Gaussian Dynamics of Folded Proteins

Turkan Haliloglu, Ivet Bahar,* and Burak Erman

Polymer Research Center and School of Engineering, Bogazici University, Bebek 80815, Istanbul, Turkey and TUBITAK Advanced Polymeric Materials Research Center, Bebek 80815, Istanbul, Turkey

(Received 7 April 1997)

Vibrational dynamics of folded proteins is studied using a Gaussian model in which the protein is viewed as a network, residues representing the junctions, and the connectivity being established by a single parameter harmonic potential. Application to seven proteins showed that the local packing density plays a major role in determining the vibrational spectrum at time scales of picoseconds. At later times, the secondary structure and tertiary context of each residue comes into play. The vibrational frequencies obey a universal distribution, confirming previous normal mode analyses. [S0031-9007(97)04215-4]

PACS numbers: 87.15.By, 87.15.He

Two recent papers [1,2] on the vibrational motions of globular proteins have led to significant simplification in our understanding of the dynamics of folded proteins. ben-Avraham [1] found that the density distribution of slow vibrational modes of globular proteins follows a characteristic, universal curve when expressed as a function of frequency. Slow vibrational modes are those resulting from the cooperative fluctuations of the α carbons on the backbone chain. The existence of such a universal curve is attributed to the main structural similarities between proteins [1]. Amplitudes of these atomic fluctuations are determined experimentally, and reported in the Protein Data Bank (PDB) [3]. The observation of a universal dispersion curve by ben-Avraham is significant because it directs attention to the generic nature of interactions in proteins in the native state. Tirion [2] showed, by normal mode analysis, that a single parameter harmonic potential reproduces in good detail the large amplitude motions of proteins in the native state. This is consistent with our recent analysis of 302 nonhomologous structures from PDB, in which the stability of native proteins is shown to be imparted predominantly by uniform nonspecific potentials of mean force, representative of average interresidue interactions in folded proteins [4].

In line with these plausible conjectures, we proposed [5] a model for the folded protein in which interactions between residues in close proximity are replaced by linear springs, in analogy with the elasticity theory of random polymer networks [6–8]. The model assumes that the protein in the folded state is equivalent to a three dimensional elastic network. The junctions are identified with the C^α atoms in the protein. These undergo Gaussian distributed fluctuations. We refer to this model as the Gaussian network model (GNM). By adopting a single parameter for the harmonic potential, following Tirion [2], we were able to predict the equilibrium fluctuations of the C^α atoms of several proteins with almost perfect agreement with experiments [5]. The specific aim of the present work is to explore the vibrational dynamics of proteins by the GNM.

According to GNM, the equilibrium correlation between fluctuations $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_j$ of two α carbons i and j is given by

$$\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle = (k_B T/\gamma)[\Gamma^{-1}]_{ij}\,, \tag{1}$$

where Γ is a symmetric matrix known as Kirchhoff or connectivity matrix [9], the subscript ij denotes the ijth element, $\mathbf{R}_i$ is the position vector of the ith α carbon, k_B is the Boltzmann constant, T is the absolute temperature, and γ is the single parameter (force constant) of the Hookean pairwise potential used by Tirion [2] for representing the interresidue interactions in the folded structure. The elements of Γ are given by

$$\Gamma_{ij} = \left\{ \begin{array}{ll} -1 & \text{if } i \neq j \text{ and } R_{ij} \leq r_c\,, \\ 0 & \text{if } i \neq j \text{ and } R_{ij} > r_c\,, \\ -\sum\limits_{i,i\neq j} \Gamma_{ij} & \text{if } i = j\,. \end{array} \right\}. \tag{2}$$

Here r_c is the cutoff separation defining the range of interaction of nonbonded α carbons, and R_{ij} is the distance between the ith and jth C^α atoms. A reasonable cutoff distance including all residue pairs within a first interaction shell is 7.0 Å [5,10]. The ith diagonal element of Γ characterizes the *local packing density* or the *coordination number* of residue $i, 1 \leq i \leq n$, for a protein of n residues. The inverse of Γ may be written as

$$\Gamma^{-1} = \mathbf{U}(\Lambda^{-1})\mathbf{U}^T. \tag{3}$$

Here $\mathbf{U}$ is an orthogonal matrix whose columns $\mathbf{u}_i, 1 \leq i \leq n$, are the eigenvectors of Γ, and Λ is the diagonal matrix of the eigenvalues (λ_i) of Γ, usually organized in ascending order $\lambda_1 = 0 < \lambda_2 < \cdots < \lambda_n$. Mean-square fluctuations of the C^α atoms and the cross correlations between the fluctuations of the C^α atoms are found using the respective diagonal and off-diagonal elements of Γ^{-1} in Eq. (1).

