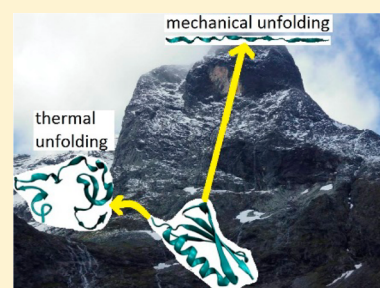


Fully Atomistic Simulations of Protein Unfolding in Low Speed Atomic Force Microscope and Force Clamp Experiments with the Help of Boxed Molecular Dynamics

Jonathan J. Booth and Dmitrii V. Shalashilin*

School of Chemistry, University of Leeds, Leeds LS2 9JT, U.K.

ABSTRACT: The results of boxed dynamics (BXD) fully atomistic simulations of protein unfolding by atomic force microscopy (AFM) in both force clamp (FC) and velocity clamp (VC) modes are reported. In AFM experiments the unfolding occurs on a time scale which is too long for standard atomistic molecular dynamics (MD) simulations, which are usually performed with the addition of forces which exceed those of experiment by many orders of magnitude. BXD can reach the time scale of slow unfolding and sample the very high free energy unfolding pathway, reproducing the experimental dependence of pulling force against extension and extension against time. Calculations show the presence of the pulling force “humps” previously observed in the VC AFM experiments and allow the identification of intermediate protein conformations responsible for them. Fully atomistic BXD simulations can estimate the rate of unfolding in the FC experiments up to the time scale of seconds.



INTRODUCTION

Proteins are parts of various molecular machines where their mechanical properties are crucially important.¹ For example, titin, which has been related to diseases such as heart failure, is a molecular spring playing an important role in muscle functioning.²

In recent years proteins have been extensively studied by various force spectroscopy techniques such as AFM which was used to pull and unfold single chains of protein domains called concatemers. The AFM experiments were motivated by the hope to shed some light on why some protein folds are more mechanically robust than others² via investigation of the mechanical unfolding pathways, which in principle should be possible by comparison of experiments with MD simulations. The problem is that the time scale of a typical experiment is not accessible for standard unbiased fully atomistic molecular dynamics. The constant pulling speed (VC) experiments^{1,3–6} are usually performed with the force in the order of tens or hundreds of pN on a microsecond or even millisecond time scale. The constant force (FC) experiments,^{7–11} where the protein chain is pulled at a constant force and extension as a function of force and time is recorded, are often performed at an even longer time scale of seconds or even minutes, which is of course prohibitive for standard MD. To speed up the MD simulations, they are often performed with unrealistically high pulling speed and force.^{12–14} To put it differently, free energy or potential of mean force (PMF) of protein extension is extremely high, and therefore protein extension is a slow and rare event, very hard to reach by standard MD. In addition, the dynamics of AFM unfolding at low and high pulling speed can be different. At low speed the process is very much stochastic exploring all available regions of the energy landscape, while at high pulling speed it can be dynamical taking a completely

different pathway. Despite significant efforts to extend the time scale of simulations and reduce the experimental time scale by increasing the speed of protein pulling, atomistic MD simulations and experiment have not yet met in the middle.

Most simulations of the AFM experiments involve the application of an artificial force to the system, a technique known as steered molecular dynamics (SMD).¹² A virtual harmonic spring is attached to each end of the protein. Moving the springs apart at constant velocity applies a force which pulls the ends of the protein apart in a similar way to how the AFM experiment operates. The extension of the spring is recorded as the protein unfolds, and this provides a plot of force versus extension which mimics the VC experiments. SMD can also mimic FC experiments by applying a constant force along the vector between the ends of the protein, pulling them apart with a constant force rather than at a constant speed.

SMD has been used to investigate the mechanical unfolding of a number of protein domains. The I27 domain of titin is an all beta domain which is well studied by experiment and simulation. VC experiments have found that I27 has a high mechanical strength and consequently unfolds at a high pulling force of around 150–200 pN.^{4,15} As the force builds up, I27 extends by a very small amount until the peak unfolding force is reached, after which a sudden and rapid collapse of the structure leads to large extensions and a decrease of the force.^{4,15} Theoretical studies^{16–20} have used SMD to pull I27 and have found that the mechanical behavior comes from the backbone hydrogen bonds between the terminal β -sheets. As force is applied, these hydrogen bonds are stressed and share

Received: November 25, 2015

Revised: January 12, 2016

Published: January 13, 2016

the load which leads to a high mechanical resistance. When the applied force reaches the unfolding force, the hydrogen bonds rupture simultaneously, which leads to the complete failure of the rest of the domain.

Protein L is a mixed α - β domain which shows similar mechanical strength and unfolding behavior to I27.²¹ SMD was used to investigate the mechanical unfolding of protein L¹² which was found to be similar to I27 in that a cluster of backbone hydrogen bonds in the terminal β -sheets withstand a high force and then suddenly fail leading to complete unfolding. The main difference between them was found to be that I27 populates a number of intermediate states along the unfolding pathway while protein L unfolds via a simple two-state system.¹²

All α domains such as IM9 unfold at much lower forces than all β or mixed α - β domains.^{4,12} This is thought to be due to the absence of a cluster of backbone hydrogen bonds which are loaded all at once and share the force as is found in β -sheets. Instead, the backbone hydrogen bonds are loaded sequentially and fail one at a time, leading to a gradual unravelling at low force rather than the initial resistance followed by sudden failure displayed by I27 and protein L.¹² This view is confirmed by computational studies where SMD was used to unfold a number of all α domains.^{22–24} In summary, SMD has been used to show that the secondary structure of a protein or domain is the biggest factor in determining the resistance to pulling.¹

SMD applies a pulling speed or force that is several orders of magnitude greater than that of the corresponding experiment. The effect of this on the validity of the simulations is under debate. Experiment has shown that for I27 the excessive pulling speeds and forces used have not negatively impacted the results of the simulations; however, it is possible that it is a problem for systems that do not share the same mechanical characteristics as I27.²⁴ BXD differs in that no force or pulling speed is applied; instead, the system diffuses along the reaction coordinate of end-to-end distance with no modification of the potential.

In this paper we report an application of boxed molecular dynamics to an investigation of the existing protein pulling experiments. Slow pulling, the most challenging case for straightforward MD and SMD, will be considered as it is accessible to BXD. The following results are reported:

(1) We apply boxed dynamics (BXD) to calculate the free energy often also called potential of mean force (PMF) for three protein domains previously studied experimentally, and the pulling force is estimated as a gradient of PMF.

