Velocity convergence of free energy surfaces from single-molecule measurements using Jarzynski's equality

Nolan C. Harris and Ching-Hwa Kiang*
Department of Physics and Astronomy, Rice University, Houston, Texas 77005, USA
(Received 22 April 2008; revised manuscript received 9 March 2009; published 10 April 2009)

We studied the velocity dependence of mechanical unfolding of single protein molecules with the atomic force microscope. We showed that with enough realizations, the free energy surfaces reconstructed from Jarzynski's equality converge with respect to pulling velocity, in good agreement with theory. Using the I27 domain of titin as an example, we estimated the required number of realizations for a given pulling velocity, and suggested the optimal range of velocities for single-molecule experiments. The results demonstrate that Jarzynski's equality is a powerful and practical tool for reconstructing free energy landscapes.

DOI: 10.1103/PhysRevE.79.041912 PACS number(s): 87.15.−v, 05.20.−y, 05.40.−a, 05.70.Ln

The free energy surface for a biomolecular process, such as unfolding of a protein or a nucleic acid, provides insight into the biological function associated with that process. The development of single-molecule manipulation techniques, typically done using atomic force microscope (AFM) [1,2] and optical tweezers [3,4], combined with the nonequilibrium work theory [5,6] offers an opportunity for experimentally mapping the free energy surface of such molecular processes for the first time.

These methods, however, are usually performed out of equilibrium, and relating the results to meaningful molecular properties has been difficult. The recently derived Jarzynski's equality [6,7], which relates nonequilibrium measurements to equilibrium free energies, has shown promise for extracting equilibrium information from single-molecule experiments [2–4,8–10]. Jarzynski's equality recovers equilibrium free energies by preferentially weighting the rare smaller work realizations in a nonequilibrium work distribution via exponential averaging [11]. The equality holds for processes performed arbitrarily far from equilibrium. However, the farther from equilibrium a process is performed, the more realizations are required to ensure the dominant events of a work distribution are sampled.

Questions remain concerning the rate of convergence of Jarzynski's equality as experiments are performed farther from equilibrium. Some studies found that practical applications of Jarzynski's equality in simulations are limited [12,13]. On the other hand, under certain conditions or when incorporating strategies for improving sampling efficiency for faster convergence, Jarzynski's equality may be advantageous [14–17]. The free energy surfaces from forced protein unfolding simulations have been shown to compare well with experimental results [18,19]. In these simulations, Jarzynski's equality converged fast enough that it was useful for relatively slow pulling velocities, comparable to those used in single-molecule experiments. It has recently been demonstrated that it is possible to use Jarzynski's estimator to determine the free energy surfaces of biomolecule unfolding from nonequilibrium single-molecule experiments [2,9,10]. In experiments practical restrictions such as instrument stability and the time a molecule can be stably attached to the cantilever tip limit the feasibility of performing an experiment infinitely slowly. As a result, few systems have been successfully studied using single-molecule techniques under equilibrium conditions, which make application of Jarzynski's equality particularly attractive.

Debate over the practicality of Jarzynski's equality in determining experimental free energy landscapes stems from the large number of realizations required for single-molecule measurements. Thus, it is important to examine the convergence behavior of the Jarzynski estimator as the process moves farther from equilibrium. Specifically, the number of samples sufficient to determine the nonequilibrium work distribution of a particular process is related to the pulling velocity. Here we report studies of unfolding a single protein domain, via AFM, over a range of constant pulling velocities. We found that the results from Jarzynski's estimator are consistent with the expectations that equilibrium free energy is independent of pulling velocity and the measurement noise does not pose a significant problem. Jarzynski's equality can recover equilibrium free energy efficiently in experiments where it would otherwise be impossible.

