PHYSICAL REVIEW E **74**, 061120 (2006)

Free-energy reconstruction from experiments performed under different biasing programs

David D. L. Minh*
Center for Theoretical Biological Physics and Department of Chemistry and Biochemistry, University of California at San Diego, San Diego, California 92093, USA

(Received 23 August 2006; published 28 December 2006)

Recently developed nonequilibrium statistical physics relationships, including Jarzynski's equality and the Crooks fluctuation theorem, have been used to calculate equilibrium thermodynamic properties using data from both laboratory and computational experiments. Although Jarzynski's derivation does not include an explicit time dependency, prior work utilizing the relationship to reconstruct free-energy surfaces has combined data from experiments performed under identical conditions. Here, a formalism is developed for combining data from a variety of biasing protocols, as in dynamic force spectroscopy experiments. The method is then demonstrated on data from simulations conducted under a wide range of pulling velocities and with a random biasing protocol.

DOI: 10.1103/PhysRevE.74.061120 PACS number(s): 05.70.Ln, 82.20.Wt, 87.10.+e, 87.15.La

I. INTRODUCTION

Molecular scientists are interested in free-energy differences. Recently developed statistical physics relationships, including Jarzynski's equality [1] and the Crooks fluctuation theorem [2], can be used to calculate this key thermodynamic property along a reaction coordinate using data from nonequilibrium laboratory and computational experiments [3,4]. These theorems have been experimentally verified through single-molecule force spectroscopy experiments [5,6], in which atomic force microscopes or optical tweezers are used to apply mechanical force to individual molecules. Jarzynski's equality has also been applied to reconstruct free-energy surfaces from steered molecular dynamics [7–9] and fast-growth thermodynamic integration [10], which are analogous numerical simulation techniques. The analysis of steered molecular dynamics simulations has been further improved by applying the Crooks fluctuation theorem to forward and reverse pullings along a reaction coordinate [11].

Nonequilibrium biasing experiments are performed with a time-dependent Hamiltonian in which the position of a transducer, i.e., an atomic force microscopy cantilever, changes the external potential experienced by the molecule. I will refer to the transducer position as a function of time as the biasing program Jarzynski's equality,

$$e^{-\beta\Delta F} = \langle e^{-\beta W_t} \rangle, \tag{1}$$

relates the exponentially weighted average of nonequilibrium accumulated work W_t to the equilibrium free-energy difference ΔF between states [1]. Noting that free energy is a state function, Atilgan and Sun devised methods of optimizing the biasing program to calculate the free-energy difference between two states [12]. The Bennett acceptance ratio method [13], originally intended for two equilibrium ensembles, has been generalized to nonequilibrium work [2] and multiple pathways between multiple states [14]. However, prior applications of the nonequilibrium work relation to surface reconstruction (free energy as a function of coordinate position instead of state [4,5,15]) have not exploited path independence, choosing instead to combine data from experiments performed under identical conditions. These experiments are most often conducted with linear force ramping, but cyclic [16] and periodic [17] processes have also been considered.

A statistical mechanics formalism which combines data from experiments performed under different biasing protocols is of considerable interest to both laboratory and theoretical scientists. Molecules perturbed at different loading rates exhibit varied hysteresis behavior [5], including different rupture forces [18]. In dynamic force spectroscopy [19], pulling experiments are conducted under a wide range of loading rates. In spite of the vast store of data currently available from force-extension curves along the entire range of molecular extension, practitioners of dynamic force spectroscopy typically make use of one data point: the rupture force [20]; essentially, they measure tensile strength on a microscopic scale. Manosas and Ritorti calculate free-energy surfaces based on the fraction of molecules which unfold more than once during a pulling path, an alternate procedure which also makes scarce use of available data [21]. In the computational realm, this formalism is useful for interpreting simulations which can be performed under different biasing protocols, such as steered molecular dynamics [22–24] and adiabatic bias molecular dynamics [25–27] simulations.

In this paper, I develop a formalism for combining data from a variety of pulling protocols to calculate free-energy surfaces, expanding the range of applicability of Jarzynski's equality.