(2) We show that the PMFs of the two β -sheet proteins have similar characteristic features. Near the equilibrium their PMFs show a deep minimum followed by an inflection point toward the region of less steep PMF. As a result, a strong initial force is needed to break out of this minimum. At longer extensions, the gradient of the PMF becomes smaller and the force rapidly decreases. After that, the force generally increases, showing “humps” again while the domains extend. Eventually when the domains are completely unfolded, the PMF rapidly increases, reflecting the high strength of the linear protein chain. This is consistent with experiment for rigid β -sheet proteins, which shows “humps” at intermediate extensions^{3,25} and a peak which is related to quick unfolding of native structure when the pulling force reaches a critical value.

(3) The forces obtained from our BXD calculation roughly reproduce the magnitude of experimental forces and unlike

SMD^{12–14} do not require artificially high additional pulling forces.

(4) Calculations confirm the experimentally observed correlation between mechanical strength of a protein and its structure. For proteins composed of α -helices, the force required along the extension coordinate is substantially lower than for those composed of β -sheets, in agreement with experimental observations. The PMF of the α -helix protein has a broader minimum and is not followed by an inflection point to a flatter region. This is compatible with experimental observations that the unfolding force for α -helix proteins is lower than that of β -sheet proteins and does not have a pronounced peak.

(5) Snapshots of unfolding proteins along the reaction coordinate suggest that the peaks and the minima in the plots of the force vs extension (i.e., the AFM spectrum) are related with structural changes in proteins induced by pulling and allow visualization of these changes.

(6) As BXD provides not only thermodynamical information (PMF) but also kinetic rate constants, the kinetic rates of protein unfolding in the constant force (FC) experiments can be estimated. The reported BXD rate constants and characteristic times, obtained on the basis of fully atomistic simulations using standard force field and no parameter adjustment, are of the order of seconds to minutes, way out of reach of standard molecular dynamics and within 2 or 3 orders of magnitude of the experimental observations. Possible modifications of the force field, which may help to reach quantitative agreement with experiment, are discussed.

THEORY

Boxed Molecular Dynamics. Although this paper is focused on the application of BXD, in this section we provide an outline for readers' convenience. Boxed molecular dynamics^{26–28} (BXD) is a simple and straightforward technique which extends the time scale of atomistic molecular dynamics (MD) simulations and facilitates simulation of rare events. It has recently been reviewed²⁹ along with several recent applications. In BXD's simplest implementation, we assume that a chemical reaction or some other atomistic physical process can be described by a reaction coordinate or an appropriate order parameter. It is then possible to split this coordinate into several boxes and subsequently lock the dynamics in each box by inverting the velocity of the trajectory in the direction of the reaction coordinate (or order parameter) every time the trajectory hits a boundary between two subsequent boxes. Then one can calculate the rate constant of exchange between the boxes simply by calculating average time between the “hits”

$$k_{m,m+1} = \frac{h_{m,m+1}}{t_m} \quad (1)$$

$$k_{m+1,m} = \frac{h_{m+1,m}}{t_m}$$

where $h_{m,m+1}$ and $h_{m+1,m}$ are the number of “hits” of the left and right boundary of the m th box and t_m is the time spent in this box. After accumulating sufficient statistics, the trajectory is allowed into the neighboring box where the procedure is repeated. Eventually a set of rate constants for exchange between the boxes is accumulated, and the dynamics is reduced to a set of kinetic equations. In matrix form

$$\frac{d\mathbf{n}(t)}{dt} = \mathbf{M}\mathbf{n}(t) \quad (2)$$

In eq 2, \mathbf{n} is the vector of box populations and the elements of the sparse matrix \mathbf{M} are expressed via the rate constants (eq 1). Then the MD simulation of the system is replaced by the solution of the master equation (3).

$$\mathbf{n}(t) = \mathbf{U}\mathbf{A}\mathbf{U}^{-1}\mathbf{n}(0) \quad (3)$$

where $\mathbf{n}(0)$ contains the initial conditions for populations in each box; \mathbf{U} is the eigenvector matrix obtained from diagonalization of \mathbf{M} , where \mathbf{A} is a diagonal matrix whose elements, $\Lambda_{ij} = e^{\lambda_j t}$, are determined by λ , the eigenvalue vector corresponding to \mathbf{M} .

BXD relies on the assumption that the motion within the box is stochastic and that sequential “hits” and velocity inversions are uncorrelated; i.e., the time between the “hits” must be bigger than the correlation time. In general, this is not the case. The requirement of uncorrelated dynamics also imposes a restriction on the box size; i.e., it should be bigger than the so-called correlation length, the length at which a trajectory loses the memory of its initial conditions. For large anharmonically coupled systems, the correlation time and correlation length are usually quite short. Nevertheless, even for a large box a trajectory reflected from a boundary can sometimes turn back rather quickly. The procedure developed in ref 27 removes any contribution to the rate coefficients from correlated velocity inversions separated by very short time and allows BXD calculations to be properly converged.

With the box-to-box rate coefficients in hand, it is possible to estimate the free energy profile along the reaction coordinate:

$$K_{m-1,m} = \frac{k_{m-1,m}}{k_{m,m-1}} = \exp\left(-\frac{\Delta G_{m-1,m}}{k_b T}\right) \quad (4)$$

In refs 26–29 a number of tricks have been developed to improve both accuracy and resolution of BXD results.

BXD has its origin in the intramolecular dynamics diffusion theory (IDDT)^{30–34} which has demonstrated that long time reaction rates can be recovered from a set of short time MD simulations.³⁴ Similarly in BXD longer time scale dynamics may be calculated from the set of rate constants obtained from short-time simulations within each box. The BXD master equation can be easily recast as a diffusional equation and vice versa. In IDDT, the dynamics was not restricted by the boundaries of the box. The idea of the box was introduced in ref 35, where the first BXD simulation of bond fission was carried out using our accelerated dynamics (AXD) with only two boxes and a single boundary between them. Locking the trajectory within a box enables the system to visit the regions of high free energy and low probability, where an unrestricted trajectory would rarely visit.

A number of techniques of accelerated molecular dynamics, such as Milestoning^{36–41} and particularly Milestoning with Voroni tessellations (MVT),^{39,42} have similarities with BXD. See ref 29 for more details. BXD also shares similarities with a number of other rare event acceleration methods. For example, BXD in its simplest form³⁵ is related to Hyperdynamics^{43,44} which similarly introduces constraints into a molecular system's configuration space to encourage it to visit regions of low probability. BXD also has similarities to umbrella sampling^{45–48} with the primary difference that the former uses rectangular boxes instead of parabolic energy restraints. Compared to these

methods, BXD's main advantage is that it is capable of extracting simultaneously high-resolution thermodynamic and kinetic information. BXD's formulation is very straightforward, based upon a simple rewriting of the classical transition state theory formula, which is common in chemistry. While other methods such as forward flux sampling,⁴⁹ nonequilibrium umbrella sampling,⁵⁰ and the adaptive weighted ensemble method⁵¹ also provide simultaneous kinetic and thermodynamic information, BXD does this in a simpler way with no need for multiple trajectories sampling areas of phase space in parallel, biasing of the potential energy function or prior knowledge of the states of the system that lie along the reaction coordinate.

