<arttitle> Valid and efficient formula for free energy difference from nonequilibrium work

<aug> L. Y. Chen[1]

<aff> [1]Department of Physics, University of Texas, One UTSA Circle, San Antonio, Texas 78249, USA

<footnote>

<abs> Atomic force microscopes and optical tweezers afford direct probe into the inner workings of single biomolecules by mechanically unfolding them.[1-10] Critical to the success of this type of probe is to correctly extract the free energy differences between the various conformations of a protein/nucleic acid along its forced unfolding pathways.[11-15] Current studies rely on the Jarzynski equality[16] (JE) or its undergirding Crooks fluctuation theorem[17] (CFT), although questions exist on its validity[17-19] and on its accuracy/efficiency.[13,20-21] The validity of JE relies on the assumption of microscopic reversibility.[17,18] The dynamics of biomolecules, however, is Langevin stochastic in nature. The frictional force in the Langevin equation breaks the time reversal symmetry and renders the dynamics microscopically irreversible, even though detailed balance holds true. The inaccuracy of JE has largely been attributed to the fact that one cannot sample a large enough number of unfolding paths in a given study, experimental or computational.[13,15] Here I show that both of these questions can be answered with a new equation relating the nonequilibrium work to the equilibrium free energy difference. The validity of this new equation requires detailed balance, but not microscopic reversibility. Taking into the new equation equal numbers of unfolding and refolding paths, the accuracy is enhanced ten folds in comparison to a JE study based on a similar but larger number of unfolding paths.

<p> The beautifully simple Jarzynski equality[16] (JE)

<fd> $$\exp\left(-\frac{\Delta G}{k_B T}\right)=\left\langle \exp\left(-\frac{W_{A\to B}}{k_B T}\right)\right\rangle=\frac{1}{N_U}\sum_{p=1}^{N_U}\exp\left(-\frac{W_{A\to B}^{(p)}}{k_B T}\right) \qquad (1)$$

relates the free energy difference $\Delta G \equiv G_B - G_A$ between equilibrium states B and A to the nonequilibrium work $W_{A\to B}^{(p)}$ done to the system along the *p*-th of the N_U unfolding paths from A to B. Here k_B is the Boltzmann constant, T is the temperature, and the brackets represent the statistical average over the unfolding paths sampled in a given experimental or computational study. JE has been shown to be a consequence of the Crooks fluctuation theorem[17] (CFT)

<fd> $$\frac{P_U(W)}{P_R(-W)}=\exp\left(\frac{W-\Delta G}{k_B T}\right), \qquad (2)$$

which is valid when the dynamics of the system is microscopically reversible. Namely, for each unfolding path from A to B, one can sample a refolding path back from B to A that is exactly the inverse of the unfolding path. In that case, of course,

<fd> $$W_{A\to B}=W\,,\;\; W_{B\to A}=-W\,. \qquad (3)$$

Otherwise, CFT, Equation (2), and JE, Equation (1), do not hold. In experiments or simulations, however, this reversibility is not there unless the process is reversibly slow, so that the system is allowed to fully equilibrate at each step along the unfolding/refolding path.

<p> Without invoking the microscopic reversibility, required by JE and CFT, I show, in the Methods section, that the equilibrium free energy difference between State B and State A is related to the nonequilibrium work by the following equation:

<fd> $$\exp\left(-\frac{\Delta G}{k_B T}\right)=\frac{\left\langle \exp\left(-\frac{W_{A\to B}}{2k_B T}\right)\right\rangle_U}{\left\langle \exp\left(-\frac{W_{B\to A}}{2k_B T}\right)\right\rangle_R}=\frac{\frac{1}{N_U}\sum_{p=1}^{N_U}\exp\left(-\frac{W^{(p)}{}_{A\to B}}{2k_B T}\right)}{\frac{1}{N_R}\sum_{p=1}^{N_R}\exp\left(-\frac{W^{(p)}{}_{B\to A}}{2k_B T}\right)}\ . \qquad (4)$$

This simple relation is the central result of this Letter. The brackets in the numerator stand for the statistical average over the N_U unfolding paths sampled in a given experimental or computational study. The brackets in the denominator represent the same, but over the N_R refolding paths. $W^{(p)}$ is the work done to the system when it is unfolded/refolded along the *p*-th unfolding/refolding path.

