Alma Mater Studiorum - Università di Bologna

Dottorato di Ricerca in:
Biologia Funzionale dei Sistemi Cellulari e Molecolari - Ciclo XXV

Settore concorsuale: 03/C1 (Chimica organica)
Settore scientifico-disciplinare: CHIM/06 (Chimica organica)

# The effect of osmolytes on protein fold stability at the single-molecule level

Presentata da: Daniel Aioanei

| Coordinatore: | Relatore: | Tutore: |
|---|---|---|
| Prof. Vincenzo Scarlato | Prof. Bruno Samorì | Dr. Marco Brucale |

Esame finale anno 2013

# Preface

## Abstract

By pulling and releasing the tension on protein homomers with the Atomic Force Miscroscope (AFM) at different pulling speeds, dwell times and dwell distances, the observed force-response of the protein can be fitted with suitable theoretical models. In this respect we developed mathematical procedures and open-source computer codes for driving such experiments and fitting Bells model to experimental protein unfolding forces and protein folding frequencies.

We applied the above techniques to the study of proteins GB1 (the B1 IgG-binding domain of protein G from Streptococcus) and I27 (a module of human cardiac titin) in aqueous solutions of protecting osmolytes such as dimethyl sulfoxide (DMSO), glycerol and trimethylamine N-oxide (TMAO). In order to get a molecular understanding of the experimental results we developed an Ising-like model for proteins that incorporates the osmophobic nature of their backbone. The model benefits from analytical thermodynamics and kinetics amenable to Monte-Carlo simulation.

The prevailing view used to be that small protecting osmolytes bridge the separating beta-strands of proteins with mechanical resistance, presumably shifting the transition state to significantly higher distances that correlate with the molecular size of the osmolyte molecules. Our experiments showed instead that protecting osmolytes slow down protein unfolding and speed-up protein folding at physiological pH without shifting the protein transition state on the mechanical reaction coordinate. Together with the theoretical results of the Ising-model, our results lend support to the osmophobic theory according to which osmolyte stabilisation is a result of the preferential exclusion of the osmolyte molecules from the protein

backbone.

The results obtained during this thesis work have markedly improved our understanding of the strategy selected by Nature to strengthen protein stability in hostile environments, shifting the focus from hypothetical protein-osmolyte interactions to the more general mechanism based on the osmophobicity of the protein backbone.

# Publications

The following peer-reviwed journal articles in chronological order are the result of work either leading to or performed during the PhD studies:

1. Aioanei D, Samorì B, Brucale M. “Maximum likelihood estimation of protein kinetic parameters under weak assumptions from unfolding force spectroscopy experiments”, Phys Rev E Stat Nonlin Soft Matter Phys. 2009 Dec;80(6 Pt 1):061916

2. Aioanei D, Brucale M, Samorì B. “Open source platform for the execution and analysis of mechanical refolding experiments”, Bioinformatics. 2011 Feb 1;27(3):423-5

3. Aioanei D, Lv S, Tessari I, Rampioni A, Bubacco L, Li H, Samorì B, Brucale M. “Single-molecule-level evidence for the osmophobic effect”, Angew Chem Int Ed Engl. 2011 May 2;50(19):4394-7

4. Aioanei D, Tessari I, Bubacco L, Samorì B, Brucale M. “Observing the osmophobic effect in action at the single molecule level”, Proteins. 2011 Jul;79(7):2214-23

5. Aioanei D, Brucale M, Tessari I, Bubacco L, Samorì B. “Worm-like Ising model for protein mechanical unfolding under the effect of osmolytes”, Biophys J. 2012 Jan 18;102(2):342-50

6. Aioanei D. “Lazy Shortest Path Computation in Dynamic Graphs”, Computer Science, 13(3), AGH University of Science and Technology Press, 2012, pp. 113-137, ISSN 1508-2806

7. Pivato M, De Franceschi G, Tosatto L, Frare E, Kumar D, Aioanei D, Brucale M, Tessari I, Bisaglia M, Samorì B, Polverino de Laureto P, Bubacco L. “Covalent $\alpha$-Synuclein Dimers: Chemico-Physical and Aggregation Properties”, PLoS ONE 7(12): e50027

# Contents

# Chapter 1

# Introduction

## 1.1 Atomic Force Microscopy

The atomic force microscope (AFM) has found countless uses for imaging nanostructures and measuring and manipulating matter at the nanoscale.

As an imaging device the AFM works as a scanning probe microscope with resolution three orders of magnitude better than the optical diffraction limit, producing true three-dimensional measurement profiles. Moreover, it is amenable to the measurement of biological samples since it does not require any special treatments such as carbon/metal coatings that may damage such samples, and it works under both ambient air and liquid environments. The AFM supports two main imaging modes of operation. *a*) In *contact mode* the force applied by the tip on the surface is maintained constant during scanning by a feedback mechanism, but it can easily result in damage to the sample. *b*) In *tapping mode* instead the cantilever is driven to oscillate near its resonance frequency by a small piezo incorporated in the AFM tip holder. Tapping mode is gentle enough to afford the measurement of single polymer molecules or supported lipid bylayers. With properly tuned scanning parameters single molecules can maintain their conformation even for hours [2], allowing accurate measurements to be performed.

The other common use of the AFM is to measure and induce mechanical tension in the field of force-spectroscopy. In this case force-distance curves are recorded by measuring the tip-sample interaction as a function of the distance between the tip and the sample. The AFM has been employed as a force-spectroscopy tool to measure the disjoining pressure of lubricant nanofilms [2], the mechanical stability of globular proteins [3], membrane proteins [4], mechano-chemical switches [5] and even to assess the differentiation potential of stem cells [6].

### 1.1.1 AFM architecture

Crucial to AFM operation is a cantilever with a sharp tip pointing out at its end (see Fig. 1.1).

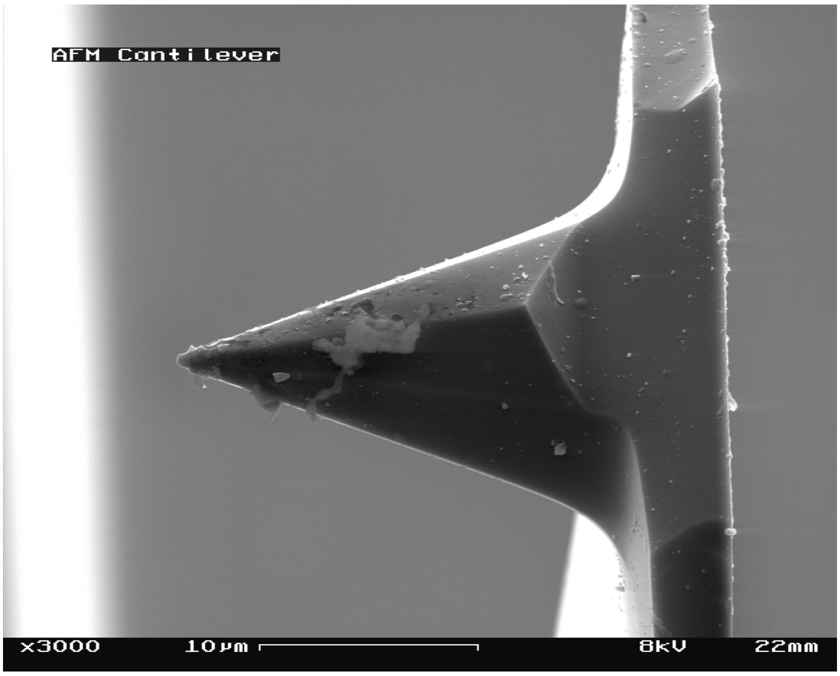


Figure 1.1: Used AFM cantilever. Image from Wikimedia Commons.

The cantilever is commonly made up of silicon nitride and coated with gold, with a radius of curvature in the nanometres range. When the tip is brought in the proximity of the sample, forces between the two lead to a deflection of the cantilever that is assumed to abide by Hooke's law. In the most common setup the cantilever deflection is measured using a laser spot reflection onto an array of photodiodes (see Fig. 1.2), and a beam path several centimetres long serves as an angle amplifier.

Commercial cantilevers come with spring constants from a few pN/nm to a few hundred pN/nm. By measuring the deflection with nanometre precision, the

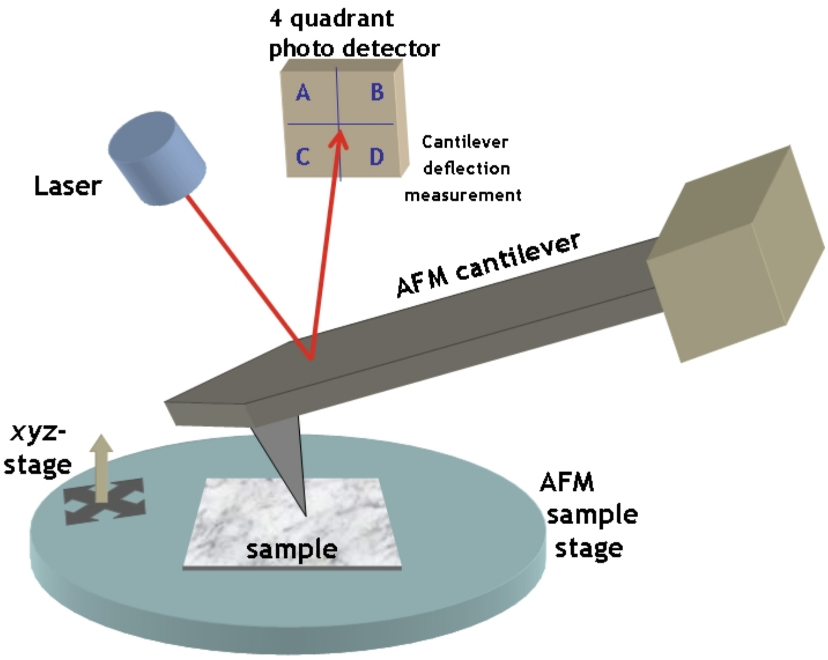


Figure 1.2: Atomic force microscope setup. Image from Wikimedia Commons.

AFM therefore is a force sensor in the piconewton range.

### 1.1.2 Cantilever calibration

In order to obtain precise force measurements the optical lever sensitivity as well as the spring constant of the cantilever must be accurately determined.

The optical lever sensitivity is the ratio between cantilever deflection and the photodiode output voltage difference. Since it depends on the optical properties of the cantilever and the medium it needs to be measured for each cantilever by pressing the cantilever onto a hard surface (e.g. glass, mica) at high forces. Assuming that the substrate does not deform, the cantilever deflection must be equal to the piezo movement, thereby affording the calculation of the desired ratio.

The Hookean spring constant of the cantilever is expected to be strongly dependent on its thickness, which is very difficult to control during manufacture. Therefore the *thermal tune* method [7] is used to estimate the spring constant of the cantilever starting from the equipartition theorem that relates the average fluctuation of the cantilever to the thermal energy:

$$k = \frac{k_b T}{\langle z_c^2 \rangle}\,, \tag{1.1}$$

where $k_b$ is the Boltzmann constant, $T$ is the absolute temperature and $\langle z_c^2 \rangle$ is the square mean cantilever deflection. The value $\langle z_c^2 \rangle$ is found by fitting the power spectrum of the cantilever thermal noise, hence avoiding interference from either non-thermal oscillation at other discrete frequencies or white noise. The typical relative error is about ± 20% [7, 8], and other calibration methods have also been developed [9, 10, 11].

### 1.1.3 Stretching homomeric protein chains

In order to study protein folding and unfolding kinetics with the AFM, constructs consisting of a tandem repeat of the protein under study are often employed. There are two main advantages that such a construct offers. *a*) First, in order to move the interesting events far away from the surfaces where aspecific interactions dominate, a longer construct is preferred. *b*) Second, the tip curvature radius is about one order of magnitude larger than the studied proteins, and in a homomeric construct the protein modules forming the bridge with the tip or surface can be sacrifised, leaving the protein modules in-between for study.

Fig. 1.3 shows some example curves obtained using protocols such as velocity-clamp, force-clamp, force-ramp or double-pulse velocity clamp with homomeric $(\mathrm{GB1})_8$ or $(\mathrm{GB1})_{16}$ constructs, where GB1 stands for the immunoglobulin-binding domain of streptococcal protein G.

## 1.2 The osmophobic effect

Throughout all kingdoms of life denaturing stresses are opposed by small organic molecules known as *protecting osmolytes* or *osmoprotectants.* Their presence is characteristic of both plants and animals that have adapted to environmental extremes, having been naturally selected to help avoid life-threatening widespread protein unfolding. Notably, protecting osmolytes make it possible for bdelloid rotifers, tardigrades, brine shrimp and nematodes to survive under conditions of complete dessication and enter a state of suspended animation know as *cryptobiosis.*

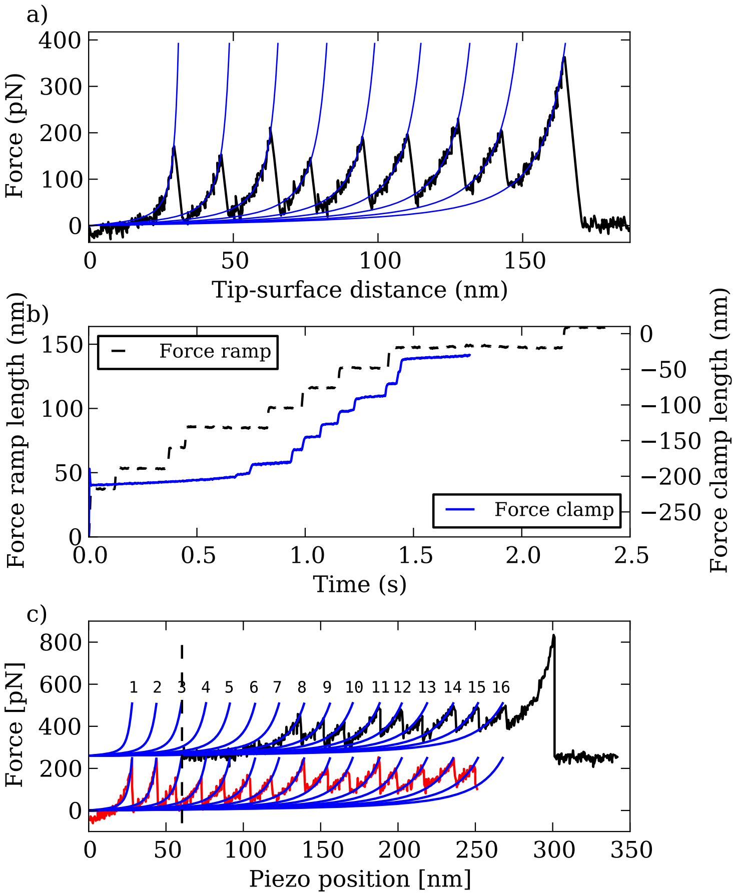


Figure 1.3: Representative examples of the type of data captured under different AFM modes of force-spectroscopy operation with $(\mathrm{GB1})_8$ or $(\mathrm{GB1})_{16}$: protein unfolding under *a*) velocity-clamp, *b*) force-clamp (continuous line) and force-ramp (dashed line), and *c*) protein refolding using a double-pulse protocol [12].