In common with all of the aforementioned techniques, BXD does not address the question of finding a good reaction coordinate or order parameter; its usage relies on being able to define these in a sensible way. In many instances, definition of an appropriate reaction coordinate or order parameter is a substantial challenge, but there are a range of systems where these quantities are reasonably well-defined. If a good reaction coordinate exists, then BXD provides long time scale dynamics/kinetics and PMF profiles in regions of very high energies (see the table of contents graphic). For protein pulling, end-to-end distance represents a natural choice of reaction coordinate.

Simulation Details. In this paper we present calculations performed using the BXD subroutine implemented on CHARMM. The reaction coordinate was chosen as the distance between the two termini of the protein domain as this would correspond to the coordinate sampled by the AFM pulling experiments. This end-to-end distance coordinate was later translated to extension by subtracting the equilibrium value.

The EEF1 implicit solvent model and Charmm 19 force field were used for the simulations along with the Langevin thermostat set to 303 K and a friction coefficient of 50 ps⁻¹ in order to replicate bulk water. Three protein domains I27, protein L, and IM9 were investigated. The PDB structures for all of them were equilibrated under the force field and solvent model for 500 ns before BXD simulations commenced.

For IM9 and protein L boxes were placed at intervals of 0.75 Å from 11.25 to 320.25 Å and from 21 to 320.25 Å, respectively. For I27 the boxes were at intervals of 0.5 Å from 20 to 330 Å. Between 5000 and 2000 inversion events were required in each box before a boundary could be passed.

Initially, the reaction coordinate was sampled downward from zero extension to the lowest boundary in order to fully explore the local minimum before the direction was reversed and the protein domain began to extend. Sampling was continued until a linear conformation had been achieved.

Simulations were also carried out with boxes at intervals of 0.5, 0.25, and 0.125 Å in order to ascertain that the result was independent of box size and placing. This was found to be the case. For each protein domain 30 unfolding BXD runs were obtained, and a similar sequence of unfolding events was observed for each domain despite the different box distributions. The total simulation time for each domain was of the order of tens of nanoseconds. The decorrelation time used in the procedure, which removes correlated short time events, was found to be 600 fs for each system.

The set of box-to-box rate constants yields both PMF and all information necessary for description of the kinetics of unfolding along the pulling reaction coordinate.

RESULTS AND DISCUSSION

BXD Calculation of PMF and Description of Constant Speed AFM Experiments. PMFs as a function of the end-to-end distance, and their gradients are shown in Figure 1 for the three protein domains investigated in this work.

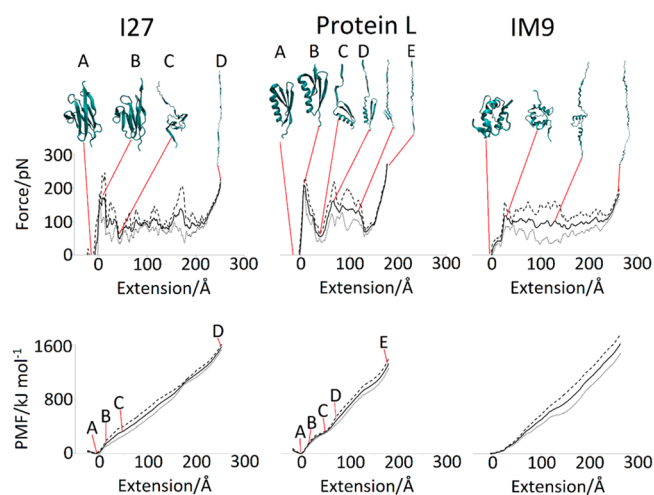


Figure 1. Potential of mean force (lower frames) and pulling force as its derivative (upper frame) are shown for proteins I27, L, and IM9. The dashed line and dotted line represent the upper and lower limit within the error bar. Points A–D indicate particularly important regions of the PMF, corresponding to minima and maxima of the force. The snapshots of the structures at those regions are also shown. The dashed and dotted lines show the boundaries of standard deviation from the calculated average. For the free energy profiles (below) this was calculated by taking the standard deviation of the ensemble of free energies from each individual unfolding trajectory. Each free energy profile was converted into force by differentiation, and then the average force and standard deviation were taken from the ensemble of individual force versus extension plots.

It should be noted that mechanical unfolding is a completely different process to thermal unfolding¹⁴ *in vivo*. While thermal unfolding PMFs feature small barriers separating stable states, the AFM probe pulls the protein apart, forming an unnatural linear conformation with a very high free energy. This is supported by the fact that there is no correlation between the mechanical and thermal stability of a protein.⁵²

As potential of mean force represents thermodynamical free energy along the reaction coordinate x , by definition the change of PMF is equivalent to mechanical work required for displacement along the reaction coordinate (protein extension in this case) and mechanical force is given by the gradient of PMF

$$F = \frac{dW}{dx} \approx \frac{dG}{dx} \quad (5)$$

provided that the extension is slow enough for equilibrium thermodynamics to be valid. In the limit of higher speed the kinetics of protein pulling should be considered in more detail. As the kinetics of mechanical unfolding changes with pulling speeds, a number of models have been developed to describe this dependence.^{53–55} In principle, BXD theory can be applied to a wide range of pulling speeds and the relationship of the pulling force and pulling speed can be obtained, but in this paper we will focus on the limit of slow speed where simple thermodynamical argument leads to eq 5.

Some features of the PMFs are similar for all three protein domains. At low extensions near equilibrium there is a well, which is steep for the beta containing domains I27 and L, and relatively smooth and broad for the all alpha IM9 domain. This difference is due to the fact that for I27 and protein L the initial force is loaded onto β -sheets which hold out to high forces before suddenly rupturing, whereas with IM9 the initial force is loaded onto α -coils which rupture gradually at low forces.^{2,4,15} This well is situated around the global minimum along the extension coordinate and corresponds roughly to the equilibrium native structure of the domain under the conditions of the simulation.

For I27 the force increases sharply to a high value without any significant change in the equilibrium structure, before a sudden rupture between β -sheets leads to a reduction in the force. This corresponds to breaking out of the steep sided equilibrium well to less steep regions of the PMF (going from the minimum point A to the inflection point B in Figure 1). This is due to the “brittle” nature of β -sheets; initially, the force is shared between multiple hydrogen bonds which fail suddenly, allowing the β -sheet to unravel easily.^{4,15} This is responsible for the sudden reduction in the force going from point B where the force is maximal to point C in Figure 1. The hydrogen bonds responsible for the initial resistance are shown in Figure 2.

