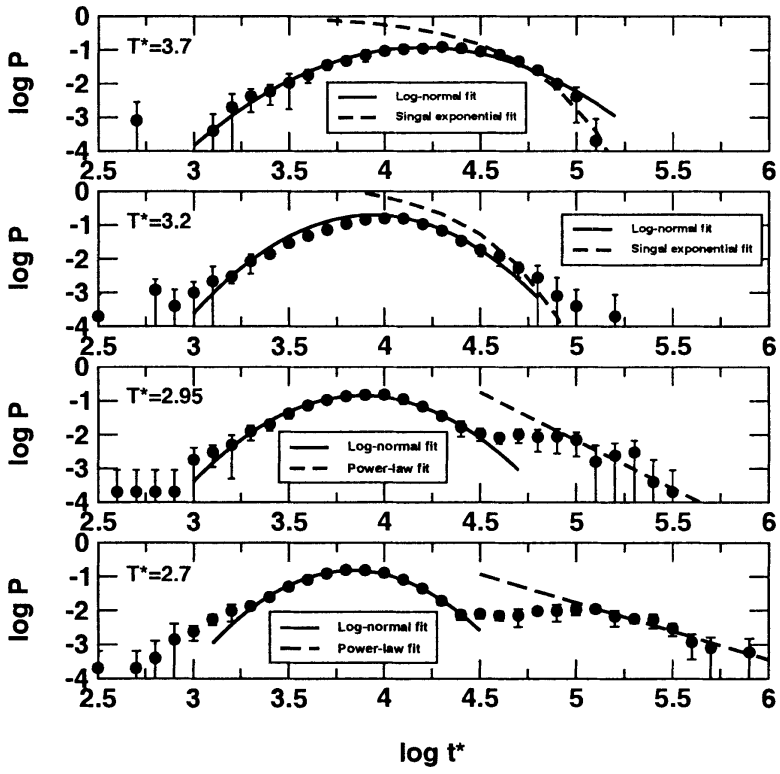
Article

Temperature Dependence of the Distribution of the First Passage Time: Results from Discontinuous Molecular Dynamics Simulations of an All-Atom Model of the Second β-Hairpin Fragment of Protein G

Yaoqi Zhou, Chi Zhang, George Stell, and Jin Wang

J. Am. Chem. Soc., **2003**, 125 (20), 6300-6305• DOI: 10.1021/ja029855x • Publication Date (Web): 26 April 2003



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 4 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article

J|A|C|S
ARTICLES
Published on Web 04/26/2003

Temperature Dependence of the Distribution of the First Passage Time: Results from Discontinuous Molecular Dynamics Simulations of an All-Atom Model of the Second β-Hairpin Fragment of Protein G

Yaoqi Zhou,*[,†,‡] Chi Zhang,[†] George Stell,[§] and Jin Wang*[,||,§,⊥]

Contribution from the Howard Hughes Medical Institute Center for Single Molecule Biophysics, Department of Physiology and Biophysics, State University of New York at Buffalo, 124 Sherman Hall, Buffalo, New York 14214, T. D. Lee Physics Laboratory and Research Center for Theoretical Physics, Fudan University, Shanghai, China, Department of Chemistry, State University of New York at Stony Brook, Stony Brook, New York 11794, State Key Laboratory of Electro-Analytical Chemistry, Changchun Institute of Applied Chemistry, Academic Sinica, Jilin 130022, China, and Global Strategic Analytics Unit, Citigroup, One Huntington Quadrangle, Suite 1N16, Melville, New York 11747

Received December 20, 2002; E-mail: yqzhou@buffalo.edu; jinwang@sprynet.com

Abstract: More than 22 000 folding kinetic simulations were performed to study the temperature dependence of the distribution of first passage time (FPT) for the folding of an all-atom Gō-like model of the second β-hairpin fragment of protein G. We find that the mean FPT (MFPT) for folding has a U (or V)-shaped dependence on the temperature with a minimum at a characteristic optimal folding temperature T_{opt}*. The optimal folding temperature T_{opt}* is located between the thermodynamic folding transition temperature and the solidification temperature based on the Lindemann criterion for the solid. Both the T_{opt}* and the MFPT decrease when the energy bias gap against nonnative contacts increases. The high-order moments are nearly constant when the temperature is higher than T_{opt}* and start to diverge when the temperature is lower than T_{opt}*. The distribution of FPT is close to a log-normal-like distribution at $T^* \geq T_{opt}^*$. At even lower temperatures, the distribution starts to develop long power-law-like tails, indicating the non-self-averaging intermittent behavior of the folding dynamics. It is demonstrated that the distribution of FPT can also be calculated reliably from the derivative of the fraction not folded (or fraction folded), a measurable quantity by routine ensemble-averaged experimental techniques at dilute protein concentrations.

1. Introduction

Theoretical[1−6] and experimental[7−10] studies of protein folding have led to the conclusion that the problem posed by the Levinthal paradox is solved by a bias in the free-energy surface toward the native conformation that extends far enough into the denatured region to reduce the conformational space search sufficiently for folding to occur in a reasonable time.[11] In this view, it is an ensemble of pathways rather than specific pathways leading toward the native state of proteins. Most theoretical studies of folding kinetics have focused on the ensemble-averaged quantities to facilitate the comparison with experimental data. However, with the rapid advances in optical technology, it is now possible to measure reaction dynamics of individual molecules in the laboratory.[12,13] This opens the door for the direct comparison of the distribution rather than the simple average of reaction events. This is important because the dynamics of the barrier crossing on a fluctuating energy landscape itself may lead to non-Poissonian statistics for individual molecules.[14,15]

This work is motivated by several recent single-molecule folding experiments[16−18] in which multiexponential folding kinetics were observed. In addition, a single-molecule experi-

[†] State University of New York at Buffalo.
[‡] Fudan University.
[§] State University of New York at Stony Brook.
[||] Changchun Institute of Applied Chemistry.
[⊥] Citigroup.

