# Large-Amplitude Nonlinear Motions in Proteins

Angel E. García

*Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545*

(Received 30 December 1991)

A molecular-dynamics calculation on a hydrated protein, crambin, demonstrates that (i) neighboring dihedral angles are correlated to local transitions in the protein backbone, and that (ii) the amplitude of collective excitations, representing correlated global motions in the protein, samples multicentered distributions. The time dependence of the multicentered dihedral and collective excitations show rapid transitions from the center of one distribution to another, followed for some time by damped, low-amplitude motions around one center. The global nonlinear collective excitations are responsible for most of the atomic fluctuations of the molecule. An analysis appropriate to multimodal conformations is reported.

PACS numbers: 87.15.By, 87.15.He

The thermodynamic stability of proteins is best understood in terms of an ensemble of nearly degenerate substates spanning a large configurational space [1]. Evidence for the existence of these substates has been obtained experimentally [2] and theoretically [3]. However, the dynamical characteristics of such systems have not been studied. Thermally driven protein should exhibit transitions between configurational substates. Therefore, the dynamics of a protein will be characterized by nonlinear modes of oscillations where the protein exhibits fast transitions from one substate to another, in a nonperiodic fashion [4,5]. These transitions should be responsible for the multiple-time-scale processes observed in proteins [6]. The character of these nonlinear modes may involve the correlated motions of a few atoms, i.e., localized modes, or collective motions involving the whole system. We will show that multicentered, nonlinear, collective motions are responsible for most of the atomic fluctuations. These motions are not describable in terms of harmonically or anharmonically [7] perturbed motions around one particular conformation. As a consequence, the comparison of the mean-square displacements of atoms during a simulation with the experimentally observed Debye-Waller factors is termed incorrect [8]. That relationship requires the motions of the proteins to be unimodal and harmonic [9].

In this work (a) we will show that a hydrated protein samples multicentered distributions of dihedral angles that are correlated to local transitions in the protein backbone; (b) in addition, we observe collective nonlinear excitations that exhibit multiple-minima transitions. These motions are extracted from a 240.0-ps, constant-temperature (300 K), molecular-dynamics simulation on a hydrated plant protein, crambin. Crambin is a 46-amino-acid amphipathic protein for which high-resolution x-ray [10], neutron diffraction [11], 1D and 2D NMR studies [12,13], and theoretical studies [14–17] have been reported. A *united-atom* [18] representation of the protein consists of 408 atoms, immersed in a box of approximate dimensions 30 Å×36 Å×42 Å containing 1315 water molecules modeled by a TIP3P potential [19]. A residue-based spherical cutoff distance of 12.5 Å was used to truncate all nonbonding interactions. Periodic boundary conditions were used for Coulombic and van der Waals interactions between atoms. We used potential-energy parameters widely used in the literature [18]. The first 24 ps of the simulation are not included in any averaging, and all the analyses shown in this work are done with the last 216 ps of the trajectory sampled 20 times/ps. The role of specific water molecules in the protein dynamics will not be discussed here, but we simply regard the protein as an open system exchanging energy with the surrounding water.

The analyses of the trajectories obtained from this simulation indicate that the protein motion consists of two main elements.

*Localized nonlinear motions.*—The backbone dihedral angles [20] of the protein $(\phi,\psi)$ show a behavior typical of a system with multiple potential-energy minima. Figure 1 shows histograms of the occupancy of the $\psi$ dihedral angles during the last 216 ps of the simulation. The $\psi$ angles corresponding to the amino acids 2 and 3 and 33 to 35, which form $\beta$-strand [20] conformations, and amino acids 19 to 21 and 40 to 43, which form parts of *turns* [20], show bimodal distributions. The $\alpha$-helical regions of the protein, amino acids 7 to 18 and 23 to 30, show sharp unimodal distributions. The $\phi$ dihedral angles, not shown, exhibit similar multimodal distributions.

