15	309	0.80 (0.53)	0.90 (0.60)
15–18	145	0.78 (0.47)	1.05 (0.70)
18–21	69	0.94 (0.58)	1.00 (0.57)

*Backbone and sidechain atoms.
†Average over shells with the standard deviation in parentheses.

TABLE IV. Anharmonicity: $|\alpha_3|$, $|\alpha_4|$, and α_4*

Atom type	U_x	U_y	U_z
a. $\|\alpha_3\|$			
All	0.38 (0.32)	0.25 (0.23)	0.18 (0.16)
Backbone	0.34 (0.28)	0.22 (0.20)	0.17 (0.14)
Sidechain	0.42 (0.36)	0.28 (0.26)	0.20 (0.18)
b. $\|\alpha_4\|$			
All	0.56 (0.52)	0.39 (0.49)	0.31 (0.36)
Backbone	0.50 (0.43)	0.33 (0.37)	0.27 (0.25)
Sidechain	0.61 (0.59)	0.46 (0.58)	0.35 (0.44)
c. α_4			
All	−0.12 (0.75)	0.05 (0.63)	−0.01 (0.47)
Backbone	−0.16 (0.64)	−0.05 (0.49)	−0.03 (0.37)
Sidechain	−0.08 (0.84)	0.15 (0.73)	0.02 (0.56)

*Averages over atom types with the standard deviation in parentheses.

the largest principal axis. The rms value of $\sigma_{0r} = [\sigma_0^2 + \langle U_y^2\rangle + \langle U_z^2\rangle]^{1/2}$ is 0.57 Å compared with that of $\langle u_r^2\rangle^{1/2}$, which is 0.76 Å. Since 88% of the atoms have $\sigma_0 < 0.5$ Å and 89% have $\sigma_{0r} < 0.75$ Å (57% have $\langle U_x^2\rangle^{1/2} < 0.5$ Å and 72% have $\langle u_r^2\rangle^{1/2} < 0.75$ Å), the results here indicate that atoms with $\langle U_x^2\rangle^{1/2}$ larger than 0.5 Å or with $\langle u_r^2\rangle^{1/2}$ larger than about 0.75 Å will often have double-peak pdfs.

A_2 measures the similarity of the U_y and U_z directions. The average value of A_2 is 0.14 ± 0.07 (Table I), which means that the fluctuations in the U_y directions on the average differ by only about 14% of the magnitude of the average of the U_y and U_z directions. This indicates that the fluctuations are close in magnitude in the two smallest principal axis directions. In fact, 81% of the atoms have $A_2 < 0.2$ and the

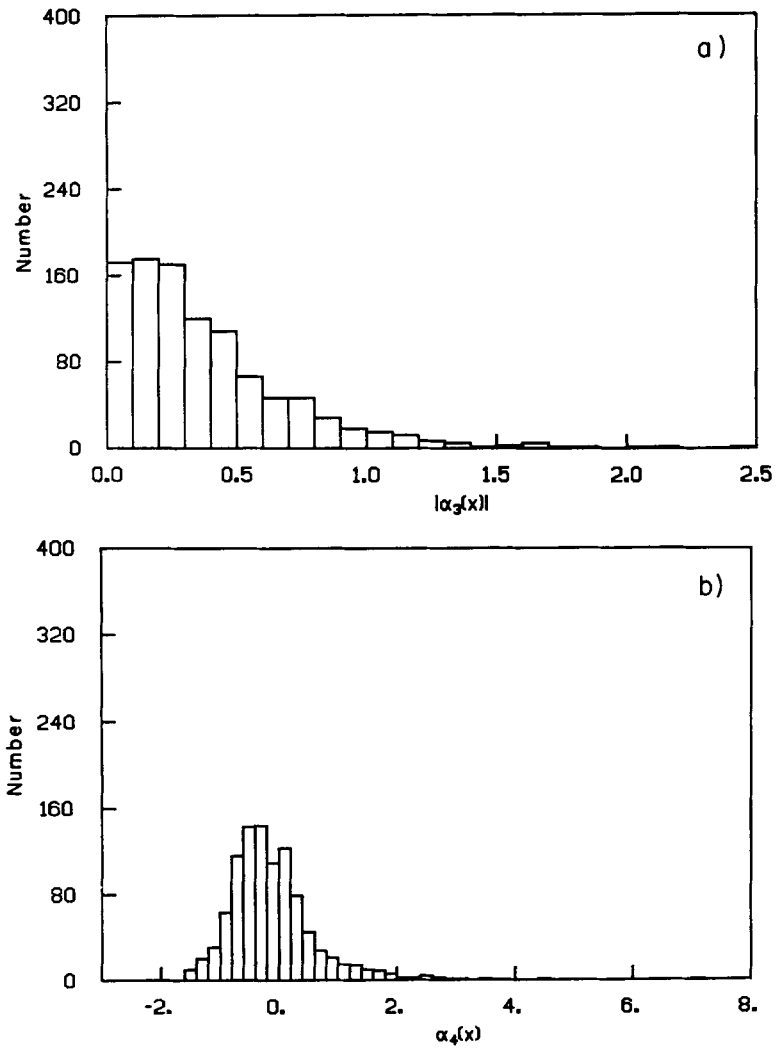

Fig. 6. Distributions of (a) skewness, $|\alpha_3|$, and (b) kurtosis, α_4, in the lysozyme trajectory in the U_x direction of the principal axis system. Histograms of the number of atoms with a given value are shown.

TABLE V. Anharmonicity by Atom Type (U_x Only)*

Atom type	$\lvert\alpha_3\rvert$	$\lvert\alpha_4\rvert$	α_4
N	0.30 (0.25)	0.46 (0.38)	−0.15 (0.58)
C	0.33 (0.27)	0.48 (0.38)	−0.22 (0.57)
O	0.38 (0.33)	0.57 (0.56)	−0.08 (0.79)
C_α	0.33 (0.25)	0.50 (0.38)	−0.19 (0.60)
C_β†	0.32 (0.24)	0.48 (0.37)	−0.16 (0.59)
γ†	0.40 (0.36)	0.57 (0.48)	−0.14 (0.73)
δ†	0.45 (0.38)	0.65 (0.53)	−0.10 (0.83)
ϵ	0.53 (0.51)	0.85 (1.07)	0.24 (1.34)

*Averages over atom types with the standard deviation in parentheses.
†Not including prolines.

TABLE VI. Anharmonicities in Shells (U_x for All Atoms)*

Shell (Å)	$\lvert\alpha_3\rvert$	α_4
0–6	0.36 (0.30)	0.19 (0.17)
6–9	0.38 (0.35)	0.25 (0.22)
9–12	0.34 (0.28)	0.22 (0.21)
12–15	0.37 (0.32)	0.28 (0.25)
15–18	0.41 (0.34)	0.28 (0.26)
18–21	0.39 (0.40)	0.22 (0.21)

*Averages over shells with the standard deviation in parentheses.

largest value for any atom is 0.33 (see Fig. 4c). The utility of the small values of A_2 in crystallographic refinement procedures is discussed in a separate paper.[28]

The results presented here can be compared only qualitatively with those for the molecular dynamics simulation[14] of cytochrome c since different measures of anisotropy are used (see Eq. 5a,b and below). In the cytochrome c simulation, the fluctuations in the U_x directions are about 1.4 and 2 times the fluctuations in the U_y and U_z directions, respectively, for all atoms, which is in accord with our results for the average value of A_1. Furthermore, the same general trends are found for backbone vs. sidechains atoms. For sidechain atoms as a function of distance from the backbone, and in shells about the geometric center.

Anharmonicities from α_3 and α_4

The anharmonicities of the local potentials of mean force were studied by calculating a_3 and α_4 in the three principal axis directions for all the atoms. The average values of $\lvert\alpha_3\rvert$ for all atoms for the U_x,U_y,U_z directions are 0.38 ± 0.32, 0.25 ± 0.23, and 0.18 ± 0.16, respectively (Table IV,a) and the average values of $\lvert\alpha_4\rvert$ for all atoms in the U_x,U_y,U_z directions are 0.56 ± .52, 0.39 ± .49, 0.31 ± .36, respectively (Table IV,b); that is, the direction of the largest fluctuations is most anharmonic. The most probable values for $\lvert\alpha_3\rvert$ are closer to zero than the average values for the U_x direction (see Fig. 6a). The average value of α_4 for all atoms is close to zero in the three principal axis directions; i.e., $\alpha_{4x} = -0.12 \pm 0.75$, $\alpha_{4y} = 0.05 \pm 0.63$, and $\alpha_{4z} = -0.01 \pm 0.47$, respectively (Table IVc). The large standard deviation indicates that there is a distribution of positive and negative values, as is apparent in the histogram for α_4 in Figure 6b. The most probable value of α_4 is ~ −0.4 in the U_x direction and ~ −0.1 in the U_y and U_z directions, but the distributions are skewed so that the largest value is ~7, whereas the smallest value is ≈ −1.6; the smallest possible value is −2 (Fig. 2).

The magnitudes of the anharmonicity measured by $\lvert\alpha_3\rvert$ and $\lvert\alpha_4\rvert$ (Table IV,a,b) for sidechain atoms are larger than for backbone atoms and increase as the atom is further away from the backbone (Table V). The C_α and C_β have similar average values, and the carbonyl O is closer to the γ-atoms than to C_β. The values of α_4 (Table V) increase slightly in going away from the backbone. For a single minimum well, this indicates that the atomic potentials are softer to larger fluctuations for sidechains atoms further away from the backbone.