 0031-9007/97/79(16)/3090(4)$10.00 © 1997 The American Physical Society

It is possible to decompose Γ^{-1} as the sum of contributions from individual modes as

$$\Gamma^{-1} = \sum_{k=2}^{n} \lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T = \sum_{k=2}^{n} \mathbf{A}^{(k)}. \quad (4)$$

Here $\mathbf{A}^{(k)}$ is the $n \times n$ matrix describing the contribution of the kth vibrational mode to atomic fluctuations. The first eigenvalue of Γ, identically equal to zero, is not included in the summation of Eq. (4).

The correlation of the fluctuations of ith and jth C^α atoms can be expressed as the sum of the contribution of individual modes as

$$\langle \Delta\mathbf{R}_i(0) \cdot \Delta\mathbf{R}_j(t) \rangle = \sum_k A_{ij}^{(k)} \exp\{-\lambda_k t/\tau_0\}, \quad (5)$$

where $A_{ij}^{(k)}$ is the ijth element of $\mathbf{A}^{(k)}$, and τ_0 is a characteristic time, which may be expressed in terms of the effective friction coefficient ζ and force constant γ as

$$\tau_0 = \zeta/\gamma. \quad (6)$$

Equations (5) and (6) follow directly from the application of the Langevin equation of motion to the vibrational dynamics of residues [11]. The time τ_0 will be shown below to have a fixed value of $\tau_0 = 6$ ps, which is characteristic of the vibrational dynamics of *all* proteins in their folded state.

Previous calculations demonstrated that the GNM yields a satisfactory description of the thermal fluctuations of α carbons of proteins in their native state [5]. We display in Fig. 1(a) the mean-square fluctuations of C^α atoms of apomyoglobin, $\langle \Delta R_i^2 \rangle$ as a function of residue index i, together with the Debye-Waller or temperature factors $B_i = 8\pi^2 \langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_i \rangle/3$, measured by x-ray crystallography [12]. Bold and dashed curves depict the theoretical and experimental results, respectively. The theoretical curve is normalized by taking $\gamma(\text{Å}^{-2}) = 2.06\ k_B T$, so as to match the area enclosed by the two curves. The agreement between the theoretical and experimental curves is remarkable.

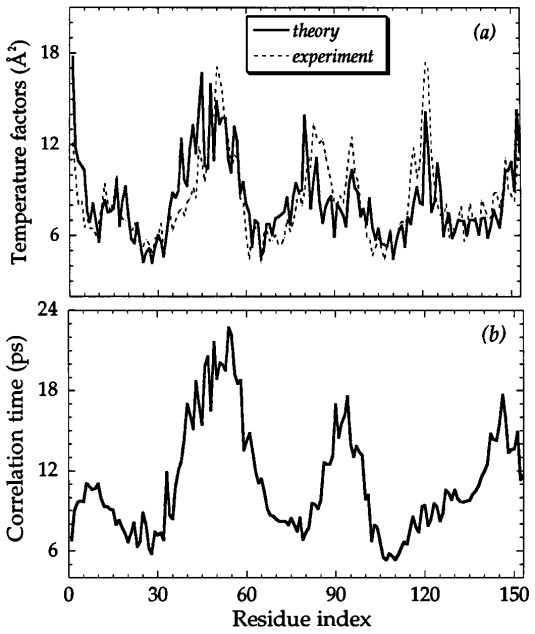

FIG. 1. (a) Temperature factors for apomyoglobin as a function of residue index. Boldface curve is obtained from present theory, dashed curve from experimental data [12]. (b) Correlation times τ_{ii} for the relaxation of C^α atoms.