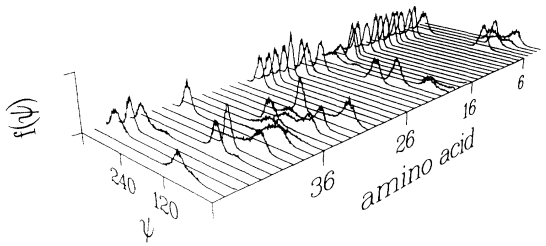


FIG. 1. Histograms of the fraction of occurrences of the backbone dihedral angles $\psi$, in degrees, of crambin during the last 216 ps of simulation. The amino acids are labeled by number.

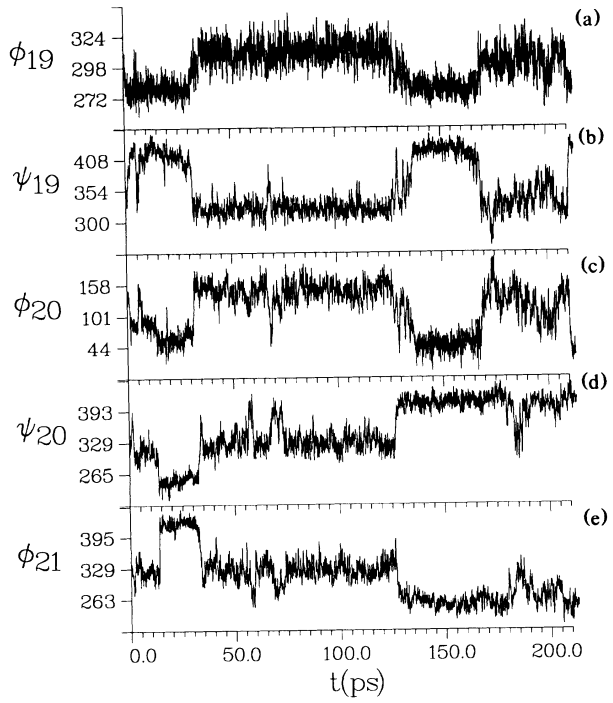


FIG. 2. Dihedral angles $\phi$ and $\psi$ (in degrees) as a function of time (in ps) for the amino acids Pro(19), Gly(20), and Thr(21).

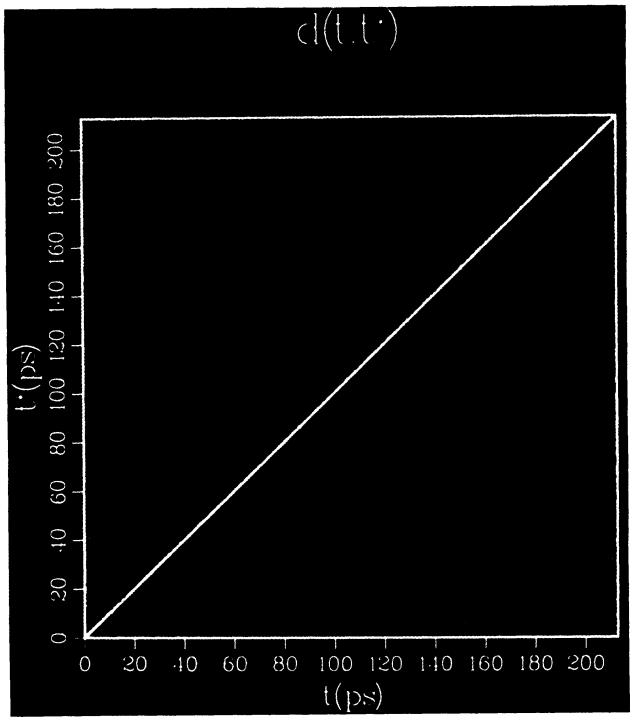


FIG. 3. Contour of the *root-mean-square* distance between pairs of conformations adopted by the protein every 0.5 ps along the last 216 ps of the molecular-dynamics trajectory. The contours are colored according to the distance between two structures sampled at times $t$ and $t^*$ (in ps) [cyan, 0.25 Å $< d \leq$ 0.75 Å; green, 0.75 Å $< d \leq$ 1.25 Å; blue, 1.25 Å $< d \leq$ 1.75 Å; magenta, 1.75 Å $< d \leq$ 2.25 Å; and red, $d >$ 2.25 Å]. The largest rms distance between two configurations is 2.38 Å.

