JCTC Journal of Chemical Theory and Computation

Adaptive Steered Molecular Dynamics of the Long-Distance Unfolding of Neuropeptide Y

Gungor Ozer,[†] Edward F. Valeev,[‡] Stephen Quirk,[§] and Rigoberto Hernandez*,[‡]

Center for Computational and Molecular Science and Technology, School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332-0400, Department of Chemistry, Virginia Tech, Blacksburg, Virginia 24061, and Kimberly-Clark Corporation, Atlanta, Georgia 30076-2199

Received June 11, 2010

Abstract: Neuropeptide Y (NPY) has been found to adopt two stable conformations in vivo: (1) a monomeric form called the PP-fold in which a polyproline tail is folded onto an α-helix via a β-turn and (2) a dimeric form of the unfolded proteins in which the α-helices interact with each other via side chains. The transition pathway and rates between the two conformations remain unknown and are important to the nature of the binding of the protein. Toward addressing this question, the present work suggests that the unfolding of the PP-fold is too slow to play a role in NPY monomeric binding unless the receptor catalyzes it to do so. Specifically, the dynamics and structural changes of the unfolding of a monomeric NPY protein have been investigated in this work. Temperature accelerated molecular dynamics (MD) simulations at 500 K under constant (N,V,E) conditions suggests a hinge-like unraveling of the tail rather than a random unfolding. The free energetics of the proposed unfolding pathway have been described using an adaptive steered MD (SMD) approach at various temperatures. This approach generalizes the use of Jarzynski's equality through a series of stages that allows for better convergence along nonlinear and long-distance pathways. Results acquired using this approach provide a potential of mean force (PMF) with narrower error bars and are consistent with some of the earlier reports on the qualitative behavior of NPY binding.

1. Introduction

The true nature of conformational changes undertaken by a given bioactive ligand during its binding to a receptor remains an elusive and important target for the development of novel drugs. The binding of a small ligand to a large membrane receptor is a dynamic process and is therefore difficult to observe using classical experimental approaches. Although atomic force microscopy (AFM) presents insightful information regarding the force required to unfold a particular protein, the detailed unfolding process is generally not observed in AFM experiments. Computer simulation techniques—to the limit of the accuracy of the model potentials and the integration of the equations of motion—provide a useful approach for elucidating the complete unfolding pathway of the protein of interest.[1−5] For example, in silico simulation using forced molecular dynamics agrees well with the corresponding AFM force pulling experiments.[6−9]

The neuropeptide Y (NPY) ligand has been a primary target of many recent pharmacological studies because of its implicated function in the brain.[10−14] Consisting of 36 amino acids, NPY is the most abundant neuropeptide in the mammalian central nervous system[10] and widely expressed in the peripheral nervous system.[11] Several important physiological activities such as induction and control of food intake, inhibition of anxiety, increase in memory retention, presynaptic inhibition of neurotransmitter release, vasoconstriction, and regulation of ethanol consumption have been attributed to NPY.[12] The multifunctionality of NPY is the

* To whom correspondence should be addressed E-mail: hernandez@chemistry.gatech.edu.

[†] Georgia Institute of Technology.

[‡] Virginia Polytechnic Institute and State University.

[§] Kimberly-Clark Corporation.

10.1021/ct100320g
Published on Web 09/14/2010

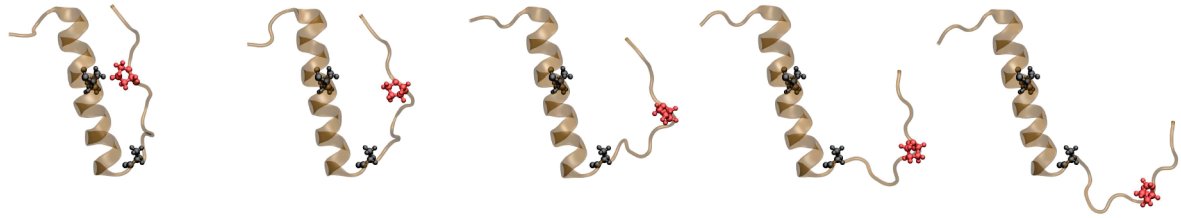

Figure 1. The structure of NPY (with the water solvent hidden) shown at 5 points along a single steered MD unfolding pathway at 500 K. The NPY backbone is illustrated using a brown ribbon. The first frame shows a stable folded NPY protein derived and equilibrated over the coordinates from the PDB file, 1PPT, and the last frame is an illustration of the unfolded protein (1RON). Both structures have the same 36 amino acid sequence: Y−P−S−K−P−D−N−P−G−E−D−A−P−A−E−D−M−A−R−Y−Y−S−A−L−R−H−Y−I−N−L−I−T−R−Q−R−Y. The first eight of these residues comprise the tail. The remaining frames illustrate the unhinging of the tail from the α helix as obtained from a particular unfolding trajectory in this work. The three residues most clearly marking the unhinging of the tail are shown in atomistic detail: LEU24 (in black on the helix), ALA12 (in black on the turn), and PRO5 (in red on the tail).

result of its affinity to bind to at least six receptor subtypes—enumerated as Y1 through Y6—belonging to the rhodopsin-like superfamily of G protein-coupled receptors. It has been shown that receptors Y1, Y4, and Y6 are closely related to each other.[13] A recent study on the evolution of neuropeptide Y receptors (Y3 was not investigated) has led to a partitioning into three subfamilies of receptors: Y1/Y4/Y6, Y2, and Y5.[14]

NPY is a member of the pancreatic polypeptide (PP) hormone family that includes also pancreatic polypeptide (PP) and peptide YY (PYY).[15] All three ligands share a common hairpin-like structure in tertiary form called the PP-fold. Therein, the N-terminal residues (1−8) adopt a polyproline type II helical conformation (tail). Residues 9−13 form a loop that allows the tail to fold onto an α-helix (residues 14−31), and the C-terminal residues (32−36) are so flexible that they do not participate in the α-helical conformation (14−31).[16−18] NMR studies have shown that NPY adopts a different conformation in the dimeric form[19−21] or when bound to membrane mimetic, dodecylphosphocholine (DPC) micelles.[22,23] In this particular state, the NPY tail is observed to be destabilized and positioned away from the α-helix. Recently, Bettio et al. reported, in contrast to earlier reports,[16−18] that at low concentrations monomeric NPY favors a less ordered structure in which the β-turn of NPY is more destabilized.[24]

The numerical study described herein aims to provide a dynamical explanation for the mechanism performed by an NPY molecule during its structural transition between the reported open (PDB[25] ID: 1PPT[26]) and closed (PDB ID: 1RON[21]) conformations. Figure 1 demonstrates a reduced representation of the unfolding of NPY using only five ribbon diagrams in order: the folded form (pp-fold), three intermediate structures, and an unfolded form. Knowledge of the pathway may be of use in the design of ligands to stimulate NPY toward the desired fold in vivo, regulators for the binding of NPY to lipid membranes, and alternative receptors. The present work, in particular, provides some insight into the likely form—PP-fold or free tail—adopted by NPY as it binds to a receptor. This article is structured as follows: High temperature MD simulations are used to accelerate the unfolding process and to observe a possible unfolding pathway for said process. The proposed unfolding pathway is investigated using steered molecular dynamics (SMD) simulations. The free energy along this path is generally obtained from the SMD trajectories through the use of Jarzynski's nonequilibrium work relation. Unfortunately, the standard application of this approach did not converge within available computational resources. An auxiliary central result of this work is the development of a stepwise adaptive SMD scheme for the calculation of the free energy along a nonlinear and large-distance pathway, in section 2.3. The time scale of the structural stability of NPY is obtained by way of a determination of the transition state theory rates on the computed surfaces. We observe that the NPY tail follows a hinge-like motion while folding/unfolding. We have also confirmed that the PP-fold conformation of NPY is favored in monomeric form, which was previously proposed by Nordmann et al.[16−18]

2. Methods

2.1. Accelerated MD Simulations at Elevated Temperature. The relative dynamics of the α-helix and tail in NPY immersed in a periodic box of water molecules have been simulated using several computational protocols to overcome the long times needed to follow simulations of the folding process. The focus of the simulations is the unfolding of NPY, as it is faster than the folding process while still revealing the folding pathway(s). The initial state of the unfolding process—namely, the protein's crystal structure—is also more clearly defined than the structures of the unfolded protein basin, and this offers additional numerical advantages in attempts to map out the pathway.[27]

Molecular dynamics simulations have been carried out using the NAMD[28] molecular dynamics integrator with the forces in NPY specified through the CHARMM force field.[29] The water molecules are treated using the TIP3P[30] model, and 13 178 water molecules are included in the cube. A time step of 1 fs has been employed in all simulations. Electrostatic interactions have been calculated through the particle mesh Ewald (PME) method.[31] Solvated structures are initialized by inserting NPY into an appropriately sized cavity created within an equilibrated neat water box. These are equilibrated at 50 K for 5 ps and subsequently heated gradually to the temperature of interest. An NPT equilibration run (at the desired final temperature) is then performed to ensure that the cubic box has a density consistent with 1.0

atm of pressure. Temperature control is realized within the NAMD program by integrating the Langevin equation with the Brunger−Brooks−Karplus (BBK) method, which is a natural extension of Verlet integration. This results in an ensemble of structures in which NPY is constrained to its folded state within an equilibrated solvent inside of a cubic box with sides roughly between 70 and 75 Å.

