Milwaukee for the use of an SLM 8000 spectrofluorometer.

Registry No. L-Glutamine, 56-85-9.

REFERENCES

Alexander, S. S., Jr., Colonna, G., & Edelhoch, H. (1979) *J. Biol. Chem. 254*, 1501–1505.
Chen, R. F. (1968) *Anal. Biochem. 25*, 412–416.
Engvall, E., & Ruoslahti, E. (1977) *Int. J. Cancer 20*, 1–5.
Erickson, H. P., & Carrell, N. A. (1983) *J. Biol. Chem. 258*, 14539–14544.
Fairclough, R. H., & Cantor, C. R. (1978) *Methods Enzymol. 48*, 347–379.
Hermans, J. (1985) in *Hematology* (McDonagh, J., Ed.) Vol. 5, Chapter 4, Dekker, New York.
Jullien, M., & Garel, J.-R. (1983) *Biochemistry 22*, 3829–3836.
Lai, C.-S., Tooney, N. M., & Ankel, E. G. (1984) *Biochemistry 23*, 6393–6397.
Lai, C.-S., Homandberg, G., Miziorko, H., & Wolff, C. (1987) *Biopolymers 26*, 1381–1389.
Lakowicz, J. R. (1983) *Principles of Fluorescence Spectroscopy*, Plenum, New York.
McDonagh, J. (1985) *Hematology*, Vol. 5, Dekker, New York.
McDonagh, R. P., McDonagh, J., Petersen, T. E., Thøgersen, H. C., Skorstengaard, K., Sottrup-Jensen, L., Magnusson, S., Dell, A., & Morris, H. R. (1981) *FEBS Lett. 127*, 174–178.
Mosesson, M. W., Chen, A. B., & Huseby, R. M. (1975) *Biochim. Biophys. Acta 386*, 509–524.
Mosher, D. F., Shad, P. E., & Vann, J. M. (1980) *J. Biol. Chem. 255*, 1181–1188.
Odermatt, E., Engel, J., Richter, H., & Hörmann, H. (1982) *J. Mol. Biol. 159*, 109–123.
Robinson, R. M., & Hermans, J. (1984) *Biochem. Biophys. Res. Commun. 124*, 718–725.
Scott, T. G., Spencer, R. D., Leonard, N. G., & Weber, G. (1970) *J. Am. Chem. Soc. 92*, 687–695.
Skorstengaard, K., Jensen, M. S., Sahl, P., Petersen, T. E., & Magnusson, S. (1986) *Eur. J. Biochem. 161*, 441–453.
Stryer, L., & Haugland, R. P. (1967) *Proc. Natl. Acad. Sci. U.S.A. 58*, 719–726.
Williams, E. C., Janmey, P. A., Ferry, J. D., & Mosher, D. F. (1982) *J.* Biol. *Chem. 257*, 14973–14978.
Wu, C. W., & Stryer, L. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1104–1108.
Yamada, K. M. (1983) *Annu. Rev. Biochem. 52*, 761–799.

Anisotropy and Anharmonicity of Atomic Fluctuations in Proteins: Implications for X-ray Analysis†

Toshiko Ichiye‡ and Martin Karplus*
Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138
Received June 3, 1986; Revised Manuscript Received October 20, 1987

ABSTRACT: The effects of anisotropy and anharmonicity of the atomic fluctuations on the results of crystallographic refinement of proteins are examined. Atomic distribution functions from a molecular dynamics simulation for lysozyme are introduced into a real-space (electron density) refinement procedure for individual atoms. Several models for the atomic probability distributions are examined. When isotropic, harmonic motion is assumed, the largest discrepancies between the true first moments (means) and second moments (*B* factors) of the positions calculated from the dynamics and the fitted values occur for probability densities with multiple peaks. The refined mean is at the center of the largest peak, and the refined *B* factor is slightly larger than that of the largest peak, unless the distance between the peaks is small compared to the peak width. The resulting values are often significantly different from the true first and second moments of the distribution. To improve the results, alternate conformations, rather than anharmonic corrections, should be included.

A knowledge of the functional form of the probability density functions (pdfs) of the fluctuations of the atomic position is essential for determining the structures of protein crystals by X-ray diffraction (Willis & Pryor, 1975). Moreover, the nature of the fluctuations is of considerable interest because of their possible role in protein function (Karplus & McCammon, 1981, 1983). Also, it has been suggested recently that the antigenicity and atomic mobility of proteins may be related (Westhof et al., 1984; Tainer et al., 1984), although alternative explanations of the observed correlations have been given (Novotný et al., 1986).

For the refinement of protein crystal diffraction data (Hendrickson, 1985), it is assumed in most cases that the atomic displacements are harmonic and isotropic, i.e., that the atomic fluctuations obey a three-dimensional isotropic Gaussian distribution. Anisotropic and anharmonic effects have been introduced in the refinement of small molecules (Willis & Pryor, 1975), such as Li_3N, which usually have temperature factors corresponding to root-mean-square (rms) displacements, $\langle u_r^2 \rangle^{1/2}$, of less than 0.3 Å at 294 K (Zucker & Schultz, 1982). The motion of an atom in these molecules is generally restricted to a single minimum in the local potential, and anisotropic effects can be included by the use of an anisotropic three-dimensional Gaussian pdf. There are several possible models for anharmonic distributions, which have been compared by Zucker and Schulz (1982). They considered the performance and implementation of an expansion of the potential in a power series; Edgeworth and Gram–Charlier expansions, which are expansions of a harmonic distribution in terms of Hermite polynomials; and the

† Supported in part by a grant from the National Science Foundation.
‡ Present address: Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720.

0006-2960/88/0427-3487$01.50/0

Table I: Model Probability Density Functions[a]

isotropic Gaussian	$p_G(\mathbf{u};\mu,V) = \sigma^{-3}(2\pi)^{-3/2}\exp(-{}^1/_2u_r^2/\sigma^2)$
anisotropic Gaussian	$p_G(\mathbf{u};\mu,V) = \lvert V\rvert^{-1/2}(2\pi)^{-3/2}\exp(-{}^1/_2\mathbf{u}^TV^{-1}\mathbf{u})$
Gram–Charlier	$p_{GC}(\mathbf{u};\mu,V,\alpha_3,\alpha_4) = p_G(\mathbf{u};\mu,V)[1 + \frac{1}{3!}\sum\kappa_{jkl}H_{jkl}(\mathbf{u}) + \frac{1}{4!}\sum\kappa_{jklm}H_{jklm}(\mathbf{u})]$
Edgeworth	$p_E(\mathbf{u};\mu,V,\alpha_3,\alpha_4) = p_G(\mathbf{u};\mu,V)[1 + \frac{1}{3!}\sum\kappa_{jkl}H_{jkl}(\mathbf{u}) + \frac{1}{4!}\sum\kappa_{jklm}H_{jklm}(\mathbf{u}) + \frac{10}{6!}\sum\kappa_{jkl}\kappa_{mno}H_{jklmno}(\mathbf{u})]$
double isotropic peaks[b]	$p_\mu(\mathbf{u};\mu,\mathbf{u}_\pm,\sigma_0) = \sigma_0^{-3}(2\pi)^{-3/2}[w_-\exp(-{}^1/_2\lvert\mathbf{u}-\mathbf{u}_-\rvert^2/\sigma_0^2) + w_+\exp(-{}^1/_2\lvert\mathbf{u}-\mathbf{u}_+\rvert^2/\sigma_0^2)]$
double anisotropic peaks[b]	$p_\mu(\mathbf{u};\mu,\mathbf{u}_\pm,V_0) = \lvert V_0\rvert^{-1/2}(2\pi)^{-3/2}[w_-\exp[-{}^1/_2(\mathbf{u}-\mathbf{u}_-)^TV_0^{-1}(\mathbf{u}-\mathbf{u}_-)] + w_+\exp[2^1/_2(\mathbf{u}-\mathbf{u}_+)^TV_0^{-1}(\mathbf{u}-\mathbf{u}_+)]]$

[a] The transpose of a matrix is denoted by a superscript T. [b] In the principal axis system, $\sigma_x^2 = \sigma_0^2{}_x + u_+u_-$, $\alpha_3 = u_+u_-(u_+ - u_-)/\sigma_x^3$, and $\alpha_4 = -2u_+^2u_-^2/\sigma_x^4 + \alpha_3^2\sigma_x^2/u_+u_-$.

α formalism or quasi-orthogonal expansion (Johnson 1980; Cava et al., 1980), in which a parameter α is introduced to reduce the correlations between odd-rank and between even-rank moments. Good results were obtained by using the Gram–Charlier expansion up to sixth-rank tensors so that the perturbation approach embodied in these expansions appears to be satisfactory. However, even in small peptides and some other organic molecules, exceptional structures show multiple sites that cannot be treated by such perturbation methods.

Proteins are considerably more flexible than small molecules. Temperature factors based on the assumption of isotropic harmonic motion yield values of $\langle u_r^2\rangle^{1/2}$ as large as about 1.1 Å for myoglobin (Frauenfelder et al., 1979) and about 2 Å for lysozyme (Artymiuk et al., 1979) at 300 K; the estimated lattice disorder contribution in the case of myoglobin is 0.37 Å. Simulation results are similar; e.g., the average side chain value of $\langle u_r^2\rangle^{1/2}$ has a maximum value of 1.6 Å for cytochrome *c* (Northrup et al., 1981), about 2.5 Å for lysozyme (Ichiye et al., 1986), and about 1.2 Å for bovine pancreatic trypsin inhibitor (van Gunsteren & Karplus, 1982a,b; Levitt, 1983a,b). Thus, multiple sites are likely to be present which may not be adequately described as perturbations to Gaussian distributions.

