Conformational Free-Energy Difference of a Miniprotein from Nonequilibrium Simulations

ARTICLE *in* JOURNAL OF PHYSICAL CHEMISTRY LETTERS · JULY 2010

Impact Factor: 7.46 · DOI: 10.1021/jz1005016

3 AUTHORS, INCLUDING:

Martin Spichty
Ecole normale supérieure de Lyon
31 PUBLICATIONS **684** CITATIONS
SEE PROFILE

Marco Cecchini
University of Strasbourg
44 PUBLICATIONS **1,099** CITATIONS
SEE PROFILE

THE JOURNAL OF PHYSICAL CHEMISTRY Letters

pubs.acs.org/JPCL

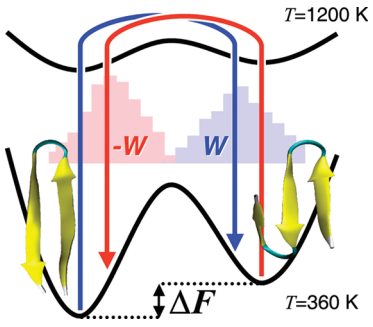
Conformational Free-Energy Difference of a Miniprotein from Nonequilibrium Simulations

Martin Spichty,*[,∥,§] Marco Cecchini,[∥] and Martin Karplus*[,∥,†]

[∥]Laboratoire de Chimie Biophysique, Institut de Science et d'Ingénierie Supramoléculaires, Université de Strasbourg, BP 70028, 8 allée Gaspard Monge, F-67083 Strasbourg Cedex, France, and [†]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge 02138, Massachusetts

ABSTRACT Conformational free-energy differences are essential thermodynamic quantities for understanding the function of many biomolecules. They are accessible from computer simulations, but their accurate calculation is a challenging task. Here nonequilibrium computer simulations and the differential fluctuation theorem are used to evaluate the free-energy difference between two conformational states of a structured miniprotein, the β-hairpin of protein G, with an implicit treatment of the solvent. A molecular dynamics-based protocol is employed for the simulation of rapid switches between the conformational states in both the forward and the reverse direction. From the work performed on the system in the individual switches, the conformational free-energy difference is determined by use of the differential fluctuation theorem. The results are in excellent agreement with reference calculations from a long molecular dynamics simulation and from the confinement method. The nonequilibrium approach is a computationally efficient method for the calculation of conformational free-energy differences for biological systems.

SECTION Biophysical Chemistry



As part of their function, many biomolecules undergo conformational transitions between structurally distinct states; examples include allosterically regulated proteins,[1,2] motor proteins,[3,4] chaperones,[5,6] and enzymes.[7] These states interconvert dynamically, and the probability of finding the system in a given conformation at equilibrium depends on its thermodynamic stability (free energy) relative to that of other conformations. Altering the relative stabilities by the binding of small molecules or by chemical reactions (e.g., ATP→ADP, P_i) in the binding site provides a mechanism for signaling[8] or doing work.[9] Thus, the knowledge of the difference in free energy among the various conformational states in the absence and presence of ligands is an essential aspect for understanding biomolecular function. Considerable effort has been devoted to the development of efficient methods for the calculation of such free-energy differences.[10] Recent examples are the "confinement" approach,[11,12] the "reference system" method,[13] "deactivated morphing",[14] "orthogonal space random walking",[15] as well as computational "single-molecule" approaches[16−26] that are based on nonequilibrium identities such as Jarzynski's equality[27] and Crooks' fluctuation theorem.[28] Recently, we derived the differential fluctuation theorem[21] (a generalization of Crooks' fluctuation theorem), with which free-energy differences can be estimated from conditional work data of nonequilibrium processes (switches); i.e., from the work performed on the system[27,29] in the course of switches that start and end in specific subensembles of the Hamiltonian system. In its original presentation, the differential fluctuation theorem was applied to the calculation of the free-energy difference between two conformations of the alanine dipeptide.[21] The present study reports an application to a much more realistic and challenging system; namely, the 16-residue β-hairpin substructure of the immunoglobin binding domain of streptococcal protein G, which is stable in the absence of the rest of the protein.[30] The β-hairpin is an excellent test case for free energy methods because it is a flexible system with many degrees of freedom that preserves the complexity of the free-energy landscape of a real protein. Also, its relatively small size makes it possible to determine the difference in conformational free energy between pairs of states from conventional equilibrium molecular dynamics (MD),[12] which provides a benchmark for the nonequilibrium results.