Each member of the ensemble of solvated folded-NPY structures is allowed to freely propagate for 5 ps under constant (N, V, E) conditions. It is common practice to run such simulations under constant (N, V, T) conditions using thermostats on all the atoms in the system. However, this has the possible negative side effect of suppressing fluctuations in energy that lead to correlated energy flows between molecules and therefore overly lose correlation as the system evolves in the heat bath.[32] For example, several popular MD packages, including NAMD, have recently been shown to suffer from a serious problem associated with the random number generators implemented in their thermostat algorithms.[33] In order to differentiate the unfolding mode from any other mode in the system, the alternative is to run the simulation under constant (N, V, E) conditions at an energy that is thermodynamically consistent with the temperature. This has the disadvantage that the total energy of the box is constant, but with a sufficiently large water box the effective dynamics of the NPY protein will still be that of an open system at constant temperature. The results from a small number of (N, V, T) and (N, V, E) MD simulations are described later, but the conclusion is that all of the remaining simulations could be performed using constant (N, V, E) conditions without losing the notion of temperature along the unfolding path.

Although we are primarily interested in the unfolding dynamics of NPY at 310 K, the duration of such trajectories is so long that it would entail simulations that are cost-prohibitive. Among several accelerated dynamics approaches now available in the literature, we chose to overcome this obstacle using temperature acceleration,[34] as it has been previously reported to accelerate the unfolding process without altering the pathway.[35] This is valid assuming that thermal unfolding of proteins demonstrates Arrhenius behavior.[5] However, some reports claim that protein unfolding can show non-Arrhenius behavior,[36,37] and therefore temperature acceleration may result in losing some intermediate states.[38,39] In the present context of NPY unfolding, the resulting potentials of mean force (shown below) contain only a single barrier at both low and high temperatures and hence are consistent with the requirement of Arrhenius behavior. Preliminary runs were tested at $T = 300$ K, 367 K, 433 K, and 500 K in a cubic box with sides of 75 Å solvated with equilibrated water (TIP3P) molecules. As will be shown below, NPY unfolded only at 500 K within 100 ps, and hence it became the temperature of choice for the temperature accelerated MD simulations in this work. A temperature of 500 K is well above any natural biological temperature and is also above the protein melting temperature, T_m. An experimental system under these conditions would exhibit different dynamics than the biological case. The water system in the computer model, however, remains as a metastable and superheated liquid because neither chemical bond breaking-and-making or evaporation pathways are available to it. The key assumption is that the dynamical pathways also remain in the same universality class, and thus we require additional tests to confirm the predictions of correlation functions using temperature acceleration. As will be shown below, the model system exhibits the appropriate chemical structures (in the same universality class) as those of the lower temperature.

2.2. Measurables and Correlation Functions of the Trajectories. Analysis of the trajectories was carried out by several methods. Both pepstat, which is our own code, and the NAMD/VMD package were used for trajectory analysis, with the latter focusing on the graphical representations of the trajectories.

Although the tail section exhibits the most dramatic dynamical changes, structural metrics were collected throughout the protein simulations. Within the polyproline tail (residues 1−12), the time dependence of the end-to-end distance and radius of gyration, $R_g^2 = 1/N \sum_{k=1}^{N}(r_k - r_{mean})^2$, are measured. The time dependence in the tail-to-helix distance is inferred by way of the pairwise distances between residue pairs, 1−31, 4−27, 5−24, 7−20, and 8−16.

The results shown below [cf. Figure 3b] suggest that the unfolding pathway involves the unhinging of the tail away from the α-helix instead of sliding. This unhinging occurs about the pivot represented by the ALA12 residue and is measurable through a so-called tail-turn-helix angle. While the α-helix is relatively stiff through this unfolding, the N-terminal of the polyproline tail—and particularly TYR1 to LYS4—is much floppier. The remaining residues (PRO5 to ASP11) on the tail follow a smoother unhinging and can be used to define the tail-turn-helix angle.

2.3. The Unfolding Path and the Potential of Mean Force (PMF). The domain of the energy landscape of even a small protein such as NPY has a high dimensionality. The identification of an unfolding pathway is therefore useful because it greatly reduces this dimensionality. Once identified, the energetics along this pathway are determined by the potential of mean force (PMF). [See, e.g., ref 40.] The importance of the PMF as well as the difficulty in calculating it has led to the development of far too many approaches to list here. Instead, we focus on those approaches which rely on sampling the states directly from trajectories. Unfortunately, the use of unconstrained trajectories is cost prohibitive when the processes of interest are very slow and dominated by deep minima. Instead, SMD can accelerate such processes by applying steering forces along the chosen unfolding pathway. Such a nonequilibrium process would not seem to provide the unconstrained structures required to obtain the equilibrium PMF. This problem was resolved by Jarzynski when he showed that an appropriately weighted average of the nonequilibrium work over many such SMD trajectories leads to the PMF.[41,42] Jarzynski's equality has been validated numerically on several systems such as deca-alanine stretching by Park and Schulten,[43] Ace−ALA_8−NMe unfolding and ligand diffusion in globins by Xionget et al.,[44] and Angeli's salt decomposition by Torras et al.[45] It has been compared to existing biased MD techniques, such as to umbrella

sampling[46] and to targeted MD,[47] yielding comparable results. It has also been verified in the context of experimental results such as RNA unfolding by Liphardt[48] and a mechanical oscillator.[49] Below, we provide a review of Jarzynski's inequality and its implementation within SMD to obtain the PMF along a selected pathway. In so doing, we also introduce a generalization of this approach to account for long-range nonlinear unfolding pathways.

2.3.1. Review of Jarzynski's Equality. Jarzynski's equality was originally expressed in terms of classical Hamiltonian systems.[41,42] It was extended to thermostatted stochastic systems by Crooks.[50] Crooks' introduction of a heat bath ensures that after sufficient time upon reaching a given nonequilibrium state, the system will reach an equilibrium with the environment at no additional cost of work. Jarzynski's equality for *dissipated* Hamiltonian systems can be stated as follows. Suppose a classical mechanical system consists of N particles, denoted by the phase space variables z, which are surrounded by a large enough heat bath. A constraint on the configuration space z_x is imposed through the projection $\xi_x = \xi_x(z_x)$ acting in configuration space alone. The constrained Hamiltonian may be written as

$$H_{\xi}^{\mathrm{SEB}}(\Gamma, \Theta) = H^{\mathrm{SE}}(z;\Theta_x) + H^{\mathrm{B}}(\Theta) \tag{1a}$$

$$= T^{\mathrm{S}}(\xi) + H_{\xi_x}^{\mathrm{E}}(\Gamma;\Theta_x) + H^{\mathrm{B}}(\Theta) \tag{1b}$$

where S, E and B denote the constrained system, environment, and bath, respectively; the subscript x (p) refers to the position (momentum) components; and T^{S} is the kinetic energy for the constrained system variables. The system variables not constrained by ξ—viz., the environment—comprise a space of dimension lower than $6N$, and its phase space variables are represented through Γ. The phase space variables Θ comprise the positions Θ_x and momenta Θ_p of the bath, and their dynamics are weakly coupled to Γ in the $H_{\xi_x}^{\mathrm{E}}$ term. The constraint ξ_x is typically one-dimensional and serves as an order parameter or reaction path that defines a state of the system. The space defined by Γ_x is orthogonal to ξ_x and denotes the environment exclusive of the bath Θ. The nonequilibrium process between two points in the constrained space is driven by the addition of a time-dependent Hamiltonian

$$H' = H'(\xi_x, t) \tag{2}$$

that acts only on ξ_x. That is, the total time-dependent Hamiltonian is $H^{\mathrm{T}} = H_{\xi}^{\mathrm{SEB}} + H'$. In what follows, we will not generally distinguish between the phase space ξ and configuration space ξ_x variables, for simplicity.

The change in the energy as the system is carried from an initial state ξ_0 to a final state ξ_t corresponds to the work done by H' (ξ, t) through this $\xi_t \leftarrow \xi_0$ process,

$$W^{\xi_t \leftarrow \xi_0}(\Gamma_t, \Theta_t, \Gamma_0, \Theta_0) = H_{\xi_t}^{E}(\Gamma_t;\Theta_t) - H_{\xi_0}^{E}(\Gamma_0;\Theta_0) \tag{3a}$$

$$= H'(\xi_t, t) - H'(\xi_0, 0) \tag{3b}$$

where Γ_t and Θ_t are connected to Γ_0 and Θ_0 through the propagator during the $\xi_t \leftarrow \xi_0$ process for a time t.