The autocorrelation functions associated with the fluctuations of C^α atoms evaluated from Eq. (5) showed a single exponential decay for $t \leq 1$ ps, followed by a stretched exponential decay with exponent $\beta = 0.5$ at longer times. In the case of cross correlations, on the other hand, the same stretched exponential behavior is observed in the range $t \geq 1$ ps, while the initial slope of the decay curves approaches zero. The correlation time τ_{ij} associated with time decay of cross correlations $\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle$ can be evaluated from

$$\tau_{ij} = \int_0^\infty \frac{\langle \Delta\mathbf{R}_i(0) \cdot \Delta\mathbf{R}_j(t) \rangle}{\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle} dt = \frac{\tau_0}{\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle} \sum_{k=2}^{n} \frac{A_{ij}^{(k)}}{\lambda_k}. \quad (7)$$

Figure 1(b) displays the correlation times τ_{ij} evaluated for the time-delayed autocorrelations ($i = j$) of all residues in apomyoglobin. Residues exhibiting smaller amplitude fluctuations [Fig. 1(a)] have generally shorter correlation times, although there is not necessarily a one-to-one correspondence. In order to understand the relationship between the correlation times and amplitudes of equilibrium fluctuations, the time evolution of vibrational motions of individual residues is analyzed below in relation to their local packing density and secondary structure.

We consider the time required for each residue to reach 5% of its complete loss of correlation. Let us denote this time as τ_{ii} (0.05), the argument denoting the fractional extent of relaxation. Figure 2(a) shows the τ_{ii} (0.05) values obtained for apomyoglobin residues as a function of the mean-square amplitude $\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_i \rangle$. Here, residues having the same number of neighbors (i.e., same Γ_{ii} values) are designated by the same symbol, as indicated in the legend. At this early stage of relaxation, residues lie on a single curve, approximately.

In Fig. 2(b), the times τ_{ii} (0.50) required for the autocorrelation functions to relax by 50% are shown. The points are not on the same curve anymore. Instead, a family of curves is observed, fitted by dashed lines on the figure. Each curve is distinguished by a given coordination number Γ_{ii}, i.e., residues which have the same number of nonbonded contacts lie on distinct curves. The Γ_{ii} values increase as we go from right to left along the family of curves. Among residues having a common temperature factor (i.e., same amplitude of equilibrium fluctuations, fixed value on the abscissa) those experiencing a higher packing density on a local scale undergo a

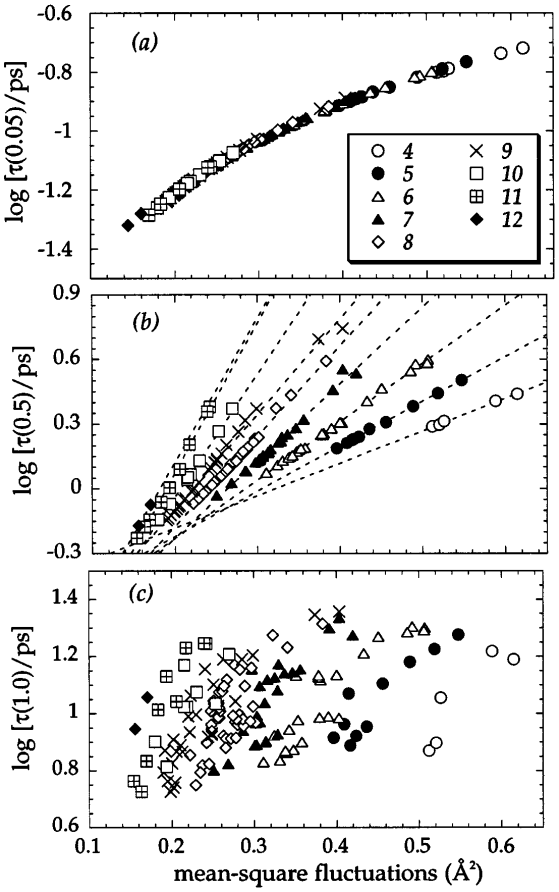

FIG. 2. Time evolution of fluctuation amplitudes as a function of mean-square amplitudes of fluctuations. (a) and (b) refer to the times τ_{ii} (0.05) and τ_{ii} (0.5) required for 5% and 50%, respectively, of the full decay of fluctuation autocorrelation function. (c) displays the correlation times for full decay. Each symbol refers to a residue in apomyoglobin. Dashed lines are drawn to guide the eye. Numbers in the legend indicate local packing density of residues.

slower loss of correlation, as evidenced by their higher correlation times, compared to those with lower packing density. Alternatively, two residues exhibiting the same rate of vibrational relaxation may differ in their equilibrium fluctuations depending on their local packing densities. The one having a higher packing density exhibits smaller amplitude fluctuations.