In this communication, we compare the conformational free energy of the native β-hairpin structure with that of the three-stranded β-sheet conformation (see Figure 1). Although these are both β-strand structures, there is a root-mean-square deviation (rmsd) of 7.1 Å between them (in terms of C_α and C_β atoms). Moreover, all of the interstrand hydrogen bonds are

Received Date: April 19, 2010
Accepted Date: May 27, 2010

DOI: 10.1021/jz1005016 | *J. Phys. Chem. Lett.* **2010**, *1*, 1922–1926

THE JOURNAL OF PHYSICAL CHEMISTRY Letters

pubs.acs.org/JPCL

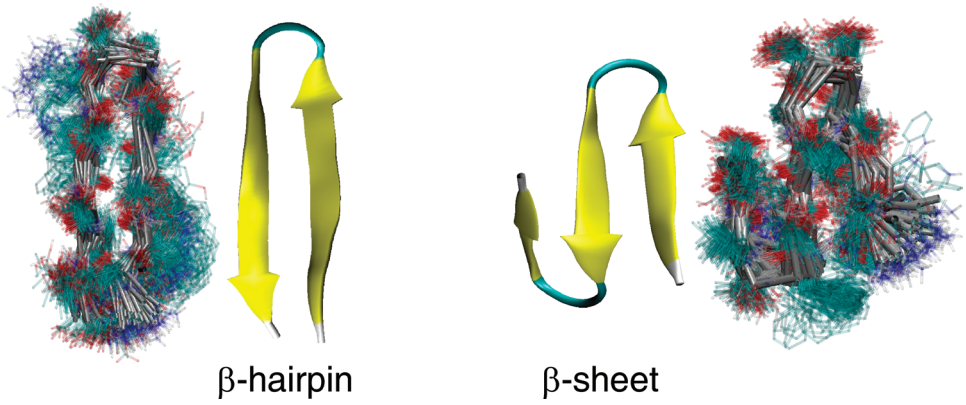


Figure 1. The conformational states of the 16-residue substructure of protein G considered in this communication. They represent the native β-hairpin state and a non-native three-stranded β-sheet state. The structures shown in sticks were sampled by MD at 360 K (see "Methods"); an overlay of 100 structures is illustrated for each conformational state. The ribbon structure (yellow) represents the average structure.

broken, and new ones are formed in the transition between the two states. The calculations were done at 360 K with an implicit treatment of the solvent so as to permit comparison with earlier work.[12,31] Under these conditions, the internal energy of the β-sheet state is lower that than that of the β-hairpin state; however, the entropic contribution to the free-energy difference makes the latter the native state (see Supporting Information).[31] From 200 μs of unbiased MD, a conformational free-energy difference of 1.89 ± 0.12 kcal/mol has been reported.[12] The high temperature and the long simulation time for the reference calculations were necessary because a free-energy barrier of about 6 kcal/mol separates the two conformational states.[31] Thus, spontaneous transitions between the β-hairpin (hereafter denoted with a) and the β-sheet (b) conformations are rare, even at 360 K; in 200 μs of MD, only 368 transitions were observed, that is, one transition every 0.5 μs on average.