It has been proposed that protecting osmolytes destabilise the protein denatured state via an unfavourable interaction of the protecting osmolyte molecules with the peptide backbone that is exposed upon protein denaturation. This unfavourable interaction, know as the *osmophobic effect*, is attenuated by an overall favourable interaction of the osmolyte molecules with the protein side-chains, which is however of lesser magnitude. While hydrophobic interactions, hydrogen bonding, electrostatic and dispersion forces have long been considered as fundamental thermodynamic forces involved in protein folding, osmolytes have little effect on these forces. Hence the osmophobic effect is separate and complementary to these more established forces [13].

However, while the thermodynamic osmophobic force is becoming a rather well accepted theory, little is known about *a*) the effect of osmolytes on protein folding and unfolding kinetics, and *b*) the molecular details by which such effects may come about. Moreover, some alternative theories have also been proposed for their mode of action, such as changing folding pathways [14] or bridging the separating beta strands of proteins under mechanical tension [15, 16]. To understand the mode of action of osmolytes we employed AFM-based single-molecule force spectroscopy (SMFS) and theoretical modelling of proteins in the presence and absence of various concentrations of osmolytes, as explained in Section 1.3.

## 1.3 The osmolyte effect studied at the single-molecule level

In order to study the effect of osmolytes at the single-molecule level we followed a phased approach with 6 stages, where the last one is a work-in-progress, while the other 5 have each resulted in a scientific publication:

### 1.3.1 We developed a Maximum-Likelihood procedure for fitting Bell's model to experimental velocity-clamp protein unfolding forces.

Before using SMFS to investigate the effect of osmolytes, we needed a method to extract energy landscape information from force-spectroscopy data. Due to a lack in the literature of fast and accurate methods suitable for this task, we developed a Maximum-Likelihood estimation procedure to extract the two main parameters that describe protein unfolding according to Bell's model, namely the spontaneous unfolding rate $k_u(0)$ and the unfolding distance $\Delta x_u$. The method is described in Chapter 2 and Ref. [17].

### 1.3.2 We developed open-source software to drive and analyse mechanical protein refolding experiments.

Protein folding complements protein unfolding to describe the energy landscape of proteins. Therefore we developed open-source software to drive and analyse protein folding under mechanical tension, and to perform Maximum-Likelihood estimation of the two main parameters that describe protein folding according to Bell's model, namely the spontaneous folding rate $k_f(0)$ and the folding distance $\Delta x_f$. The software codes and the estimation method are decribed in Chapter 3 and Ref. [18].

### 1.3.3 Dimethyl sulfoxide (DMSO) enhances the mechanical stability of protein GB1.

Once we could perform and analyse protein folding and unfolding experiments, we set off to study the effect of osmolytes on these protein kinetic processes. We started with the protecting osmolyte DMSO. Our experiments showed that DMSO accelerates the folding and decelerates the unfolding of protein GB1 without shifting the transition state of the protein on the mechanical reaction coordinate. We noted that since the transition state likely exposes more backbone than the native state, but less than the denatured state, our experimental results were explainable by applying the osmophobic effect not only to the native and denatured states, but also to the transition state of the protein. The experiments and their interpretation are described in Chapter 4 and Ref. [12].

### 1.3.4 The osmophobic hypothesis is supported by mechanical unfolding experiments of GB1 in the presence of glycerol and theoretical calculations with an Ising model.

In order to see how generic was our finding that DMSO stabilises protein GB1 without shifting its transition state, we decided to look at another osmolyte, another reaction coordinate, and later on (see next section) another protein. We started by pulling protein GB1 in the presence of the protecting osmolyte glycerol, finding that it stabilises the native state of GB1 against mechanical unfolding without changing its unfolding distance. To see if this behaviour carries over to a non-mechanical unfolding process, we extended a protein Ising model with support for group-transfer free energies, effectively incorporating the osmophobic effect into the Hamiltonian of the system. We achieved this extension while maintaining exact (polynomial-time) computation of the model's thermodynamics. Using this approach we found that the osmophobic effect does not shift the transition state of GB1 when its energy landscape is projected onto the reaction coordinate defined by the number of native peptide bonds or the weighted number of contacts. The experiments together with the theoretical developments are described in Chapter 5 and Ref. [19].

### 1.3.5 Experiments with protein I27 and glycerol supported by kinetic Monte-Carlo simulations buttress the osmophobic hypothesis.

We pulled protein I27, a human cardiac titin module, in the presence and absence of glycerol. Contrary to a previous report [15], we found that glycerol molecules do not increase the unfolding distance of protein I27. We also performed kinetic Monte-Carlo simulations of the Ising model with support for group transfer energies by first adding to it worm-like chain (WLC) behaviour while preserving exactly solvable thermodynamics. The simulations confirmed our previous hypothesis that the osmophobic effect can explain our experimental data. The experiments together with the theoretical developments are described in Chapter 6 and Ref. [20].

### 1.3.6 Mechanical unfolding of protein GB1 in the presence of Trimethylamine N-oxide (TMAO) shows that at low pH effects other than the osmophobic force dominate.

In order to look for evidence of effects other than the osmophobicity of the protein backbone, we turned to reports of bulk experiments that found protecting osmolytes to behave like destabilisers in extreme conditions. In particular, it has been previously found via bulk thermodynamic experiments that TMAO destabilises three different proteins at low pH, when TMAO is positively charged [21, 22]. By pulling protein GB1 across a set of different pH values from 3.5 to 7.0, we found that at each pH TMAO leaves the unfolding distance of GB1 largely unaffected. Moreover, we found that TMAO stabilises the native state of GB1 against mechanical unfolding around physiological pH, but at low pH it has a destabilising effect. We adopted a statistical mechanics model for osmolyte-backbone interactions [1] to estimate the energetics of the backbone-protonated TMAO interactions. Taken together with the preliminary experimental results, these findings suggest that effects other than osmophobic character of the protein backbone dominate at low pH in the presence of TMAO. More details can be found in Chapter 7.

# Chapter 2

# Maximum likelihood estimation of protein kinetic parameters under weak assumptions from unfolding force spectroscopy experiments

Adapted with permission from Aioanei D, Samorì B, Brucale M. “Maximum likelihood estimation of protein kinetic parameters under weak assumptions from unfolding force spectroscopy experiments”, Phys Rev E Stat Nonlin Soft Matter Phys. 2009 Dec;80(6 Pt 1):061916. Copyright (2009) by the American Physical Society.

## 2.1 Abstract

Single molecule force spectroscopy (SMFS) is extensively used to characterize the mechanical unfolding behavior of individual protein domains under applied force by pulling chimeric polyproteins consisting of identical tandem repeats. Constant velocity unfolding SMFS data can be employed to reconstruct the protein unfolding energy landscape and kinetics. The methods applied so far require the specification of a single stretching force increase function, either theoretically derived or experimentally inferred, which must then be assumed to accurately describe the entirety of the experimental data. The very existence of a suitable optimal force

model, even in the context of a single experimental dataset, is still questioned. Herein, we propose a Maximum Likelihood (ML) framework for the estimation of protein kinetic parameters which can accommodate all the established theoretical force increase models. Our framework does not presuppose the existence of a single force characteristic function. Rather, it can be used with a heterogeneous set of functions, each describing the protein behavior in the stretching time range leading to one rupture event. We propose a simple way of constructing such a set of functions via piecewise linear approximation of the SMFS force vs. time data, and we prove the suitability of the approach both with synthetic data and experimentally. Additionally, when the spontaneous unfolding rate is the only unknown parameter, we find a correction factor that eliminates the bias of the ML-estimator while also reducing its variance. Finally, we investigate which of several time-constrained experiment designs leads to better estimators.

## 2.2 Introduction

The kinetics of protein unfolding under mechanical stress represent a very important topic in the field of biophysics as can be seen in the large number of reviews surveying the problem [23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35]. The most widely employed model for studying mechanical protein unfolding sees bond rupture as a decay of a metastable state with reaction kinetics given by

$$\frac{d\eta}{dt} = -k(f(t))\eta(t) \tag{2.1}$$

where $\eta(t)$ is the survival probability up to time $t$, $f(t)$ stands for the force at time $t$ and $k(f(t))$ is the dissociation rate [36]. The dependence of the dissociation rate on force was given in [37] the analytical formula

$$k(f) = k_0 e^{\alpha f} \tag{2.2}$$

where $\alpha = x_\beta/(k_B T)$ with $x_\beta$ standing for the position of the transition state along the mechanical reaction coordinate and $k_0$ being the spontaneous dissociation rate.

The two parameters $k_0$ and $\alpha$ are usually extracted by either of two approaches. The first one, sometimes called the “standard method” (see e.g. [38, 39]), involves gaussian fits of the rupture force distributions for various loading rates and a linear regression between the most probable rupture forces and the logarithm of the loading rate, without taking into account the joint effects of multiple modules that unfold sequentially in the context of polyproteins. The second approach is based on Monte Carlo simulations (see e.g. [40, 23, 41, 42, 43, 44, 45]), and in this case the mentioned joint effects are properly accounted for. However both traditional methods have intrinsic shortcomings: they either throw away useful information by summarizing the data into statistics that are not *sufficient*, or geometrically fitting as closely as possible quantities that are not of prime interest, such as linear dependencies or rupture force distributions, rather than focusing directly on finding the most probable kinetic parameters [46]. To overcome these shortcomings a Maximum Likelihood (ML) approach has been previously proposed [46, 47].

We have further developed the ML approach in order to address the following problems:

1. The probability to observe an unfolding event is a contextual feature of homomeric polyproteins.

2. In real experiments a unique force-time, and likewise force-displacement, characteristic does not exist (see e.g. [47]). When either the cantilever tip or the surface is not functionalized, which is often the case, this is in fact predicted by the theoretical models since they depend on microscopic parameters that vary based on the length of the subrange under mechanical stress and even from spot to spot depending on the local properties of the soft protein layer [48, 49].

3. Choosing one among the many existing theoretical force models for idealized polymeric chains, as reviewed e.g. in [46, 50, 51, 52, 53], with various corrections of the interpolation formulas [54, 55], is not trivial, and neither is deriving an empirical force model from the experimental data itself [47].

We tackle the first problem by taking into account the number of not-yet-unfolded modules when computing the survival probability in Eq. (2.3). We solve the second problem by allowing in Eq. (2.4) a different force-time function to describe the stretching time range leading to each unfolding event, as long as they are considered known (i.e., not introducing *nuisance parameters* into the likelihood function). Finally we address the third problem by constructing the force-time functions in an automated, fully objective way as increasing, continuous piecewise linear approximations to the AFM-recorded data points, using Eq. (2.13).

We show that $\alpha$ can be estimated just by maximizing the univariate function in Eq. (2.7), after which $k_0$ immediately comes out from Eq. (2.6). Since the statistical estimation procedure would not be complete without a way to compute the uncertainty of the estimated values [56], with Eq. (2.9) we show how to extract a *Bayesian credible region*, i.e., a fixed two-dimensional area that contains with a given probability the random point $(k_0, \alpha)$ (see e.g. [57]). The approach is computationally feasible even for complex theoretical models such as Worm-Like Chain (WLC) [58, 59] that require numeric integration for the evaluation of the likelihood function. In fact WLC has never been used before in the context of ML estimation of kinetic parameters, but it can be easily applied with our framework by solving Eq. (2.12).

It should be emphasized that it is common practice to fix $\alpha$ to a known value and estimate only $k_0$ in situations that are believed not to alter the position of the transition state: replacing water by deuterium oxide [60], certain protein mutations [61, 62, 44], and stretching proteins under the effect of chemical denaturants [63]. For this particular case we propose the *unbiased* and *more efficient* estimator given by Eq. (2.8).

## 2.3 Theory

Next we are going to present the analytical form of the likelihood function, we'll explain how it can be maximized and how to compute a credible region for the two parameters.

