# Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability

Ivet Bahar, Ali Rana Atilgan, Melik C. Demirel, and Burak Erman

*Polymer Research Center, Bogazici University, and TUBITAK Advanced Polymeric Materials Research Center, Bebek 80815, Istanbul, Turkey*

(Received 10 November 1997)

A single-parameter harmonic Hamiltonian based on local packing density and contact topology is proposed for studying residue fluctuations in native proteins. The internal energy obeys an equipartition law, and free energy changes result from entropy fluctuations only. Frequency–wave-number maps show communication between residues involved in slow and fast modes. Fast modes are strongly localized, resulting from the geometric irregularity of the structure. Comparison with experiments shows that slow and fast modes are associated, respectively, with function and stability. Specifically, domain motions and folding cores of HIV-1 protease are accurately identified. [S0031-9007(98)05626-9]

PACS numbers: 87.15.By, 87.15.He

The equilibrium dynamics of proteins in the folded state result from a delicate balance between large opposing effects. On the entropic side, there is a configurational entropy gain upon unfolding, whereas, there is a decrease in entropy due to hydrophobic effect and vibrational fluctuations away from mean atomic positions. As to the internal energy or enthalpy of the system, the folded state is known to be stabilized by a network of hydrogen bonds, along with attractive inter-residue interactions favoring a hydrophobic core. These effects drive the protein to maximize its density, and lead to residue fluctuations in the folded state, which bear close resemblance to density fluctuations in disordered condensed matter [1]. Such fluctuations are well described by the Gaussian network model (GNM) of proteins [2,3] based on the internal Hamiltonian [4]

$$\mathcal{H} = \tfrac{1}{2}\gamma[\Delta\mathbf{R}^{\mathrm{T}}(\Gamma \otimes \mathbf{E})\,\Delta\mathbf{R}], \tag{1}$$

where $\gamma$ is the single parameter (force constant) of the Hookean potential originally proposed by Tirion [5] for representing all inter-residue interactions in the folded structures. $\{\Delta\mathbf{R}\}$ represents the $3N$-dimensional column vector of the $X$, $Y$, and $Z$ components of the fluctuation vectors $\Delta\mathbf{R}_1, \Delta\mathbf{R}_2, \ldots, \Delta\mathbf{R}_N$ of the $C^\alpha$ atoms, $N$ in number, the superscript T denotes the transpose, $\otimes$ is the direct product, $\mathbf{E}$ is the third order identity matrix, and $\Gamma$ is the $N \times N$ symmetric Kirchhoff matrix [6], the elements of which are [2]

$$\Gamma_{ij} = \left\{ \begin{array}{ll} -H(r_c - r_{ij}), & i \neq j \\ -\sum\limits_{i(\neq j)}^{N} \Gamma_{ij}, & i = j \end{array} \right\}. \tag{2}$$

Here, $r_{ij}$ is the distance between the $i$th and $j$th $C^\alpha$ atoms, $H(x)$ is the Heaviside step function given by $H(x) = 1$ for $x > 0$ and $H(x) = 0$ for $x \leq 0$, and $r_c$ is the upper limit for the separation between two "contacting" residues. The value $r_c = 7.0$ Å was adopted in previous GNM studies [2,3], this distance including all neighbors within the first coordination shell around a central residue [7,8].

In the present Letter, the modal decomposition of the vibrational dynamics of proteins using the GNM is considered. We propose, using HIV-1 protease as an illustrative example, that the slow modes of motion revealed by the GNM are associated with the function of the protein, and the fast modes underlie stability. Furthermore, a cooperativity among residues involved in slow and fast modes emerges from the examination of the frequency–wave-number maps of vibrational modes. This suggests an energy flow from kinetically hot residues, which are active in the fastest modes of motion, to those undergoing slower, but functionally relevant, motions.