At later stages of relaxation the classification of residues according to their packing densities becomes less pronounced. Figure 2(c) depicts the correlation times τ_{ii} (1.00) calculated from the full decay of the autocorrelations as a function of equilibrium mean-square fluctuations. Residues having the same packing densities are not aligned anymore on well-defined distinct curves. Evidently, other factors, such as the particular secondary structure and tertiary context come into play. Further examination shows that residues belonging to distinct helices indeed exhibit a preference to form clusters in the figure. In the case of apomyoglobin, a protein consisting of eight helices *A*–*H* arranged in parallel, residues belonging to helix *G* lie in the lower left portion of the correlation time vs fluctuation amplitude diagram; whereas in the other extreme region of high fluctuation amplitudes and high correlation times, we observe a clustering of residues belonging to helix *D*, followed by helix *C*. These two helices are observed by NMR spectroscopy to be involved in either the early unfolding stage [13] or the late folding stage [14] of apomyoglobin.

The eigenvalue decomposition of the inverse Kirchhoff matrix Γ^{-1} has been carried out for a set of proteins (Table I) to study the general properties of the relaxation modes. The density distribution $g(\lambda)$ of the eigenfrequencies is plotted for different proteins in Fig. 3. The solid best fitting curve is obtained by considering all data points. The data appear to collapse into a universal curve similar to the one first proposed by ben-Avraham [1], which is derived from a classical normal mode analysis. A major difference is the portion of our distribution curve in the lowest frequency region of the spectrum ($\lambda < 5$), where an increase is observed with decreasing λ. By matching the peak of our distribution curve with that obtained by ben-Avraham [1], the characteristic time τ_0 appearing in Eq. (5) is evaluated as 6.0 ps. This value may be viewed as a universal parameter characteristic of the vibrational dynamics of residues in folded proteins.

The cumulative density $G(\lambda)$ of the modes is plotted in Fig. 4 as a function of eigenfrequencies. Points display results obtained for the same set of seven proteins (Table I). Over the complete frequency range the points fall on a sigmoidal curve which is fitted by a cubic

TABLE I. Proteins used in calculations.

PDB code	N[a]	Name	Resolution (Å)	Ref.
5pti	58	Bovine pancreatic trypsin inhibitor	1.0	[15]
1hoe	74	α–Amylase inhibitor	2.0	[16]
1tho	108	*E. coli* thioredoxin	1.68	[17]
1bni	108	Barnase (wild type, *p*H 6)	2.0	[18]
1ccr	112	Rice ferrytochrome *c*	2.0	[19]
1bvc	153	Biliverdin apomyoglobin complex (D)	1.5	[12]
3lzm	164	T4 lysozyme	1.7	[20]

[a]Number of residues.

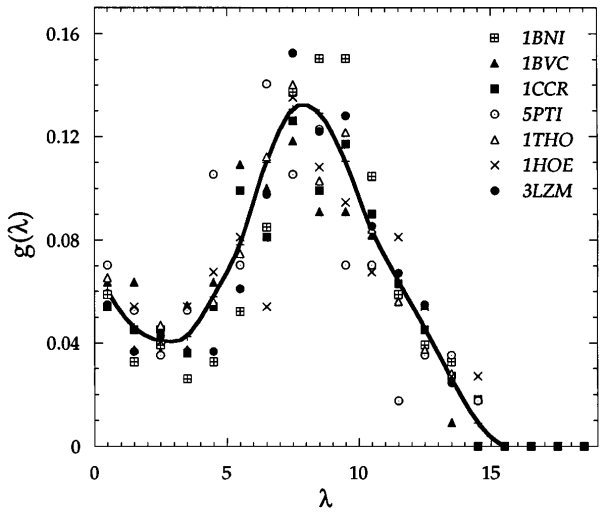

FIG. 3. Density of relaxation modes for different proteins, $g(\lambda)$. PDB codes of the proteins are indicated (see Table I).