Some studies have 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 (Karplus & McCammon, 1979; Northrup et al., 1981; van Gunsteren & Karplus, 1982a,b; Ichiye & Karplus, 1987) and in a few X-ray studies (Artymiuk et al., 1979; Konnert & Hendrickson, 1980; Haneef et al., 1985). Theoretical studies have shown that the distributions tend to be highly anisotropic, particularly for side chain atoms. Further, it has been demonstrated that the orientation of the anisotropy tensors found in molecular dynamics simulations does not have a simple relation to the static X-ray structure (Yu et al., 1985).

The anharmonicity has been analyzed by examining the temperature dependence of atomic fluctuations in crystallographic analyses of myoglobin (Frauenfelder et al., 1979; Hartmann et al., 1982) and in a theoretical study of α-helix (Levy et al., 1982). Another approach has been to study the third and fourth moments of the atomic distributions in molecular dynamics simulations of proteins (Mao et al., 1982; van Gunsteren & Karplus, 1982a,b). In a detailed analysis of the anharmonicity of local atomic pdfs in a lysozyme simulation (Ichiye & Karplus, 1987), it has been shown that 59% of the backbone and 41% of the side chain atoms are close to harmonic in all three principal axis directions; an additional 9% of the backbone and 15% of the side chain atoms are close to harmonic in the largest principal axis direction although not in both of the two smaller principal axis directions. Of the remaining atoms, 31% of the backbone and 41% of the side chain atoms can be described as multiple harmonic peaks. Thus, very few atoms with significant anharmonicity are described well as single anharmonic peaks. Although simulations are known to yield only approximate results for systems as large as proteins, it is likely that the general character of the distributions is correct. Furthermore, in a study of myoglobin between 220 and 300 K, Frauenfelder et al. (1979) noted possible alternate side chain conformations, especially at 220 K, and in some cases, the mean position of atoms appeared to shift with temperature. More recently, a number of theoretical and experimental analyses of alternate conformations of side chains have been reported (Summers et al., 1987; Kuriyan et al., 1986b; Hendrickson et al., 1986).

It has been shown in a recent isotropic harmonic refinement of simulated X-ray data for myoglobin at 1.5-Å resolution that the positional errors are in the range 0.10–0.20 Å for backbone atoms and 0.28–0.33 Å for side chain atoms and that, particularly for side chains, the mean-square fluctuations are consistently underestimated (Kuriyan et al., 1986a). However, in that paper no analysis was made of the relative importance of anisotropic and harmonic effects nor were suggestions made as to how to improve the refinement procedure. The present paper concerns itself with these questions. The errors in refinements that use an isotropic harmonic model are analyzed by fitting dynamics data. We also derive expressions for the error in fit value of the mean and the second moment from isotropic, harmonic refinement of model distributions, as a function of their third and fourth moments. To determine which correction gives the most improvement over the isotropic Gaussian model, higher order models (i.e., anisotropic, Gram–Charlier, and double peak) for refinement are compared by fitting the pdfs obtained from the simulation. We first outline the methology and then present the results. The discussion evaluates the results and considers the advantages of a simulation approach over experimental crystallographic analyses that are possible with the available data.

Methods

This section outlines the methods used in the present analysis.

Model Probability Density Functions. The different model pdfs used in the analysis of refinement procedures are shown in Table I. The pdfs are expressed in a general local coordinate system in terms of the atomic fluctuations $\mathbf{u} = \mathbf{r} - \langle\mathbf{r}\rangle$, where $\langle\mathbf{r}\rangle$ is the time-averaged mean position. The dynamics results are given in the local principal axis system for a given atom; its fluctuations in this system are designated as $\mathbf{u} = (u_x,u_y,u_z)$ with u_x along the direction of the largest mean-square displacement and u_z along the direction of the smallest. In this coordinate system the covariances $\langle u_iu_j\rangle$, $i \neq j$, are equal to zero. Details concerning the choice of these pdfs and their analysis are given in Ichiye (1985) and Ichiye and Karplus (1987).

The argument of each pdf includes all of the parameters needed to determine the distribution; therefore, the mean position $\mu = \langle\mathbf{r}\rangle$, is listed even though the pdfs do not depend

explicitly on it. The parameters include **V**, the covariance matrix with elements $\sigma_{ij} = \langle u_i u_j \rangle$ and determinant $|\mathbf{V}|$, and the matrices α_3 and α_4 with elements $\kappa_{jkl}/(\sigma_j\sigma_k\sigma_l)$ and $\kappa_{jklm}/(\sigma_j\sigma_k\sigma_l\sigma_m)$, respectively, where $\sigma_j^2 = \sigma_{jj}$, and $\kappa_{jkl} = \langle u_j u_k u_l \rangle$ and $\kappa_{jklm} = \langle u_j u_k u_l u_m \rangle - \sigma_{jk}\sigma_{lm} - \sigma_{kl}\sigma_{mj} - \sigma_{mk}\sigma_{jl}$ are third and fourth cumulants (Kendall & Stuart, 1977). The diagonal elements of α_3 and α_4 are the coefficients of skewness and kurtosis, respectively. The Hermite polynomials for the Gram–Charlier and Edgeworth expansions can be defined by $H_{jk\ldots}(\mathbf{u})p_G(\mathbf{u}) = (-1)^i(\partial/\partial u_j)(\partial/\partial u_k)\ldots p_G(\mathbf{u})$ (Johnson & Levy, 1974). For the double-peak pdf, $\mathbf{u}_\pm = (u_{\pm,x}, u_{\pm,y}, u_{\pm,z})$ are the coordinates of the maxima relative to μ and $w_\pm$ are the relative probabilities of being in each well ($w_+ + w_- = 1$). The two peaks are assumed to have the same half-width, $\mathbf{V}_0$, to reduce the number of parameters; this is equivalent to assuming that the force constants are the same for the two potential wells. In the principal axis system, $\mathbf{u}_\pm = (\pm u_\pm, 0, 0)$ and $w_-u_- = w_+u_+$. When one of the peaks goes to zero (either $w_+ \rightarrow 0$ or $w_- \rightarrow 0$) or when the separation between the peaks approaches zero ($\mathbf{u}_+ + \mathbf{u}_- \rightarrow 0$), the pdf approaches a Gaussian.

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; the details of the simulation are presented elsewhere (Ichiye et al., 1986). Starting with a refined X-ray structure (D. E. P. Grace and D. C. Phillips, unpublished results) based on the 2.0-Å resolution crystal data of the tetragonal form (Blake et al., 1965) and an equilibration period of 26 ps, the simulation was continued for an additional 34 ps of which the first 30 ps were used for analysis. The time step was 0.001 ps, and the mean temperature was equal to 304 K. The moments and the probability densities were obtained from the coordinates at intervals of 0.05 and 0.01 ps, respectively. The probability densities were determined by calculating normalized histograms of the coordinates with an interval size of 0.1 Å.

Refinement Procedure. Refinement of a protein structure with current procedures (Watenpaugh et al., 1979; Diamond, 1971, 1974; Konnert, 1976; Konnert & Hendrickson, 1980; Dodson et al., 1976; Jack & Levitt, 1978) involves a simultaneous fit of the electron densities of all the atoms to a model for the electron densities that may include constraints and/or restraints on the molecular geometry. For the most common model in which the electron density due to a given atom is assumed to correspond to a harmonic and isotropic distribution, the parameters that are varied in the fit are the coordinates of every atom and the isotropic temperature factor B, which may be constrained to the same value for groups of atoms or even the entire molecule, depending on the resolution of the data. In the harmonic and isotropic model, the temperature factor is related to the fluctuations by

$$B = 8\pi^2\langle u_i^2 \rangle = \frac{8\pi^2\langle u_r^2 \rangle}{3} \qquad (1)$$

where $u_r^2 = \sum u_i^2$ and it is assumed that $\langle u_i^2 \rangle = \langle u_x^2 \rangle = \langle u_y^2 \rangle = \langle u_z^2 \rangle = {}^1/_3\langle u_r^2 \rangle$.

The "R-factor" least-squares refinement procedure minimizes the function (Cruickshank, 1974)

$$R = \sum w_a(hkl)[|F_o(hkl)| - |F_c(hkl)|]^2 \qquad (2)$$

where F_o and F_c are the observed and calculated structure factors, respectively, w is the weight associated with each term, and h, k, and l correspond to the reciprocal lattice points of the crystal. For proteins, the weights generally used are equal or vary linearly with $(h^2 + k^2 + l^2)^{1/2}$ so that low-resolution structure factors are weighted more heavily than the high-resolution data. (Note that the R value given in many references is often not the same quantity as in eq 2, even if the refined quantity is that in eq 2.)

The refinement procedure used here minimizes the quantity

$$R' = [p_c(\mathbf{u};\mu_f,\mathbf{V}_f,\alpha_{3f},\alpha_{4f}) - p_o(\mathbf{u})]^2 d\mathbf{u} \qquad (3)$$

where p_o and p_c are the observed and calculated pdfs and the subscript f refers to values obtained by the fitting procedure. This is the quantity minimized in Diamond real-space refinement (Diamond, 1971), which he has shown to be equivalent to

$$R' = \frac{1}{\nu}\sum|F_o(hkl) - F_c(hkl)|^2 \qquad (4)$$

where ν is the unit cell volume. Thus, real-space refinement is equivalent to a reciprocal-space refinement in which phase information is included and all of the data are weighted equally. Equations 2 and 4 are expected to behave similarly especially if the calculated phases are close to the real values. Thus, the general features of the results obtained here are likely to be applicable to reciprocal-space refinement; this has been demonstrated recently in an analysis of myoglobin (Kuriyan et al., 1986a). In what follows, we analyze each atomic pdf separately; i.e., each atom is treated as an isolated system to simplify the calculation. The parameters μ_f and σ_f are allowed to vary simultaneously.