(1) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167−195.
(2) Karplus, M.; Šali, A. *Curr. Opin. Struct. Biol.* **1995**, *5*, 58−73.
(3) Shakhnovich, E. I. *Curr. Opin. Struct. Biol.* **1997**, *7*, 29−40.
(4) Thirumalai, D.; Klimov, D. K.; Woodson, S. A. *Theor. Chem. Acc.* **1997**, *96*, 14−22.
(5) Skolnick, J.; Kolinski, A.; Ortiz, A. R. *J. Biomol. Struct. Dyn.* **1998**, *16*, 381−396.
(6) Dill, K. A. *Protein Sci.* **1999**, *8*, 1166−1180.
(7) Baldwin, R. L. *Folding Des.* **1996**, *1*, R1−R8.
(8) Fersht, A. R. *Curr. Opin. Struct. Biol.* **1997**, *7*, 3−9.
(9) Van Nuland, N. A. J.; Forge, V.; Balbach, J.; Dobson, C. M. *Acc. Chem. Res.* **1998**, *31*, 31773−780.
(10) Capaldi, A. P.; Radford, S. E. *Curr. Opin. Struct. Biol.* **1998**, *8*, 86−92.
(11) Dobson, C. M.; Šali, A.; Karplus, M. *Angew. Chem., Int. Ed.* **1998**, *37*, 868−893.
(12) Moerner, W. E. *Acc. Chem. Res.* **1996**, *29*, 563.
(13) Lu, H. P.; Xun, L.; Xie, X. S. *Science* **1998**, *282*, 1.
(14) Wang, J.; Wolynes, P. G. *Phys. Rev. Lett.* **1995**.
(15) Kaya, H.; Chan, H. S. *J. Mol. Biol.* **2002**, *315*, 899−909.
(16) Zhuang, X.; Barley, L. E.; Babcock, H. P.; Russell, R.; Ha, T.; Herschlag, D.; Chu, S. *Science* **2000**, *288*, 2048.
(17) Jia, Y.; Talaga, D. S.; Lau, W. L.; Lu, H. S. M.; Degrado, W. F.; Hochstrasser, R. M. *Chem. Phys.* **1999**, *247*, 69.
(18) Schuler, B.; Lipman, E. A.; Eaton, W. *Nature* **2002**, *419*, 743−747.

10.1021/ja029855x CCC: $25.00

ment on enzymatic reaction dynamics[19,20] found that the reaction times of proteins follows a Lévy-like distribution (or power-law decay) at low temperatures. Similar behavior was found in a study of protein-folding dynamics using an analytical theory.[21,22] In this paper, we examine the distribution of the folding times of an all-atom Gō-like model of the second β-hairpin fragment of protein G using discontinuous molecular-dynamics simulation technique. It has been shown[23] that this semirealistic model yields a hydrophobic collapse mechanism that is consistent with high-temperature unfolding simulations and equilibrium free-energy surface analysis based on established all-atom empirical force fields in explicit or implicit solvent. The all-atom model further predicts that the collapse is initiated by two nucleation contacts (a hydrophilic contact between D46 and T49 and a hydrophobic contact between Y45 and F52); these four residues have one-to-one correspondence to the four residues that were found to be critical for the folding stability of the β-hairpin in recent nuclear magnetic resonance measurements.[24]

In this study, we calculate the mean and the distribution of the first passage time (FPT) as a function of temperature and as a function of the contact-energy bias gap. The mean FPT (MFPT) has a U (or V)-shaped dependence on the temperature and becomes smaller as the energy bias gap increases. Above a kinetic transition temperature T_{opt}*, the FPT is well behaved, and the distribution tends to be log-normal. On the other hand, as the temperature decreases, the distribution of FPT starts to become broader and has a power-law-like tail. In other words, the MFPT can no longer be a good parameter to describe the folding reaction at low temperatures. Within this temperature regime, the behavior of the distribution function is often needed to adequately capture the entire kinetic behavior. We found that T_{opt}* is larger than the solidification temperature T_s* defined by the Lindemann criterion but is smaller than T_f*, the thermodynamic folding transition temperature. Moreover, we also show that the distribution of FPT can be evaluated reliably from the derivative of fraction not folded with respect to time. Thus, it is possible to obtain the distribution of folding time without performing single-molecule experiments.

2. Model and Methods

The setup for the all-atom (except nonpolar hydrogen atoms) Gō-like model of the second β-hairpin fragment of protein G has been described previously.[23] Here, we give a brief summary for completeness.

The model contains 16 residues and 163 atoms. The bond lengths between two bonded atoms, bond angles, and improper dihedral angles were maintained by an infinitely deep square-well potential.[23] Interactions between nonbonded atoms are given by a square-well potential in which the hard-core and square-well diameters were derived from the van der Waals parameters from the CHARMM polar hydrogen parameter set 19.[25] The square-well depth is $-\epsilon$, if the atomic pair is in contact in the global minimum native structure derived from X-ray data,[26] and $B_o\epsilon$, otherwise. We defined a contact-energy bias gap g as $1 + B_o$ with $g = 1.0$ ($B_o = 0$) corresponding to the original Gō model.[27] In this paper, we studied the models with $g = 0.7$, 0.8, 0.9, and 1.0 (or $B_o = -0.3$, -0.2, -0.1, 0), respectively.