<p> Several points should be noted here. First, Equation (4) is fundamentally different from Equation (1), the JE. The two only agree in a trivial case when the system is dominated by one unfolding path that is reversible. In such a limit, the free energy difference is expectedly equal to the reversible work done on the system. Second, the validity of Equation (4) does not rely on the assumption of microscopic reversibility. It just requires detailed balance between equilibrium states A and B. This will be clearly illustrated in its derivation in the Methods section. Third, if the nonequilibrium work $W_{A\to B}$ were Gaussian and so were $W_{B\to A}$, then Equation (4) would become $\Delta G=\left(\langle W_{A\to B}\rangle_U-\langle W_{B\to A}\rangle_R\right)/2$. This over-simple limit may not hold in real systems, but it sheds light on why Equation (4) is more accurate and more efficient than the JE. The dissipative work (energy spent to overcome friction) along the refolding path and the dissipative work along the unfolding path end up cancelling one another in Equation (4).[20,22-24] This cancellation works to enhance the efficiency of Equation (4). There are no such workings in the JE, Equation (1).

<p> To illustrate the efficiency/accuracy of Equation (4) in realistic applications, I have performed computer simulations of unfolding a deca-alanine poly-peptide, using the all-atom CHARMM 27 force fields[25] implemented within the NAMD[26] molecular

dynamics software package. This system is simple enough for which an exact solution is achievable by unfolding and refolding it at a reversibly slow speed, yet complex enough to serve as a prototype for studies of proteins and nucleic acids. Fig. 1 shows the work done to the poly-peptide when it is unfolded at a speed of 10^{-4} Å/ps along with the work done to the system when it is refolded at the same speed. The negligible difference between the two work curves suggests that unfolding/refolding at 10^{-4} Å/ps is indeed slow enough. The processes can be regarded as reversible and the work as a function of the end-to-end distance is a good approximation to the exact solution of free energy. Fig. 1 also shows the results of this research work and the results of JE in comparison with the exact solution. The free energy of this work is obtained through Equation (4) with 10 unfolding and 10 refolding paths at an irreversible speed of 10^{-2} Å/ps. The JE free energy is obtained through Equation (1) from 20 unfolding paths sampled at the same irreversible speed. In Ref.[13], Park et al give a thorough study of the same system based on JE. Sampling ten blocks of ten unfolding paths, they find that the deviation of the JE value of the free energy is 1.9 kcal/mol from the exact solution in the fully unfolded state with an end-to-end distance of 33.42 Å. They also estimate that the error bar is 1.6 kcal/mol. My simulations confirm these results. Park et al also employ the second and third order moments of work in place of the exponential average of JE. They conclude that the second order moment gives the best result and that the third order moment causes greater deviations from the exact solution and greater error bars, indicating that the nonequilibrium work distribution is not Gaussian.

<p> I have sampled 10 unfolding and 10 refolding paths at the irreversible speed of 10^{-2} Å/ps. These 20 paths are divided and arranged into four sets of five unfolding and five refolding paths. Fig. 2 shows one of the four sets of nonequilibrium work curves. The work curves along the refolding paths obviously do not overlap with their counterparts along the unfolding paths, evidencing the lack of reversibility required by the JE. Fig. 2 also shows the free energy computed through Equation (4) from these four sets of

unfolding and refolding paths. The free energy so obtained is found to be very close to the exact value (Fig. 1). The deviation from the exact value is found to be 0.1 kcal/mol in the fully unfolded state and the error bar is approximately 0.2 kcal/mol. Indeed, Equation (4) is ten folds more accurate than the JE, Equation (1), for deca-alanine.

<p> Based on the validity and efficiency of Equation (4), it is reasonable to expect that its applications will yield accurate estimations of free energy differences from nonequilibrium work for mechanical unfolding experiments and simulations of biomolecules, even when a limited number of unfolding and refolding paths can be sampled.

