Exact Low-Force Kinetics from High-Force Single-Molecule Unfolding Events

Jeremiah Nummela and Ioan Andricioaei
Department of Chemistry and Center for Computational Medicine and Biology, University of Michigan, Ann Arbor, Michigan

ABSTRACT Mechanical forces play a key role in crucial cellular processes involving force-bearing biomolecules, as well as in novel single-molecule pulling experiments. We present an exact method that enables one to extrapolate, to low (or zero) forces, entire time-correlation functions and kinetic rate constants from the conformational dynamics either simulated numerically or measured experimentally at a single, relatively higher, external force. The method has twofold relevance: 1), to extrapolate the kinetics at physiological force conditions from molecular dynamics trajectories generated at higher forces that accelerate conformational transitions; and 2), to extrapolate unfolding rates from experimental force-extension single-molecule curves. The theoretical formalism, based on stochastic path integral weights of Langevin trajectories, is presented for the constant-force, constant loading rate, and constant-velocity modes of the pulling experiments. For the first relevance, applications are described for simulating the conformational isomerization of alanine dipeptide; and for the second relevance, the single-molecule pulling of RNA is considered. The ability to assign a weight to each trace in the single-molecule data also suggests a means to quantitatively compare unfolding pathways under different conditions.

INTRODUCTION

The effect of mechanical forces on biomolecules is a topic of forthright interest in several important instances. In single-molecule experiments, external forces are routinely applied in vitro to individual protein or nucleic acid molecules by novel micromanipulation techniques. Using atomic-force microscopy, optical or magnetic tweezers, such manipulations can reveal, one molecule at a time, the dynamics underlying the kinetics of stretch-induced unfolding (1,2) or ligand unbinding (3). Furthermore, biomolecules often respond in vivo to mechanical forces as part of their biological function in such crucial cellular processes as gene regulation, cell adhesion, protein import, muscle contraction, or signal transduction (4).

There are two outstanding issues that demand a theoretical treatment to enable the extrapolation of conformational kinetics observed at higher forces to lower or no forces.

The first issue has to do with the computer simulations of the underlying processes. Major conformational transitions typically occur over milliseconds or longer, while the time range of state-of-the-art molecular dynamics (MD) simulations (5) is limited to microseconds at best. Several techniques have been developed to speed up the simulations. Targeted (6), biased (7), or steered (8) MD, while most useful as computational tools to observe the conformational transitions, often render the underlying kinetics unphysical. This is because such in silico methods, to compensate for the time gap, employ external forces that are much higher than the forces of relevance in vivo or in vitro.

The second issue is experimental in nature. The forces experienced by most biomolecules, in their physiological contexts, are typically low, seldom exceeding a few picoNewtons. The single-molecule experiments that unfold biomolecules often involve attaching long molecular handles (typically of at least a few hundred nm), which makes it difficult to deconvolute the effect of low forces on the biomolecule from that on the handle ($k_BT \approx 4.1$ pN nm at room temperature). As for the simulations, a rigorous force extrapolation framework would be an important asset for the experiments, too.

A significant breakthrough aiding in the interpretation of both simulation and experimental data has been achieved in a landmark article by Hummer and Szabo (9). They have utilized the Jarzynski identity (10) in the context of single-molecule manipulations to reconstruct the free energy profile along the pulling coordinate. Relatedly, the Jarzynski identity has also been used by Liphardt et al. (11) to analyze, from an energetical standpoint, force-extension curves obtained by unfolding RNA, and by the Schulten group to extract, from steered MD, free energy profiles for glycerol conduction through aquaporin (12), and for decaalanine unfolding (13). Using the original Hummer-Szabo method or subsequent developments (14–16), kinetic rates for unfolding can be calculated for the zero-force system under the assumption that the pulling direction is the reaction coordinate (i.e., the only coordinate that is much slower than the other degrees of freedom). However, this assumption may be violated for complex protein folds. Moreover, even if the pulling direction is the reaction coordinate when the pulling force is on, it might not be the reaction coordinate in the absence of the force (17) (e.g., because of passage through force-induced intermediates that are absent in thermal/chemical unfolding (18)).

Here, using a different approach, we develop a theoretical framework that allows simulations performed at a large force to be post-processed and reweighted such that the correct

Submitted April 27, 2007, and accepted for publication July 10, 2007.

Address reprint requests to I. Andricioaei, E-mail: andricio@umich.edu.

Editor: John E. Straub.

0006-3495/07/11/3373/09 $2.00

doi: 10.1529/biophysj.107.111658

kinetics at the lower values of force (including zero) are recuperated. It is important to note that the proposed method dispenses with the requirement to define a reaction coordinate. Instead, it needs the less stringent definition of ''reactant'' and ''product'' basins (e.g., folded and unfolded states), belonging, from this standpoint, to the general class of transition path descriptions for the statistical mechanics of trajectories developed by Chandler and co-workers (19).

The technique we put forth exploits the stochastic properties of Langevin dynamics, which enable one to statistically reweight the contribution of each trajectory entering the expression of a conformational time-correlation function, resulting in the correlation function (and thereby the kinetics) expected in the presence of a different applied force. The theoretical design of the method is presented in the next section, followed by two application sections, concerning the extrapolation of kinetics from computer simulations and from experimental single-molecule force-extension curves. We conclude with a summary discussion.

THEORETICAL DESIGN OF THE METHOD

Treatment of relaxation and kinetics invariably begins with time-correlation functions (20). When one seeks to calculate a kinetic rate constant for a conformational transition (e.g., the unfolding of a biomolecule), the time correlation function of interest can be written in terms of two time products of Heaviside functions, $h_{\mathcal{F},\mathrm{U}}$, which are 1 in the folded, $\mathcal{F}$, and unfolded, $\mathcal{U}$, macrostates, respectively, and zero otherwise,

$$C(t) \equiv \frac{\langle h_{\mathcal{F}}(x_0) h_{\mathcal{U}}(x_\mathrm{t})\rangle_0}{\langle h_{\mathcal{F}}(x_0)\rangle_0} = \frac{\int P(x_\mathrm{t}, t|x_0, 0) h_{\mathcal{F}}(x_0) h_{\mathcal{U}}(x_\mathrm{t}) d\rho(x_0)}{\int h_{\mathcal{F}}(x_0) d\rho(x_0)}, \tag{1}$$

with $\langle\cdot\rangle_0$ denoting averaging over the Boltzmann distribution of initial phase-space points. The second equality in Eq. 1 writes $C(t)$ as an average of the conditional probability $P(x_\mathrm{t}, t|x_0, 0)$ to be at x_t at time t, given that the trajectory started at x_0 at time $t = 0$, over an ensemble of phase-space trajectories, $x(t)$, that are initiated with the proper measure $d\rho(x_0)$ (21). When a defined distinction between the states exists (i.e., transitions between $\mathcal{F}$ and $\mathcal{U}$ are rare when compared to molecular motions), $C(t)$ reaches its equilibrium value, the concentration $\langle h_\mathrm{U}\rangle$ of unfolded species, in an exponential fashion, $C(t) \approx \langle h_\mathrm{U}\rangle(1 - \exp[-(k_{\mathcal{F}\to\mathcal{U}} + k_{\mathcal{U}\to\mathcal{F}})t])$, with a decay constant that is the sum of the forward and backward reaction rate constants. For times smaller than the equilibration time of the system, but greater than the time required for transient behavior to relax, $C(t)$ is a linear function and its time-derivative (i.e., its slope) corresponds to the forward rate constant $k_{\mathcal{F}\to\mathcal{U}}$ (19,22).