The folding thermodynamics and kinetics of the $g = 1.0$ model have been studied in detail using constant-temperature discontinuous molecular-dynamics (DMD) techniques.[23] The algorithm and implementation of DMD are described in earlier publications.[28,29] In the thermodynamic studies, initial structures are the native structure, and initial random velocities were generated from the Maxwell distribution. Thermodynamic studies were conducted at 13 different reduced temperatures ($T^* = k_BT/\epsilon$) from $T^* = 1.0$ to $T^* = 8.0$. At each temperature, five independent runs with different initial velocities were performed to estimate errors. The thermodynamic properties at temperatures that were not simulated were obtained by using the weighted histogram method.[30] The folding transition temperature T_f* was found at about 4.0.

The folding kinetics of the $g = 1.0$ model were studied in more details at nine different temperatures ($T^* = 2.70$, 2.80, 2.95, 3.00, 3.10, 3.20, 3.30, 3.50, and 3.70). At each temperature, a minimum of 100 folding simulations were carried out with the initial coillike configurations and velocities that were sampled from the equilibrium simulations at $T^* = 6.0$ in every 100 reduced time units (or about 500 000 collisions). Simulations were stopped when the β-hairpin was folded, and the simulation times were recorded as FPT. The peptide was considered to be folded if its all-heavy atom rmsd (root-mean-squared deviation) from the global minimum structure is less than 2.5 Å.[23] The rmsd was evaluated every 10 reduced time units ($t^* = t\sqrt{\epsilon/M\sigma_L^2}$, where M is the average mass of the atoms and σ_L is 1 Å). To obtain a reasonably accurate distribution of FPT, a total of 5000 folding simulations each was conducted at $T^* = 2.70$, 2.95, 3.20, and 3.70, respectively.

To investigate the dependence of MFPT on the contact-energy bias gap, g, the folding kinetics were also studied at $g = 0.7$, 0.8, and 0.9. At $g = 0.9$, the 100 folding simulations each were performed at $T^* = 3.2$, 3.3, 3.4, 3.5, and 3.6. At $g = 0.8$, the simulation temperatures were 3.3, 3.4, 3.5, 3.6, and 3.7. At $g = 0.7$, the temperatures were 3.4, 3.5, 3.6, 3.7, and 3.8. The initial coillike configurations and velocities for all folding kinetics studies were obtained from the equilibrium simulations of the respective gap models at $T^* = 6.0$.

The kinetic results are described in terms of MFPT, $\langle\tau\rangle$, as well as higher moments $\langle\tau^2\rangle$ and $\langle\tau^3\rangle$, where $\langle\ \rangle$ denotes the ensemble average. The MPFT is reported in reduced time units. The distribution of FPT, $P(t^*)$, is obtained using a bin width of 0.1 on a logarithmic time scale. (Other bin widths were also used, and the results were essentially the same.)

The distribution of FPT, $P(t^*)$, can also be calculated from the fraction not folded [$f_{nf}(t^*)$]. This is because

$$f_{nf}(t^*) = 1 - \int_0^{t^*} P(t)\,dt.$$

That is, $P(t^*) = -\,df_{nf}(t^*)/dt^*$. In other words, the distribution of FPT can also be obtained from the time dependence of the fraction not folded [or the fraction folded, $1 - f_{nf}(t^*)$], which is a quantity measurable using routine ensemble averaged experimental methods. For a log-normal distribution, we have

$$P^{\log}(t) = \frac{a}{\sqrt{\pi}t}e^{-(a\ln t + b)^2} \qquad (1)$$

The corresponding $f^{\log}(t)$ is given by

$$f^{\log}(t) = 1 - \int_0^t P^{\log}(t)\,dt = \frac{1}{2}[1 - \mathrm{erf}(a\ln t + b)] \qquad (2)$$

where erf(x) is the error function. Thus, if a distribution can be fitted by a log-normal distribution, the fraction not folded should be

(19) Yang, H.; Xie, X. S. *Chem. Phys.* **2002**, *284*, 423−437.
(20) Yang, H.; Xie, X. S. *J. Chem. Phys.* **2003**, *117*, 10965−10979.
(21) Lee, C. L.; Stell, G.; Wang, J. *J. Chem. Phys.* **2003**, *118*, 959−968.
(22) Lee, C. L.; Lin, C. T.; Stell, G.; Wang, J. *Phys. Rev. E* **2003**, in press.
(23) Zhou, Y.; Linhananta, A. *Proteins* **2002**, *47*, 154−162.
(24) Kobayashi, N.; Honda, S.; Yoshii, H.; Munekata, E. *Biochemistry* **2000**, *39*, 6564−6571.
(25) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902−1921.
(26) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. *Science* **1991**, *253*, 657−661.
(27) Ueda, Y.; Taketomi, H.; Gō, N. *Biopolymers* **1978**, *17*, 1531−1548.