There is very little correlation between either $\lvert\alpha_3\rvert$ or α_4 and distance from the center of the molecule (or of the lobes) (Table VI) or $\langle u_r^2\rangle^{1/2}$.[32] There is, however, a high degree of correlation between $\alpha_3^2$and α_4 (Fig. 7). The correlation coefficients for the fit of $\alpha_4 = m\alpha_3^2 + b$ with respect to the U_x, U_y, U_z directions, respectively, are 0.70, 0.75, and 0.65. If there are two peaks, this correlation is explained by Eq. 15c since α_4 is a function of α_3^2 with positive slope $\sigma^2/\lvert u_+u_-\rvert$.

In the scatter plots of α_3^2 vs. α_4 (Fig. 7a–c), lines indicating the range of allowed values of α_3 and α_4 ($-2+\alpha_3^2 < \alpha_4 < 6+3\alpha_3^2$), the range of harmonic values ($\Delta\chi_{GC}^2 < N/24$), the range which must be single peaks ($\xi^2 < 0.3$), and the range which cannot be single peaks ($\alpha_4 < -1.2$) are also shown. It is apparent from these plots that many atoms are essentially harmonic and that those that are not harmonic usually are within the range which can be described as double peak. The largest values of α_3 and α_4 are 2.47 and 7.20 ($\Delta\chi_G^2 = 3.17$), respectively, both for Gln 41 $N_{\epsilon 2}$. The percentages of atoms with various values of $\lvert\alpha_3\rvert$, $\lvert\alpha_4\rvert$, $\Delta\chi_G^2$, $\Delta\chi_{dp}^2$, and ξ^2 are summarized in Table VII. By the criterion $\Delta\chi_G^2 < N/24$, approximately half of the atoms are harmonic in all three principal axis directions; i.e., 59% of backbone and 41% of sidechain atoms satisfy the criterion along all three axes. An even higher percentage are harmonic in at least the U_x direction (68% of the backbone and 56% of the sidechain atoms). Since the fluctuations in the two smallest principal axis directions are about half the size of those in the largest direction from the values of A_1, the anharmonicities in the U_y and U_z

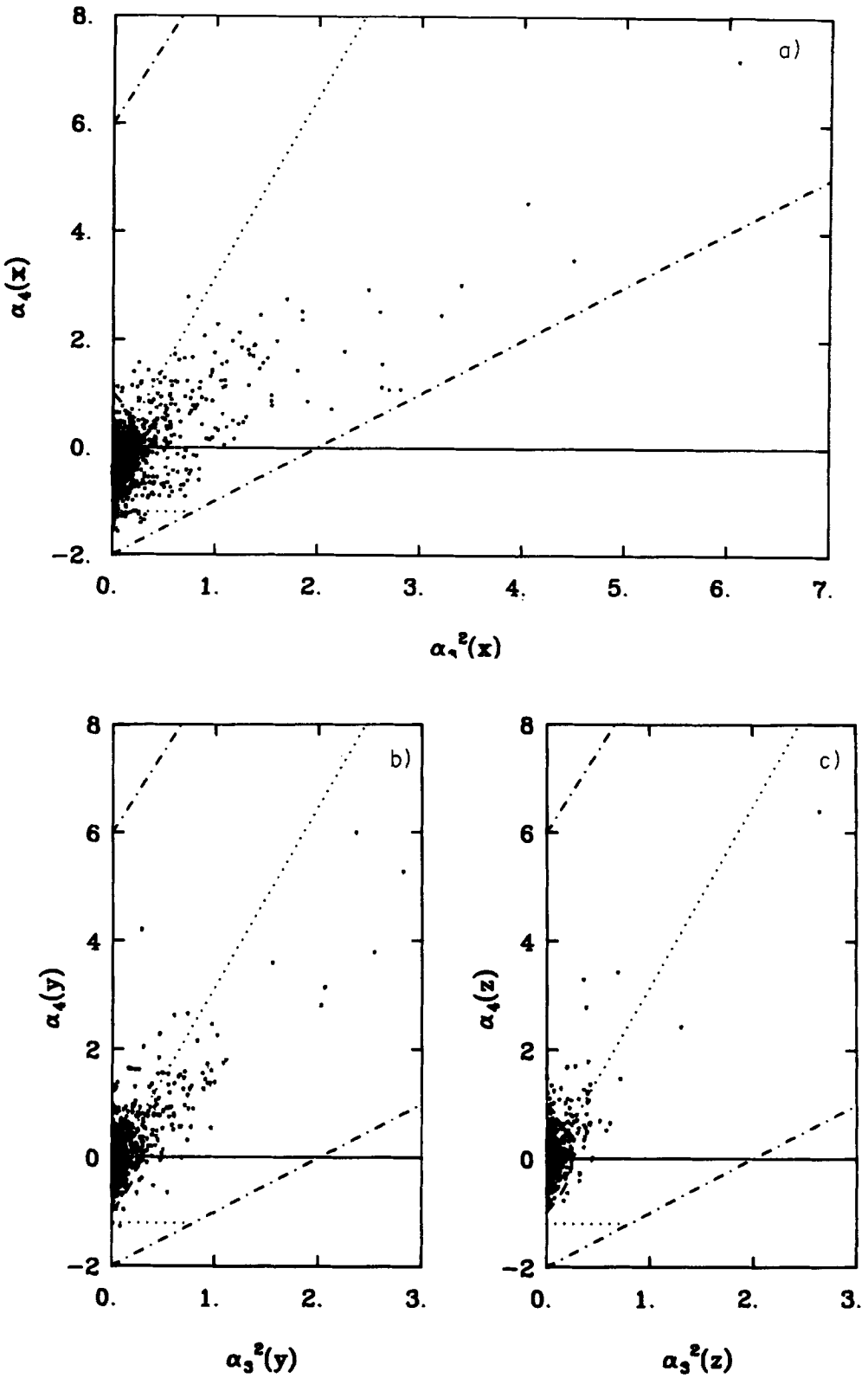

Fig. 7. Correlation of α_3^2 and α_4 in the (**a**) U_x, (**b**) U_y, and (**c**) U_z directions of the principal axis system. The various limiting values of α_3 and α_4 are also drawn (see Fig. 2 for key).

TABLE VII. Degree of Anharmonicity

Atoms with:	Direction	Backbone (%)	Sidechain (%)
$\|\alpha_3\| < ½$	U_x	78	71
$\|\alpha_4\| < 1$	U_x	90	84
$\|\alpha_3\| < ½$ and $\|\alpha_4\| < 1$	U_x	73	64
$\|\alpha_3\| < ½$ and $\|\alpha_4\| < 1$	U_x,U_y&U_z	66	50
$\Delta\chi^2_{GC} < N/24$	U_x	68	56
$\Delta\chi^2_{GC} > N/24$ and $\xi^2 > 0.3$ in U_x	U_x	31	41
$\Delta\chi^2_{GC} < N/24$	U_x,U_y&U_z	59	41
$\Delta\chi^2_{GC} > N/24$ and $\xi^2 > 0.3$ in U_x	U_x,U_y&U_z	31	41

directions are less important. Of the remaining atoms, another 31% of the backbone and 41% of the sidechain atoms are considered to have double peak distributions since they have $\xi^2 > 0.3$

DISTRIBUTIONS FROM THE SIMULATIONS: COMPARISON WITH MODEL DISTRIBUTIONS

PDFs and Fits to Model Distributions

Since the molecular dynamics trajectory determines the complete pdfs as well as its moments, we can evaluate the ability of the various models to predict the shape of pdfs, given the values of σ, α_3, and α_4 calculated from the trajectory. This is important because it provides a means of interpreting the values of α_3 and α_4. From Figures 1 and 3, it is evident that pdfs with different shapes may have the same α_3 and $\alpha 4$, and that the deviation from a Gaussian pdf for a given value of α_3 and α_4 depends on the type of distribution. The detailed comparison presented here, in which the moments are known and only the model of the pdf is being tested, is not possible in an X-ray crystallographic experiment, since the actual moments of the distributions are not known; instead, they are parameters (so far only **m** and either σ or, in a few cases, the anisotrophy tensor σ^2, for proteins) to be varied in the specific models used to interpret the electron density.

The fit of a Gaussian pdf was compared with the three higher-order pdfs, the Gram-Charlier, Edgeworth, and double peak. We concentrate on these models since they include skewness as well as kurtosis, whereas the power-law pdf is independent of skewness. Also, these three pdfs can be parameterized exactly, whereas the potential expansion cannot (see the earlier section on Probability Density Functions). For the analysis, the pdfs of 20 atoms (9 backbone and 11 sidechain) were calculated in the local principal axis system (i.e., a total of 60 pdfs). These include the atoms with the largest values of α_3 and α_4. Eight of the backbone atoms were chosen because of large anharmonicity along one or more axis; these were the C_α of Gly 4, Leu 17, Ala 31, Thr 47, Gly 71, Gly 104, Met 105, Thr 118. Leu 83 C_α, which has relatively low anharmonicity, was also examined. Eight of the sidechain atoms were chosen because of large anharmonicity; these were Glu 7 $O_{\epsilon 2}$, His 15 $N_{\delta 1}$, Glu 35 $O_{\epsilon 1}$, Gln 41 $N_{\epsilon 2}$, Gln 57 $O_{\epsilon 1}$ and $N_{\epsilon 2}$, and Gln 121 $O_{\epsilon 1}$ and $N_{\epsilon 2}$. Gln 7 $O_{\epsilon 1}$ and Gln 41 $O_{\epsilon 1}$, which has relatively low anharmonicity despite the high value for the other ϵ atom, were also examined.

Figure 8 shows the pdfs for several atoms calculated from the dynamics trajectories along with the various model pdfs with parameters given by the moments for the dynamics. Of the pdfs examined, if $\sigma \lesssim 0.5$ Å, the pdf usually has only one peak and appears close to harmonic. Glu 7 $O_{\epsilon 1}$, U_x (Fig. 8a) is an example of a pdf with low values of α_3 and α_4; i.e., a single maximum without skewness. A Gaussian pdf fits this case very well, and furthermore, the Gram-Charlier, Edgeworth, and double peak pdfs are very close to the Gaussian.