<meth1ttl> Methods

<meth1hd> In this section, I present a rigorous derivation of Equation (4) and show that its validity only requires detailed balance, far less restrictive a constraint on the dynamics of a system than the microscopic reversibility that is required by CFT and JE. Consider a biomolecule consisting of *N* atoms having *3N* degrees of freedom. *L* degrees of freedom will be controlled in an experimental or computational unfolding/refolding process from State A to State B. For example, deca-alanine has *N=104* atoms and *312* degrees of freedom. In the simulations discussed in the previous section, one terminal nitrogen atom is fixed while the other terminal nitrogen atom is pulled with a constant velocity *v*. For this case, *L*=6 degrees of freedom are controlled while the other *306* degrees of freedom are subjected to stochastic dynamics. In State A, the *L* coordinates of the two terminal atoms are fixed to one set of values, $x_l = x_l^A$ for $l = 1,2,\cdots L$. In State B, they are fixed to another set of values, $x_l = x_l^B$ for $l = 1,2,\cdots L$. The *3N-L* coordinates assume stochastic values according to the statistical weight factors determined by the energy and the entropy of the system. The probability for the system to be in the

macroscopic states A and B are related to its equilibrium free energy G_A and G_B as follows:[27]

<fd> $$p(A)=\frac{Z_A}{Z}=\frac{1}{Z}\exp\left(-\frac{G_A}{k_BT}\right)\ ,\quad p(B)=\frac{Z_B}{Z}=\frac{1}{Z}\exp\left(-\frac{G_B}{k_BT}\right). \tag{5}$$

In this equation, Z_A and Z_B are the partition functions of the system when it is constrained to the macroscopic states A and B respectively. Z is the total partition function of the system without any constraints.

<p> In equilibrium, the system satisfies the detailed balance[27,28] between the macroscopic states A and B. Namely,

<fd> $$p(A)P(A\to B)=p(B)P(B\to A)\ . \tag{6}$$

In this way, the equilibrium free energy difference between the macroscopic states A and B is related to the transition probabilities $P(A\to B)$ from A to B and $P(B\to A)$ from B to A as follows:

<fd> $$\exp\left(-\frac{\Delta G}{k_BT}\right)=\frac{P(A\to B)}{P(B\to A)}\ . \tag{7}$$

The transition probabilities are determined by the stochastic dynamics of the system that is governed by the Langevin equation,[28]

<fd> $$m_i\frac{d^2x_i}{dt^2}+m_i\gamma\frac{dx_i}{dt}+\frac{\partial}{\partial x_i}V=\xi_i\ . \tag{8}$$

Here m_i and x_i are the atomic mass and coordinate of the *i*-th degree of freedom respectively. γ is the damping (frictional) constant. V is the potential energy of the system that is a function of all coordinates. ξ_i is the stochastic force acting on the *i*-th degree of freedom. It is assumed to be Gaussian with the following characteristics:

<fd> $$\langle\xi_i(t)\rangle=0\ ,\quad \langle\xi_i(t)\xi_j(t')\rangle=2m_i\gamma k_BT\delta_{ij}\delta(t-t')\ . \tag{9}$$

Here δ_{ij} is the Kronecker delta and $\delta(t-t')$ is the Dirac delta function. For each sample of the stochastic force { ξ_i } generated according the following probability

<fd> $$P[\{\xi_i(t)\}] = \frac{1}{\Omega}\exp\left(-\frac{1}{2}\sum_{i=1}^{3N}\frac{1}{2m_i\gamma k_B T}\int_0^\tau dt[\xi_i(t)]^2\right) , \qquad (10)$$

the Langevin equation (8) produces one transition path { $x_i^{(p)}(t)$ } starting from one initial condition { $x_l^{(p)}(0) = x_l^A$ }(unfolding) or { $x_l^{(p)}(0) = x_l^B$ } (refolding). Ω is the normalization factor. Time *t* is taken to be in the interval of $(0,\tau)$. In terms of transition paths, the transition probabilities are[28]

<fd> $$p(A \to B) = \frac{1}{N_U}\sum_{p=1}^{N_U}\int[D\xi]\prod_{l=1}^{L}\delta\left(x_l^{(p)}(0) - x_l^A\right)P[\{\xi_i(t)\}]\prod_{l=1}^{L}\delta\left(x_l^{(p)}(\tau) - x_l^B\right), \qquad (11)$$

<fd> $$p(B \to A) = \frac{1}{N_R}\sum_{p=1}^{N_R}\int[D\xi]\prod_{l=1}^{L}\delta\left(x_l^{(p)}(0) - x_l^B\right)P[\{\xi_i(t)\}]\prod_{l=1}^{L}\delta\left(x_l^{(p)}(\tau) - x_l^A\right). \qquad (12)$$

Involved in the above equations are Wiener path integrals with noise (stochastic force) measure $[D\xi]$. Implicit are the integrals over the *3N-L* degrees of freedom.