In the native state, the protein assumes a conformation (of the multidimensional energy surface) that is energy minimized with respect to all residue fluctuations. The protein may be regarded as a canonical ensemble in contact with a heat bath. The vibrational contribution to the Helmholtz free energy of the protein is [9,10]

$$A = -k_B T \ln Z_N = -(3k_B T/2)\ln[(\pi/\gamma^*)^{N-1}\det(\Gamma^{-1})], \tag{3}$$

where $Z_N$ is the configurational integral part of the vibrational partition function given by $Z_N = \int \exp\{-\mathcal{H}/k_B T\}\, d\{\Delta\mathbf{R}\}$, and $\gamma^* = \gamma/2k_B T$. The last equality in Eq. (3) follows from the integration of the single parameter multivariate Gaussian function in the configurational integral, originally given by Flory [4]. Note that the generalized inverse of the Kirchhoff matrix is taken here after eliminating the zero eigenvalue. On the other hand, the average Hamiltonian in $A = \langle\mathcal{H}\rangle - TS$ may be expressed in terms of the matrices $\mathbf{U}$ and $\Lambda$ of the eigenvectors ($\mathbf{u}_i$) and eigenvalues ($\lambda_i$) of $\Gamma$ as

$$\begin{aligned} \langle\mathcal{H}\rangle &= \tfrac{1}{2}\gamma\langle\Delta\mathbf{R}^{\mathrm{T}}\mathbf{U}\Lambda\mathbf{U}^{\mathrm{T}}\Delta\mathbf{R}\rangle = \tfrac{1}{2}\gamma\sum_{i=2}^{N}\lambda_i\langle\Delta r_i^2\rangle \\ &= \tfrac{3}{2}(N-1)k_B T, \end{aligned} \tag{4}$$

  

where $\Delta r_i$ refers to the fluctuations $\Delta r_i \equiv \mathbf{U}^{\mathrm{T}} \Delta \mathbf{R}_i$ in the mode space spanned by the eigenvectors. Equation (4) reveals the equipartition law for the energies over the $(N-1)$ internal modes of motion. The last equality in Eq. (4) follows from the fact that $\langle \Delta r_i^2 \rangle$ is the $i$th diagonal element of the correlation matrix $\langle \Delta \mathbf{r}\, \Delta \mathbf{r}^{\mathrm{T}} \rangle = \mathbf{U}^{\mathrm{T}} \langle \Delta \mathbf{R}\, \Delta \mathbf{R}^{\mathrm{T}} \rangle \mathbf{U} = \mathbf{U}^{\mathrm{T}} \langle (3k_B T/\gamma) \Gamma^{-1} \rangle \mathbf{U} = (3k_B T/\gamma) \Lambda^{-1}$. In other words, the mean-square displacements in mode space are inversely proportional to the eigenvalues, as $\langle \Delta r_i^2 \rangle = (3k_B T/\gamma) \lambda_i^{-1}$.

The Helmholtz free energy change accompanying vibrational motions is, thus, fully determined by the entropic effect, $\langle \mathcal{H} \rangle$ being invariant with respect to changes in conformation. Our recent interpretation of hydrogen exchange data using GNM also supports the dominance of entropic effects [11]. The vibrational free energy may alternatively be written as a sum over the contributions of the individual modes, after rearranging Eq. (3) as

$$A = \sum_{i=2}^{N} (A)_i = (3k_B T/2) \sum_{i=2}^{N} [1 - \ln(\lambda_i^{-1} \pi e/\gamma^*)], \quad (5)$$

which, using Eq. (4), yields $S_i = (3k_B/2) \ln(\lambda_i^{-1} \pi e/\gamma^*)$ for the entropic contribution of the $i$th mode.

In the scope of the free energy landscape characterizing the conformational state of the protein in mode space, we conclude that each mode in the GNM is represented by an equally deep energy well, $3k_B T/2$, while the curvature at the minima (or the steepness of the walls surrounding these wells) accounting for by the entropic term, differs from one mode to another, and fully determines the dynamics in the folded state. Modes characterized by steeper energy walls, i.e., larger $\lambda_i$ values, are more localized, in the sense originally proposed by Anderson [12,13]: The fluctuations of a given amplitude associated with these modes are accompanied by a larger decrease in entropy, compared to those driven by the slower (smaller $\lambda_i$) modes. Residues active in the fastest modes are, therefore, referred to as kinetically hot residues. Their strong resistance to conformational changes implies their important role in maintaining the structure, or in underlying the stability of the folded state. Residues active in the slowest modes, on the other hand, are susceptible to large scale (global) motions. It is reasonable to conceive that such motions are associated with the collective dynamics of the overall tertiary structure, and thereby relevant to biological function, although in some proteins, such as the electron transport proteins, fast modes may also be operative in the biological function.