Various types of pdfs with higher anharmonicity are shown in Figure 8b–f. The pdfs examined with $\xi^2 > 0.3$ are often better characterized as multiple harmonic peaks rather than as a single anharmonic peak. For instance, the pdf of Gln 57 $N_{\epsilon 2}$, U_x has two approximately equally populated peaks (Fig. 8b); that of Gln 41 $N_{\epsilon 2}$, U_x has a large peak with a small peak about 2.5 Å from the mean (Fig. 8c); and that of Thr 47 C_α, U_x has multiple peaks (Fig. 8d). For these cases, the double-peak pdf appears closest to the dynamics pdf, even when the population of one of the peaks is very small as in Gln 41 $N_{\epsilon 2}$ (Fig. 8c) or when there are more than two peaks as in Thr 47 C_α (Fig. 8d).

There are also some pdfs with large anharmonicity but only one apparent peak; examples are Gln 41 $O_{\epsilon 1}$, U_z, which has a distribution with a single skewed peak with small kurtosis (Fig. 8e), and Gln 121 $O_{\epsilon 1}$, U_z, which has a single peak with positive kurtosis with small skewness (Fig. 8f). For single peaks that are skewed, none of the distributions consistently give the best fit and the differences between the distributions are small. The skewness in some cases (e.g., Gln 41 $O_{\epsilon 1}$, U_z, Fig. 8e; Gln 57 $O_{\epsilon 1}$, U_y not

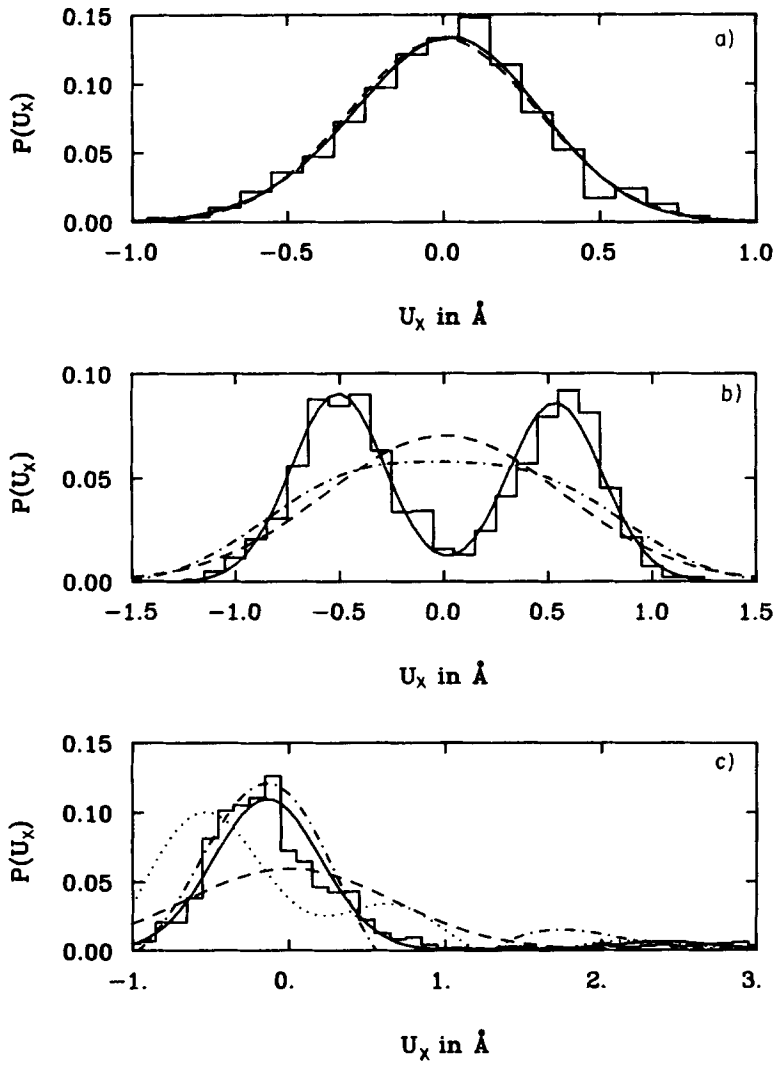

Fig. 8. Atomic distributions. The histograms correspond to distributions calculated from the lysozyme trajectory: **a**: Gln 7 $O_{\epsilon 1}$, U_x (σ = 0.301, α_3 = −0.124, α_4 = 0.061). **b**: Gln 57 $N_{\epsilon 2}$, U_x (σ = 0.567, α_3 = 0.039, α_4 = −1.414). **c**: Gln 41 $N_{\epsilon 2}$, U_x (σ = 0.669, α_3 = 2.470, α_4 = 7.196). **d**: Thr 47 C_α, U_x (σ = 1.592, α_3 = 0.527, α_4 = −0.696). **e**:Gln 41 $O_{\epsilon 1}$, U_z (σ = 0.200, α_3 = 0.453, α_4 = 0.095). **f**: Gln 121 $O_{\epsilon 1}$, U_z (σ = 0.390, α_3 = −0.063, α_4 = 1.472). **g**: Gln 121 $O_{\epsilon 1}$, U_x (σ = 1.779, α_3 = −2.120, α_4 = 3.486). The remainder of the lines correspond to model pdfs parameterized with the actual moments of the distribution. For each figure, the dashed line corresponds to a Gaussian pdf, the solid line to a double-well pdf, the dotted line to a Edgeworth pdf, and the dot-dashed line to a Gram-Charlier pdf.

shown) may be due to a slight shift in the mean position which is of short duration in the trajectory. If $|\alpha_3| \to 0$) and $\alpha_4 \gg 0$, the pdf is described better as a single anharmonic pdf with a softening of the potential at large distances from the mean rather than as multiple harmonic peaks (Fig. 8f), and the Gram-Charlier and Edgeworth distributions give better results than the double-peak distribution. No cases of single peaks with significant negative kurtosis were observed.

The largest values of $|\alpha_3|$ and $|\alpha_4|$ are usually in the principal axis U_x direction and are due to multiple peaks, while the U_y and U_z directions are usually single Gaussian peaks. Correspondingly, the distributions which appear to be a single peak with relatively large skewness or kurtosis are usually in the U_y or U_z direction and have multiple peaks in the U_x direction; e.g., Gln 121 $N_{\epsilon 2}$, U_x (Fig. 8g). Since the line which connects multiple peaks may not lie exactly along a principal axis, the large anharmonicity of what appear to be single peaks in one direction may be due to multiple peaks along another direction.

Quantitative Analysis

To obtain a more quantitative characterization of the agreement between the various models and the pdf from the simulation, we calculated the goodness

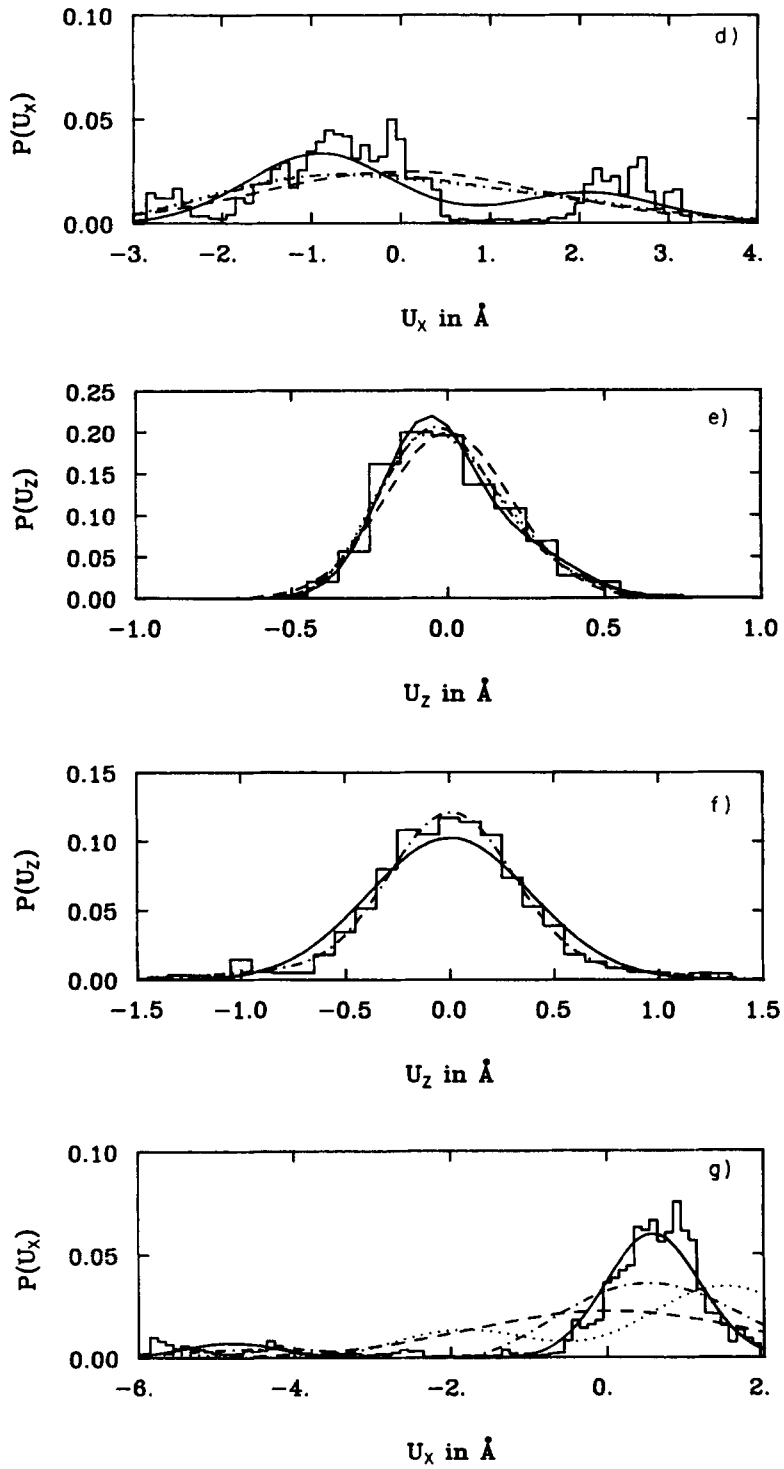

Fig. 8. Continued.

of fit (Eq. 19) of the models to the dynamics of pdf's for each direction separately. Since the number of degrees of freedom, ν, differs from each case, the quantity $\chi_\nu^2 = \chi^2/\nu$ is defined. In addition, rather than estimating N, the number of independent data points, for each atom, results are presented as χ_ν^2/N, which is independent of N (Eq. 19). A plot of the fit of the Gaussian model to the dynamics data for the atoms examined is shown as a function of σ in Figure 9 for U_x, U_y, and U_z. Although there is not a clear trend, the cases which have $\sigma < 0.5$ Å give smaller values of χ_ν^2/N. If χ_ν^2/N is 10^{-2} or less, the Gaussian distribution appears to be a good representation of the actual distribution.