Results

We begin the analysis of the X-ray refinement method by first testing the isotropic Gaussian assumption. We fit model pdfs in section a and dynamics pdfs in section b. In section c we give a method for estimating the fit values of the mean position and second moment found by refinement, which often differ considerably from the time-average moments of the trajectory. We test higher order models for refinement using pdfs from dynamics in section d.

(a) Three-Dimensional Isotropic Gaussian Refinement of Model Probability Density Functions. Here we examine three-dimensional, isotropic Gaussian refinement of various model pdfs which are anisotropic and/or anharmonic. Since the "observed" data correspond to a model pdf, we can examine the effects of anisotropy and anharmonicity separately. This contrasts with the dynamics pdfs that have a combination of anisotropic and anharmonic effects as well as errors due to sampling. The refinement minimizes R' in eq 3 where p_c is an isotropic Gaussian, $p_G(\mathbf{u};\mu_f,\sigma_f)$, and p_o is one of the model pdfs given under Methods. The fitting parameters μ_f and σ_f are thus the results of the refinement procedure, as opposed to the true mean $\mu = (0,0,0)$ and second moment σ, where $\sigma^2 = (\sigma_x^2 + \sigma_y^2 + \sigma_z^2)/3$. For simplicity, the observed (model) pdfs are in a local principal axis system. By taking derivatives of R' with respect to μ_f and σ_f, expressions for these two parameters as a function of μ, **V**, α_3, and α_4 may be obtained for each model (see Appendixes A and B). We evaluate these expressions over the range of values of the anisotropy and anharmonicity (α_3 and α_4) found in the lysozyme simulation (Ichiye & Karplus, 1987).

Anisotropy. The first case we consider is isotropic, Gaussian refinement of an anisotropic Gaussian pdf. The "observed" pdf is thus $p_o(\mathbf{u}) = p_G(\mathbf{u};\mathbf{0},\mathbf{V})$, where **V** has elements $V_{ii} = \sigma_i^2$, $\sigma_x \neq \sigma_y$ and/or $\sigma_y \neq \sigma_z$, and $V_{ij} = 0$ in the principal axis system. The values of μ_f and σ_f are found by simultaneously solving eq A6 and A8 in Appendix A with $\alpha_3 = \alpha_4 = 0$. Under

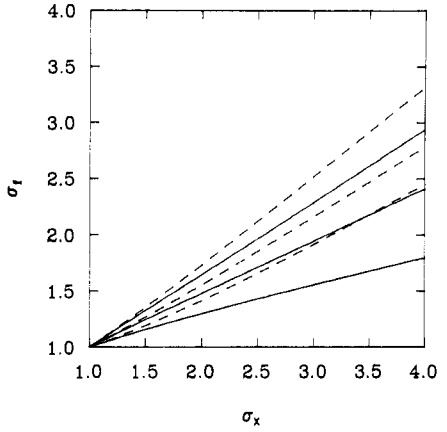

FIGURE 1: Least-squares fit of anisotropic model distribution, where the solid line is σ_f versus σ_x and the dashed line is σ versus σ_x. From top to bottom, $\sigma_y = \sigma_x$, $\sigma_y = (\sigma_x + \sigma_z)/2$, and $\sigma_y = \sigma_z$.

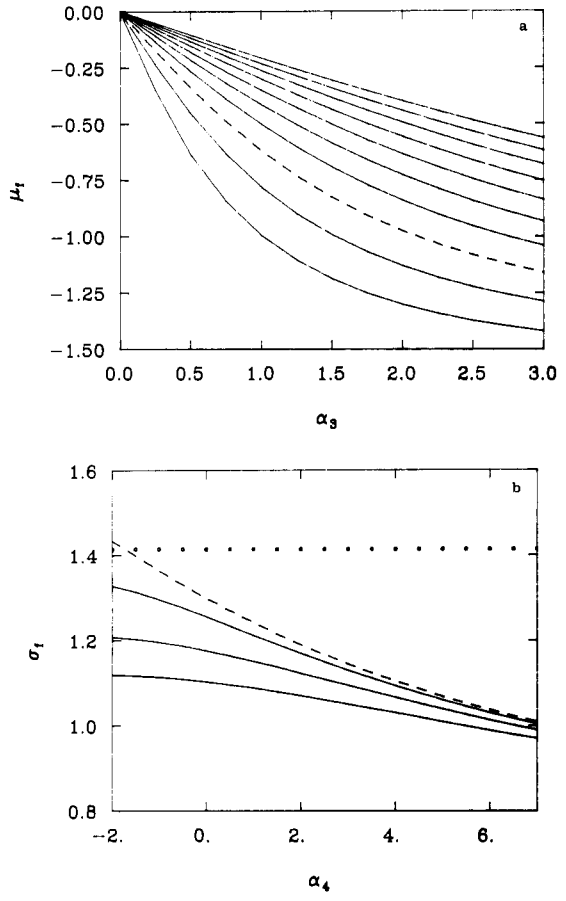

FIGURE 2: Least-squares fit of Gram–Charlier distribution by a Gaussian distribution when $\sigma = (2,1,1)$. (a) Deviation of μ_f from the true value versus $|\alpha_3|$; from top to bottom, $\alpha_4 = 7$ to -2 (dashed line, $\alpha_4 = 0$). (b) Values of σ_f versus α_4; from top to bottom, $\alpha_3 = 0$ to 3 (dashed line, $\alpha_3 = 0$).

these conditions, the correct mean position is obtained by the refinement; i.e., $\mu_f = (0,0,0)$, but there are errors in σ_f. In Figure 1 we plot σ_f versus σ_x when $\sigma_z = 1$ and σ_y is chosen in three different ways ($\sigma_y = \sigma_x$, $\sigma_y = (\sigma_x + \sigma_z)/2$, or $\sigma_y = \sigma_z$). We also plot the correct rms fluctuation σ for the same values of σ_y and σ_z. The fit values, σ_f, are approximately equal to σ for small anisotropies ($1 \leq \sigma_x/\sigma_z < 1.5$) but are too small by 0.4–$0.6\sigma_z$ at $\sigma_x/\sigma_z = 4$.

Anharmonicity, Unimodal. To consider the isotropic, Gaussian refinement of an anharmonic unimodal pdf, the "observed" pdf is given by the Gram–Charlier expansion

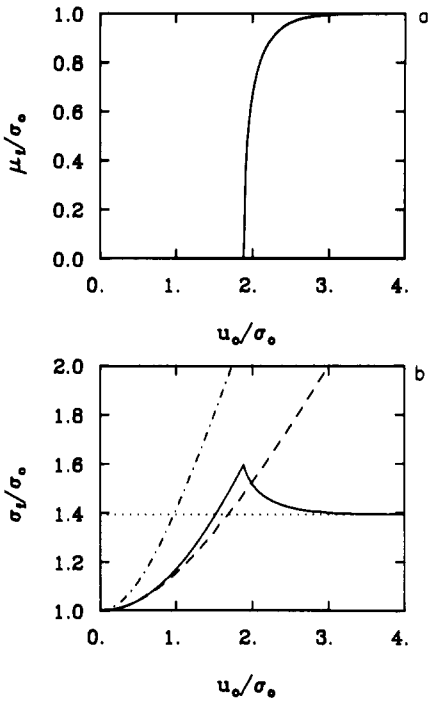

FIGURE 3: Least-squares fit of double-peak distribution (a) μ_f or (b) σ_f versus peak separation u_0 in units of the width of each peak, σ_0 (solid line), and the actual value of σ (dashed line). The approximate value of σ_f at large separations from eq B9 (dotted line) and the actual value of σ_x (dotted–dashed line) are also shown.

(Table I) in the x direction and Gaussian pdfs in the y and z directions; i.e., $p_o(\mathbf{u}) = p_{GC}(\mathbf{u};\mathbf{0},\mathbf{V},\alpha_3,\alpha_4)$, where $\mathbf{V}$ may be anisotropic and α_3 and α_4 are the anharmonicities in the x direction. This distribution is fit by an isotropic Gaussian by use of the method given in Appendix A. The fit value of the mean for this case is $\mu_f = (\mu_f,0,0)$; i.e., the correct mean is obtained for the y and z directions but not necessarily for the x direction with anharmonicity. From Figure 2a, when $\sigma_x = 2\sigma_y = 2\sigma_z$, the deviation of μ_f from the true value of the mean increases with increasing $|\alpha_3|$, or skewness, and decreases with α_4 if $|\alpha_3| > 0$. Thus, μ_f can have an error of up to $-1.5\sigma_z$ when $|\alpha_3| = 3$ and $\alpha_4 = -2$. From the plot of σ_f versus α_4 in Figure 2b, σ_f decreases with α_4 (increased peakedness) and decreases slightly with $|\alpha_3|$. In the range of $-2 \leq \alpha_4 \leq 7$, $0 \leq |\alpha_3| \leq 3$, σ_f ranges from being too small by $0.45\sigma_z$ to too large by $0.02\sigma_z$. The dependence of μ_f on α_4 and of σ_f on α_3 is less if $\sigma_x = \sigma_y = 2\sigma_z$ and still less if $\sigma_x = \sigma_y = \sigma_z$ (Ichiye, 1985).

Anharmonicity, Bimodal. Since many of the pdfs found in the dynamics are bimodal (Ichiye & Karplus, 1987), we consider isotropic, Gaussian refinement of a double-peak pdf in which each peak is isotropic (Appendix B). The "observed" pdf for this case is $p_o(\mathbf{u}) = p_\mu(\mathbf{u};\mathbf{0},\mathbf{u}_\pm,\sigma_\pm)$ (Table I) in the principal axis system, so the positions of the two peaks are $\mathbf{u}_\pm = (\pm u_\pm,0,0)$. We concentrate on the refinement of two equally weighted isotropic peaks with the same σ_0 since this case seems most likely to be fit by an envelope of the two peaks. For this case, the parameters of the double-peak pdf (Table I) are $u_+ = u_- \equiv u_0$, $w_+ = w_- \equiv w_0$, and $\sigma_+ = \sigma_- \equiv \sigma_0 = \sigma_y = \sigma_z$.

