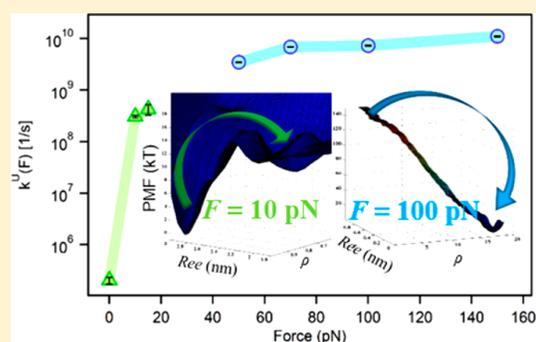


Simulated Force Quench Dynamics Shows GB1 Protein Is Not a Two State Folder

Ronen Berkovich,^{*,†,⊕} Jagannath Mondal,^{*,‡,⊥} Inga Paster,[†] and B. J. Berne^{*,§}[†]Department of Chemical Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel[‡]Tata Institute of Fundamental Research, Centre for Interdisciplinary Sciences, Hyderabad, India[§]Department of Chemistry, Columbia University, New York, New York 10027, United States

Supporting Information

ABSTRACT: Single molecule force spectroscopy is a useful technique for investigating mechanically induced protein unfolding and refolding under reduced forces by monitoring the end-to-end distance of the protein. The data is often interpreted via a “two-state” model based on the assumption that the end-to-end distance alone is a good reaction coordinate and the thermodynamic behavior is then ascribed to the free energy as a function of this one reaction coordinate. In this paper, we determined the free energy surface (PMF) of GB1 protein from atomistic simulations in explicit solvent under different applied forces as a function of two collective variables (the end-to-end-distance, and the fraction of native contacts ρ). The calculated 2-d free energy surfaces exhibited several distinct states, or basins, mostly visible along the ρ coordinate. Brownian dynamics (BD) simulations on the smoothed free energy surface show that the protein visits a metastable molten globule state and is thus a three state folder, not the two state folder inferred using the end-to-end distance as the sole reaction coordinate. This study lends support to recent experiments that suggest that GB1 is not a two-state folder.



INTRODUCTION

Protein conformational dynamics is often mechanistically interpreted through the protein's underlying multidimensional free energy landscape.^{1,2} Theory and simulation can be used to determine the free energy surface as a function of specified collective variables and thus can be useful in ferreting out mechanistic aspects of folding and unfolding.³ A major problem is determining a minimum set of collective variables capable of describing the essential features of the slow dynamics of a given protein. The free energy hypersurface suggests possible folding pathways involving intramolecular (amino-acid) and external (solvent) rearrangements. In general, we expect the pathways to involve many degrees of freedom and thus a description in terms of a multidimensional conformation space. A resolution to the Leventhal paradox⁴ is given by funnel theory,^{5,6} which relies on the assumptions, that the free energy hypersurface can be reduced to a lower dimensional of conformation space to describe folding. McLeish's theory suggests an organization of the protein high dimensional energy surfaces into hypergutter structures, and their low dimensional projections represent the funneled folding landscape of the protein.⁷ Projecting the multidimensional energy landscape onto a low dimensional conformational space therefore becomes a useful approach for understanding minimal folding routes.^{6,8}

Conformational free energy funnels,^{5,9} high-resolution single molecule experiments^{10–13} and computer simulations^{14–17} have

helped in understanding protein folding and unfolding in response to changing mechanical forces. Manipulating proteins at the single molecule level using magnetic tweezers, optical tweezers, and atomic force microscopy (AFM)^{10–12,18–22} differs from bulk experiments in which temperature and solvent alteration mediate the structural perturbation.²³ Single molecule force spectroscopy experiments (SMFS) probe protein conformations by measuring the end-to-end distance, but in ensemble experiments the probed coordinates are often harder to define. Interestingly, the unfolded and intermediate states explored following chemical denaturation or thermal denaturation do not necessarily correspond with each other²⁴ or with mechanical denaturation by force. In SMFS, the dynamical evolution of the end-to-end distance of the protein is directly observed and this well-defined coordinate is often chosen as the reaction coordinate. This one-dimensional coordinate is assumed to capture the transitional states of the protein free energy landscape under the implicit assumption that they evolve along a single path, as the applied force averages out faster (short-lived) configurational states, ending up showing the dominant transitions.^{25,26} Nevertheless, interpreting protein dynamics using two-dimensional energy

Received: January 19, 2017

Revised: April 3, 2017

Published: April 28, 2017

landscapes (given by a length, and an enthalpy coordinate, characterized by the structural number of contacts) turns out to be more insightful and proves to be sufficient to capture the main mechanistic features of folding of the small globular protein.^{14,15,24,26–35}

The B1 segment of streptococcal Protein G (GB1), a small α/β fold protein has been investigated experimentally^{36–41} and by numerical simulations as well.^{14,15,26,40,42–46} Its folding, however, is subjected to some debate. While some works consider it as a simple two-state folding protein,^{26,39} other suggest a low-energy on-pathway intermediate,^{47,48} supported by Go-like lattice simulations.^{49,50} However, a recent work by Gianni and co-workers,⁴⁰ identified a distinctly different high-energy partially unfolded state than the one previously suggested. Unlike the previous works, the later combines experimental measurements of GB1 protein with all-atom molecular dynamics (MD) simulations in explicit solvent. In order to address this discrepancy and to better understand these processes, we studied the presence of GB1 conformations from the perspective of single molecule force spectroscopy by performing all-atom MD simulations of GB1 in explicit solvent, from which the free energy surface (or potential of mean force (PMF)) were determined. To explore the kinetics of GB1, we then performed Brownian dynamics (BD) simulations on these potential surfaces to determine the dynamical transitions that the protein undergoes when external force is applied and when this force is relaxed, such as in unfolding, collapse, and refolding. We find that solvent plays an important role in the formation of a local molten-globule state due to hydrophobic interactions and that this has a dramatic effect on the protein dynamics. We find that to determine the full dynamical response we must use not only the collective end-to-end coordinate but also the fraction of native contacts.

■ COMPUTATIONAL METHODS

Atomistic MD Simulations. The major focus of the current work is to quantify the conformational free energy landscape of a globular protein along certain relevant collective variable. Toward this end, we have chosen GB1 protein as a system of interest by starting with its NMR structure (PDB 3GB1). The GB1 protein was solvated using atomistic water and charge-neutralized by adding four sodium ions. The delineation of free-energy landscape required sampling protein conformation of diverse number of native contacts and accordingly the system size was adjusted to fit the dimension of the protein conformations commensurate with the certain native contacts and extension. The globular conformations of the protein were solvated by 5600 water molecules in a cubic box of dimension 5.6 nm along each direction and the extended unfolded conformations were solvated by 69789 water molecules in a rectangular box of dimension 10 nm \times 10 nm \times 20 nm. The overall system size ranged between 17674 particles for globular conformations and 210226 particles for unfolded extended conformation. The Charmm36 force-field⁵¹ was used for the protein all-atom parameters and TIP3P⁵² water model was used to describe the water potentials.

We have employed the fraction of native contacts, ρ as the collective variable in the umbrella sampling simulation. The choice of fraction of native contact as the collective variable of interest was majorly guided by the previous studies.^{26,42} The evolution of the fraction of native contacts, ρ , was monitored during the simulation by counting only those native pairs of C-alpha atoms of the GB1 domain, which retain a separation

within 0.65 nm. The fraction of native contacts ρ was then calculated by following normalization:

$$\rho = \frac{N - N_u}{N_f - N_u} \quad (1)$$

where N is the instantaneous number of native contacts, N_f is the number of native contacts in native conformation, and N_u is the number of native contacts in the unfolded extended conformation of the protein. Hence $\rho = 1$ in the native configuration and $\rho = 0$ in the unfolded extended configuration. We note that, in our simulation, we could reach an unfolded extended state which has the total number of native contacts of $N_u = 63$. Since all of the simulations were done in the presence of explicit water molecules, modeling an extended state, which has an all-trans structure, would have necessitated usage of prohibitively large box to accommodate the number of water required to solvate the completely extended state. However, the most unfolded extended state used in our study (corresponding to a number of native contacts of 63) depicts qualitatively very similar feature of typical all-trans extended structure. The number of native contacts in the native folded structure was $N_f = 178$.

The initiation of umbrella sampling at each ρ necessitated the generation of a large database of different conformations corresponding to a particular ρ . Toward this end we employed different simulation techniques to facilitate the generation of conformation of all ρ . The conformations corresponding to relative large ρ were generated by equilibrating the solvated system of protein for 100 ns at 300 K. Subsequently, the system was equilibrated at 450 K for 100 ns to generate a database of configurations representative of the intermediate native fraction. Finally, to obtain configurations with lower native contacts corresponding to unfolded extended conformations, the protein was subjected to a constant external force (value mentioned below) that was applied between the Nitrogen atom of the N-terminal residue and the Carbon atom of the C-terminal residue.