The equilibrium partition functions associated with the initial and final points associated with the $\xi \leftarrow 0$ process can be rewritten in terms of the original system variables as[51,52]

$$Z_{\xi}^{\mathrm{SE}} = \int \mathrm{d}z\, \mathrm{e}^{-\beta H^{\mathrm{SE}}(z)}\, \delta(\xi(z) - \xi) \tag{4a}$$

$$= \int \mathrm{d}z\, \mathrm{d}\Theta\, \mathrm{e}^{-\beta\{H_{\xi}^{\mathrm{SEB}}(\Gamma,\Theta) + H'(\xi_x,t)\}}\, \delta(\xi(z) - \xi) \tag{4b}$$

which is related to the potential of mean force, $G(\xi)$, through the reversible work theorem, $\ln Z_{\xi}^{\mathrm{S}} = -\beta G(\xi)$. In terms of these free energies, Jarzynski's equality,[41,42,50] is

$$G(\xi_t) = G(\xi_0) - \frac{1}{\beta}\ln\langle \mathrm{e}^{-\beta W^{\xi_t \leftarrow \xi_0}}\rangle_0 \tag{5}$$

where the ensemble average is taken over the initial variables (z, Θ) satisfying the constraint, $\xi(z_x) = \xi_0$. Note that, similar to the ground-state dominance in the calculation of a partition function, the Jarzynski average is dominated by the trajectories with the lowest work change.

Jarzynski's inequality follows from eq 5 through the use of Jensen's inequality:

$$G(\xi_t) - G(\xi_0) \le \langle W^{\xi_t \leftarrow \xi_0}\rangle_0 \tag{6}$$

Alternatively, the use of a cumulant expression provides the second-order cumulant (SOC) expression

$$G(\xi_t) - G(\xi_0) \approx \langle W^{\xi_t \leftarrow \xi_0}\rangle_0 - \frac{1}{2}\beta(\langle [W^{\xi_t \leftarrow \xi_0}]^2\rangle_0 - \langle W^{\xi_t \leftarrow \xi_0}\rangle_0^2) \tag{7}$$

which is surprisingly accurate for small nonequilibrium processes or environments with a Gaussian response.[43,53,54]

2.3.2. Adaptive Scheme for Jarzynski's Equality. As will be seen below, the application of the Jarzynski equality for the extended motion of a finite number of NPY unfolding trajectories provides a very weak upper bound to the PMF. In fact, it is so weak that the cumulant expansion of eq 5 presents a dramatically large deviation between the second-order cumulant and the exponential average, as demonstrated in Figure 8. In order to treat such extended systems, we have developed an adaptive version of Schulten's algorithm[43] in which the Jarzynski equality is applied through a series of shorter steps. It is *adaptive* in the sense that the initial configuration for a given step is obtained (or adapted) from the trajectories of the previous step.

The overall unfolding path is initially partitioned into N steps marked by its end points, ξ_0, ξ_1, ..., ξ_N. The ith iteration is initiated at ξ_{i-1} and Γ_{i-1} while the bath Θ_{i-1} is sampled from the appropriate canonical ensemble. Each such bath, $\Theta_{\alpha}^{\xi_i \leftarrow \xi_{i-1}}(t_{i-1})$, leads to M trajectories labeled by α for the $\xi_i \leftarrow \xi_{i-1}$ process. This, in turn, leads to a distribution of values in the work $W_{\alpha}^{\xi_i \leftarrow \xi_{i-1}}(t)$, environment $\Gamma_{\alpha}^{\xi_i \leftarrow \xi_{i-1}}(t)$, and bath $\Theta_{\alpha}^{\xi_i \leftarrow \xi_{i-1}}(t)$ for times t within the ith step. At the end of the iteration, the average work $W^{\xi_i \leftarrow \xi_{i-1}}(t_i)$ is computed according to the Jarzynski equality (eq 5). There then exists a trajectory α' for which its work $W_{\alpha'}^{\xi_i \leftarrow \xi_{i-1}}(t_i)$ is closest to the average work $W^{\xi_i \leftarrow \xi_{i-1}}(t_i)$. The initial value of the environment Γ_i for the $(i + i)$th iteration is then taken to be the corresponding $\Gamma_{\alpha'}^{\xi_i \leftarrow \xi_{i-1}}(t_i)$. Meanwhile the algorithm is initiated with values

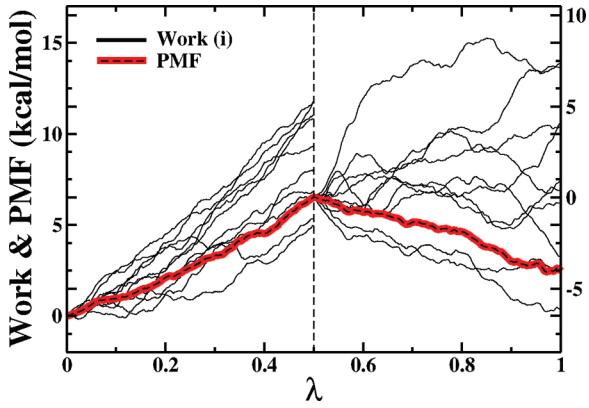

Figure 2. Illustration of the adaptive scheme applied to a system where the unfolding path is divided into two steps (with 10 trajectories α each). Black solid curves are the work for each of the 10 trajectories at each step. The PMF along a given substep is shown with a thick-red highlighted black-dashed cure. The right *y* axis tick marks are labeled for the second-step work trajectories with the 0 position located at the final average work value of the first substep. The left *y* axis tick marks are labeled for both the first step work trajectories and the overall PMF. (This figure is drawn for illustration purposes only and is not based on real physical data.)

(ξ_0, Γ_0) matching the initial structure of the system and environment. In the case of NPY, this amounts to the crystal structure of the entire protein, while ξ refers only to the constrained angle spanned by the helix and tail.

A proof of this algorithm begins by considering the application of the adaptive procedure to divide a single step into two substeps, as illustrated in Figure 2, along a specific unfolding path λ for the corresponding system variables ξ. For simplicity, but without a loss of generality, we suppose that the system is carried along by a nonequilibrium process from state $\xi = 0$ at initial time 0 to a final state $\xi = 1$ at a final time t. For each of M realizations labeled by α of the $1 \leftarrow 0$ process, the trajectories of the environment $\Gamma_\alpha^{1\leftarrow 0}(t)$ and the bath $\Theta_\alpha^{1\leftarrow 0}(t)$ can be formally constructed. The work done along each of these trajectories is $W_\alpha^{1\leftarrow 0}[\Gamma_\alpha^{1\leftarrow 0}(t), \Theta_\alpha^{1\leftarrow 0}(t), \Gamma_\alpha^{1\leftarrow 0}(0), \Theta_\alpha^{1\leftarrow 0}(0)]$ as specified by eq 3. The PMF of this process is

$$\Delta G^{1\leftarrow 0} = -\frac{1}{\beta}\ln\langle e^{-\beta W_\alpha^{1\leftarrow 0}}\rangle_\alpha \tag{8}$$

where the average is taken over the M realizations starting with the same initial ξ_0 and Γ_0 and various initial bath configurations $\Theta_\alpha(0^{1\leftarrow 0})$.

The single step can now be partitioned into two steps in which the system is stopped at an intermediate time t' and the corresponding position ξ'. For each of the original M trajectories in the $1 \leftarrow 0$ process, this partitions the work into two components:

$$W_\alpha^{\xi'\leftarrow 0} = H_{\xi'}^{\mathrm{SB}}[\Gamma_\alpha^{1\leftarrow 0}(t'), \Theta_\alpha^{1\leftarrow 0}(t')] - H_{\xi'}^{\mathrm{SB}}[\Gamma_\alpha^{1\leftarrow 0}(0), \Theta_\alpha^{1\leftarrow 0}(0)] \tag{9a}$$

$$W_\alpha^{1\leftarrow \xi'} = H_{\xi'}^{\mathrm{SB}}[\Gamma_\alpha^{1\leftarrow 0}(t), \Theta_\alpha^{1\leftarrow 0}(t)] - H_{\xi'}^{\mathrm{SB}}[\Gamma_\alpha^{1\leftarrow 0}(t'), \Theta_\alpha^{1\leftarrow 0}(t')] \tag{9b}$$

from which the free energy change for the first step can be easily obtained using Jarzynski's equality

$$\Delta G_{\xi'\leftarrow 0} = -\frac{1}{\beta}\ln\langle e^{-\beta W_{\xi'\leftarrow 0}^{\alpha}}\rangle_\alpha \tag{10}$$

For the second substep, however, each trajectory specified by eq 9b starts at a different value of the environment, $\Gamma_\alpha^{1\leftarrow 0}(t')$. We now introduce a $\xi'\leftarrow\xi'$ process during which ξ' is held fixed and the environment $\Theta_\alpha^{\tau}(t')$ relaxes in time τ from 0 to τ_α for some arbitrary final time τ_α, which is likely different for each trajectory α. The work to move the system from the state at the end of the process described in eq 9a along this $\xi'\leftarrow\xi'$ process is

$$\Delta W_\alpha^{\xi'\leftarrow\xi'} = H_{\xi'}^{\mathrm{SB}}[\Gamma_\alpha^{\tau_\alpha}(t'), \Theta_\alpha^{\tau_\alpha}(t')] - H_{\xi'}^{\mathrm{SB}}[\Gamma_\alpha^{\xi'\leftarrow 0}(t'), \Theta_\alpha^{\xi'\leftarrow 0}(t')] \tag{11}$$

and the work to return to the final point of the $1 \leftarrow 0$ process is

$$W''^{1\leftarrow\xi'}_\alpha = H_{\xi'}^{\mathrm{SB}}[\Gamma_\alpha^{1\leftarrow\xi'}(t), \Theta_\alpha^{1\leftarrow\xi'}(t)] - H_{\xi'}^{\mathrm{SB}}[\Gamma_\alpha^{\tau_\alpha}(t'), \Theta_\alpha^{\tau_\alpha}(t')] \tag{12}$$