<p> Now, for a controlled path of the *L* degrees of freedom, $\{x_l(t)\}$, that goes from $\{x_l(0) = x_l^A\}$ to $\{x_l(\tau) = x_l^B\}$, one can obtain the unfolding transition probability from Equation (11), using Equations (8) and (10),

<fd> $$p(A \to B) = \frac{1}{N_U}\sum_{p=1}^{N_U}\frac{1}{\Omega}\exp\left(-\sum_{l=1}^{L}\frac{1}{4m_l\gamma k_B T}\int_0^\tau dt\left[m_l\ddot{x}_l(t) + m_l\gamma\dot{x}_l(t) + \frac{\partial V}{\partial x_l}\right]^2\right). \qquad (13)$$

When the unfolding velocity is uniform, $\ddot{x}_l(t) = 0$. In the over-damped limit, $\gamma \to \infty$, the transition probability can be approximated as

<fd> $$p(A \to B) = \frac{1}{N_U}\sum_{p=1}^{N_U}\frac{1}{\Omega}\exp\left(-\frac{1}{2k_B T}\sum_{l=1}^{L}\int_0^\tau dt\dot{x}_l(t)\frac{\partial V}{\partial x_l}\right), \qquad (14)$$

which has no explicit dependence upon the damping parameter γ. Now it is obvious that the unfolding transition probability is related to the work done on the system:

<fd>

$$p(A \to B) = \frac{1}{N_U} \sum_{p=1}^{N_U} \frac{1}{\Omega} \exp\left(-\frac{W_{A \to B}^{(p)}}{2k_B T} \right) . \tag{15}$$

In the same manner, one can prove that the refolding transition probability is also related to the work done on the system:

<fd>

$$p(B \to A) = \frac{1}{N_R} \sum_{p=1}^{N_R} \frac{1}{\Omega} \exp\left(-\frac{W_{B \to A}^{(p)}}{2k_B T} \right) . \tag{16}$$

With Equations (7), (15), and (16), I complete the rigorous proof for the central result of this Letter in Equation (4).

<received>

<bibcit> 1. Liphardt, J., Onoa, B., Smith, S. B., Tinoco, I. Jr & Bustamante, C. Reversible unfolding of single RNA molecules by mechanical force. *Science* **292**, 733–737 (2001).

<bibcit> 2. Gerland, U., Bundschuh, R. & Hwa, T. Mechanically probing the folding pathway of single RNA molecules, *Biophys. J.* **84**, 2831–2840 (2003).

<bibcit> 3. Dietz, H. & Rief, M. Exploring the energy landscape of GFP by single-molecule mechanical experiments, *Proc. Natl. Acad. Sci. USA* **101**, 16192–16197 (2004).

<bibcit> 4. Fernandez, J. M. & Li, H. Force-Clamp Spectroscopy Monitors the Folding Trajectory of a Single Protein, *Science,* **303**, 1674 – 1678 (2004).

<bibcit> 5. Best, R.B. & Hummer, G. Comment on ‘Force-clamp spectroscopy monitors the folding trajectories of a single protein.’, *Science* **308**, 498b (2005).

<bibcit> 6. Cecconi, C., Shank, E.A., Bustamante, C. & Marqusee, S. Direct observation of the three-state folding of a single protein molecule, *Science* **309**, 2057–2060 (2005).

<bibcit> 7. Brockwell, D. J., Beddard, G.S., Paci, E., West, D. K., Olmsted, P.D., Smith, D.A. & Radford, S.E. Mechanically unfolding the small, topologically simple protein L, *Biophys. J.* **89**: 506-519 (2005).

<bibcit> 8. Hyeon, C. & Thirumalai, D. Forced-unfolding and force-quench refolding of RNA hairpins, *Biophys. J.* **90**, 3410–3427 (2006).

<bibcit> 9. Green, L., Kim, C.H., Bustamante, C. & Tinoco, I. Characterization of the mechanical unfolding of RNA pseudoknots, *J. Mol. Bio.* **375**, 511-528 (2008).

<bibcit> 10. Bornschlog, T. & Rief, M. Single-molecule dynamics of mechanical coiled-coil unzipping, *Langmuir*, **24**, 1338 -1342 (2008).