Nonequilibrium switching simulations can make the transition about 250 times faster by perturbing the Hamiltonian function (and external conditions) during the simulation; i.e., barrier crossing is facilitated by a temperature increase followed by the application of an external force to switch from a to b and vice versa (see Figure 2a). The work (W) performed on the system during the forward and reverse switches is related to the equilibrium free-energy difference (ΔF) between a and b by the differential fluctuation theorem:[21]

$$P_F(W|a \rightarrow b)P_F(b|a)e^{-\beta W} = P_R(-W|b \rightarrow a)P_R(a|b)e^{-\beta \Delta F} \quad (1)$$

where $P_F(W|a \rightarrow b)$ is the probability of performing the work W in a successful forward switch, $P_R(-W|b \rightarrow a)$ is the corresponding probability for reverse switching, and $P_F(b|a)$ and $P_R(a|b)$ are the forward and reverse arrival probabilities, respectively, which are obtained from the fraction of switches that ended successfully; i.e., those that arrive in b (a), given that they started in a (b). We rewrite eq 1 as

$$P_F(W|a \rightarrow b)e^{-\beta W} = P_R(-W|b \rightarrow a)e^{-\beta \Delta F'} \quad (2)$$

with

$$\Delta F' = \Delta F + kT \ln[P_F(b|a)] - kT \ln[P_R(a|b)] \quad (3)$$

Equation 2 states that at the crossing point of the conditional forward and reverse probability distributions (normalized histograms) of the work, i.e., where $P_F(W|a \rightarrow b)$ equals $P_R(-W|b \rightarrow a)$, the work value corresponds to $\Delta F'$. The conditional work probability distributions obtained from about 1950 successful forward and reverse switches are shown in Figure 2b. The distributions have a large overlap and a value of $\Delta F' \sim$ 2 kcal/mol can be read directly from the crossing point of the histogram plots. A statistically robust estimate of $\Delta F'$ can be obtained by a maximum likelihood estimation (MLE), which solves the differential fluctuation theorem for the given set of work values from forward and reverse switches[32] and provides an analytical error estimate for $\Delta F'$ (see also refs 21 and 33). The result is $\Delta F' = 1.82 \pm 0.08$ kcal/mol. Among the 1963 forward switching attempts (starting from basin a) and 1952 reverse attempts (starting from basin b), 1925 and 1943 switches successfully arrived at the destination, respectively. Thus, the arrival probabilities are $P_F(b|a) = 0.981 \pm 0.003$ in the forward and $P_R(a|b) = 0.995 \pm 0.002$ in the reverse direction; the errors correspond to the analytical error of the binomial distribution (see Supporting Information). Introducing these values into eq 3 yields $\Delta F = 1.83 \pm 0.08$ kcal/mol. Since the arrival probabilities for forward and reverse switching are almost identical, the correction term $kT \ln [P_F(b|a)/P_R(a|b)]$ is negligible; we note that this is not always the case and that the correction term can contribute significantly to ΔF.[21] An alternative estimate of ΔF along with its statistical error was obtained from a block analysis (see Supporting Information). The result is identical to the MLE, which indicates that the there is no correlation in the work data and that the bias of the MLE is negligible.

The calculated value of ΔF is in excellent agreement with the value obtained from sampling at equilibrium (i.e., by conventional MD) and also with the free energy estimate of 1.88 ± 0.13 kcal/mol obtained from a confinement calculation of 4 μs.[12] The agreement validates the nonequilibrium result from the differential fluctuation theorem. The error is