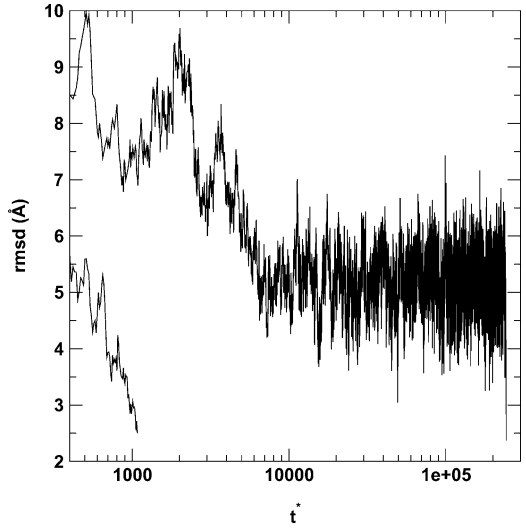


Figure 1. The results of the all heavy-atom rmsd from the native structure as a function of reduced time (t^*) for two typical trajectories at $T^* = 2.7$. The folding simulations are stopped when rmsd $\leq$ 2.5 Å.

automatically fitted by $f^{\log}(t)$. This method is used to test the accuracy of the FPT distribution generated from the histogram analysis. For a log-normal distribution, we have $\langle\tau^2\rangle/\langle\tau\rangle^2 = e^{a^{-2/2}}$ and $\langle\tau^3\rangle/\langle\tau\rangle^3 = e^{3a^{-2/2}}$. In other words, $\langle\tau^3\rangle/\langle\tau\rangle^3 = (\langle\tau^2\rangle/\langle\tau\rangle^2)^3$.

3. Results

Two typical folding trajectories at $T^* = 2.70$ for the $g = 1.0$ model are shown in Figure 1. One is a fast-folding trajectory with FPT = 1080 reduced time units, while the other is much slower with FPT = 243 410 reduced time units. For the slow-folding trajectory, it appears that the peptide first folded into a structure whose rmsd is about 5 Å from the native structure and, subsequently, was trapped there for a long time. The folding was complete only after many unfolding and refolding attempts. Several snapshots of the slow-folding trajectory are shown in Figure 2. The folding was initiated by the hydrophobic contacts between F52 and Y45 at $t^* = 1000$ (Figure 2b). The folding intermediate (Figure 2c) is characterized by misfolded side chains of Y45 and F52 that contact on the wrong side of the β-hairpin. This long-lived intermediate is a folding trap that was escaped by the reorientation of Y45 and F52 side chains (Figure 2d) to their native conformations (Figure 2e).

The results of MFPT $\langle\tau\rangle$ for the folding of the four different gap models are shown as a function of inverse temperature in Figure 3. Each data point was obtained from an average of 100 kinetic simulations. The error due to the limited sampling is likely large, particularly at low temperatures. Nevertheless, we detected a U (or V)-shaped curve that suggests the existence of an optimal folding temperature T_{opt}^* for each model. For $g = 1.0$, $T_{opt}^* = 3.2$. The existence of an optimal folding temperature reflects the fact that at high temperature, the native state is less stable, whereas at low temperature, there is an onset of the low energy nonnative trapped states. This optimal temperature is between the folding temperature ($T_f^* = 4.0$[23]) and the solidification temperature of $T_s^* = 2.0$, where the equilibrium state of the entire protein becomes solidlike based on the Lindemann

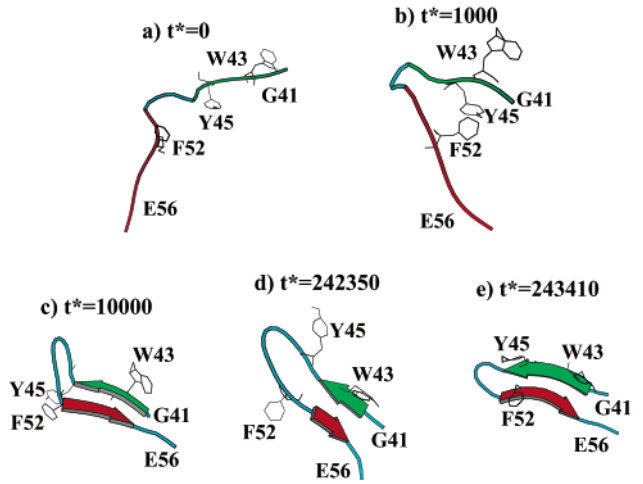


Figure 2. Snapshots of the slow-folding trajectory shown in Figure 1: (a) the initial coil state ($t^* = 0$); (b) the contact of hydrophobic residues (Y45 and F52, $t^* = 1000$); (c) the formation of the secondary structure and misfolded side chain hydrophobic contacts ($t^* = 10\,000$); (d) the reorientation of side chains ($t^* = 242\,350$); and (e) the native state. Drawn with MOLSCRIPT.[31]

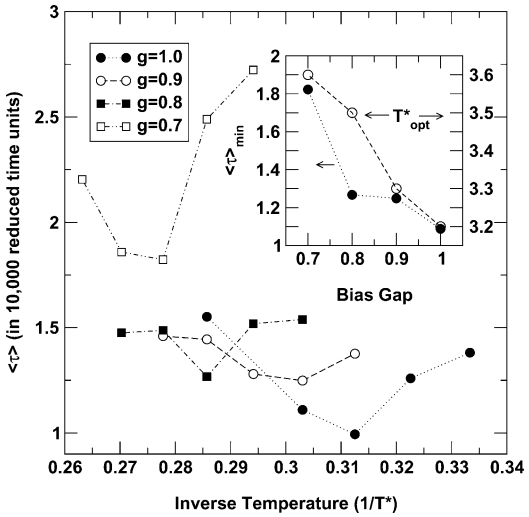


Figure 3. MFPT versus inverse reduced temperature $1/T^*$ for various gap models. In each model, the curve has a U (or V) shape. The temperature at which MFPT reaches its minimum ($\langle\tau\rangle_{min}$) is the optimal folding temperature T_{opt}^*. In the insert, $\langle\tau\rangle_{min}$ and T_{opt}^* are shown as a function of the gap parameter. Both monotonically decrease as the gap parameter increases. Each point represents an average value of 100 folding simulations.