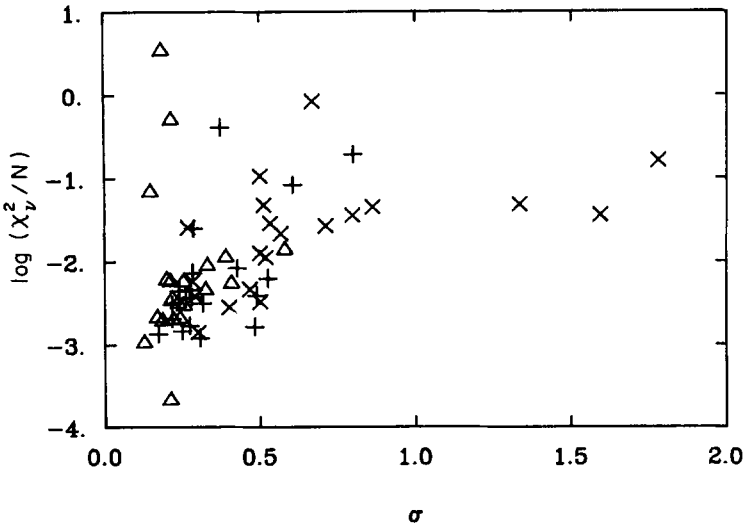

Fig. 9. χ^2_ν/N (Eq. 19) for Gaussian fit to dynamics data vs. σ. Results for 20 atoms are shown (see text).

TABLE VIII. Values of χ^2_ν/N

Atom	Interval size	$\Delta\chi^2_{GC}$/N	$\Delta\chi^2_{dp}$/N	χ^2_G/N	χ^2_{GC}/N	χ^2_{dp}/N
Glu 7 $O_{\epsilon 1}$, U_x	0.1	0.003	0.003	0.023	0.023	0.023
Gln 57 $N_{\epsilon 2}$, U_x	0.1	0.084	0.047	0.474	0.383	0.038
Gln 41 $N_{\epsilon 2}$, U_x	0.2	3.2	6.9	15.3	*	0.2
Thr 47 C_α, U_x	0.3	0.07	0.24	0.69	0.64	0.45
Gln 41 $O_{\epsilon 1}$, U_x	0.1	0.035	0.046	0.061	0.021	0.023
Gln 121 $O_{\epsilon 1}$, U_x	0.2	0.091	0.000	0.136	0.020	0.165
Gln 121 $N_{\epsilon 2}$, U_x	0.5	1.3	2.1	2.3	*	8.7

*Indicates p_c has negative values within the range of the dynamics pdf.

The values of χ^2_ν/N for Figure 8a–g are given in Table VIII. $\Delta\chi^2_{GC}$/N (Eq. 20a) and $\Delta\chi^2_{dp}$/N represent the systematic contribution to χ^2/N, the fit of the Gaussian to the dynamics. The systematic contribution is the error due to the assumption that the distribution is Gaussian when it is in fact either a Gram-Charlier or double peaked.

To examine the improvement obtained with the higher-order models over a Gaussian pdf, scatterplots of χ^2_ν/N of the fit to the dynamics pdf for the higher-order models vs. the Gaussian are shown in Figure 10a–c for the atoms examined. A plot of the Gram-Charlier fit vs. the Gaussian is shown in Figure 10a; the plot for the Edgeworth is similar so it is not shown. The Gram-Charlier and Edgeworth expansions improve χ^2_ν/N slightly for about a third of the cases. However, they do worse than the Gaussian for a third of the cases and fail for several cases because they have negative probabilities within the range of the simulation data (χ^2_ν/N set equal to 10). Since the Edgeworth expansion (Eq. 8) has negative probabilities more often, the Gram-Charlier appears to be better than the Edgeworth expansion for applications in which there may be a range in skewness. This is in agreement with the results of Zucker and Schulz.[36]

A plot of χ^2_ν/N for the double peak fit vs. the Gaussian fit to the dynamics pdf is shown in Figure 10b. There is a significant improvement for most cases and there are no negative probabilities for this model. In about a fourth of the cases the Gaussian does slightly better than the double peak model, but these generally have low χ^2_ν for the double-peak model so the double peak still gives a good fit. More importantly, the double peak mode has lower χ^2_ν for those atoms which are fit poorly by the Gaussian (i.e., have a high χ^2_ν for the Gaussian fit); i.e., the double peak fits have a value of χ^2_ν/N $\leq 10^{-2}$ for all but the U_x and U_y direction of Gln 121 $N_{\epsilon 2}$. Most of the cases which are fit equally well by the double peak model and the Gaussian model have $\xi^2 < 0.3$ and/or $\Delta\chi^2_{GC}$/N < 1/24; i.e., the distributions are close to harmonic (see Eq. 20a). However, some cases which have $\xi^2 < 0.3$ or $\Delta\chi^2_{GC}$/N < 1/24 are still considerably improved by the double-peak fit. The one case in which the double peak has a significantly larger χ^2_ν than the Gaussian is Gln 121 $N_{\epsilon 2}$. Although the dou-

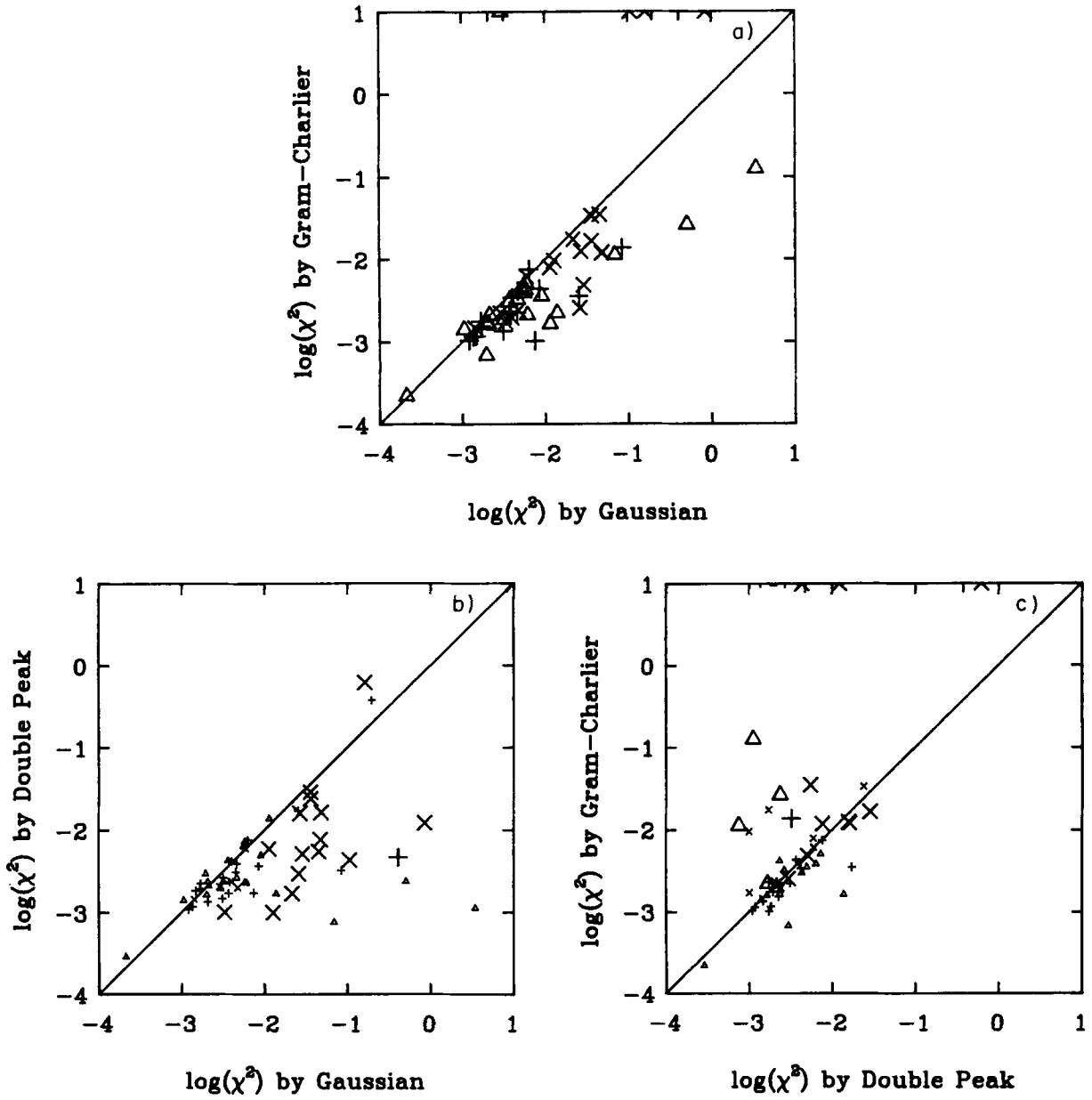

Fig. 10. Scatterplots of log (χ_ν^2/N) for the U_x, U_x, and U_z directions, represented by x, +, and Δ, respectively. **a**: Gaussian vs. Gram-Charlier. **b**: Gaussian vs. double well. **c**: Double well vs. Gram-Charlier. χ_ν^2/N set is equal to 10 if negative probabilities occur. Results for 20 atoms are shown (see text).

ble peak pdf correctly predicts the positions of the two major peaks and "looks" similar the distribution (Fig. 8g), there are several additional peaks that lead to larger χ^2.