DOI: 10.1021/jz1005016 | J. Phys. Chem. Lett. 2010, 1, 1922–1926

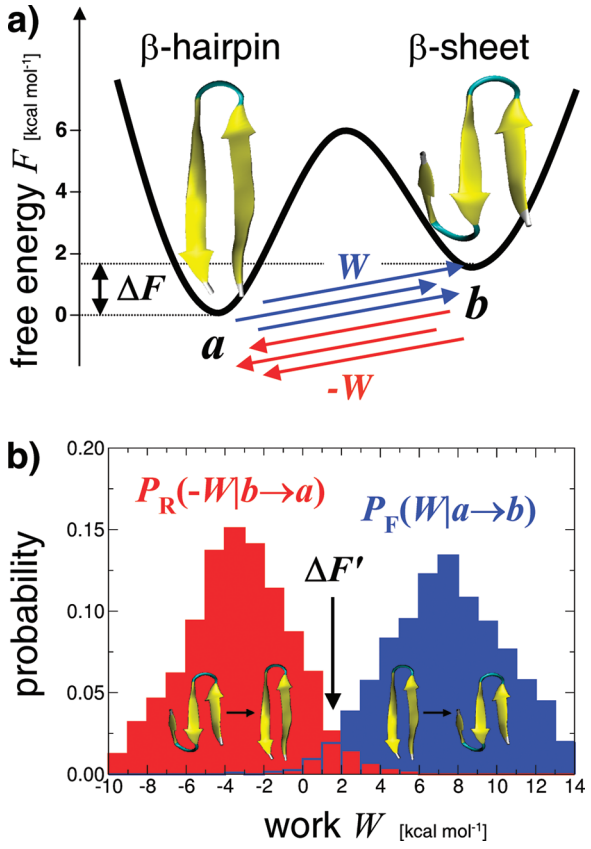


Figure 2. Conformational free-energy differences from nonequilibrium simulations. (a) Schematic representation of the free-energy profile of the 16-residue substructure of protein G. Starting from different equilibrium structures of the β-hairpin basin (denoted with *a*, left) we performed several independent, fast MD transitions (schematically indicated by the blue arrows) to the β-sheet basin (*b*, right). A specially designed, time-dependent Hamiltonian transformed the molecular structure from *a* to *b* in a short time interval of about 2 ns. The transitions in reverse direction (red arrows), i.e., starting from a random equilibrium structure of *b* and ending in a nonequilibrium ensemble of *a*, were performed in the same time interval, but the time-dependence of the switching protocol was reversed so that the structure was transformed from *b* to *a*. (b) Conditional probability distributions of the work. The graph shows normalized histograms of the work for the forward (blue) and reverse (red) switching direction constructed with a bin size of 1 kcal/mol. The arrow indicates the crossing point of the two distribution functions for which it is $W = \Delta F'$.

1.5 times smaller than the error obtained from conventional MD. Further, the nonequilibrium approach required about 25 times less CPU time and is therefore 50 times more efficient; the gain in efficiency drops to 30 if one includes the CPU time required for the optimization of the switching protocol. Interestingly, the performance of the nonequilibrium and the confinement approach are essentially identical.

In this letter, we have demonstrated that the nonequilibrium approach based on the recently developed differential fluctuation theorem[21] is an efficient method for calculating the conformational free energy between a pair of conformational substates of the β-hairpin peptide of protein G, a realistic system studied previously. We have shown that the nonequilibrium results are in excellent agreement with the free-energy estimate obtained from 200 μs MD but at a much lower computational cost. The comparison with the confinement method,[12] a recently introduced efficient technique,[11] demonstrates that the two approaches have similar performance. Thus, the nonequilibrium approach provides a valuable alternative to the calculation of conformational free-energy differences for biological systems. Bridging the size gap to larger biomolecules, as well as extending the scope of the method to explicit water simulations, are important challenges left for the future.

METHODS

Sampling of Equilibrium Ensembles. The equilibrium ensembles of the β-hairpin and the β-sheet basins were sampled at 360 K by two independent MD simulations of 200 ns in length. Molecular snapshots were saved every 100 ps along the trajectory. The sampling was focused on the basin of interest by adding a boundary potential centered on a molecular structure that was used as a reference (see Supporting Information). The reference structures chosen for basins *a* and *b* correspond to structures **1** and **2** of ref 31, which were identified as the most populated microstates of the corresponding free energy basins at 360 K. From the saved snapshots, all structures with a rmsd of the C_α and C_β atoms smaller than 2.5 Å from the reference were considered. Such a simple structure-based definition of the free-energy basin has been shown to be sufficient for the accurate determination of ΔF between these two β-hairpin states;[12] it is stable over a wide range of cutoffs. With these definitions, 1963 and 1952 structures (out of 2000) were obtained for the equilibrium ensembles of *a* and *b*, respectively.