criterion.[32,33] (The procedure for calculating the Lindemann index is as in ref 49.) The solidification temperature is likely higher than the actual glass-transition temperature discussed in the literature[34] because we found that a significant fraction of the folding simulations of the model β-hairpin can still fold to the native state at the temperature even lower than T_s^*. The solidification temperature T_s^* may be related to the transition temperature for partial freezing of the states, which are frozen to the basin of attractions, but within each basin there are still considerable degrees of freedom (called T_A in the density

(28) Zhou, Y.; Karplus, M.; Wichert, J. M.; Hall, C. K. *J. Chem. Phys.* **1997**, *107*, 10691−10708.
(29) Zhou, Y.; Karplus, M. *J. Mol. Biol.* **1999**, *293*, 917−951.
(30) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195−1197.
(31) Kraulis, P. *J. Appl. Crystallogr.* **1991**, *24*, 946−950.
(32) Zhou, Y.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 14429−14432.
(33) Zhou, Y.; Vitkup, D.; Karplus, M. *J. Mol. Biol.* **1999**, *285*, 1371−1377.
(34) Wolynes, P. G. In *Proceedings International Symposium on Frontiers of Science*; Frauenfelder, H., Chan, H. S., Debrunner, P. G., Eds.; American Inst. Physics: Boca Raton, New York, 1989; pp 38−65.

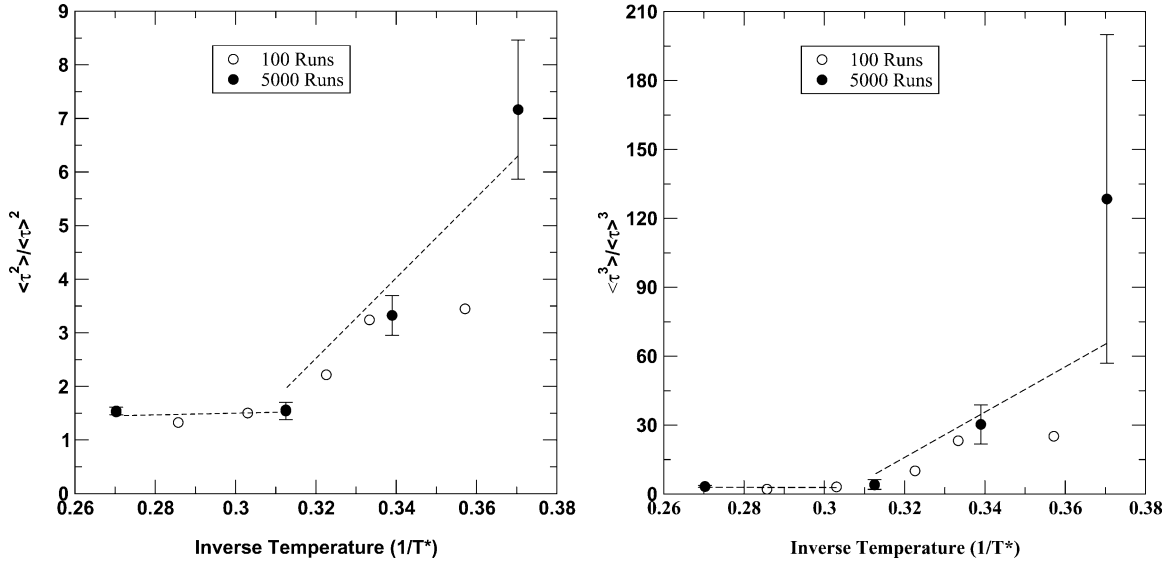


Figure 4. The reduced second ($\langle\tau^2\rangle/\langle\tau\rangle^2$, left) and third ($\langle\tau^3\rangle/\langle\tau\rangle^3$, right) moments as functions of the reduced inverse temperature $1/T^*$ for the $g = 1.0$ model. Both reduced moments are nearly constant at high temperature and start to diverge at low temperatures (high $1/T^*$). The error bars at three temperatures (●) are the 95% confidence limit estimated from five independent 1000 folding simulations (student *t*-tests). Dashed lines serve as a guide for the eyes.

functional and mode coupling studies[34,35]). The results of the optimal folding temperature T_{opt}* versus the bias gap are shown in the insert of Figure 3. As the gap increases, the optimal temperature decreases along with the minimum value of MFPT, $\langle\tau\rangle_{min}$. This validates the criterion that selects the subset (subspace) of the whole sequence space leading to well-designed fast-folding proteins by maximizing the difference between the average energy of native contacts and that of nonnative contacts.[36] In other words, one has to choose the sequence subspace such that the global bias overwhelms the roughness of the energy landscape.[37]

The reduced second ($\langle\tau^2\rangle/\langle\tau\rangle^2$) and third ($\langle\tau^3\rangle/\langle\tau\rangle^3$) moments of FPT are shown in Figure 4. At high temperature (low $1/T^*$), the reduced second and third moments change little with temperature. However, both the second and the third reduced moments start to diverge at $T^* < T_{opt}^* = 3.2$ (i.e., $1/T^* > 0.3125$). Thus, at low temperature, the MFPT, $\langle\tau\rangle$, is no longer a good representative of the system. In other words, the dynamics exhibit the onset of non-self-averaging behavior around and below T_{opt}*.