Representative configurations corresponding to different fractions of native contacts ranging from 0–1 were chosen from these equilibrium simulations (where 0 means that the protein is completely unfolded and 1 represents the folded configuration). To obtain the free energy landscape as a function of ρ , an umbrella sampling approach⁵³ was employed. This required the coarse-graining of the number of native contacts making it a continuous differentiable collective variable which was then implemented using a free energy simulation namely plugin plumed 1.0.4⁵⁴ which we used for performing all umbrella sampling simulations in the current project in combination with the MD simulation program gromacs 4.5.4.⁵⁵ Toward this end, we used the following coarse-grained description of the number of native contacts

$$N = \sum_i \sum_j n_{ij} \quad (2)$$

$$n_{ij} = \begin{cases} \frac{1 - (r_{ij}/r_0)^n}{1 - (r_{ij}/r_0)^m}, & r_{ij} > r_0 \\ 1, & \text{otherwise} \end{cases} \quad (3)$$

where $d_0 = 0.05$ nm, $r_0 = 0.65$ nm, $n = 6$, $r_{ij} = |r_i - r_j| - d_0$, and $m = 12$. i and j are pairs of C- α atoms constituting the native contact.

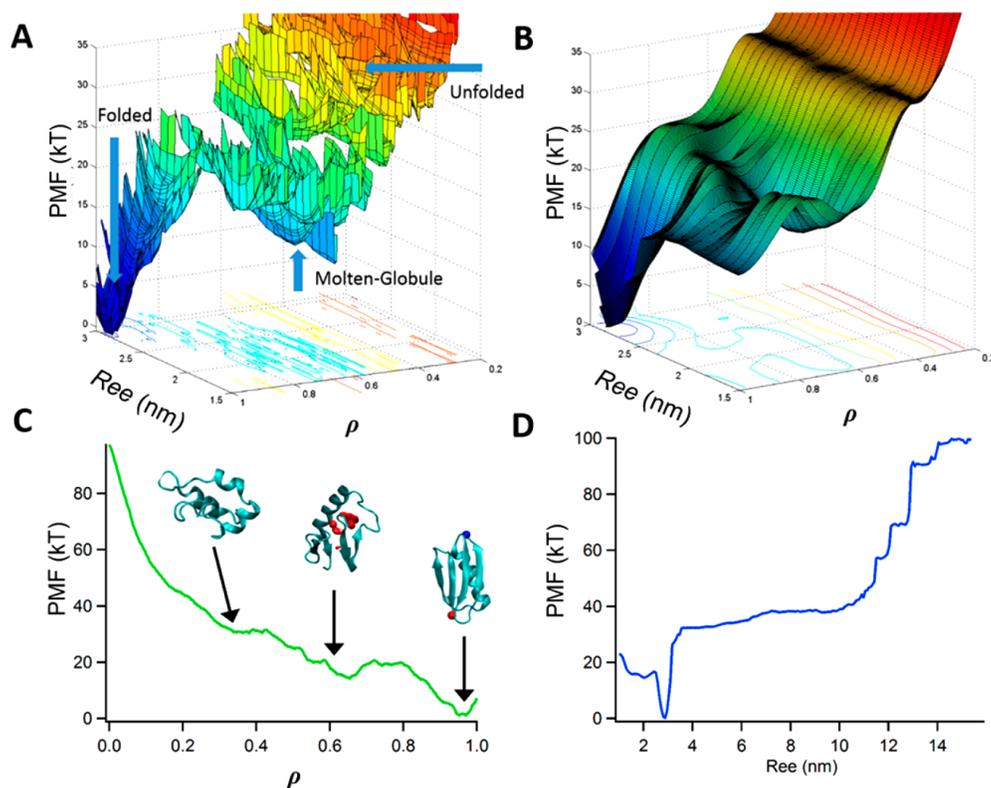


Figure 1. PMF calculated from MD simulations of GB1 in explicit solvent as a function of the fraction of native contacts (ρ), and the end-to-end distance (Ree). (A) Two-dimensional PMF profile along Ree and ρ , where the presence of several intermediates having different distinct values of ρ are located at low Ree . (B) Extrapolated and smoothed version of the PMF shown in part A. (C) 1-d Projection of the PMF on the ρ coordinate (green), where three states are apparent—the native state ($\rho \sim 1$), molten-globule/collapsed state ($\rho \sim 0.65$), completely unfolded, but coiled ($\rho \sim 0.35$) and fully extended state ($\rho \sim 0$). The various states in part C correspond to different values of ρ collapse to almost the same value of Ree in 1A. (D) 1-d projection of the PMF on the Ree coordinate (blue). Here the states are distributed differently than in C, where the folded and molten globule states are localized at low Ree (~ 3 and ~ 2 nm respectively), while the unfolded (coiled) state can take wider range of values. Snapshots of the protein at different conformations (from the MD simulations) are illustrated along part C, corresponding to their representative regions along the PMF landscapes.

Here, 114 uniformly spaced (with a spacing of 0.0087) umbrella-sampling windows were employed to cover the range of native fractions from 0 to 1. As is customary in umbrella sampling, each value of ρ was harmonically constrained with a force constant of 50 kJ/mol and simulated at a constant pressure of 1 bar and a temperature of 300 K for 10 ns (hence a total simulation length of 1.14 μ s). The harmonic force constant and interval between two adjacent “windows” were chosen to ensure a Gaussian probability distribution of the collective variable with appreciable overlap among the distributions in adjacent windows (see Supporting Information, Figure S1). The weighted histogram analysis method (WHAM)⁵⁶ was used to determine the free energy as a function of ρ by combining all the umbrella-sampling windows. The cumulative time series of Ree , shown in Figure S2a) along with the histograms of Ree , shown in Figure S2b). These were obtained from the umbrella sampled trajectories along ρ . They clearly show that all values of end-to-end distances ranging from globular to the extended conformations have been effectively sampled in the simulations. Finally, the effect of a finite biasing force on the free energy landscape was explored by repeating the calculation with similar umbrella sampling protocols for constant biasing forces $F = 10, 15, 50, 70, 100,$ and 150 pN applied between the Nitrogen atom of the N-

terminal residue and the Carbon atom of the C-terminal residue. Hence, the cumulative simulation time for all umbrella sampling at all biasing forces was 7.98 μ s.

In all simulations, the Nose–Hoover thermostat^{57,58} and the Parrinello–Rahman barostat⁵⁹ were employed to maintain the average desired temperature and pressure. A group-based cutoff scheme was implemented within gromacs⁵⁵ with the Lennard-Jones interaction being curtailed beyond 1 nm. The electrostatic interaction was treated using the particle mesh Ewald summation technique.⁶⁰ The integration time step in all simulations was 0.002 ps. The water molecules were simulated as rigid triatomic molecules using the SETTLE algorithm.⁶¹ All the bonds involving hydrogen atoms of protein were constrained using LINCS algorithm.⁶² The interpretation of single-molecule force spectroscopy experiments is usually based on the end-to-end distance of the protein as the reaction coordinate, but we wish to determine the free energy as a function of both Ree and ρ . This was done as follows; First, using the trajectories of each umbrella-sampled window, we determined the conditional probability $P(Ree|\rho)$ of finding Ree at a given ρ . The joint probability is then recovered using the relation

$$P(Ree, \rho) = P(Ree|\rho)P(\rho) \quad (4)$$

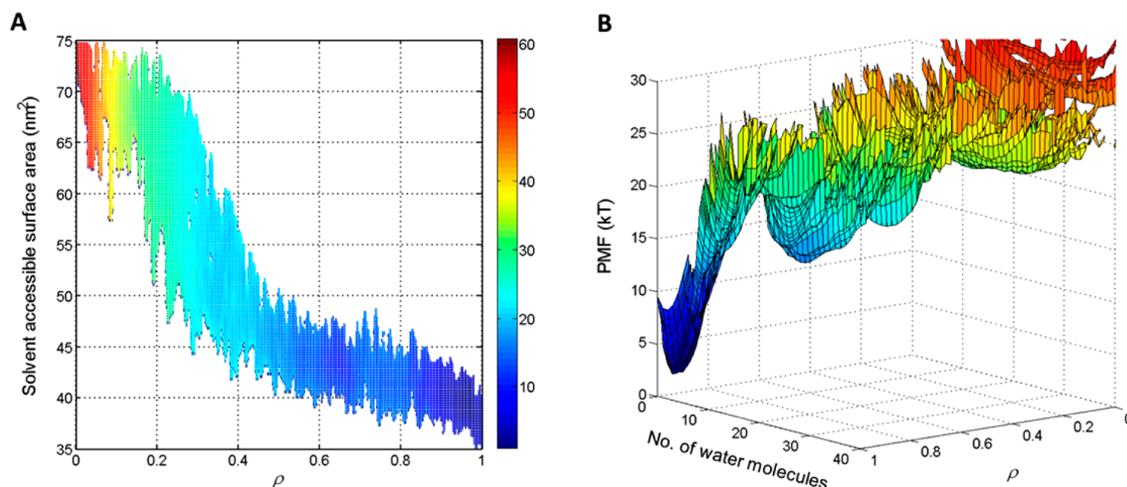


Figure 2. Solvent accessibility effect on the states along the free energy surface. (A) Free energy surface along solvent accessible surface area (SASA) and ρ . A sharp decrease in SASA is observed when collapsing from the extended configuration ($\rho = 0$) to the molten globule state ($\rho \sim 0.35$) is followed by very slow decrease in SASA as folding takes place when ρ increases and the native folded configuration is reached. (B) Free energy profile along number of water molecules and ρ . Water molecule depletion is observed inside the globular core of the protein and further more in the folded state. The globular structure of the protein was ensured by considering only configurations with radius of gyration of less than 1.5 nm.