Figures 2(a)–2(e) show the time dependence of the $(\phi,\psi)$ angles near the residues Pro(19) to Thr(21) that form a *turn*, and sample bimodal distributions. These variables show characteristics typical of nonlinear systems [4,5]. That is, there are many fast flips from one conformation to another, with rapid underdamped oscillations between. Equal-time transitions in $\psi_{19}$ and $\phi_{20}$, shown in Figs. 2(b) and 2(c), are *anticorrelated*; i.e., when $\psi_{19}$ changes abruptly from one conformation to another, $\phi_{20}$ also changes abruptly in the opposite sense. The ring nature of the proline residue Pro(19) does not allow $\phi_{19}$, Fig. 2(a), to change from *gauche*$^-$ to *gauche*$^+$ or *trans* [20], but two separate values of $\phi_{19}$ are sampled around *gauche*$^-$. Notice that the pairs of angles $(\phi_{19},\psi_{19})$, $(\psi_{19},\phi_{20})$, and $(\psi_{20},\phi_{21})$ are anticorrelated, but $(\phi_{20},\psi_{20})$ are not anticorrelated over the whole trajectory. These motions have the overall effect of changing the *turn* conformation. At this point we prefer to analyze the protein motions in terms of nonstructural variables that can describe collective, large-amplitude motions.

*Delocalized nonlinear motions.*—In addition to localized, anticorrelated large motions we analyzed the simulations for extended large-amplitude motions. To do so we first employed the $N$-particle *root-mean-square (rms) distance* [21] $d(t,t^*)$ between evolving protein configurations; if this rms distance per particle is large it implies substantial overall motion. The *distance* $d(t,t^*)$ between pairs of conformations at $t,t^*$, respectively, sampled every 0.5 ps, is displayed by color coding in Fig. 3. Note first that the distance (magenta, 1.75 Å $< d \leq$ 2.25 Å) between the initial conformation and those after 40.0 ps is quite large, indicating a substantial overall motion away from it. Proceeding further in time we see a repeated pattern of flipping between green (0.75 Å $< d \leq$ 1.25 Å) and blue (1.25 Å $< d \leq$ 1.75 Å) approximately every 60 ps. This period is much larger than any calculated by normal-mode analysis [7,14], as well as larger than the average time between localized dihedral transitions discussed above. These features indicate extended collective nonlinear motions. It is desirable to find a representation of the collective modes in terms of a set of directions, $\mathbf{m}^{3N}$, which most efficiently describe the atomic fluctuations of the specific protein under study. The directions $\mathbf{m}^{3N}$ are determined by using the following ansatz: Minimize the mean-square distances of the $\{\mathbf{r}_i^{3N}\}$ configurations *normal* to $\mathbf{m}^{3N}$, such that most of the fluctuations will then be along $\mathbf{m}^{3N}$. Introducing a Lagrange multiplier $\lambda$, such that a constraint $\mathbf{m}^{3N}\cdot\mathbf{m}^{3N}=1$ is satisfied, we minimize the functional

$$f(\mathbf{m},\mathbf{y}_0,\lambda)=\frac{1}{L}\sum_{i=1}^{L}\{(\mathbf{r}_i-\mathbf{y}_0)^2-[(\mathbf{r}_i-\mathbf{y}_0)\cdot\mathbf{m}]^2\}+\lambda[\mathbf{m}\cdot\mathbf{m}-1], \quad (1)$$

with respect to $(\mathbf{m},\mathbf{y}_0,\lambda)$. This is equivalent to a generalized least-squares fitting [22] in which a line passing through $\mathbf{y}_0$, with directional cosines $m_\alpha$, is the best approximation of all sampled conformations by a line in $3N$-dimensional space. $f(\mathbf{y}_0,\{m_\alpha,\alpha=1,\ldots,3N\},\lambda)\geq 0$ gives a measure of the reliability of the fitting.