Since the pdf is harmonic in the y and z directions, the fit value of the mean is $\mu_f = (\mu_f,0,0)$. The values of μ_f and σ_f as a function of u_0, which is half of the peak separation, are given in Figure 3; the actual mean and fluctuation are $\mu = 0$ and $\sigma^2 = \sigma_0^2 + {}^1/_3u_0^2$ (Table I), respectively. For small u_0, the minimum of R' is at $\mu_f = 0$ and $\sigma_f^2 \approx \sigma_0^2 + {}^1/_3u_0^2$. Thus, the mean is correct and the fluctuations are approximately correct. As u_0 increases, σ_f^2 is larger than σ^2 but smaller than the second moment in the x direction ($\sigma_x^2 = \sigma_0^2 + u_0^2$). However,

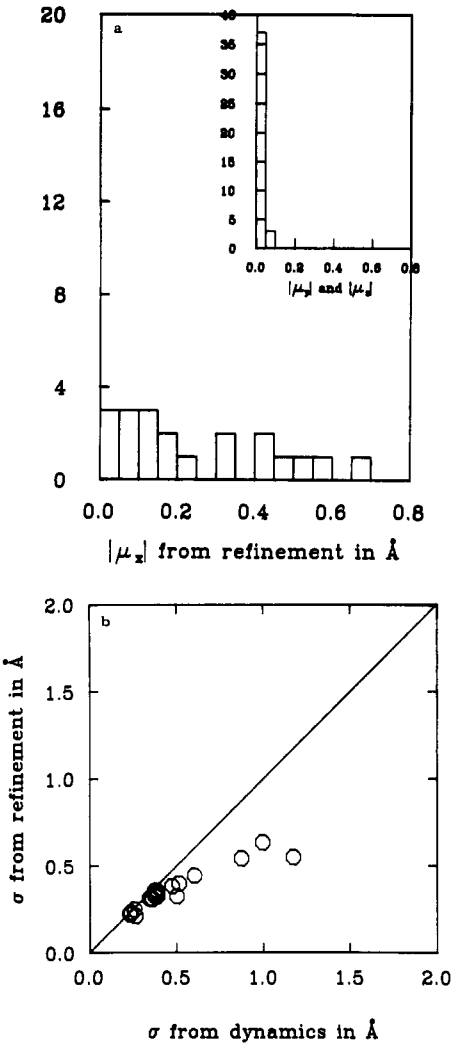

FIGURE 4: Least-squares fit of three-dimensional distributions from dynamics by an isotropic harmonic pdf. (a) Histogram of $|\mu_f|$. (b) Plot of σ_f versus σ.

when $u_0^2 \gtrsim 3.55\sigma_0^2$ so that the separation between the peaks, $2u_0$, is greater than $\sim 3.77\sigma_0$, $\mu_f = 0$ is no longer a minimum (i.e., the second derivative is less than 0); for this case, $\mu_f \rightarrow u_0$ (there are two equal solutions at each peak) and $\sigma_f^2 \rightarrow \sim 1.94\sigma_0^2$. Thus, the refinement for large peak separations gives values of the mean that are approximately equal to the center and rms fluctuations up to twice the rms fluctuations of only one peak.

(*b*) *Three-Dimensional Isotropic Gaussian Refinement of Dynamics Probability Density Functions.* We test the three-dimensional, isotropic Gaussian refinement of dynamics pdfs by simultaneously varying the mean μ_f and the second moment σ_f to minimize the least-squares differences (eq 3). We concentrate on distributions that deviate strongly from being isotropic and/or harmonic; i.e., we concern ourselves with the atomic distributions that are likely to have large errors in the refinement. In particular, we note that refinement of distributions that are anisotropic but harmonic lead to systematic errors in the *B* values but not the mean position, whereas those that are anharmonic can result in errors in both the mean positions and the *B* values. The pdfs of twenty atoms, nine backbone and eleven side chain, were examined. Eight of the backbone and eight of the side chain atoms were chosen because they had the largest anharmonicities, i.e., the largest values of either α_3 or α_4; these atoms often have the largest anisotropies as well. These were the same pdfs analyzed in a previous study (Ichiye & Karplus, 1987).

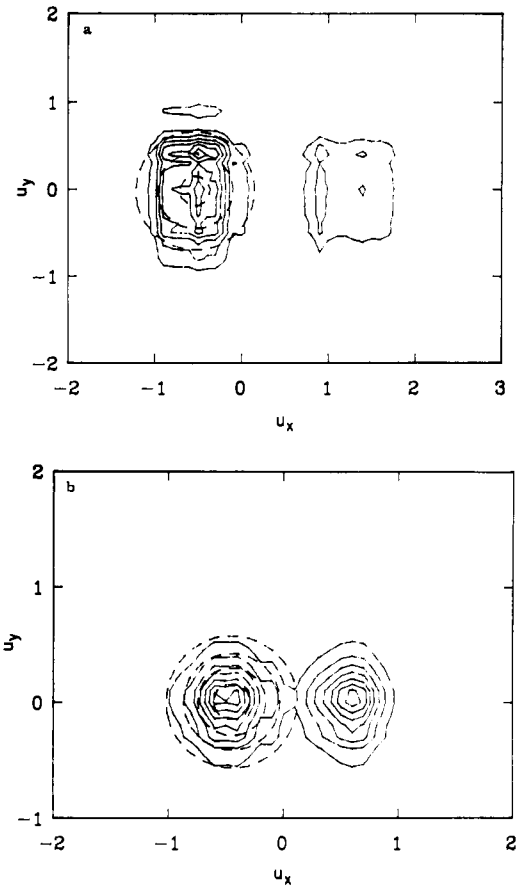

FIGURE 5: Comparison of probability density from a molecular dynamics simulation of lysozyme with the pdf from real-space "refinement" of the density in the isotropic, harmonic approximation, presented as cross sections through the $u_x - u_y$ plane when $u_z = 0$ in the principal axis system: (a) molecular dynamics results for Gly-71 C_α (solid), contoured every 0.0001, and "refinement" results (dash), contoured every 0.0002; (b) molecular dynamics results for Gln-57 $N_{\epsilon 2}$ (solid) and "refinement" results (dash), contoured every 0.0004.

Rather than calculating the actual three-dimensional dynamics pdf for an atom, we approximate it by multiplying together the pdfs for each dimension (in the principal axis system); i.e., we use $p_o(\mathbf{u}) \approx p_o(u_x)p_o(u_y)p_o(u_z)$. Some differences in the peak shape may result; however, we are interested here in testing the refinement procedure itself rather than in obtaining accurate dynamics pdfs. The correct values of the mean and fluctuations obtained independently from a time average over the dynamics trajectory in the local coordinate system are $\mu = \langle \mathbf{u} \rangle = (0,0,0)$ and $\sigma^2 = [\langle u_r^2 \rangle/3] = [(\sigma_x^2 + \sigma_y^2 + \sigma_z^2)/3]$.

The results of the least-squares refinement of the three-dimensional dynamics pdfs show that the assumption that the electron density of an atom is isotropic and Gaussian causes errors in the calculation of μ_f and σ_f if the actual density is anharmonic (Figure 4). The largest errors occur in the direction of the largest principal axis; i.e., $\mu_{f,x}$ has errors up to about 0.7 Å whereas the largest errors in $\mu_{f,y}$ and $\mu_{f,z}$ are less than 0.1 Å. The value of σ_f from the refinement consistently underestimates the value of σ from the dynamics, especially for large σ (greater that about 0.5 Å). The least-squares fit appears to place μ_f at the most probable position (i.e., the maximum probability) of the dynamics pdfs rather than at the actual mean. If there are multiple peaks, the least-squares procedure fits the largest peak as a single Gaussian with μ_f close to the center of that peak and σ_f about twice as big as that of the peak, as in the case of Gly-71 C_α (Figure 5a) a corresponding result is obtained even if there are two ap-

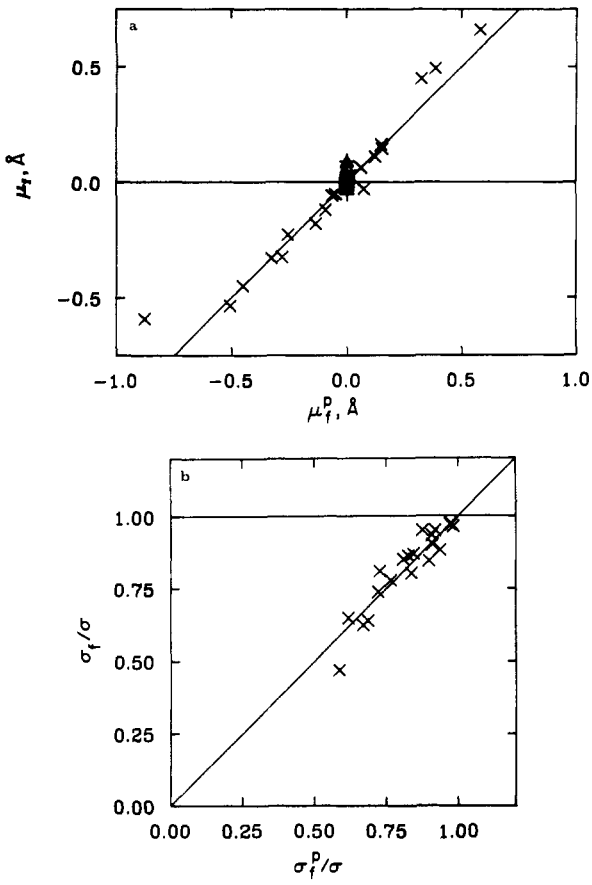

FIGURE 6: Prediction of μ_f and σ_f when an isotropic Gaussian is used to fit the dynamics distributions for the u_x, u_y, and u_z directions, represented by x, +, and Δ, respectively. (a) Predicted μ_f versus μ_f. (b) Predicted σ_f^p/σ versus σ_f^p/σ.

proximately equal peaks such as for Gln-57 $N_{\epsilon 2}$ (Figure 5b).