We used AFM to mechanically unfold the I27 domain of human cardiac titin. A schematic illustrating the single-molecule techniques is given in Fig. 1(a). Experiments were performed at pulling velocities ranging from 0.02 to 5.0 μm/s, using a cantilever with spring constant of 0.05 N/m. Representative force versus time curves are shown in Fig. 1(b). The unfolding forces were measured, and distributions at each velocity are displayed in Fig. 2(a). The most probable unfolding force as a function of pulling velocity, determined using the fit to a Gaussian distribution, is shown in Fig. 2(b) [20]. These data can be used to fit the phenomenological model based on Bell's theory [21] or the high-force microscopic model based on Kramers theory [22–24] to obtain unfolding barrier height information. Here we used nonequilibrium work theorem to analyze the results, and attempt to estimate the barrier height information to compare the results from our analysis to that from the models.

We used Jarzynski's equality to reconstruct the free energy curve directly from experimental data. Detailed procedures are described in Ref. [2]. In brief, we used the histo-

*Author to whom correspondence should be addressed; chkiang@rice.edu

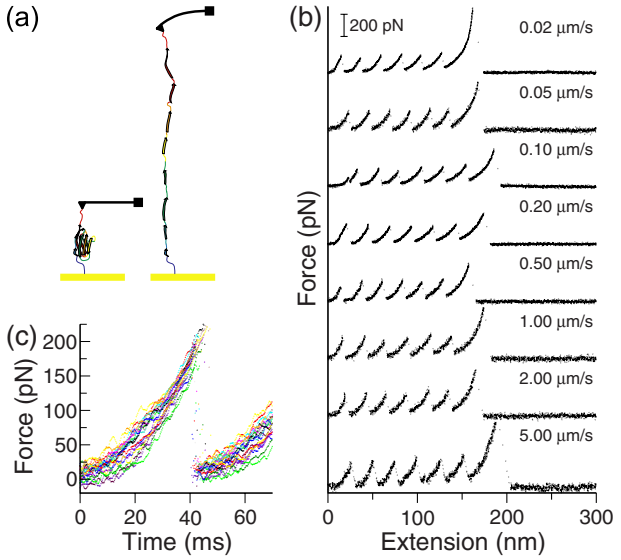

FIG. 1. (Color online). (a) A schematic of single-molecule manipulation using AFM. A protein is anchored at one end to a gold substrate and at the other end to a silicon nitride tip on a cantilever spring. The protein is pulled at a constant velocity until the force on the molecule is large enough for a single domain to rupture and unfold. (b) Representative force-extension curves of titin I27 domain pulled at different velocities. (c) Force versus time trajectories for a titin I27 domain unfolding at a pulling velocity of 1 μm/s. The curves were smoothed with a smoothing spline for display purposes.