### 2.3.1 Likelihood function

When a monomeric protein is stretched starting with time $t_s$, from Eq. (2.1) and imposing that $\eta(t_s) = 1$ we obtain

$$\eta(t) = \exp\left[-\int_{t_s}^{t} k(f(u))\,du\right],\ t \geq t_s.$$

For a multimeric construct made up of identical tandem repeats behaving independently, let's consider an unfolding event after which, chronologically, there are $m - 1 \geq 0$ more unfolding events in the SMFS curve. The probability that all $m$ modules survive becomes

$$\eta_m(t) = \exp\left[-m \int_{t_s}^{t} k(f(u))\,du\right],\ t \geq t_s\ . \tag{2.3}$$

Assuming that $f(t)$ is continuous and increasing with $f(t_s) = y_s$, we can change the integration domain to force:

$$\eta_m(y) = \exp\left[-m \int_{y_s}^{y} k(z)(f^{-1})'(z)\,dz\right],\ y \geq y_s\ .$$

The probability density $r_m(y)$ to observe a rupture event at force $y \geq y_s$ is

$$\begin{aligned} r_m(y) = \quad & -\frac{d}{dy}\eta(y) = mk(y)(f^{-1})'(y) \\ & \times \exp\left[-m \int_{y_s}^{y} k(z)(f^{-1})'(z)\,dz\right]\ . \end{aligned}$$

**Notation 1** *Let $n$ be the total number of unfolding peaks in the whole data set, and for each unfolding event $1 \leq i \leq n$ we denote by $t_{si}$ and $y_{si}$ the time point and force at which we consider the stretching to start, by $t_i > t_{si}$ and $y_i > y_{si}$ the rupture time instant and force of rupture, and by $m_i$ the number of modules that will unfold after $i$ in the same curve, plus one. The force-time function for peak $i$, from $t_{si}$ to $t_i$, is described by $f_i$.*

Since the unfolding events are independent of each other, the joint probability density function associated to the rupture forces $\vec{y} = (y_1 \ldots y_n)$ is

$$\begin{aligned} L(\vec{y}; k) = \quad & \exp\left[-\sum_{i=1}^{n} m_i \int_{y_{si}}^{y_i} k(z)(f_i^{-1})'(z)\,dz\right] \\ & \times \prod_{i=1}^{n} m_i k(y_i)(f_i^{-1})'(y_i)\ . \end{aligned} \tag{2.4}$$

At this point we introduce $\alpha$ and $k_0$ explicitly into the joint probability density function by using Eq. (2.2):

$$\begin{aligned} L(\vec{y}; k_0, \alpha) = \quad & \exp\left[-k_0 \sum_{i=1}^{n} m_i \int_{y_{si}}^{y_i} e^{\alpha z}(f_i^{-1})'(z)\,dz\right] \\ & \times k_0^n \prod_{i=1}^{n} m_i e^{\alpha y_i}(f_i^{-1})'(y_i)\ . \end{aligned} \tag{2.5}$$

Note that $f_i$ can be any continuous increasing function. Appendix 2.8.1 contains more details about the computation of the derivative of the inverse force-time function when the WLC model is assumed (Eq. (2.12)), and explicit formulas for the likelihood function when applied to the linear force-displacement characteristic $f(t) = \kappa v t$ (Eq. (2.10)) or the piecewise linear force-time approximation (Eq. (2.13)).

Briefly, the piecewise linear force-time approximation is a linear interpolation of a *longest increasing subsequence* (see e.g. [64, 65]) of the force values reported by the AFM during stretching, which is computed by removing the minimal number of data points such that the remaining ones show increasing force with time, and breaking ties by calling for increased time resolution towards the rupture event (see Appendix 2.8.1). This approach eliminates most of the noise and it has the nice theoretical property that if applied to a set of forces that is already increasing, it becomes a simple linear interpolation.

### 2.3.2 Point estimation

Regarding $L(\vec{y}; k_0, \alpha)$ as a function of $k_0$, the *conditional maximum-likelihood estimate* of $k_0$ is the argument for which the function achieves the global maximum on $(0, \infty)$, and can be computed as

$$\widehat{k_0}(\alpha) = \frac{n}{\sum_{i=1}^{n} m_i \int_{y_{si}}^{y_i} e^{\alpha z} (f_i^{-1})'(z)\, dz} \; . \tag{2.6}$$

Substituting in Eq. (2.5) we obtain the *profile likelihood* for $\alpha$, which needs to be maximized numerically to obtain the estimator $\widehat{\alpha}$:

$$L_p(\alpha) = \quad \widehat{k_0}(\alpha)^n exp \left[ \alpha \sum_{i=1}^{n} y_i \right] e^{-n} \prod_{i=1}^{n} m_i (f_i^{-1})'(y_i) \; . \tag{2.7}$$

While ML-estimators are known to have very good asymptotic properties when used with *i.i.d.* (independent and identically distributed) random variables (see e.g. [66, 67]), our rupture forces $y_i$ are not identically distributed because each unfolding event $i$ is assigned its own $y_{si}$, $m_i$ and $f_i$. As a result a more complex theory, such as perhaps that developed in [68] would be needed to study the asymptotic behavior of our ML-estimators, but a rigorous treatment of the problem would exceed the scope of the present paper.

However we do show in Appendix 2.8.2 that when $\alpha$ is fixed and known, and under the conditions of Proposition 2, which can be shown to hold for the WLC interpolation formula in Eq. (2.11) and for any force-time function that increases linearly after an arbitrary time point, the estimator $\widehat{k_0}(\alpha)$ is biased. For this situation we propose the following unbiased estimator of $k_0$:

$$\widetilde{k_0}(\alpha) = (n-1)\widehat{k_0}(\alpha)/n, \; n \geq 2 \; . \tag{2.8}$$

Since $Var[\widetilde{k_0}(\alpha)] = [(n-1)/n]^2 Var[\widehat{k_0}(\alpha)]$ the unbiased estimator is also *more efficient*.

### 2.3.3 Bayesian credible region

We show here that the particular shape of our likelihood function makes it feasible to numerically compute a (rectangular) credible region for $(k_0, \alpha)$ containing the respective point estimates.

A key operation in the numerical computation of credible regions is the ability to efficiently integrate the likelihood function over (potentially infinite) rectangular regions. For this purpose we make the following observation:

$$\begin{aligned} \int_a^b x^n e^{-cx}\, dx = & \quad \frac{1}{c^{n+1}} \int_{ca}^{cb} y^n e^{-y}\, dy \\ = & \quad \frac{n!}{c^{n+1}} \left[ P(n+1, bc) - P(n+1, ac) \right] \end{aligned}$$

for any $0 \leq a < b \leq \infty$, $c > 0$ and integer $n \geq 0$, where $P$ is the *incomplete gamma function* defined as

$$P(h, x) = \frac{1}{\Gamma(h)} \int_0^x t^{h-1} e^{-t}\, dt, \; h > 0 \; .$$

The integral of the likelihood function on a rectangular region then simplifies as:

$$\begin{aligned} L_{(k_{0s},\alpha_s)}^{(k_{0e},\alpha_e)} &= \int_{k_{0s}}^{k_{0e}} \int_{\alpha_s}^{\alpha_e} L(\vec{y}; k_0, \alpha)\, d\alpha\, dk_0 \\ &= \int_{\alpha_s}^{\alpha_e} \frac{P(n+1, k_{0e}c(\alpha, \vec{y})) - P(n+1, k_{0s}c(\alpha, \vec{y}))}{c(\alpha, \vec{y})^{n+1}} \\ &\quad \times \exp\left[\alpha \sum_{i=1}^{n} y_i\right] d\alpha \\ &\quad \times n! \prod_{i=1}^{n} m_i (f_i^{-1})'(y_i) \end{aligned} \tag{2.9}$$

with

$$c(\alpha, \vec{y}) = \sum_{i=1}^{n} m_i \int_{y_{si}}^{y_i} e^{\alpha z} (f_i^{-1})'(z)\, dz \ .$$

A $(1-p)$ credible region of $(k_0, \alpha)$ can then be found as a rectangular area $(k_{0s}, k_{0e}) \times (\alpha_{0s}, \alpha_{0e})$ that includes the point estimates $(\widehat{k_0}, \widehat{\alpha})$ such that $L_{(k_{0s},\alpha_s)}^{(k_{0e},\alpha_e)} / L_{(0,0)}^{(\infty,\infty)} = 1 - p$. That would imply the usage of an improper, unbounded uniform prior distribution.

Alternatively $k_0$ and $\alpha$ can be restricted to a finite, more physically feasible region via a (proper) uniform prior distribution, and it is indeed common practice to do so with Monte Carlo methods which sample only a particular domain of interest (see e.g. [44]).

## 2.4 Validation

We present below three applications: a synthetic experiment for the situation when $\alpha$ is fixed and known, another synthetic experiment to check the suitability of the linear force-displacement model and the piecewise linear approximation with WLC-conforming data under a few time-constrained design strategies, and finally a real SMFS experiment with a well characterized protein.

### 2.4.1 The unbiased estimator $\widetilde{k_0}(\alpha)$ is indeed a better estimator than the biased one

In order to confirm the theoretical prediction that $\widetilde{k_0}(\alpha)$ is not only unbiased, but also a better estimator than $\widehat{k_0}(\alpha)$ in terms of showing smaller Root Mean

Square Error (RMSE), we simulated pulling a multimeric construct made up of 20 identical modules whose length, spontaneous unfolding rate and position of the transition state were chosen to match those previously reported for a real protein, namely the B1 immunoglobulin-binding domain of protein G from Streptococcus (GB1) [69].

Therefore we used $k_0 = 0.039\ s^{-1}$, $x_\beta$=0.17 nm, T=301.15 K, and for each $n$ in 2, ..., 101 we generated 10000 datasets of $n$ unfolding events each following the linear force vs. displacement characteristic with a cantilever spring constant of 0.07 N/m. The $n$ events were generated giving roughly equal shares to each of the velocities $2^{-i}$ 2180 nm/s, i = 0, . . . , 5. The starting force of pulling was randomly chosen within a range compatible with what is commonly observed experimentally, and the number of not-yet-unfolded modules was varied between 1 and 20.

Under these conditions the observed mean of $\widetilde{k_0}(\alpha)$ was always very close to the theoretical expected value of 0.039, even for n=2, while the observed RMSE went down from about 0.0744 $s^{-1}$ for n=2 to about 0.0039 $s^{-1}$ for n=101 (data not shown). The biased estimator instead showed large bias for small values of $n$ and higher RMSE all throughout (see Fig. 2.1) thus confirming the theoretical prediction that the unbiased estimator $\widetilde{k_0}(\alpha)$ is better than the biased estimator $\widehat{k_0}(\alpha)$.

### 2.4.2 Synthetic WLC data is well approximated by the piecewise-linear function

We performed a more comprehensive simulation in order to compare the ability to recover the kinetic parameters with the various approaches discussed so far, when the data are generated by using the WLC model. We focused specifically on the importance of the intermediate data points from the start of the stretching process up to the rupture event, which are roughly approximated by the piecewise linear force-time function, but not taken into account by the widespread linear force-displacement model.

A second goal of this simulation was to investigate the efficiency of a few different experiment designs in terms of spreading a fixed amount of experimental time across different pulling velocities, and checking which strategy leads to estimates with better performance.

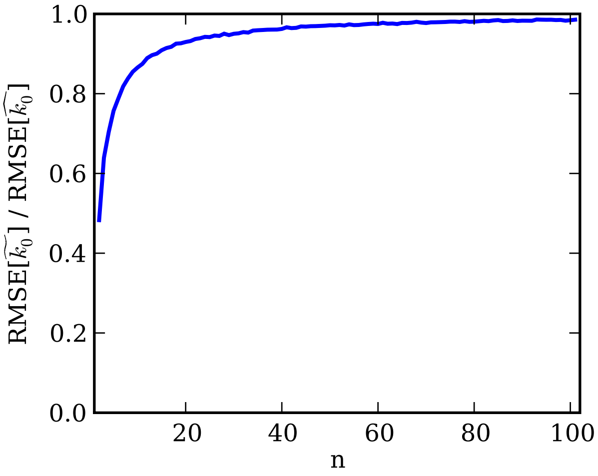


Figure 2.1: Ratio of $RMSE[\widetilde{k_0}(\alpha)]$ over $RMSE[\widehat{k_0}(\alpha)]$ as a function of the number of unfolding events.

The kinetic parameters and the cantilever spring constant were kept the same as in the previous synthetic data experiment, while the protein pick-up rate was set to 100%. Each trial simulated about 107.87 seconds of experimental time with a surface delay of 200 ms, an approach speed of 4360 nm/s, four unfolding modules in each curve, a piezo range of 500 nm and 2048 sample points per curve, all of which are very reasonable values commonly used in real experiments. Six retraction speeds have been used, namely 125, 249, 545, 1090, 2180 and 4360 nm/s, and six experiment design strategies were covered: Lowest Speed Only (LSO) with 25 curves at the lowest velocity, Highest and Lowest Equal Number (HLEN) with 22 curves at the lowest speed and 23 at the highest one, Highest and Lowest Equal Time (HLET) with 12 curves at the lowest speed and 126 at the highest one, All Speeds Equal Number (ASEN) with 11, 11, 11, 12, 12, and 12 curves respectively, in increasing velocity order, All Speeds Equal Time (ASET) with 4, 7, 15, 24, 34, 42 curves respectively, in increasing velocity order, and finally Highest Speed Only (HSO) with 251 curves at the highest velocity. The number of trials was 1000 and all the data were generated using the WLC interpolation formula of Eq. (2.11) with a persistence length of 0.35 nm.

The data were analyzed using our ML framework with the following force function types: linear force-displacement (L), piecewise linear force (PL), WLC and finally Gaussian piecewise linear force (GPL). For the last-mentioned one we kept the rupture force unchanged, to allow for a reasonably fair comparison to the other approaches, but added noise with a standard deviation of 20 pN [44] to all the other data points in order to check how well the piecewise linear force approximation is able to tackle noise by selecting only the longest increasing subsequence of force values, or indeed how much the remaining inaccuracies matter.

Additionally, we also analyzed the HLEN, HLET, ASEN and ASET data sets using the "standard method" (SM) as reviewed in [38] and the Monte Carlo method (MC) analysing the speed dependence of the unfolding force as reviewed in [40, 45, 42, 43, 44]. Briefly, the "standard method" consists in fitting a linear dependence between the most probable rupture force and the logarithm of the loading rate $\kappa v$. The two kinetic parameters are then computed from the slope and intercept of the fitting line implicitly adopting the assumption that the force behavior during stretching can be satisfactorily approximated by the linear force-displacement characteristic. The Monte Carlo method instead consists in the simulation of thousands of synthetic curves on a two-dimensional grid of $k_0$ and $x_\beta$ parameters, and then selecting the combination of parameters that best match the experimental mean unfolding force dependence on velocity.