Nonequilibrium Switches. Given two conformational ensembles at equilibrium (*a*, *b*), the nonequilibrium switches aimed to transform any microstate of *a* (the initial structure) into a microstate of *b* (the target structure) and vice versa (from *b* to *a*) by introducing an external biasing potential. In this study, a switching protocol was employed that made use of a high-temperature transformation before biasing the conformational transition; the high temperature was used to lower the barriers between the initial and final states so as to facilitate the transition.[34] Thus, the individual nonequilibrium switches were composed of an initial heating step, a transition part, and a final cooling step. In the actual implementation, a "slow" switching protocol was used; i.e., a linear interpolation was employed in which the Hamiltonian was adjusted every 1000 dynamic steps. In such a scheme, the fraction of steps used for heating and cooling (p), the temperature value employed to speed up the conformational change (T_{hot}), and the force constant of the biasing potential (f) are free parameters that can be optimized. Here, we used $p = 0.4$, $T_{hot} = 1200$ K, and $f = 8$ kcal/mol/Å^2. Details on the switching protocol and the determination of the parameters are given in the Supporting Information.

A forward (reverse) switch starting in basin *a* (*b*) was considered successful if the final structure was within 2.5 Å from the reference of basin *b* (*a*); the rmsd between the reference structures of the two basins is 7.1 Å. For each successful switch, the work performed on the system was

DOI: 10.1021/jz1005016 | J. Phys. Chem. Lett. 2010, 1, 1922–1926

calculated by summing up the contributions that resulted during the switch from incremental and instantaneous (fixed coordinates) changes of the generalized Hamiltonian, i.e., the potential energy function and the temperature. For details, see refs 29 and 35 and the Supporting Information.

Simulation System. The β-hairpin from protein G is a 16-residue peptide with the following amino-acid sequence: Gly-Glu-Trp-Thr-Tyr-Asp-Asp-Ala-Thr-Lys-Thr-Phe-Thr-Val-Thr-Glu.[30] The Hamiltonian of the peptide was modeled with a polar-hydrogen function,[36] whereas solvation effects were approximated by the solvent model EEF1,[37] which contains screened electrostatic interactions and a Gaussian term to represent the hydrophobic interactions. The simulations were performed in contact with a heat bath; we used Langevin dynamics and a friction coefficient of 4 ps^{-1} to control the temperature. SHAKE[38] constraints for bonds to hydrogens were used to allow for an integration time-step of 2 fs. All simulations were performed with the program CHARMM[39] (version c33b2); trajectory analyses were carried out with the program Wordom (version 0.20g).[40]

SUPPORTING INFORMATION AVAILABLE

Definitions and terminology and detailed descriptions of the equilibrium sampling simulations, switching protocol, calculation of the work, and error analysis. This material is available free of charge via the Internet at http://pubs.acs.org/.

AUTHOR INFORMATION

Corresponding Author:

*To whom correspondence should be addressed. E-mail: martin.spichty@ens-lyon.fr (M.S.); marci@tammy.harvard.edu(M.K.) Phone: +33 (0)-472-72-8052 (M.S.); (617)-495-4018 (M.K.). Fax: +33 (0)-472-72-8080 (M.S.); (617)-496-3204 (M.K.).

Present Addresses:

§ Current address: Ecole Normale Supérieure, Laboratoire de Biologie Moléculaire de la Cellule, Differentiation and Cell Cycle, 46 allée d'Italie, 69394 Lyon Cedex 7, France.

ACKNOWLEDGMENT We thank Sergei Krivov for valuable discussions, and Francesco Rao for the maintenance of the local computer cluster. This work was supported in part by computational resources from GENCI (Grand Equipement National de Calcul Intensif, France) and by a grant from the National Institutes of Health. M.C. was supported by the Human Frontier Science Program; M.S. thanks ISIS and the University of Strasbourg for hosting him as guest assistant professor.