Next the other β -sheets are loaded and fail sequentially in a similar brittle manner, until they have all unfolded and the force again increases as a linear conformation is reached (point D). The shape of the PMF curve calculated from fully atomistic simulations is very much in line with the suggestions made to explain experimental results²⁵ and theoretical models.⁵⁶

For the α -protein IM9 after the initial increase the force remains flatter and shows smaller peaks and troughs than those of the I27 protein. This is due to the fact that the connections within α -helices, and the helices themselves fail more gradually, leading to lower forces. This is in agreement with the literature; Brockwell reports that the all alpha IM9 domain unfolds below the noise limit of the experiment,² and SMD simulations^{22–24} suggest that the mechanical weakness of α -coils is thought to be due to the fact that only one backbone hydrogen bond is loaded at a time, leading to sequential failure and an unravelling of the α -coil.

Protein L is a combination of β -sheets and α -helices. Initially, the force profile and PMF is similar to that of I27 as the force is resisted by β -sheets in both proteins. After the first β -sheet fails (see Figure 2) a strong intermediate structure remains. The resistance of a β -sheet on this intermediate is responsible for the large force peak at moderate extensions of protein L (going from points C to D). After this β -sheet fails the force curve flattens out as an α -coil unravels gradually before the remaining structure fails and a linear conformation is reached (point E).

In experiment it is not a single protein domain but their sequence, a concatemer that is pulled. However, knowing the PMF of an individual domain and its gradient and assuming that the domains extend independently and sequentially one by one allows reconstruction of the experimentally observed dependence of the force vs protein extension, shown by Figure 3.

This assumption that the unfolding events witnessed in the concatemer are equivalent to the unfolding in isolated domains is common.^{7,57} However, the extent to which it is valid is not fully resolved as cooperative motion of the concatemer and domain–domain interactions can affect the results of the experiment;⁵⁸ in certain systems domains have been shown to

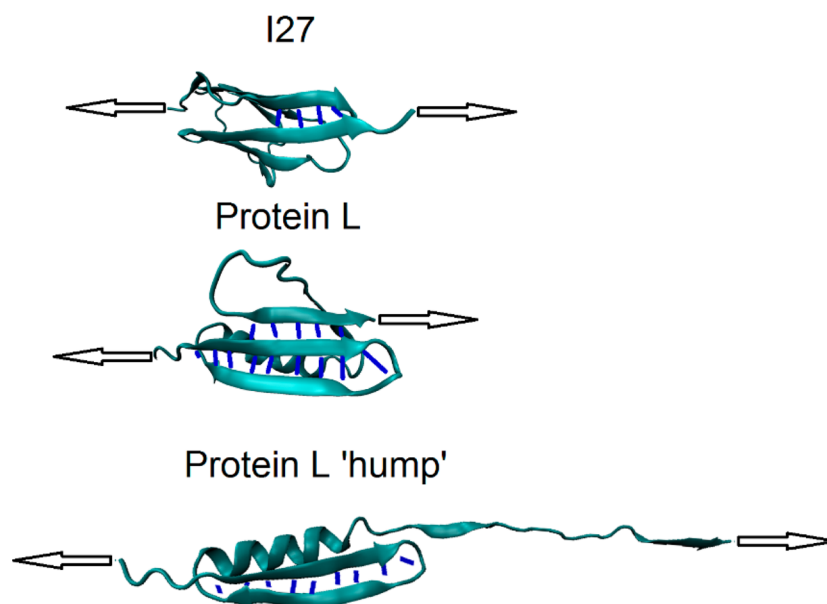


Figure 2. Structures responsible for force peaks in I27 and protein L. Breaking the native structures (top two) requires the maximum pulling force as they represent the bottom of steep free energy wells around the native structure. The hydrogen bonds responsible for the initial resistance are shown as blue lines. These bonds rupture simultaneously, causing a large structural change and rapid extension of the domain. In protein L there are two systems of hydrogen bonds, which “unzip” sequentially. The structure of the intermediate responsible for the “hump” in the force curve of protein L is shown at the bottom, corresponding to point D in the middle frame of Figure 1. Arrows indicate the direction and points of application of the experimental pulling force.

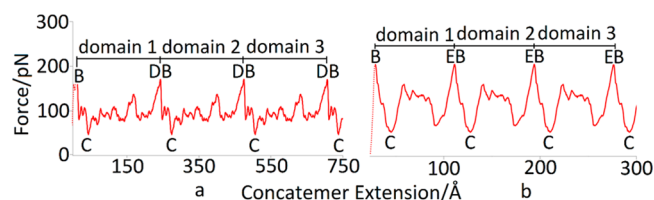


Figure 3. “Teeth” due to extension of a chain sequence of the proteins. When one protein is extended to the point where the force reached the peak in the native well of the next protein, the next protein’s native structure quickly passes the inflection point and breaks down leading to the sharp fall of the pulling force. Then the process is repeated. Frames a and b correspond to I27 and protein L, respectively. The “teeth” are generated by repeating the section of the force curve beyond the initial peak force (Figure 1, point B) and the point where the force reaches this value again (Figure 1, point D). This mimics the experimental force trace where a chain of domains unfolds one at a time.

unfold together rather than independently,²³ and in VC experiments the unfolding forces depend on the number of domains in the concatemer.^{57,58} As shown by Figure 3a for I27 BXD qualitatively reproduces the overall structure of the experimentally observed “teeth” in the dependence of force on extension. The peak force and the length of the tooth are both well reproduced. Figure 3b is the same as Figure 3a but for protein L. The main peak DB in both Figures 3a and 3b is reached when an expanding member of the concatemer reaches a linear extension and the force becomes sufficient to break the native structure of the next member of the chain. Then the process is repeated.

No multiple “teeth” can be obtained during the extension of a sequence of the α -protein IM9 as there is no sharp decrease of the force at the point of inflection. The slope of the PMF for the IM9 α -protein in frame c of Figure 1 is lower than that of the Figures 1a and 1b for (protein L and I27, respectively).

These observations are in agreement with experiment where α -proteins were found to be less robust and showed neither significant peaks nor “teeth” in the dependence of the force vs extension.