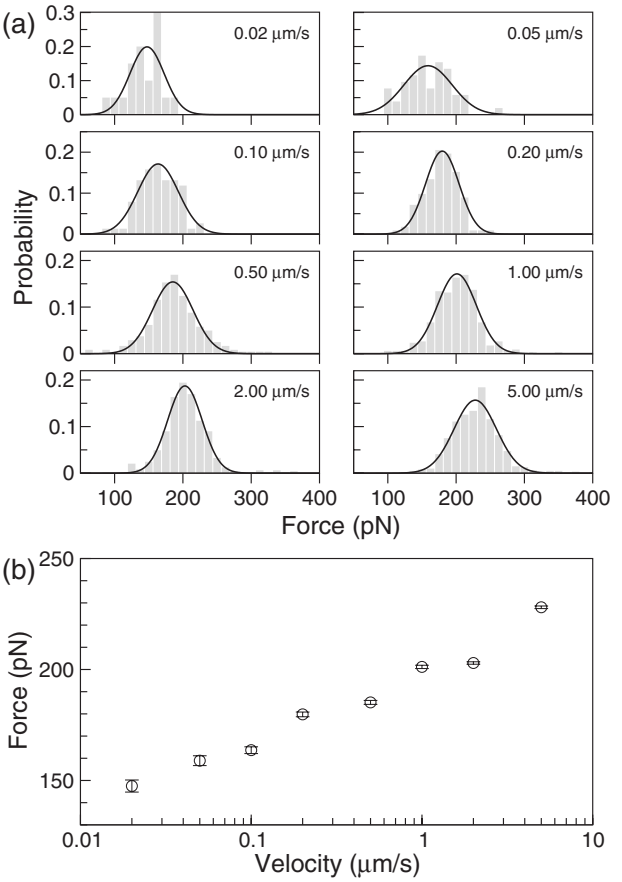

FIG. 2. (a) Histogram of rupture forces for pulling velocities v ranging from 0.02 to 5 μm/s. Solid lines are fits to Gaussian distribution. (b) The most probable rupture force as a function of v. Error bars indicate the standard statistical error of the most probable force as determined by fitting to a Gaussian distribution.

gram method derived from the exact formula [2,6] to reconstruct the free energy curve G as a function of molecular end-to-end distance z. To calculate $G(z)$, we averaged the work from each time step $T=\tau/\delta t$ of duration δT,

$$\exp\{-\beta G[z^{(m)}]\} \approx \frac{1}{NT}\sum_{n=1}^{N}\sum_{s=1}^{T}\delta_{\epsilon}[z^{(m)}-z_{n,s}] \times \exp\{-\beta[W_{n,s}-U(z_{n,0},\lambda_A)]\}, \quad (1)$$

where n is the nth trajectory, s is the sth time step, $W_{n,s}$ is the work performed up to t_s, and $z_{n,s}$ is the value of z at time $t_s=s\delta t$. $U=\frac{1}{2}kx^2$ is the energy stored in the spring, where k is the cantilever spring constant and x is the distance the cantilever tip moved from its equilibrium position. The z axis is divided into bins of width ϵ with $z^{(m)}$ as the midpoint of the mth bin. The process is repeated for each pulling velocity.

The calculated free energy surfaces at various pulling velocities converge to within 10%, as shown in Fig. 3(a). For comparison, the profiles of the average work done on the system, which include dissipated work, are shown in Fig. 3(b). The average work increases with the pulling velocity, which is evidenced in Fig. 3(b) and the non-Gaussian distributions of work in Fig. 3(c). However, Jarzynski's equality exponentially weights these work distributions such that the resulting free energy surfaces are independent of pulling velocity, as shown in inset of Fig. 3(b). Unlike the phenomenological approach, our approach does not require assuming explicitly the apparent stiffness of the system, which includes both the spring and the molecule. This demonstrates the feasibility of using Jarzynski's equality to reconstruct free energy surfaces provided that the work distribution is properly sampled. Consistency in the surfaces reconstructed from different velocities further verifies the use of Jarzynski's equality for free energy calculations from single-molecule manipulation experiments.

For pulling at near-equilibrium velocities, thermal fluctuations will have equal probability of lowering and raising the forces, with a very narrow distribution. In other words, the Jarzynski equation reduces to the thermodynamic statement that "free energy equals reversible work" in the low velocity limit. For example, at the speed of 0.01 μm/s, the Jarzynski and thermodynamic work estimates are within 10 kcal/mol for the free energy. Since most of the experiments were done far from equilibrium, there was significant dissipation during each process. Therefore, the work done on the system will be on average larger than the free energy difference. This is confirmed in Fig. 3(c), which shows the higher the pulling velocity, the more work, and therefore, the higher the force is required to pull the molecule over the transition state. However, using Jarzynski's equality, we were able to recover the equilibrium information. The reconstruction procedure can be verified by comparing Jarzynski-derived results from results generated at low pulling velocities. In this limit, the free