We illustrate the role of modal decomposition of entropy in free energy [Eq. (5)] using HIV-1 protease as an example. This is a dimeric enzyme consisting of two identical monomers of $N = 99$ residues. Figure 1 displays the mean-square fluctuations for the individual residues predicted by the GNM (thick curve), and those

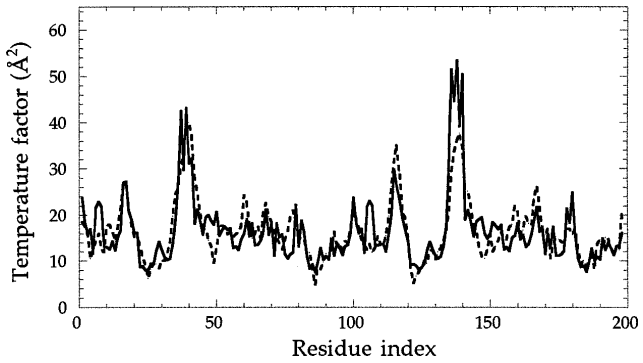


FIG. 1. Mean-square fluctuations of $C^{\alpha}$ atoms for HIV-1 protease. The solid curve is obtained using the GNM, and the dashed curve represents the crystallographic temperature factors [14].

observed in experiments (thin curve). Experimental results refer to the x-ray crystallographic temperature factors $B_i = 8\pi^2 \langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle/3$ of the high resolution structure of Wlodawer and collaborators [14]. The theoretical curve is normalized by suitable choice of the parameter $\gamma$. The good agreement between the conventional x-ray crystallographic $B$ factors and the predictions of a simple model based on Gaussian fluctuations of residues may be explained by the fact that the observed thermal fluctuations (or Debye-Waller factors) are dominated by unimodal quasiharmonic motions, rather than anharmonic, multimodal jumps between conformational isomers, as elegantly demonstrated by Garcia, Krumhansl, and Frauenfelder [15].

Figure 2 displays the fluctuations of the HIV-1 protease residues induced by subsets of slow (a) and fast (b) modes of motion. The fluctuations $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle_k$ associated with the $k$th mode of motion are found from [2,3]

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle_k = (3k_B T/\gamma) [\lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^{\mathrm{T}}]_{ii} = (3k_B T/\gamma) \lambda_k^{-1} [\mathbf{u}_k]_i [\mathbf{u}_k]_i , \quad (6)$$

where the subscripts designate the elements of the matrices (or vectors) enclosed in brackets. Part (a) of Fig. 2 displays the mean-square fluctuations driven by the slowest mode, $k = 2$; part (b) refers to the result of the subset of the fastest six modes $N - 5 \le k \le N$. The fluctuations from individual modes are combined therein using $\lambda_k^{-1}$ as a weighting factor. Results are displayed in normalized form (divided by $3k_B T/\gamma$), such that the areas enclosed by the curves are equal to unity. The results for only one monomer of the protein are displayed, those for the second monomer being approximately the same.

A relatively smooth curve is observed in Fig. 2(a), consistent with the fact that slow modes are more cooperative, and manifested by long-wavelength motions embodying a large number of residues. Minima in such curves correspond to hinge regions, whereas maxima describe regions of highest mobility in the global motion of the molecule.

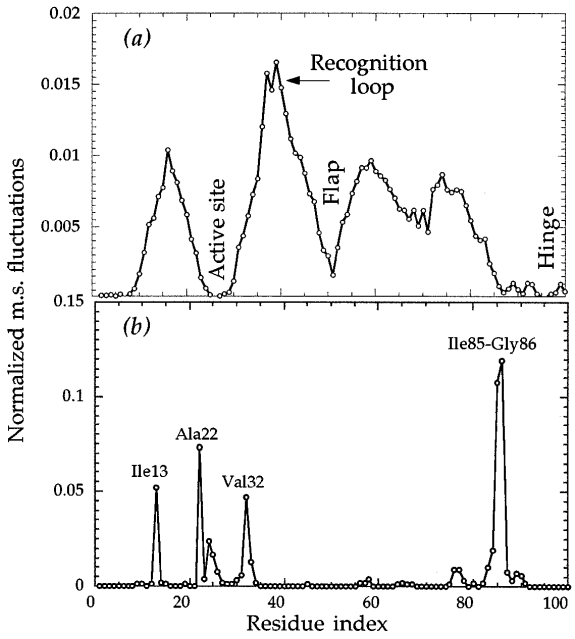


FIG. 2. Calculated normalized mean-square fluctuations of $C^{\alpha}$ atoms for HIV-1 protease monomers, in (a) the slowest mode, and (b) the fastest six modes of motion, of the dimer. Results for the two monomers are almost identical. In part (a), cooperative residue fluctuations are distinguished in consistency with biological function: The active site, the recognition loop, the flap, and the hinge regions of HIV-1 protease are known to lie between residues 25–27, 36–42, 46–56, and 96–99, respectively. The peaks in part (b), on the other hand, refer to residues that participate in the folding core and are highly conserved among the different retroviral sequences.