REFERENCES

(1) Kern, D.; Zuiderweg, E. The Role of Dynamics in Allosteric Regulation. *Curr. Opin. Struct. Biol.* **2003,** *13*, 748–757.
(2) Cui, Q.; Karplus, M. Allostery and Cooperativity Revisited. *Protein Sci.* **2008,** *17*, 1295–1307.
(3) Rice, S.; Lin, A.; Safer, D.; Hart, C.; Naber, N.; Carragher, B.; Cain, S.; Pechatnikova, E.; Wilson-Kubalek, E.; Whittaker, M.; et al. A Structural Change in the Kinesin Motor Protein that Drives Motility. *Nature* **1999,** *402*, 778–784.
(4) Karplus, M.; Gao, Y. Biomolecular Motors: The F1-ATPase Paradigm. *Curr. Opin. Struct. Biol.* **2004,** *14*, 250–259.
(5) Ellis, J. Proteins as Molecular Chaperones. *Nature* **1987,** *328*, 378–379.
(6) van der Vaart, A.; Ma, J.; Karplus, M. The Unfolding Action of GroEL on a Protein Substrate. *Biophys. J.* **2004,** *87*, 562–573.
(7) Hammes, G. Mechanism of Enzyme Catalysis. *Nature* **1964,** *204*, 342–343.
(8) Bourne, H.; Sanders, D.; McCormick, F. The GTPase Superfamily: Conserved Structure and Molecular Mechanism. *Nature* **1991,** *349*, 117–127.
(9) Geeves, M.; Holmes, K. Strucrural Mechanism of Muscle Contraction. *Annu. Rev. Biochem.* **1999,** *68*, 687–728.
(10) Meirovitch, H. Recent Developments in Methodologies for Calculating the Entropy and Free Energy of Biological Systems by Computer Simulation. *Curr. Opin. Struct. Biol.* **2007,** *17*, 181–186.
(11) Tyka, M.; Clarke, A.; Sessions, R. An Efficient, Path-Independent Method for Free-Energy Calculations. *J. Phys. Chem. B* **2006,** *110*, 17212–17220.
(12) Cecchini, M.; Krivov, S.; Spichty, M.; Karplus, M. The Calculation of Free-Energy Differences between Peptide Conformers by Confinement Simulations. *J. Phys. Chem. B* **2009,** *113*, 9728–9740.
(13) Ytreberg, F.; Zuckerman, D. Simple Estimation of Absolute Free Energies for Biomolecules. *J. Chem. Phys.* **2006,** *124*, 104105.
(14) Park, S.; Lau, A.; Roux, B. Computing Conformational Free Energy by Deactivated Morphing. *J. Chem. Phys.* **2008,** *129*, 134102.
(15) Zheng, L.; Chen, M.; Yang, W. Random Walk in Orthogonal Space to Achieve Efficient Free-Energy Simulation of Complex Systems. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105*, 20227.
(16) Hummer, G.; Szabo, A. Free Energy Surfaces from Single-Molecule Force Spectroscopy. *Acc. Chem. Res.* **2005,** *38*, 504–513.
(17) Adib, A. Free Energy Surfaces from Nonequilibrium Processes without Work Measurement. *J. Chem. Phys.* **2006,** *124*, 144111.
(18) Oberhofer, H.; Dellago, C.; Boresch, S. Single Molecule Pulling with Large Time Steps. *Phys. Rev. E* **2007,** *75*, 061106.
(19) Paramore, S.; Ayton, G. S.; Voth, G. A. Extending the Fluctuation Theorem to Describe Reaction Coordinates. *J. Chem. Phys.* **2007,** *126*, 051102.
(20) Minh, D.; Adib, A. Optimized Free Energies from Bidirectional Single-Molecule Force Spectroscopy. *Phys. Rev. Lett.* **2008,** *100*, 180602.
(21) Maragakis, P.; Spichty, M.; Karplus, M. A Differential Fluctuation Theorem. *J. Phys. Chem. B* **2008,** *112*, 6168–6174.
(22) Luccioli, S.; Imparato, A.; Torcini, A. Free-Energy Landscape of Mechanically Unfolded Model Proteins: Extended Jarzinsky versus Inherent Structure Reconstruction. *Phys. Rev. E* **2008,** *78*, 031907.
(23) Calderon, C. P.; Janosi, L.; Kosztin, I. Using Stochastic Models Calibrated from Nanosecond Nonequilibrium Simulations to Approximate Mesoscale Information. *J. Chem. Phys.* **2009,** *130*, 144908.
(24) Junier, I.; Mossa, A.; Manosas, M.; Ritort, F. Recovery of Free Energy Branches in Single Molecule Experiments. *Phys. Rev. Lett.* **2009,** *102*, 070602.
(25) Nicolini, P.; Chelli, R. Improving Fast-Switching Free Energy Estimates by Dynamical Freezing. *Phys. Rev. E* **2009,** *80*, 041124.
(26) Ytreberg, F. M. Absolute FKBP Binding Affinities Obtained via Nonequilibrium Unbinding Simulations. *J. Chem. Phys.* **2009,** *130*, 164906.
(27) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997,** *78*, 2690–2693.