In a number of experiments the humps of the force were observed and also attributed to the structural changes in small proteins,^{3,25} which are also present in our simulations. The experimental dependence of the force on the extension is typically less structured than those shown by Figure 4 as BXD

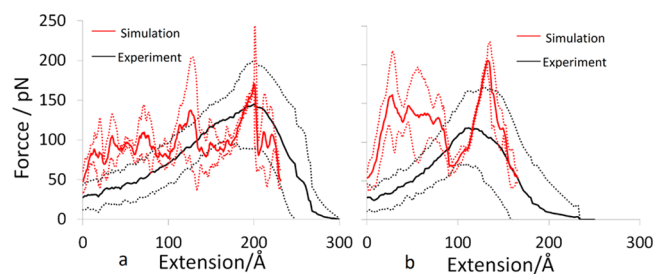


Figure 4. To illustrate the spread of experimental data, this figure shows theoretically calculated forces (red, a single “tooth” at the Figure 3) superimposed with experimental “teeth” (black) provided by David Brockwell, for I27 (frame a) and protein L (frame b). Simulated force teeth (red line) were calculated by taking the gradient of the PMF for each trajectory and averaging. Thin red lines represent a standard deviation of the force from each trajectory; thin black lines represent one standard deviation in the average of the experimental peaks.

provides the PMF and force curves at a higher resolution than the AFM experiments. This has allowed the identification of an intermediate structure in the mechanical unfolding of protein L (see Figure 2).

The lower resolution of the experiment is because, when a sequence of protein molecules is pulled and the extension is a collective motion of several protein molecules in the sequence,

the details shown by theoretical calculations of the PMF of a single molecule are washed away. Only the main peak due to initial destruction of native structure and perhaps occasionally some other most pronounced humps survive in experiment. Very clear force peaks and “humps” corresponding to sequential breaking of the protein structure were observed for larger species.^{59–62}

The measured unfolding force is somewhat dependent on the experimental conditions such as cantilever stiffness and loading rate^{4,21} and on how many domains in the concatamer have unfolded already¹⁵ and on many other factors. Figure 4 illustrates the uncertainty of the experiment by showing the average of a number of experimental force peaks (black lines, dotted lines show one standard deviation) against our calculated peaks (red) for I27 and protein L. Given that these effects are not included in the simulations, one cannot expect a detailed agreement with experiment. The fact that the magnitude of the main peak of the force is well reproduced is encouraging.

BXD Kinetic Description of Force Clamp Experiments.

Force clamp is another mode of AFM pulling in which proteins are pulled with a constant force and the distribution of unfolding times is measured, giving the rate constant for unfolding at that particular force. Another quantity reported in the FC experiments is the protein extension as a function of the pulling force. Typical measured unfolding times (inverse rate constants) are in the order of seconds, which nevertheless can be within the reach of BXD. Figure 5a,b summarize the results of modeling the FC experiment.⁹

The PMFs calculated from BXD simulations increase monotonically to very high values, as shown in Figure 1. It is clear from this that unfolded states would never be reached as the free energies are too high. It is thought that the application of force along the pulling coordinate tilts the free energy profile and allows mechanical unfolding to occur.^{63,64} Similarly to this the PMFs obtained with BXD were adjusted to take a pulling force into account.

Frame A of Figure 5a,b show the force adjusted PMF obtained by modifying box-to-box rate constants with the factor $\exp(\pm\Delta xF/k_B T)$ so that

$$k'_{n,n-1} = k_{n,n-1} \exp\left(\frac{\Delta x F}{k_B T}\right)$$

$$k'_{n-1,n} = k_{n-1,n} \exp\left(-\frac{\Delta x F}{k_B T}\right) \quad (6)$$

where Δx is the box size. For each free energy profile an end point was chosen to correspond to the maximum extension under that pulling force. The choice of end point depended on the shape of the free energy. For low forces (<30 pN) the free energy rises monotonically with no transition state or well present, which implies no or extremely slow unfolding. In this case the end point was defined as the inflection point in the PMF in the initial native well at which peak force is reached. For intermediate forces (30–50 pN) the free energy is tilted to show two wells: the native structure well at zero extension and a well at extensions of 50–100 Å. In this case the second well was chosen as the final state because it represents a transition from the native structure to a stable unfolded conformation. At higher forces (>60 pN) the force adjusted PMF shows three wells, and the final well was chosen as a final state. The kinetic end points are shown by frame A of Figure 5a,b by the arrows,

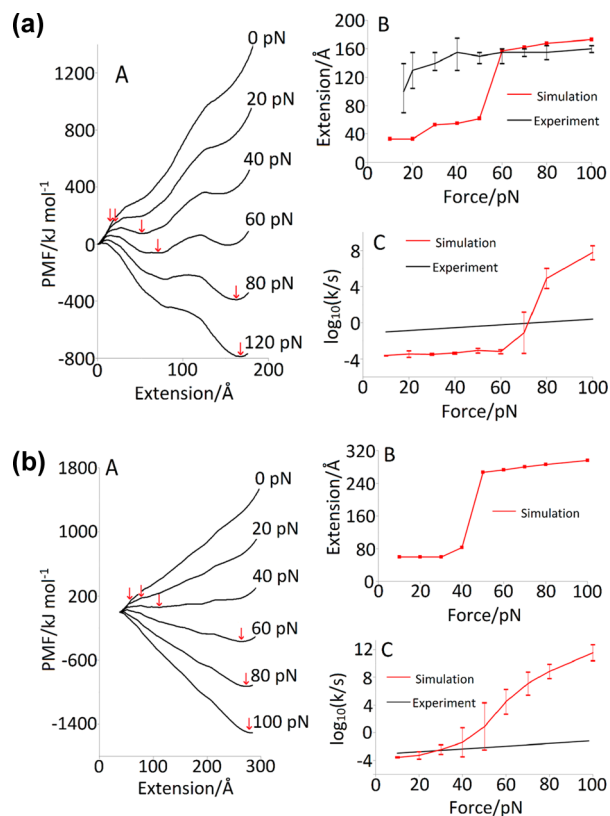


Figure 5. (a) PMFs for protein L obtained with modified rate constants eq 6 which include an additional factor taking into account for the external force in the FC experiment (frame A). The assumed end points are shown by arrows. They are also shown in frame B as a function of the applied force (red squares) and compared with the experimental data (black line). The estimated rate constant of unfolding is shown by frame C by the red line and compared with experiment shown by the black line. For the force below 60 pN the theoretical rate constant remains flat underestimating reported experimental data by 2 orders of magnitude and grows fast at higher forces as the unfolding becomes barrierless. It must be noted that the experimental error bar is not known but may be significant. (b) PMFs for the protein I27 obtained with modified rate constants eq 6 which include additional factor taking into account external force in the FC experiment (frame A). The assumed end points are shown by arrows. They are also shown at the (frame B) as a function of the applied force (red squares). Experimental data are absent here. The estimated rate constant of unfolding is shown in frame C by the red line and compared with experiment shown by black line. For the force above 60 pN the theoretical rate constant starts growing fast as the unfolding becomes barrierless. It must be noted that the experimental error bar is not known but may be significant as experiment is unable to detect very fast unfolding events.

and the dependence of end point on the external force, that is, the extent to which the domain unfolds at each force, is shown by frame B of the same figure in comparison with that of experiment.⁹

In order to estimate the unfolding times, the matrix **M** in eq 2, containing the rate constants for a particular force, was truncated by removing all boxes beyond the end point corresponding to the maximum extension at that force. The eigenvalues of the kinetic matrix determine the kinetic rate constants in the system. Usually the lowest eigenvalue of a kinetic matrix **M** is separated from all other eigenvalues and determines the reaction characteristic time and the rate constant of the unfolding. Frame C of Figures 5a,b compares

the calculated rate constant with experiment. For medium force the calculated characteristic times are within the range of 10^2 – 10^3 s, which is approximately 2–3 orders of magnitude higher than the reported experimental result. At higher forces (~ 50 pN) the reaction becomes barrierless, unfolding becomes very fast, and the rate constant increases dramatically.