On the contrary, Fig. 2(b) is characterized by sharp peaks, which indicate the centers of localization of energy.

The peaks observed in Fig. 2(b) describe the kinetically hot residues, playing a key role in the stability of the protein. We note that these peaks lie within the folding cores identified using a detailed statistical analysis with residue-specific potentials [16]. Therein, the sequences of residues 22–32 (comprising the active site residues 25–27), 74–78, and 84–91 were pointed out to form the folding cores, or foldons. These are structural units possessing nativelike structure in the unfolded state, and participating in the initial folding event. Among them, the sequences 22–32 and 84–91, which coincide with our peaks, were further shown to be highly conserved [16] among the different retroviral sequences.

In addition to residue conservation due to stability requirements, which are associated with the peaks in Fig. 2(b), other conserved residues are observed in HIV-1 protease, relevant to functionally important aspects of the protein. The flap region between residues 46–56 is an example. The minima in Fig. 2(a) reflect the regions constrained severely during the global motion of the protein relevant to function. In this respect, it is interesting to observe that the region 46–56, despite being located on the relatively flexible flap, is distinguished by its reduced mobility, in consistency with its low tolerance to mutations. The region between residues 36–42, on the other hand, located at the solvent-exposed parts of the flap, exhibits the highest mobility. These highly mobile arms are, indeed, required for surrounding and anchoring the peptide substrate in the active site cleft between the two monomers. The loop containing residues 36–42 directly controls substrate binding and product release [17,18].

An interesting result of the present Letter is that the kinetically hot residues, which are tightly packed, are apparently conserved during evolution because that part of the structure would be very sensitive to mutations which could disrupt the packing. This has implications for folding as it may be difficult for a protein to achieve such tight packing kinetically. The GNM can efficiently recognize such residues. In support of the correlation between tightly packed residues and conservation during evolution, it has been pointed out that the regions of restricted mobility may generally be associated with disruptive mutation sites [19,20]. A typical example is T4 lysozyme: Of the 2015 single amino acid substitutions carried out by Poteete and co-workers [21], only a small fraction (10%–15%) were found to be deleterious, and the residues

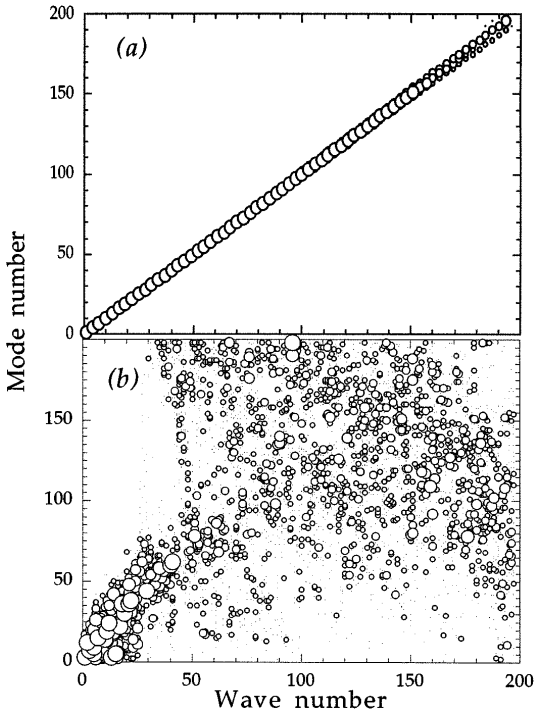


FIG. 3. Wave-number–mode-number maps for (a) a Rouse-like randomly coiled polymer of the same size as HIV-1 protease, and (b) HIV-1 protease. In part (a), most of the diagonal elements exhibit $W(q,k)$ values above 0.9. In the lower map, six different size symbols are used to designate $W(q,k)$ values lying in different ranges, the largest open circles refer to $W(q,k) \geq 0.35$, and the points to the range $0.10 \leq W(q,k) \leq 0.15$. Intermediate size symbols are assigned to each successive interval of 0.05.

with the highest sensitivity to mutation were all at local minima of the $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle$ curve. Our recent simulations of T4 lysozyme [22] also verify that most of the residues whose substitution has been pointed out to be deleterious in mutation studies exhibit highly restrictive thermal fluctuations.