We also extracted the one-dimensional free energy along Ree by integrating over the fraction of native contacts.

Brownian Dynamics (BD) Simulations. We studied the dynamics of the unfolding and refolding processes on the 2-d PMFs by numerically solving the coupled overdamped Langevin equations:⁶³

$$Ree_{i+1} = Ree_i - \frac{D_{Ree}\Delta t}{k_B T} \frac{\partial W(Ree_i, \rho_i; F)}{\partial Ree} + (2D_{Ree}\Delta t)^{1/2}\Gamma_{Ree}(t) \quad (5)$$

$$\rho_{i+1} = \rho_i - \frac{D_\rho\Delta t}{k_B T} \frac{\partial W(Ree_i, \rho_i; F)}{\partial \rho} + (2D_\rho\Delta t)^{1/2}\Gamma_\rho(t) \quad (6)$$

(Ree_i, ρ_i) and (Ree_{i+1}, ρ_{i+1}) are respectively the values of the two collective coordinates at t and $t + \Delta t$, and $W(Ree, \rho; F)$, is the free PMF containing the applied force. The time increment, is $\Delta t = 1.10 \cdot 10^{-10}$ s. $\Gamma_{Ree}(t)$ and $\Gamma_\rho(t)$ represent Gaussian random white noise stochastic forces. BD simulations were performed over the interpolated-smoothed free energy surfaces for force protocols that are typical of single molecule force-clamp experiments^{64–66} with a diagonal diffusion tensor where the diffusion coefficients along Ree and along ρ are different. For simplicity, we ignore both position dependence and cross terms in this tensor. D_{Ree} was taken as $1.5 \cdot 10^{-8}$ nm²/s, which is of the same order of magnitude as experimentally reported⁶⁷ and D_ρ was taken to be $1.5 \cdot 10^{-7}$ s⁻¹ in line with reference.²⁶ In previous SMFS experiments, the evolution of Ree is recorded without providing any information about the collective behavior of other orthogonal coordinates.

RESULTS AND DISCUSSION

Thermodynamics. We used umbrella sampling to determine the PMF along the two reaction coordinates, ρ , the fraction of native contacts, and Ree , the end-to-end distance (See Supporting Information for the methodological details of free energy simulation). Our simulation strategy allows us to

explore the role played by water (often ignored in some structure based simulations,²⁶ although it was readily shown that MD simulations in explicit solvent give significant differences from simulations without explicit solvent such as Go-like models⁴⁵). Water can have significant effects on the energy landscape. The PMF is determined from $W(\rho, Ree; F)$, $= -k_B T \ln[P(\rho, Ree; F)]$, where $P(\rho, Ree; F)$ is the probability density of these coordinates under a constant force along the end-to-end distance. In addition to the unperturbed (zero force) equilibrium PMF (Figure 1A), the PMFs corresponding to six external forces, $F = 10, 15, 50, 70, 100$, and 150 pN were calculated. Ree is the coordinate along which the protein is pulled in single-molecule force spectroscopy experiments,²¹ and ρ provides complementary information that could not be accessed in such experiments.^{25,26} These PMFs provide key insights into cooperative dynamical processes measured in single molecule experiments and in ensemble experiments.²⁴ A continuous representation of the PMF is shown in Figure 1B, where the undersampled regions were compensated by interpolation (see Supporting Information). Subtleties of the force quench dynamics have been addressed in depth using a simple RNA hairpin,^{68–71} nevertheless, these studies use Go-like models and no explicit solvent for the study the force-induced unfolding of RNA hairpins, a very different physical system than GB1, which involves water dynamics in forming its intermediate metastable state.

The two-dimensional PMF, as depicted in parts A and B of Figure 1, identifies several important conformational states. These are more clearly seen in the calculated one-dimensional projections of the PMF surface onto ρ (Figure 1C) and Ree (Figure 1D) separately. Three key thermodynamic states are quite apparent from the one-dimensional projection of the PMF onto ρ : the native state at $\rho \sim 1$, the unfolded state at $\rho \sim 0.35$, and a molten globule/collapsed intermediate at $\rho \sim 0.65$. As suggested in previous papers,^{25,26} it can be clearly seen that along ρ the separated native and intermediate states are distinguishable but that, in the 1-d PMF as a function of Ree , the native and intermediate states are not resolved. Experiments probing Ree therefore would be incapable of resolving these

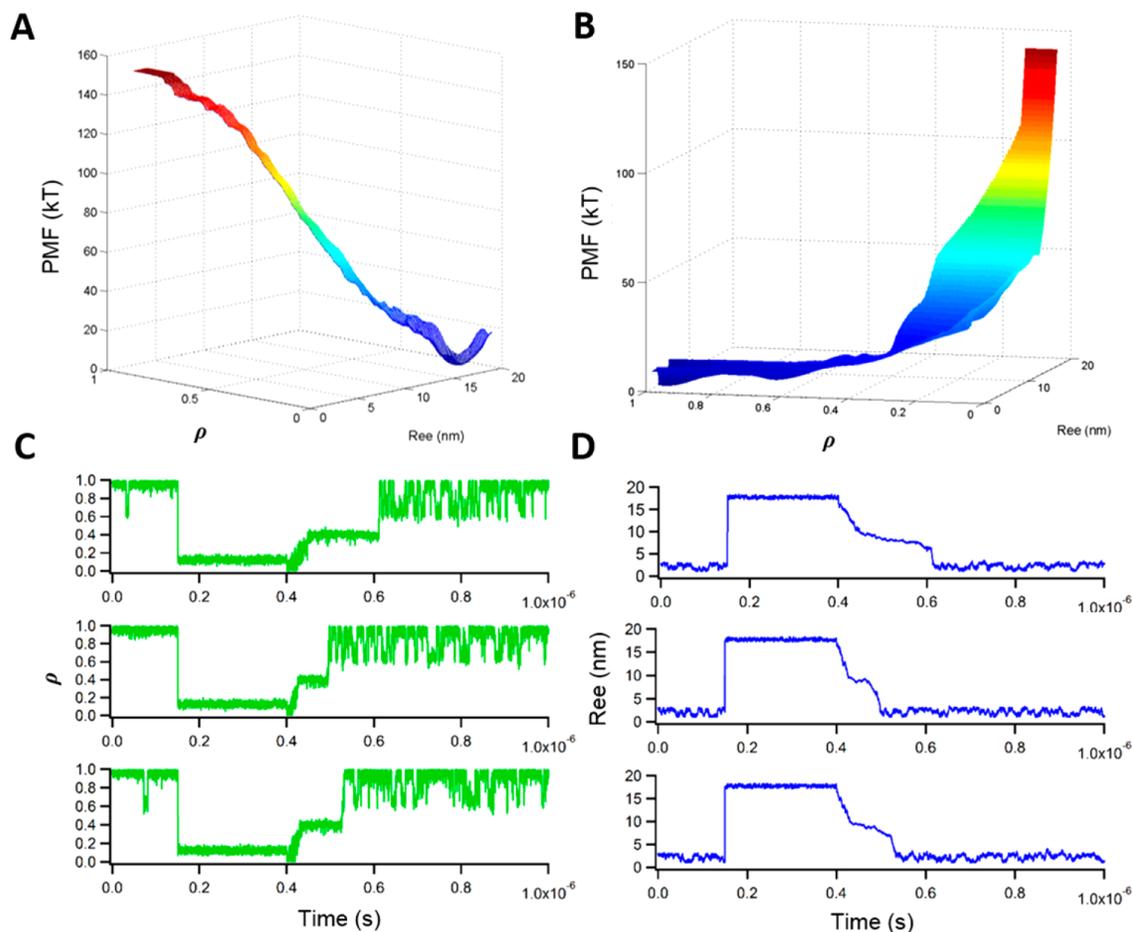


Figure 3. Brownian dynamics simulations on 2-d PMFs according to (A) PMF calculated from MD simulations at 100 pN, (B) PMF calculated from MD simulations at 15 pN, (C) three exemplary trajectories calculated over three 2-d PMFs and projected onto the ρ coordinate, and (D) the Ree coordinate following a force protocol that starts at zero force, goes to 100 pN, and then is quenched down to 15 pN.

states and would thus lead one to interpret the data in terms of a two state model as shown in Figure 1D.^{65,66} The presence of explicit solvent gives rise to the formation of a collapsed/molten-globule state between the folded and unfolded states. Surprisingly, the value of Ree for the intermediate state is slightly smaller than that of the native folded state. This occurs because the molten state is more flexible than the folded state and is free to explore a range of conformations leading to smaller end-to-end distances. Projecting the PMF onto the Ree coordinate, the folded, intermediate (molten-globule) and unfolded states overlap to form a “broad” single minimum.

Recently, MacKerrel and co-workers have introduced a revised version of the charmm36 force-field, and coined it “charmm36m”⁷² to reduce the overstabilization of collapsed states, a problem that generally plagues the existing force-fields. To check the robustness of our result for GB1 we repeated free energy computations in the absence of force using the revised “charmm36m” force-field. Figure S3 compares respective free energy profiles obtained using the “charmm36” and “charmm36m”, showing that the PMFs are quite similar.