We can formally interpret an external force f on the system as a modification of its underlying potential energy, i.e., a transformation $V \to V' = V - fx$. Barring cases when V' is not conservative, one could extrapolate the thermodynamics of the system with potential V from a simulation done on V' by importance sampling techniques (23). This can be done by employing a multiplicative reweighting factor in each term of the ensemble averaging sum, in effect dividing by the equilibrium weight each sampled conformation x had on V', and multiplying by what its weight would have been on V. However, to extrapolate the kinetics on V from dynamics simulated on V' is nontrivial (19,24–27). Directly applying the above-described reweighting will not be sufficient for the calculation of time-dependent averages such as $C(t)$ in Eq. 1 because the entire phase-space trajectory that goes from $x(0)$ to $x(t)$ would be different on the modified potential. For a kinetic extrapolation to work, one has to generalize the weights of single conformational points x to weights of entire trajectories $x(t)$, i.e., to find the functional equivalent of the weight function. Such a functional weight for trajectories can be found in the case of Langevin dynamics, assumed sufficient for the observed conformational dynamics, $m\ddot{x} = -\gamma m\dot{x} - \nabla V(x) + f(t) + \xi(t)$, where m is the mass, γ the coefficient of friction, $-\nabla V(x)$ the force arising due to the potential, $\xi(t)$ a force due to random collision, and f the externally applied force (which may have an explicit time dependence). Assuming stochastic dynamics, the conditional probability $P(x_\mathrm{t}, t|x_0, 0)$ in Eq. 1 can be obtained from knowledge of the joint probability $W(\xi(t))$ of a particular time series of random ''kicks'' $\xi(t)$ that conspires to lead to x_t at time t, and from subsequent functional integration over all possible realizations of $\xi(t)$:

$$P(x_\mathrm{t}, t|x_0, 0) = \int D\xi W[\xi]\delta(x(t) - x_\mathrm{t}). \tag{2}$$

For Langevin dynamics, when $\xi(t)$ is white noise with zero mean and obeying a fluctuation-dissipation relation, $\langle\xi(t)\xi(t')\rangle = 2\,k_\mathrm{B}Tm\gamma\delta(t - t')$, this probability follows from the Gaussian nature of ξ,

$$W[\xi(t)] = \exp\left(-\frac{1}{2w}\int_0^\mathrm{t} \xi(t')^2 dt'\right), \tag{3}$$

where $w = 2\,k_\mathrm{B}Tm\gamma$. As such, the conditional probability can also be written, using the Wiener formalism of path integrals (28) as

$$P(x_\mathrm{t}, t|x_0, 0) = \int_{(x_0,0)}^{(x_\mathrm{t},t)} Dx|J[x]|\exp\left(-\frac{S_\mathrm{f}[x(t)]}{2w}\right), \tag{4}$$

where $S_\mathrm{f}[x(t)] = \int_0^\mathrm{t}(m\ddot{x} + m\gamma\dot{x} + \nabla V - f(t))^2 dt$ is a force-dependent generalization of what is often called the Onsager-Machlup action (29,30); see also (31)). The functional integration in Eq. 4 is performed over all possible paths connecting the initial and final points, and contains the key quantity we seek, i.e., the functional weight of each trajectory, under external force conditions $f(t)$,

$$W_\mathrm{f}[x(t)] \equiv \exp\left(-\frac{S_\mathrm{f}[x(t)]}{2w}\right). \tag{5}$$

The functional Jacobian $|J(x)|$ arises from the ξ to x coordinate transformation. Using, as done here, a time-slicing discretization of the trajectory corresponding to the Ito formalism for stochastic calculus (32), the Jacobian is unity (28) (see also (33)). An appropriate Langevin finite-difference algorithm to advance positions and velocities that meets this criterion is given by the straightforward scheme of Ermak (34), which is based on the assumption that the force remains essentially constant during the time step discretization; we have used its implementation as described in Allen and Tildesley ((35); see Eq. 9.19 therein). An overdamped limit case of the algorithm can also be used, and is described in Extrapolation of Experiments: Single Molecule Pulling of RNA below.

The correction functional for the relative weight of a hypothetical trajectory $x(0\to\mathrm{t})$ generated under force condition $f(t)$ that would have passed through exactly the same configurations as those corresponding to the actual trajectory $x'(0\to t)$ generated under $f'(t)$ (i.e., $x(t) \equiv x'(t)$ for all times from 0 to t) is

$$\frac{W_\mathrm{f}[x(t)]}{W_{\mathrm{f}'}[x(t)]} = \exp\left(-\frac{S_\mathrm{f}[x(t)] - S_{\mathrm{f}'}[x(t)]}{2w}\right). \tag{6}$$

Using the correction functional in Eq. 6, time-correlation functions such those in Eq. 1, for systems under applied force f, can be expressed from trajectories generated under a distinct force f', without approximation, as

$$\langle h_{\mathcal{F}}(x_0)h_{\mathcal{U}}(x_t)\rangle_0 = \frac{\int Dx' h_{\mathcal{F}}(x'_0)h_U(x'_t)W_f[x']/W_{f'}[x']e^{\beta(V'(x_0)-V(x_0))}}{\int Dx' W_f[x']/W_{f'}[x']e^{\beta(V'(x_0)-V(x_0))}}, \quad (7)$$

where $V'(x_0) = V(x_0) - f(0)x_0$ and the corresponding Boltzmann factors are corrections for the different initial distributions. Equation 7 is the central formula of this method; the prime notation emphasizes that the averaging summations are over trajectories generated under applied force f'. As such, kinetics under a particular force condition f', either simulated or measured in single-molecule experiments (provided the action functional is accessible), may be rigorously extrapolated to yield the exact kinetics under a range of other (usually lower) forces f.