To compare the fits resulting from the higher-order models, the χ_ν^2/N of the fits by Gram-Charlier and double peak pdfs to the dynamic data are shown in Figure 10c. Of the cases in which there are not negative regions (i.e., when α_3 and α_4 are small and $\Delta\chi^2_{GC}$/N < 1/24), the Gram-Charlier often does slightly better than the double peak. χ_ν^2/N for these two models for Figure 8a–g are also given in Table VII. Thus, the double-peak model appears to fit very anharmonic pdfs better, while the Gram-Charlier appears to be a small improvement over the double-peak model for slightly anharmonic pdfs.

We also determined the significance of the improvement by the higher-order models over the Gaussian for the 60 pdfs examined (U_x, U_y, U_z of 20 atoms) by using an F-test.[42] The test quantity is

$$F = \frac{[\chi^2_G - \chi^2_{ho}]/2}{\chi^2_{ho}/\nu}$$

where χ^2_G is the fit of a Gaussian to the dynamics χ^2_{ho} is the fit of a higher-order pdf to the dynamics, and ν is the number of degrees of freedom for the higher-order fit. Improvement is considered significant if the probability of observing such a large value of F in a random sample is less than 0.005. There were 34 pdfs, including 5 in the U_x direction, which had $\Delta\chi^2_{GC}/N < 1/24$. Of these, only His 15 $N_{\delta 1}$, U_x had a value of F indicating significant improvement by either the Gram-Charlier or the double-peak pdf over the Gaussian. Of the 26 pdfs with $\Delta\chi^2_{GC}/N > 1/24$, only Thr 47 C_{ϵ}, U_x and Gln 121 $N_{\epsilon 2}$, U_x and U_y were not improved significantly by either the Gram-Charlier or double-peak pdfs. Both Thr 47 C_α, U_x and Gln 121 $N_{\epsilon 2}$, U_x have more than two peaks (Fig. 8d,g). Although there is only one peak apparent for the U_y direction of Gln 121 $N_{\epsilon 2}$, the projection of the multiple peaks on the U_x direction onto the U_y direction apparently causes the poor fits. Of the nine pdfs which have lower χ^2 values for the Gram-Charlier fit than for the double peak, only two showed significantly smaller values for the Gram-Charlier; that is, they had values of $F = \chi^2_{dp}/\chi^2_{GC}$ corresponding to a less than 0.1 probability that a random sample would show as large values of F. Both had small values of $|\alpha_3|$ and large positive α_4; i.e., Gln 57 $N_{\epsilon 2}$, U_y had $\alpha_3 = -0.23$ and $\alpha_4 = 1.4$ and Gln 121 $O_{\epsilon 1}$, U_z had $\alpha_3 = 0.06$ and $\alpha_4 = 1.5$.

Number of Independent Data Points

Most values of χ^2_ν/N for the fit of the dynamics data by the double-peak model shown in Figure 10b, which includes the most anharmonic atoms in the protein, are between 10^{-2} and 10^{-3}. Since $\chi^2_\nu \approx 1$ when the model is a good representation of the parent pdf, this implies that there are between 100 and 1,000 independent data points for each atom. Furthermore, since the correlation time important to determining N is given by $\tau = T/N$, the values of τ (0.03 to 0.3 ps) indicate that local fluctuation randomizes the motions sufficiently even when larger time-scale correlations are present.[41]

Time Series

To provide a physical interpretation of the positional distributions of an atom, the values of the Cartesian and related internal coordinates as a function of time were examined. The time series of the Cartesian coordinates generally consists of oscillations with amplitudes corresponding to the overall $\langle u^2 \rangle^{1/2}$ (about 0.5 Å in the U_x direction) with periods of about 1 ps, which are due to combinations of dihedral angle and bond angle vibrations and faster, smaller fluctuations with amplitudes of less than 0.05 Å with periods of less than 0.05 ps, corresponding to bond vibrations. In addition, some atoms had much slower oscillations with periods greater than 4 ps which arise from more collective motions.[41]

The time series for some of the pdfs illustrated in Figure 8 are shown in Figure 11. A typical time series for an atom with a harmonic pdf, Glu 7 $O_{\epsilon 1}$, U_x, is shown in Figure 11a. Multiple peaks in the positional pdf (Fig. 8b,d) indicate conformational transitions, usually due to changes in dihedral angles near the atom, which occur for main-chain as well as sidechain atoms. Within the 30 ps trajectory such dihedral transitions generally occur only once or twice, if at all, for a given dihedral. The transitions in the backbone angles near Thr 47 C_α, U_x (Fig. 11b) are the cause of the multiple peaks of the pdf illustrated in Figure 8d whereas the double peaks for Gln 57 $N_{\epsilon 2}$, U_x are due to a slight shift in the backbone position which occurs twice during the simulation (Fig. 11c) since there are no transitions apparent in the dihedrals near residue 57. For Gln 41 $N_{\epsilon 2}$, there is a single larger fluctuation in the U_x direction lasting 2 ps (Fig. 11d). This is three to six times larger in amplitude than the 1-ps fluctuations of the rest of the time series and skews the pdf so that $|\alpha_3|$ and α_4 are large (Fig. 8c). The probability of such a fluctuation (~4–5 σ) in a Gaussian pdf for a given atom is on the order of 10^{-5}. The small probability indicates that the large fluctuation may be due to a softening of the potential to large fluctuations rather than to an improbable large fluctuation in a Gaussian potential.

Comparison With Other Simulations

The trends in α_3 and α_4 presented here for lysozyme are mostly in agreement with those found by Mao et al.[26] in their study of cytochrome c. The major difference is in the behavior of α_4. They find a significant decrease in α_4 with increasing distance from the geometric center, which is not seen here. Lysozyme, in contrast to cytochrome c, is not close to being spherical or even ellipsoidal since it has an active site cleft which divides the molecule into two lobes. However, this trend is not found for α_4 vs. distance relative to the center of each lobe. Overall, the distinction between exterior and interior atoms is not as clear for lysozyme as in the case of cytochrome c, although there is a definite trend of increasing σ with distance from the center for lysozyme.[27] Mao et al. find that while α_4 increases with distance from the backbone up to the γ-atoms, it decreases for δ- and ϵ-atoms while the results here indicate that α_4 increases up to the ϵ-atoms.

More importantly, the present analysis leads to an interpretation of the degree of anharmonicity that is different from that of Mao et al.[26] They classified the significantly anharmonic atoms (i.e., those with $|\alpha_3| > 0.5$ and/or $|\alpha_4| > 0.5$ in the direction of the largest principal axis) into ones with negative and positive kurtosis which they call classes I and II, respectively. Based on the correlation of α_3 and α_4, class I was said to involve symmetric square-well potentials and class II was said to involve potentials that are asymmetric due to multiple minima. However, the analysis here indicates that when there are multiple peaks, the kurtosis is dependent on the relative asymmetry and

separation of the peaks; that is, two equal peaks have negative kurtosis whereas two unequal peaks may have positive, negative, or zero kurtosis (Eq. 15c). Thus, the sign of the kurtosis does not indicate the number of peaks and both class I and class II may be described by multiple peaks. Figure 8b is an example of two well-defined peaks with large negative kurtosis. In fact, the plot given by Mao et al. of an atomic pdf typifying class I (their Fig. 3b) may be described as two peaks within the limits of error of the simulation results.

In a different approach, Levy et al.[11] studied the temperature dependence of the atomic fluctuations of an α-helix in a molecular dynamics simulation. The α-helical results showed that the distributions were approximately harmonic at each temperature but the force constants were a function of temperature.[43] This corresponds to the quasi-harmonic model which has been suggested for the analysis of protein fluctuations.[25]

Comparison With Crystal Studies

Frauenfelder et al.[21] used the temperature dependence of the temperature factors of myoglobin in an attempt to determine the shape of the local atomic potential. They assumed that the local potential w(**r**) can be approximated by a temperature-independent isotropic potential (Eq. 9), where ν is a measure of the anharmonicity. The temperature dependence of the fluctuations is then given by Eq. 11a, i.e., $\langle u_i^2 \rangle$ varies as $T^{2\nu}$. They also assumed that the temperature factors are related to the fluctuations by $B = 8\pi\langle u_i^2 \rangle = 8\pi \langle u_r^2 \rangle/3$; since this is a harmonic approximation, it breaks down for very anharmonic fluctuations. Applying these assumptions to the experimental isotropic harmonic B-factors for myoglobin at four temperatures (220°K, 250°K, 275°K, 300°K), they suggested that there is a "condensed" core in which $\langle u_r^2 \rangle$ is less than 0.04 Å^2 and ν is greater than 2 and a "semiliquid" outer shell in which $\langle u_r^2 \rangle$ ranges from 0.04 to 0.25 Å^2 and ν is less than 2.