(c) Prediction of Errors Due to the Assumption of Isotropic Gaussian Distributions. The molecular dynamics simulation shows that the largest deviations from the single isotropic Gaussian peak model can be attributed to the presence of secondary peaks (Ichiye & Karplus, 1987). Furthermore, the above results have demonstrated that refinement procedures fit only one peak if the peaks are separated by a critical distance ($\sim 3.77\sigma_0$ for equal peaks). In this section we estimate the fit values of μ_f and σ_f which would be obtained from least-squares (real-space) refinement of the dynamics probability densities of atomic positions. We assume that the pdf of any atom from a trajectory can be represented as two Gaussians with the same width but different weights (Table I); the pdf is then specified by μ, V, α_3, and α_4. We choose the axes so that the double peaks are situated along the u_x direction; i.e., the peaks are located at $\mu = (\pm u_\pm, 0, 0)$. Further, we assume for the refinement that the probability density of any atom can be treated independently of the other atoms. This is clearly not correct but makes it possible to identify the important factors in a straightforward manner. In particular, only the first four moments for every atom need to be calculated rather than the three-dimensional probability density for all of the atoms in the molecule.

Our procedure for estimating the values of μ_f and σ is to solve for the minimum of eq 3 when p_0 is an unequal double-peak pdf in which each peak can be anisotropic but each has the same V_0 (Appendix B). The predicted value is denoted by a superscript p. The estimated values, μ_f^p and σ_f^p, are plotted against μ_f and σ_f, respectively, in panels a and b of Figure 6. Almost all of the predicted values, μ_f^p, are within 0.1 Å of the refined value μ_f, even when they differ from the true mean by up to 0.7 Å. The predicted values σ_f^p/σ are within about 15% of the refined values of σ_f/σ, even when σ_f is reduced to about half of σ. A simpler estimate in which $\mu_f = (u_{max},0,0)$, u_{max} being the u_x coordinate of the larger peak, and $\sigma_f^2 = (\sigma_0^2 + \sigma_y^2 + \sigma_z^2)/3$ also gives good results (Ichiye, 1985).

(d) Refinement Using Higher Order Probability Density Functions. In this section we evaluated the efficacy of the various model pdfs of Table I in the refinement of dynamics pdfs. We examined the pdfs along the u_x, u_y, and u_z axes separately (i.e., one-dimensional refinement). This allows for anisotropy but is only approximately equal to an anisotropic three-dimensional refinement since the pdfs are refined independently in each direction. Thus, R' in eq 3 was minimized by using, in turn, the Gaussian, Gram–Charlier, Edgeworth, and double-peak pdfs for p_c (Figure 7).

This type of "anisotropic" refinement with a Gaussian p_c also fits the larger peak, as does the isotropic refinement, so that there are still considerable deviations of μ_f and σ_f from the actual values of μ and σ. The refinement by a Gram–Charlier pdf gives better values for μ_f and σ_f than a Gaussian but still may give large deviations if there are two or more widely separated peaks with significant populations. The double-peak pdf gives the best results (less than ~0.1-Å deviation of μ_f and σ_f from the actual values in most cases)—it fails in two of the twenty cases examined, both cases having more than two peaks. For the refinements using the Gram–Charlier or the double-peak pdf, more than one minimum in R' was found in some cases. The values for μ_f and σ_f corresponding to the lowest value of R' are shown in Figure 7, although the other minimum sometimes gave values closer to the correct first two moments.

Discussion

Under ideal conditions X-ray diffraction gives the distribution of electron density of a molecule in a crystal, which can be related to the probability distributions of the atoms in the molecule (Willis & Pryor, 1975). The probability distributions of atomic positions from a dynamics simulation and from X-ray analysis both contain contributions from fluctuations within a single minimum and, if they take place, from transitions among several minima. However, in the X-ray results, diffusion and lattice disorder also affect the atomic probability distributions. Although there are methods for the approximate separation of the various contributions (Frauenfelder et al., 1979), they are generally difficult to apply so that most experiments measure, in principle, the overall effective probability distribution; i.e., they include diffusion and lattice disorder as well as the local motions. Also, in an X-ray study, the mean position and second moment are parameters fit in the refinement procedure, so the results are model dependent.

In the work presented here, we examine the errors in the refinement of proteins by studying simulations. In the simulation, the true mean position and second moment can be calculated independently of the refinement, and they then can be compared with those obtained from the refinement. Also, the simulation is free from errors due to diffusion and lattice disorder so that one can focus on the errors in the refinement due solely to the internal atomic fluctuations. Furthermore, in an X-ray experiment, the data are in reciprocal space and are due to the entire molecule. In the method used here, we focus on the distribution functions of individual atoms in real space and examine those that are most poorly fit by an isotropic Gaussian. Thus, the present approach provides a simple

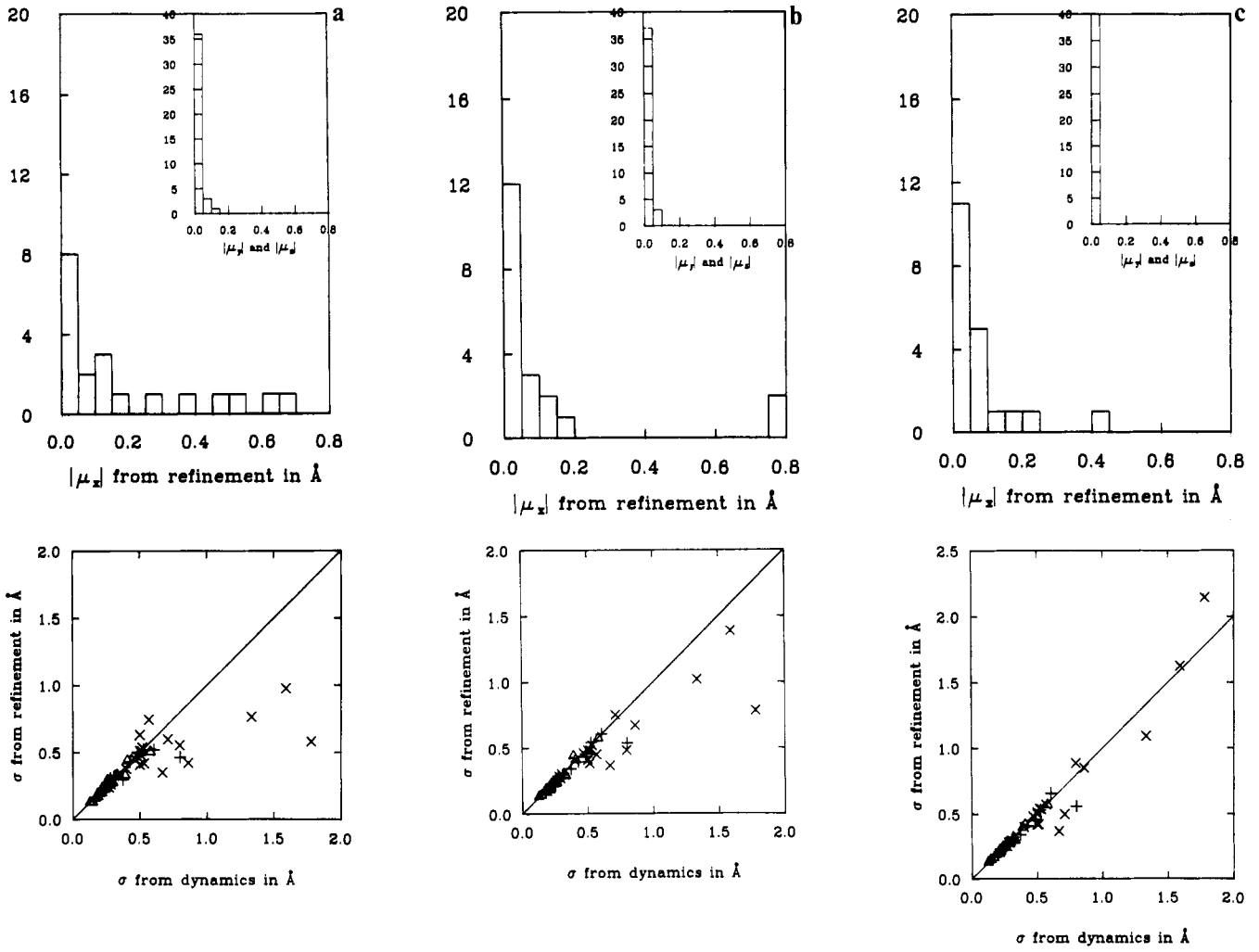

FIGURE 7: Least-squares fit of distributions from dynamics by (a) Gaussian, (b) Gram–Charlier, and (c) double-peak distributions, fitting each principal axis direction separately. For each model distribution, there is (top) a histogram of μ_f and (bottom) a plot of σ_f versus σ.

method for studying two problems of interest; i.e., the determination of the major errors due to the isotropic harmonic assumption and the evaluation of higher order corrections to this model.