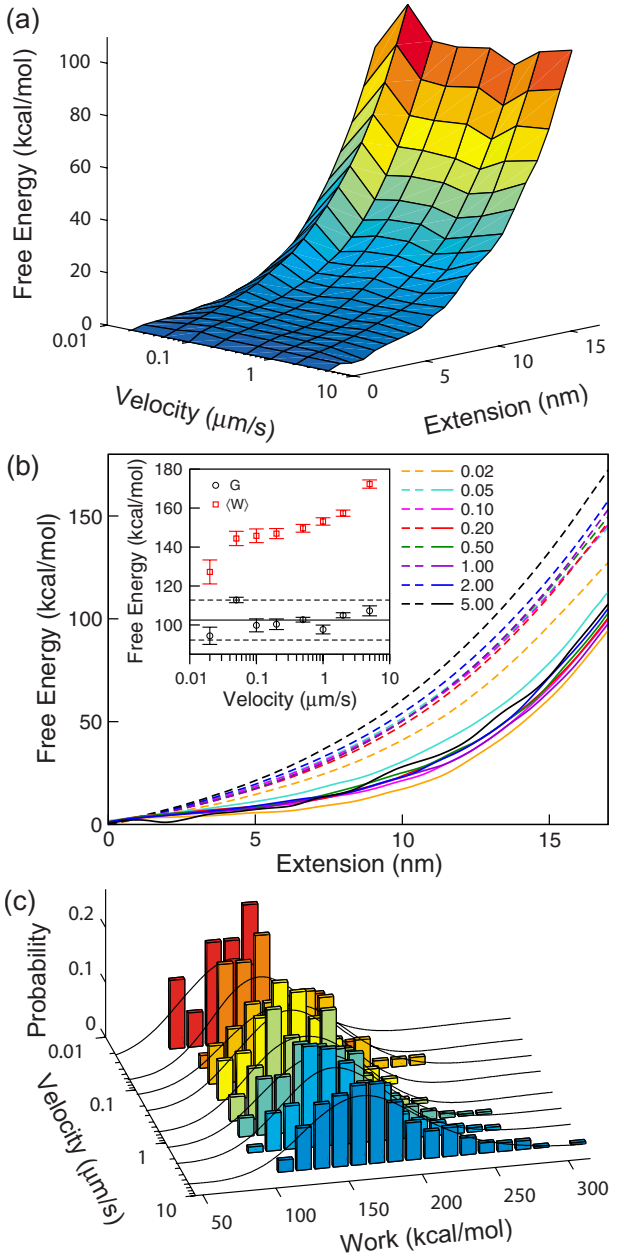

FIG. 3. (Color online). (a) Free energy surfaces for unfolding of a single titin I27 domain at various pulling velocities. (b) Average work $\langle W \rangle$ performed for each pulling velocity v (in μm/s) and the free energy G reconstructed from Jarzynski's equality. Solid lines: G determined from Jarzynski's equality. Dashed lines: $\langle W \rangle$ calculated by weighting each trajectory equally. Inset: G and $\langle W \rangle$ at 17 nm as a function of v. Error bars are statistical uncertainties calculated using the bootstrap method. Dashed line represents the mean of free energies $\langle G \rangle = 100$ kcal/mol from all velocities. Shaded region indicates 10% uncertainty in G. (c) Distributions of W at 17 nm as a function of v. Solid lines are smoothing spline fits to each distribution, which are non-Gaussian.

energy as a function of extension is simply the integral of the reversible work. Our curves obey this limit, which implies the reconstruction procedure is correct.

To investigate the convergence of free energies G for different pulling velocities, we plot the estimated G as a function of number of realizations. Since the experimental data consist of both statistical and systematic errors such as instrument drift and uncertainty in cantilever spring constant, we estimated the error in averaged ΔG to be at least 10%. The estimated number of realizations required for converging to within 10% of the averaged final value for each velocity is shown in Fig. 4. For a sufficiently slow pulling velocity such as 0.1 μm/s, fewer than 50 realizations are required. For common pulling velocities used in AFM experiments, e.g., 1.0 μm/s, greater than 200 realizations are required to ensure proper sampling of the work distributions at the low values, which dominate the resulting values of the free energies.