<bibcit> 11. Hummer, G. & Szabo, A. Free-energy reconstruction from nonequilibrium single molecule experiments. *Proc. Natl Acad. Sci. USA* **98**, 3658–3661 (2001).

<bibcit> 12. Liphardt, J., Dumont, S., Smith, S. B., Tinoco, I. Jr & Bustamante, C. Equilibrium information from nonequilibrium measurements in an experimental test of the Jarzynski equality. *Science* **296**, 1832–1835 (2002).

<bibcit> 13. Park, S., Khalili-Araghi, F., Tajkhorshid, E. & Schulten, K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality, *J. Chem. Phys.* **119**, 3559-3566 (2003).

<bibcit> 14. Collin, D., Ritort, F., Jarzynski, C., Smith, S.B., Tinoco, I. & Bustamante, C. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies, *Nature*, **437**, 231-234 (2005).

<bibcit> 15. Harris, N.C., Song, Y. & Kiang, C.H. Experimental free energy surface reconstruction from single-molecule force spectroscopy using Jarzynski's equality, *Phys. Rev. Lett.* **99**, 068101 (2007).

<bibcit> 16. Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **78**, 2690–2693 (1997).

<bibcit> 17. Crooks, G. E. Entropy production fluctuation theorem and the nonequilibrium work relation for free-energy differences. *Phys. Rev. E* **60**, 2721–2726 (1999).

<bibcit> 18. Cohen, E.G.D. & Mauzerall, D. The Jarzynski equality and the Boltzmann factor, *Mol. Phys.* **103**, 2923-2926 (2005).

<bibcit> 19. Vilar, J.M.G. & Rubi, J.M.F. Failure of the work-Hamiltonian connection for free-energy calculations, *Phys. Rev. Lett.* **100**, 020601 (2008).

<bibcit> 20. Kosztin, I., Barz, B. & Janosi, L. Calculating potentials of mean force and diffusion coefficients from nonequilibrium processes without Jarzynski's equality, *J. Chem. Phys.* **124**, 064106 (2006).

<bibcit> 21. Jarzynski, C. Rare events and the convergence of exponentially averaged work values, *Phys. Rev. E* **73**, 045101 (2006).

<bibcit> 22. Bennett, C. H. Efficient estimates of free energy differences from Monte Carlo data. *J. Comp. Phys.* **22**, 245–268 (1976).

<bibcit> 23. Shirts, M. R. & Pande, V. S. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration, *J. Chem. Phys.* **122**, 144107 (2005).

<bibcit> 24. Ytreberg, F.M., Swendsen, R.H.& Zuckerman, D.M. Comparison of free energy methods for molecular systems, *J. Chem. Phys.* **125**, 184114 (2006).

<bibcit> 25. MacKerell, A.D., Banavali, N. & Foloppe, N. Development and current status of the CHARMM force field for nucleic acids, *Biopolymers* **56**: 257-265 (2001).

<bibcit> 26. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L. & Schulten, K. Scalable molecular dynamics with NAMD, *J. Comp. Chem.* **26**, 1781-1802 (2005).

<bibcit> 27. Chandler, D. Introduction to Modern Statistical Mechanics, (Oxford Univ. Press, New York 1987).

<bibcit> 28. Risken, H. The Fokker-Planck Equation, (Springer-Verlag, New York 1989).