The free energy surface suggests that protein folding proceeds by passing through a metastable intermediate state known as the molten globule state,^{13,22} in which the complete tertiary structure is not yet achieved.^{73–76} Our explicit solvent

simulation allowed us to explore the role of water in the stabilization of intermediates. Figure 2A shows the free energy surface as a function of solvent accessible surface area (SASA) of the protein and ρ . Quite expectedly, we find that the global minimum in the free energy corresponds to a high fraction of native contacts, and a small SASA (characteristic of the folded state), while the higher free energy configurations correspond to a very low fraction of native contacts, and a significantly higher SASA (characteristic of unfolded or extended states). We find that the surface area rapidly decreases when the fraction of native contacts increases from $\rho \sim 0$ to $\rho \sim 0.3$, but then decreases significantly more slowly in the intermediate range of ρ (0.3–0.6). Interestingly, this behavior of the SASA is uncharacteristic of a two state folding/unfolding description. A clearer interpretation of the stabilized intermediates is represented in Figure 2B, which shows the free energy profile versus the number of water molecules and ρ , taking only configurations with radius of gyration < 1.5 nm (i.e., excluding the extended conformation). This illustration suggests that the so-called molten globular intermediate (corresponding to $\rho = 0.65$) contains a significant number of water molecules in the protein core, as opposed to the dry protein core in the folded configuration of the protein (hence the global minima). Note that the free energy as a function of end-to-end distance does

not properly characterize the size of the protein, which is generally represented by the radius of gyration. The projection of the free energy landscape along the radius of gyration and ρ (see Figure S9 in Supporting Information) shows that the molten globule conformation is larger in size relative to the more compact native folded conformation, and swells due to the presence of water molecules. The molten-globule state, indicated by the intermediate values of ρ , shown by the time-averaged density profile of water, is stabilized by the interactions of protein core and water molecules present inside the core.

Unfolding Kinetics. How do the various states described by the 2-d PMF surface manifest themselves in constant force (i.e., force-clamp) experiments? Performing BD simulations (consistent with the Smoluchowski equation) on the smoothed 2-d force dependent PMF surfaces (cf. Figures 1B, S3B, S4B and S5) allows us to study how the application of force affects the unfolding and refolding kinetics (see Supporting Information for details and for a discussion on the effect of force on the PMF surfaces).

In more elaborate simulations, it would be possible, but difficult, to use umbrella sampling methods to determine accurate and position dependent diffusion tensors,^{77,78} something we hope to do in the future to better understand the sensitivity of the results to the diffusion tensor. The external force protocol used for the BD simulation was applied as follows: initially, zero force is set for 0.175 s, it is then instantly increased to 100 pN (Figure 3A) for 0.225 s and then quenched to 15 pN (Figure 3B). This allowed us to probe both unfolding (from low to high forces), and refolding (from high to low forces). Figures 3C and 3D shows the dynamical behavior of the trajectories projected onto the ρ and Ree coordinates respectively over the whole time span of the simulations.

In Figure 3C, we see that in the zero force case fluctuations are observed in the folded-native state ($\rho \sim 0.972$, $Ree = 2.87$), with infrequent thermally activated transitions to the partially folded molten-globule state, which can be detected only at $\rho \sim 0.65$. In the simulations when the force is instantaneously changed to 100 pN, the PMF is also changed instantaneously (in contrast to real AFM experiments, where the force is changed quickly but on a finite time scale and the PMF also changes on a finite time scale). Concomitantly, the protein unfolds and stretches to the unfolded-extended conformation with $\rho = 0.125$, with almost no native contacts remaining. In this conformation, $Ree = 17.89$ nm, which agrees with the prediction from the worm-like chain (WLC) model of polymer entropic elasticity⁷⁹ at 100 pN (with a persistence length, $l_p = 0.381$ nm). The refolding process begins to take place with a recoil followed by the collapse of the unfolded protein chain once the force is instantaneously reduced to 15 pN. The apparent jump observed at the beginning of the 15 pN stage appears because the position of the unfolded-extended state at 100 pN does not overlap with the low force position. After the initial elastic recoil, the chain relaxes to a collapsed state ($\rho \sim 0.4$, $Ree \sim 5$ to 10 nm), where it resides until it stochastically collapses to the molten-globule state ($\rho \sim 0.65$, $Ree \sim 2.2$ nm), from which it folds ending up in the state ($\rho \sim 0.98$, $Ree = 2.84$). From the BD simulations, it is evident that in the absence of force, occasional thermally induced transitions occur between the folded and intermediate states. When refolding against 15 pN, it is evident that the probability to populate the intermediate state increases, as this state expands due to the

application of force. Such a similar behavior was recently shown for a cold-shock protein.⁸⁰ Although GB1 was not measured in the force-clamp mode (as was its computationally designed fast-folding mutant, NuG2⁸¹), its simulated unfolding and refolding pattern along the Ree coordinate (Figure 3D) remarkably resembles actual force-clamp experiments on ubiquitin and I91.^{65,66} In these experiments, the unfolding is characterized by a typical “step” length, and refolding upon diminishing or totally quenching the forces, is observed through an initial recoiling to the collapsed state, from which folding takes place to the native state. Generally, a single macromolecule can undergo a cooperative, first-order-like transition under the application of force from a stretched coil into a collapsed globule.⁸²

As already mentioned, experimental data from force spectroscopy measurements has been interpreted on the basis of a two-state state unfolding model and analyzed using Kramers theory for the diffusion of a single Brownian particle over a one-dimensional energy barrier.^{5,83,84} According to this theory, the rate of crossing a potential barrier, ΔE_i , with a diffusion coefficient D_i is given by

$$k_i^{1-d} = \frac{(K_i^U K_i^N)^{1/2}}{2\pi k_B T} D_i e^{-\Delta E_i/k_B T} = k_0 e^{-\Delta E_i/k_B T} \quad (7)$$

where the index i denotes the specific reaction coordinate on which the dynamics takes place, K_i^U and K_i^N are given by the second derivatives of the potential at the minimum preceding the barrier and at the maximum of the barrier respectively, k_B is Boltzmann's constant, T is the absolute temperature, and k_0 is the Arrhenius factor or attempt rate for crossing the unperturbed barrier. In practice, it has become common to interpret force dependent experiments by assuming that the energy barrier depends linearly on the applied force such that $\Delta E(F) \sim \Delta E^0(1 - F\Delta x/\Delta E^0)$,^{85,86} where ΔE^0 is the unperturbed (zero force) intrinsic energy barrier, and Δx denotes the transition length (the distance from the minimum of the folded state to the maximal height of the unfolding barrier). A different approach^{87–89} describes the height of the barrier in terms of a reduced biasing force, $\Delta E(F) \sim \Delta E^0(1 - F/F_C)^{3/2}$, where F cannot be greater than the force F_C at which the barrier vanishes (typically taken as $\Delta E^0/\Delta x$). These two different approaches can be expressed in the general form $\Delta E(F) \sim \Delta E^0(1 - F\Delta x/\Delta E^0)^n$, where $n = 1$ accounts for the phenomenological approach and $n = 3/2$ for the high force approximation (see Supporting Information). For a thermally activated high damping dynamics, the rate of crossing the transition barrier is accordingly,⁸⁷

$$k(F) = k_0 \left(1 - \frac{F\Delta x}{\Delta E_0}\right)^{n-1} \exp\left[-\frac{\Delta E_0}{k_B T} \left(1 - \frac{F\Delta x}{\Delta E_0}\right)^n\right] \quad (8)$$

Fitting experimental data to eq 5 determines k_0 in addition to ΔE_0 and Δx .^{12,22,32} The fitted values thus represent the parameters of the one-dimensional potential projected onto the end-to-end length coordinate, Ree .

Alternatively let us consider kinetic behavior of the protein under force³⁴ for the two-dimensional reaction coordinate (Ree , ρ).^{30,90} Langer's theory for the rate of diffusively crossing a barrier on a multidimensional surface^{35,91,92} is

$$k_L = \frac{1}{2\pi k_B T} \left(\frac{\det(\underline{K}_U)}{|\det(\underline{K}_\rho)|} \right)^{1/2} \lambda e^{-\Delta E_0/k_B T} \quad (9)$$

here \underline{K}_U and \underline{K}_ρ are the curvatures of the potential, given by the matrices of the second derivatives of the PMF at the minimum of the folded state that precede the saddle point and at the peak of the unfolding barrier respectively, and λ is the positive root of $\det(\lambda \underline{I} + \underline{K}_U \underline{D})$, with \underline{D} being the diagonal diffusion tensor. If *Ree* is a faster coordinate than ρ , as in our BD simulations, it would be necessary to use the multidimensional Langer theory otherwise eq 8 would be sufficient. The multidimensional k_L can be related to k_0 when D_{Ree} is at least ten times larger than D_ρ by $k_0^{-1} = (k_L^{-1} + k_\rho^{-1})$,^{35,92} where k_ρ is given by eq 7 with $i = \rho$.