Minimizing with respect to $\mathbf{y}_0$, i.e., $\nabla_{\mathbf{y}_0} f=0$, yields $\mathbf{y}_0=(1/L)\sum_{i=1}^{L}\mathbf{r}_i$, which equals the geometric center of all the sampled conformations, and is the best approximation of all sampled conformations by a point [22]. Minimizing with respect to $\lambda$ yields $\mathbf{m}\cdot\mathbf{m}=1$. Minimizing with respect to the set of $\{m_\alpha\}$, $\alpha=1,\ldots,3N$, yields the matrix eigenvalues equation

$$\vec{\sigma}\cdot\mathbf{m}=\lambda\mathbf{m}\,, \tag{2}$$

where

$$\sigma_{\alpha,\beta}=\frac{1}{L}\sum_{i=1}^{L}(r_i-y_0)_\alpha(r_i-y_0)_\beta\,. \tag{3}$$

Equation (2) has eigenvalues $\lambda_\nu$ and eigenvectors $\mathbf{m}_\nu$, $\nu=1,\ldots,3N$, corresponding to the various eigendirections which characterize the motions seen in the simulation. Using $\mathbf{m}_\nu$ as a basis it is straightforward to show that the $\nu$th eigendirection has mean-square fluctuations normal to it given by

$$\langle d^2(\mathbf{m}_\nu)\rangle=f(\mathbf{m}_\nu,\mathbf{y}_0,\lambda_\nu)=\mathrm{Tr}(\sigma)-\lambda_\nu\,. \tag{4}$$

That is, the eigendirection with largest $\lambda_\nu$ best represents the predominant motion of the protein. Eigenvectors and eigenvalues can be computed from the simulation data. Having done so the projection of the motion $\mathbf{r}^{3N}(t)$ along a given eigendirection $\mathbf{m}_\nu$, $p_\nu(t)=\mathbf{r}(t)\cdot\mathbf{m}_\nu$, yields a picture of the motion in a set of generalized natural coordinates for the particular protein. There is no assumption that this decomposition is harmonic or quasiharmonic. Indeed, Figs. 4(a)–4(e) show the projections for the first five directions (left) and the histograms (right) of the occurrence of all the values of $p(t)$ for the same vectors. The histograms of the population distributions of $p(t)$ can be fitted by multicentered distributions with two or three clearly distinguishable centers. The projection corresponding to the largest eigenvalue, $\lambda_1=131.2$ Å$^2$, Fig. 4(a), shows a bimodal distribution. The distribution centered at $p_1(t)\approx 20.0$ Å is only sampled during the first 40 ps of the dynamic trajectory. A transition from that minimum to the distribution centered at $\approx -7.0$ Å occurs during the next 20 ps. Similar behavior was shown in Fig. 3. This bimodal distribution could be interpreted as either a nonlinear transition from one conformation to another or, since we observed only one transition, a transient in the dynamics from a metastable high-energy initial conformation to a lower-energy conformation. Given the length of our simulation we cannot distinguish one case from the other, but it would be proper to say that it is due to a metastable initial conformation.

The distribution corresponding to the second largest eigenvalue, $\lambda_2=44.4$ Å$^2$, is shown in Fig. 4(b). This distribution has three centers with maxima at $-7$, 1, and 10 Å. The time history of the projection of the trajectories along this eigenvector shows frequent transitions from one center to another at time intervals larger than 60 ps. Transitions to each center occur at least twice, and, therefore, the motions along this direction describe equilibrium transitions from one protein conformation to another. Similar curves are shown for the third ($\lambda_3=37.9$ Å$^2$), fourth ($\lambda_4=17.0$ Å$^2$), and fifth ($\lambda_5=16.7$ Å$^2$) eigenvalues. The distributions for vectors six and seven ($\lambda_6=10.7$ Å$^2$, $\lambda_7=8.9$ Å$^2$), not shown, exhibit multicentered distributions, but the maxima of the distributions are very close to one another and cannot be clearly distinguished. We do not observe markedly multicentered distributions for vectors with eigenvalues smaller than $\lambda_7$, indicating unimodal, though possibly quasiharmonic, motions.