Quantitative agreement with experiment has not been achieved here, but it must be taken into account that BXD estimates the reaction rates in the order of minutes based on fully atomistic simulations without any modifications of the force field. Comparison of BXD calculations with FC experiment allows modifications of the force field to be suggested which may improve the agreement with experiment. If the equilibrium well in the PMF is steeper and the region after the inflection point is flatter, then the agreement between experiment and theory in frame C of Figure 5a,b should improve. The former can be achieved by adjusting the parameters of hydrogen bonds responsible for the gradient of the PMF on the right-hand side of the equilibrium well. Also in our simulations an implicit solvent model was used to reduce the cost of the simulations, common practice in mechanical unfolding simulations.¹² There is evidence that the barrier to unfolding is lowered by the presence of water molecules. For example, with protein such as I27 where unfolding is resisted by multiple hydrogen bonds in terminal β -sheets it is thought that ruptured hydrogen bonds can re-form with water rather than within the protein, lowering the energetic cost of rupture and hence reducing the height of the unfolding barrier. Using explicit water instead of implicit solvent model also may improve the agreement with experiment, but this is computationally expensive because as the protein extends a larger water box must be used.

Perhaps an approach similar to that outlined by ref 67 could be used, which treats explicitly only the water molecules which are in close proximity to the protein, and the rest of the water box by a hydrodynamic approach.

Similarly to the velocity clamp experiments, force clamp experiments can be affected by the fact that AFM experiments pull a concatemer of many protein domains. It is assumed that each domain unfolds fully and independently of the others. Based on this assumption, each of the steps in the extension vs time trace of FC pulling is assumed to be from a single domain, and the data for each domain are averaged.

However, if this independence of unfolding events is not wholly correct, then the situation becomes complicated. Cooperative motion of the concatemer, interactions between domains, and energy storage in the chain would make analysis of the experimental data much more difficult and inconclusive. The debate as to whether this central assumption is valid is ongoing,⁶⁵ and AFM studies of the refolding of a concatemer show that cooperative motion drives the process.⁶⁶ Simulation of a concatemer and investigation of cooperative effects will be our future goal.

It is also possible that the measured force at which unfolding events occur does not exactly correspond to reality. Thermal fluctuations in the position of the AFM tip may cause protein domains to unfold prematurely at forces other than that registered by the experiment,¹⁰ and if the concatemer is not pulled perpendicular to the solid surface to which it is attached, then the force along the pulling coordinate could be less than what is measured.

After a domain unfolds the AFM tip recoils, and the force is no longer applied to the concatemer for a certain amount of

time.⁴⁷ Any unfolding events that take place in less time than this recoil time will be missed. In the experiments from which the data are taken for comparison to the BXD simulation, this detection threshold is roughly 5 ms, corresponding to $\log(k_{\text{unfold}}) \approx 2.3$. This effect would affect the experimental $\log(k_{\text{unfold}})$ vs force curve as fast unfolding events would be missed.

Recently, it has been discussed in the literature whether the kinetics of the FC experiment can be described by a single rate.⁶⁶ In principle, BXD allows reproduction of complicated kinetics. For example, it has been shown that given the initial conditions peptide cyclization kinetics can exhibit both simple single exponential and complicated power law behavior.²⁸ Describing this phenomena in AFM experiments will require more work. In this first paper, however, we only show that BXD is capable of recovering realistic rates and very long characteristic times observed in the FC experiments. In the future we will attempt to reach quantitative agreement with experiment by addressing the issues discussed above.

CONCLUSIONS

The above discussion can be summarized as follows:

1. BXD reproduces the realistic pulling forces (in the order hundreds of pN) observed in VC experiments.
2. Initially the force increases without affecting the equilibrium structure significantly. When a threshold force is reached, a structural unit fails, followed by the sequential failure of the remaining structural units at lower forces until a linear conformation is reached.
3. β -Sheets resist the force well before failing abruptly leading to pronounced peaks and troughs in the force vs extension curves, whereas α -coils fail gradually at lower forces, leading to a flatter force curve.
4. The force vs extension curves reproduced by BXD reveal intermediate structures along the mechanical unfolding pathway. These structure can easily be visualized.
5. Qualitative agreement with the FC experiments has been achieved on unfolding processes up to a time scale of seconds. The theory that the application of force tilts the free energy landscape, thus allowing mechanical unfolding to proceed,^{63,64} has been confirmed.

All the above results were obtained with the help of boxed molecular dynamics, a fully atomistic technique of accelerated MD, without additional forces or adjustable parameters. We believe that this is the first time that the mechanical unfolding of proteins has been simulated under realistic conditions, providing reliable confirmation of the factors affecting mechanical stability as well as the unfolding pathways and free energies

FUTURE WORK

The following interesting and important issues will be addressed later:

1. The parameters of the force field will be varied in particular the strength of the hydrogen bond. This should make the features of the PMF, such as deep well and inflection point for protein L and I27, more pronounced, so that a better quantitative agreement could perhaps be achieved. Explicit water may also be introduced.
2. An attempt to go beyond the assumption of slow pulling will be made. This may allow us to look at the dependence of

the maximal pulling force on the pulling speed observed in experiment. BXD kinetics is well suited for that.

3. By simulating the extension of a concatomer, we may be able to assess the importance of cooperative effects and also answer the question whether the “humps” which correspond to the breaking of the structures within a single member of the concatomer are washed away.

Thus, BXD is an efficient theoretical tool to study long time processes in the dynamics of protein unfolding by pulling.

AUTHOR INFORMATION

Corresponding Author

*E-mail D.Shalashilin@leeds.ac.uk (D.V.S.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge David Glowacki for his valuable feedback on the work as well as Simon Scheuring, David Brockwell, and Lorna Dougan for their insights into the AFM experiments and Katie Lamb for fixing a malfunctioning reference section. J.B.'s studentship was provided by EPSRC grant EP/J0001481/1.

ABBREVIATIONS

AFM, atomic force microscopy; FC, force clamp; VC, velocity clamp; MD, molecular dynamics; SMD, steered molecular dynamics; BXD, boxed molecular dynamics.