DOI: 10.1021/jz1005016 | *J. Phys. Chem. Lett.* **2010,** *1*, 1922–1926

(28) Crooks, G. E. Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences. *Phys. Rev. E* **1999,** *60*, 2721–2726.
(29) Crooks, G. Comment Regarding "On the Crooks Fluctuation Theorem and the Jarzynski Equality"[J. Chem. Phys. 129, 091101 (2008)] and "Nonequilibrium Fluctuation-Dissipation Theorem of Brownian Dynamics"[J. Chem. Phys. 129, 144113 (2008)]. *J. Chem. Phys.* **2009,** *130*, 107101.
(30) Gronenborn, A.; Filpula, D.; Essig, N.; Achari, A.; Whitlow, M.; Wingfield, P.; Clore, G. A Novel, Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein G. *Science* **1991,** *253*, 657–661.
(31) Krivov, S. V.; Karplus, M. Hidden Complexity of Free Energy Surfaces for Peptide (Protein) Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** *101*, 14766–14770.
(32) Shirts, M.; Bair, E.; Hooker, G.; Pande, V. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Phys. Rev. Lett.* **2003,** *91*, 140601.
(33) Bennett, C. H. Efficient Estimation of Free-Energy Differences from Monte-Carlo Data. *J. Comput. Phys.* **1976,** *22*, 245–268.
(34) Christen, M.; van Gunsteren, W. On Searching in, Sampling of, and Dynamically Moving through Conformational Space of Biomolecular Systems: A Review. *J. Comput. Chem.* **2007,** *29*, 157–166.
(35) Maragakis, P.; Spichty, M.; Karplus, M. Optimal Estimates of Free Energies from Multistate Nonequilibrium Work Data. *Phys. Rev. Lett.* **2006,** *96*, 100602.
(36) Neria, E.; Fischer, S.; Karplus, M. Simulation of Activation Free Energies in Molecular Systems. *J. Chem. Phys.* **1996,** *105*, 1902–1921.
(37) Lazaridis, T.; Karplus, M. Effective Energy Function for Proteins in Solution. *Proteins: Struct., Funct., Genet.* **1999,** *35*, 133–152.
(38) Ryckaert, J. P.; Ciccotti, G.; C., B. H. J. Numerical-Integration of Cartesian Equations of Motion of a System with Constrains. Molecular-Dynamics of *N*-alkanes. *J. Comput. Phys.* **1977,** *23*, 327–341.
(39) Brooks, B.; Brooks, C., III; Mackerell, A., Jr; Nilsson, L.; Petrella, R.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009,** *30*, 1545–1614.
(40) Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. Wordom: A Program for Efficient Analysis of Molecular Dynamics Simulations. *Bioinformatics* **2007,** *23*, 2625–2627.

DOI: 10.1021/jz1005016 | *J. Phys. Chem. Lett.* **2010,** *1*, 1922–1926