The distributions of FPT at four different temperatures are shown in Figure 5. The error bars were estimated from five independent 1000 folding simulations. At high temperatures ($T^* = 3.20$ and 3.70), the distribution can be fitted by a log-normal distribution very well. At low temperatures ($T^* = 2.95$ and 2.70), the population maximum is shifted slightly to a lower value of FPT. This indicates that some trajectories fold faster as a result of the gain of folding stability at a lower temperature. The short time behavior can still be fitted very well by a log-normal distribution. However, at a large FPT, there appears to be a transition to a linear decay in the log−log plot. This suggests a power-law-like tail. The distribution of FPT at $T^* = 3.70$ can be fitted very well by a single exponential for log $t^* > 4.5$, but is not so at lower temperatures.

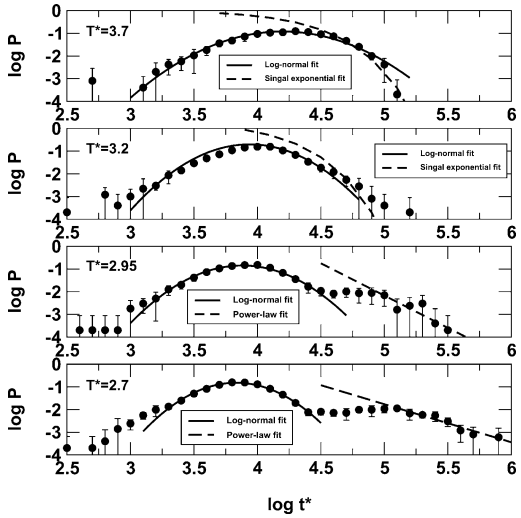


Figure 5. The log−log plot of the distribution of FPT at four different temperatures. The error bars are estimated from five independent 1000-simulation data sets. A log-normal distribution can fit the first peak very well at all four temperatures. A linear dependence at large log t^* at $T^* = 2.95$ and $T^* = 2.7$ suggests a power-law-like decay. A single-exponential fits well for log $t^* > 4.5$ at $T^* = 3.7$, but poorly at $T^* = 3.2$.

In Figure 6a, the fraction not folded (which is one, or 100% of coil structures, at $t^* = 0$) is shown as a function of time at four temperatures. At the highest temperature simulated ($T^* = 3.70$), $f(t^*)$ can be fitted by a single exponential very well except at the very beginning ($t^* < 8000$, Figure 6b). (Note that the fitted time range is an order of magnitude longer than 8000, which is often less than the experimental dead time. Further studies are needed to assess if the missing short-time amplitude from the fitted curve is responsible for the burst phase observed in some folding kinetic experiments.[38]) However, there is no such a good fit even at $T^* = 3.20$. In the two low-temperature curves ($T^* = 2.95$ and 2.70), slowly decaying tails become visually obvious. Thus, as the temperature decreases, the kinetic process becomes more and more non-single-exponential. There is an onset of a slower decay in the long time multiexponential

(35) Kirkpatrick, T. R.; Wolynes, P. G. *Phys. Rev. B* **1987**, *36*, 8552−8564.
(36) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *Folding Des.* **1996**, *1*, 221−230.
(37) Goldstein, R. A.; Luthey-Shulten, Z. A.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4918.
(38) Roder, H.; Colon, W. *Curr. Opin. Struct. Biol.* **1997**, *7*, 15−28.

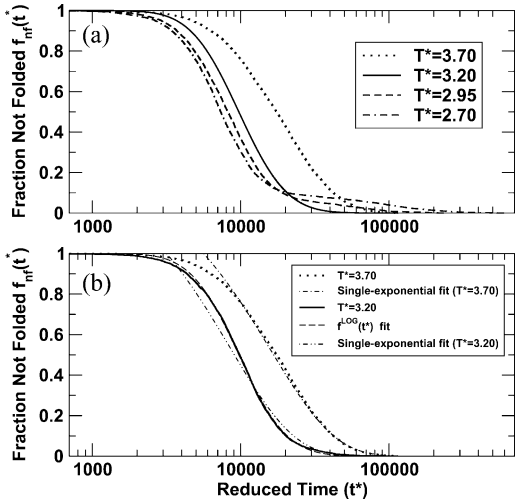


Figure 6. (a) The fraction not folded f_{nf} as a function of reduced time at four different temperatures ($T^* = 3.7$, 3.2, 2.95, and 2.70). Each curve was obtained from 5000 folding simulations. (b) Curve fitting: A single exponential can fit the data very well except for $t^* < 8000$ at the highest temperature ($T^* = 3.7$), but not at $T^* = 3.2$. Both curves ($T^* = 3.2$ and 3.7), however, can be fitted very well by $f^{\log}(t)$. For clarity, only the fit to the curve at $T^* = 3.2$ is shown. The fitting parameters a and b (eq 2) are used to generate the log-normal distribution shown in Figure 5. Note that the slow and nonexponential decay at long times occurs at low temperatures.

kinetic folding process at a temperature near or lower than T_{opt}*, indicating multiple local kinetic barriers encountered during the folding process. Figure 6b further shows that a log-normal fit to the distribution leads to a good fit of $f^{\log}(t^*)$ to $f(t^*)$. This indicates the distribution obtained from the histogram analysis is reasonably accurate. It also suggests that one can profitably calculate $P(t)$ from $f(t)$ if $df(t)/dt$ can be accurately evaluated.

It is of interest to know the free-energy landscapes revealed from these kinetic simulations. The structures obtained from kinetic simulations were analyzed by the method of principal components.[39,40] Results at optimal folding temperatures ($T^* = 3.2$ for the $g = 1.0$ model and $T^* = 3.6$ for the $g = 0.7$ model) are shown in Figure 7. There are three well-defined clusters of structures that belong to the coillike, intermediate, and native states, respectively. The location of the kinetic intermediate is well separated from that of either the native or the coillike state for the large gap model ($g = 1.0$), but much less so for the small gap model. Moreover, there is a significant increase in population of intermediates from $g = 1.0$ to $g = 0.7$. This verifies that the landscape of a small gap model is more rugged than that of a large gap model.