For an assessment of the size of chain segments involved in modes of different frequencies, we performed the following Fourier analysis. The cosine transform of the eigenvectors $[\mathbf{u}_k]_j$ yields a wave-number–mode-number matrix in the form of $W(q,k) = \sum_{j=1}^{N} \cos[q(\pi j/N)][\mathbf{u}_k]_j$. Here, $q$ represents the wave number which is inversely proportional to the number of sequential residues, and $k$ is the mode number increasing with frequency. The $W$ map for a Rouse-like polymer of the same size as HIV-1 protease, and that obtained for HIV-1 protease are shown in Figs. 3(a) and 3(b), respectively. Figure 3(a) contains nonzero entries lying on or near the diagonal only. Figure 3(b), on the other hand, contains significant off-diagonal contributions.

The low-frequency band (lower part) of the map is responsible for large amplitude collective motions related to function, as illustrated in Fig. 2(a). These modes obey a frequency density of the form $g(\lambda) \sim \lambda^{\gamma}$, where the exponent $\gamma$ is anomalously low ($\gamma \leq 1$) [1,3,23]. The high-frequency band (upper part), on the other hand, refers to small amplitude motions of individual residues [see Fig. 2(b)]. These are related to stability. The nonzero portion of the map associated with the low-frequency and short-wavelength motions (lower right portion of the figure) is due to collective motions of residues which are relatively close along the backbone. The upper left region of the figure, on the other hand, is due to high-frequency communication between residues which are far apart from each other along the backbone. These off-diagonal regions of the map apparently form a bridge between residues active in high- and low-frequency modes, which are presently proposed to be related to stability and function.

Partial support from Bogazici University Research Funds 97A0401 is gratefully acknowledged.

[1] D. Ben-Avraham, Phys. Rev. B **47**, 14 559 (1993).
[2] I. Bahar, A. R. Atilgan, and B. Erman, Fold. Des. **2**, 173 (1997).
[3] T. Haliloglu, I. Bahar, and B. Erman, Phys. Rev. Lett. **79**, 3090 (1997).
[4] P. J. Flory, Proc. R. Soc. London A **351**, 351 (1976).
[5] M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
[6] F. Harary, *Graph Theory* (Addison-Wesley, Reading, MA, 1971).
[7] S. Miyazawa and R. L. Jernigan, Macromolecules **18**, 534 (1985).
[8] I. Bahar and R. L. Jernigan, J. Mol. Biol. **266**, 195 (1997).
[9] M. Karplus and J. Kushick, Macromolecules **14**, 325 (1981).
[10] R. M. Levy, M. Karplus, J. Kushick, and D. Perahia, Macromolecules **17**, 1370 (1984).
[11] I. Bahar, A. Wallquist, D. G. Covell, and R. L. Jernigan, Biochemistry **37**, 1067 (1998).
[12] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).
[13] P. W. Anderson, Rev. Mod. Phys. **50**, 191 (1978).
[14] N. Thanki *et al.*, Protein Sci. **1**, 1061 (1992).
[15] A. E. Garcia, J. A. Krumhansl, and H. Frauenfelder, Proteins **29**, 153 (1997).
[16] A. Wallqvist, G. W. Smythers, and D. G. Covell, Protein Sci. **6**, 1627 (1997).
[17] D. D. Loeb, R. Swanstrom, L. Everitt, M. Manchester, S. E. Stamper, and C. A. Hutchison III, Nature (London) **340**, 397 (1989).
[18] L. K. Nicholson *et al.*, Nat. Struct. Biol. **2**, 274 (1995).
[19] T. Alber, S. Dao-Pin, J. A. Nye, D. C. Muchmore, and B. W. Matthews, Biochemistry **26**, 3754 (1987).
[20] B. W. Matthews, Adv. Protein Chem. **46**, 249 (1995).
[21] D. Rennell, S. E. Bouvier, L. W. Hard, and A. R. Poteete, J. Mol. Biol. **222**, 67 (1991).
[22] I. Bahar, B. Erman, T. Haliloglu, and R. L. Jernigan, Biochemistry **36**, 13 512 (1997).
[23] K. Elber and M. Karplus, Phys. Rev. Lett. **56**, 394 (1986).