The $\xi'\leftarrow\xi'$ process can be allowed to propagate for as long as it takes for $\Gamma_\alpha^{\tau_\alpha}(t')$ to be equal to some $\Gamma^{1\leftarrow\xi'}(t')$ which is independent of α. The existence of such a common end point is assured if the process is ergodic and the system is found in a single local basin of attraction. The requirement of ergodicity is a weak constraint given that the environment is coupled to a bath. The requirement for a single basin is also weak because the environment must access all possible such basins with zero-work paths. This motivates a new path for a *restricted* $1\leftarrow\xi'$ process starting at the fixed end point $\Gamma^{1\leftarrow\xi'}(t')$, and its work is given by

$$W''^{1\leftarrow\xi'}_\alpha = H_{\xi'}^{\mathrm{SB}}[\Gamma_\alpha^{1\leftarrow\xi'}(t), \Theta_\alpha^{1\leftarrow\xi'}(t)] - H_{\xi'}^{\mathrm{SB}}[\Gamma^{1\leftarrow\xi'}(t'), \Theta_\alpha^{1\leftarrow\xi'}(t')] \tag{13}$$

where the stochastic $\Theta_\alpha^{1\leftarrow\xi'}(t')$ has replaced the formally propagated $\Theta_\alpha^{\tau}(t')$. That is, the bath decoherence time is sufficiently fast so that the detailed propagation can be ignored while the initial bath $\Theta_\alpha^{1\leftarrow\xi'}(t')$ in the $1\leftarrow\xi'$ process is Gaussian random. The PMF of the restricted $1\leftarrow\xi'$ process is

$$\Delta G^{1\leftarrow\xi'} = -\frac{1}{\beta}\ln\langle e^{-\beta W''^{1\leftarrow\xi'}_\alpha}\rangle_\alpha \tag{14}$$

The average in eq 8 can thus be written as

$$\langle e^{-\beta W_\alpha^{1\leftarrow 0}}\rangle_\alpha = \langle e^{-\beta\{W''^{1\leftarrow\xi'}_\alpha + \Delta W_\alpha^{\xi'\leftarrow\xi'} + W_\alpha^{\xi'\leftarrow 0}\}}\rangle_\alpha \tag{15}$$

where it should be noted that the sum in the exponent in the RHS is not equal to $W_\alpha^{1\leftarrow 0}(t)$, nor is the trajectory the same after t'. However, the averages are equal because they are both nonequilibrium $1 \leftarrow 0$ processes between the same initial and final points satisfying Jarzynski's equality. Meanwhile, the work in the $\xi'\leftarrow\xi'$ process is

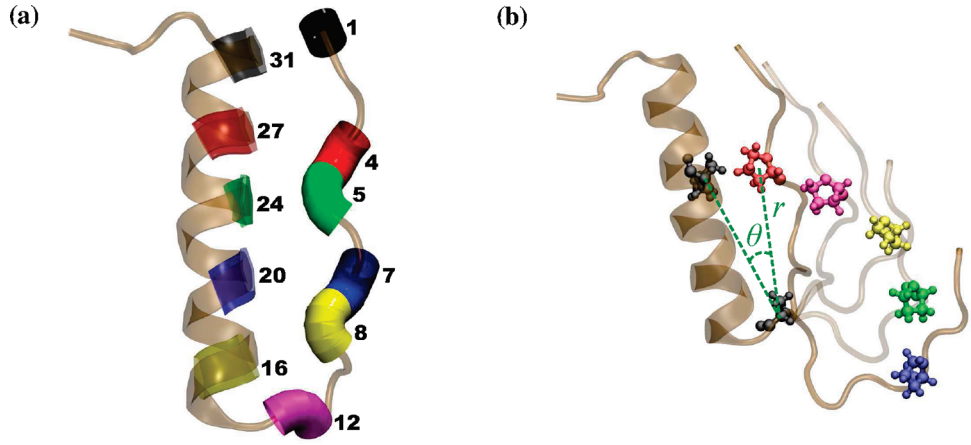

Figure 3. (a) A backbone ribbon diagram of NPY shown in brown with the helix emphasized by the thick ribbons as usual. The residue, ALA12, at the turn is shown in magenta and acts as the hinge. Pairs of residues that are in contact in the folded NPY and whose relative distances and angles are tracked in the following are color-coded as in the following scheme: black for TYR1 and ALA31, red for LYS4 and TYR27, green for PRO5 and LEU24, blue for ASN7 and TYR20, and yellow for PRO8 and ASP16. Note that residue positions 14−31 correspond to the helix. (b) The unfolding path is illustrated on the right, wherein the helix and hinge regions are held fixed while five images of the tail are overlaid. The PRO5 residue, which is explicitly used for steering relative to the fixed residues LEU24 and ALA12 (shown in black), is shown in five different colors along the unfolding path: red → magenta → yellow → green → blue.

zero because the system was allowed to relax freely. Hence,

$$\begin{aligned}\langle e^{-\beta W_\alpha^{1\leftarrow 0}}\rangle_\alpha &= \langle e^{-\beta\{W_\alpha''^{1\leftarrow\xi'}+W_\alpha^{\xi'\leftarrow 0}\}}\rangle_\alpha \\ &= \langle e^{-\beta W_\alpha''^{1\leftarrow\xi'}}\rangle_\alpha \times \langle e^{-\beta W_\alpha^{\xi'\leftarrow 0}}\rangle_\alpha\end{aligned} \tag{16}$$

The second equality follows from the fact that the trajectories in the $1\leftarrow\xi'$ and $\xi'\leftarrow 0$ processes are uncoupled and independently sampled. Combining eqs 8, 10, 14, and 16, we obtain the desired result:

$$\Delta G^{1\leftarrow 0} = \Delta G^{1\leftarrow\xi'} + \Delta G^{\xi'\leftarrow 0} \tag{17}$$

where the initial value of the environment $\Gamma^{1\leftarrow\xi'}(t')$ at the beginning of the $1\leftarrow\xi'$ process, in principle, can be chosen to be any arbitrary (but the same) state that is accessible to a $\xi'\leftarrow\xi'$ process. However, the choice of that intermediate state will affect the accuracy and convergence of the approach insofar as better choices would be more easily accessible and thus require less numerical relaxation in the evolution of the $1\leftarrow\xi'$ process. The best such choice is one that corresponds to a typical structure (not the minimum energy state) associated with the nonequilibrium process. To this end, we choose $\Gamma^{1\leftarrow\xi'}(t')$ according to the $\Gamma_\alpha^{1\leftarrow\xi'}(t')$ corresponding to the trajectory α, which minimizes the work difference, $|\Delta G^{\xi'\leftarrow 0} - W_\alpha^{\xi'\leftarrow 0}(t')|$.

Repeated application of eq 17 and the associated proscription for the choice of intermediate environment variables Γ for N steps gives rise to the desired final expression for the adaptive free energy difference:

$$\Delta G = \sum_{i=1}^{N} \Delta G^{i\leftarrow(i-1)} \tag{18}$$

where i labels the corresponding steps. In the limit that the "environment variables" are empty—i.e., that the dimensionality of the Γ space is zero—the adaptive procedure reduces to the use of the Jarzynski equality with the additivity trivially arising from the fact that the free energy is a state function.

In so far as the bath has been assumed to be Gaussian, the adaptive procedure should fail if the second-order cumulants in the work of a given set of trajectories begin to be nontrivial. As is shown below, the adaptive procedure does indeed satisfy this requirement.

2.3.3. Implementation of the Method. In this work, steered MD simulation is performed by pulling PRO5 at a constant velocity relative to the α helix on NPY. The choice of PRO5 is motivated both by experiment and computation. It has been previously reported that amino acids 1−4 of NPY (TYR1 to LYS4) form salt bridges with corresponding receptors.[55] Recent studies have indicated that binding hot spots at protein−protein interfaces exhibit high frequency fluctuation.[56] This suggests that the four residues from TYR1 to LYS4 of the NPY tail fluctuate faster than the other tail residues. Therefore, the choice of PRO5, rather than one of these other residues, allows us to drive the unfolding of the semirigid tail (including residues PRO5 to ASP11) while allowing the residues from TYR1 to LYS4 to fluctuate freely. Meanwhile the α helix must be represented by at least two fixed points so as to define the requisite hinge motion. These residues are LEU24 on the α helix and ALA12 on the hinge connecting the helix to the polyproline tail. The constrained system can therefore be designated through two variables: the LEU24−ALA12−PRO5 angle and the ALA12−PRO5 distance.