The most populated structure of intermediates at $g = 1.0$ that lead to the onset of the slow decay observed in Figure 6 is found to involve two nucleating hydrophobic residues (F52 and Y45)[23] that contact on the wrong side of the β-hairpin (Figure 2c). We further calculated the rmsd between the structure of this most populated intermediate and those structures in fast-folding trajectories (FPT < 10 000 reduced time units) at $T^* = 2.95$ and the structures of folding trajectories at $T^* = 3.2$ for the $g = 1.0$ model. We found that there is no structure within a rmsd value of 2.5 Å from the intermediate structure in fast-folding trajectories at $T^* = 2.95$. However, similar structures were observed at high temperatures even in the fast-folding trajectories (FPT < 10 000) at $T^* = 3.2$. Thus, it is clear that reducing temperature selectively "freezes" some pathways, while it leaves other fast-folding pathways untouched. The most populated structure of intermediates at $g = 1.0$ is also present in the structures of intermediates at $g = 0.7$ where there is no single dominant intermediate structure.

There is a question if the location of the initial coillike configuration on the landscape correlates with the FPT. A correlation analysis between FPT and the first two principal components does not find any significant correlation. This is further verified by 100 kinetic folding simulations of the same initial coillike configuration with different initial velocities and different initial seeds for random number generators. For the $g = 1.0$ model, we simulated two structures at $T^* = 2.7$, two at $T^* = 3.2$, and three at $T^* = 3.3$. For the $g = 0.8$ model, we simulated two structures at $T^* = 3.3$. We found that an

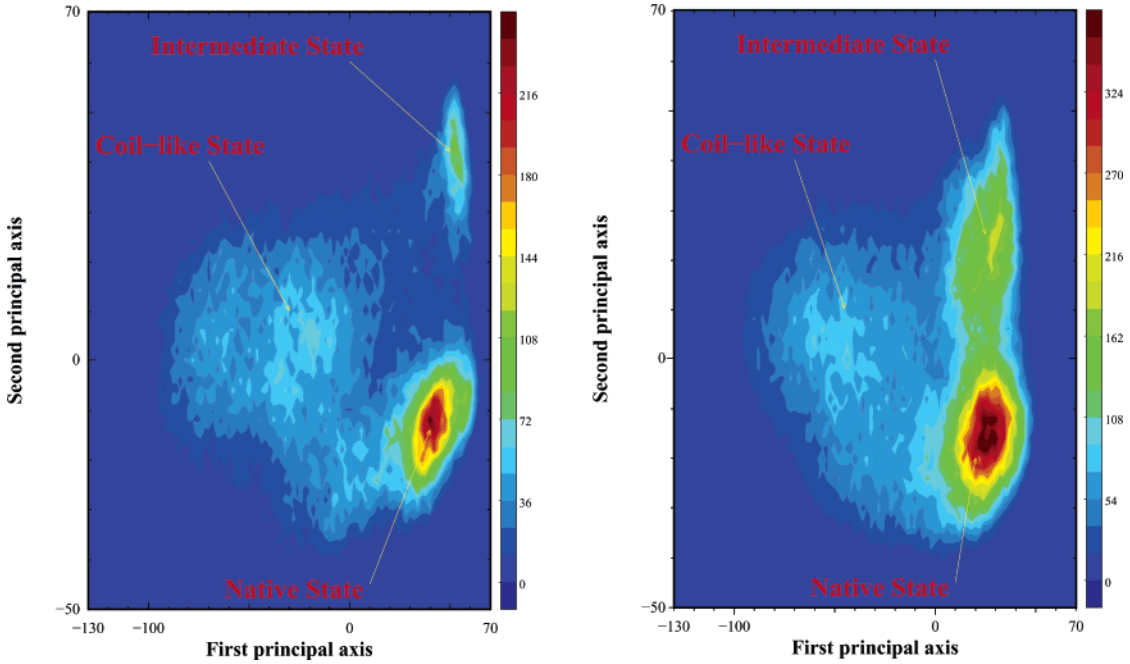


Figure 7. The free-energy surface map at $T^* = T_{opt}$* as a function of the first two principal components. (Left) $T^* = 3.2$ and $g = 1.0$. (Right) $T^* = 3.6$ and $g = 0.7$.

essentially same distribution of FPT can be generated from a single configuration with different initial velocities for this β-hairpin. It is not clear if this result is also true for proteins which have more complex structures than the β-hairpin investigated here.

4. Discussion

In this report, we studied the kinetics of the folding process by using all-atom molecular-dynamics simulations of a β-hairpin fragment of protein G. We investigated the mean and the distribution of the FPT and their behavior at different gap parameters and temperatures. It is shown that both optimal folding temperature and MFPT are lower for a model with a larger bias gap or a smoother folding landscape.

The higher-order moments of FPT were also studied. For temperatures above T_{opt}*, the folding process is self-averaging, and its FPT distribution obeys a simple log-normal-like distribution. The values of the fitted parameter a at $T^* = 3.2$ and 3.7 are close to a constant ($a = 1.21$ and 0.94 at $T^* = 3.2$ and 3.7, respectively). This explains near constant values of $\langle\tau^2\rangle/\langle\tau\rangle^2$ ($\sim$1.5) and $\langle\tau^3\rangle/\langle\tau\rangle^3$ ($\sim$3.2) at high temperature. The value of 3.2 is close to the cube of 1.5 as predicted by the log-normal distribution. As the temperature is below T_{opt}*, the high-order moments start to diverge. This means that the actual folding process may happen in multiple time scale, and the non-self-averaging behavior emerges. In this case, the distribution of FPT develops a long tail at low temperatures.