The SM approach produced the worst results where applicable, next followed by the L approach, the results of both being displayed in Fig. 2.2 and 2.3. The observed bias was around one-two orders of magnitude for $k_0$ and not too small for $x_\beta$ either, therefore raising a signal flag about the dangers of applying the wrong theoretical force model, in this case using the linear force-displacement characteristic when the underlying data has been generated using the WLC model. The assumed general applicability of the linear force-displacement characteristic has also been previously disproved with experimental data in [47].

Figures 2.4 and 2.5 contain the results for the other methods. Except for a little bias at the lowest speed, the ML-based WLC approach worked very well, as expected since the data were generated with the same model, and no noise was added.

Surprisingly the performance of the PL approach was almost indistinguishable from that of the WLC-assuming approach. This is quite significant since the PL approach does not take into account the fact that the data were generated with

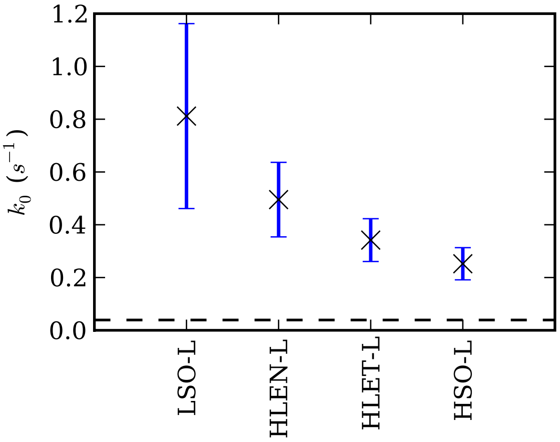


Figure 2.2: Estimation of $k_0$ using the linear force-displacement model. Data were generated using the strategies LSO, HLEN, HLET, ASEN, ASET and HSO, from left to right. Estimation was performed using maximum likelihood (L) and the "standard method" (SM). The dashed horizontal line indicates the actual $k_0$ value used for data generation. The crosses mark the mean and the error bars extend one standard deviation in both directions.

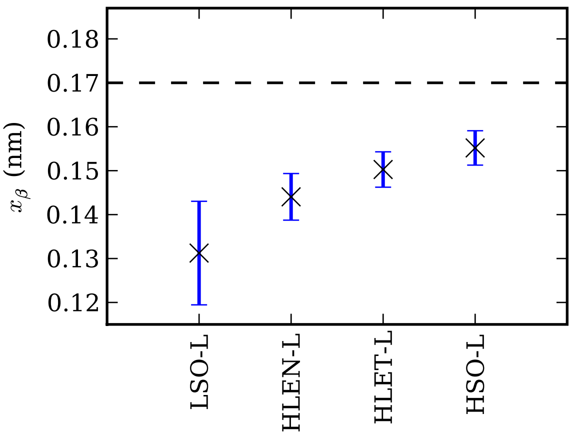


Figure 2.3: Estimation of $x_\beta$ using the linear force-displacement model. Data were generated using the strategies LSO, HLEN, HLET, ASEN, ASET and HSO, from left to right. Estimation was performed using maximum likelihood (L) and the "standard method" (SM). The dashed horizontal line indicates the actual $x_\beta$ value used for data generation. The crosses mark the mean and the error bars extend one standard deviation in both directions.

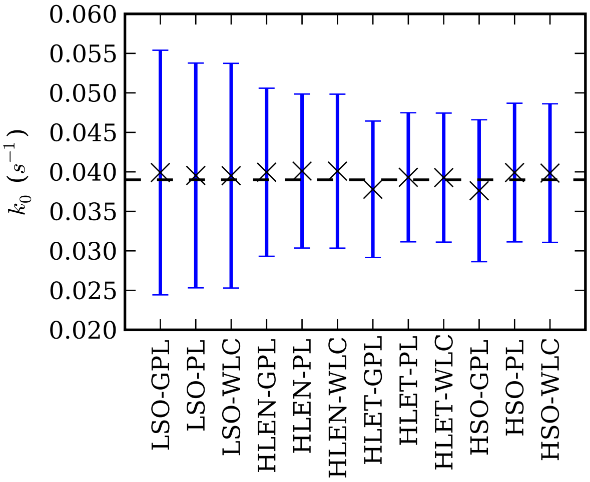


Figure 2.4: Estimation of $k_0$ using GPL, PL and WLC. Data were generated using the strategies LSO, HLEN, HLET, ASEN, ASET and HSO, from left to right. The dashed horizontal line indicates the actual $k_0$ value used for data generation. The crosses mark the mean and the error bars extend one standard deviation in both directions.

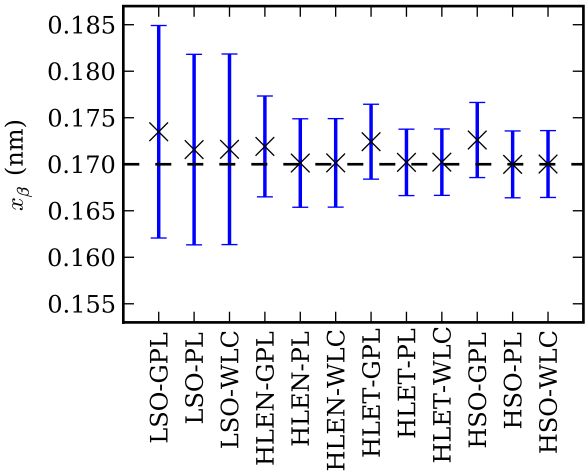


Figure 2.5: Estimation of $x_\beta$ using GPL, PL and WLC. Data were generated using the strategies LSO, HLEN, HLET, ASEN, ASET and HSO, from left to right. The dashed horizontal line indicates the actual $x_\beta$ value used for data generation. The crosses mark the mean and the error bars extend one standard deviation in both directions.

WLC, suggesting that the results might be just as good with data conforming to any other theoretical model. It means that the time resolution in our synthetic data, which is typical of AFM instrumentation, is high enough so that the error performed by making a piecewise linear approximation to the WLC curve is negligible.

GPL, which is identical to PL except that it receives noisy input, also gave good results, although the estimates were noticeably more biased than the WLC or PL ones throughout all design strategies, while the variance was slightly larger. The bias in the GPL $k_0$ estimate ranged from 2.3% to 3.5%, while for $x_\beta$ from 1.1% to 2.0%. which in the presence of noise can be considered as very small.

Excluding the unsatisfactory L approach, for each of the other three ML-based approaches (GPL, PL, WLC) the HLET experiment design strategy showed the smallest RMSE for both $k_0$ and $\alpha$ when compared to the other five design strategies (LSO, HLEN, ASEN, ASET, HSO) covered in our simulation. That suggests that a very efficient experiment design consists in equally splitting the experimental time across two velocities, one very high and one very low.

Since the Monte Carlo method implies the same WLC model also used to generate the synthetic data of the simulated time-constrained experiments, the Monte Carlo method performed quite well (Fig. 2.6 and 2.7). The best experiment design strategy for MC turned out to be HLEN instead of HLET, followed closely by the latter one, for both $k_0$ and $x_\beta$. Comparing the results of the best design strategy of each approach, the Monte Carlo method achieved an RMSE about 50% higher for $k_0$ and 128% higher for $x_\beta$ compared to our ML-based WLC and PL approaches, and about 40% higher for $k_0$ and and 74% higher for $x_\beta$ compared to GPL, that is, the piecewise linear method with noisy input. We'd like to stress once more that while the MC approach in our simulation had access to the exact force model (WLC) with the exact parameters (contour length, persistence length) used for data generation, for GPL noise was present in the data and no information whatsoever about the underlying model was available. The difference in statistical performance is only expected to increase in real experimental settings where the force response is generated by the (linker-)protein-cantilever system rather than a unique, known theoretical model.

We can draw four conclusions from this synthetic data experiment. Firstly, the general applicability of the linear force-displacement characteristic is disproved,

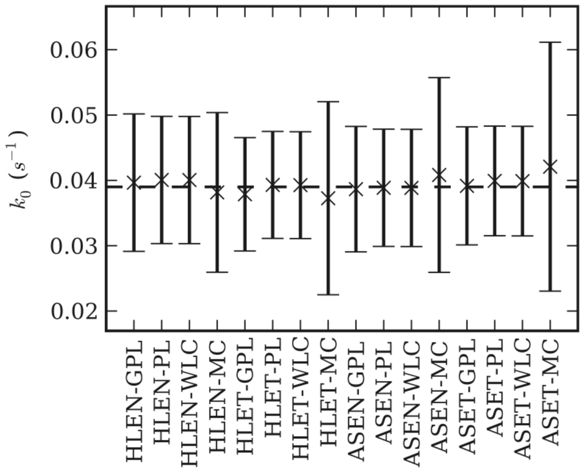


Figure 2.6: Estimation of $k_0$ using GPL, PL, WLC and MC. The GPL, PL and WLC data are the same as in Fig. 2.4 and are reproduced here for easy visual comparison. Data were generated using the strategies HLEN, HLET, ASEN and ASET, from left to right. The dashed horizontal line indicates the actual $k_0$ value used for data generation. The crosses mark the mean and the error bars extend one standard deviation in both directions.

whether applied through either the "standard method" or the maximum likelihood framework. Secondly, the piecewise linear approximation works well with WLC-conforming data. Thirdly, for a fixed amount of experimental time, using only the highest feasible velocity to get as many unfolding events as possible is not the best experiment design strategy; instead it is better to allot half of the experimental time to a much lower velocity. Fourthly, the ML approach, including the model-independent piecewise linear approximation, is better than the Monte Carlo method even when a unique force model exists and is known, as was the case in our simulation, and the advantage remains solid when noise is added only to the input of the piecewise linear approach.

### 2.4.3 GB1 kinetic parameters were correctly recovered from an SMFS experiment with polyprotein $(GB1)_8$

We further tested our ML approach with experimental data we obtained by pulling a multimeric construct consisting of eight GB1 modules [69, 49, 70].

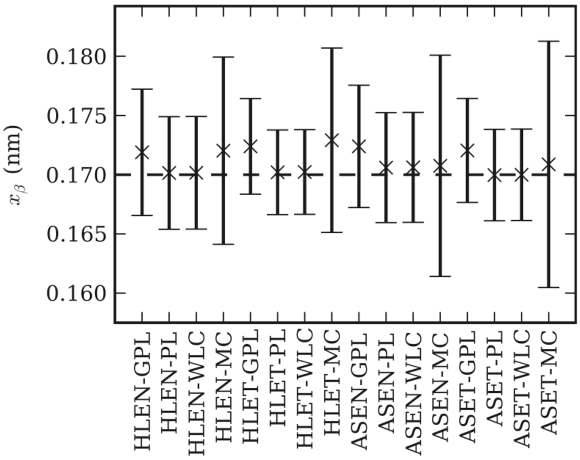


Figure 2.7: Estimation of $x_\beta$ using GPL, PL, WLC and and MC. The GPL, PL and WLC data are the same as in Fig. 2.5 and are reproduced here for easy visual comparison. Data were generated using the strategies HLEN, HLET, ASEN and ASET, from left to right. The dashed horizontal line indicates the actual $x_\beta$ value used for data generation. The crosses mark the mean and the error bars extend one standard deviation in both directions.

For the experiment, a drop of the $(GB1)_8$-containing solution (20 $\mu$L, ~0.1 g/L) was deposited on a flame cleaned glass coverslip for about 30 min. The velocity-clamp mechanical unfolding SMFS experiment was performed using Picoforce AFM with Nanoscope IIIa controller (Digital Instruments, Plainview, NY, USA) with a V-shaped silicon nitride cantilever (NP; Digital Instruments) whose spring constant was calibrated by the thermal noise method [8]. The buffer used was Tris/HCl (10 mM, pH 7.5).

We used the open source project Hooke [71] with locally made modifications to extract the relevant information from the AFM-recorded files, after which we applied an automated filtering step mostly based on the protocol specified in [43]. A total of 250 unfolding events passed the filtering stage, about a quarter of them at a retraction velocity of 125 nm/s, a quarter at 249 nm/s and half at 2180 nm/s.

Making the WLC assumption we obtained the point estimates $k_0 \simeq 0.0475\ s^{-1}$, $x_\beta \simeq 0.1661$ nm and a 70%-credible region of $(0.0415, 0.0653)\ s^{-1} \times (0.1583, 0.1687)$ nm, in very good agreement with the values reported in the literature of 0.039 $s^{-1}$ and 0.17 nm that had been previously extracted from a larger data set consisting of 1826 unfolding events via WLC-assuming Monte Carlo simulations [69]. Using the piecewise linear force-time approach we obtained instead the point estimates $k_0 \simeq 0.0622\ s^{-1}$, $x_\beta \simeq 0.1627$ nm and a 70%-credible region of (0.0466, 0.0777) $s^{-1}\times$ (0.1576, 0.1677) nm. Each of the two credible regions contains the point estimates obtained with both approaches. Figure 2.8 shows the piecewise linear approximation and the WLC fits for a $(GB1)_8$ curve with eight unfolding events from our experimental data set.

Using the WLC-based Monte Carlo method we obtained similar values of $k_0 \simeq 0.05\ s^{-1}$ and $x_\beta \simeq 0.16$ nm. Instead, using the "standard method" for the estimation task we obtained $k_0 \simeq 8.1440\ s^{-1}$, which is two orders of magnitude larger than expected, and also a smaller distance to the transition state $x_\beta \simeq 0.1302$ nm.

We conclude that when WLC describes well the experimental data, as is the case with GB1 [69], the estimates obtained via the piecewise-linear approximation and those extracted by making the WLC assumption are compatible within statistical uncertainty, while the $k_0$ estimate computed via the "standard method" can be off by a couple of orders of magnitude.

## 2.5 Summary

We have set forth an ML framework for the analysis of SMFS experiments with homomeric polyproteins, where the protein kinetic parameters of the monomeric module are of interest. For the restricted case when only the spontaneous dissociation rate is unknown, we found an unbiased estimator that is also more efficient than the plain ML estimator.

To account for the heterogeneity of force behaviors seen in SMFS experiments we propose a piecewise linear approximation to the forces recorded by the AFM during stretching and showed via extensive simulation that the approach is able to correctly recover both kinetic parameters. That obviates the need to assume a pre-determined force increase model which, as our tests show, can result in large estimation errors if the wrong one is chosen, thus disproving the widespread practice of assuming the linear force-displacement characteristic.