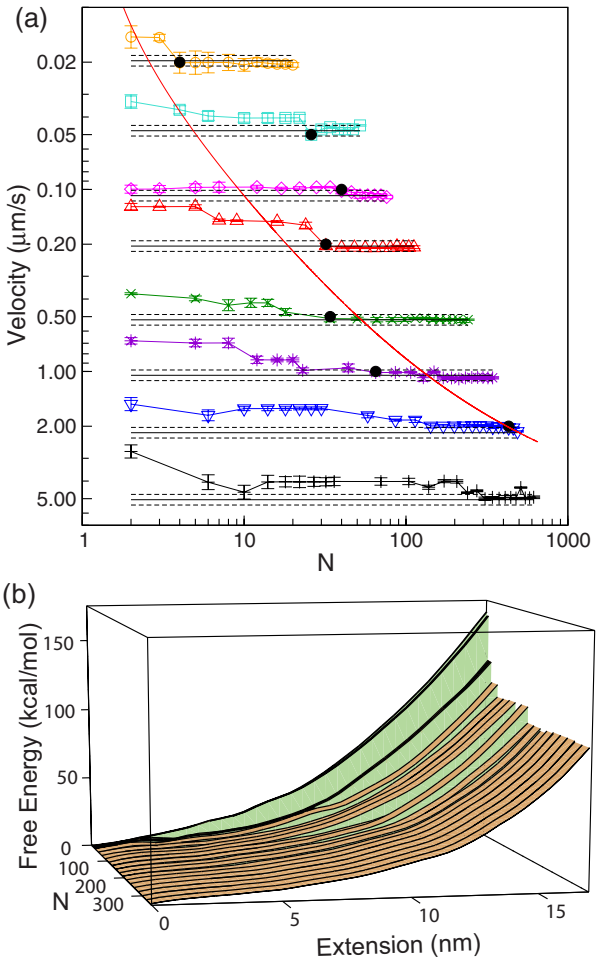

FIG. 4. (Color online). (a) The convergence of G with respect to the number of realizations N for different pulling velocities v. Solid circles represents N_c, the estimated number of realizations required for the free energy to converge to within 10% of the averaged free energy $\langle \Delta G \rangle$, which is represented by solid lines. The labels on the Y axis show the v for each curve. Due to hydrodynamic drag, the N_c for 5.0 μm/s data was not determined because the data contain systematic error that are not included in the error bars, which include only statistical error. Solid curve is an empirical equation $N_c = 1 + 20(e^{1.5v/v_0} - 1)$, where $v_0 = 1$ μm/s. The parameters were determined by least square fit to the estimated N_c with the constraint $N = 1$ when $v \to 0$. (b) The reconstructed free energy surfaces as a function of N at $v = 1.0$ μm/s.

We found that the number of pulling trajectories required

to sample the work distribution properly to obtain an accurate free energy is reasonable. In the lower pulling velocity regime, the number of realizations required for velocities such as 0.02 μm/s is small. However, they are difficult to obtain experimentally because of instrument drift, which introduces systematic errors that may significantly affect the Jarzynski's average. The high-velocity experiments require more trajectories, and the extra time associated with each realization and the computing time for integration grows rapidly with pulling velocity. In addition, hydrodynamic drag may affect the results at higher pulling velocities [25]. Thus we concluded that 0.2–1.0 μm/s is the optimal range of pulling velocity. The number of realizations required to obtain an accurate estimate of free energy is reasonable, and the systematic errors are kept at a minimum. Figure 4(b) shows the convergence of the free energy surface with increased sampling at a pulling velocity of 1.0 μm/s. The estimated profiles resemble the final surface within 100 trajectories, and converge within 10% after 200 realizations.

In conclusion, we study the free energy reconstruction using Jarzynski's equality under different pulling velocities and found that with enough realizations, the free energy surface does not depend on the pulling velocity. Using the unfolding of the titin I27 domain as a demonstration, for a typical puling velocity of 0.3 μm/s, fewer than 100 trajectories are required for convergence of the free energy surface. The convergence with respect to velocity validates the application of Jarzynski's equality for reconstructing biomolecular free energy landscapes. The method allows us to obtain the folding landscape in the region not probed by chemical denaturant studies, i.e., from the relaxed unfolded state to the extended unfolded state, where the proteins are likely to be when first synthesized *in vivo*. The method is also useful for studying protein and nucleic acid folding and translocation as well as receptor-ligand binding free energy landscapes [26–29], where quantitative information of free energy may help us to understand the fundamentals of biological interactions.