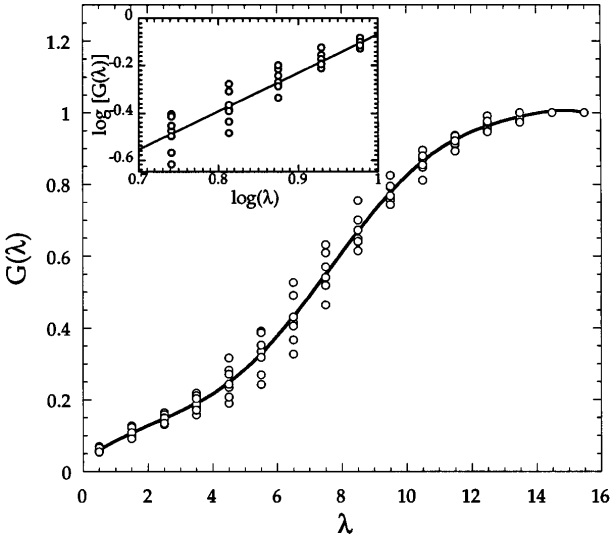

FIG. 4. Cumulative distribution of vibrational modes, $G(\lambda)$, obtained from Fig. 3. The inset is the log-log plot of the intermediate region.

relationship in the figure. The part of the curve for $\lambda < 4$ corresponds to the initial portion of the curve shown in Fig. 3. A log-log plot of this part of the curve yields a straight line with slope equal to unity. The intermediate portion of the curve for the range $5 < \lambda < 10$, which is the counterpart of the slow mode region discussed by ben-Avraham [1], is shown in the inset in Fig. 4 on a logarithmic scale. The slope of the best fitting line is found as 1.63, indicating a power law of the form $G(\lambda) \sim \lambda^{1.63}$. This exponent is lower than the value 2 obtained by ben-Avraham by normal mode analysis [1]. We have also observed that the exponent is dependent on the size of the protein, approaching a value of 2 in the case of larger proteins. Our calculations, as well as the previous normal mode analyses, reflect the anomalous nature of the slow vibrational modes of proteins, which is different from regular crystals where $G(\lambda)$ scales with λ^3.

In conclusion, the present Letter introduces a simple method to analyze the dynamic correlations in native proteins, based on a Gaussian model of nonbonded interactions. The study of protein dynamics is hampered by the size of the systems and the complexity of the potential functions in use. With this model, one can study the slow motions of even the largest known proteins. The proposed method directs attention to novel findings on protein dynamics: At early times the relaxation of modes is universal, at intermediate times it is dominated by the local density of packing, and at longer times the effects of the particular secondary structure appear.

Partial support from Bogazici University Research Funds Project No. 97P003 is gratefully acknowledged.

*Electronic address: bahar@prc.bme.boun.edu.tr

[1] D. ben-Avraham, Phys. Rev. B **47**, 14 559 (1993).
[2] M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
[3] F. C. Bernstein *et al.*, J. Mol. Biol. **112**, 535 (1977).
[4] I. Bahar and R. L. Jernigan, J. Mol. Biol. **266**, 195 (1997).
[5] I. Bahar, A. R. Atilgan, and B. Erman, Folding Des. **2**, 173 (1997).
[6] P. J. Flory, Proc. R. Soc. London A **351**, 351 (1976).
[7] D. S. Pearson, Macromolecules **10**, 696 (1977).
[8] A. Kloczkowski, J. E. Mark, and B. Erman, Macromolecules **22**, 1423 (1989).
[9] F. Harary, *Graph Theory* (Addison-Wesley, Reading, MA, 1971).
[10] S. Miyazawa and R. L. Jernigan, Macromolecules **18**, 534 (1985).
[11] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon, Oxford, 1986), Vol. 73.
[12] U. G. Wagner *et al.*, J. Mol. Biol. **247**, 326 (1995).
[13] M. J. Cocco and J. T. Lecomte, Protein Sci. **3**, 267 (1994).
[14] P. A. Jennings and P. E. Wright, Science **262**, 892 (1993).
[15] M. M. Teeter, Proc. Natl. Acad. Sci. U.S.A. **81**, 6014 (1984).
[16] J. W. Pflugrath *et al.*, J. Mol. Biol. **189**, 383 (1986).
[17] S. K. Katti, D. M. LeMaster, and H. Eklund, J. Mol. Biol. **212**, 167 (1990).
[18] A. M. Buckle, K. Henrick, and A. R. Fersht, J. Mol. Biol. **234**, 847 (1993).
[19] H. Ochiet *et al.*, J. Mol. Biol. **166**, 407 (1983).
[20] L. H. Weaver and B. W. Matthews, J. Mol. Biol. **193**, 189 (1987).