Protein fluctuations and breakdown of time-scale separation in rate theories

Jianhua Xing[1,2,*] and K. S. Kim[1]
[1]*Chemistry and Material Science Directorate, University of California & Lawrence Livermore National Laboratory, Livermore, California 94550, USA*
[2]*Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0406, USA*
(Received 10 May 2006; revised manuscript received 3 October 2006; published 27 December 2006)

A long-time fluctuation correlation function with a power-law form has been observed in recent single-molecule experiments by the Xie group. By analyzing the dynamics of an elastic network model (ENM) under white noise, we show that the observed long-time memory kernel can be explained by the discrepancy between the experimentally measured coordinate (or the coordinate directly coupled to protein function) and the minimum energy path of the system. Consequently, the dynamics of the measured collective coordinate has contributions from degrees of freedoms with a broad distribution of time scales. Our study also implies that the widely used ENM Hamiltonian should be viewed as a coarse-grained model of a protein over a rugged energy landscape. Large effective drag coefficients are needed to describe protein dynamics with the ENM's.

DOI: 10.1103/PhysRevE.74.061911 PACS number(s): 87.15.He, 87.15.Ya, 87.16.Ac, 82.37.−j

I. INTRODUCTION

Rate processes are ubiquitous in physics, chemistry, and biology. The development of reaction rate theories is a classical topic of theoretical physics and chemistry [1]. A basic physical picture is given in the seminal paper by Kramers [2] (see Ref. [1] for earlier and subsequent work). For a dynamical many-body system, a basic premise behind the construction of many reaction rate theories is the existence of one special degree of freedom (DOF) called the reaction coordinate (RC), whereby the trajectory of the species along this coordinate results in the chemical reaction. Mathematically it is normally the minimum-energy path (MEP) (or the intrinsic reaction coordinate) connecting the reactant and the product along the mass-weighted multidimensional potential energy surface (see Fig. 1) [3]. One usually assumes a separation of time scales between the dynamics along the MEP and along the remaining DOF orthogonal to the MEP. Consequently, the system dynamics can be well described by a Langevin-type or generalized Langevin-type dynamics with-short time memory kernels [4],

$$m\frac{d^2x}{dt^2} = -\frac{dU}{dx} - \zeta_g \int_0^t d\tau K(t-\tau)\frac{dx(\tau)}{d\tau} + f(t), \qquad (1)$$

where m is the reduced mass, x represents the coordinates of the MEP, U is the potential along MEP, ζ_g is the drag coefficient, K is the memory kernel which usually take the form of a Dirac δ function (for Langevin dynamics) or a fast decaying function of $(t-\tau)$ (e.g., an exponentially decaying function in the Grote-Hynes treatment [5]), and f is the fluctuation force. The generalized Langevin equation (1) can be derived formally using the Mori-Zwanzig projection operator formalism with temporal-spatial coarse graining [4,6,7]. The basic idea is that a system can be divided into two subspaces with the dynamics treated explicitly and implicitly, respectively. The projection formalism involves the construction of an operator which projects the full dynamics of system onto a subspace spanned by the explicit DOF. The projection formalism does not eliminate the implicit DOF, but preserves their influence or back reaction on the reduced dynamics through the appearance of a memory term and a stochastic forcing term. Under the condition that the dynamics of the explicit and implicit DOF be slow and fast modes, respectively, and that there exist a clear time-scale separation, the memory term decays quickly.

Proteins are flexible entities. Numerous experimental and theoretical studies have found that the structural fluctuations of proteins are strongly correlated with their function [8–12]. A widely used technique to study protein fluctuations is to calculate the potential of mean force (PMF) along the RC. On obtaining the PMF, one assumes that all the DOF or-