In practical implementations, the integral over paths in Eq. 7 is calculated as a sum over all the discrete trajectories generated by the finite difference solution to the Langevin equation. Accordingly, the action is calculated using the corresponding finite difference algorithm,

$$S_f[x(t)] \approx \sum_i \left(m\frac{\Delta v_i}{\Delta t} + m\gamma\frac{\Delta x_i}{\Delta t} + \nabla V_i - f_i \right)^2 \Delta t, \quad (8)$$

where Δv_i and Δx_i are the changes in velocity and position during the i^{th} time step, and $-\nabla V_i$ and f_i are the systematic and applied forces evaluated at the beginning of the step.

Because, given the stochastic force term in the Langevin equation, any trajectory is possible on either potential surface (i.e., with and without the applied force), the method is exact even when the location of product and reactant basins and transition states move significantly as the force is applied. The dividing surface used in the calculation of Eq. 1 should be defined to match the conditions under which the kinetics is desired. Given sufficient sampling, the relatively fewer trajectories exhibiting typical low force behavior will be weighted heavily and will dominate the average. In practice, the method is limited by finite trajectory sampling as the perturbation to the potential energy surface becomes large.

For each of the two possible applications, i.e., for both simulations and experiment, we present results pertaining to biomolecular unfolding in the following two sections. In this section, we continue with the derivation of the form of the correction functional for two typically encountered force conditions: the constant force and the constant loading rate modes of pulling. The fourth section contains a third derivation for the particular case of a time-dependent harmonic pulling potential and overdamped Langevin dynamics. Here, to exemplify, we start with a pedagogical test case of a particle in a one-dimensional potential, $V(x) = x^2(x-2)^2$, providing minima at $x = 0$ and $x = 2$, joined by a barrier of height 1 at $x = 1$. A constant applied force modifies the potential as shown in Fig. 1 *a*. Rate constants for traversal of the barrier from left to right are calculated from the slope of the linear regime of the time-correlation function in Eq. 1 under a variety of applied force conditions.

The first representative case considered for the form of $f(t)$ is that of a constant applied force. In this case, the results of a simulation run at one force are reweighted to yield correlation functions at an array of other forces, including zero force. For trajectories propagated under a constant applied force, f_t, the desired correction functional allowing the calculation of the correlation function for some other applied force f, using Eq. 6, is

$$\frac{W_f[x]}{W_{f_t}[x]} = \exp\left[\frac{(f-f_t)}{w}\left(-\frac{f_t+f}{2}\int dt + m\int d\left(\frac{dx}{dt}\right) + m\gamma\int dx + \int\frac{dV}{dx}dt\right)\right]. \quad (9)$$

Using Eq. 9, and the fact that $V(x_0) - V'(x_0) = (f - f_s)x_0$, the correlation function in Eq. 7 can be expressed as a function of a constant applied force, f, where f_s is the force at which initial states have been sampled, and f_t is the force at which the trajectories have been run.

Trajectories were run by integrating the Langevin equation with $\beta = 4$, $\gamma = 10$, $m = 1$, and f values of 0.0, 0.1, 0.2, 0.3, and 0.4 (in arbitrary units). One million trajectories were run at each value of f. Initial positions for each trajectory were sampled from the paths of single long time trajectories run under appropriate force conditions. All trajectories originate from the left side of the barrier (the potential well near $x = 0$). To serve as references, correlation functions (Eq. 1) were calculated directly on the basis of the simulations at each force value. The correlation functions for each value of f were then recalculated using a single simulation in which the initial states are sampled at $f = 0.0$ and trajectories are propagated at $f = 1.0$, by reweighting the results according to Eq. 7. A comparison of the correlation functions is shown in Fig. 1 *b*. The correlation functions obtained by reweighting the results of the high force simulation are in exact agreement with the direct simulations performed at the lower forces. Rate constants calculated from the reweighted correlation functions also agree exactly with those obtained from the directly-calculated correlation functions (Fig. 1 *c*).

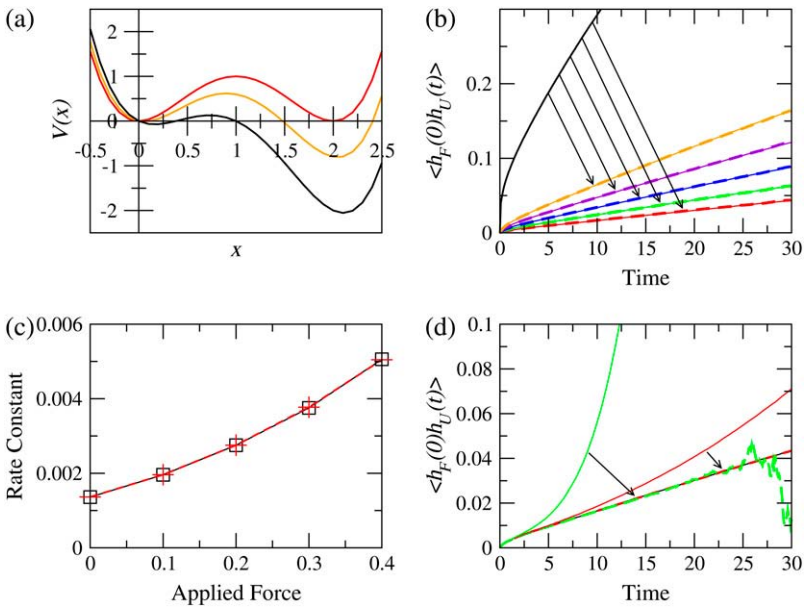

FIGURE 1 (*a*) Constant applied force, f, lowers barrier to forward transition for bistable oscillator (see text); $f = 0$, 0.4, 1.0 shown in red, orange, and black, respectively. (*b*) Correlation functions for bistable oscillator at various applied forces, $f = 0.0$, 0.1, 0.2, 0.3, 0.4 and 1.0 in red, green, blue, violet, orange and black. Solid lines: data from simulations at the actual applied force. Dashed lines (overlapping exactly onto *solid lines*) are data obtained by reweighting of trajectories run at $f = 1.0$. (*c*) Rate constants for bistable oscillator at various applied forces. Values based on non-reweighted data are plotted in black. Values based on reweighted data are plotted in red. (*d*) Correlation function for the bistable oscillator at zero applied force calculated from simulations run with a time dependent applied force. Solid lines are plots of the correlation function based on raw simulated data. Dashed lines have been reweighted to provide the correlation function in the absence of an applied force. The black line is the result of simulations without any applied force. The red lines are the result of simulations with an applied force $f(t) = 0.01t$. The green lines are the result of simulations with an applied force $f(t) = 0.1t$.