We compare our simulated results with the predictions of these models. First, we determine the intrinsic properties, Δx and ΔE^0 directly from the unperturbed PMF surface, and compared them to the values fitted by eq 8. Unfolding probabilities were calculated from the simulated unfolding trajectories at several high forces ($F = 50, 70, 100$, and 150 pN), and then fitted to single exponentials⁶⁴ thereby yielding approximate unfolding rate constants corresponding to each applied force (Figure 4A). The force-dependent unfolding rates were fitted using eq 8 thereby to obtain corresponding k^U vs F . (Figure 4B).

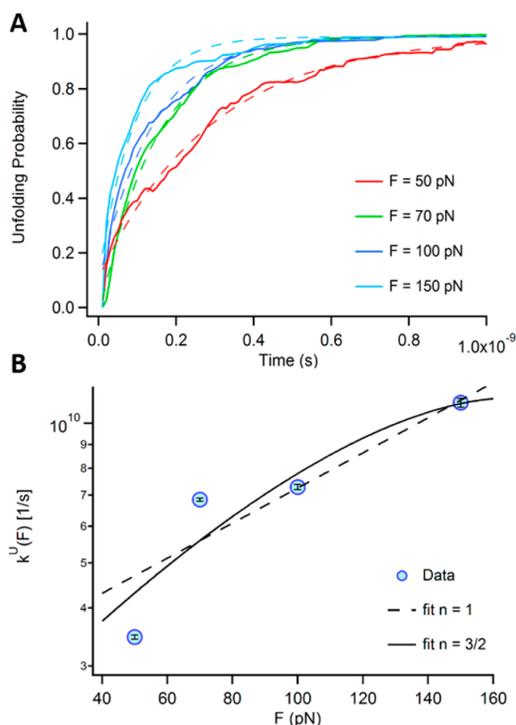


Figure 4. Unfolding at high force. (A) Unfolding probabilities calculated from trajectories that transitioned between zero force to 50, 70, 100, and 150 pN (solid curves). The unfolding probabilities were fitted with a single exponential to estimate the transition rate at the given force (dashed curves). (B) Unfolding force dependency fitted with eq 8 for $n = 1$ and $n = 3/2$ approximations of the unfolding rates to estimate the potential parameters.

Table 1 summarizes the values of Δx , ΔE^0 and k_0 that were obtained directly from the 2-d PMF surface and alternatively by fitting to the rate theories. The values of ΔE^0 and Δx were directly measured from the unperturbed 2-d PMF surface, while the unfolding rate at zero force was calculated from 700 simulated trajectories using the relation $k_0 = -\tau^{-1} \ln(1 - P_0)$. P_0 is the probability to transition to the unfolded state and is determined from the dwell-time distribution measured from transitional “hopping” across the unfolding barrier of the unperturbed PMF (which can be observed from the initial part of the trajectories at zero force shown in Figure 3C), and τ is the total time over which each trajectory is determined. The 1-d data were obtained by fitting the force dependent rates (Figure 4B) with eq 8 for the for the phenomenological linear force dependency, where $n = 1$,^{85,86} and for the power law dependency with $n = 3/2$.^{87–89} The 2-d unfolding rate was calculated using eq 9 and eq 7. The off-rate using Langer’s theory was calculated by combining $k_L = 1.53 \times 10^7$ s⁻¹ that was determined from eq 9 using the parameters measured directly from the unperturbed 2-d PMF, together with the value $k_\rho = 9.05 \cdot 10^5$ s⁻¹ determined from eq 7 (with $i = \rho$) according to $k_0^{-1} = (k_L^{-1} + k_\rho^{-1})$.^{35,92}

The 1-d models with different n underestimate ΔE_0 by ~35%, and Δx by an order of magnitude, and moreover predict an unfolding rate constant 5 orders of magnitude larger than found from the simulations. This discrepancy can result from several factors, including reduced dimensionality as well as the absence of coordinate dependence of the diffusion coefficient. This means that for some cases, like the one investigated here, the kinetic data determined from *Ree* alone could lead to inexact results, and that the pulling coordinate should not be taken as the sole reaction coordinate.^{26,35,92,93} eq 8 was formulated under high damping assumption around the inflection point between the minimum that precedes the energy barrier and its maximum.⁸⁷ Around the inflection point, where the energy barrier vanishes at sufficiently high forces, the conformational distributions evolve differently compared to the low (or none) force regime, where ΔE_0 is intact and can have a large value. Therefore, due to this considerable variation, the extrapolated value of k_0 (at zero force) can yield misleading results,^{93,94} and also lead to unreliable force dependent unfolding pathways^{95–99} (see also Supporting Information for a discussion on the force dependency of the unfolding pathways). On the other hand, calculating k_0 from eq 9 for the two-dimensional reaction coordinate resulted in a value of k_0 of the same order of magnitude as determined from simulations on the unperturbed PMF surface. Although the one-dimensional kinetic models were incapable of accounting for the rate constants found by simulation, one should bear in mind that the dynamical representation presented in this work did not include other important experimental features that have to be considered, such as the effect of linkers,³⁵ hydrodynamic effects resulting from the tethering,¹⁰⁰ the nonzero conformational relaxation times following changes in force, etc. Each one of these features might have a large effect on the data interpretation.

Collapse and Refolding from the Extended State.

Unfolding from relatively well-defined initial conditions in the native state can be described as a two-state process, but refolding displays a more complex behavior^{65,66,101} and can be initiated in various unfolded conformations (particularly in force spectroscopy experiments). Figure 5A shows the time evolution of a BD simulated refolding traces from high-applied

Table 1. Intrinsic Parameters Measured from the Unperturbed PMF and Their Corresponding 1-d Calculated Values

	Δx (nm)	$\Delta E^0(kT)$	k_0 (s^{-1})
unperturbed PMF	0.30	18.48	2.02×10^5
(1-d) eq 8 with $n = 1$	0.04 ± 0.01	12.1 ± 9.2	$5.7 \times 10^{10} \pm 1.3 \times 10^{16}$
(1-d) eq 8 with $n = 3/2$	0.05 ± 0.04	11.4 ± 3.4	$3.2 \times 10^{10} \pm 1.2 \times 10^{10}$
(2-d) eq 7 and eq 9	–	–	8.54×10^5

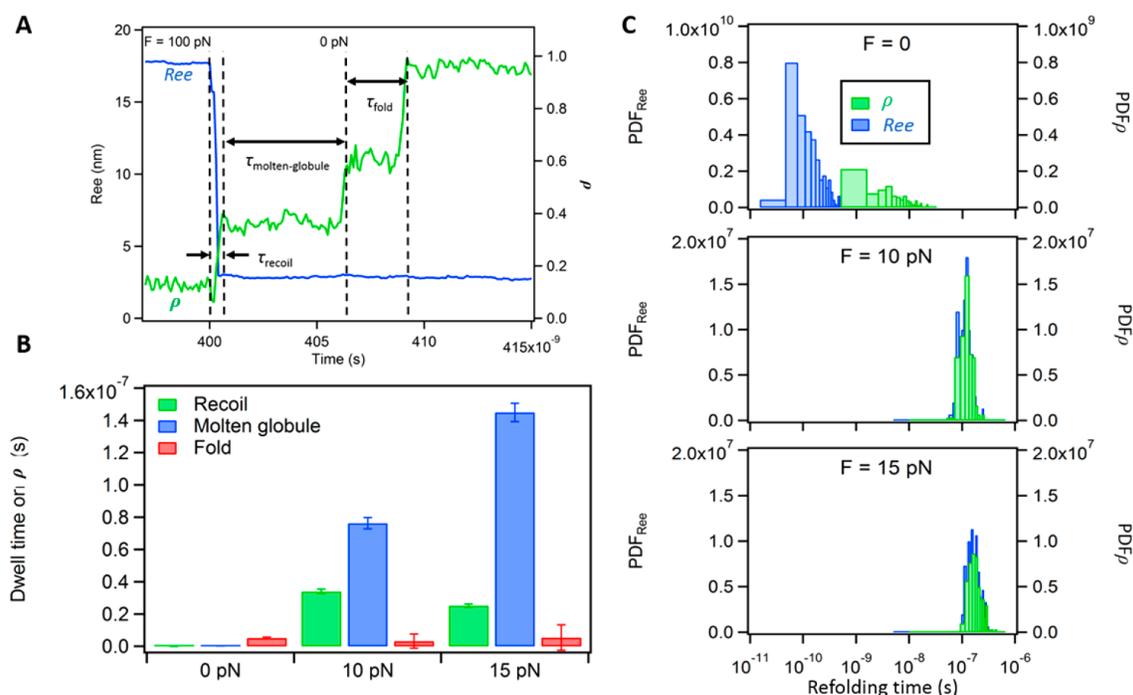


Figure 5. Effect of force on refolding from high extension. (A) Definition of the dwell times for each of the collapse/molten-globule/folded states when the quenching from highly extended conformation at high force (100 pN) to zero force for Ree (blue) and ρ (green). (B) Dwell times separation along the ρ coordinate at each conformational state during the quenching from 100 pN down to 0, 10 pN and 15 pN. Here the effect of the force on the refolding process is mostly observed through the dwell time changes during the initial recoiling and the transition between the unfolded state and the molten globule state, indicating the deepening of the latter with increasing the force. (C) Overall refolding time distributions at three quenching forces. Quenching to zero force shows a time separation between the Ree and ρ coordinates (blue and green respectively), where the refolding process on Ree is faster by about an order of magnitude than ρ . Increasing the quenching force against which refolding occurs, result with an increase and overlap in the refolding times in both coordinates.