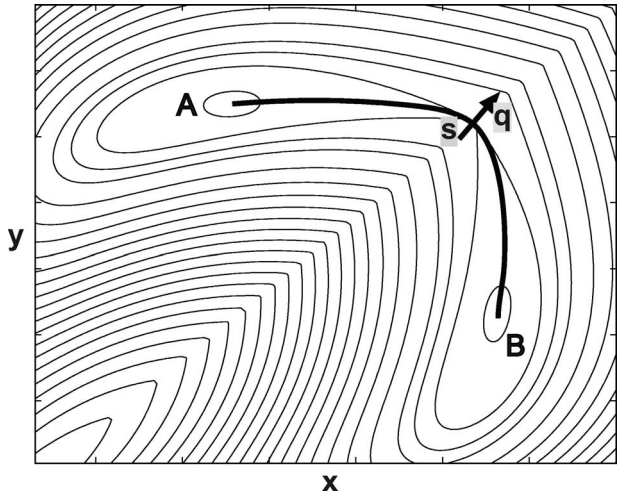

FIG. 1. Schematic illustration of a potential energy surface and the minimum-energy path (s). The shown 2D system can be described by either a coordinate system with the MEP and the orthogonal coordinate q or an x-y coordinate system. A projection along the MEP may result in a 1D generalized Langevin equation with short-time memory kernels. However, if the projection coordinate (e.g., x, set by either experiments or functional relevance) deviates from the MEP, a long-tail memory kernel may be expected.

*Corresponding author. Electronic address: xing3@llnl.gov

thogonal to the RC adjust to motion along the RC adiabatically. In other words, there is clear time-scale separation between dynamics along the RC and along the remaining DOF. Recent single-molecule measurements by the Xie group have shown a power-law memory kernel for the fluctuations within proteins of various systems described by a generalized Langevin equation [13,14]. The existence of a long-time memory effect implies the breakdown of the time-scale separation assumption, thus initiating several theoretical investigations to clarify the mechanism behind the power-law decay. Granek and Klafter explain the observation by using a fracton dynamics model [15]. Debnath and co-workers [16] and Tang and Marcus [17] show that distance fluctuations of one-dimensional polymers give the observed power-law decay of the memory kernel. Here, we propose a dynamic version of the widely used elastic network model which is simple and yet captures the essential physics. The experimentally measured coordinate (or the coordinate directly coupled to the protein function) may not coincide with the minimum-energy path of the system; e.g., consider the projection is along x in Fig. 1. Then the above-mentioned time-scale separation condition is not satisfied and a long-time memory term will be expected.

II. THEORY AND NUMERICAL RESULTS

In the present work, we adopted the widely used elastic network model (ENM) to analyze the experiment by Min *et al.* [14]. The ENM is a coarse-grained model and gives a reasonable description of protein fluctuations [18–20]. An ENM represents a protein by a network of elastically coupled N nodes (usually the C_α atom positions), $\mathbf{q}=(\mathbf{q}_1,\ldots,\mathbf{q}_N)$, with the following simple interaction form:

$$V=\frac{1}{2}c\sum_{i\neq j}h(r_{cut}-r_{ij}^e)(r_{ij}-r_{ij}^e)^2, \tag{2}$$

where the superscript e represents the equilibrium structure, c is a universal spring constant, h is a Heaviside function, r_{cut} is a cutoff distance, and $r_{ij}=|\mathbf{q}_i-\mathbf{q}_j|$ and $r_{ij}^e=|\mathbf{q}_i^e-\mathbf{q}_j^e|$.

The above potential form can be approximated by expanding to quadratic terms. By diagonalizing the Hessian, one can transform to normal-mode coordinates with $\Delta\mathbf{q}\equiv\mathbf{q}-\mathbf{q}^e=\mathbf{\Gamma Q}$, where $\mathbf{\Gamma}$ is composed of the eigenvectors of the Hessian. To study the dynamics, we include the effects of solvent and coarse-grained degrees of freedom in the ENM as dissipation terms, so the equations of motion are given by a set of overdamped Langevin equations (see also [21–23])

$$-\lambda_{i\alpha}Q_{i\alpha}-\zeta\frac{dQ_{i\alpha}}{dt}+f(t)=0, \tag{3}$$

where $Q_{i\alpha}$ and $\lambda_{i\alpha}$ are the $[(i-1)\times 3+\alpha]$th normal-mode coordinate and the corresponding eigenvalue. Here, for convenience of discussion, we use two indices to label the normal mode, with $i=1,\ldots,N$ and $\alpha=1$, 2, 3. From Eq. (3), the coordinate autocorrelation function is given by

$$C_Q^{i\alpha}(t)\equiv\langle Q_{i\alpha}(t)Q_{i\alpha}(0)\rangle=\frac{k_BT}{\lambda_{i\alpha}}\exp\left(-\frac{\lambda_{i\alpha}}{\zeta}t\right), \tag{4}$$

with k_B the Boltzmann constant and T the temperature. The Laplace transform is

$$\tilde{C}_Q^{i\alpha}(s)=\frac{k_BT}{\lambda_{i\alpha}}\frac{1}{s+\lambda_{i\alpha}/\zeta}. \tag{5}$$