Our framework does however allow one to specify a pre-determined force model, and we validated this use case in the context of the WLC model with both synthetic and experimental data. The latter was obtained by pulling a polyprotein made up

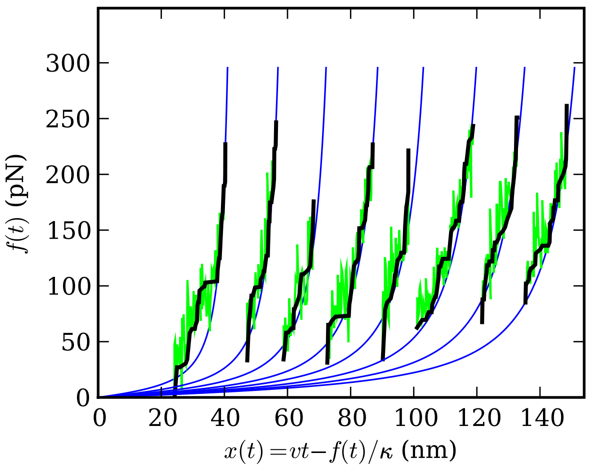


Figure 2.8: (Color online) Experimental curve obtained by pulling the $(GB1)_8$ construct at v=2180 nm/s. The light blue (light gray) lines meeting in the origin represent WLC fits. The points interpolated by the thin red (medium gray) lines represent AFM force readings. The thick black lines are piecewise linear approximations of the longest increasing subsequences of force values. The detachment peak and nonspecific interactions at the start are not shown.

of identical tandem repeats of a protein domain that had been previously characterized via WLC-assuming Monte Carlo simulations. By imposing the same WLC model, from our data set we recovered almost the same kinetic parameters under the ML framework and also by applying the Monte Carlo method. Since WLC describes well the behavior of $(GB1)_8$ [69], this confirms the correctness of both our Monte Carlo implementation and the ML framework we propose. In order to compare the statistical properties of our approach against the Monte Carlo method we turned to synthetic data experiments which show that the WLC-assuming ML estimators are better than the WLC-assuming Monte Carlo estimators in terms of RMSE.

Without imposing a theoretical force model instead we obtained from the experimental data a slightly larger spontaneous unfolding rate, but still within the 70%-credible region of the WLC based estimator. The compatibility, but with some difference, between the two sets of estimates can be at least partly attributed to the ability of the WLC model (strictly speaking, the approximation formula of Eq. (2.11)) to describe well, but not perfectly, the behavior of the studied protein. This constitutes a new approach for testing the applicability of a theoretical force model to any protein when geometric curve fits by themselves do not provide a definitive answer.

To do a comparison in terms of statistical performance with the Monte Carlo method we turned again to synthetic data experiments which proved that our ML approach, even when used without any information about the underlying force increase model, and with noise added to the data, performs better than the Monte Carlo method configured with the correct force increase model that was used for data generation, with the correct parameters so that no fitting is necessary. This is a clear proof of the superiority of the ML estimation: it requires less information as input while at the same time leading to better estimators, even under clearly disadvantageous conditions.

Finally we approached the problem of long experimental times in two ways. Firstly, by using ML estimation one is likely to need fewer rupture events for the estimation task compared to traditional approaches that do not benefit from the *likelihood principle* that guarantees that no information is lost. Secondly, our synthetic data experiments suggest that rather than using the highest feasible pulling velocity to get as many unfolding events as possible, it's more efficient to also use one much lower retraction velocity, and to allot to the two velocities equal shares of the experimental time.

## 2.6 Conclusions

We conclude with a short review of the advantages our proposed method brings over existing ones:

1. It is particularly well suited for the analysis of experiments where a master curve cannot be easily identified, such as those involving ligand-receptor complexes [47] (see Subsection 2.8.1).

2. It is the only method for which an unbiased estimator of the dissociation rate has been provided when the distance to the transition state is known, thus making it particularly attractive for the analysis of protein unfolding under the effect of certain chemical denaturants [63], protein mutations [61, 62, 44] and different solvents of equal molecule size [60] (see Eq. (2.8)).

3. In all the tested settings it leads to better estimators of the kinetic parameters in terms of RMSE when compared to existing methods, even under disadvantageous conditions (see Subsection 2.4.2).

4. It comes with a clear recommendation about how to design experiments based on the well accepted statistical criterion of reducing the RMSE of the obtained estimators (see Subsection 2.4.2).

5. As a simple numerical maximization of a univariate function (see Eq. (2.7)), point estimation is very fast in practice, in our experience orders of magnitude faster than the more established Monte Carlo method which instead requires extensive data generation on a two-dimensional grid of parameters.

Because of the generality of the last three mentioned advantages we recommend our approach as the method of choice in the analysis of all velocity-clamp experiments with polymers made up of one or more identical domains.

## 2.7 Acknowledgments

We thank Prof. Hongbin Li, University of British Columbia, Vancouver, Canada for kindly providing the $(GB1)_8$ plasmid, Dr. Isabella Tessari and Prof. Luigi Bubacco, University of Padova, Italy for providing the $(GB1)_8$ construct and Dr. Aldo Rampioni, University of Bologna, Italy for helpful comments on the manuscript.

## 2.8 Appendix

### 2.8.1 Application to some force-time functions

We show next how the likelihood function can be computed with two force models and the piecewise linear approximation.

**Linear force-displacement characteristic**

Under the linear force vs. displacement characteristic $f(t) = \kappa v t$ the likelihood function in Eq. (2.5) becomes

$$L(\vec{y}; k_0, \alpha) = \exp\left[-\frac{k_0}{\alpha}\sum_{i=1}^{n}\frac{m_i}{\kappa_i v_i}(e^{\alpha y_i} - e^{\alpha y_{si}})\right] \times k_0^n \prod_{i=1}^{n}\frac{m_i}{\kappa_i v_i}e^{\alpha y_i} \ . \tag{2.10}$$

**Worm-like chain**

The worm-like chain describes the force dependence on the distance over contour length ratio, and we adopt the well known interpolation formula with less than 10% error proposed in [59]:

$$f(t) = \frac{k_B T}{4p}\left[(1 - x(t)/L_c)^{-2} + 4x(t)/L_c - 1\right] \tag{2.11}$$

where $x(t) = vt - f(t)/\kappa$ is the distance at time $t$ [72]. By substitution we get the cubic equation

$$\begin{aligned}
&-4az^3 + [(12a/b + 4)y + 9a]z^2 \\
&-[y^2(12a/b + 8)/b + y(18a/b + 8) + 6a]z \\
&+y^3(4a/b + 4)/b^2 + y^2(9a/b + 8)/b + y(6a/b + 4) = 0
\end{aligned} \tag{2.12}$$

where $a = k_BT/p$, $b = \kappa L_c$ and $z = (f^{-1})(y)\, v/L_c$.

It's possible to show that there is exactly one root in the interval of interest $(y/b, 1 + y/b)$, so $(f^{-1})(y)$ can be obtained without ambiguity. Many ways to solve the cubic polynomial equation exist, including closed form solutions [73]. Then $(f^{-1})'(y)$ can be computed by implicit differentiation, thus making possible the numerical computation of the likelihood function using Eq. (2.5).

### Piecewise linear force

The problem of finding the longest increasing subsequence of a sequence is classical in computer science, and for this purpose we use an $\mathcal{O}(n\, log\, n)$ algorithm as described e.g. in [64]. For completeness we reproduce here our Python adaptation of the algorithm:

```
def extract_path(a, s):
    max_score = max(s)
    path = [0] * max_score
    next_score = max_score
    for i in range(len(s)):
        if s[len(s) - 1 - i] == next_score:
            assert next_score == max_score or \
                a[len(s) - 1 - i] < a[path[next_score]], (
                next_score, max_score, i, a, s, path)
            next_score -= 1
            path[next_score] = len(s) - 1 - i
    assert next_score == 0, next_score
    assert mostly_ok(a, path), path
```

```
    return tuple(path)

def longest_increasing_subsequence(a):
    m = []
    s = []
    for (ii, t) in enumerate(a):
        assert (not m) == (not ii)
        bs = bisect.bisect_left(m, t)
        j = bs - 1
        if 0 <= bs < len(m) and m[bs] == t:
            j = bs - 1
            assert j == -1 or m[j] < t
        else:
            assert -1 <= j
        if j == len(m) - 1:
            m.append(t)
        else:
            assert t <= m[j + 1]
            m[j + 1] = t
        s.append(j + 2)
    path = extract_path(a, s)
    return path
```

Now let's consider one unfolding event with the longest increasing subsequence of forces $y_1, \ldots, y_p$ at increasing times $t_1, \ldots, t_p$, with $y_p$ being the rupture force. The piecewise linear force-time function is then assembled as

$$f(t) = \begin{cases} y_j + (t - t_j)(y_{j+1} - y_j)/(t_{j+1} - t_j), \\ \qquad \text{if } t_j \le t < t_{j+1}, 1 \le j < p-1 \\ y_{p-1} + (t - t_{p-1})(y_p - y_{p-1})/(t_p - t_{p-1}), \\ \qquad \text{if } t \ge t_{p-1} \end{cases}.$$

Assuming that unfolding event $i$ has longest increasing subsequence of force values $(t_{i1}, y_{i1}), (t_{i2}, y_{i2}) \ldots (t_{ip_i}, y_{ip_i})$ with the connection to the notation throughout the rest of the paper being that $(t_{i1}, y_{i1}) = (t_{si}, y_{si})$ and $(t_{ip_i}, y_{ip_i}) = (t_i, y_i)$, the likelihood function can be written as:

$$L(\vec{y};k_0,\alpha) = \quad \exp\left[-\frac{k_0}{\alpha}\sum_{i=1}^{n} m_i \sum_{j=1}^{p_i-1}\frac{t_{ij+1}-t_{ij}}{y_{ij+1}-y_{ij}}(e^{\alpha y_{ij+1}}-e^{\alpha y_{ij}})\right] \\ \times k_0^n \prod_{i=1}^{n} m_i \frac{t_{p_i}-t_{p_i-1}}{y_{p_i}-y_{p_i-1}} e^{\alpha y_{ip_i}} \ . \tag{2.13}$$

### 2.8.2 Expectation of $\widehat{k_0}(\alpha)$ for $\alpha$ fixed

**Proposition 2** *For $\alpha > 0$ fixed and $n \geq 2$, if $\int_{y_{si}}^{\infty} e^{\alpha z}(f_i^{-1})'(z)\,dz = \infty$ for all $1 \leq i \leq n$, then $E[\widehat{k_0}(\alpha)] = nk_0/(n-1)$.*

To prove the above result we start from the definition

$$E[\widehat{k_0}(\alpha)] = \int_{y_1 \geq y_{s1} \ldots y_n \geq y_{sn}} \widehat{k_0}(\alpha) L(\vec{y};k_0,\alpha)\,dy_1 \ldots dy_n \ .$$

Using Eq. (2.5), (2.6), and making the changes of variables

$$x_i = k_0 m_i \int_{y_{si}}^{y_i} e^{\alpha z}(f_i^{-1})'(z)\,dz,\ 1 \leq i \leq n$$

we obtain

$$E[\widehat{k_0}(\alpha)] = nk_0 \int_{x_1 \geq 0 \ldots x_n \geq 0} \frac{exp\left[-\sum_{i=1}^{n} x_i\right]}{\sum_{i=1}^{n} x_i} dx_1 \ldots dx_n \ ,$$

where the multiple integral is equal to $1/(n-1)$ according to Lemma 7.

**Notation 3** *We'll denote by $E_1$ the well-known exponential integral with its equivalent forms:*

$$E_1(u) = \int_u^{\infty} \frac{e^{-x}}{x}\,dx = \int_1^{\infty} \frac{e^{-ux}}{x}\,dx$$

**Lemma 4** *For any real $a > 0$ and natural number $n >= 0$ the following holds:*

$$\int_0^{\infty} (a+x)^n E_1(a+x)dx = -\frac{a^{n+1}E_1(a)}{n+1} + \frac{e^{-a}}{n+1}\sum_{k=0}^{n}\frac{n!}{k!}a^k \ .$$

**Proof.** From

$$E_1'(u) = -\frac{e^{-u}}{u}$$

we can compute

$$(a+x)^n E_1(a+x) = \left(\frac{(a+x)^{n+1}E_1(a+x)}{n+1}\right)' + \frac{(a+x)^n e^{-(a+x)}}{n+1}. \tag{2.14}$$

Note that

$$E_1(u) < \int_1^\infty e^{-ux}\,dx = \frac{e^{-u}}{u}$$

from which we deduce that

$$\lim_{u\to\infty} u^m E_1(u) = 0 \qquad \text{for all real } m. \tag{2.15}$$

Integrating (2.14) and using (2.15):

$$\int_0^\infty (a+x)^n E_1(a+x)dx =$$
$$-\frac{a^{n+1}E_1(a)}{n+1} + \frac{e^{-a}}{n+1}\sum_{k=0}^{n}\binom{n}{n-k}a^k\int_0^\infty x^{n-k}e^{-x}\,dx$$

and knowing that $\Gamma(k+1) = k!$ for $k \geq 0$ we get the desired result. ■

**Lemma 5** *For any integers $i, n$ such that $0 \leq i \leq n-2$ the following holds:*

$$\sum_{k=0}^{i}\frac{(-1)^k k!}{(n-1)\dots(n-1-k)} = \frac{1}{n} + \frac{(-1)^i(i+1)!}{n\dots(n-i-1)} \tag{2.16}$$

**Proof.** By induction on $i$. ■

**Lemma 6** *For any positive real $a$ and natural $n \geq 1$:*

$$I_n(a) = \int_{x_1\geq 0\dots x_n\geq 0}\frac{e^{-\sum_{i=1}^n x_i}}{a+\sum_{i=1}^n x_i}dx_1\dots dx_n =$$
$$\frac{(-1)^{n-1}e^a a^{n-1}E_1(a)}{(n-1)!} + \sum_{k=0}^{n-2}\frac{(-a)^k}{(n-1)\dots(n-1-k)}$$