The MFPT of the model β-hairpin has a U (or V)-shaped dependence on temperature. Similar results have been found in previous theoretical studies[22,41] as well as the simulations of lattice models.[42,43] Although this curve is similar to that found in the non-Arrhenius temperature dependence of experimental kinetic results, the physics are different.[44] This is because, unlike the model peptide, proteins are often less stable at low temperatures.[45] It has been argued that the temperature-dependent behavior obtained here is more suitable for "chevron plot" on denaturant.[15,44,46] In their studies, a crossover from a single exponential at $T^* > T_{opt}$* to multiple exponentials at $T^* < T_{opt}$* was also observed. The qualitatively similar results obtained from various methods ranging from analytical theories to simulations of lattice and off-lattice models indicate that the behavior is likely the generic feature of model proteins.

The distribution of FPT obtained here is in qualitative agreement with that obtained from the analytical theory.[22] For example, both theoretical and simulation results show that the reduced high-order moments are nearly constant at high temperatures and start to diverge below the optimal folding temperature. Some quantitative differences are also observed. The FPT distribution at high temperature is a Poisson distribution in the analytical theory but is log-normal from simulations. The MFPT and reduced high-order moments predicted from the theory also change much more drastically with temperature than those given by the simulations. This is in part due to the fact that Gō-like models studied here have a smoother landscape than the models studied in analytical theories. The roughness of a Gō-like model can be estimated as follows. The density of states obtained from the weighted histogram analysis is separated into the contribution of the native state (structures with rmsd $\leq$ 2.5 Å) and that of the nonnative state (structures with rmsd $>$ 2.5 Å). The distributions of the logarithm of density of states were approximately fitted by two Gaussian functions. The roughness of the underline landscape can be estimated from the ratio of the difference in the average energies of the Gaussians and the square root of the sum of the variances of the two Gaussians. For the $g = 1.0$ model, a ratio of 9.1 indicates that the roughness is relatively small as compared with the native driving force. This is because, unlike analytical models, a Gō-like model does not have fluctuation in the energies of native and nonnative atomic contacts. That is, the roughness of the landscape for a Gō-like model is mainly caused by nonspecific collapse rather than energetic frustration. As the bias gap g decreases, the roughness increases (the ratio is 7.7 at $g = 0.8$). Although this study is limited to the highly optimized Gō model, it provides a microscopic detail and foundation for the existence of a long-decay tail.

Traditional ensemble-averaged experimental studies on the folding kinetics of this β-hairpin have been made,[47,48] where single-exponential kinetics were observed over a temperature range from 270 to 310 K. It would be of interest to perform similar experiments on the single-molecule level in a wider temperature range and compare them with the results presented here.

Acknowledgment. We would like to thank Professors Peter Wolynes, Hue Sun Chan, and reviewers for critical reading and helpful comments. Y.Z. would like to thank the hospitality of Professors Ruibao Tao, Zhifang Lin, and Jian Zi while he was a visiting professor of T. D. Lee Physics Laboratory and Research Center for Theoretical Physics, Fudan University, during which part of this work was completed. The work at Buffalo was supported by a grant from HHMI to SUNY Buffalo and by the Center for Computational Research and the Keck Center for Computational Biology at SUNY Buffalo. J.W. would like to thank Professors Erkang Wang, Shangjun Dong, Xiurong Yang, Jinyue Gao, and Tao Zhang for their hospitality while he was an adjunct professor of State Key Laboratory of Electro-Analytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences and Department of Physics at Jilin University, during which part of this work was completed. G.S. gratefully acknowledges the support of the Division of Chemical Sciences, Office of Basic Sciences, Office of Energy Research, U.S. Department of Energy.

Note Added in Proof. A just-published experimental study of a FBP WW domain provided the first example of a multiple-exponential behavior that can be induced by decreasing the temperature below the midpoint of unfolding (Nguyen, H.; Jäger, M.; Moretto, A.; Gruebele, M.; Kelly, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3948−3953), similar to what is reported here.

JA029855X

(39) Garcia, A. E.; Sanbonmatsu, K. Y. *Proteins* **2001**, *42*, 345−354.
(40) Kamiya, N.; Higo, J.; Nakamura, H. *Protein Sci.* **2002**, *11*, 2297−2307.
(41) Saven, J. G.; Wang, J.; Wolynes, P. G. *J. Chem. Phys.* **1994**, *101*, 11037.
(42) Gutin, A. M.; Abkevich, V. I.; Shakhnovich, E. I. *Phys. Rev. Lett.* **1996**, *77*, 5433.
(43) Klimov, D. K.; Thirumalai, D. *J. Chem. Phys.* **1998**, *109*, 4119.
(44) Chan, H. S.; Dill, K. A. *Proteins* **1998**, *30*, 2−33.
(45) Privalov, P. L. *Adv. Protein Chem.* **1979**, *33*, 167−241.
(46) Kaya, H.; Chan, H. S. *J. Mol. Biol.* **2003**, *326*, 911−931.
(47) Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196−199.
(48) Muñoz, V.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5872−5879.
(49) Zhou, Y.; Linhananta, A. J. *J. Phys. Chem. B* **2002**, *106*, 1481−1485.