To compare with the experimental results of Min *et al.* [14], one needs to calculate the autocorrelation function of the distance between two ENM nodes, which is approximately given by

$$\begin{aligned}C_r(t)&\equiv\langle(r_{ij}(t)-r_{ij}^e)(r_{ij}(0)-r_{ij}^e)\rangle\\&\approx\left(\frac{1}{r_{ij}^e}\right)^2\langle(\mathbf{q}_i^e-\mathbf{q}_j^e)\cdot[\Delta\mathbf{q}_i(t)-\Delta\mathbf{q}_j(t)]\\&\quad\times(\mathbf{q}_i^e-\mathbf{q}_j^e)\cdot[\Delta\mathbf{q}_i(0)-\Delta\mathbf{q}_j(0)]\rangle\\&=\sum_{k\beta}\left\{\left(\frac{1}{r_{ij}^e}\right)^2\sum_\alpha(q_{i\alpha}^e-q_{j\alpha}^e)^2[(\bar{\Gamma}_{i\alpha,k\beta})^2+(\bar{\Gamma}_{j\alpha,k\beta})^2]\right.\\&\quad\left.\times\langle Q_{k\beta}(0)Q_{k\beta}(0)\rangle\right\}\frac{\langle Q_{k\beta}(0)Q_{k\beta}(t)\rangle}{\langle Q_{k\beta}(0)Q_{k\beta}(0)\rangle}\\&=\sum_{k\beta}p_{k\beta}\frac{C_Q^{k\beta}(t)}{C_Q^{k\beta}(0)}.\end{aligned} \tag{6}$$

In the above expression, the matrix $\bar{\Gamma}$ is the reduced Γ after the elimination of the zero-frequency modes corresponding to the three translational degrees of freedom of the center of mass and three rotations. The normal-mode correlation functions $C_Q^{k\beta}(t)$ are given by Eq. (4). Equation (6) reveals that the distance correlation function, which is given by a linear combination of the single-exponential decaying normal-mode correlation functions, may show multiexponential decay. If the generalized Langevin equation (1) with a harmonic potential is used to model the residue-residue distance fluctuation, the memory kernel in Laplace space is given by [14]

$$\tilde{K}(s)=\frac{m\omega^2}{\zeta_g}\frac{\tilde{C}_r(s)}{C_r(0)-s\tilde{C}_r(s)}, \tag{7}$$

where $\tilde{C}_r(s)$ is the Laplace transform of the distance correlation function $C_r(t)$,

$$\tilde{C}_r(s)\equiv\sum_{k\beta}p_{k\beta}\frac{\tilde{C}_Q^{k\beta}(s)}{C_Q^{k\beta}(0)}. \tag{8}$$

Equations (6)–(8) are the central results of this work. They show that a long-time memory kernel for the distance fluctuation can exist even if all the normal modes are described by the Langevin dynamics with δ memory kernels. The reason is that the distance fluctuation coordinate is not along any of the normal-mode coordinates and has contributions from a large number of normal modes [see Fig. 3(b) below].