force along Ree and ρ separately. The two trajectories start at high extension (at 100 pN force), and then travel downhill along the zero force 2-d PMF funnel until they reach the folded native state of the protein. As in the force-clamp experiments at forces below F_C ⁶⁶ the actual folding pathway cannot be determined from the relaxation of the Ree coordinate alone (Figure 5A, Ree blue trace). The ρ coordinate, on the other hand, shows that folding takes place from the unfolded states, to the molten-globule state, and then to the folded state (Figure 5A, green trace). These states are represented by the local basins on the free energy surface (PMF). Each basin has a characteristic dwell time distribution at the applied quenching force (defined by the black arrows in Figure 5A) and these characterize the time dependent population of the different states (Figure 5B). At zero force, the unperturbed PMF has the steepest gradient with Ree , in the extended state, which gives rise to a rapid average refolding time during which the protein explores the whole range for the end-to-end distance. The time evolution of the ρ coordinate, however, shows a multistage transition between the conformational states along the energy surface. Refolding from 100 pN against low forces of 10

and 15 pN is accompanied by a variation in the average dwell-times compared to refolding in the absence of force. The average dwell-times in each state are shown in Figure 5B, for different final forces. These dwell-times were divided into three transitional regimes: τ_{recoil} , the time for initial entropic contraction from high extension to the unfolded coiled state; $\tau_{molten-globule}$, the time spent in the unfolded coiled before reaching the molten-globule state for the first time; and τ_{fold} , the time spent in the molten-globule state before the crossing the final barrier to reach the native folded state. Unlike the dwell times in the recoil and molten-globule stages, the $\langle \tau_{fold} \rangle^p$ does not appear to show distinct variation with the quenching force. On the other hand, $\langle \tau_{recoil} \rangle^p$, and $\langle \tau_{molten-globule} \rangle^p$ are dramatically increased as quenching forces increase from F = 0 to 10 and 15 pN (see Supporting Information for details). The gradient from the extended state (at 100 pN) to the molten globule state at zero force becomes less steep with the increase of the quenching force to 10 and 15 pN. As a result, the recoil dwell times increase. Likewise, the dwell time spent in the molten globule state increase, as this state gets deeper with the applied quenching force. The barrier separating the molten

globule state from the folded state, however, scarcely changes with the quenching force, hence $\langle \tau_{\text{fold}} \rangle^{\rho}$ hardly changes.

The cooperative effect of the route in the $Ree-\rho$ plane is portrayed through the overall refolding time distributions (Figure 5C), which sums all the dwell time contributions described above. The BD simulated refolding traces from 100 to 0 pN showed a distinct time separation where the refolding time on Ree is faster by about an order of magnitude than on ρ . The downhill entropic driven motion on the length coordinates happens rapidly with the removal of the force while bonds are formed. Once the protein reached a local coiled conformation, bonds begin to form, and the motion takes place on the ρ orthogonal coordinate with very little change in Ree . This motion involves crossing energy barriers, and transient residing in local states over the PMF, which slows the dynamics down. However, when the force is quenched down to lower values, rather than being completely removed, the overall refolding times increase and coincide. The elongation in the overall refolding time is an outcome of the alteration in the morphological shape of the PMF under the application of force.³⁴ As indicated in Figure 5B, the presence of the low forces alters the PMFs by expanding the basins of the unfolded and molten globule states (see Supporting Information, Figures S5 and S6) to such extent that their corresponding dwell times exhibits similar duration on both coordinates.¹⁰²

CONCLUSIONS

In this paper we show that calculating two-dimensional PMF surfaces for proteins under force using Molecular Dynamics simulations and then generating Brownian dynamic paths on these surfaces provides a useful tool for understanding of the folding and unfolding pathways of a biomolecule under the application of an external force. In this paper, we determined the free energy surface (PMF) of GB1 protein from atomistic simulations in explicit solvent under different forces as a function of two collective variables: the end-to-end-distance, Ree , and the fraction of native contacts, ρ . The calculated 2-d free energy surfaces from these simulations showed several distinct states, or basins, mostly visible along the ρ coordinate. The presence of these states, recently pointed out experimentally, support the conclusion that GB1 is not a two-state folder.⁴⁰ Single molecule force spectroscopy measures the evolution over the single Ree coordinate, and thus can be insensitive to the transitioning between these states, as they overlay each other while being projected onto the Ree coordinate. BD simulations on the force dependent 2-d PMF surfaces, gives reasonable agreement with generic patterns observed in force-clamp experiments,^{65,66} and readily demonstrates the different unfolding/refolding pathways. When ρ is taken as the slower coordinate, reconstruction of the PMF from pulling experiments can result in an energy profile that only partially represents the kinetics and thermodynamics of biomolecule. However, careful interpretation of the experimental data can reproduce the actual features of interest.^{35,103,104} Our current work have specifically established that the approach of exploring native fraction as a collective variable^{26,42} in addition to end-to-end distance, has great potential in unraveling the interesting intermediates of protein in its pathway to folded structure both in the presence of a constant force and absence of it. The atomistic nature of the models makes the approach more accurate and straightforward to compare with experiments. Our simulation approach is practical and easy to implement using existing simulation

package. We note that in principle it would be possible to use umbrella sampling methodology to determine the position dependent diffusion tensors^{77,78} (which we plan to explore in the future), nevertheless, determining the optimum set of collective variable capable of describing protein thermodynamics and dynamics remains an open question in chemical physics. Some progress has been made recently using a method called SGOOP (spectral gap optimization of order parameters).^{105,106} The 1-d kinetic model of eq 6 based on Ree as the reaction coordinate, could in principle be improved by choosing a one-dimensional reaction coordinate $f(Ree, \rho)$ optimized using SGOOP in a Brownian dynamics simulation. In the future, we plan to apply such methods to check the robustness of our current choice of collective variables.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpbc.7b00610.

Details on the umbrella sampling, PMF adaptation, kinetic models from the thermodynamics of the PMF surfaces, BD simulations, unfolding pathways, and refolding pathways(PDF)

AUTHOR INFORMATION

Corresponding Authors

*(R.B.) E-mail: berkovir@bgu.ac.il.

*(J.M.) E-mail: jmondal@tifrh.res.in.

*(B.J.B.) E-mail: bb8@columbia.edu.

ORCID

Ronen Berkovich: 0000-0002-0989-6136

Author Contributions

[†]These authors contributed equally.

Author Contributions

R.B., J.M., and B.J.B. contributed to the design of the research; J.M. and B.J.B. performed and analyzed the molecular dynamics simulations. I.P. numerically extrapolated the 3d PMFs. R.B. performed and analyzed the Brownian dynamics simulations. R.B., J.M., and B.J.B. contributed to the writing of the paper.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to thank M. Urbakh for useful discussions. This work was supported by grants to B.J.B. from the National Institutes of Health [NIH-GM4330] and the Extreme Science and Engineering Discovery Environment (XSEDE) [TG-MCA08 × 002]. J.M. would like to acknowledge financial support from Tata Institute of Fundamental research and DST funding under Ramanujan fellowship program and early career research program, R.B. acknowledge the financial support of the I-CORE Program of the Planning and Budgeting Committee and The Israel Science Foundation (Grant No. 152/11).

ABBREVIATIONS

PMF, potential of mean force; BD, Brownian dynamics; AFM, atomic force microscopy; SMFS, single molecule force spectroscopy experiments; GB1, B1 segment of streptococcal Protein G; MD, molecular dynamics; WHAM, weighted histogram analysis method; SASA, solvent accessible surface area