**Proof.**

$$I_1(a) = e^a \int_u^\infty \frac{e^{-x}}{x}\,dx = e^a E_1(a)$$

$$\begin{aligned} I_2(a) &= \int_0^\infty e^{-x_2} I_1(a+x_2)dx_2 \\ &= \int_0^\infty e^{-x_2} e^{a+x_2} E_1(a+x_2)dx_2 \qquad \text{(by Lemma 4)} \\ &= e^a(-aE_1(a)+e^{-a}) = -ae^aE_1(a)+1 \end{aligned} \tag{2.17}$$

$$\begin{aligned} I_3(a) &= \int_0^\infty e^{-x_3} I_2(a+x_3)dx_3 \\ &= \int_0^\infty e^{-x_3}(1-(a+x_3)e^{a+x_3}E_1(a+x_3))dx_3 \\ &= 1 - e^a\left(-\frac{a^2E_1(a)}{2} + \frac{e^{-a}}{2}(1+a)\right) \\ &= \frac{e^a a^2 E_1(a)}{2} + \frac{1}{2} - \frac{a}{2} \end{aligned}$$

For $n \geq 3$ we proceeed by induction on $n$:

$$I_{n+1}(a) = \int_0^\infty e^{-x_{n+1}} I_n(a+x_{n+1})\,dx_{n+1} = A + B \tag{2.18}$$

where:

$$\begin{aligned} A &= \int_0^\infty e^{-x_{n+1}} \frac{(-1)^{n-1}e^{a+x_{n+1}}(a+x_{n+1})^{n-1}E_1(a+x_{n+1})}{(n-1)!}\,dx_{n+1} (by\ (4)) \\ &= \frac{(-1)^n e^a a^n E_1(a)}{n!} + \frac{(-1)^{n-1}}{n}\sum_{i=0}^{n-1}\frac{a^i}{i!} \end{aligned}$$

and

$$\begin{aligned}
B &= \sum_{k=0}^{n-2} \frac{(-1)^k}{(n-1)\dots(n-1-k)} \int_0^\infty e^{-x_{n+1}} (a + x_{n+1})^k \, dx_{n+1} \\
&= \sum_{k=0}^{n-2} \frac{(-1)^k k!}{(n-1)\dots(n-1-k)} \sum_{i=0}^{k} \frac{a^i}{i!} = \sum_{i=0}^{n-2} \frac{a^i}{i!} \sum_{k=i}^{n-2} \frac{(-1)^k k!}{(n-1)\dots(n-1-k)} \\
&= \frac{1}{n} + \frac{(-1)^{n-2}(n-1)!}{n!} + \sum_{i=1}^{n-2} \frac{a^i}{i!} \sum_{k=i}^{n-2} \frac{(-1)^k k!}{(n-1)\dots(n-1-k)} \\
&= \frac{1}{n} + \frac{(-1)^{n-2}(n-1)!}{n!} + \sum_{i=1}^{n-2} \frac{a^i}{i!} \left( \frac{1}{n} + \frac{(-1)^{n-2}(n-1)!}{n!} - \frac{1}{n} - \frac{(-1)^{i-1} i!}{n\dots(n-i)} \right) \\
&= \frac{1}{n} + \frac{(-1)^{n-2}(n-1)!}{n!} + \frac{(-1)^{n-2}}{n} \sum_{i=1}^{n-2} \frac{a^i}{i!} + \sum_{i=1}^{n-2} \frac{(-a)^i}{n\dots(n-i)} \qquad (2.19)
\end{aligned}$$

It follows that

$$\begin{aligned}
I_{n+1}(a) &= A + B \\
&= \frac{(-1)^n e^a a^n E_1(a)}{n!} + \frac{(-1)^{n-1}}{n} + \frac{(-a)^{n-1}}{n!} + \frac{1}{n} + \frac{(-1)^{n-2}(n-1)!}{n!} + \sum_{i=1}^{n-2} \frac{(-a)^i}{n\dots(n-i)} \\
&= \frac{(-1)^n e^a a^n E_1(a)}{n!} + \sum_{i=0}^{n-1} \frac{(-a)^i}{n\dots(n-i)} \qquad (2.20)
\end{aligned}$$

■

**Lemma 7** *For any integer $n >= 2$, the following holds:*

$$\int_{x_1 \geq 0 \dots x_n \geq 0} \frac{e^{-\sum_{i=1}^n x_i}}{\sum_{i=1}^n x_i} \, dx_1 \dots dx_n = \frac{1}{n-1}$$

**Proof.** According to Lemma 6 the left hand side integral is equal to:

$$I = \int_0^\infty e^{-x_1} \left( \frac{(-1)^{n-2} e^{x_1} {x_1}^{n-2} E_1(x_1)}{(n-2)!} + \sum_{k=0}^{n-3} \frac{(-x_1)^k}{(n-2)\dots(n-2-k)} \right) dx_1 \qquad (2.21)$$

Let's compute the first part.

$$\int_0^{\infty} x^{n-2} E_1(x)\, dx = \int_1^{\infty} \left( \int_0^{\infty} \frac{x^{n-2} e^{-ux}}{u}\, dx \right) du = \tag{2.22}$$
$$= \int_1^{\infty} \left( \int_0^{\infty} \frac{y^{n-2} e^{-y}}{u^n}\, dy \right) du = \int_1^{\infty} \frac{(n-2)!}{u^n}\, du = \frac{(n-2)!}{n-1}$$

Meeting together (2.21) with (2.22):

$$I = \frac{(-1)^{n-2}}{n-1} + \sum_{k=0}^{n-3} \frac{(-1)^k k!}{(n-2)\dots(n-2-k)}\,. \tag{2.23}$$

Putting $i = n-3$ in (5), together with (2.23) we have:

$$I = \frac{(-1)^{n-2}}{n-1} + \frac{1}{n-1} + \frac{(-1)^{n-3}(n-2)!}{(n-1)!} = \frac{1}{n-1}\,. \tag{2.24}$$

■

### 2.8.3 Monte Carlo simulations

**Standard approaches**

There are two Monte Carlo methods that can be used for the estimation of kinetic parameters. One is used when unfolding data is available at only one retraction speed, and involves generation of unfolding events on a grid of $k_0$, $x_b$ values, followed by fitting the experimentally obtained histogram of unfolding forces against the Monte Carlo simulated one. The other one is used when multiple retraction velocities are available, and in this case only the most probable unfolding force is fitted. The combination of parameters that lead to the best fit is then taken as the solution. A bootstrap Monte Carlo procedure can be used to approximately normalize the experimental distribution of unfolding forces and to estimate confidence intervals for the two parameters. Alternatively, the mode of the experimental distribution can be used directly when confidence intervals are not required.

Usually a Monte Carlo setup which is similar to the average experimental curve is employed, matching the average contour length and the average number of unfolded modules as in the real experiment. In principle it is also possible to have a one to one mapping between every single experimental peak to a simulated peak, matching them in terms of contour length and the starting force of the pulling process (after the unfolding of the previous peak), but that is often too computationally expensive and it is not usually done.

The Monte Carlo simulation is performed as described in [23]. Briefly, the time range leading to each unfolding event is broken down into a set of small time steps during which the force is assumed to remain constant. The probability to observe an unfolding event during time $\Delta t$ is:

$$dP_u = m\alpha(F)\Delta t \tag{2.25}$$

where $m$ is the number of modules that are still folded and $\alpha(F)$ is the force dependent unfolding rate:

$$\alpha(F) = k_0 exp\left[\frac{Fx_b}{k_bT}\right] \tag{2.26}$$

Using a random number generator the folded-to-unfolded transition is executed with probability $dP_u$ by the increasing the polymer contour length by the force-hidden length of one module and decreasing $m$ by one. The time step needs to be kept small enough so that $dP_u$ always stays below one, but generally the smaller it is, the better. After each time step the updated force is calculated according to the WLC formula of Eq. 2.11.

Herein we used the Monte Carlo procedure with multiple velocities and we performed the simulation using one setup chosen to match perfectly the conditions used for data generation in Subsection 2.4.2. That was possible because all curves had the same maximum contour length, and peaks have predefined contour lengths in the synthetic data. The correct kinetic parameters were included in the center of the grid, with exponentially increasing steps for $k_0$ and a fixed step size of 0.01 nm for $x_b$.

### Our approach to synthetic data generation

Little is known about the tradoffs of chosing a smaller or larger time step in the standard Monte Carlo simulation. In order to make the simulation more accurate, we used numerical integration instead of assuming that the force is constant over small time steps. To also speed up the computation, we split the force range from zero to a maximum, never seen in real experiments, unfolding force into

8196 equal intervals. For each of the intervals we computed via numerical integration the probability to observe an unfolding event during the time period when the force would increase over that particular force range. As compared to the standard Monte Carlo approach, we move from time quantization to force quantization, which guarantees us a pre-set unfolding force accuracy. That is important because via numerical integration we avoid the drawback of approximating the WLC formula by constant steps all the way until unfolding; instead we use the WLC formula without any changes, and do not assume that the force stays constant over any amount of time, except for the actual unfolding step. Besides that the unfolding force is what we aim to extract, not the time of unfolding. Thus it is better to control the force accuracy rather than the time accuracy.

For each force range, from $f_1$ to $f_2$, we compute:

$$I(f1, f2) = \int_{f1}^{f2} e^{\alpha z}(f^{-1})'(z)dz \ . \tag{2.27}$$

We then compute a cumulative array of the $I$ values that we'll name $C$. Given this array, every time we need to generate an unfolding event, we compute:

$$y = -log(1 - min(UniformRandom(0, 1)\ for\ i\ in\ range(m)))/k_0 \tag{2.28}$$

and then we extract the unfolding force by binary-searching $y$ into the (sorted) array $C$.

# Chapter 3

# Open source platform for the execution and analysis of mechanical refolding experiments

Adapted with permission from Aioanei D, Brucale M, Samorì B. “Open source platform for the execution and analysis of mechanical refolding experiments”, Bioinformatics. 2011 Feb 1;27(3):423-5. Copyright (2011) by Oxford University Press.

## 3.1 Abstract

### 3.1.1 Motivation:

Single-molecule force spectroscoy (SMFS) has facilitated the experimental investigation of biomolecular force-coupled kinetics, from which the kinetics at zero force can be extrapolated via explicit theoretical models. The atomic force microscope (AFM) in particular is routinely used to study protein unfolding kinetics, but only rarely protein folding kinetics. The discrepancy arises because mechanical protein refolding studies are more technically challenging.

### 3.1.2 Results:

We developed software that can drive and analyse mechanical refolding experiments when used with the commercial AFM setup “Picoforce AFM”, Bruker (previously Digital Instruments). We expect the software to be easily adaptable to other AFM setups. We also developed an improved method for the statistical characterisation of protein folding kinetics, and implemented it into an AFM-independent software module.

### 3.1.3 Availability:

Software and documentation are available at
http://code.google.com/p/refolding under Apache License 2.0.

## 3.2 Introduction

Biochemical reactions commonly proceed via large conformational changes, resulting in a well defined mechanical reaction coordinate on which they can be monitored. Since force is a determinant factor in the rate of such reactions, single-molecule force spectroscopy (SMFS) emerged as an invaluable tool in their investigation under mechanical tension [3, 74]. Thanks to its remarkable ability to stretch and monitor one molecule at a time, SMFS seeks to achieve the long-standing goal of mapping the energy landscape of biomolecules (e.g. proteins, RNA) on a well defined reaction coordinate, even for proteins which show irreversible thermal or chemical unfolding [75].

In particular, protein folding kinetics can be studied at the single molecule level using the atomic force microscope (AFM). To this end protein modules can be first unfolded and subsequently allowed to refold while subjected to a “force-clamp” [76, 77, 78], or they can be directly observed to refold at fixed extension via “lock-in force spectroscopy” [79]. Such techniques depend on recent technological advances implemented in custom-built AFMs with very limited availability. Alternatively, the AFM can be used in the more traditional “velocity-clamp” mode to drive protein modules to fold under mechanical tension via the “double-pulse protocol” (see Suplementary Data). Shortly, the distance between the base of the cantilever and the surface, rather than the stretching force, is maintained constant for some amount of time allowing previously unfolded protein modules to refold [49, 80, 41].

Despite the widespread availability of AFM instrumentation supporting the velocity-clamp mode of operation, single molecule folding kinetics studies remain rather scarce in the scientific literature, likely due to the unavailability of mandatory software technology. We fill this gap by making such software freely available. To validate our software we studied the folding kinetics of protein GB1 [69] and obtained a kinetics characterisation similar to the previously published one [49].

## 3.3 Approach

Our software contains three main components:

1. An automated procedure for driving refolding experiments through Nanoscope v6 software, in conjunction with Picoforce AFM and Nanoscope IIIa controller, Bruker. For each execution of the double-pulse protocol, our software instructs the Nanoscope software to execute a "Nanoscope script" and capture a "Nanoscope strip chart". Importantly, the actual bending of the cantilever is detected from the strip chart file and the starting position is adjusted accordingly for the next double-pulse so as to counterbalance accumulating drift [81]. Section 1 in Supplementary Data contains more information on the implemented double-pulse protocol.

2. Offline tools for automated peak identification, force measurements, Worm-like chain [59] fits and data filtering. Tools are also included for manually improving the results of some of the automated tasks such as zero-force baseline and contact point identification.

3. A standalone, AFM-independent offline procedure for the statistical characterisation of protein folding kinetics from mechanical refolding experiments with homomeric polyproteins, based on the analytical model of Section 3.4.1.

## 3.4 Methods

### 3.4.1 Maximum likelihood estimation of folding kinetic parameters

We adopt Bell's approximation [82, 37] to Kramers' reaction-rate theory [83], which describes the force-dependent folding rate as $k_f(F) = k_f^0 exp[-F\Delta x_f/(K_bT)]$, where $K_b$ is the Boltzmann constant, $T$ is the temperature in Kelvin, $k_f^0$ is the spontaneous folding rate and $\Delta x_f$ is the folding distance. We aim to extract the last two mentioned parameters from refolding experiments.