We apply the above analysis here to the protein complex formed between fluorescein (FL) and monoclonal antifluo-

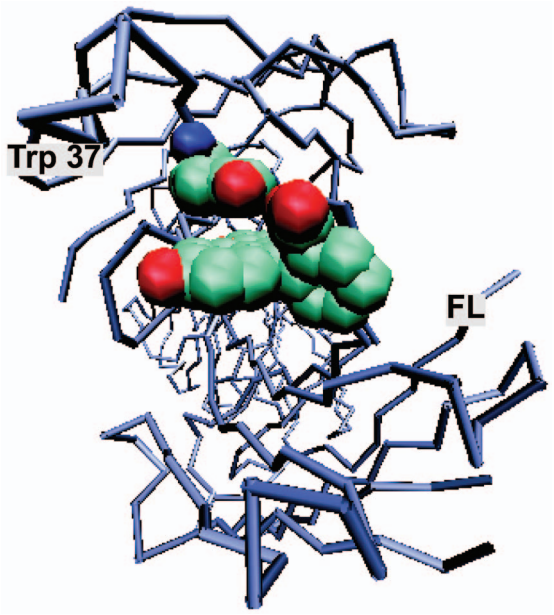

FIG. 2. (Color) Structure of the FL–anti-FL protein complex. In ENM, the protein structure 1FLR of anti-FL is modeled by 437 nodes representing the Ca atoms with an extra node representing the center of the aromatic ring of Trp37 and the fluorecein is modeled by nodes locating at the centers of the three aromatic rings. The distance fluctuation between the center of Trp37 ring and the FL ring formed by C3-C8 is calculated to compare with the experiment.

rescein 4-4-20 (anti-FL) (PDB label 1FLR; see Fig. 2). The spring constant c in Eq. (2) is determined by fitting to $C_r(0)$ and the drag coefficient ζ by fitting the experimental data of the distance correlation function [14]. In the calculations, the temperature is assumed as $T=298$ K and a cutoff distance $r_{cut}=10$ Å used in the literature is adopted [19], $c=1.4$ kcal/(mol/Å^2), comparable to the value 1.0 ± 0.5 kcal/(mol Å^2) used in the ENM literature [19], $\zeta=1k_BT$ s/Å^2. Figure 3(a) compares the calculated distance correlation function using Eq. (6) and the experimental data of Min *et al.* Given the simple form of the ENM, the agreement is remarkable. At large t, there is larger discrepancy between the theoretical result and the experimental data in Fig. 3(a). This discrepancy implies that contributions from some low-frequency modes are underestimated with the current model.[1] Models with different cutoff distances also give reasonable fittings. Figure 3(b) shows the normalized contribution of each normal mode given by Eq. (6). While several low-frequency modes make significant contributions, those from other modes cannot be neglected due to the large number of degrees of freedom involved. Figure 3(c) shows the Laplace transform of the memory kernel calculated by Eq. (7), which has approximately a power-law form. Figure 3(c) also shows that distance correlations between the FL and several other residues have similar behaviors, which was also