From analysis of the dynamics simulation of lysozyme, the atomic distributions have been found to be highly anisotropic and somewhat anharmonic mainly due to multiple peaks (Ichiye & Karplus, 1987). However, refinement procedures currently used for proteins assume that the distribution for an atom is a single isotropic harmonic peak, although in a few experiments multiple occupancy or anisotropy has been included. To determine the errors introduced by this approximation, we have considered (section a under Results) how isotropic harmonic pdfs fit various model distributions that are anisotropic and/or anharmonic. In the lysozyme simulation, $\langle \sigma_x^2/\sigma_z^2 \rangle^{1/2}$ averaged over backbone atoms is 2.2 and over side chain atoms is 2.6, and $\langle \sigma_y^2/\sigma_z^2 \rangle^{1/2}$ averaged over backbone atoms is 1.4 and over side chain atoms is 1.5. If the simulation distributions were anisotropic but harmonic, the fit of σ_f to the model distribution for the average atom would be too small by about $0.1–0.2\sigma_z$. Only for highly anisotropic atoms, i.e., $\sigma_x/\sigma_z \approx 4$, are the errors as large as about $0.4–0.6\sigma_z$. When the "observed" pdf is a Gram–Charlier pdf, the maximum error is small if $\alpha_{3x}^2/6 + \alpha_{4x}/24 \lesssim 1/24$ (our criterion for harmonicity; Ichiye & Karplus, 1987); i.e., the mean obtained by fitting is in error by less than about $0.3\sigma_z$ and the value of σ is too small by less than about $0.2\sigma_z$ for $\sigma = (2,1,1)\sigma_z$. Thus, errors in the 68% of backbone and 56% of side chain atoms in lysozyme which the simulation shows to be harmonic in at least the u_x direction (Ichiye & Karplus, 1987) are expected to be small and errors due to anharmonicity in either the u_y or u_z direction will be still smaller since the fluctuations are by definition smaller.

Since most of the atoms with large values of α_3 and α_4 have multiple peaks, we have considered what happens in this case. It has been generally assumed that both real-space refinement and reciprocal-space refinement using a harmonic model fit an envelope to multiple peaks. Diamond (1971) noted that gradient/curvature real-space refinement techniques (Lipson & Cochran, 1966; Cruickshank, 1974) tend to fit the maximum rather than the actual mean and suggested that his refinement procedure would fit the entire density. However, the results obtained here indicate that even the Diamond real-space refinement technique tends to fit the maximum rather than the mean.

Atoms with the largest anharmonicities also have the largest anisotropies and can best be described as multiple harmonic peaks (Ichiye & Karplus, 1987). We examined in detail dynamics pdfs for lysozyme atoms with the largest anharmonicities (section b under Results). When the average atomic positions and fluctuations from a dynamics simulation are compared with the values obtained from real-space least-squares refinement of the atomic pdfs with the isotropic, harmonic approximation, we find that if $|\alpha_3|$ and $|\alpha_4|$ are small, the refined values agree well with the dynamics values. However, if $|\alpha_3|$ and/or $|\alpha_4|$ are large, the refined values may

be very different from the actual averages. For instance, for multiple peaks, refinement gives a peak centered at the most probable value, or mode, with the width corresponding to that of the fitted peak rather than the actual second moment. If the weight of one peak is much greater than that of the other, this result is appropriate since the position and width of the major peak are close to the true mean and second moment. However, the refinement of a pdf from the dynamics with two peaks which appeared almost equal (Figure 5b) still resulted in only one peak being fit. In this case, the difference between the position of the major peak and the mean can be quite large. The results obtained here are expected to be similar to those for reciprocal-space refinement since the differences in the quantities refined (see eq 2 and 4) cannot correct for such errors.

Previously, when dynamics simulations gave different structures and larger fluctuations than X-ray crystallographic results, the deviations were attributed solely to inadequacies of the simulation such as lack of solvent and crystal forces. However, the results presented here indicate that the differences may also be due in part to the failure of the X-ray refinement when multiple peaks are involved. The coordinates and temperature factors from X-ray refinement correspond to a single potential well at its minimum whereas those from dynamics will be the average position and second moment due to all the wells. This has been confirmed by Kuriyan et al. (1986a) for the reciprocal-space refinement of a molecular dynamics simulation of myoglobin. The simulation analyses showed that the crystallographic results will tend to underestimate the magnitude of the fluctuations; for the side chains, in particular, there can be large errors. This has important implications for studies of biological activity (Elber & Karplus, 1987), since multiple conformational states that could be important in the function of a protein may not be detected by using current X-ray crystallographic techniques.

In a previous paper (Ichiye & Karplus, 1987) we have shown that we can predict the position and width of two peaks in the pdfs for each atom given the first four moments without resorting to actually calculating the pdf. Furthermore, we have shown in section c under Results that we can use these values to predict the positions and temperature factors that would be obtained in a real-space refinement within the isotropic, harmonic approximation. We suggest that these are the appropriate quantities from simultations (rather than the true means and second moments) to compare with results from experimental refinements that use the isotropic harmonic assumption. An alternative approach would be to calculate the total probability densities due to all of the protein atoms from the simulation and then refine them by using the same programs as used for the crystal data (Kuriyan et al., 1986a). This requires large amounts of computer time and storage. Since our conclusions agree with those of Kuriyan et al., and since simulation results are only qualitative, the method given here is a useful approximation for most comparisons of simulation and experimental results.

With higher resolution data becoming available for proteins, it is of interest to evaluate corrections to the temperature factor models that significantly improve the fits with the smallest number of added parameters. The number of parameters needed per atom for the various pdfs considered in this paper are given in Table II. Isotropic harmonic temperature factors add one parameter, in addition to the mean position. Anisotropic harmonic temperature factors require six coefficients for the symmetric **B** and β matrix [see Willis and Pryor (1975) for notation]. If anharmonic corrections are to be included as well, the number of parameters increases further; e.g., to include skewness (α_3) and kurtosis (α_4) via expansions, we need 10 and 15 unique elements, respectively. Thus, it is clear that any refinements introduced with the limitations set by the available protein data require simplifications in the description of the motion.

Table II: Number of Parameters per Atom for Refinement

level of correction	parameters	no. of parameters
position	**r**	3
isotropic B	**r**, σ	4
anisotropic B	**r**, σ	9
anisotropic B	$\sigma_y = \sigma_z$, **r**, σ_x, ($\sigma_y = \sigma_z$)	7
skewness	**r**, σ, α_3	19
kurtosis	**r**, σ, α_3, α_4	34
isotropic kurtosis	**r**, σ, α_4	5
double well	**r**, σ, **r**′, σ', w'	9
double well	**r**, $\sigma = \sigma'$, **r**′, w'	8

Within the harmonic approximation, attempts have been made to relate the orientations of the anisotropy tensors to the static structure (Konnert & Hendrickson, 1980). However, it has been demonstrated by molecular dynamics simulations of proteins that there is no simple relation (Yu et al., 1985). Assuming that $\sigma_y \approx \sigma_z$ in the principal axis system, we could reduce the temperature factor parameters to four: σ_x, $\sigma_y = \sigma_z$, θ, and ϕ, where the last two give the orientation of u_x in the crystal coordinate system. Results of previous analyses [Table 1 of Ichiye and Karplus (1987)] indicated that the difference between σ_y and σ_z is only about 14% of the average value of σ_y and σ_z for most atoms.

To reduce the number of parameters needed to include anharmonic effects, one possibility is to describe the potentials as isotropic nonskewed anharmonic potentials (Frauenfelder et al., 1979). There are then only two isotropic parameters, σ and α_4. However, the molecular dynamics results show that the distributions are quite anisotropic and only somewhat anharmonic. Thus, it seems unreasonable to include anharmonic effects without including anisotropic effects. Moreover, the deviations from $\alpha_4 = 0$ are usually associated with skewness (Ichiye & Karplus, 1987). A better approach may be to assume that the major deviations from isotropy and harmonicity of a potential are due to multiple minima. If we consider that an atom may be in either of two isotropic harmonic wells, we need the position, σ_2, and a weight factor for the second well in addition to the position and σ_1 of the first well. Thus, anisotropic and anharmonic corrections are included by using a total of nine parameters. We can further reduce the number of parameters to eight by assuming the second potential has the same force constant and therefore the same width σ_0. 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. Results of a previous analysis (Ichiye & Karplus, 1987) indicate that this model parametrized by the moments gives a significant improvement in describing the dynamics pdfs. The present work is closer to the true experimental situation since it involves fitting pdfs to data so that the moments are fitting parameters whereas in the previous work the pdfs were parametrized by the actual calculated moments.

We have looked at real-space refinement of the parameters for each of the model pdfs against the dynamics probability densities (section d under Results). Our results indicate that the best model for including anisotropic and anharmonic effects for X-ray crystal refinement is to include multiple occupancy. The anisotropic Gaussian and Gram–Charlier expansion do not approach the accuracy of the double peak when the distributions are highly anisotropic and anharmonic. Although

in our analysis we allowed each peak to be anisotropic, the results should be little changed for isotropic peaks since the anisotropy of each peak of a double-peak distribution is small (Ichiye & Karplus, 1987). Thus, by including multiple occupancy, the most significant contributions to the anisotropy and anharmonicity of the atomic pdfs are included by using only eight or nine parameters (depending on whether or not the wells are assumed to have equal force constants). For more than two peaks, higher moments can be related similarly (Ichiye, 1985).

In the previous paper (Ichiye & Karplus, 1987), we showed that we can predict positions and relative weights of secondary peaks in the simulation distributions given the first four moments of atomic fluctuations. Of course, the resulting parameters are not expected to be highly accurate for an actual protein due to approximations in the potential (even when solvent and crystal effects are included) and the short time scale of current simulations. However, the information from the simulation may be used as an initial assumption in X-ray crystal refinement to reduce the number of parameters needed to include some of the major anharmonic effects by adding multiple occupancy factors only for atoms predicted by dynamics to have a second peak with a significant weight and separation relative to the first peak.