The external forces that carry the system along the unfolding path, $\xi(z_x)$, are imposed by way of a predefined potential $H'(\xi(z_x);\lambda)$. With the addition of this new potential, the extended time-dependent Hamiltonian, H^{ext}, becomes

$$H^{\text{ext}}(z,\Theta;\lambda) = H_\xi^{\text{SEB}}(\Gamma,\Theta) + H'(\xi(z_x);\lambda) \tag{19}$$

where

$$H'(\xi(z_x);\lambda) = \frac{1}{2}k[\xi_x(z_x) - \lambda]^2 \quad (20)$$

for a specified time-dependent process $\lambda(t)$. In the case of NPY, the pulling process is staged in N linear steps so as to approximate the circular unbinding process in a coordinate frame in which the center of mass of the hinge (ALA12) is the origin. The position of the center of mass of PRO5 $\vec{r}$ encodes the radius and angle of the system $\xi_x(z_x)$. Thus, for each step i, the auxiliary potential in eq 20 becomes

$$U_i(\vec{r}) = \frac{1}{2}k[\vec{r}(t) - (\vec{r}_i + v_i\vec{n}_i t)]^2 \quad (21)$$

where $\vec{r}_i$ is the position of PRO5 at the beginning of the interval, v_i is the velocity to move the particle to the end in the fixed time step, and $\vec{n}_i$ is the direction between the initial and final positions of PRO5. The position $\lambda_i \equiv (\vec{r}_i + v_i\vec{n}_i t)$ can be associated with an auxiliary particle (or dummy atom) that follows smoothly the prescribed unfolding path. As it does so, it exerts a work on the system ξ that is given by

$$\Delta W_i(t) = \int_{t_{i-1}}^{t} \vec{F}_i \cdot \vec{n}_i v \, dt \quad (22)$$

where the force $\vec{F}_i = -\nabla U_i(\xi_x(z_x))$ is related to the corresponding potential of eq 21. The corresponding free energy change, $\Delta G^{t\leftarrow 0}$, at time t within the ith interval can now be calculated using the adaptive work expression in eq 18, i.e.,

$$e^{-\beta\Delta G_{t\leftarrow 0}} = \langle e^{-\beta\Delta W_i(t)}\rangle_i \times \prod_{j=1}^{i-1}\langle e^{-\beta\Delta W_j(t_j)}\rangle_j \quad (23)$$

where the subscript on the angle brackets denotes the averaging over the trajectories in the corresponding interval.

2.4. Transition State Theory and Rates. The experimental results, unfortunately, do not provide a potential of mean force that can be used to compare directly to the computational work. Instead, we use the relative stability of the folded and unfolded states (as suggested by the calculated $\Delta G^{u\leftarrow f}$) to compare to the experimentally known stable structures. In addition, the rates of the unfolding and folding processes can be determined using transition state theory for the PMF determined along the unfolding path. These will be compared to the findings from both the molecular dynamics trajectories and experiment.

The simple transition state rate is

$$K = \frac{k_B T}{h} e^{-\frac{\Delta G^{\ddagger}}{k_B T}} \quad (24)$$

where $\Delta G^{\ddagger}$ is the free energy barrier of the transition. Although much work has been done to go beyond this simple estimate,[57−59] it is reasonably accurate for the order of magnitude of the rate.

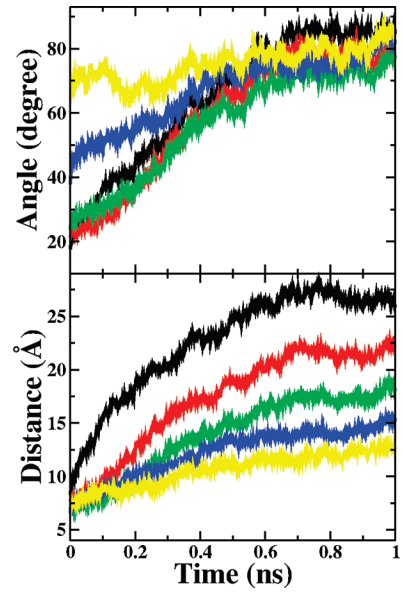

Figure 4. (Bottom) Average time-dependent displacements as NPY unfolds at 500 K, shown for several tail−helix residue pairs defined in Figure 3 using the same labeling scheme. (Top) Corresponding time-dependent angles spanned by a given pair of residues with respect to ALA12.

3. Results and Discussion

3.1. Identifying the Unfolding Path of NPY. Unfolding of NPY was first investigated through unconstrained MD simulations. MD trajectories were propagated using NAMD with the CHARMM force field in an explicit water solvent (TIP3P). At each of several temperatures, 300, 367, 433, and 500 K, 50 independent free MD simulations were integrated for 1 ns. At low temperatures, no unfolding was observed within the 1 ns observation window of the trajectories (not shown). At 500 K, all of the 50 generated trajectories unfolded in less than 1 ns.

Detailed analysis of the time dependence of the helix−tail separation in the 500 K unfolding trajectories reveals a hinge-like motion. The distance between the five pairs of residues initially in contact within the folded NPY are shown in the bottom panel of Figure 4. Pairs of residues farther from the turn (ALA12) move to more distant positions as the protein unfolds. All but the farthest residue pairs sweep a similar angle relative to the turn (ALA12), as shown in the top panel of Figure 4. This suggests that the tail hinges away from the helix about the turn during the unfolding process. It does not, however, follow this path linearly. The farthest residue pairs violate the quantitative agreement because the residues at the end of the tail are much more mobile and exhibit large fluctuations in position.

Both experimental[60,61] and computational[55] studies have suggested that residues 1−4 of the NPY tail form a pharmacophore that plays an active role during NPY binding to receptors. As postulated by Ertekin et al.,[56] interface residues that are in close contact with binding protein residues have a higher packing density and exhibit high frequency fluctuation.[56] The dynamics of the tail shown in Figure 5 are in good agreement with these previous reports. The part of the tail that is proximal to the hinge (including

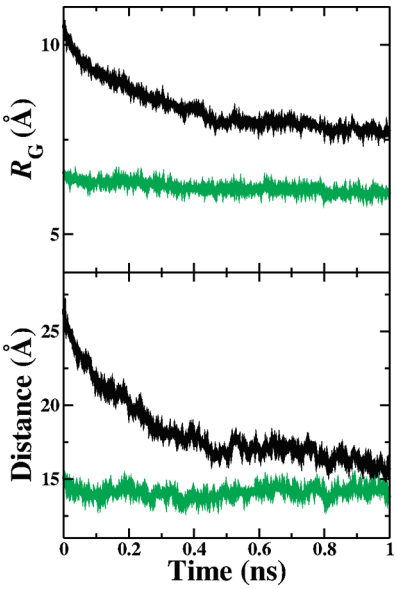

Figure 5. (Top and bottom) Radius of gyration and end-to-end displacement as a function of simulation time of two different segments of the NPY tail: ALA12 to TYR1 (black) and ALA12 to PRO5 (green). [cf. Figure 3b for the identification of the residues.]

all the residues up to PRO5) have nearly similar geometric properties (in terms of the length and radius of gyration) through the entire unfolding process. The rest of the tail, however, exhibits a significant geometric change through the unfolding process. It appears to be relaxation of the tail end toward a more compact structure in the vicinity of the proximal part of the tail.

The unfolding path thus appears to be primarily following the unhinging of the proximal part of the tail about the ALA12 hinge. Through this process, the tail appears to be nearly rigid up to PRO5, while the more distant residues are much more mobile. Hence, PRO5 is associated in the remainder of this work with the unfolding (reaction) path illustrated in Figure 4. Following Daggett and co-workers,[2] we therefore suppose that this unfolding path is followed not just at the elevated temperature of 500 K but also at experimentally accessible temperatures.

3.2. The PMF along the Unfolding Path. Our objective is to learn about the dynamics of NPY at temperatures relevant to the experimental systems. The temperature accelerated MD simulations provided us rates only at the locally stable temperature of 500 K. They also suggested an unfolding path along which we can calculate the PMF at lower temperatures for the purpose of obtaining relative rate information as will be done in the next subsection. The PMF must be calculated at 500 K for comparison with the MD simulations. For the lower temperature, we chose 310 K, as it is the body or *in vivo* temperature and is the temperature at which several experimental studies have explored the NPY dynamics.[17,18,22] The determination of the PMF at these two temperatures is nontrivial because the models are quite large (consisting of 40 123 atoms), for which a single nanosecond trajectory takes approximately 100 h on one computer core. Nevertheless, the nonequilibrium SMD approaches described in the previous section were used to obtain the PMFs. The

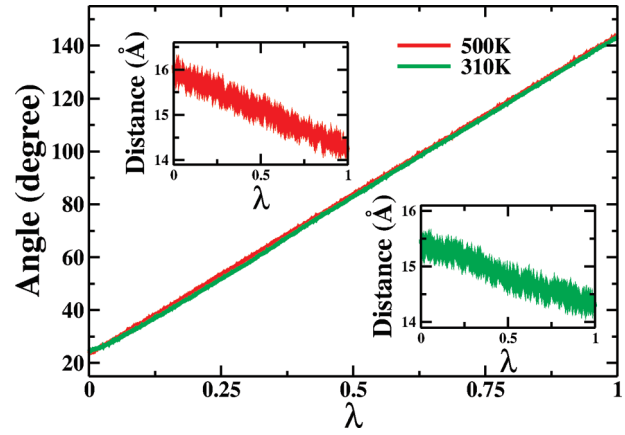

Figure 6. The average displacements in the system shown along the parametrized path λ for the adaptive SMD at 310 K and 500 K. The displacements $\xi_x = (r, \theta)$ fixed by the nonequilibrium process correspond to the radial distance r of PRO5 from the hinge (ALA12) and the angle θ spanned by the PRO5−ALA12 and LEU24−ALA12 vectors.

nonequilibrium simulations were realized using NAMD with the CHARMM force field for NPY in an explicit water solvent (TIP3P). All standard configuration parameters were the same as in the unconstrained MD simulations. The PMFs determined by either SMD approach required 110 h running on 48 2.33 GHz Intel 64 CPUs for 144 1-ns trajectories at a cost of 5280 CPU hours.