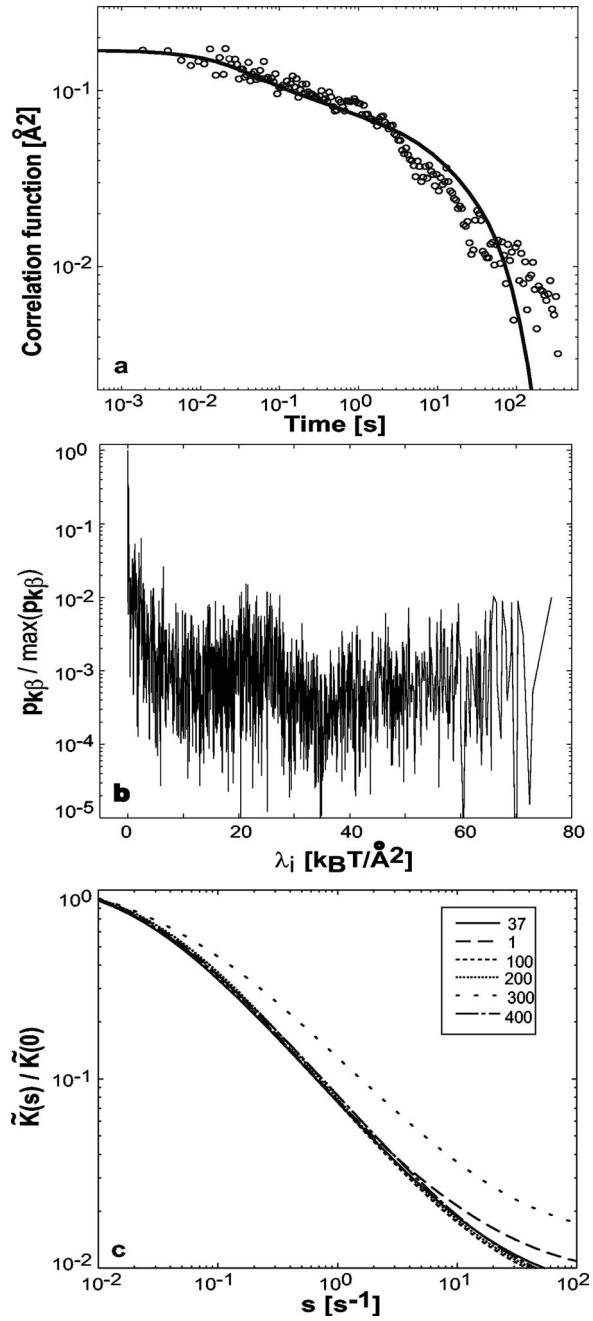

FIG. 3. The distance correlation function between Tryp 37 and the FL. (a) Comparison of the calculated distance correlation function with Eq. (6) (solid line) and the experimental data by Min *et al.* [14] (circles). (b) The normalized contribution of each normal mode given by Eq. (6). (c) Laplace transform of the memory kernel calculated by Eq. (7). Here the calculated memory kernels from the distance correlation functions between the FL and residue 37 as well as residues 1, 100, 200, 300, and 400 are shown.

observed in the one-dimensional polymer study of Tang and Marcus [17]. This implies that the lack of time-scale separation is a general phenomenon in protein dynamics.

III. DISCUSSION

The phenomenon of dynamic disorder, or rate constant fluctuations, has been widely studied [24–27]. The experi-

[1]To focus on the essential physics, we adopted the simplest version of the ENM. Improvement of the fitting is expected with some refined but more complex models discussed in the literature.

mental observations of the Xie group further demonstrate dynamic disorder at single-molecule levels. The experiments reveal that caution should be taken on applying rate theories to complex biological systems. The legitimacy of the underlying assumptions of a rate theory should be reexamined. The reaction coordinate framework discussed in the Introduction can break down in two respects. First, a reaction coordinate with slow dynamics may not exist if the system involves broad and continuous time scales. Second, even if a slow reaction coordinate (the MEP) can be defined, the coordinate relevant to experimental measurements or to protein functions may not coincide with the MEP, and this discrepancy may result in the breakdown of the time-scale separation between the tagged coordinate and the remaining DOF. Consequently, to describe the system dynamics, a long-time memory term may be necessary or extra DOF need to be treated explicitly. These observations may have significant relevance in any study involving the dynamics of biological macromolecules. Here we want to mention a similar situation in protein motor studies. Protein motors use chemical and electrochemical energies to perform mechanical work and are essential for many biological processes [28–32]. A protein motor is described by chemical reaction coordinates and mechanical motion coordinates which are coupled together, analogous to the electron-transfer coordinate and the protein fluctuation coordinate in the FL–anti-FL system. There is usually no time-scale separation between the different degrees of freedom. Min *et al.* related the observed long-time memory kernel to the observation that there may be no well-defined single-valued rate constant for an enzymatic reaction [33,34]. Studies show that theoretical treatments beyond discrete rate equations are necessary to understand some mechanochemical properties of a protein motor [35–37]. For example, a distribution of rate constants is essential to explain the long-standing puzzle of the motor torque-speed relationship of the bacterial flagellar motor [38,39] [comparing Eq. (6) of Ref. [39] and Eq. (5) of Ref. [33]]. Further studies are needed to examine the implications of the experimental observations of Min *et al.* to the understanding of other biological systems.