II. FORMALISM FOR DATA COMBINATION

Individual trajectories from single-molecule pulling experiments exhibit a range of accumulated work values, with data sets collected at faster pulling speeds leading to a larger mean and variance of dissipated work [6]. In spite of the differences between work distributions from experiments conducted at different force loading rates, data can still be combined to calculate free-energy differences because the free energy is a state function and expectation value for all

---

*Electronic address: dminh@mccammon.ucsd.edu; URL: http://mccammon.ucsd.edu/~dminh/

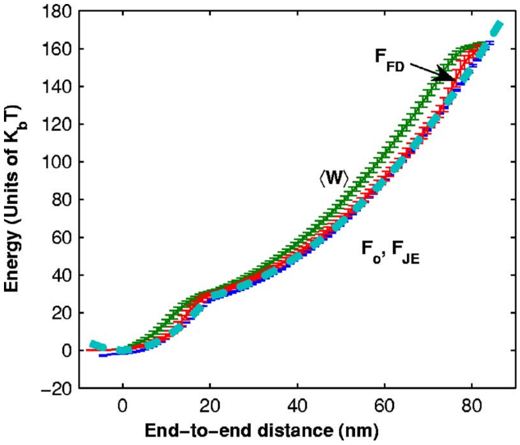

FIG. 1. (Color online) Average free-energy surfaces reconstructed from 200 trajectories loaded at 10 pN/s, using mean work $\langle W\rangle$, fluctuation-dissipation theorem F_{FD}, and Jarzynski equality F_{JE} metrics. Standard deviations are indicated by error bars and the original potential F_o by the thick dashed line.

the distributions is the same, namely $e^{-\beta\Delta F}$. By the laws of probability, random sampling from multiple distributions with the same expectation value will lead to another distribution with this expectation value.

Consider a system in which the Hamiltonian can be written as $H(\mathbf{x},t)=H_0(\mathbf{x})+V[\mathbf{x},t]$, where $H_0(\mathbf{x})$ is the original Hamiltonian and $V[\mathbf{x},t]$ is the perturbing bias. If the system evolves according to an explicitly time-dependent Liouville operator with the Boltzmann distribution as a stationary solution, then

$$\frac{e^{-\beta\{H_0(\mathbf{x})+V[\mathbf{x},t]\}}}{\int e^{-\beta\{H_0(\mathbf{x}')+V[\mathbf{x}',0]\}}d\mathbf{x}'}=\langle\delta[\mathbf{x}-\mathbf{x}(t)]e^{-\beta W_t}\rangle, \qquad (2)$$

where W_t is the accumulated work [4,10]. In a typical single-molecule pulling experiment, a harmonic potential with spring constant k_s between the bias center transducer position, $b(t)$, and the reaction coordinate, $z_t=z[\mathbf{x}(t)]$, leads to the bias $V[\mathbf{x},t]=V[z_t,b(t)]=k_s[z_t-b(t)]^2/2$. The accumulated work $W_t=\int_0^t\frac{\partial H(\mathbf{x}(t'),t')}{\partial t'}dt'$, along each nonequilibrium trajectory is $\int_0^t k_s\{z[\mathbf{x}(t')]-b(t')\}\frac{\partial b(t')}{\partial t'}dt'$ for a generic time-dependent harmonic potential. Under a linear force ramp $b(t)=b_0+vt$, the transducer moves from its position b_0 at time $t=0$ with a velocity v, and $W_t=k_s v\{vt^2/2-\int_0^t z[\mathbf{x}(t')]dt'\}$.

As in Hummer and Szabo's work, the symbol $\langle\cdots\rangle$ in Eq. (2) denotes the average over an ensemble of trajectories which satisfy a few conditions: (i) they start from the equilibrium distribution at $t=0$, (ii) evolve according $\frac{\partial f(\mathbf{x},t)}{\partial t}=\mathfrak{L}_t f(\mathbf{x},t)$, where the operator $\mathfrak{L}_t$ has the stationary solution $\mathfrak{L}_t e^{-\beta H(\mathbf{x},t)}=0$, and (iii) conclude at time t [10]. These requirements are unnecessarily strict. As the left hand side of Eq. (2) does not include any explicit dependence on history of the $H(\mathbf{x},t)$, any time-dependent Hamiltonian with the same final bias will have the same ensemble average. The time it takes to reach this bias is also irrelevant. Thus I use $\langle\cdots\rangle^*$ to denote an average over all trajectories satisfying (i), evolving according to all operators and biasing protocols satisfying (ii), ending at all time points $0<t<\infty$, and concluding at the same bias, $V[z_t,b(t)]$. This trajectory ensemble is vastly generalized from previous work [10,17].