Traditionally data was collected in a few fixed configurations, where a configuration is defined by the amount of time allowed for refolding and the inferred force at the start of the waiting period. It was therefore assumed that the same configuration can be reproduced exactly multiple times, ignoring any variation between double-pulse protocol executions. It was also assumed that the residual force remains constant during the waiting time [49, 41], ignoring the fact that it increases after each folding event. Furthermore, it was assumed that the total number of modules that could refold is limited by the extension during the waiting period [49], breaking the assumption of the ideal spring cantilever. The mean and standard deviation of the refolding ratio would then be computed for each configuration and then fitted to an exponential formula based on Bell's equation or to Monte-Carlo simulations based on it, ignoring the fact that such summary statistics are not *sufficient* [84], i.e., they do not capture all possible information about the parameters.

We overcome all the above limitations by introducing a Maximum Likelihood estimation procedure. Shortly, the stretching force is computed by solving the WLC cubic equation [17], and the likelihood function is computed as the product of the probability to observe the actual number of folding events for each double-pulse protocol execution. The maximum likelihood is then located over a grid of $(k_f^0, \Delta x_f)$ values, and estimation errors are computed by case resampling (see Section 2 in Supplementary Data).

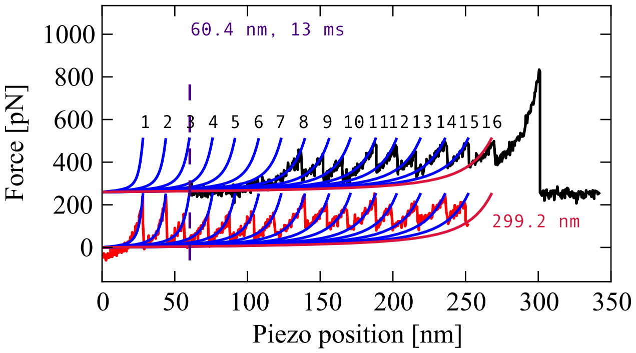


Figure 3.1: (Colour online) A force-extension trace according to the double-pulse protocol with the homomeric polyprotein $(GB1)_{16}$. The lower curve represents the protein fetching phase, during which the polyprotein attached nonspecifically to the cantilever and then a total of 15 modules have been subsequently unfolded. The higher curve is shifted by 260 pN just for display purposes, and it represents the phase where the same molecule is pulled for the second time. Note that only 15 out of the 16 modules could have refolded, since one module was not unfolded during the fetching phase. The vertical dashed line represents the piezo position during the waiting time-lapse relative to the resting position of the cantilever tip, and its numerical value is shown together with the length of the time-lapse at the top of the figure. Each WLC fit is shown redundantly shifted higher for display purposes. The contour length at the start of the waiting time-lapse (before any refolding) is indicated in the bottom-right position.

### 3.4.2 Folding kinetics of protein GB1

We estimated the kinetic parameters of protein GB1 in buffer Tris/HCl (10 mM, pH 7.5) by performing mechanical refolding experiments with homomeric polyproteins $(GB1)_8$ and $(GB1)_{16}$ (see Section 3 in Supplementary Data for experimental data statistics). A sample trace can be seen in Figure 3.1.

We obtained the kinetic parameters $\Delta x_f = 2.53 \pm 0.12$ nm and $k_f^0 = 500 \pm 85\,s^{-1}$, errors representing one standard deviation. Our kinetics characterisation is roughly compatible with previously published values of $\Delta x_f = 2.1$ nm and $k_f^0 = 720 \pm 120\,s^{-1}$ [49].

## 3.5 Discussion

Mechanical refolding experiments can be performed with typical commercial velocity-clamp AFM instrumentation, and we provide an out-of-the-box software solution for performing and analysing such experiments in conjunction with Picoforce AFM, Bruker. We expect our software to be easily adaptable to other AFM setups. In fact since the analytical model of Section 3.4.1 is not specific to a particular AFM, its implementation can already be used with refolding data obtained with any other AFM. Furthermore, we developed all the software in the Java and Python programming languages to ensure its portability across all major operating systems.

## Acknowledgement

We thank Prof. Hongbin Li, University of British Columbia, Vancouver, Canada for kindly providing the $(GB1)_8$ plasmid, Dr. Isabella Tessari and Prof. Luigi Bubacco, University of Padova, Italy for providing the $(GB1)_{16}$ construct.

**Funding:** This work was supported by Ministero dellUniversitá e della Ricerca Fondo per gli Investimenti della Ricerca di Base (MIUR FIRB) RBNE03PX83/001; MIUR FIRB Progetto NG-lab (G.U. 29/07/05 n.175); Progetti di Ricerca di Interesse Nazionale (PRIN) 2008 (Prot. 2008KZ3E5).

# Chapter 4

# Single-molecule-level evidence for the osmophobic effect

Adapted with permission from Aioanei D, Lv S, Tessari I, Rampioni A, Bubacco L, Li H, Samorì B, Brucale M. “Single-molecule-level evidence for the osmophobic effect”, Angew Chem Int Ed Engl. 2011 May 2;50(19):4394-7. Copyright (2011) WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

## 4.1 Abstract

Protecting osmolytes play a crucial role in preventing protein denaturation in harsh environmental conditions of living organisms. Experimental evidence is provided for a mechanism of protein-fold stabilisation by these molecules that is in accord with the hypothesis of a backbone-based osmophobic effect.

## 4.2 Introduction

Organic osmolytes are low-molecular-weight osmotically active compounds, which are ubiquitous in living systems and are able to modulate protein stability. Among them, those that act as folding agonists, enhancing the stability of the native structure of proteins, such as trimethylamine N-oxide, betaine, sarcosine, proline, trehalose, sucrose, glycerol, sorbitol, and dimethyl sulphoxide (DMSO), are collectively called protecting osmolytes or “chemical chaperones” [85, 86, 87]. One

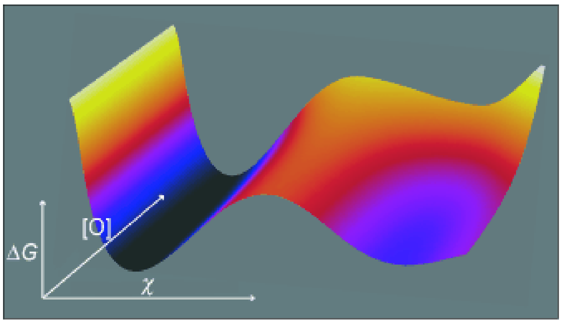


Figure 4.1: $\Delta G$=free energy, [O]=osmolyte concentration, $\chi$=unfolding reaction coordinate.

rather puzzling feature of these compounds is that they are able to affect the folding of very diverse proteins in similar ways, suggesting that they might act according to a general mechanism [88, 89, 90, 91, 92, 13], in contrast to the more specific mechanisms employed by chaperone proteins [93, 94]. In fact, the most widely accepted theory to rationalize their mode of action proposes that the addition of a protecting osmolyte to water as a co-solvent results in diminished solvent quality for the protein backbone, thus making intra-peptide backbonebackbone hydrogen bonds energetically more favorable than those between the backbone and the solvent [85].

This effect, known as the osmophobic effect, implies that protecting osmolytes have a universal, indirect mode of action, which does not entail the presence of any specific binding sites for the osmolyte on the protein in any of its states, including the folding/unfolding transition state. On the other hand, evidence of such a direct participation have been recently provided by experimental studies for specific proteinosmolyte combinations [76, 15, 95, 96].

Herein, we provide experimental evidence, at the single-molecule level, that the osmolyte DMSO protects the native state of a globular protein against mechanical unfolding without any active complexation of the osmolyte molecules into its unfolding transition state. Apart from slowing down the spontaneous unfolding rate of the protein, we show that the osmolyte also simultaneously accelerates its folding rate. The kinetic description of the observed stabilization mechanism strongly supports a backbone-based theory of the osmophobic effect.

## 4.3 Results and Discussion

We employed atomic force microscopy (AFM)-based single-molecule force spectroscopy (SMFS) [72] to characterize the effect of DMSO on the folding and unfolding kinetics of a globular protein domain, namely the B1 immunoglobulin binding domain of protein G from streptococcus (herein referred to as GB1), which behaves as a two-state folding protein on AFM experimental timescales [49, 63, 97, 98].

SMFS mechanical unfolding and refolding [18] experiments were performed on polyprotein constructs made up of either eight or sixteen tandem repeats of the GB1 domain [49]. We used buffered solutions with five different concentrations of DMSO ranging from 0% to 50% v/v (see Section 4.4.1 for the detailed SMFS experimental methods, and Section 4.4.2 for preparation of protein constructs).

The unfolding experiments were performed using three independent and complementary SMFS modes of operation (Fig. 4.2), which consistently led to the same result: in the presence of DMSO, the spontaneous unfolding rate of GB1 at zero applied force is negatively correlated with DMSO concentration, meaning that DMSO kinetically protects the folded state against unfolding. This protection manifests itself as an increase of the average mechanical unfolding forces at all loading rates in the velocity clamp SMFS mode (Fig. 4.3a), and a decrease of the force-dependent unfolding rates in force ramp and force clamp experiments (Section 4.4.4, Fig. 4.6). The refolding experiments were instead performed using a variable time lapse, double pulse procedure and showed that DMSO increases the spontaneous folding rate of GB1 (please refer to Section 4.4.5 for the experimental details on this procedure).

The data (Fig. 4.3a) make it possible to map the mechanical unfolding energy landscape of GB1 at each investigated DMSO concentration, extracting (as detailed elsewhere [17]) the two fundamental kinetic parameters of the widely employed Kramer two-state model: the distance between the native state and the transition state along the reaction coordinate $\Delta x_u$ (Fig. 4.4) and the spontaneous unfolding rate $k_u^0$ in absence of an applied force (see Section 4.4.3 for the data analysis methods). Importantly, as the reaction coordinate coincides with the direction of elongation of the protein in SMFS mechanical unfolding experiments, the extracted value of $\Delta x_u$ is highly sensitive to changes in the geometry of the transition state [76, 15]. The other parameter of interest, namely $k_u^0$, is exponen-

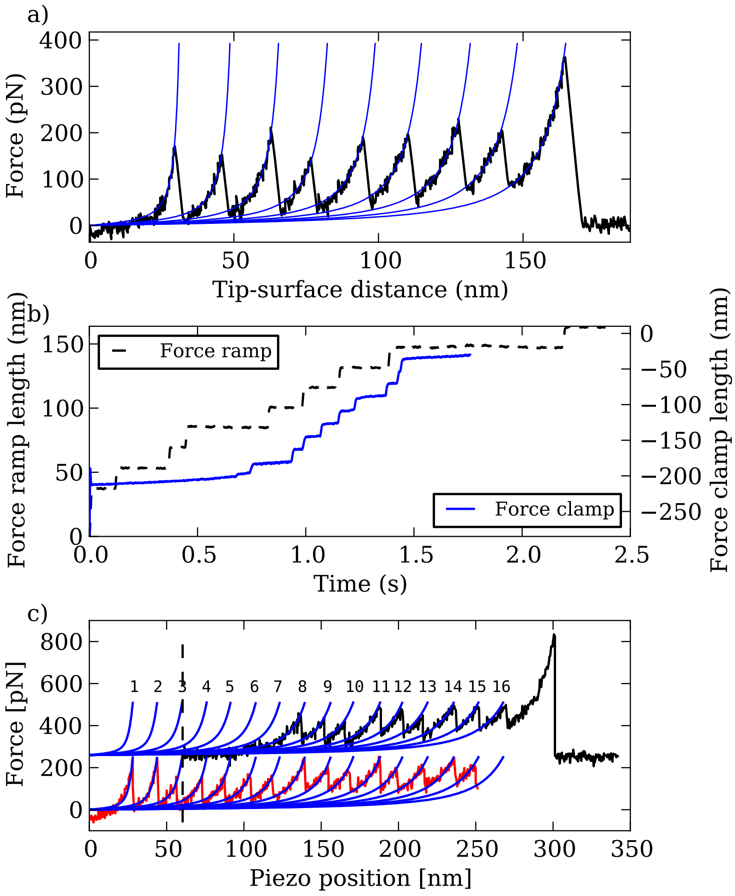


Figure 4.2: Representative examples of the type of data captured under different AFM modes of operation used in our study: protein unfolding under *a*)velocity-clamp, *b*)force-clamp (continuous line) and force-ramp (dashed line), and *c*)protein refolding using a custom double-pulse method [18] (described in detail in Ref. [18] and Section 4.4.5).

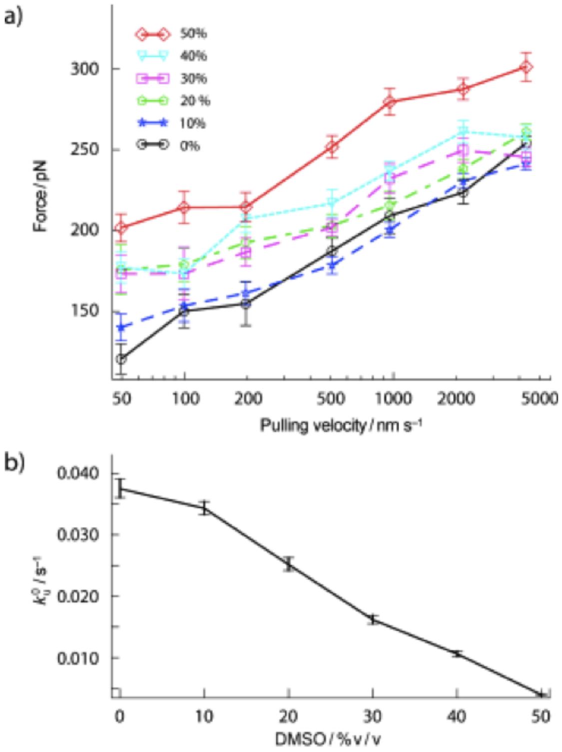


Figure 4.3: *a*)Average mechanical unfolding force by pulling velocity at various DMSO concentrations (values shown as %v/v). The error bars extend two standard errors of the mean unfolding force above and below the mean value. (The number of unfolding events are given in the Table 4.1.) *b*)Spontaneous dissociation rate and 68.3% confidence intervals with varying DMSO concentration, calculated with a fixed $\Delta x_u$ value of 0.165nm.