REFERENCES

- (1) Brockwell, D. J. Probing the mechanical stability of proteins using the atomic force microscope. *Biochem. Soc. Trans.* **2007**, *35*, 1564–1568.
- (2) Hedberg, C.; Toledo, G.; Gustafsson, C.; Larson, G.; Oldfors, A.; Macao, B. Hereditary myopathy with early respiratory failure is associated with misfolding of the titin fibronectin III 119 subdomain. *Neuromusc. Disord.* **2014**, *24*, 373–379.
- (3) 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*, 100–103.
- (4) Crampton, N.; Brockwell, D. J. Unravelling the design principles for single protein mechanical strength. *Curr. Opin. Struct. Biol.* **2010**, *20*, 508–517.
- (5) Hann, E.; Kirkpatrick, N.; Kleanthous, C.; Smith, D. A.; Radford, S. E.; Brockwell, D. J. The effect of protein complexation on the mechanical stability of Im9. *Biophys. J.* **2007**, *92*, L79–L81.
- (6) Sadler, D. P.; Petrik, E.; Taniguchi, Y.; Pullen, J. R.; Kawakami, M.; Radford, S. E.; Brockwell, D. J. Identification of a mechanical rheostat in the hydrophobic core of Protein L. *J. Mol. Biol.* **2009**, *393*, 237–248.
- (7) Brujic, J.; Hermans, R. I.; Walther, K. A.; Fernandez, J. M. Single-molecule force spectroscopy reveals signatures of glassy dynamics in the energy landscape of ubiquitin. *Nat. Phys.* **2006**, *2*, 282–286.
- (8) Garcia-Manyes, S.; Brujic, J.; Badilla, C. L.; Fernández, J. M. Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of I27 and ubiquitin. *Biophys. J.* **2007**, *93*, 2436–2446.
- (9) Liu, R.; Garcia-Manyes, S.; Sarkar, A.; Badilla, C. L.; Fernández, J. M. Mechanical characterization of Protein L in the low-force regime by electromagnetic tweezers/evanescent nanometry. *Biophys. J.* **2009**, *96*, 3810–3821.
- (10) Oberhauser, A. F.; Hansma, P. K.; Carrion-Vazquez, M.; Fernandez, J. M. Stepwise unfolding of titin under force-clamp atomic force microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 468–472.

(11) Schlierf, M.; Li, H.; Fernandez, J. M. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 7299–7304.

(12) Brockwell, D. J.; Beddard, G. S.; Paci, E.; West, D. K.; Olmsted, P. D.; Smith, D. A.; Radford, S. E. Mechanically unfolding the small, topologically simple protein L. *Biophys. J.* **2005**, *89*, 506–519.

(13) Paci, E.; Karplus, M. Forced unfolding of fibronectin type 3 modules: An analysis by biased molecular dynamics simulations. *J. Mol. Biol.* **1999**, *288*, 441–459.

(14) Yew, Z. T.; Olmsted, P. D.; Paci, E. Free energy landscapes of proteins: insights from mechanical probes. *Single-Molecule Biophysics: Experiment and Theory*; John Wiley and Sons: 2012; Vol. 146, pp 395–417.

(15) Kawakami, M.; Byrne, K.; Brockwell, D.; Radford, S.; Smith, A. Viscoelastic study of the mechanical unfolding of a protein by AFM. *Biophys. J.* **2006**, *91*, L16–L18.

(16) Fowler, S.; Best, R.; Herrera, J.; Rutherford, T.; Steward, A.; Paci, E.; Karplus, M.; Clarke, J. Mechanical unfolding of a titin Ig domain: Structure of unfolding intermediate revealed by combining AFM, molecular dynamics simulations, NMR and protein engineering. *J. Mol. Biol.* **2002**, *322*, 841–849.

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

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

(19) Lu, H.; Schulten, K. The key event in force-induced unfolding of titin's immunoglobulin domains. *Biophys. J.* **2000**, *79*, 51–65.

(20) Gao, M.; Lu, H.; Schulten, K. Unfolding of titin domains studied by molecular dynamics simulations. *J. Muscle Res. Cell Motil.* **2002**, *23*, 513–521.

(21) Crampton, N.; Alzahrani, K.; Beddard, G.; Connell, S.; Brockwell, D. Mechanically unfolding Protein L using a laser-feedback-controlled cantilever. *Biophys. J.* **2011**, *100*, 1800–1809.

(22) Chng, C.; Kitao, A. Mechanical unfolding of bacterial flagellar filament protein by molecular dynamics simulation. *J. Mol. Graphics Modell.* **2010**, *28*, 548–554.

(23) Law, R.; Carl, P.; Harper, S.; Dalhaimer, P.; Speicher, D.; Discher, D. Cooperativity in forced unfolding of tandem spectrin repeats. *Biophys. J.* **2003**, *84*, 533–544.

(24) Lee, E.; Hsin, J.; Sotomayor, M.; Comellas, G.; Schulten, K. Discovery through the computational microscope. *Structure* **2009**, *17*, 1295–1306.

(25) Rico, F.; Gonzalez, L.; Casuso, I.; Puig-Vidal, M.; Scheuring, S. High-speed force spectroscopy unfolds Titin at the velocity of molecular dynamics simulations. *Science* **2013**, *342*, 741–743.

(26) Glowacki, D. R.; Paci, E.; Shalashilin, D. V. Boxed Molecular Dynamics: A simple and general technique for accelerating rare event kinetics and mapping free energy in large molecular systems. *J. Phys. Chem. B* **2009**, *113*, 16603–16611.

(27) Glowacki, D. R.; Paci, E.; Shalashilin, D. V. Boxed Molecular Dynamics: decorrelation time scales and the kinetic master equation. *J. Chem. Theory Comput.* **2011**, *7*, 1244–1252.

(28) Shalashilin, D. V.; Beddard, G. S.; Paci, E.; Glowacki, D. R. Peptide kinetics from picoseconds to microseconds using boxed molecular dynamics: Power law rate coefficients in cyclisation reactions. *J. Chem. Phys.* **2012**, *137*, 165102.

(29) Booth, J.; Vazquez, S.; Martinez-Nunez, E.; Marks, A.; Rodgers, J.; Glowacki, D. R.; Shalashilin, D. V. Recent applications of boxed molecular dynamics: a simple multiscale technique for atomistic simulations. *Philos. Trans. R. Soc. A* **2014**, *372*, 20130384.

(30) Guo, Y.; Shalashilin, D. V.; Krouse, J. A.; Thompson, D. L. Intramolecular dynamics diffusion theory approach to complex unimolecular reactions. *J. Chem. Phys.* **1999**, *110*, 5521–5525.