To fit the experimental data, we used a drag coefficient $\zeta = 1 k_B T$ s/Å^2. This value is orders of magnitude higher than the typical drag coefficient of a polymer [17]. This discrepancy may call into question the validity of the distance fluctuation models as discussed in the present work and in the work of Tang and co-workers [17,40]. However, the discrepancy can be reconciled by the fact that the elastic network model is a coarse-grained model. Many experimental and theoretical studies (especially in the protein folding community) show that a system complex such as a protein possesses a rugged energy [41,42]. The ruggedness of the energy landscape may lead to the conclusion that a normal-mode analysis cannot work since it only characterizes the potential near one local minimum. On the other hand, the smooth potential used by an elastic network model should be understood as an effective potential after averaging out the local rugged fluctuations [43,44]. That explains why the elastic network models are surprisingly successful on describing large-scale protein fluctuations (notice that the short-range fluctuations of the potentials have a marginal effect on the equilibrium properties such as the B factor at experimental resolutions) [19]. Consequently, the effective drag coefficients used in a dynamic version of the elastic network model are different from the bare drag coefficients and the difference can be large. To further illustrate this, we refer to the work of Zwanzig [45]. Zwanzig proposed a model describing diffusion in a rugged potential and derived an expression for the effective diffusion coefficient. His results show that the ruggedness of the potential can dramatically reduce the diffusion coefficient at low temperatures. Zwanzig looked at diffusion on a length scale much larger than the ruggedness and, in effect, replaced the original rugged potential with an effective smooth potential, integrating out the rapid small fluctuations. Carrying this analogy to our case, the effective drag coefficients used for an elastic network model should be normal-mode frequency dependent (see Fig. 4). For high-frequency modes with a length scale comparable to the characteristic length scale of the rugged potential, the values should approach the bare drag coefficients. For low-frequency modes with a length scale much larger than the characteristic length scale of the rugged potential, the drag coefficients reach renormalized values according to the analysis of Zwanzig. A set of normal-mode frequency-dependent drag coefficients may help fit the experimental data. Our current treatment with a single value of the drag coefficient is oversimplified and requires further study. Another prediction of the present model is that the effective drag coefficients and the distance autocorrelation function are expected to be highly temperature dependent. The experiment setup by the Xie group may serve as a tool to detect ruggedness of the protein energy landscape (see also Ref. [46]).

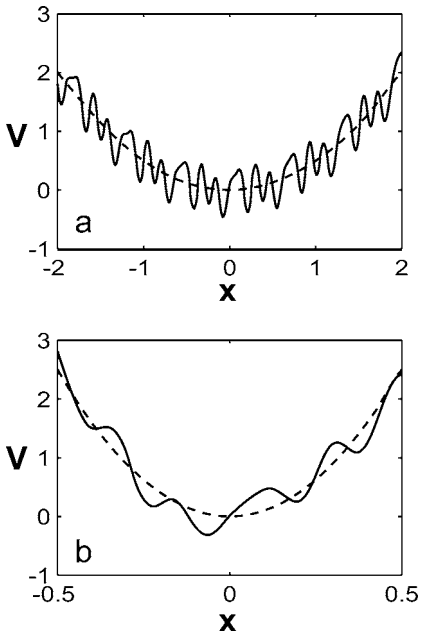

FIG. 4. Schematic illustration of rugged energy landscapes (solid lines). Also shown are renormalized smooth potentials with high (a) and low (b) frequencies (dashed lines). The units of distance and energy are arbitrary and only for illustration purposes. The existence of potential roughness may have different effects on the effective diffusion along the potential landscapes in the two cases.

ACKNOWLEDGMENTS

We would like to thank Wei Min (Harvard University) for providing us his data and pointing to Ref. [17], Professor Robert Jernigan and Dr. Andrzej Kloczkowski (Iowa State University) for pointing to Refs. [21–23], and Dr. Wenjun Zheng (NIH), Dr. Daniel Barsky (LLNL), and Dr. Michael Surh (LLNL) for many fruitful discussions. J.X. is supported by a Lawrence Livermore National Laboratory Directed Research and Development grant. K.K. would like to acknowledge the support by the DOE/The University of California Merced Center for a Computational Biology Grant No. DE-FG01–04ER04–15. This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory, under Contract No. W-7405-Eng-48.