The second case considered is that of a linear time dependence, $f(t) = rt$, corresponding to the experimental realization of a constant force-loading rate, r. In this case, no Boltzmann correction is needed for the initial positions, as they are always sampled at zero external force (see Eq. 6). The correction functional to calculate the zero-force kinetics from constant force loading becomes

$$\frac{W[x]}{W_f[x]} = \exp\left(\frac{r\beta}{2\gamma}\int td\left(\frac{dx}{dt}\right) - \frac{r\beta}{2}\int tdx - \frac{r\beta}{2m\gamma}\int t\frac{dV}{dx}dt\right). \quad (10)$$

Trajectories were run with the same parameters previously used for β, γ, and m. One million trajectories were run at each of the following three force conditions: $f = 0$, $f = 0.01t$, and $f = 0.1t$. In Fig. 1 *d*, the correlation function calculated directly from Eq. 1 in the absence of the force is compared with the reweighted correlation functions calculated using the correction functional Eq. 10 in Eq. 7 at $r = 0.01$ and 0.1. The correlation function for $r = 0.01$ is nearly exact, but for $r = 0.1$ significant deviations appear at longer times. Although the formalism is exact, these deviations occur due to insufficient sampling, at large loading rates, of trajectories that would be probable in the absence of the applied force. The deviations reflect the trajectory-space equivalent of the problem of distribution overlap in conformational space when calculating, say, the equilibrium free energy change between two states (23). While a drawback in itself when one wishes to extrapolate between widely different force conditions, the lack of overlap in the trajectory distributions at the two forces can be used, on the other hand, as a gauge for how representative are the trajectories at the higher force for the ensemble of single-molecule trajectories at the lower (or zero) force.

EXTRAPOLATION OF SIMULATIONS: CONSTANT-FORCE PEPTIDE UNFOLDING

Alanine dipeptide is perhaps the simplest model exhibiting properties relevant to the general problem of studying protein conformational changes (36,37). As such, it was used here as a system of biological relevance and sufficient complexity to allow for a more stringent test of reweighted force-induced kinetics, while remaining computationally undemanding so that, for comparison, zero-force kinetics can be calculated directly.

Simulations were done with CHARMM (38), using parameter set 19 and the ACE implicit solvation model (39). This model provides well-defined minima for the C_{7eq} and α_r conformations. The transitions between them, involving rotation about the dihedral angle ψ, serve to represent the $\mathcal{F} \rightarrow \mathcal{U}$ conformational kinetics under study. Fig. 2 *a* shows a simplified Ramachandran plot for the system indicating the two basins. The ψ-values of the transition structures in the absence of an applied force are taken as cutoffs defining the initial $\mathcal{F}$ and final $\mathcal{U}$ basins: $\mathcal{U}$ is defined by $-117.568° < \psi < 39.034°$, and $\mathcal{F}$ by the values of ψ completing a full circle.

To effect the C_{7eq} to α_r transition, a constant pulling force was applied to the C-terminal methyl group (see *inset* in Fig. 2 *a*). To constrain rotational and translational motions of the molecule when the force is applied, the positions of the three heavy atoms composing the N-terminal acetyl group were fixed. The force direction was defined by the difference between the equilibrium position vectors of the methyl group in the C_{7eq} and α_r conformations when the fixed atoms of the

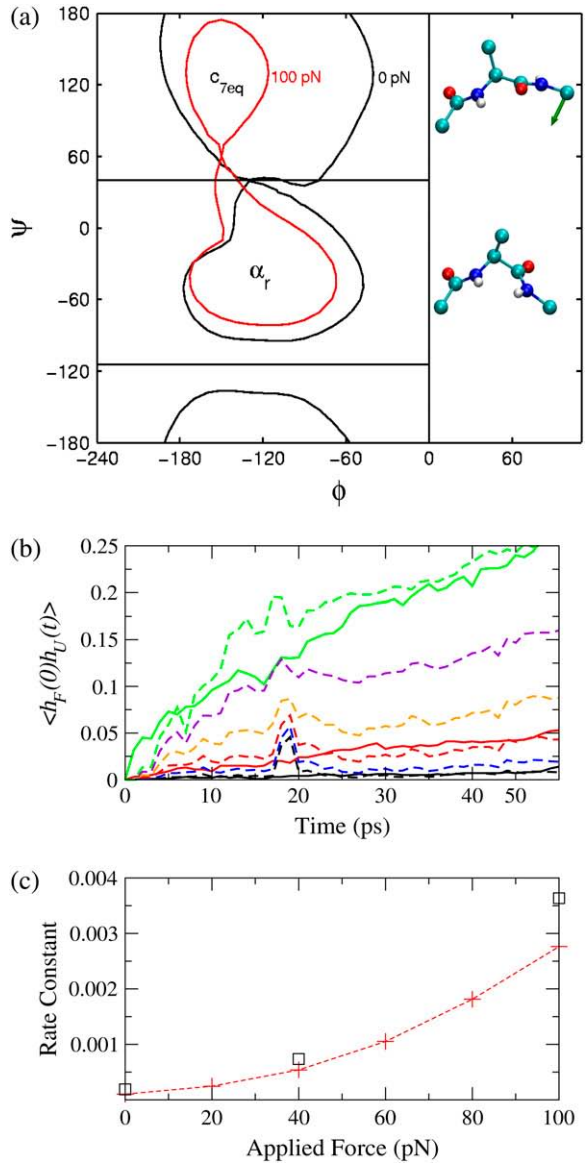

FIGURE 2 (*a*) A single adiabatic potential energy contour in ϕ–ψ space is plotted indicating the location of C_{7eq} and α_r minima, and the transition structure joining them, at both zero (*black*) and 100 pN (*red*) applied forces. The boxes indicate the definitions of the product and reactant regions used in calculating the correlation function (Eq. 1). Equilibrium geometries for each of the minima are depicted on the right, and the green arrow indicates the direction of the applied force. (*b*) Correlation functions for the conformational transition of alanine dipeptide at various applied forces. Solid lines (*black*, *red*, *green*) are from simulations at the actual applied force (0, 40, 100 pN, respectively; see Eq. 1). Dashed lines (*black*, *blue*, *red*, *orange*, *violet*, and *green*) are obtained by reweighting the trajectories of a simulation at 100 pN, using initial states sampled at zero force, for extrapolation to forces of 0, 20, 40, 60, 80, and 100 pN, respectively (Eq. 7). (*c*) Rate constants for conformational transitions of alanine dipeptide at various applied forces. Values based on actual simulations at 0, 40, and 100 pN applied forces are plotted in black. Values plotted in red are determined by reweighting the results of a single set of simulations carried out at 100 pN from initial states generated in the absence of an applied force.

acetyl group are overlaid. We showcase how, using a simulation with a high force (that effects the transition on a rapid timescale), we can extrapolate rigorously the kinetics at any of the lower values of the force (which can also be calculated directly in this model because of its simplicity).