<supp>

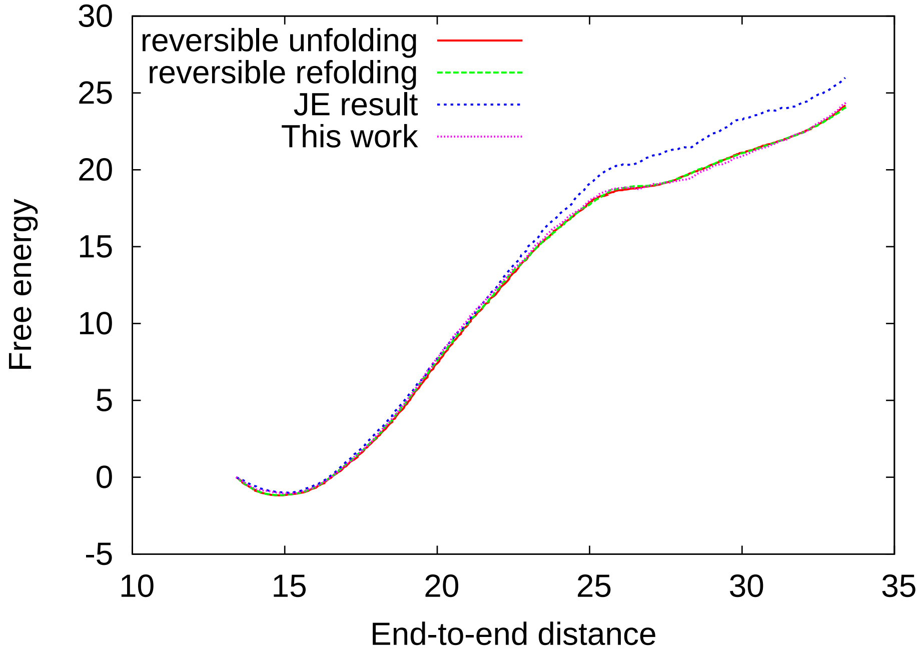
<ack> This work has been supported in part by an NIH grant, by a Faculty Research Award of UTSA, and by Texas Advanced Computing Centre with computing time. The author declares no competing financial interests.

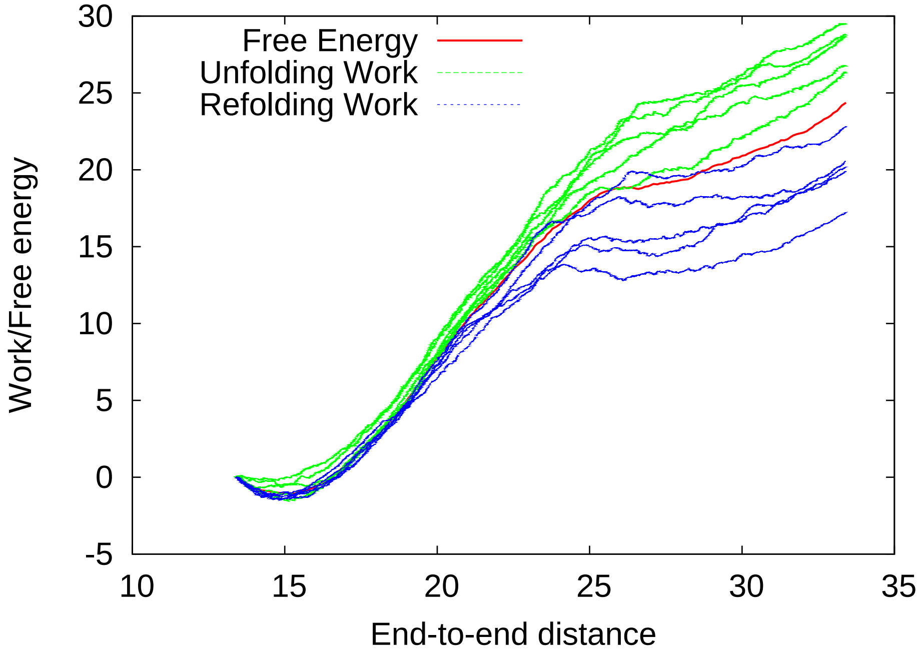
<corr> Correspondence and requests for materials should be addressed to L.Y.C. (Liao.Chen@utsa.edu).

<LEGEND> Fig. 1, Free energy (kcal/mol) as a function of the end-to-end distance (Å): The exact results are the red and green curves that overlap with one another completely. The result of this work is the pink curve that nearly overlaps with the exact results. The JE result is the blue curve. The zero point of the free energy is set to where the end-to-end distance is equal to 13.42 Å.

<LEGEND> Fig. 2, Free energy/nonequilibrium work (kcal/mol) as a function of the end-to-end distance (Å): The red curve is the free energy computed with Equation (4). Five green curves are the nonequilibrium work done along five unfolding paths ($W_{A\to B}$)and five blue curves are the nonequilibrium work done

along five refolding paths ($-W_{B\rightarrow A}$). The state when the end-to-end distance is 13.42 Å is chosen as the reference point for the free energy and the work.


30
25
20
15
10
5
0
-5
10
15
20
25
30
35
Free energy
End-to-end distance
reversible unfolding
reversible refolding
JE result
This work



Free Energy
Unfolding Work
Refolding Work
Work/Free energy
End-to-end distance
30
25
20
15
10
5
0
-5
10
15
20
25
30
35