We find instead that the local atomic fluctuations are close to harmonic, in disagreement with the results of Frauenfelder et al. One source of the different results is in the form of the potential assumed. The assumption by Frauenfelder et al. that the potential is isotropic should not cause significant error for a single-well potential since the temperature dependence of $\langle u_r^2 \rangle$ is the same as $\langle u^2 \rangle$. If there are multiple wells the temperature dependence may be unpredictable since, at low temperatures, only one well may be accessible while, at high temperatures, there may be hopping between wells. Furthermore, the form of the potential used to describe anharmonicity is not in agreement with our findings. Results from dynamics pdfs indicate that if $|\alpha_4| < 1$ and $\alpha_3 = 0$ (i.e., $\Delta\chi^2_{GC} < N/24$), the pdf is harmonic whereas a potential with $\nu \to 0$ is far from harmonic. For $\nu \geqslant 1$, the potential is an inverse power curve which has an unphysical cusp at the minimum. Furthermore, the error in ν for Frauenfelder et al.'s measurements is inversely proportional to $\langle u_r^2 \rangle$ so we expect greater error for small $\langle u_r^2 \rangle$ which are associated with large ν.

Other possible soures of error in Frauenfelder et al.'s findings are that quantum and thermal expansion effects were neglected in the analysis. The measurements of Frauenfelder et al. cover the range from 220 to 300°K and the simulation study of the temperature dependence of the atomic fluctuations in an α-helix have shown that quantum effects are not significant above 50°K.[11] A more likely source of the nonlinear temperature dependence is thermal expansion. An approximate expression for including the effects of thermal expansion in the fluctuations is [7]

$$\langle u^2 \rangle = \langle u^2 \rangle_h (1 + 2\gamma_G \chi T)$$

where $\langle u^2 \rangle_h$ is from the harmonic approximation and γ_G is the Grüneisen constant and χ is the volume coefficient of expansion.

In a later study of myoglobin in which a structure at 80°K was added to the data of Frauenfelder et al., Hartmann et al.[44] reported that 30% of the residues were harmonic, 33% were quasi-harmonic, and the rest were anharmonic. This is in agreement with our results which are for a single temperature since a quasi-harmonic potential appears harmonic at a given temperature, and so we cannot distinguish between harmonic and quasi-harmonic.

Finally, Frauenfelder et al. noted possible alternate sidechain conformations in the myoglobin crystal structure, especially at 220°K, as well as some cases in which the mean position of atoms appeared to shift with temperature. This is in agreement with our finding of multiple peak pdfs since the first of these observations corresponds to multiple peaks in the pdf while the second could also correspond to multiple peaks with differing populations at different temperatures.

DISCUSSION AND CONCLUSION

The anisotropy and anharmonicity of the atomic motions in a molecular dynamics simulation of hen egg-white lysozyme have been analyzed. The anisotropy results show that the positional fluctuations are highly anisotropic (only for 3.1% of the atoms is the anisotropy, A_1, less than 0.2) and that sidechains have higher anisotropies than the main-chain atoms, in agreement with previous studies.[13–16] The anharmonicity results show that most atoms have small values of the skewness $|\alpha_3|$ and kurtosis $|\alpha_4|$; i.e., 66% of the backbone and 50% of the sidechain atoms have both $|\alpha_3| < 0.5$ and $|\alpha_4| < 1$ in all three of the prinicpal axis directions, also in agreement with previous studies.[15,16,26] However, in contrast to the earlier work, we find that it is important to determine not only the moments α_3 and α_4 but also the type of

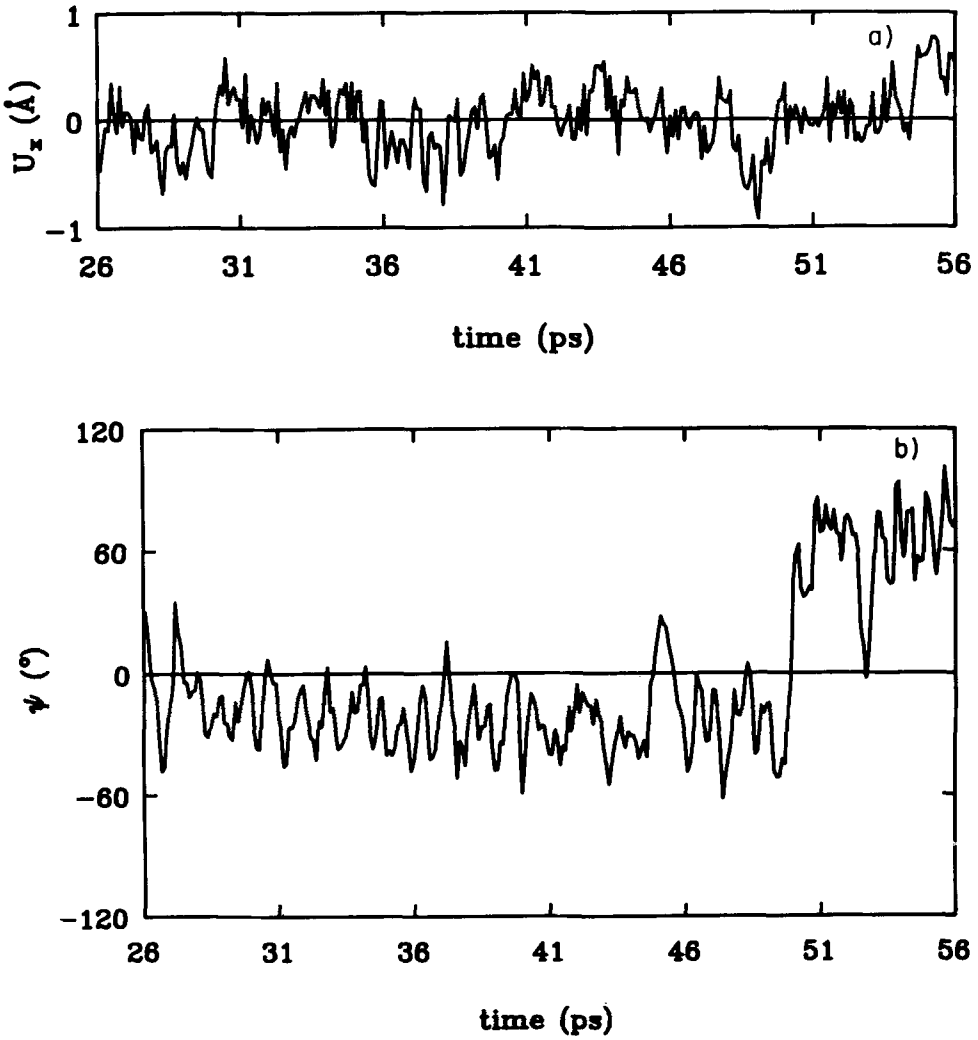

Fig. 11. Time series for (a) Glu 7 $O_{\epsilon 1}$, U_x; (b) Thr 47, ψ dihedral; (c) Gln 57 $N_{\epsilon 2}$, U_x; (d) Gln 41 $N_{\epsilon 2}$ U_x.

deviation of the dynamics pdf from a harmonic pdf to define what constitutes significant anharmonicity; that is, the anharmonicity is best related to a large systematic value of the χ^2 deviation between the dynamics and a harmonic pdf. From examination of sample dynamics pdfs, atoms with small values of $|\alpha_3|$ and $|\alpha_4|$ have pdfs described well by Gaussian pdfs, those with intermediate values by Gram-Charlie expansions, and those with large values by double-peak distributions. Furthermore, the largest principal axis tends to lie along the line joining double peaks, if present, and the two other principal axis directions tend to show harmonic distributions. Thus, the 59% backbone and 41% sidechain atoms with $\Delta\chi^2_{GC}/N < 1/24$ in all three directions are considered to be harmonic. Another 31% of the backbone and 41% of the sidechain atoms have $\xi^2 > 0.3$ and thus may be described as double wells.

We have examined how well we can predict the shape of atomic pdfs calculated from dynamics trajectories assuming various models for the pdf with parameters obtained from the first four moments of the fluctuations from the dynamics trajectory. The expressions used for the pdfs and the potentials all depend only on these four independent parameters. Of the three higher-order approximations (Gram-Charlier, Edgeworth and double peak), the double-peak pdf gives the largest improvement in the fit for atoms with distributions which are fit poorly by a Gaussian (Fig. 10b). Where the double peak does worse than the Gram-Charlier or Edgeworth, it usually is only slightly worse and the value of χ^2_ν is small for all of the models, including the Gaussian. This is because the Gram-Charlier and Edgeworth expansions are best suited for small perturbations relative to a harmonic distribution whereas the double peak is best for a large perturbation, i.e., two peaks. Furthermore, the low values of χ^2_ν found generally in fitting the dynamics by the double-peak model indicate that we can accurately predict the shape of the