Conclusions

The results presented here indicate that refinement of proteins using the isotropic, harmonic approximation can lead to large errors in the mean positions and fluctuations of some atoms due to the presence of multiple peaks. An estimate is given for the errors in the refined values of these two quantities. They are also the appropriate results from simulation to compare with experimental refinements that use the isotropic, harmonic approximation. Furthermore, the analysis indicates that the most significant improvement to the isotropic harmonic approximation is to include multiple occupancy even when two separate positions for a residue are not evident in the electron density map. This can account for both anisotropy and anharmonicity while adding relatively few parameters.

Acknowledgments

We thank John Kuriyan for many helpful discussions and B. Montgomery Pettitt and Neena Summers for comments on the manuscript.

Appendix

(*A*) *Minimizing an Anisotropic pdf with Anharmonicity Only along the u_x Direction.* Here eq 3 is minimized when p_c is given by an isotropic Gaussian and the "observed" pdf is given by an anisotropic and anharmonic Gram–Charlier expansion (Table I). The case for anisotropy but no anharmonicity is given by setting $\alpha_3 = \alpha_4 = 0$. For this discussion, we use $\mathbf{m}$ and s for μ_f and σ_f to reduce the number of subscripts. The quantity to be minimized is given by eq 3 where

$$p_c(\mathbf{u};\mathbf{m},s) = \frac{1}{(2\pi)^{3/2}s^3}\exp\left(-\frac{|\mathbf{u}-\mathbf{m}|^2}{2s^2}\right) \quad \text{(A1)}$$

$$p_o(\mathbf{u}) = p_G(\mathbf{u};\mathbf{0},\mathbf{V})\left[1 + \frac{\alpha_3}{3!}H_3\left(\frac{u_x}{\sigma_x}\right) + \frac{\alpha_4}{4!}H_4\left(\frac{u_x}{\sigma_x}\right)\right] \quad \text{(A2)}$$

in which $\sigma_{ii} \equiv \sigma_i^2$ and $\sigma_{i\neq j} = 0$. The conditions for extrema are $\partial R/\partial m = \partial R/\partial s = 0$. The necessary integrals are

$$\int p_c^2\,d\mathbf{u} = \frac{1}{(2s)^3\pi^{3/2}} \quad \text{(A3)}$$

$$\int p_c p_o\,d\mathbf{u} = \frac{1}{2^{3/2}}p_G(\mathbf{u};\mathbf{0},\mathbf{S})\left[1 + \frac{\alpha_3\sigma_x^3}{3!S_x^3}H_3\left(\frac{m_x}{S_x}\right) + \frac{\alpha_4\sigma_x^4}{4!S_x^4}H_4\left(\frac{m_x}{S_x}\right)\right] \quad \text{(A4)}$$

where $S_i^2 = \sigma_i^2 + s^2$. Since the derivatives of R with respect to m_y and m_z are zero when $m_y = m_z = 0$, we substitute these values and drop the subscript x wherever it occurs. The derivative of R with respect to m is

$$\frac{\partial R}{\partial m} = \frac{-1}{2^{3/2}}p_G(\mathbf{u};\mathbf{0},\mathbf{S})\left[\frac{m}{S} + \frac{\alpha_3\sigma^3}{3!S^3}H_4\left(\frac{m}{S}\right) + \frac{\alpha_4\sigma^4}{4!S^4}H_5\left(\frac{m}{S}\right)\right] = 0 \quad \text{(A5)}$$

and the second derivative of R with respect to m is

$$\frac{\partial^2 R}{\partial m^2} = \frac{1}{2^{3/2}}p_G(\mathbf{u};\mathbf{0},\mathbf{S}) \times \left[H_2\left(\frac{m}{S}\right) + \frac{\alpha_3\sigma^3}{3!S^3}H_5\left(\frac{m}{S}\right) + \frac{\alpha_4\sigma^4}{4!S^4}H_6\left(\frac{m}{S}\right)\right] \quad \text{(A6)}$$

Equation A5 may be rewritten as

$$\frac{m}{S} + \frac{\alpha_3\sigma^3}{3!S^3}H_4\left(\frac{m}{S}\right) + \frac{\alpha_4\sigma^4}{4!S^4}H_5\left(\frac{m}{S}\right) = 0 \quad \text{(A7)}$$

The derivative of R with respect to s, when $\partial R/\partial\mu = 0$, is

$$\frac{\partial R}{\partial s} = \frac{-3}{8\pi^{3/2}s^4} + \frac{s}{2^{1/2}}p_G(\mathbf{u};\mathbf{0},\mathbf{S})\left[\sum\frac{1}{S_i^2} + \frac{\alpha_3\sigma^3}{3!S^3}H_3\left(\frac{m}{S}\right)\left(\frac{3}{S^2} + \sum\frac{1}{S_i^2}\right) + \frac{\alpha_4\sigma^4}{4!S^4}H_4\left(\frac{m}{S}\right)\left(\frac{4}{S^2} + \sum\frac{1}{S_i^2}\right)\right] = 0 \quad \text{(A8)}$$

where $\sum S_i^{-2}$ is over $i = x,y,z$, remembering that explicit reference to x has been dropped. Thus, m and s can be obtained by simultaneously solving eq A7 and A8.

(*B*) *Real-Space Refinement of a Double-Peak pdf by a Gaussian pdf.* Here eq 3 is minimized when p_c is given by an isotropic Gaussian and the "observed" pdf is given by double-well pdf (Table I).

$$p_o(\mathbf{r}) = \frac{1}{(2\pi)^{3/2}\sigma_{0x}\sigma_{0y}\sigma_{0z}}\left[w_+\exp\left[-\frac{(x-u_+)^2}{2\sigma_{0x}^2} + \frac{y^2}{2\sigma_{0y}^2} + \frac{z^2}{2\sigma_{0z}^2}\right] + w_-\exp\left[-\frac{(x+u_-)^2}{2\sigma_{0x}^2} + \frac{y^2}{2\sigma_{0y}^2} + \frac{z^2}{2\sigma_{0z}^2}\right]\right] \quad \text{(B1)}$$

R is defined as in eq 3. The integrals needed are eq A4 and

$$\int p_o p_c\,d\mathbf{r} = \frac{1}{(2\pi)^{3/2}SS_yS_z}\exp\left(-\frac{m_y^2}{2S_y^2} + \frac{m_z^2}{2S_z^2}\right) \times \left[w_+\exp\left[-\frac{(m-u_+)^2}{2S^2}\right] + w_-\exp\left[-\frac{(m+u_-)^2}{2S^2}\right]\right] \quad \text{(B2)}$$

where $S_i^2 = \sigma_i^2 + \sigma_0^2$ and we have dropped the subscript x wherever it occurs. When the derivatives of R with respect to m_y and m_z are zero, $m_y = m_z = 0$. The conditions for a minimum of R with respect to m and s are

$$\frac{\partial R}{\partial m} = \frac{2}{(2\pi)^{3/2}S^3S_yS_z}\left[(m-u_+)w_+ \exp\left[-\frac{(m-u_+)^2}{2S^2}\right] + (m+u_-)w_- \exp\left[-\frac{(m+u_-)^2}{2S^2}\right]\right] = 0 \quad \text{(B3a)}$$

$$\frac{\partial R}{\partial s} = \frac{-3}{8\pi^{3/2}s^4} + \frac{2s}{(2\pi)^{3/2}SS_yS_z}\left[\left[\sum\frac{1}{S_i^2} - \frac{(m-u_+)^2}{S^4}\right]w_+ \exp\left[-\frac{(m-u_+)^2}{2S^2}\right] + \left[\sum\frac{1}{S_i^2} - \frac{(m+u_-)^2}{S^4}\right]w_- \exp\left[-\frac{(m+u_-)^2}{2S^2}\right]\right] = 0 \quad \text{(B3b)}$$

where $\sum S_i^{-2}$ is over $i = x,y,z$ and the second derivative with respect to m is

$$\frac{\partial^2 R}{\partial m^2} = \frac{2}{(2\pi)^{3/2}S^3S_yS_z}\left[\left[1 - \frac{(m-u_+)^2}{S^2}\right]\exp\left[\frac{-(m-u_+)^2}{2S^2}\right] + \left[1 - \frac{(m+u_-)^2}{S^2}\right]\exp\left[\frac{-(m+u_-)^2}{2S^2}\right]\right] \quad \text{(B4)}$$

with $\partial^2 R/\partial m^2 > 0$.

When $u_+ \neq u_-$, the conditions for extrema with respect to m and s are

$$\frac{u_- + m}{u_+ - m} = \frac{u_-}{u_+}\exp\left[\frac{-(m-u_+)^2}{2S^2} + \frac{(m+u_-)^2}{2S^2}\right] \quad \text{(B5a)}$$

and

$$3 = \frac{2^{5/2}s^3}{SS_yS_z}\left[\sum_i\frac{s^2}{S_i^2} + \frac{s^2(m+u_-)(m-u_+)}{S^4}\right]\times\left[\frac{u_+ + u_-}{m+u_-}\right]w_+ \exp\left[\frac{-(m-u_+)^2}{2S^2}\right] \quad \text{(B5b)}$$

The condition that an extremum be a minimum with respect to m is

$$s^2 + \sigma_0^2 > (u_- + m)(u_+ - m) \quad \text{(B6)}$$

When $u_+ = u_- \equiv u_0$ and $\sigma_{0x} = \sigma_{0y} = \sigma_{0z} \equiv \sigma_0$, the conditions for a minimum with respect to m (eq B5a and B6) become

$$\frac{2mu_0}{S^2} = \log\frac{u_0 + m}{u_0 - m} \quad \text{(B7)}$$

$$s^2 > u_0^2 - m^2 - \sigma_0^2 \quad \text{(B8)}$$

respectively. Given eq B7, the condition for a minimum with respect to s (eq B5b) is

$$\frac{3}{2^{5/2}s^5} = \frac{\exp\left(\frac{m^2+u_0^2}{2S^2}\right)}{S^5}\left[3 - \frac{(u_0^2-\mu^2)}{S^2}\right]\cosh\frac{mu_0}{2S^2} \quad \text{(B9)}$$

The condition for a minimum with respect to m at $m = 0$ can be found by substituting eq B9 into eq B8:

$$u_0^2 < (1 - \tfrac{1}{2}(\tfrac{9}{4}e)^{1/5})^{-1}\sigma_0^2 \approx 3.55\sigma_0^2 \quad \text{(B10)}$$

We can obtain an approximate expression for s from eq B9:

$$s^2 \approx \sigma_0^2 + \tfrac{1}{3}u_0^2, \quad m = 0 \quad \text{(B11)}$$

$$s^2 \approx (2^{3/5} - 1)^{-1}\sigma_0^2 \approx 1.94\sigma_0^2, \quad m \to u_0$$

References

Artymiuk, P. J., & Blake, C. C. F. (1981) *J. Mol. Biol. 152*, 737–762.