REFERENCES

- (1) Tsai, C. J.; Ma, B. Y.; Sham, Y. Y.; Kumar, S.; Nussinov, R. Structured Disorder and Conformational Selection. *Proteins: Struct., Funct., Genet.* **2001**, *44* (4), 418–427.
- (2) Henzler-Wildman, K.; Kern, D. Dynamic Personalities of Proteins. *Nature* **2007**, *450* (7172), 964–972.
- (3) Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **2012**, *338* (6110), 1042–1046.
- (4) Levinthal, C. Are There Pathways for Protein Folding? *J. Chim. Phys. Phys.-Chim. Biol.* **1968**, *65* (1), 44–45.
- (5) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, Pathways and the Energy Landscape of Protein Folding: A Synthesis. *Proteins: Struct., Funct., Genet.* **1995**, *21* (3), 167–195.
- (6) Dill, K. A.; Chan, H. S. From Levinthal to Pathways to Funnels. *Nat. Struct. Biol.* **1997**, *4* (1), 10–19.
- (7) McLeish, T. C. B. Protein Folding in High-Dimensional Spaces: Hypergutters and the Role of Nonnative Interactions. *Biophys. J.* **2005**, *88* (1), 172–183.
- (8) Onuchic, J. N.; Wolynes, P. G. Theory of Protein Folding. *Curr. Opin. Struct. Biol.* **2004**, *14* (1), 70–75.
- (9) Bryngelson, J. D.; Wolynes, P. G. Intermediates and Barrier Crossing in a Random Energy Model (with Applications to Protein Folding). *J. Phys. Chem.* **1989**, *93* (19), 6902–6915.
- (10) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. Reversible Unfolding of Individual Titin Immunoglobulin Domains by Afm. *Science* **1997**, *276* (5315), 1109–1112.
- (11) Yang, G. L.; Ceconi, C.; Baase, W. A.; Vetter, I. R.; Breyer, W. A.; Haack, J. A.; Matthews, B. W.; Dahlquist, F. W.; Bustamante, C. Solid-State Synthesis and Mechanical Unfolding of Polymers of T4 Lysozyme. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (1), 139–144.
- (12) Borgia, A.; Williams, P. M.; Clarke, J. Single-Molecule Studies of Protein Folding. *Annu. Rev. Biochem.* **2008**, *77*, 101–125.
- (13) Hofmann, H.; Nettels, D.; Schuler, B. Single-Molecule Spectroscopy of the Unexpected Collapse of an Unfolded Protein at Low Ph. *J. Chem. Phys.* **2013**, *139* (12), 121930.
- (14) Sheinerman, F. B.; Brooks, C. L. Calculations on Folding of Segment B1 of Streptococcal Protein G. *J. Mol. Biol.* **1998**, *278* (2), 439–456.
- (15) Zhou, R. H.; Berne, B. J.; Germain, R. The Free Energy Landscape for Beta Hairpin Folding in Explicit Water. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (26), 14931–14936.
- (16) Levy, Y.; Papoian, G. A.; Onuchic, J. N.; Wolynes, P. G. Energy Landscape Analysis of Protein Dimers. *Isr. J. Chem.* **2004**, *44* (1–3), 281–297.
- (17) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334* (6055), 517–520.
- (18) Marszalek, P. E.; Lu, H.; Li, H. B.; Carrion-Vazquez, M.; Oberhauser, A. F.; Schulten, K.; Fernandez, J. M. Mechanical Unfolding Intermediates in Titin Modules. *Nature* **1999**, *402* (6757), 100–103.
- (19) Carrion-Vazquez, M.; Oberhauser, A. F.; Fowler, S. B.; Marszalek, P. E.; Broedel, S. E.; Clarke, J.; Fernandez, J. M. Mechanical and Chemical Unfolding of a Single Protein: A Comparison. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (7), 3694–3699.
- (20) Dietz, H.; Rief, M. Exploring the Energy Landscape of Gfp by Single-Molecule Mechanical Experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (46), 16192–16197.
- (21) Neuman, K. C.; Nagy, A. Single-Molecule Force Spectroscopy: Optical Tweezers, Magnetic Tweezers and Atomic Force Microscopy. *Nat. Methods* **2008**, *5* (6), 491–505.
- (22) Elms, P. J.; Chodera, J. D.; Bustamante, C.; Marqusee, S. The Molten Globule State Is Unusually Deformable under Mechanical Force. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (10), 3796–3801.
- (23) Bartlett, A. I.; Radford, S. E. An Expanding Arsenal of Experimental Methods Yields an Explosion of Insights into Protein Folding Mechanisms. *Nat. Struct. Mol. Biol.* **2009**, *16* (6), 582–588.
- (24) Stirnemann, G.; Kang, S.-g.; Zhou, R.; Berne, B. J. How Force Unfolding Differs from Chemical Denaturation. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (9), 3413–3418.
- (25) Berkovich, R.; Garcia-Manyes, S.; Klafter, J.; Urbakh, M.; Fernandez, J. M. Hopping around an Entropic Barrier Created by Force. *Biochem. Biophys. Res. Commun.* **2010**, *403* (1), 133–137.
- (26) Dudko, O. K.; Graham, T. G. W.; Best, R. B. Locating the Barrier for Folding of Single Molecules under an External Force. *Phys. Rev. Lett.* **2011**, *107* (20), 208301.
- (27) Brooks, C. L.; Gruebele, M.; Onuchic, J. N.; Wolynes, P. G. Chemical Physics of Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (19), 11037–11038.
- (28) Dobson, C. M.; Sali, A.; Karplus, M. Protein Folding: A Perspective from Theory and Experiment. *Angew. Chem., Int. Ed.* **1998**, *37* (7), 868–893.
- (29) Zhou, R. H. Trp-Cage: Folding Free Energy Landscape in Explicit Water. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (23), 13280–13285.
- (30) Best, R. B.; Hummer, G. Reaction Coordinates and Rates from Transition Paths. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (19), 6732–6737.
- (31) Yang, S.; Onuchic, J. N.; Levine, H. Effective Stochastic Dynamics on a Protein Folding Energy Landscape. *J. Chem. Phys.* **2006**, *125* (5), 054910.
- (32) Morfill, J.; Neumann, J.; Blank, K.; Steinbach, U.; Puchner, E. M.; Gottschalk, K.-E.; Gaub, H. E. Force-Based Analysis of Multidimensional Energy Landscapes: Application of Dynamic Force Spectroscopy and Steered Molecular Dynamics Simulations to an Antibody Fragment-Peptide Complex. *J. Mol. Biol.* **2008**, *381* (5), 1253–1266.
- (33) Hatch, H. W.; Stillinger, F. H.; Debenedetti, P. G. Computational Study of the Stability of the Miniprotein Trp-Cage, the Gb1 B-Hairpin, and the Ak16 Peptide, under Negative Pressure. *J. Phys. Chem. B* **2014**, *118* (28), 7761–7769.
- (34) Sun, L.; Noel, J. K.; Sulkowska, J. I.; Levine, H.; Onuchic, J. N. Connecting Thermal and Mechanical Protein (Un)Folding Landscapes. *Biophys. J.* **2014**, *107* (12), 2950–2961.
- (35) Cossio, P.; Hummer, G.; Szabo, A. On Artifacts in Single-Molecule Force Spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (46), 14248–14253.
- (36) Gronenborn, A. M.; Filipula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. A Novel Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein G. *Science* **1991**, *253* (5020), 657–661.
- (37) McCallister, E. L.; Alm, E.; Baker, D. Critical Role of [Beta]-Hairpin Formation in Protein G Folding. *Nat. Struct. Biol.* **2000**, *7* (8), 669–673.
- (38) Cao, Y.; Li, H. Polypeptide of Gb1 Is an Ideal Artificial Elastomeric Protein. *Nat. Mater.* **2007**, *6* (2), 109–114.
- (39) Chung, H. S.; Louis, J. M.; Eaton, W. A. Experimental Determination of Upper Bound for Transition Path Times in Protein Folding from Single-Molecule Photon-by-Photon Trajectories. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (29), 11837–11844.
- (40) Morrone, A.; Giri, R.; Toofanny, R. D.; Travaglini-Allocatelli, C.; Brunori, M.; Daggett, V.; Gianni, S. Gb1 Is Not a Two-State Folder: Identification and Characterization of an on-Pathway Intermediate. *Biophys. J.* **2011**, *101* (8), 2053–2060.
- (41) Li, Y. D.; Lamour, G.; Gsponer, J.; Zheng, P.; Li, H. The Molecular Mechanism Underlying Mechanical Anisotropy of the Protein Gb1. *Biophys. J.* **2012**, *103* (11), 2361–2368.
- (42) Sheinerman, F. B.; Brooks, C. L. Molecular Picture of Folding of a Small A/B Protein. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (4), 1562–1567.
- (43) Shimada, J.; Shakhnovich, E. I. The Ensemble Folding Kinetics of Protein G from an All-Atom Monte Carlo Simulation. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (17), 11175–11180.
- (44) Li, P. C.; Huang, L.; Makarov, D. E. Mechanical Unfolding of Segment-Swapped Protein G Dimer: Results from Replica Exchange Molecular Dynamics Simulations. *J. Phys. Chem. B* **2006**, *110* (29), 14469–14474.