To now apply our method to this system, one thousand 50-ps long trajectories were run for each of three applied forces (0, 40, and 100 pN). Langevin dynamics were used, with a temperature of 300 K, and friction coefficient of 91 ps^{-1} for all atoms. Initial positions for the trajectories were sampled, after discarding an initial equilibration period of 500 ps, by saving geometries every picosecond provided they fell within the initial basin. Correlations calculated from these simulations are depicted as solid lines in Fig. 2 *b*.

In addition, 5000 trajectories were run with an applied force of 100 pN, using initial states sampled at 0 pN. These trajectories were reweighted to recover the correlation function at an array of lower forces, from 0–80 pN. The reweighted correlations functions are shown as dashed lines in Fig. 2 *b*, showing agreement with the directly calculated correlation functions, and permitting us to plot the force dependence of the rate constant in Fig. 2 *c*.

This application to alanine dipeptide has showed that this method can be used to extrapolate kinetics over a wide range of forces, lower than the force at which the simulation has been performed, including zero. While the application pertained to a small portion of a protein backbone, the formalism should be generally applicable to simulations of larger biomolecules using massively (but trivially) parallel computing.

EXTRAPOLATION OF EXPERIMENTS: SINGLE MOLECULE PULLING OF RNA

We now describe the application of the trajectory reweighting strategy to the force-extension curves typically reported in single molecule pulling experiments (see Fig. 3 *a*). The statistical weight of each pulling trajectory is to be calculated from each measured force-extension trace data and adjusted to provide a time correlation under some different condition. Ideally, such a technique could be applied to recover any time-dependent biomolecular property at in vivo conditions from pulling experiment data. However, the calculation of a statistical weight from force-extension traces is complicated by the time resolution of experimental data, and by the inability to measure the total microscopic force on the system. Here we begin by undertaking a less ambitious calculation: the recovery, from force-extension curves for RNA unfolding, of a time-correlation function for low pulling velocity on the basis of a experiments carried out at high pulling velocity. Because the extrapolation is to a condition with a time-dependent Hamiltonian, the calculation of the correlation function, Eq. 1, will not provide a rate constant, but will merely serve to demonstrate how the method can be applied experimentally to recover correlation functions under conditions other than those of the experiment.

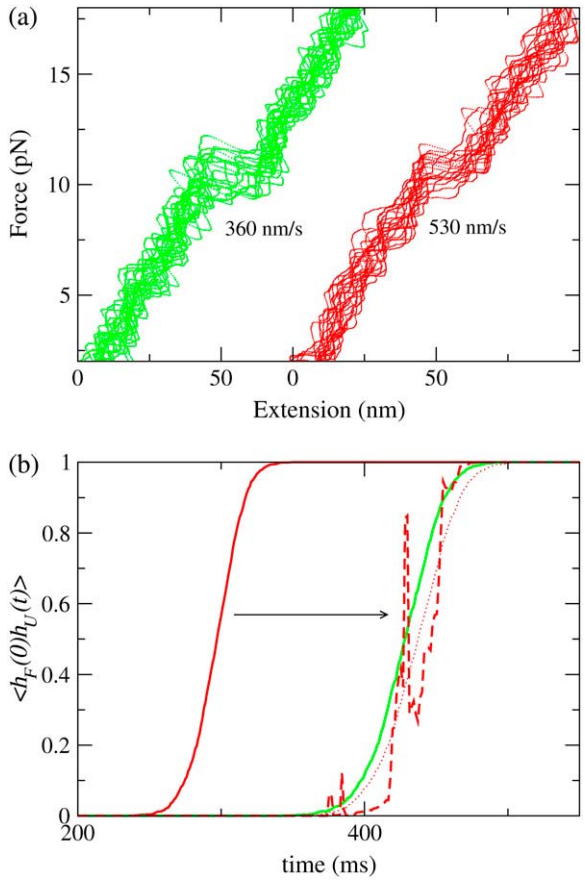

FIGURE 3 (*a*) Force-extension curves for the P5abc RNA unfolding experiments produced from the simulated data using Langevin dynamics on the potential energy surface described by Eq. 14. (*b*) The correlation function calculated directly from single molecule pulling data using Eq. 1 (*continuous red line* for $v = 530$ nm/s, *continuous green line* for $v = 360$ nm/s), by reweighting the high-speed data to the low speed conditions using Eqs. 7 and 13 (*dashed red line*), and under the $t' = \alpha t$ scaling (*dotted red line*). Borderline between $\mathcal{F}$ and $\mathcal{U}$ states at 50 nm extension.

Consider a trajectory propagated by overdamped Langevin dynamics, shown to be appropriate for characteristic timescales longer than ps (40), and for which the displacement is written in discretized form involving the sum of systematic and random displacements,

$$\Delta x_j = \frac{\Delta t}{m\gamma} F_j + \Delta x_j^R, \tag{11}$$

where $F_j = -\nabla V_j + f_j(t)$ includes the external force f applied in the experiment. As the force is evaluated at the beginning of each discrete interval, this constitutes an Ito-discretized integration algorithm for the stochastic differential equation. The probability $w(\Delta x_j^R)$ of observing Δx_j^R for a particular time step j has a Gaussian form, and the probability functional $W(x)$ of experimentally observing a particular trajectory, $x(t)$, is simply the joint probability of occurrence of the series of random displacements, Δx_i^R, necessary to produce that trajectory,

$$w(\Delta x_j^R) = \sqrt{\frac{\beta m\gamma}{4\pi\Delta t}}\exp\left(-\frac{\beta m\gamma}{4\Delta t}(\Delta x_j^R)^2\right),$$
$$W[x] = \prod_{j=1}^{n} w(\Delta x_j^R). \quad (12)$$

This is a general equation for the relative weight of an overdamped Langevin trajectory generated with a discrete time step, Δt. We wish to apply the resulting correction functional, i.e., the ratio W'/W of the above weight, to extrapolate the kinetics from single molecule pulling experiments in which a biomolecule is stretched by a harmonic pulling spring (an AFM cantilever) moving at a constant velocity. As an initial effort, we will calculate a correlation function expected at a low pulling velocity, from force-extension traces measured at a high pulling velocity. In this case the position variable, x, corresponds to the extension along the pulling coordinate, and the change in sampling conditions corresponds to change in the time-dependent component of the force experienced during the trajectory. Given the effective force $F_j = F_{0j} - k_s(x - vj\Delta t)$ under which the experiment is conducted, where $F_0 = -\nabla V$ is the force due to the time-independent molecular potential, we seek the correlation function under the conditions $F_j = F_{0j} - k_s(x - v'j\Delta t)$, where $v' = v/\alpha$ is a new (lower) pulling velocity ($\alpha > 1$). However, for widely distinct pulling velocities, encountered in the experiments, these two sets of conditions give rise to ensembles of trajectories with very poor overlap, resulting in numerical difficulties. In particular, for the problem of single molecule pulling experiments, it is likely that for any reasonable number of samples, all trajectories occurring at a particular low pulling velocity will remain folded while trajectories at high velocity are unfolding, and all trajectories at high velocity will be completely unfolded while trajectories at low velocity are unfolding. This situation makes the calculation of certain correlation functions by reweighting impossible.