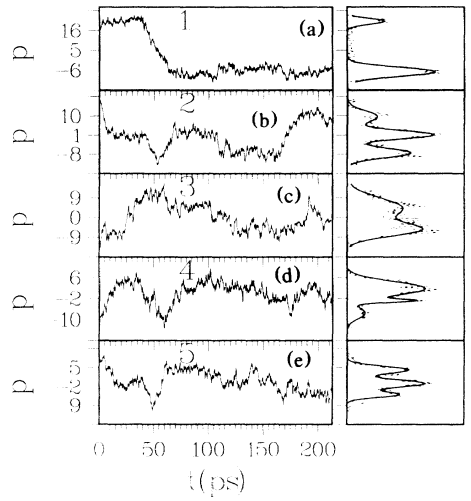


FIG. 4. Projections $p_\nu(t)$ of the protein trajectory along the five largest eigenvalue vectors of Eq. (4) are shown in left-hand plots. The right-hand-side plots show histograms of the populations of occurrence of all the values of $p(t)$ for the corresponding vectors. The solid line shows a fitting of the distributions by sums of two [(a) and (c)] or three [(b),(d), and (e)] Gaussians. The dashed line shows the distributions calculated from the simulation. $p_\nu(t)$ is given in Å, and $t$ in ps.

It should be added that the *mean-square* displacements of atoms around their average positions are described by $\langle u^2\rangle=(1/N)\mathrm{tr}(\sigma)=(1/N)\sum_{\nu=1}^{3N}\lambda_\nu$. In this calculation, the trace of $\sigma$ is 386.0 Å$^2$, which gives a value for $\langle u^2\rangle$ of 0.946 Å$^2$. Note that the first configurational eigenvector contributes 34% of the total $\langle u^2\rangle$. The first five directions contribute 64% of $\langle u^2\rangle$, while the remaining directions, which show unimodal displacements, contribute 36%. That is, $\langle u^2\rangle$ has contributions from markedly nonharmonic motions. Therefore, the comparison of $\langle u^2\rangle$ to the Debye-Waller factors, which is equivalent to assuming a Gaussian distribution of the fluctuations around one conformation, is not correct [5].

From the results presented here we can conclude that nonlinear motions describing oscillations around multicentered distributions are responsible for most of the atomic fluctuations in proteins. These atomic fluctuations may be used by proteins during catalysis and, indeed, biological evolution may have optimized and exploited the role of what has been called the time dimension of protein structure [6].

I thank J. A. Krumhansl, J. Berendzen, J. Hay, B. Goldstein, G. Gupta, and A. Redondo for comments and suggestions. This work was funded by the AICD Biocatalysis Program at the U.S. Department of Energy. All computations were done at the Advanced Computing Laboratory (ACL) facilities at Los Alamos National Laboratory.

[1] H. Frauenfelder, H. A. Siglar, and R. D. Young, Science **254**, 1598 (1991); H. Frauenfelder, P. J. Steinbach, and R. D. Young, Chem. Scr. **29A**, 145 (1989).

[2] A. Ansari *et al.*, Proc. Natl. Acad. Sci. (U.S.A.) **82**, 5000 (1985); I. E. T. Iben *et al.*, Phys. Rev. Lett. **62**, 1916 (1989).

[3] R. Elber and M. Karplus, Science **235**, 318 (1987); N. Go and T. Noguti, Chem. Scr. **29A**, 151 (1989).

[4] E. Helfand, Physica (Amsterdam) **118A**, 123 (1983).

[5] J. A. Krumhansl, in *Proceedings in Life Sciences: Protein Structure, Molecular, and Electronic Reactivity*, edited by R. H. Austin *et al.* (Springer-Verlag, New York, 1987).

[6] S. W. Englander and N. R. Kallenbach, Quarterly Rev. Biophys. **16**, 521 (1984).

[7] M. Karplus and J. N. Kushick, Macromolecules **14**, 325 (1981); T. Noguti and N. Go, Nature (London) **296**, 433 (1982); T. Ichiye and M. Karplus, Biochemistry **27**, 3487 (1988).