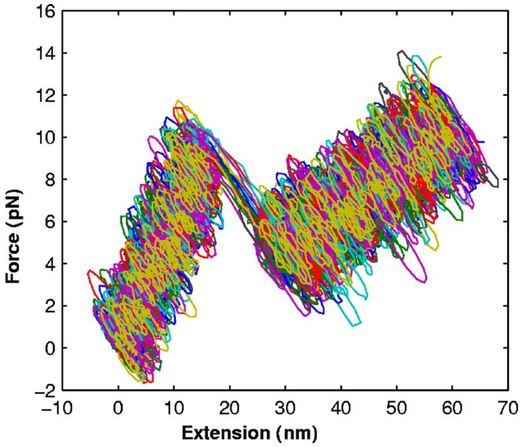

FIG. 2. (Color online) Smoothed force-extension profiles from trajectories at 20 evenly spaced velocities.

To obtain a free-energy reconstruction formula, take Eq. (2), replace $\langle\cdots\rangle$ with $\langle\cdots\rangle^*$ and $V[\mathbf{x},t]$ by $V[z_t,b(t)]$, multiply both sides by $e^{-\beta V[z_t,b(t)]}\delta(z-z_t)$, and integrate over x, leaving

$$e^{-\beta F_0(z)}=\langle\delta(z-z_t)e^{-\beta(W_t-V[z_t,b(t)])}\rangle^*. \qquad (3)$$

Although in principle one can reconstruct the complete free-energy surface $F_0(z)$ from any ensemble of measurements, in practice most sampling will be limited to a small region of z around b. Accuracy can be improved by adapting the weighted histogram analysis method [28] to combine free-energy profiles from different bias centers,

$$F_o(z)=-\beta^{-1}\ln\frac{\sum_{i=1}^{N_w}\langle\delta[z-x(t)]e^{-\beta W_{t,i}}\rangle^*/\langle e^{-\beta W_{t,i}}\rangle^*}{\sum_{i=1}^{N_w}e^{-\beta V(z,b(t))}/\langle e^{-\beta W_{t,i}}\rangle^*}, \qquad (4)$$

where accumulated work measurements $W_{t,i}$ with close bias centers are binned together in a window, and the sum is over all of these bias slices. This differs from the procedure of Hummer and Szabo because the average is taken over bias slices rather than time slices.

As with histogram construction, a key parameter in this procedure is the choice of bias center bin size. One must strike a data-dependent balance, weighing bias potential precision and sufficient sampling in each window. For the numerical simulations in this paper, the number of bias center bins is set to the number of discrete time points, yielding time slices in the special case of a set of experiments performed at the same constant pulling velocity. The main benefit of this formalism, however, is the promising property that biasing windows can be determined *a posteriori*, allow-

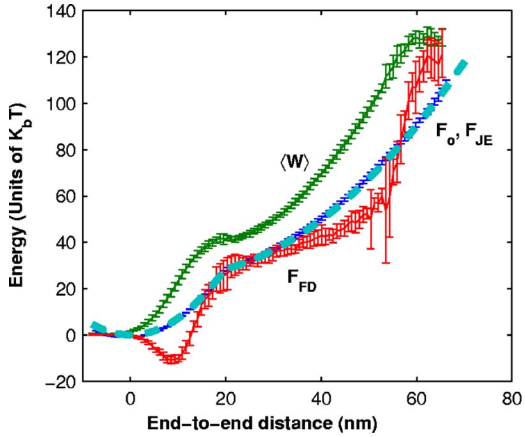

FIG. 3. (Color online) Average free-energy surfaces reconstructed from ten blocks of trajectories loaded at 20 different velocities, annotated as in Fig. 1.

ing data to be combined even when loading rates are adjusted from experiment to experiment, or during an experiment.