- (45) Glyakina, A. V.; Balabaev, N. K.; Galzitskaya, O. V. Mechanical Unfolding of Proteins L and G with Constant Force: Similarities and Differences. *J. Chem. Phys.* **2009**, *131* (4), 045102.
- (46) Best, R. B.; Hummer, G. Diffusion Models of Protein Folding. *Phys. Chem. Chem. Phys.* **2011**, *13* (38), 16902–16911.
- (47) Park, S.-H.; O'Neil, K. T.; Roder, H. An Early Intermediate in the Folding Reaction of the B1 Domain of Protein G Contains a Native-Like Core. *Biochemistry* **1997**, *36* (47), 14277–14283.
- (48) Park, S. H.; Shastray, M. C. R.; Roder, H. Folding Dynamics of the B1 Domain of Protein C Explored by Ultrarapid Mixing. *Nat. Struct. Biol.* **1999**, *6* (10), 943–947.
- (49) Kmiecik, S.; Kolinski, A. Folding Pathway of the B1 Domain of Protein G Explored by Multiscale Modeling. *Biophys. J.* **2008**, *94* (3), 726–736.
- (50) Graham, T. G. W.; Best, R. B. Force-Induced Change in Protein Unfolding Mechanism: Discrete or Continuous Switch? *J. Phys. Chem. B* **2011**, *115* (6), 1546–1561.
- (51) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive Charmm All-Atom Protein Force Field Targeting Improved Sampling of the Backbone Φ , Ψ and Side-Chain X(1) and X(2) Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8* (9), 3257–3273.
- (52) 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* (2), 926–935.
- (53) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23* (2), 187–199.
- (54) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; et al. Plumed: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. *Comput. Phys. Commun.* **2009**, *180* (10), 1961–1972.
- (55) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. Gromacs 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447.
- (56) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I: The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021.
- (57) Nose, S. Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52* (2), 255–268.
- (58) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A: At, Mol., Opt. Phys.* **1985**, *31* (3), 1695–1697.
- (59) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190.
- (60) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N-Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.
- (61) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13* (8), 952–962.
- (62) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. Lincs: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472.
- (63) Ermak, D. L.; McCammon, J. A. Brownian Dynamics with Hydrodynamic Interactions. *J. Chem. Phys.* **1978**, *69* (4), 1352–1360.
- (64) Schlierf, M.; Li, H. B.; Fernandez, J. M. The Unfolding Kinetics of Ubiquitin Captured with Single-Molecule Force-Clamp Techniques. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (19), 7299–7304.
- (65) Fernandez, J. M.; Li, H. B. Force-Clamp Spectroscopy Monitors the Folding Trajectory of a Single Protein. *Science* **2004**, *303* (5664), 1674–1678.
- (66) Berkovich, R.; Garcia-Manyses, S.; Urbakh, M.; Klafter, J.; Fernandez, J. M. Collapse Dynamics of Single Proteins Extended by Force. *Biophys. J.* **2010**, *98* (11), 2692–2701.
- (67) Nettels, D.; Gopich, I. V.; Hoffmann, A.; Schuler, B. Ultrafast Dynamics of Protein Collapse from Single-Molecule Photon Statistics. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (8), 2655–2660.
- (68) Hyeon, C.; Thirumalai, D. Mechanical Unfolding of Rna Hairpins. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (19), 6789–6794.
- (69) Hyeon, C.; Thirumalai, D. Forced-Unfolding and Force-Quench Refolding of Rna Hairpins. *Biophys. J.* **2006**, *90* (10), 3410–3427.
- (70) Hyeon, C. T.; Thirumalai, D. D., Measuring the Energy Landscape Roughness and the Transition State Location of Biomolecules Using Single Molecule Mechanical Unfolding Experiments. *J. Phys.: Condens. Matter* **2007**, *19* (11), 113101.
- (71) Hyeon, C.; Thirumalai, D. Multiple Probes Are Required to Explore and Control the Rugged Energy Landscape of Rna Hairpins. *J. Am. Chem. Soc.* **2008**, *130* (5), 1538–1539.
- (72) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr Charmm36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14* (1), 71–73.
- (73) Ohgushi, M.; Wada, A. 'Molten-Globule State': A Compact Form of Globular Proteins with Mobile Side-Chains. *FEBS Lett.* **1983**, *164* (1), 21–24.
- (74) Pitsyn, O. B. Molten Globule and Protein Folding. *Adv. Protein Chem.* **1995**, *47* (47), 83–229.
- (75) Pande, V. S.; Rokhsar, D. S. Is the Molten Globule a Third Phase of Proteins? *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (4), 1490–1494.
- (76) Sherman, E.; Haran, G. Coil-Globule Transition in the Denatured State of a Small Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (31), 11539–11543.
- (77) Straub, J. E.; Berne, B. J.; Roux, B. Spatial Dependence of Time-Dependent Friction for Pair Diffusion in a Simple Fluid. *J. Chem. Phys.* **1990**, *93* (9), 6804–6812.
- (78) Hummer, G. Position-Dependent Diffusion Coefficients and Free Energies from Bayesian Analysis of Equilibrium and Replica Molecular Dynamics Simulations. *New J. Phys.* **2005**, *7* (1), 34.
- (79) Marko, J. F.; Siggia, E. D. Stretching DNA. *Macromolecules* **1995**, *28* (26), 8759–8770.
- (80) de Sancho, D.; Best, R. B. Reconciling Intermediates in Mechanical Unfolding Experiments with Two-State Protein Folding in Bulk. *J. Phys. Chem. Lett.* **2016**, *7* (19), 3798–3803.
- (81) Cao, Y.; Kuske, R.; Li, H. Direct Observation of Markovian Behavior of the Mechanical Unfolding of Individual Proteins. *Biophys. J.* **2008**, *95* (2), 782–788.
- (82) Halperin, A.; Zhulina, E. B. On the Deformation Behaviour of Collapsed Polymers. *EPL (Europhysics Letters)* **1991**, *15* (4), 417.
- (83) Kramers, H. A. Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions. *Physica* **1940**, *7* (4), 284–304.
- (84) Klimov, D. K.; Thirumalai, D. Viscosity Dependence of the Folding Rates of Proteins. *Phys. Rev. Lett.* **1997**, *79* (2), 317–320.
- (85) Bell, G. I. Models for the Specific Adhesion of Cells to Cells. *Science* **1978**, *200* (4342), 618–627.
- (86) Evans, E.; Ritchie, K. Dynamic Strength of Molecular Adhesion Bonds. *Biophys. J.* **1997**, *72* (4), 1541–1555.
- (87) Garg, A. Escape-Field Distribution for Escape from a Metastable Potential Well Subject to a Steadily Increasing Bias Field. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1995**, *51* (21), 15592–15595.
- (88) Dudko, O. K.; Filippov, A. E.; Klafter, J.; Urbakh, M. Beyond the Conventional Description of Dynamic Force Spectroscopy of Adhesion Bonds. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (20), 11378–11381.
- (89) Dudko, O. K.; Hummer, G.; Szabo, A. Intrinsic Rates and Activation Free Energies from Single-Molecule Pulling Experiments. *Phys. Rev. Lett.* **2006**, *96* (10), 108101.
- (90) Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. Diffusive Dynamics of the Reaction Coordinate for Protein Folding Funnels. *J. Chem. Phys.* **1996**, *104* (15), 5860–5868.
- (91) Langer, J. S. Statistical Theory of the Decay of Metastable States. *Ann. Phys.* **1969**, *54* (2), 258–275.

- (92) Berezhkovskii, A. M.; Szabo, A.; Greives, N.; Zhou, H.-X. Multidimensional Reaction Rate Theory with Anisotropic Diffusion. *J. Chem. Phys.* **2014**, *141* (20), 204106.
- (93) Hyeon, C.; Hinczewski, M.; Thirumalai, D. Extracting Folding Landscape Characteristics of Biomolecules Using Mechanical Forces. *arXiv* **2015**, *1501*, 03448.
- (94) Tshiprut, Z.; Urbakh, M. Exploring Hysteresis and Energy Dissipation in Single-Molecule Force Spectroscopy. *J. Chem. Phys.* **2009**, *130* (8), 084703.
- (95) Wright, C. F.; Lindorff-Larsen, K.; Randles, L. G.; Clarke, J. Parallel Protein-Unfolding Pathways Revealed and Mapped. *Nat. Struct. Biol.* **2003**, *10* (8), 658–662.
- (96) Mickler, M.; Dima, R. I.; Dietz, H.; Hyeon, C.; Thirumalai, D.; Rief, M. Revealing the Bifurcation in the Unfolding Pathways of Gfp by Using Single-Molecule Experiments and Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (51), 20268–20273.
- (97) Sosnick, T. R.; Barrick, D. The Folding of Single Domain Proteins - Have We Reached a Consensus? *Curr. Opin. Struct. Biol.* **2011**, *21* (1), 12–24.
- (98) Kotamarthi, H. C.; Sharma, R.; Narayan, S.; Ray, S.; Ainaravapu, S. R. K. Multiple Unfolding Pathways of Leucine Binding Protein (Lbp) Probed by Single-Molecule Force Spectroscopy (Smfs). *J. Am. Chem. Soc.* **2013**, *135* (39), 14768–14774.
- (99) Zhuravlev, P. I.; Hinczewski, M.; Chakrabarti, S.; Marqusee, S.; Thirumalai, D. Force-Dependent Switch in Protein Unfolding Pathways and Transition-State Movements. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (6), E715–E724.
- (100) Berkovich, R.; Hermans, R. I.; Popa, I.; Stirnemann, G.; Garcia-Manyes, S.; Berne, B. J.; Fernandez, J. M. Rate Limit of Protein Elastic Response Is Tether Dependent. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (36), 14416–14421.
- (101) Hyeon, C.; Morrison, G.; Pincus, D. L.; Thirumalai, D. Refolding Dynamics of Stretched Biopolymers Upon Force Quench. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (48), 20288–20293.
- (102) Thirumalai, D.; O'Brien, E. P.; Morrison, G.; Hyeon, C. Theoretical Perspectives on Protein Folding. *Annu. Rev. Biophys.* **2010**, *39*, 159–183.
- (103) Hinczewski, M.; Gebhardt, J. C. M.; Rief, M.; Thirumalai, D. From Mechanical Folding Trajectories to Intrinsic Energy Landscapes of Biopolymers. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (12), 4500–4505.
- (104) Zhang, Y.; Dudko, O. K. A Transformation for the Mechanical Fingerprints of Complex Biomolecular Interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (41), 16432–16437.
- (105) Tiwary, P.; Berne, B. J. Spectral Gap Optimization of Order Parameters for Sampling Complex Molecular Systems. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (11), 2839–2844.
- (106) Tiwary, P.; Berne, B. J. Kramers Turnover: From Energy Diffusion to Spatial Diffusion Using Metadynamics. *J. Chem. Phys.* **2016**, *144* (13), 134103.