This difficulty can be surmounted by scaling the time step, Δt, between observed extensions, x, by a factor of α so that $\Delta t' = \alpha\Delta t$. We no longer consider the relative probability of the same trajectory occurring on two time-dependent potentials. Instead, we calculate the relative probability that two distinct trajectories will occur on two distinct time-dependent potentials, where the new trajectory passes through the same extensions as the original, but at time intervals that are longer by a factor of α. Setting $v' = v/\alpha$ and $\Delta t' = \alpha\Delta t$ results in

$$\frac{W[x']}{W[x]} = \alpha^{-\frac{n}{2}}\exp\sum_{j=1}^{n}\left(\frac{\beta m\gamma}{4\Delta t}\left(1-\frac{1}{\alpha}\right)\Delta x_j^2 + \frac{\beta\Delta t}{4m\gamma}(1-\alpha)F_j^2\right), \quad (13)$$

where Δx_j is the change in extension for the j^{th} step of the trajectory, F_j is the force exerted by the potential that contributes to the j^{th} step, and Δt is the time resolution of the original trajectory data. Note that F_j is the same under both conditions due to cancellation between the scaling of time and pulling velocity. In this way, the portion of the correlation function occurring during the unfolding of trajectories at low pulling velocities is reconstructed from the high velocity pulling data collected during the period of time when high pulling velocity trajectories are unfolding.

Force versus extension curves for the single molecule RNA unfolding experiments of Liphardt et al. (41) were generated following the method used by Hummer and Szabo (42). Overdamped Langevin motion is simulated on a potential energy surface consisting of the lower in energy of two harmonic potentials, and an additional time-dependent harmonic potential:

$$V(x) = \min\left\{\frac{k_0}{2}x^2, \frac{k_0}{2}(x-\Delta x)^2 + \Delta G_u\right\} + k_s(x-vt)^2. \quad (14)$$

The first two potential wells are due to the stretching of the linkers that anchor the RNA to beads, and $k_0 = 0.22$ pN/nm corresponds to the spring constant for the linkers. The first potential corresponds to the folded state of the RNA. The second corresponds to the unfolded state, and is shifted by change in extension due to unfolding, $\Delta x = 15$ nm, and the free energy of unfolding, $\Delta G_u = 34\ k_BT$. The two potentials cross at ~50 nm extension, where the unfolding transition occurs. This serves as the dividing surface between products and reactants. The pulling force is represented by the time-dependent potential well with a force constant $k_s = 0.1$ pN/nm. The pulling potential is moved at a constant velocity, v, to exert an increasing unfolding force by controlling the total extension (handle, molecule, bead), which corresponds to what is called the mixed ensemble (43,44). The parameters used are those in Hummer and Szabo (42) to correspond to the unfolding measurements on a P5abc RNA molecule in Liphardt et al. (41). Fig. 3 *a* shows the overlaid force versus extension curves for 20 simulations at each of two pulling velocities (360 nm/s and 530 nm/s). Force and extension are reported every millisecond, consistent with the time resolution reported by Liphardt et al. (41). As in the experimental reports, the raw data is smoothed by convolution with a window function (5 ms in width). The model used in this section is very schematic, collapsing the multidimensional potential energy surface of biomolecular unfolding onto a single pulling coordinate (although more refined path-integral models using native-contact coordinates for folding exist (45)). The model in Eq. 14 was chosen because it has appeared previously in the literature to describe biomolecular unfolding in single molecule pulling experiments, and appears to provide relatively realistic force-extension curves. Moreover, because we are extrapolating to a condition involving a time-dependent Hamiltonian, $C(t)$ is no longer related to a rate constant, but is simply the time-correlation function describing the appearance of product from reactant that one would observe for the low pulling rate. The idea was simply to demonstrate the possibility, and difficulties inherent,

in applying our method (i.e., Eq. 7 with the associated reweighting formulae, Eqs. 9, 10, or 13) to real single molecule experiments that measure force and extension time-series under constant-force, constant loading rate, or, respectively, constant-velocity modes of operation of the corresponding experimental apparatus.

On the basis of extension data from 2500 simulated experiments performed at each pulling rate, the time correlation function in Eq. 1 is calculated directly. The force-extension traces generated at the high pulling rate are then reweighted by applying the correction functional in Eq. 13, to obtain results pertinent to an experiment performed at the low pulling rate, according to Eq. 7. As seen Fig. 3 *b*, the reweighted curve is able to reproduce the result from direct slow-pulling data. As for the simulations in the previous section, the reweighting here is made possible by mapping each trace sampled at the high pulling rate into a new trace that passes through the same extension points, but under the influence of a pulling rate, $v' = v/\alpha$, that is slower by a factor of α, and at time intervals, $\Delta t' = \alpha \Delta t$, that are greater by a factor of α.

The values $F_j = -\nabla V_j + f_j(t)$ from Eq. 13 are not available experimentally (because the force on the molecule-linker system, $-\nabla V$, cannot be directly measured). However, for the pulling velocities used, the contracting force on the linkers, is on average matched almost evenly by the pulling force of the cantilever spring, resulting in values for F_j reasonably approximated as zero. The correlation function calculated using accurate F_j values for reweighting (known to us from the simulation but hidden in real experiments) is very similar to the one shown in Fig. 3 *b*, which was calculated by approximating the resultant forces as equal to zero.

The use of timescaling in addition to force reweighting yields good results for the tested example. Rather than simply reweighting the contribution of a trajectory to the ensemble average with the probability that it would instead occur at a lower pulling velocity, each trajectory is mapped onto a new trajectory that passes through the same extensions as the original, but at longer time intervals (i.e., time is dilated). This demonstrates the possibility to calculate time correlation results for timescales longer than those of the experiment (or simulation) upon which they are based. As applied to the reweighting of single molecule pulling experiments from high to low pulling rate conditions, this time dilation will also significantly enhance the probability overlap between experimental and target trajectories. This is because the behavior of high pulling velocity trajectories at short times is similar to the behavior of low pulling velocity trajectories at long times, and allows for improved convergence of the reweighted correlation function. In fact, for the studied system, good results for the correlation function at slower pulling rates can be accurately recovered simply by scaling the time coordinates of the high pulling velocity correlation function by a factor of α, i.e., by performing the transformation $t' = \alpha t$, without any statistical reweighting (see Fig. 3 *b*). This is due to the fact that trajectories closely follow the minima created by the interaction of the pulling potential and the time-independent potential of the system. Unfolding occurs when this minima crosses the extension at which the unfolded RNA becomes more stable than the folded RNA. Statistical reweighting is limited by sample size, overlap between probable trajectories under the two conditions, and, in the case of time-step reweighting, by the ability of Langevin dynamics to produce a correct distribution of trajectories for the larger time step.