Steered MD trajectories have been obtained at a high temperature (500 K) as well as at body temperature (310 K). The unhinging of the tail was steered by pulling PRO5 (coupled to a dummy atom through a spring constant as per eq 20) relative to the virtually fixed residues ALA12 at the turn of the loop and LEU24 on the α helix. The unfolding path, which the dummy atom follows, is a discretization of the pseudocircular path shown in Figure 3b with each of the N finite steps taken to be linear. Specifically, the external force was applied on PRO5 to steer it from an initial configuration of the PRO5−ALA12−LEU24 angle θ_{initial} and radius r_{initial} to the final values, θ_{final} and r_{final}. At 500 K, $\theta_{\text{initial}} = 24.36°$ and $r_{\text{initial}} = 16.09$ Å. At 310 K, $\theta_{\text{initial}} = 24.41°$ and $r_{\text{initial}} = 15.49$ Å. At both temperatures, the final configuration is $\theta_{\text{final}} = 144.4°$ and $r_{\text{final}} = 14.3$ Å. The initial configurations for the two temperatures differ because each was prepared from equilibration runs at the respective temperatures. All control parameters, such as the pulling velocity ($v = 33$ Å/ns) and the spring constant ($k = 7.2$ kcal mol^{-1} Å^{-2}), were kept identical to each other so as to render comparable results. The degree to which the PRO5 residue followed the unfolding path through the SMD simulations is shown in Figure 6. On average, both θ and r follow the linear displacement well, as expected for a constant velocity pulling SMD simulation. The fluctuations around the average are small and also consistent with this conclusion.

At each temperature, 144 independent SMD trajectories were generated. (The number is 144, not 100, because of technical reasons related to the architecture of the particular computer cluster and the number of simultaneous trajectories—three—that could be run per core without increasing the wall clock time.) This number was sufficient to converge the

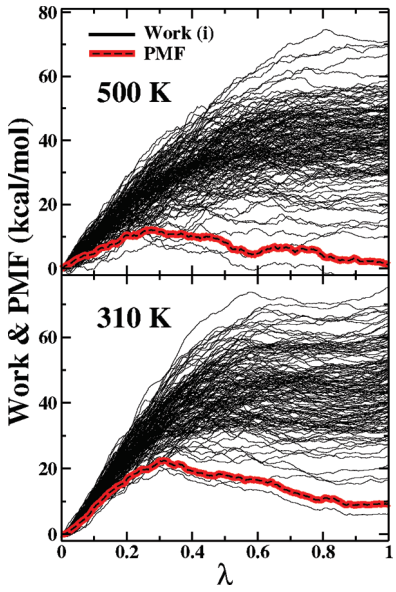

Figure 7. The work for 144 individual trajectories α (in black) and the PMF (in thick-red highlighting of a black-dashed curve) obtained using the Jarzynski equality displayed as a function of the parametrized unfolding path at 500 K (top panel) and 310 K (bottom panel).

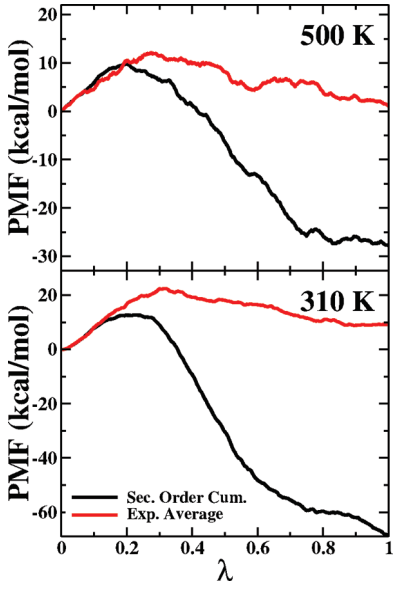

Figure 8. The PMF obtained using Jarzynski's equality (in red, cf. eq 5) and second-order cumulant expression (in black, cf. eq 7) obtained from a standard SMD calculation with 144 trajectories displayed as a function of the parametrized unfolding path at 500 K (top panel) and 310 K (bottom panel).

adaptive SMD trajectories and therefore serves as a good foil for the comparison of the two methods utilizing a similar amount of computational resources. Figure 7 shows the work and the averaged PMF using Jarzynski's relation at both 500 K (top) and 310 K (bottom). There are only a limited number of trajectories contributing to the PMF of the system at each temperature. This suggests a need for many more trajectories in order to converge the Jarzynski average. Indeed, the original deca-alanine in vacuum SMD PMFs calculated by

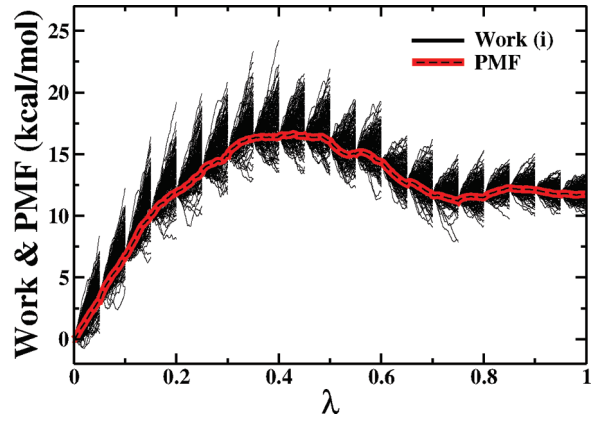

Figure 9. The work for 144 individual trajectories α (in black) and the PMF (in thick-red highlighting of a black-dashed curve) obtained using *adaptive* SMD displayed as a function of the parametrized unfolding path at 500 K.

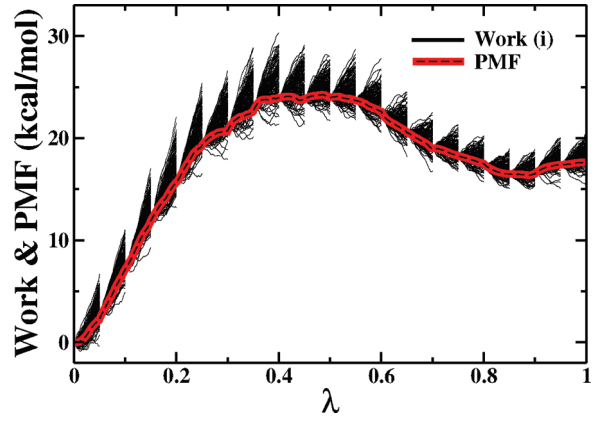

Figure 10. The work for 144 individual trajectories α (in black) and the PMF (in thick-red highlighting of a black-dashed curve) obtained using *adaptive* SMD displayed as a function of the parametrized unfolding path at 310K.

the Schulten group[43] required over 10 000 trajectories on this much smaller system.

The lack of convergence of this approach (using a limited number of trajectories) is also illustrated by the comparison of the PMF between Jarzynski's average and the second-order cumulant expression shown in Figure 8. The two expressions are equal in the limit that the work distribution is Gaussian because of the well-known Marcinkiewicz's theorem.[62] The lack of agreement between the two expressions is due both to the use of too few trajectories and also the fact that the observed trajectories were able to stray far from the relevant configurations. The consequence of the latter is that the statistics of the work contributions are far from Gaussian, and hence the second-order cumulant expression deviates greatly from Jarzynski's average.

The adaptive SMD method described in section 2.3 preempts the work distribution of such high barrier PMFs from losing their Gaussian nature by partitioning the unfolding path into several steps over which the PMF undergoes smaller changes. For the curved unfolding path illustrated in Figure 3b, we found convergence when we used 20 steps and a mere 144 trajectories per step. As noted earlier, the total computational cost is almost the same, excluding the

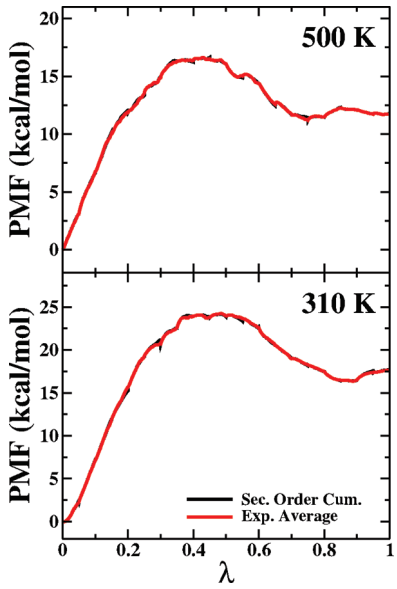

Figure 11. The PMF obtained using Jarzynski's equality (in red, cf. eq 5) and second-order cumulant expression (in lack, cf. eq 7) obtained from an *adaptive* SMD calculation with 144 trajectories displayed as a function of the parametrized unfolding path at 500 K (top panel) and 310 K (bottom panel). Note that the black curves are nearly entirely covered by the red curves and hence not very visible.

negligible cost required for trajectory comparison at the end of each step. As before, 144 independent adaptive SMD trajectories were generated for each of the two temperatures, 310 and 500 K.