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*, 563–568.

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., & Sarma, V. R. (1967) *Nature (London) 206*, 757.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187–217.

Cava, R. J., Reidinger, F., & Wuensch, B. J. (1980) *Solid State Commun. 24*, 411–416.

Cruickshank, D. W. J. (1974) in *International Tables for X-ray Crystallography*, Vol. 2, Kynoch Press, Birmingham.

Diamond, R. (1971) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A27*, 436–452.

Diamond, R. (1974) *J. Mol. Biol. 82*, 371–391.

Dodson, E. J., Issacs, N. W., & Rollett, J. S. (1976) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A32*, 311–315.

Elber, R., & Karplus, M. (1987) *Science (Washington, D.C.) 235*, 318–321.

Frauenfelder, H., Petsko, G. A., & Tsernoglou, D. (1979) *Nature (London) 280*, 558–563.

Gelin, B. R., & Karplus, M. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2002.

Haneef, I., Glover, I., Tickle, I., Moss, D., Pitts, J., Wood, S., Blundell, T., Hermans, J., & van Gunsteren, W. (1985) *Molecular Dynamics and Protein Structure* (Hermans, J., Ed.) pp 85–91, Polycrystal Book Service, Western Springs, IL.

Hartmann, H., Parak, F., Steigemann, W., Petsko, G. A., Ringe Ponzi, D., & Frauenfelder, H. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4967–4971.

Hendrickson, W. A. (1985) *Methods Enzymol. 115*, 252–270.

Ichiye, T. (1985) Ph.D. Thesis, Harvard University, Cambridge, MA.

Ichiye, T., & Karplus, M. (1987) *Proteins 2*, 236–259.

Ichiye, T., Olafson, B. D., Swaminathan, S., & Karplus, M. (1986) *Biopolymers 25*, 1909–1937.

Jack, A., & Levitt, M. (1978) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A34*, 931–935.

Johnson, C. K. (1980) *Thermal Motion Analysis*, Report 1980 DOE/TIC-11068, Oak Ridge National Laboratory, Oak Ridge, TN.

Johnson, C. K., & Levy, H. A. (1974) in *International Tables for X-ray Crystallography*, Vol. 4, Kynoch Press, Birmingham.

Karplus, M., & McCammon, J. A. (1979) *Nature (London) 277*, 578.

Karplus, M., & McCammon, J. A. (1981) *CRC Crit. Rev. Biochem. 9*, 293–349.

Kendall, M. G., & Stuart, A. (1977) *The Advanced Theory of Statistics*, Vol. I, Griffin, London.

Konnert, J. H. (1976) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A32*, 614–617.
Konnert, J. H., & Hendrickson, W. A. (1980) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A36*, 344–350.
Kuriyan, J., Karplus, M., Levy, R. M., & Petsko, G. A. (1986a) *J. Mol. Biol. 190*, 227–254.
Kuriyan, J., Wilz, S., Karplus, M., & Petsko, G. A. (1986b) *J. Mol. Biol. 192*, 133–154.
Levitt, M. (1983a) *J. Mol. Biol. 168*, 595–620.
Levitt, M. (1983b) *J. Mol. Biol. 168*, 621–657.
Levy, R. M., Perahia, D., & Karplus, M. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 1346–1350.
Lipson, H. S., & Cochran, W. (1966) *The Determination of Crystal Structures*, Bell, London.
Mao, B., Pear, M. R., & McCammon, J. A. (1982) *Biopolymers 21*, 1979–1989.
McCammon, J. A., Gelin, B. R., & Karplus, M. (1977) *Nature (London) 267*, 585.
McCammon, J. A., Wolynes, P. G., & Karplus, M. (1979) *Biochemistry 18*, 927.
Northrup, S. H., Pear, M. R., Morgan, J. D., McCammon, J. A., & Karplus, M. (1981) *J. Mol. Biol. 153*, 1087–1109.
Novotný, J., Handschumacher, M., Haber, E., Bruccoleri, R. E., Carlson, W. B., Fanning, D. W., Smith, J. A., & Rose, G. D. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 226–230.
Summers, N. L., Carlson, W. D., & Karplus, M. (1987) *J. Mol. Biol. 196*, 175–198.
Tainer, J. A., Getzoff, E. D., Alexander, H., Houghten, R. A., Olson, A. J., Lerner, R. A., & Hendrickson, W. A. (1984) *Nature (London) 312*, 127–134.
van Gunsteren, W. F., & Karplus, M. (1982a) *Biochemistry 21*, 2259–2274.
van Gunsteren, W. F., & Karplus, M. (1982b) *Macromolecules 15*, 1528–1544.
Watenpaugh, K. D., Sieker, L. C., Herriott, J. R., & Jensen, L. H. (1973) *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. B29*, 943–956.
Watenpaugh, K. D., Sieker, L. C., & Jensen, L. H. (1979) *J. Mol. Biol. 131*, 509–522.
Westhof, E., Altschuh, D., Moras, D., Bloomer, A. C., Mondragon, A., Klug, A., van Regenmortel, M. H. V. (1984) *Nature (London) 311*, 123–126.
Willis, B. T. M., & Pryor, A. W. (1975) *Thermal Vibrations in Crystallography*, Cambridge University Press, Cambridge.
Yu, H.-A., Karplus, M., & Hendrickson, W. A. (1985) *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. B41*, 191–201.
Zucker, U. H., & Schulz, H. (1982) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A38*, 563–568.

Photocycles of Bacteriorhodopsins Containing 13-Alkyl-Substituted Retinals†

Wolfgang Gärtner,‡,§ Dieter Oesterhelt,*,‡ Jürgen Vogel,‖ Robert Maurer,‖ and Siegfried Schneider‖
Max-Planck-Institut für Biochemie, D-8033 Martinsried, FRG, and Institut für Physikalische Chemie der Technischen Universität München, D-8046 Garching, FRG

Received December 3, 1987

ABSTRACT: Three retinal analogues, carrying at position 13 a hydrogen atom, an ethyl group, or a *n*-propyl group instead of the naturally occurring methyl group, were incorporated into bacterioopsin. The absorption maxima of the newly formed analogue bacteriorhodopsins (BRs) are in the range of 545–565 nm; the positions of those maxima decrease with increasing length of the alkyl substituent. No light/dark adaptation was found for the analogue BRs. Flash photolytic experiments revealed the presence of separate cis and trans photocycles. Each of the analogue BRs forms, within several microseconds, an intermediate with a strongly blue-shifted absorption maximum, comparable to the M-intermediate of the trans photocycle of BR. The M-intermediates of the ethyl and the propyl derivative decay within a few milliseconds, as in BR, and their proton translocating ability correlates with the proportion of the all-trans isomer in the binding sites. The 13-demethylretinal BR (13-dm BR), however, is an exception. Although 15% of the retinal is in the all-trans isomeric state [Gärtner, W., Towner, P., Hopf, H., & Oesterhelt, D. (1983) *Biochemistry 22*, 2637–2644], in the binding site, maximally 3% of proton translocating activity was found. This discrepancy can be resolved if the delay time of about 20 ms of the M-intermediate in the trans photocycle of 13-dm BR is taken into account. The cis photocycle of 13-dm BR apparently consists of only one red-shifted intermediate absorbing around 610 nm and decaying with a half-time of 250 ms back to the initial state. At increasing irradiance of the monitoring light beam secondary photochemistry leads from this intermediate to an M-like state (λ_{max} about 420 nm), which has a half-life of 1.7 s.

The most abundant membrane protein of *Halobacterium halobium*, bacteriorhodopsin (BR),[1] acts as a light-driven proton pump that allows the halobacteria to grow phototrophically [for review see Stoeckenius and Bogomolni (1982) and Oesterhelt and Krippahl (1983)]. The isomeric state of retinal in bacteriorhodopsin can be either all-trans or 13-cis. These species equilibrate in the dark to yield a 1:1 ratio, whereas in the light-adapted form of the chromoprotein nearly

† This work was financially supported by the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 143).
‡ Max-Planck-Institut für Biochemie.
§ Present address: Institut für Biologie I (Zoologie), D-7800 Freiburg, FRG.
‖ Institut für Physikalische Chemie der Technischen Universität München.

[1] Abbreviations: BR, bacteriorhodopsin; BO, bacterioopsin; dm, demethyl; 13-Et, 13-ethyl; 13-Pr, 13-*n*-propyl; 13-dm BR, 13-demethylretinal-containing bacteriorhodopsin; OD, optical density; S/N, signal to noise ratio; I_{560}, intermediate absorbing at 560 nm.

0006-2960/88/0427-3497$01.50/0