In addition to providing a formalism for calculating time correlations functions for conditions other than those of the experiment, this analysis provides the statistical weight of each observed trajectory. This information could be a useful bias for computational studies of the system which try to reproduce in atomic detail trajectories conforming to those found from experimental data to be the most probable.

CONCLUDING DISCUSSION

We have presented a method that relies on the calculation of conditional probabilities for systems undergoing Langevin dynamics as a function of an applied force. From a single set of simulated trajectories, recorded at a high force, a family of exact correlation functions can then be reconstructed for an entire range of lower forces of interest (including zero force). The context in which the method is useful requires some overlap to exist between trajectories occurring within the range of force conditions. Useful application to the simulation of biomacromolecules will require a balance to be found between the acceleration of rare event sampling, and the preservation of trajectory overlap. In principle, the method is also applicable in experimental context, and could be useful to analyze and extrapolate, to a wide range of force conditions, the kinetics measured using single molecule pulling experiments in which biomolecules are pulled from a native state into a denatured state by the application of an external force. A successful application to such experiments would be particularly useful because the option of performing the experiment in the absence of force does not exist. When a Heaviside-Heaviside time correlation function is used, the kinetic rate constants can be calculated as a function of force.

Reweighting formulae were derived and applied for three force conditions. The first two corresponded to constant applied force, and constant loading rate. The third corresponded to the case of constant velocity pulling, and required in addition the development of a time dilation scheme to improve the overlap between trajectories relevant under different velocities.

Our approach is not restricted, however, to these three force conditions. Any form of (time- or spatial-dependent) external force can be incorporated in general. For example, one can also derive the formalism for reweighting upon application of biasing forces (7) or holonomic constraints (6).

While exact in theory for any value of the force, in numerical applications the method will be accurate only if

the trajectories generated at high-force visit the conformational pathways populated by trajectories evolving for the low-force regime. This need of trajectory overlap is the functional analogy of the need for overlap of conformational distributions required by free energy perturbation methods.

On the other hand, the ability to assess the weight of each trace in the single-molecule data (and the weight overlap W'/W in Eq. 13) is likely to aid a quantitative comparison of the unfolding pathways under different conditions (46). For instance, despite the fact that temperature- or denaturant-induced unfolding at zero force leads to collapsed structures, whereas forced unfolding results in an extended chain, an extrapolation to zero force of atomic force microscopy experiments for a β-sandwich protein found an unfolding rate comparable to that for chemical denaturation (47). This is similar to what has been found recently, both in DNA unzipping (48) and ligand-unbinding (49) experiments which reported that the detachment force depends linearly on the logarithm of the loading rate and that by an extrapolation to zero force the off-rate in solution can be determined.

Using the reweighting approach presented, if a conformational property of an intermediate along a slow (un)folding route is known, one can also calculate the force dependence of kinetic partitioning coefficients, which are useful means to address the relative count of fast and slow folding trajectories for both proteins and RNA (50).

The probabilistic weight of each recorded unfolding curve that the method can calculate is also useful in comparing quantitatively the pathways between experiment and simulation. The ability to calculate weights leads to the possibility to single-out, from the multitude of single-molecule traces, those kinds of traces (or trajectories) that have significant importance. In addition to yielding physical insight into the nature of that particular realization of the trajectory, this knowledge can lead to novel strategies to generate transition paths that will have a desired weight or weight overlap. For example, in the Jarzynski identity, which has been used both in simulation and experiment, only low-work trajectories count significantly. Finding correlations between the statistical weight functional W and the work distribution (work being roughly the area under the force-extension curves) can lead to insight into ways of generating low work trajectories in simulations and perhaps also in experiments.

While we delimitated our approach from that of Hummer and Szabo by not requiring a transition coordinate, if a such a coordinate is known and one wishes to compute a free energy profile along it, a combination of that method with this one is feasible. One can use the Hummer-Szabo approach to extrapolate the free energy profile at other forces by running simulations in which one accumulates a statistical weight for each trajectory as well as the work, and then reweights the works in the average involved in the Jarzinski identity. Furthermore, while extrapolation to loading rates other than zero can be done by generalization of the Hummer-Szabo approach, that formalism, as currently presented, does not provide for it; such a generalization is possible within this formalism by combining the Jarzynski-based reconstruction techniques with action-functional rescaling.

We have showcased only Heaviside correlation functions, but the proposed method can also be applied to the calculation of any equilibrium time-correlation function. This is important, for instance, in relating to biomolecular NMR measurements (51) for cases when one wishes to understand conformational relaxation with contributions from macrostates separated by barriers so large that driving forces would be required in simulations to effect the conformational exchange.

J.N. was supported by a National Institutes of Health Graduate Training grant. I.A. acknowledges funds from a National Science Foundation CAREER award (No. CHE-0548047), an American Chemical Society PRF-G grant, and the University of Michigan.

REFERENCES

1. Strick, T., J. F. O. Allemand, V. Croquette, and D. Bensimon. 2001. The manipulation of single biomolecules. *Phys. Today*. 54:46–51.
2. Bustamante, C., Z. Bryant, and S. B. Smith. 2003. Ten years of tension: single-molecule DNA mechanics. *Nature*. 421:423–427.
3. Merkel, R., P. Nassoy, A. Leung, K. Ritchie, and E. Evans. 1999. Energy landscape of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature*. 397:50–53.
4. Gao, M., M. Sotomayor, E. Villa, E. Lee, and K. Schulten. 2006. Molecular mechanisms of cellular mechanics. *Phys. Chem. Chem. Phys*. 8:3692–3706.
5. Karplus, M., and J. A. McCammon. 2002. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol*. 9:646–652.
6. Schlitter, J., M. Engels, P. Kruger, E. Jacoby, and A. Wollmer. 1993. Targeted molecular-dynamics simulation of conformational change—application to the T↔R transition in insulin. *Mol. Simul*. 10:291–308.
7. Paci, E., and M. Karplus. 1999. Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *J. Mol. Biol*. 288:441–459.
8. Isralewitz, B., M. Gao, and K. Schulten. 2001. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol*. 11:224–230.
9. Hummer, G., and A. Szabo. 2001. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. USA*. 98:3659–3661.
10. Jarzynski, C. 1997. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett*. 78:2690–2693.
11. Liphardt, J., B. Onoa, S. B. Smith, I. Tinoco, and C. Bustamante. 2001. Reversible unfolding of single RNA molecules by mechanical force. *Science*. 292:733–737.
12. Jensen, M., S. Park, E. Tajkhorshid, and K. Schulten. 2002. Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proc. Natl. Acad. Sci. USA*. 99:6731–6736.
13. Park, S., F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten. 2003. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J. Chem. Phys*. 119:3559–3566.
14. Hummer, G., and A. Szabo. 2003. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophys. J*. 85:5–15.
15. Dudko, O. K., G. Hummer, and A. Szabo. 2006. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett*. 96:108101.
16. Manosas, M., D. Collin, and F. Ritort. 2006. Force-dependent fragility in RNA hairpins. *Phys. Rev. Lett*. 96:218301.