[1] P. Hanggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).
[2] H. Kramers, Physica (Utrecht) **7**, 284 (1940).
[3] K. Fukui, Acc. Chem. Res. **14**, 363 (1981).
[4] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, Oxford, 2001).
[5] R. F. Grote and J. T. Hynes, J. Chem. Phys. **73**, 2715 (1980).
[6] H. Mori, Prog. Theor. Phys. **33**, 423 (1965).
[7] R. Zwanzig, in *Lectures in Theoretical Physics*, edited by W. E. Brittin, B. W. Downs, and J. Downs (Interscience, New York, 1961), p. 106.
[8] S. H. Northrup *et al.*, Proc. Natl. Acad. Sci. U.S.A. **79**, 4035 (1982).
[9] M. F. Perutz and F. S. Mathews, J. Mol. Biol. **21**, 199 (1966).
[10] H. Frauenfelder, G. A. Petsko, and D. Tsernoglou, Nature (London) **280**, 558 (1979).
[11] T. W. Allen, O. S. Andersen, and B. Roux, J. Gen. Physiol. **124**, 679 (2004).
[12] M. Karplus and G. A. Petsko, Nature (London) **347**, 631 (1990).
[13] W. Min *et al.*, Acc. Chem. Res. **38**, 923 (2005).
[14] W. Min *et al.*, Phys. Rev. Lett. **94**, 198302 (2005).
[15] R. Granek and J. Klafter, Phys. Rev. Lett. **95**, 098106(1) (2005).
[16] P. Debnath *et al.*, J. Chem. Phys. **123**, 204903 (2005).
[17] J. Tang and R. A. Marcus, Phys. Rev. E **73**, 022102 (2006).
[18] M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
[19] A. R. Atilgan *et al.*, Biophys. J. **80**, 505 (2001).
[20] W. J. Zheng and S. Doniach, Proc. Natl. Acad. Sci. U.S.A. **100**, 13253 (2003).
[21] A. Erkip and B. Erman, Polymer **45**, 641 (2004).
[22] A. Kloczkowski, J. E. Mark, and H. L. Frisch, Macromolecules **23**, 3481 (1990).
[23] W. W. Graessley, Macromolecules **13**, 372 (1980).
[24] R. Zwanzig, Acc. Chem. Res. **23**, 148 (1990).
[25] H. Frauenfelder, P. G. Wolynes, and R. H. Austin, Rev. Mod. Phys. **71**, S419 (1999).
[26] M. Karplus, J. Phys. Chem. B **104**, 11 (2000).
[27] X. S. Xie, J. Chem. Phys. **117**, 11024 (2002).
[28] G. Banting and S. Higgins, *Molecular Motors* (Portland Press, London, 2000).
[29] J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer, Sunderland, MA, 2001).
[30] R. D. Vale, J. Cell Biol. **150**, 13 (2000).
[31] T. Ogura and A. J. Wilkinson, Genes Cells **6**, 575 (2001).
[32] J. Ye *et al.*, Biochim. Biophys. Acta **1659**, 1 (2004).
[33] B. P. English *et al.*, Nat. Chem. Biol. **2**, 87 (2006).
[34] W. Min and X. S. Xie, Phys. Rev. E **73**, 010902 (2006).
[35] P. Reimann, Phys. Rep. **361**, 57 (2002).
[36] J. Xing, H.-Y. Wang, and G. Oster, Biophys. J. **89**, 1551 (2005).
[37] F. Julicher, A. Ajdari, and J. Prost, Rev. Mod. Phys. **69**, 1269 (1997).
[38] H. C. Berg, Annu. Rev. Biochem. **72**, 19 (2003).
[39] J. Xing *et al.*, Proc. Natl. Acad. Sci. U.S.A. **103**, 1260 (2006).
[40] J. Tang and S. H. Lin, Phys. Rev. E **73**, 061108 (2006).
[41] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
[42] S. S. Plotkin and J. N. Onuchic, Q. Rev. Biophys. **35**, 111 (2002).
[43] J. P. Ma, Curr. Protein Peptide Sci. **5**, 119 (2004).
[44] G. H. Li and Q. Cui, Biophys. J. **83**, 2457 (2002).
[45] R. Zwanzig, Proc. Natl. Acad. Sci. U.S.A. **85**, 2029 (1988).
[46] C. Hyeon and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **100**, 10249 (2003).