The work and the averaged PMF using adaptive SMD (eq 18) are shown in Figures 9 (500 K) and 10 (310 K). Unlike in the results for the standard SMD simulations shown above, the PMFs are not dominated by the lowest energy trajectories. On the contrary, the PMF for each step has contributions from several trajectories. The results obtained for the PMF using the adaptive SMD method (cf. eq 18) with Jarzynski's equality (cf. eq 5) shown in Figures 9 and 10 are reproduced in Figure 11. Therein, the PMFs obtained with the second-order cumulant expression (cf. eq 7) are also shown. The agreement is remarkable, as the differences are not visible at this level of resolution. Though not shown, the number of sampled trajectories was doubled, leading to no significant change in the converged PMFs. Thus, the adaptive nonequilibrium process appears to result in a better estimate of the PMF with a limited number of trajectories, i.e., computational resources.

The PMFs in Figure 11 also provide information about the energetics of the unfolding process of NPY at the two temperatures, 310 and 500 K. The barrier heights to unfolding at 310 and 510 K are 24 and 17 kcal/mol, respectively. The 7 kcal/mol lowering of the barrier height at the higher temperature presumably results from the fact that the orthogonal degrees of freedom have a higher frequency at the folded state than at the barrier to unfolding. The difference is even more dramatic when one compares the ratio of the barrier height in units of k_BT. At 310 and 510 K, these ratios are 40 and 17, respectively, which suggests that the rates at 310 K are several orders of magnitude slower than at 510 K. As NPY had exhibited only partial unfolding at the high temperature, it is therefore not surprising that the low temperature MD simulations did not unfold within the 1 ns observation window. In addition, the folded state has a lower PMF and is therefore predicted to be the more stable form for monomeric NPY at 310 K.

3.3. The Folding and Unfolding Rates of NPY. The barrier height for the transition from the folded to unfolded conformations of NPY has been found to be 24 and 17 kcal/mol for 310 and 500 K, respectively. From these activation energy values, the rates have been calculated as 5.1×10^{-5} s^{-1} and 5.5×10^{5} s^{-1} again for 310 and 500 K, respectively. The inverse of these rates corresponds to a lifetime for the NPY unfolding transition. At the elevated temperature (500 K), this lifetime is 1.8 μs and is consistent with the fact that the NPY trajectories would explore the unfolded space within 1 ns, as seen in the MD simulations. At body temperature (310 K), this would suggest a lifetime of over 5 h, which is consistent with the fact that none of the low temperature NPY proteins unfolded during the MD simulations.

4. Conclusions

We have developed an adaptive algorithm extending the Schulten−Jarzynski SMD method for the calculation of PMFs when the subsystem is dragged across long nonlinear paths. In such cases, the PMF can span many k_BT's, leading to the sampling of nonequilibrium trajectories with work functions that fluctuate over a very large energy range. Consequently, only a small fraction of the trajectories generated from the SMD contribute nontrivially to the Jarzynski average. In order to numerically converge this average, one then needs to generate a large number of trajectories, which can be cost prohibitive. The adaptive algorithm allows one to break up the SMD calculation in a series of steps. The free energy difference across each such step is much smaller, and thereby allows convergence of the Jarzynski average with significantly fewer trajectories. In this sense, the adaptive algorithm is not formally better than the standard approach, but it is significantly more numerically efficient.

The SMD approach using the Jarzynski equality has been implemented experimentally in molecular force pulling experiments by several groups[48,49] with the underlying theory having been recently clarified by Zimanyi and Silbey.[52] The numerical success of the adaptive technique developed here could also be extended to such molecular force pulling experiments. Instead of using single constant velocity force pulling, the adaptive procedure would suggest the use of staged (or stepped) force pulling events. The pauses between the stages need only be held long enough so that the environment to the constrained system can relax (while applying zero work.)

The unfolding path of NPY has been suggested by temperature accelerated MD simulations to be the unhinging of the polyproline tail away from the α helix about the turn (near ALA12.) The NPY tail maintains its overall shape between PRO5 and ASP11 while unhinging away from the NPY helix. As the NPY unfolds along the path, the first four N-terminal residues (TYR1 to LYS4) fluctuate freely when

no biasing force is applied on them. This observation is consistent with earlier reports which hypothesized that these four residues on the polyproline tail of NPY form a pharmacophore at the NPY−receptor interface during NPY bioactivity.[55] This was earlier justified by the fact that protein−protein interfaces have been seen to be enriched by the presence of high frequency fluctuating residues.[56] The observation that this unfolding process can be reduced to a dominant one-dimensional pathway is not uncommon. Several groups[63,64] have reported the possibility of such a simplication if the dynamics are dominated by passage across a single barrier, as we have seen here for NPY.

The potentials of mean force along the folding path provide a more detailed view of the dynamics. This was possible because of a generalization of SMD (also known as force-biased simulations) using the adaptive scheme introduced in this work. The barrier heights and associated rates of the NPY unfolding transition at an elevated temperature (500 K) and the *in vivo* temperature (310 K) agree well with the numerical MD simulations (reported here) and those authors[16−18] which have proposed the stability of PP-fold on the basis of their experimental findings. At the *in vivo* temperature, we have determined an unfolding rate for NPY on a time scale longer than 5 h. The typical single-domain protein folding/unfolding time scale is a few microseconds at the fastest and a couple hundred microseconds at the slowest.[65] We thus conclude that at 310 K monomeric NPY does not unfold. This conclusion is consistent with our preliminary unconstrained MD simulations in which NPY did not unfold at temperatures up to 433 K. The fact that the unfolded NPY state has a higher free energy than the folded structure also suggests that the NPY monomer in solution is folded in the pancreatic−polypeptide (PP) fold. This result is also consistent with the experimental hypothesis that the NPY dimer is biologically inactive in solution because the tail moves away from the PP-fold.[18] This indirectly suggests that the biological activity of the NPY monomer results from the stability of the folded structure in agreement with the energetic stability found in this work. Recently, Bader et al.[22] reported that the micelle-bound form of NPY demonstrates a less ordered conformation than the PP-fold. In this less-ordered conformation, the NPY tail is observed to be fluctuating (Figure 3 in Bader et al.) while the α-helix remains stable. Our results suggest that this is due to the specific contacts, formed between micelle and side chains of the NPY α-helix, replacing the favorable polyproline tail and α-helix contacts observed in the PP-fold.

Acknowledgment. This work has been partially supported by the National Science Foundation (NSF) through Grant No. CHE 0749580. The computing resources necessary for this research were provided in part by the National Science Foundation through TeraGrid resources provided by the Purdue Dell PowerEdge Linux Cluster (Steele) under grant number TG-CTS090079, and by the Center for Computational Molecular Science & Technology through Grant No. CHE 0946869.

References

(1) Galzitskaya, O. V.; Finkelstein, A. V. A theoretical search for folding/unfolding nuclei in three-dimensional protein structure. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11299–11304.

(2) Best, R. B.; Li, B.; Steward, A.; Daggett, V.; Clarke, J. Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation. *Biophys. J.* **2001**, *81*, 2344–2356.

(3) Ng, S. P.; Rounsevell, R. W. S.; Steward, A.; Geierhaas, C. D.; Williams, P. M.; Paci, E.; Clarke, J. Mechanical unfolding of TNfn3: The unfolding pathway of a fnIII domain probed by protein engineering, AFM and MD simulation. *J. Mol. Biol.* **2005**, *350*, 776–789.

(4) Hisatomi, Y.; Katagiri, D.; Neya, S.; Hara, M.; Hoshino, T. Analysis of the unfolding process of green fluorescent protein by molecular dynamics simulation. *J. Phys. Chem. B* **2008**, *112* (29), 8672–8680.

(5) Mayor, U.; Johnson, C. M.; Daggett, V.; Fersht, A. R. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13518–13522.

(6) Lu, H.; Isralewitz, B.; Krammer, A.; Vogel, V.; Schulten, K. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.* **1998**, *75*, 662–671.

(7) Grater, F.; Shen, J.; Jiang, H.; Gautel, M.; Grubmuller, H. Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations. *Biophys. J.* **2005**, *88*, 790–804.

(8) Gao, M.; Craig, D.; Vogel, V.; Schulten, K. Identifying unfolding intermediates of FN-III10 by steered molecular dynamics. *J. Mol. Biol.* **2002**, *323*, 939–950.

(9) Lu, H.; Schulten, K. Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins* **1999**, *35*, 453–463.

(10) Gray, T.; Morley, J. Neuropeptide Y: Anatomical distribution and possible function in mammalian nervous system. *Life Sci.* **1986**, *38*, 389–401.

(11) Dumont, Y.; Fournier, A.; Quirion, R. Neuropeptide Y and neuropeptide Y receptor subtypes in brain and peripheral tissues. *Progr. Neurobiol.* **1992**, *38*, 125–167.

(12) Turton, M.; O'Shea, D.; Bloom, S. *Central effects of neuropeptide Y with emphasis on its role in obesity and diabetes*; Academic Press: San Diego, CA, 1997.

(13) Larhammar, D. Structural diversity of receptors for neuropeptide Y, peptide YY and pancreatic polypeptide. *Regul. Pept.* **1996**, *65*, 165–174.

(14) Wraith, A.; Tornsten, A.; Chardon, P.; Harbitz, I.; Chowdhary, B. P.; Andersson, L.; Lundin, L.-G.; Larhammar, D. Evolution of the neuropeptide Y receptor family: Gene and Chromosome duplications deduced from the cloning and mapping of the five receptor subtype genes in pig. *Genome Res.* **2000**, *10*, 302–310.