17. Kirmizialtin, S., L. Huang, and D. E. Makarov. 2005. Topography of the free-energy landscape probed via mechanical unfolding of proteins. *J. Chem. Phys.* 122:234915.

18. Williams, P. M., S. B. Fowler, R. B. Best, J. L. Toca-Herrera, K. A. Scott, A. Steward, and J. Clarke. 2003. Hidden complexity in the mechanical properties of titin. *Nature*. 422:446–449.

19. Dellago, C., P. G. Bolhuis, F. S. Csajka, and D. Chandler. 1998. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.* 108:1964–1977.

20. Berne, B. J., and R. Pecora. 1976. Dynamic Light Scattering. Wiley, New York.

21. Chandler, D. 1987. Introduction to Modern Statistical Mechanics. Oxford University Press, New York.

22. Chandler, D. 1978. Statistical-mechanics of isomerization dynamics in liquids and transition-state approximation. *J. Chem. Phys.* 68:2959–2970.

23. Torrie, G. M., and J. P. Valleau. 1977. Non-physical sampling distributions in Monte Carlo free-energy estimation—umbrella sampling. *J. Comput. Phys.* 23:187–199.

24. Mazonka, O., C. Jarzynski, and J. Blocki. 1998. Computing probabilities of very rare events for Langevin processes: a new method based on importance sampling. *Nucl. Phys. A*. 641:335–354.

25. Mazonka, O., C. Jarzynski, and J. Blocki. 1999. Computing probabilities of very rare events for Langevin processes: new method based on importance sampling. *Nucl. Phys. A*. 650:499–500.

26. Zuckerman, D. M., and T. B. Woolf. 2001. Efficient dynamic importance sampling of rare events in one dimension. *Phys. Rev. E*. 6302:016702.

27. MacFadyen, J., and I. Andricioaei. 2005. A skewed-momenta method to efficiently generate conformational-transition trajectories. *J. Chem. Phys.* 123:074107.

28. Kleinert, H. 2004. Path Integrals in Quantum Mechanics, Statistics, Polymer Physics and Financial Markets, 3rd Ed. World Scientific.

29. Onsager, L., and S. Machlup. 1953. Fluctuations and irreversible processes. *Phys. Rev.* 91:1505–1512.

30. Machlup, S., and L. Onsager. 1953. Fluctuations and irreversible process. 2. Systems with kinetic energy. *Phys. Rev.* 91:1512–1515.

31. Chandrasekhar, S. 1943. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.* 15:1–89.

32. Ito, K. 1951. On stochastic differential equations. *Mem. Amer. Math. Soc.* 4:1–51.

33. Elber, R., J. Meller, and R. Olender. 1999. Stochastic path approach to compute atomically detailed trajectories: application to the folding of C peptide. *J. Phys. Chem. B*. 103:899–911.

34. Ermak, D. L., and H. Buckholz. 1980. Numerical integration of the Langevin equation: Monte Carlo simulation. *J. Comput. Phys.* 35:169–182.

35. Allen, M. P., and D. J. Tildesley. 1987. Computer Simulation of Liquids. Oxford University Press.

36. Neria, E., S. Fischer, and M. Karplus. 1996. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* 105:1902–1921.

37. Bolhuis, P. G., C. Dellago, and D. Chandler. 2000. Reaction coordinates of biomolecular isomerization. *Proc. Natl. Acad. Sci. USA*. 97:5877–5882.

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

39. Schaefer, M., and M. Karplus. 1996. A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.* 100:1578–1599.

40. Smith, J. C. 1991. Protein dynamics: comparison of simulations with inelastic neutron scattering experiments. *Q. Rev. Biophys.* 24:227–291.

41. Liphardt, J., S. Dumont, S. Smith, I. Tinoco, and C. Bustamante. 2002. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science*. 296:1832–1835.

42. Hummer, G., and A. Szabo. 2005. Free energy surfaces from single-molecule force spectroscopy. *Acc. Chem. Res.* 38:504–513.

43. Gerland, U., R. Bundschuh, and T. Hwa. 2001. Force-induced denaturation of RNA. *Biophys. J.* 81:1324–1332.

44. Manosas, M., and F. Ritort. 2005. Thermodynamic and kinetic aspects of RNA pulling experiments. *Biophys. J.* 88:3224–3242.

45. Wang, J., K. Zhang, H. Y. Lu, and E. K. Wang. 2005. Quantifying kinetic paths of protein folding. *Biophys. J.* 89:1612–1620.

46. Paci, E., and M. Karplus. 2000. Unfolding proteins by external forces and temperature: the importance of topology and energetics. *Proc. Natl. Acad. Sci. USA*. 97:6521–6526.

47. Carrion-Vazquez, M., A. F. Oberhauser, S. B. Fowler, P. M. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez. 1999. Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl. Acad. Sci. USA*. 96:3694–3699.

48. Strunz, T., K. Oroszlan, I. Schumakovitch, H. J. Guntherodt, and M. Hegner. 2000. Model energy landscapes and the force-induced dissociation of ligand-receptor bonds. *Biophys. J.* 79:1206–1212.

49. Schwesinger, F., R. Ros, T. Strunz, D. Anselmetti, H. J. Guntherodt, A. Honegger, L. Jermutus, L. Tiefenauer, and A. Pluckthun. 2000. Unbinding forces of single antibody-antigen complexes correlate with their thermal dissociation rates. *Proc. Natl. Acad. Sci. USA*. 97:9972–9977.

50. Thirumalai, D., D. K. Klimov, and S. A. Woodson. 1997. Kinetic partitioning mechanism as a unifying theme in the folding of biomolecules. *Theor. Chem. Acc.* 96:14–22.

51. Case, D. A. 2002. Molecular dynamics and NMR spin relaxation in proteins. *Acc. Chem. Res.* 35:325–331.