(31) Guo, Y.; Shalashilin, D. V.; Krouse, J. A.; Thompson, D. L. Predicting nonstatistical unimolecular reaction rates using Kramers' theory. *J. Chem. Phys.* **1999**, *110*, 5514–5520.

- (32) Shalashilin, D. V.; Thompson, D. L. Monte Carlo variational transition-state theory study of the unimolecular dissociation of RDX. *J. Phys. Chem. A* **1997**, *101*, 961–966.
- (33) Shalashilin, D. V.; Thompson, D. L. Method for predicting IVR-limited unimolecular reaction rate coefficients. *J. Chem. Phys.* **1997**, *107*, 6204–6212.
- (34) Shalashilin, D. V.; Thompson, D. L. Intramolecular dynamics diffusion theory: nonstatistical unimolecular reaction rates. *ACS Symp. Ser.* **1997**, *678*, 81–98.
- (35) Martinez-Nunez, E.; Shalashilin, D. V. Acceleration of classical mechanics by phase space constraints. *J. Chem. Theory Comput.* **2006**, *2*, 912–919.
- (36) Elber, R. A milestone study of the kinetics of an allosteric transition: Atomically detailed simulations of deoxy Scapharca hemoglobin. *Biophys. J.* **2007**, *92*, L85–L87.
- (37) Faradjian, A. K.; Elber, R. Computing time scales from reaction coordinates by milestone. *J. Chem. Phys.* **2004**, *120*, 10880–10889.
- (38) Kuczera, K.; Jas, G. S.; Elber, R. Kinetics of helix unfolding: molecular dynamics simulations with milestone. *J. Phys. Chem. A* **2009**, *113*, 7461–7473.
- (39) Maragliano, L.; Vanden-Eijnden, E.; Roux, B. Free energy and kinetics of conformational transitions from voronoi tessellated milestone with restraining potentials. *J. Chem. Theory Comput.* **2009**, *5*, 2589–2594.
- (40) Vanden-Eijnden, E.; Venturoli, M.; Ciccotti, G.; Elber, R. On the assumptions underlying milestone. *J. Chem. Phys.* **2008**, *129*, 174102.
- (41) West, A. M. A.; Elber, R.; Shalloway, D. Extending molecular dynamics time scales with milestone: example of complex kinetics in a solvated peptide. *J. Chem. Phys.* **2007**, *126*, 145104.
- (42) Vanden-Eijnden, E.; Venturoli, M. Markovian milestone with Voronoi tessellations. *J. Chem. Phys.* **2009**, *130*, 194101.
- (43) Voter, A. F. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.* **1997**, *78*, 3908–3911.
- (44) Voter, A. F. A method for accelerating the molecular dynamics simulation of infrequent events. *J. Chem. Phys.* **1997**, *106*, 4665–4677.
- (45) Torrie, G. M.; Valleau, J. P. Monte-Carlo study of a phase-separating liquid-mixture by umbrella sampling. *J. Chem. Phys.* **1977**, *66* (4), 1402–1408.
- (46) Torrie, G. M.; Valleau, J. P. Non-physical sampling distributions in Monte-Carlo free-energy estimation – umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (47) Kaestner, J. Umbrella sampling. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2011**, *1*, 932–942.
- (48) Frenkel, D.; Smit, B. *Understanding Molecular Simulations*, 2nd ed.; Academic Press: London, 2002.
- (49) Allen, J.; Valeriani, C.; Wolde, P. Forward flux sampling for rare event simulations. *J. Phys.: Condens. Matter* **2009**, *21*, 1–21.
- (50) Gao, Y.; Wang, G.; Williams, D.; Williams, S. Non-equilibrium umbrella sampling applied to force spectroscopy of soft matter. *J. Chem. Phys.* **2012**, *136*, 054902.
- (51) Bhatt, D.; Bahar, I. An adaptive weighted ensemble procedure for efficient computation of free energies and first passage rates. *J. Chem. Phys.* **2012**, *137*, 104101.
- (52) Irback, A.; Mitternacht, S. Thermal versus mechanical unfolding of ubiquitin. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 759–766.
- (53) Dudko, O. K.; Hummer, G.; Szabo, A. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.* **2006**, *96*, 108101.
- (54) Hummer, G.; Szabo, A. Free energy surfaces from single-molecule force spectroscopy. *Acc. Chem. Res.* **2005**, *38*, 504–513.
- (55) Hummer, G.; Szabo, A. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophys. J.* **2003**, *85*, 5–15.
- (56) Friddle, R. W.; Noy, A.; De Yoreo, J. J. Interpreting the widespread nonlinear force spectra of intermolecular bonds. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 13573–13578.
- (57) Zinober, R.; Brockwell, D.; Beddard, G.; Blake, A.; Olmsted, P.; Radford, S.; Smith, D. Mechanically unfolding proteins: The effect of unfolding history and the supramolecular scaffold. *Protein Sci.* **2002**, *11*, 2759–2765.
- (58) Brujic, J.; Hermans, R.; Garcia-Manyes, S.; Walther, K.; Fernandez, J. Dwell-time distribution analysis of polyprotein unfolding using force-clamp spectroscopy. *Biophys. J.* **2007**, *92*, 2896–2903.
- (59) Bertz, M.; Rief, M. Mechanical unfolds as building blocks of maltose-binding protein. *J. Mol. Biol.* **2008**, *378*, 447–458.
- (60) Bertz, M.; Rief, M. Ligand binding mechanics of maltose binding protein. *J. Mol. Biol.* **2009**, *393*, 1097–1105.
- (61) Dietz, H.; Rief, M. Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16192–16197.
- (62) 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*, 20268–20273.
- (63) Mitternacht, S.; Luccioli, S.; Torcini, A.; Imparato, A.; Irback, A. Changing the mechanical unfolding pathway of FnIII(10) by tuning the pulling strength. *Biophys. J.* **2009**, *96*, 429–441.
- (64) Kirmizialtin, S.; Huang, L.; Makarov, D. Topography of the free-energy landscape probed via mechanical unfolding of proteins. *J. Chem. Phys.* **2005**, *122*, 234915.
- (65) Garcia-Manyes, S.; Brujic, J.; Badilla, C.; Fernandez, J. Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of I27 and ubiquitin. *Biophys. J.* **2007**, *93*, 2436–2446.
- (66) Lannon, H.; Vanden-Eijnden, E.; Brujic, J. Force-clamp analysis techniques give highest rank to stretched exponential unfolding kinetics in ubiquitin. *Biophys. J.* **2012**, *103*, 2215–2222.
- (67) Korotkin, I.; Karabasov, S.; Nerukh, D.; Markesteijn, A.; Scukins, A.; Farafonov, V.; Pavlov, E. A hybrid molecular dynamics/fluctuating hydrodynamics method for modelling liquids at multiple scales in space and time. *J. Chem. Phys.* **2015**, *143*, 014110.