We thank NSF (Grant No. DMR-0505814), NIH (Grant No. T90DK70121), and Welch Foundation (Grant No. C-1632) for support.

[1] T. Hugel, M. Rief, M. Seitz, H. E. Gaub, and R. R. Netz, Phys. Rev. Lett. **94**, 048301 (2005).
[2] N. C. Harris, Y. Song, and C.-H. Kiang, Phys. Rev. Lett. **99**, 068101 (2007).
[3] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Science **296**, 1832 (2002).
[4] W. J. Greenleaf, K. L. Frieda, D. A. N. Foster, M. T. Woodside, and S. M. Block, Science **319**, 630 (2008).
[5] C. Jarzynski, Prog. Theor. Phys. **165**, 1 (2006).
[6] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **98**, 3658 (2001).
[7] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[8] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Nature (London) **437**, 231 (2005).
[9] R. A. Nome, J. M. Zhao, W. D. Hoff, and N. F. Scherer, Proc. Natl. Acad. Sci. U.S.A. **104**, 20799 (2007).
[10] J. Preiner, H. Janovjak, C. Rankl, H. Knaus, D. A. Cisneros, A. Kedrov, F. Kienberger, D. J. Müller, and P. Hinterdorfer, Biophys. J. **93**, 930 (2007).
[11] C. Jarzynski, Phys. Rev. E **73**, 046105 (2006).
[12] R. C. Lua and A. Y. Grosberg, J. Phys. Chem. B **109**, 6805 (2005).
[13] H. Oberhofer, C. Dellago, and P. L. Geissler, J. Phys. Chem. B **109**, 6902 (2005).
[14] B. Cleuren, C. Van den Broeck, and R. Kawai, Phys. Rev. Lett. **96**, 050601 (2006).
[15] S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten, J. Chem. Phys. **119**, 3559 (2003).
[16] F. M. Ytreberg and D. M. Zuckerman, J. Chem. Phys. **120**, 10876 (2004).
[17] D. A. Hendrix and C. Jarzynski, J. Chem. Phys. **114**, 5974 (2001).
[18] A. Imparato, A. Pelizzola, and M. Zamparo, Phys. Rev. Lett. **98**, 148102 (2007).
[19] A. Imparato, S. Luccioli, and A. Torcini, Phys. Rev. Lett. **99**, 168101 (2007).
[20] T. Strunz, K. Oroszlan, R. Schäfer, and H.-J. Güntherodt, Proc. Natl. Acad. Sci. U.S.A. **96**, 11277 (1999).
[21] G. I. Bell, Science **200**, 618 (1978).
[22] H. A. Kramers, Physica **7**, 284 (1940).
[23] O. K. Dudko, A. E. Filippov, J. Klafter, and M. Urbakh, Proc. Natl. Acad. Sci. U.S.A. **100**, 11378 (2003).
[24] O. K. Dudko, G. Hummer, and A. Szabo, Phys. Rev. Lett. **96**, 108101 (2006).
[25] H. Janovjak, J. Struckmeier, and D. J. Müller, Eur. Biophys. J. **34**, 91 (2005).
[26] R. F. Fox, Proc. Natl. Acad. Sci. U.S.A. **100**, 12537 (2003).
[27] M. Rief, H. Clausen-Schaumann, and H. E. Gaub, Nat. Struct. Biol. **6**, 346 (1999).
[28] V. Barsegov and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **102**, 1835 (2005).
[29] C. Ke, M. Humeniuk, H. S-Gracz, and P. E. Marszalek, Phys. Rev. Lett. **99**, 018302 (2007).