III. SIMULATIONS ON A ONE-DIMENSIONAL ENERGY SURFACE

As a toy demonstration of Eq. (4), Brownian dynamics simulations were run on a double-well potential, with folded (F_f) and unfolded (F_u) states represented by $F_f(z)=\frac{k_f}{2}z^2$ and $F_u(z)=\frac{k_u}{2}(z-\Delta z)^2+\Delta F_u$ where k_f (0.6 pN/nm) and k_u (0.2 pN/nm) are spring constants for the folded and unfolded states, respectively, Δz is the minima of the unfolded state (8 nm), and ΔF_u is the unfolding free energy (25 k_bT). Simulations were run at 300 K, with a 1-ms time step, and the diffusion constant was set to 1200 nm^2/s, which reflect the parameters of Hummer and Szabo [4]. All simulations were stopped when the bias center reached 87.5 nm.

To demonstrate the validity of bias slicing, free-energy surfaces were constructed from 200 trajectories with a force loading rate of 10 pN/s. The biasing spring constant k_s was set at 2.0 pN/nm. Accumulated work was numerically integrated by $W_t=kv[vt^2-\Sigma_{j=1}^{l}(t_j-t_{j-1})(z_j+z_{j-1})/2]$, where the sum is over the discrete time steps l required to reach time t [10]. Results from bias slicing F_{JE} are compared with the mean work $\langle W_t\rangle$ and fluctuation-dissipation F_{FD} calculations. The fluctuation-dissipation theorem estimator $\Delta F\approx\langle W_t\rangle-\beta\sigma^2/2$, where σ^2 is the variance of the work distribution, is equivalent to the second-order cumulant expansion of the Jarzynski estimator, valid in the near-equilibrium regime [5]. In this near-equilibrium simulation, both F_{FD} and F_{JE} are found to be good estimators of the free-energy surface (Fig. 1).

The bias slicing method was then tested under the more challenging conditions: dynamic force spectroscopy with a softer spring constant of 0.4 pN/nm. Simulations were conducted at 200 equally spaced force loading rates between 35 and 175 pN/s. Figure 2 shows representative force-extension profiles from these trajectories. Much like experimental profiles, these curves exhibit bistability and hysteresis [29] and have velocity-dependent rupture forces [18]. As the trajectories exist far from equilibrium and occur at different force loading rates, the mean work and fluctuation-dissipation theorem calculations are not expected to be correct. Their gross inaccuracy emphasizes the nontriviality of accurate surface reconstruction from bias slicing (Fig. 3).

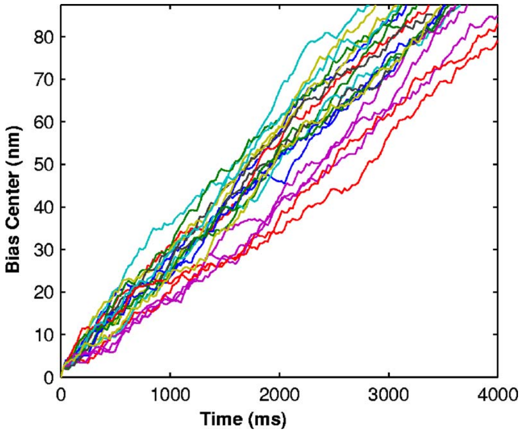

FIG. 4. (Color online) Twenty representative bias program from random force loading simulations.

Last, bias slicing is tested on a random force loading simulation. Every 25 ms, the force loading rate was changed to a uniformly distributed random number between −25 and 75 pN/s (Fig. 4), meaning that the bias center was allowed to move backwards. Accumulated work was numerically integrated by $W_t=\Sigma_{j=1}^{t}-k_s[(z_j+z_{j-1})/2-(b_j+b_{j-1})/2]\times(b_j-b_{j-1})$, where b_j is the position of the bias center at time step j. Again, the reconstructed free-energy surface using Eq. (4) is significantly more accurate than other calculations designed for quasistatic equilibrium conditions (Fig. 5).

IV. DISCUSSION AND SUMMARY

I have expounded on a method for reconstructing free-energy surfaces from single-molecule pulling experiments

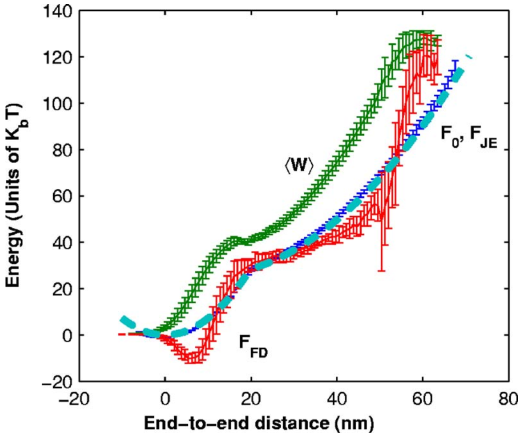

FIG. 5. (Color online) Average free-energy surfaces reconstructed from 10 blocks of 20 trajectories simulated with random force loading rates, annotated as in Fig. 1.