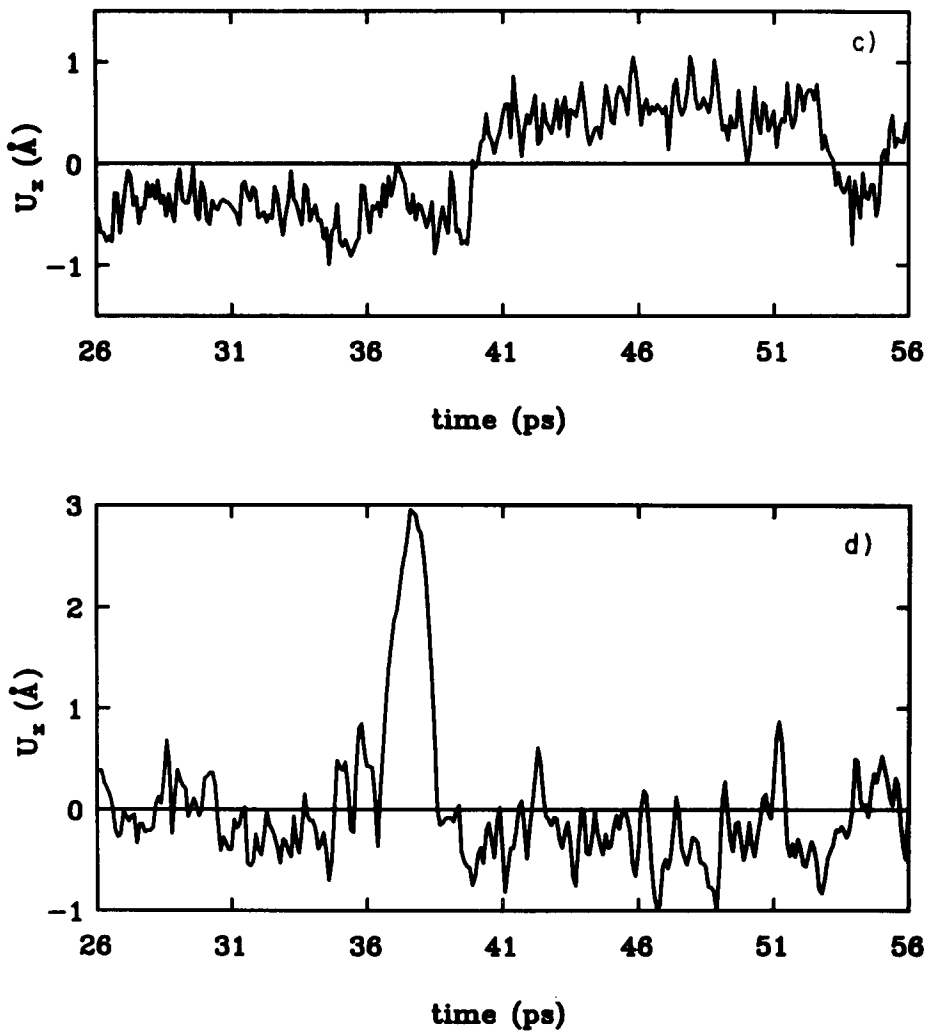

Fig. 11. Continued.

dynamics pdfs given only the values of m, σ, α_3, and α_4. These results suggest that the best approach to improving the description of the pdf in proteins, relative to the standard fit by isotropic Gaussians, is to model pdfs by Gaussians for small $|\alpha_3|$ and $|\alpha_4|$ and double-peak pdfs for large $|\alpha_3|$ and $|\alpha_4|$. The Gram-Charlier and Edgeworth may be considered as higher-order corrections to this model and might best be used to improve the double-peak fits if sufficient data were available. In experiments there is no direct measure of the form of the distribution, although multiple occupancy has been used to explain some x-ray[21] and NMR[45] results. This makes it particularly important to have the simulation information to guide the interpretation of experimental measurements.

The present findings have several important implications. First, the best model for including anharmonic effects in X-ray crystal structure refinement of proteins is the one that includes multiple occupancy with Gaussian distributions. In a separate paper,[28] we analyze anharmonicity from this point of view. Second, we can predict positions and relative weights of double peaks in the simulated distributions given the first four moments of the atomic fluctuations. The values for the larger of the two peaks are expected to be closer to the positions and temperature factors from X-ray crystal refinement in the isotropic, harmonic approximation. This information can be used as a starting point in X-ray crystal refinement to reduce the number of parameters needed to include some of the major anharmonic effects by introducing multiple occupancy factors only for atoms predicted by dynamics to have second peaks with significant weight and separation relative to the first peak. By considering two isotropic harmonic peaks with equal force constants, anharmonic as well as anisotropic effects can be included by adding parameters for only the position and relative weight of the second peak. This assumption is reasonable, in particular for atoms in the interior of collective groups in which the multiple minima are due to the motions of the entire group.

We emphasize the difference between anharmonicity due to multiple wells vs. skewness and/or kurtosis

of a single well with one minimum. In the multiple-well picture, the atom spends time localized about harmonic well minimum and then makes a transition to another position. Consequently, the nonzero skewness and kurtosis are due to the relative populations and separation of the peaks. For small time segments, the pdf will appear harmonic. This behavior is analogous to caging effects in liquids and to jump diffusion in solids. It is a very different physical picture from that which follows from the descriptions such in Mao et al.[26] and Frauenfelder et al.[21] They suggest that in any given time interval, the trajectory of the atom corresponds to an anharmonic probability distribution. Of the 20 atoms in the simulation which exhibited the largest anharmonicity, none was adequately described by a model corresponding to diffuse motion in a square-well potential; this would lead to a single broad flat peak with negative kurtosis. The cases which corresponded to a single anharmonic peak had positive kurtosis. The peaks appeared Gaussian near the maximum but had large values in one or both wings than a Gaussian. Here again the atoms move in a well which is harmonic near the center but is softer for large displacements due to the fluctuating environment. In a longer simulation, such atoms might make a transition to a neighboring well.[46]

In summary, for much of the protein, particularly the main-chain atoms, the harmonic approximation is valid. Thus, normal mode calculations with force constants taken from the potentials of mean force are expected to be a good first approximation and the use of harmonic constraints in dynamics simulations of localized regions[47] appears to be justified. The most significant anharmonic effect can be included by adding a second peak to the pdf, which has a simple expression for the potential as well. Furthermore, we can use the first four moments to estimate the positions and weights of double peaks in the simulated distribution without actually calculating the pdf, which may be of use in comparing results with X-ray crystal refinement of proteins as well as for a starting point for including multiple occupancy in the refinement of a protein. The positions and weights of the double peaks from the moments are good estimates of the actual values in the simulated distribution, although especially the weights are not quantitative for an actual protein due to inaccuracies of the potential (even when solvent and crystal effects are included) and the short time scale of current simulations.

ACKNOWLEDGMENTS

We wish to thank Charles L. Brooks, III, John Kuriyan, and B. Montgomery Pettitt for many helpful discussions. Research was supported in part by a grant from the National Institutes of Health.

REFERENCES

1. Karplus, M., McCammon, J.A.: Dynamics of proteins: Elements and function. Annu. Rev. Biochem. 53:263–300, 1983.
2. Gurd, F.R.N., Rothgeb, T.M.: Motions in proteins. Adv. Protein Chem. 33:73–165, 1979.
3. Debrunner, P.G., Frauenfelder, H.: Dynamics of proteins. Annu. Rev. Phys. Chem. 33:283–299, 1982.
4. Petsko, G.A., Ringe, D.: Fluctuations in protein structure from X-ray diffraction. Ann. Rev. Biophys. Bioeng. 13:331–371, 1984.
5. Karplus, M., McCammon, J.A.: The internal dynamics of globular proteins. CRC Crit. Rev. Biochem. 9:293–349, 1981.
6. Levitt, M.: Molecular dynamics of native proteins I. Computer simulation of trajectories. J. Mol. Biol. 168:595–620, 1983. Molecular dynamics of native proteins II. Analysis and nature of motion. J. Mol. Biol. 168:621–657, 1983.
7. Willis, B.T.M., Pryor, A.W.: "Thermal Vibrations in Crystallography." Cambridge: Cambridge University Press, 1975.
8. Boutin, H., Yip, S.: "Molecular Spectroscopy With Neutrons." Cambridge, Mass.: M.I.T. Press, 1968.
9. de Gennes, P.G., Papoular, M.: Vibrations de basse fréquence dans certaines structures biologiques. In: "Polarization, Matériere et Rayonnement." La Société Française de Physique, ed. Paris: Presses Universitaire de France, 1969:243–258.
10. Go, N.: Shape of the conformational energy surface near the global minimum and low-frequency vibrations in the native conformation of globular proteins. Biopolymers 17:1373–1379, 1978.
11. Levy, R.M., Perahia, D., Karplus, M.: Molecular dynamics of an α-helical polypeptide: temperature dependence and deviation from harmonic behavior. Proc. Natl. Acad. Sci. USA 79:1346–1350, 1982.
12. Brooks, B.R., Karplus, M.: Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc. Natl. Acad. Sci. USA 80:6571–6575, 1983.
13. Karplus, M., McCammon, J.A.: Protein structural fluctuations during a period of 100 ps. Nature (Lond.) 277:578, 1979.
14. Northrup, S.H., Pear, M.R., Morgan, J.D., McCammon, J.A., Karplus, M.: Molecular dynamics of ferrocytochrome *c*: Magnitude and anisotropy of atomic displacements. J. Mol. Biol. 153:1087–1109, 1981.
15. van Gunsteren, W.F., Karplus, M.: Protein dynamics in solution and in a crystalline environment: A molecular dynamics study. Biochemistry 21:2259–2274, 1982a.
16. van Gunsteren, W.F., Karplus, M.: Effect of constraints on the dynamics of macromolecules. Macromolecules 15:1528–1544, 1982b.
17. Artymiuk, P.J., Blake, C.C.F., Grace, D.E.P., Oatley, S.J., Phillips, D.C., Sternberg, M.J.E.: Crystallographic studies of the dynamic properties of lysozyme. Nature 280:563–568, 1979.
18. Konnert, J.H., Hendrickson, W.A.: A restrained-parameter thermal-factor refinement procedure. Acta Cryst. A36:344–350, 1980.
19. Yu, H., Karplus, M., Hendrickson, W.A.: Restraints in temperature-factor refinement for macromolecules: an evaluation by molecular dynamics. Acta Cryst. B41:191–201, 1985.
20. Haneef, I., Glover, I.D., Tickle, I.J., Moss, D.S., Pitts, J.E., Wood, S.P., Blundell, T.L., Hermans, J., van Gunsteren, W.F.: The Dynamics of Pancreatic Polypeptides: A Comparison of X-Ray Anisotropic Refinement at 0.98Å Resolution. Molecular Dynamic and Normal Mode Analysis. In: "Molecular Dynamics and Protein Structure." Hermans, J., ed. Western Springs, IL: Polycrystal Book Service, P.O. Box 27, 1985:85–91.
21. Frauenfelder, H., Petsko, G.A., Tsernoglou, D.: Temperature dependent x-ray diffraction as a probe of protein structural dynamics. Nature 280:558–563, 1979.
22. Lifshitz, I.M.: Some problems of the statistical theory of biopolymers. Soviet Phys. JETP 28:1280–1286, 1969.
23. Klein, M.L., Venables, J.A. (eds.): "Rare Gas Solids," Vol. I. London: Academic Press, 1975.
24. Klein, M.L., Venables, J.A. (eds.): "Rare Gas Solids," Vol. II. London: Academic Press, 1977.