(15) Larhammar, D. Evolution of neuropeptide Y, peptide YY, and pancreatic polypeptide. *Regul. Pept.* **1996**, *62*, 1–11.

(16) Li, X.; Sutcliffe, M. J.; Schwartz, T. W.; Dobson, C. M. Sequence-specific proton NMR assignments and solution structure of bovine pancreatic polypeptide. *Biochemistry* **1992**, *31*, 1245–1253.

(17) Darbon, H.; Bernassau, J.; Deleuze, C.; Chenu, J.; Roussel, A.; Cambillau, C. Solution conformation of human neuropeptide Y by 1H nuclear magnetic resonance and restrained molecular dynamics. *Eur. J. Biochem.* **1992**, *209*, 765–771.

(18) Nordmann, A.; Blommers, M.; Fretz, H.; Arvinte, T.; Drake, F. Aspects of the molecular structure and dynamics of neuropeptide Y. *Eur. J. Biochem.* **1999**, *261*, 216–226.

(19) Cowley, D.; Hoflack, J.; Pelton, J.; Saudek, V. Structure of neuropeptide Y dimer in solution. *Eur. J. Biochem.* **1992**, *205*, 1099–1106.

(20) Mierke, D.; Durr, H.; Kessler, H.; Jung, G. Neuropeptide Y: Optimized solid-phase synthesis and conformational analysis in trifluoroethanol. *Eur. J. Biochem.* **1992**, *206*, 39–48.

(21) Monks, S.; Karagianis, G.; Howlett, G.; Norton, G. Solution structure of human neuropeptide Y. *J. Biomol. NMR* **1996**, *8*, 379–390.

(22) Bader, R.; Bettio, A.; Beck-Sickinger, A. G.; Zerbe, O. Structure and dynamics of micelle-bound neuropeptide Y: Comparison with unligated NPY and implications for receptor selection. *Genome Res.* **2000**, *10*, 302–310.

(23) Lerch, M.; Mayrhofer, M.; Zerbe, O. Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors. *J. Mol. Biol.* **2004**, *339*, 1153–1168.

(24) Bettio, A.; Dinger, M. C.; Beck-Sickinger, A. G. The neuropeptide Y monomer in solution is not folded in the pancreatic-polypeptide fold. *Protein Sci.* **2002**, *11*, 1834–1844.

(25) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(26) Blundell, T. L.; Pitts, J. E.; Tickle, I. J.; Wood, S. P.; Wu, C.-W. X-ray analysis (1.4-A resolution) of avian pancreatic polypeptide: Small globular protein hormone. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4175–4179.

(27) Daggett, V.; Fersht, A. R. Is there a unifying mechanism for protein folding. *Trends Biochem. Sci.* **2003**, *28*, 18–25.

(28) 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. Comput. Chem.* **2005**, *28*, 1781–1802.

(29) Brooks, B.; Bruccoleri, R.; Olafson, R.; States, D.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

(30) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(31) Park, P. J.; Lee, S. Particle mesh Ewaldi: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(32) Nose, S. Constant-temperature molecular dynamics. *J. Phys.: Condens. Matter* **1990**, *2*, SA115–SA119.

(33) Cerutti, D. S.; Duke, R.; Freddolino, P. L.; Fan, H.; Lybrand, T. P. A Vulnerability in Popular Molecular Dynamics Packages Concerning Langevin and Andersen Dynamics. *J. Chem. Theory Comput.* **1999**, *4*, 1669–1680.

(34) Okur, A.; Roe, D. R.; Cui, G.; Hornak, V.; Simmerling, C. Improving Convergence of Replica-Exchange Simulations through Coupling to a High-Temperature Structure Reservoir. *J. Chem. Theory Comput.* **2007**, *3*, 557–568.

(35) Day, R.; Daggett, V. Increasing Temperature Accelerates Protein Unfolding Without Changing the Pathway of Unfolding. *J. Mol. Biol.* **2002**, *322*, 189–203.

(36) Chan, H. S.; Dill, K. A. Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. *Proteins* **1998**, *30*, 2–33.

(37) Matagne, A.; Jamin, M.; Chung, E. E.; Robinson, C. V.; Radford, S. E.; Dobson, C. M. Thermal unfolding of an intermediate is associated with non-Arrhenius. *J. Mol. Biol.* **2000**, *297*, 193–210.

(38) Khan, F.; Chuang, J. I.; Gianni, S.; Fersht, A. R. The kinetic pathway of folding of barnase. *J. Mol. Biol.* **2003**, *333*, 169–186.

(39) Nguyen, H.; Jaeger, M.; Moretto, A.; Gruebele, M.; Kelly, J. W. Tuning the free energy landscape of a WW domain by temperature, mutation, and truncation. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *352*, 370–381.

(40) Chandler, D. *Introduction to modern statistical mechanics*; Oxford: New York, 1987.

(41) Jarzynski, C. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E* **1997**, *56*, 5018–5035.

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

(43) Park, S.; Schulten, K. Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.* **2004**, *120*, 5946–5961.

(44) Xiong, H.; Crespo, A.; Marti, M.; Estrin, D.; Roitberg, A. E. Free energy calculations with non-equilibrium methods: applications of the Jarzynksi relationship. *Theor. Chem. Acta* **2006**, *116*, 338–346.

(45) Torras, J.; de M. Seabra, G.; Roitberg, A. E. A Multiscale Treatment of Angeli's Salt Decomposition. *J. Chem. Theory Comput.* **2009**, *5*, 37–46.

(46) Piccinini, E.; Ceccarelli, M.; Affinito, F.; Brunetti, R.; Jacoboni, C. Biased Molecular Simulations for Free-Energy Mapping: A Comparison on the KcsA Channel as a Test Case. *J. Chem. Theory Comput.* **2008**, *4*, 173–183.

(47) Huang, H.; Ozkirimli, E.; Post, C. B. Comparison of Three Perturbation Molecular Dynamics Methods for Modeling Conformational Transitions. *J. Chem. Theory Comput.* **2009**, *5*, 1304–1314.

(48) Liphardt, J.; Dumont, S.; Smith, S. B.; Bustamente, C. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science* **2002**, *296*, 1832–1835.

(49) Douarche, F.; Ciliberto, S.; Petrosyan, A.; Rabbiosi, L. An experimental test of the Jarzynski equality in a mechanical experiment. *Europhys. Lett.* **2005**, *70*, 593–599.

(50) Crooks, G. E. Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J. Stat. Phys.* **1998**, *90*, 1481–1487.

(51) Hummer, G.; Szabo, A. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3658–3661.

(52) Zimanyi, E. N.; Silbey, R. J. The work-Hamiltonian connection and the usefulness of the Jarzynski equality for free energy calculations. *J. Chem. Phys.* **2009**, *130*, 171102.

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

(54) Amaro, R.; Tajkhorshid, E.; Luthey-Schulten, Z. Developing an energy landscape for the novel function of a (beta/alpha)8 barrel: Ammonia conduction through HisF. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7599–7604.

(55) Sylte, I.; Andrianjara, C.; Calvet, A.; Pascal, Y.; Dahl, S. Molecular dynamics of NPY Y1 receptor activation. *Bioorg. Med. Chem.* **1999**, *7* (12), 2737–2748.

(56) Ertekin, A.; Nussinov, R.; Haliloglu, T. Association of putative concave protein-binding sites with the fluctuation behavior of residues. *Protein Sci.* **2006**, *15*, 2265-2277.

(57) Hnggi, P.; Talkner, P.; Borkovec, M. Reaction-rate theory: Fifty years after Kramers. *Rev. Mod. Phys.* **1990**, *62*, 251–341, and references therein.

(58) Pollak, E.; Talkner, P. Reaction rate theory: What it was, where it is today, and where is it going? *Chaos* **2005**, *15*, 026116−1-11.

(59) Hernandez, R.; Bartsch, T.; Uzer, T. Transition state theory in liquids beyond planar dividing surfaces. *Chem. Phys.* **2010**, *370*, 270–276.

(60) Beck-Sickinger, A. G.; Wieland, H. A.; Wittneben, H.; Willim, K.-D.; Rudolf, K.; Jung, G. Complete L-alanine scan of neuropeptide Y reveals ligands binding to Y1 and Y2 receptors with distinguished conformations. *Eur. J. Biochem.* **1994**, *225* (3), 947–958.

(61) Fournier, A.; Gagnon, D.; Quirion, R.; Dumont, Y.; Pheng, L.-H.; St-Pierre, S. Conformational and biological studies of neuropeptide Y analogs containing structural alterations. *Mol. Pharmacol.* **1994**, *45*, 93–101.

(62) Marcinkiewicz, J. Sur une propriete de la loi de Gauss. *Math. Z* **1939**, *44*, 612–618.

(63) Berezhkovskii, A.; Szabo, A. One-dimensional reaction coordinates for diffusive activated rate processes in many dimensions. *J. Chem. Phys.* **2005**, *122*, 014503−014506.

(64) Rhee, Y. M.; Pande, V. S. One-dimensional reaction coordinate and the corresponding potential of mean force from commitment probability distribution. *J. Phys. Chem. B* **2005**, *109*, 6780–6786.

(65) Kubelka, J.; Hofrichter, J.; Eaton, W. A. The protein folding speed limit. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76–88.

CT100320G