performed with different biasing protocols. The method was found to be valid for a range of cases, including near-equilibrium pulling, dynamic force loading, and random force loading. This paper opens the door for the application of transient fluctuation relations to the interpretation of a much broader array of biased experiments. A more careful efficiency and error analysis, as well as investigations into more complex systems, such as a spherical potential [30], alanine dipeptide [7], or a biological macromolecule [9], are left for future studies.

ACKNOWLEDGMENTS

I would like to thank R. Amaro, J. Gullingsrud, S. Bond, J. A. McCammon, R. D. Astumian, and K. Lindenberg, for helpful discussions. This work was funded by the NIH Molecular Biophysics Training Grant at UCSD. McCammon group resources are supported by the NSF (Grant No. MCB-0506593) and NIH (Grant No. GM31749), HHMI, CTBP, NBCR, W.M. Keck Foundation, and Accelrys, Inc.

[1] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[2] G. E. Crooks, Phys. Rev. E **61**, 2361 (2000).
[3] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **98**, 3658 (2001).
[4] G. Hummer and A. Szabo, Acc. Chem. Res. **38**, 504 (2005).
[5] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, Science **296**, 1832 (2002).
[6] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante, Nature (London) **437**, 231 (2005).
[7] S. Park, F. Khalili-araghi, E. Tajkhorshid, and K. Schulten, J. Chem. Phys. **119**, 3559 (2003).
[8] S. Park and K. Schulten, J. Chem. Phys. **120**, 5946 (2004).
[9] R. Amaro, E. Tajkhorshid, and Z. Luthey-Schulten, Proc. Natl. Acad. Sci. U.S.A. **100**, 7599 (2003).
[10] G. Hummer, J. Chem. Phys. **114**, 7330 (2001).
[11] I. Kosztin, B. Barz, and L. Janosi, J. Chem. Phys. **124** 064106(2006).
[12] E. Atilgan and S. Sun, J. Chem. Phys. **121**, 10392 (2004).
[13] C. Bennett, J. Comput. Phys. **22**, 245 (1976).
[14] P. Maragakis, M. Spichty, and M. Karplus, Phys. Rev. Lett. **96**, 100602 (2006).
[15] J. R. Gullingsrud, R. Braun, and K. Schulten, J. Comput. Phys. **151**, 190 (1999).
[16] C. Jarzynski, Phys. Rev. E **73**, 046105 (2006).
[17] O. Braun, A. Hanke, and U. Seifert, Phys. Rev. Lett. **93**, 158105 (2004).
[18] E. Evans and K. Ritchie, Biophys. J. **72**, 1541 (1997).
[19] E. Evans, Annu. Rev. Biophys. Biomol. Struct. **30**, 105 (2001).
[20] B. Heymann and H. Grubmuller, Phys. Rev. Lett. **84**, 6126 (2000).
[21] M. Manosas and R. Ritort, Biophys. J. **88**, 3224 (2005).
[22] Z. Liu, Y. Xu, and P. Tang, J. Phys. Chem. B **110**, 12789 (2006).
[23] M. Bayas, K. Schulten, and D. Leckband, Biophys. J. **84**, 2223 (2003).
[24] M. Gao, M. Wilmanns, and K. Schulten, Biophys. J. **83**, 3435 (2002).
[25] M. Marchi and P. Ballone, J. Chem. Phys. **110**, 3697 (1999).
[26] E. Paci, A. Caflisch, A. Pluckthun, and M. Karplus, J. Mol. Biol. **314**, 589 (2001).
[27] Y. M. Li, Z. G. Zhou, and C. B. Post, Proc. Natl. Acad. Sci. U.S.A. **102**, 7529 (2005).
[28] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
[29] J. Liphardt, B. Onoa, S. Smith, I. Tinoco, and C. Bustamante, Science **292**, 733 (2001).
[30] D. Bicout and A. Szabo, Protein Sci. **9**, 452 (2000).