25. Karplus, M., Kushick, J.N.: Method for estimating the configurational entropy of macromolecules. Macromolecules 14:325–332, 1981.
26. Mao, B., Pear, M.R., McCammon, J.A.: Molecular dynamics of ferrocytochrome *c*: Anharmonicity of atomic displacements. Biopolymers 21:1979–1989, 1982.
27. Ichiye, T., Olafson, B.D., Swaminathan, S., Karplus, M.: Structure and internal mobility of proteins: A molecular dynamics study of hen egg white lysozyme. Biopolymers 25:1909–1937, 1986.
28. Ichiye, T., Karplus, M.: Anisotropy and anharmonicity of atomic fluctuations in proteins: analysis of a molecular dynamics simulation. Biochemistry, 1987.
29. McQuarrie, D.A.: "Statistical Mechanics." New York: Harper & Row, 1976.
30. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M.: CHARMM: A program for macromolecular energy minimization and dynamics calcu-
31. Kendall, M.G., Stuart, A.: "The Advanced Theory of Statistics," Vol. I. London: Charles Griffin, 1977.
32. Ichiye, T.: The Internal Dynamics of Proteins. Ph.D. Thesis, Harvard University, Cambridge, Mass., 1985.
33. Cramer, H.: "Mathematical Methods of Statistics." Princeton: Princeton University Press, 1946. See also Cramer, H. "The Elements of Probability Theory and Some of its Applications." Huntington, NY: Robert E. Krieger Publishing Co., Inc., 1973.
34. Johnson, C.K.: Addition of higher cumulants to the crystallographic structure-factor equation: a generalized treatment for thermal-motion effects. Acta Cryst. A25:187–194, 1969.
35. Johnson, C.K., Levy, H.A.: Thermal-motion analysis using Bragg diffraction data. In: "International Tables for X-ray Crystallography," Vol. 4. Birmingham: Kynoch Press, 1974:313–335.
36. Zucker, U.H., Schulz, H.: Statistical approaches for the treatment of anharmonic motion in crystals. I. A comparison of the most frequently used formalisms of anharmonic thermal vibrations. Acta Cryst. A38:563–568, 1982.
37. Willis, B.T.M.: Lattice vibrations and the accurate determination of structure factors for the elastic scattering of x-rays and neutrons. Acta Cryst. A25:277–300, 1969.
38. Kittel, C.: "Introduction to Solid State Physics," 5th Ed. London: J. Wiley & Sons, Inc., 1976.
39. Noguti, T., Go, N.: Collective variable description of small-amplitude conformational fluctuations in a globular protein. Nature 296:776–778, 1982.
40. Nadler, W., Schulten, K.: Theory of Mössbauer spectra of proteins fluctuating between conformational substates. Proc. Natl. Acad. Sci. USA 81:5719–5723, 1984.
41. Swaminathan, S., Ichiye, T., van Gunsteren, W.F., Karplus, M.: Time dependence of atomic fluctuations in proteins: analysis of local and collective motions in bovine pancreatic trypsin inhibitor. Biochemistry 21:5230–5241, 1982.
42. Bevington, P.R.: "Data Reduction and Error Analysis for the Physical Sciences." New York: McGraw-Hill Book Co., 1969.
43. Perahia, D., Levy, R.M., Karplus, M.: Motions of an alpha helical polypeptide: Comparison of molecular and harmonic dynamics (to be published).
44. Hartmann, H., Parak, F., Steigemann, W., Petsko, G.A., Ringe Ponzi, D., Frauenfelder, H.: Conformational substates in a protein: Structure and dynamics of metmyoglobin at 80 K. Proc. Natl. Acad. Sci. USA 79:4967–4971, 1982.
45. Hoch, J.C., Dobson, C.M., Karplus, M.: Vicinal Coupling Constants and Protein Dynamics. Biochemistry 24:3831–3841, 1985.
46. Elber, R., Karplus, M.: Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. Science 235:318–321, 1986.
47. Brooks, C.L. III, Brünger, A., Karplus, M.: Active site dynamics in protein molecules: a stochastic boundary molecular-dynamics approach. Biopolymers 24:843–865, 1985.

APPENDIX

In this appendix we define criteria for the values of α_3 and α_4 that correspond to anharmonic behavior. A dynamics pdf, p_o, is a sample of the true distribution (i.e., the exact equilibrium distribution for the potential used in the dynamics) with a pdf, ϕ_o, so that the difference between p_o and ϕ_o is due to statistical error, with no systematic deviations. We can model the true pdf with a function, p_c, such as described in the Probability Density Functions section, parameterized with the dynamics values of the moments. If the model is incorrect, the χ^2 difference between p_o and p_c (Eq. 20) has a systematic contribution due to $\phi_o - p_c$, as well as statistical error, so that $\chi^2 \gg \nu$. We define the test quantity for anharmonicity, $\Delta\chi^2$, as the systematic contribution to χ^2 of the difference between the true distribution, ϕ_o, and a Gaussian, $p_c = p_G$. If the true pdf has the functional form of one of the higher-order pdfs, ϕ_o', the systematic contribution is

$$\Delta\chi^2 = N \int_{-\infty}^{\infty} \frac{[\phi_o'(u) - p_G(u)]^2}{P_G(u)} du \qquad (A1)$$

If the true distribution has a Gram-Charlier or Edgeworth pdf, then

$$\Delta\chi^2_{GC,E} = N\left[\frac{{\alpha_3}^2}{3!} + \frac{{\alpha_4}^2}{4!} + \frac{5{\alpha_3}^4}{3!^2}\right] \qquad (A2)$$

where the final term is present only for the Edgeworth pdf. If instead the true distribution has a double peak pdf,

$$\Delta\chi^2_{dp} = \frac{N}{(1-\xi^4)^{1/2}}\left[{w_-}^2 \exp\left(\frac{\xi_-^2}{1+\xi^2}\right) + {w_+}^2 \exp\left(\frac{\xi_+^2}{1+\xi^2}\right) + 2w_+w_- \exp\left(-\frac{\xi^2[2+\xi_+^2+\xi_-^2]}{2(1-\xi^4)}\right)\right] - N \qquad (A3)$$

where $\xi_\pm = u_\pm/\sigma$ and $\xi^2 = u_+u_-/\sigma^2 = \xi_+\xi_-$.

The value of the integral in Eq. A1 is a measure of the area of deviation between the Gaussian curve and the parent pdf curve. We define a value of $\int[\phi_o' - p_G]2/p_G\,du = 1/24$ as a significant deviation relative to the total area, 1, under either curve; i.e., $\Delta\chi^2 > N/24$ is considered significant. If the parent distribution has a Gram-Charlier pdf, $\Delta\chi^2_{GC} = N/24$ when $\alpha_3 = 0.5$ and $\alpha_4 = 0$ or when $\alpha_3 = 0$ and $\alpha_4 = 1$. Since the Gram-Charlier pdf should describe small perturbations to a harmonic pdf well, a pdf with $\Delta\chi^2_{GC} < N/24$ is considered to be harmonic. Furthermore, since in this simulation $N \approx 150$, if $\Delta\chi^2_{GC} \lessapprox N/24 = 6.25$, it

is within the range of the standard deviation of χ^2, $(2\nu)^{1/2}$;[31] this was in the range 3.5 to 6.8 since $\nu = k-3$ varies from 6 to 23. Consequently, if $\Delta\chi^2_{GC} < N/24$, the systematic contribution due to deviation from harmonicity is less than the standard deviation of χ^2 for a 30-ps simulation and thus cannot be distinguished from statistical noise.

If the value of $\Delta\chi^2_{dp}$ is small, i.e., $\Delta\chi^2_{dp} < N/24$, but the values of α_3 and α_4 are high so that $\Delta\chi^2_{GC} > N/24$, the actual distribution is assumed to have a single anharmonic peak. However, when the actual distribution is a single skewed peak, i.e., $\alpha_3 \neq 0$, the double-peak assumption may predict a spurious small peak at a large distance ($> \pm 3\sigma$) which gives a large contribution to the integral (Eq. A1) because of the weighting by $[p_G(u)]^{-1}$. Thus, while a small $\Delta\chi^2_{dp}$ indicates single peaks, a large $\Delta\chi^2_{dp}$ does not necessarily indicate double peaks. Alternatively, we can consider another quantity, $\xi^2 = u_+u_-/\sigma^2 = 1-\sigma_0^2/\sigma^2$ (i.e., the fractional difference between σ and σ_0) as a measure of double-peak character. It has a maximum value of 1 and approaches 0 either when distribution is very asymmetric so that one of the peaks is approximately equal in size and location to the single Gaussian with the same σ while the other has very low occupancy and is far from the mean or when the separation between the peaks is much less than their width. In either case, the probability density within a range of about $\pm 3\sigma$ of the mean becomes approximately a single Gaussian. Therefore, the pdf is defined as a single peak if $\xi^2 < 0.3$ and as a double peak if $\xi^2 > 0.3$ and $\Delta\chi^2_{GC} > N/24$.