[8] C. L. Brooks, M. Karplus, and B. Montgomery-Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, Advances in Chemical Physics Vol. LXXI (Wiley, New York, 1988).

[9] J. A. Krumhansl, in *Computer Analysis for Life Science*, edited by C. Kawabata and A. R. Bishop (Ohmsha, LTD, Tokyo, Japan, 1985), pp. 78–88.

[10] W. A. Hendrickson and M. M. Teeter, Nature (London) **290**, 107 (1981).

[11] M. M. Teeter, Proc. Natl. Acad. Sci. (U.S.A.) **81**, 6014 (1984).

[12] M. Llinas *et al.*, Biochemistry **19**, 1140 (1980).

[13] J. A. W. H. Vermeulen *et al.*, Proc. FEBS Meet. **219**, 426 (1987).

[14] M. Levitt, C. Sander, and P. S. Stern, J. Mol. Biol. **181**, 423 (1985).

[15] M. M. Teeter, and D. A. Case, J. Phys. Chem. **94**, 8091 (1990).

[16] W. L. Jorgensen and J. Tirado-Rives, J. Am. Chem. Soc. **110**, 1657 (1988).

[17] M. Withlow and M. M. Teeter, J. Am. Chem. Soc. **108**, 7163 (1986).

[18] S. J. Weiner *et al.*, J. Am. Chem. Soc. **106**, 765 (1984).

[19] W. L. Jorgensen *et al.*, J. Chem. Phys. **79**, 926 (1983).

[20] G. E. Schulz and R. H. Schirmer, *Principles of Protein Structure* (Springer-Verlag, New York, 1978).

[21] J. McLachalan, J. Mol. Biol. **128**, 49 (1979).

[22] D. M. Soumpasis, C-S. Tung, and A. E. García, J. Biomol. Struct. Dyn. **8**, 867 (1991).

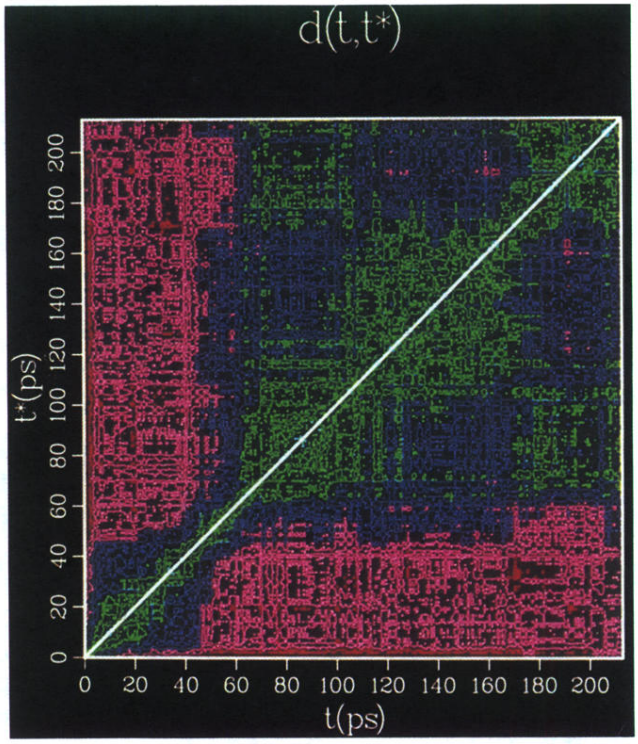


FIG. 3. Contour of the *root-mean-square* distance between pairs of conformations adopted by the protein every 0.5 ps along the last 216 ps of the molecular-dynamics trajectory. The contours are colored according to the distance between two structures sampled at times $t$ and $t^*$ (in ps) [cyan, 0.25 Å $< d \leq$ 0.75 Å; green, 0.75 Å $< d \leq$ 1.25 Å; blue, 1.25 Å $< d \leq$ 1.75 Å; magenta, 1.75 Å $< d \leq$ 2.25 Å; and red, $d >$ 2.25 Å]. The largest rms distance between two configurations is 2.38 Å.