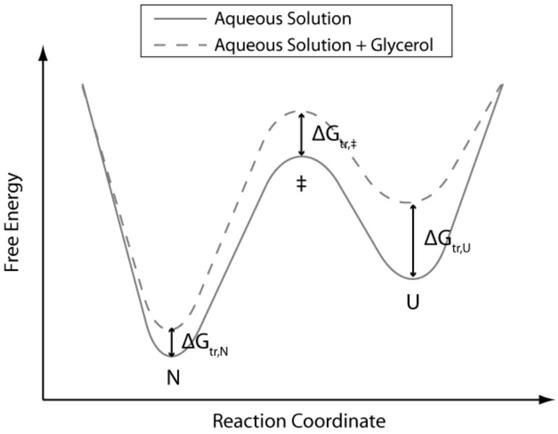


Figure 4.4: Representation of the unfoldingrefolding energy landscape of a two-state protein under the effect of a protecting osmolyte (-water only, – water+protecting osmolyte). Kinetically, a protecting osmolyte accelerates protein folding by increasing the height of the unfolding activation barrier ($\Delta\Delta G_u = \Delta G_{tr,\ddagger} - \Delta G_{tr,N} > 0$) and decreasing the height of the folding activation barrier ($\Delta\Delta G_f = \Delta G_{tr,U} - \Delta G_{tr,\ddagger} > 0$), where $\Delta G_{tr,\ddagger}$, $\Delta G_{tr,N}$ and $\Delta G_{tr,U}$ represent the free energy of transfer from water to the waterosmolyte mixture of the transition (‡), native (N), and unfolded (U) state, respectively. It follows as a thermodynamical consequence that $\Delta\Delta G = \Delta G_{tr,U} - \Delta G_{tr,N} = \Delta\Delta G_f + \Delta\Delta G_u > 0$, where $\Delta G$ is the free energy difference between the unfolded and native states, therefore generating the inequality commonly referred to as the osmophobic effect.

tially dependent on the height of the unfolding transition barrier $\Delta G_u$ according to the formula $k_u^0 \propto \exp[-\Delta G_u/(k_B T)]$. Therefore, the difference in the heights of the unfolding transition barrier ($\Delta\Delta G_u$) in two different conditions (for example, DMSO concentrations) can be computed from the logarithm of the ratio of the values in the two conditions.

The obtained $\Delta x_u$ values at each DMSO concentration are in very good accord with previously published $\Delta x_u$ values of GB1 in absence of DMSO [49]. More specifically, the $\Delta x_u$ values remained almost unchanged around a mean value of 0.165nm, without any apparent trend (see Table 4.2). The amount of variation in $\Delta x_u$ for all tested conditions is small in comparison with the size difference between the DMSO and water molecules, thus indicating that no DMSO molecules bridge the gap between the force-bearing beta strands of GB1 in the unfolding transition

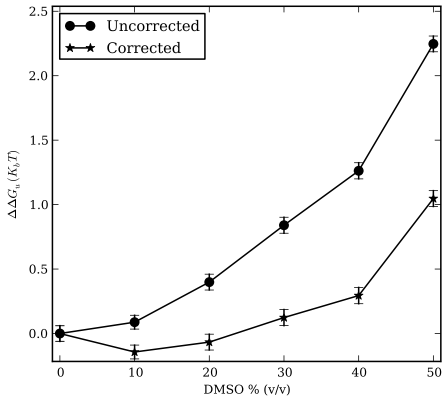


Figure 4.5: Height of the free energy barrier, by DMSO concentration, relative to the height in absence of DMSO (●), together with the viscosity-adjusted estimates ($\star$). Vertical error bars denote one standard deviation in both directions.

state.

We thus fixed $\Delta x_u$ to its mean value and performed a single-parameter fit of across all examined DMSO concentrations. The resulting spontaneous dissociation rate values showed a clear inverse correlation with DMSO concentration (Fig. 4.3b). Accordingly, a consistent decreasing trend of the computed force-dependent unfolding rate was observed in force clamp experiments (Fig. 4.6). These observations demonstrate that the height of the unfolding transition barrier increases with increasing DMSO concentration. After correcting the data for the viscosity of DMSO-water mixtures (see Section 4.4.3), the general increasing trend of $\Delta\Delta G_u$ is still maintained, with some uncertainty only in the low-concentration range (Fig. 4.5). Therefore, the increase in viscosity alone is not sufficient to explain the protective effect of DMSO on the native state of GB1 against unfolding.

Similarly, we mapped the folding-energy landscape of GB1 on the mechanical reaction coordinate in the absence of DMSO and at a DMSO concentration of 30% v/v, extracting the distance between the unfolded state and the transition state $\Delta x_f$ (see Fig. 4.4) and the spontaneous folding rate (see Section 4.4.5). The data both in the absence and presence of DMSO is explained by the common value $\Delta x_f = 2.57$ nm (which is compatible with the previously published $\Delta x_f$ value of GB1 in absence of DMSO [49]), resulting in $k_f^0 = 546 \pm 36 s^{-1}$ in absence of

DMSO and $k_f^0 = 693 \pm 64 s^{-1}$ with 30% v/v DMSO. The osmolyte increased the height of the folding activation barrier by $\Delta\Delta G_f = 0.24 \pm 0.11 k_B T$ which becomes $\Delta\Delta G_f = 0.95 \pm 0.11 k_B T$ when correcting for viscosity effects (see Section 4.4.5).

Taken together, the results outlined above demonstrate that the protecting osmolyte DMSO is able to slow down the unfolding kinetics of the globular domain GB1 by raising the free energy of the transition state with respect to that of the native state. In particular, they demonstrate that the above effect is obtained by an indirect mechanism without any apparent complexation of the osmolyte to the protein in the transition state. Moreover, DMSO also accelerates the folding rate of GB1 by raising the free energy of the unfolded state with respect to the transition state. At the same time, the fact that $\Delta x_u$ and $\Delta x_f$ values remained constant across all tested experimental conditions implies that DMSO does not change the force dependence of the folding and unfolding kinetics, showing that its mode of action is independent of the effects of mechanical tension on the energy landscape of the protein.

According to the osmophobic model, a protecting osmolyte disfavors backbone solvent hydrogen bonding, and therefore it predominantly destabilizes structural states with a higher number of such bonds, for example, the unfolded state. If this is true, it would be expected that (*i*) protecting osmolytes should act on all the conformations assumed by the protein, and (*ii*) the magnitude of their effect should be proportional to the amount of backbone solvent hydrogen bonds therein [99]. Obviously, the number of intrapeptide bonds is higher in the native state of a globular protein than in its unfolded state, while its transition state generally contains an intermediate number of such bonds [100]. Accordingly, a protecting osmolyte should mostly destabilize the unfolded state, but also to a lesser extent the transition state [101]. All of the above considerations collectively suggest that if an indirect, backbone-based osmophobic effect is acting, then the thermodynamic destabilization effect of the unfolded state should come to be through a kinetic protection against unfolding (that is, a slower spontaneous unfolding rate), together with a kinetic facilitation of folding (that is, a faster spontaneous folding rate; Fig. 4.4). The results of the SMFS experiments described above are in perfect accord with this picture.

The ability of another protecting osmolyte, glycerol, to slow down the unfolding kinetics of ubiquitin [76] and the I27 immunoglobulin module of human cardiac titin [15] was previously demonstrated by single-molecule mechanical studies. In both cases, the osmolyte molecules were found to actively participate in the transition-state structure. However, similarly to what we observed with GB1 in DMSOwater mixtures, a kinetic protection against unfolding was observed in the two mentioned experiments [76, 15] across a wide range of glycerol concentrations. As the protection resulted partly from the indirect, backbone-based osmophobic effect and partly from the more specific -strand bridging phenomenon, from these previous experiments it was not possible to numerically evaluate or even claim evidence for the osmophobic effect. Preliminary experiments performed by us show that the solvent-bridging phenomenon observed with I27 in presence of glycerol does not occur with GB1, thus reinforcing the hypothesis that the formation of direct proteinosmolyte interactions at the transition state is not an universal phenomenon.

As deviations from the osmophobic mechanism are caused by direct complexation of the osmolyte with specific residues at the transition state (thus making alternative folding/unfolding pathways available), the fold complexity of a given protein may influence its susceptibility to show such deviations. Therefore, we expect that simple and small proteins with low-fold complexity, which lack any folding/unfolding intermediates (as the GB1 domain examined in this study), should thus be less prone to show deviations from a purely osmophobic behavior under the effect of protecting osmolytes.

In summary, using a combination of SMFS-based experimental strategies we presented to the best of our knowledge the first single-molecule evidence of a protecting osmolyte slowing down the unfolding kinetics of a globular protein while concurrently accelerating its folding rate, independently of the mechanical tension applied and without any complexation in the unfolding transition state. This observation serves as evidence for a purely indirect backbone-based mechanism for the osmophobic effect [85, 88, 91].

## 4.4 Materials and Methods

### 4.4.1 Single Molecule Force Spectroscopy Experiments

For the experiments we used two AFM setups. One setup consisted in a Picoforce AFM with Nanoscope IIIa controller (Digital Instruments, Plainview, NY, USA) with V-shaped silicon nitride cantilevers (NP; Digital Instruments) whose spring constant was calibrated by the thermal noise method [8]. Using this setup we performed velocity-clamp mechanical unfolding experiments in mixed water-DMSO mixtures buffered with Tris/HCl (10 mM, pH 7.5), with DMSO concentrations of 0%, 10%, 20%, 30%, 40%, and 50% v/v, at pulling velocities of 50.1, 100, 198, 513, 969, 2180 and 4360 nm/s. At least three different cantilevers have been used with each DMSO concentration. Data extraction has been performed using Hooke [71]. With the same setup we also performed refolding experiments using a double-pulse protocol [18], in the absence of DMSO and with a DMSO concentration of 30% v/v. Another setup consisted in a custom-built AFM [78]. All the measurements were carried out in Tris/HCl (10 mM, pH 7.5) buffer. The spring constant of each individual cantilever was calibrated in solution using the equipartition theorem [8]. With this setup two kinds of experiments were performed. Force-ramp experiments were performed at a ramping speed of 300 pN/s, covering the same gamut of DMSO concentrations as the velocity-clamp experiments. Finally force-clamp experiments were performed at constant forces of 50 pN, 75 pN, 100 pN, 150 pN, 200 pN and 250 pN, and the DMSO concentrations covered were 0%, 10%, 20%, 30% v/v. However for the force-clamp experiments only some selected force-DMSO concentration combinations were used, as seen in Table 4.4 and Table 4.5.

### 4.4.2 Preparation of protein constructs

Plasmid pQE80-$(GB1)_8$ encoding GB1 polyprotein was obtained as previously described [49]. Briefly, $(GB1)_8$ gene was obtained by iterative cloning of the sequence on the basis of the identity of the sticky ends generated by BamHI and BglII restriction enzymes. Then, $(GB1)_8$ polyprotein was overexpressed in DH5$\alpha$ strain and purified by $Ni^{2+}$-affinity chromatography. The purified polyprotein sample was at a final concentration of 0.34 mg/ml and was stored at -20°C in PBS buffer with 0.2% (v/v) sodium azide, diluted with 30% (v/v) glycerol.

Table 4.1: Number of unfolding events by DMSO concentration and Velocity.

| DMSO | Velocity [nm/s] | | | | | | |
|---|---|---|---|---|---|---|---|
| v/v | 50.1 | 100 | 198 | 513 | 969 | 2180 | 4360 |
| 0% | 22 | 35 | 28 | 43 | 40 | 112 | 334 |
| 10% | 24 | 45 | 63 | 168 | 196 | 219 | 345 |
| 20% | 6 | 25 | 59 | 69 | 82 | 121 | 239 |
| 30% | 33 | 30 | 54 | 134 | 73 | 83 | 157 |
| 40% | 56 | 50 | 57 | 77 | 112 | 126 | 72 |
| 50% | 58 | 38 | 53 | 100 | 86 | 143 | 104 |

Table 4.2: Unfolding kinetic parameters of GB1 obtained by maximizing the likelihood of the unfolding data.

| DMSO [v/v] | $\Delta x_u$ [nm] | $k_u^0$ [$s^{-1}$] |
|---|---|---|
| 0% | 0.163 | 0.0419 |
| 10% | 0.167 | 0.0307 |
| 20% | 0.178 | 0.0126 |
| 30% | 0.154 | 0.0289 |
| 40% | 0.173 | 0.00686 |
| 50% | 0.154 | 0.00770 |

### 4.4.3 Analysis of multiple-speed velocity-clamp data

The number of unfolding events in our multiple-speed, multiple DMSO-concentration velocity-clamp experiments can be found in Table 4.1.

We obtained the two unfolding kinetic parameters via maximum likelihood estimation using the WLC model [17]. Shortly, for each DMSO concentration, the unfolding events at all velocities were used to find the most likely distance to the transition state $\Delta x_u$ and spontaneous dissociation rate $k_u^0$. The point estimates thus obtained can be found in Table 4.2.

Fixing the distance to the transition state to the mean value $\Delta x_u = 0.165$ nm, the spontaneous dissociate rate $k_u^0$ was then estimated, together with 68.3%

confidence intervals as described in [17] (see Fig. 4.3b).

The zero-force unfolding rate is given by the formula $k_u^0 = Aexp[-\Delta G_u/(k_B T)]$ where the prefactor A depends on the vibrational modes inside the metastable minimum and at the transition state and the friction. [83] Following the approach of [15], we analyzed our data in two ways. First we considered the prefactor A to stay constant across different DMSO concentrations, in which case the change in the height of the unfolding transition barrier relative to its height in the absence of DMSO was computed with the formula $\Delta\Delta G_u(c) = -k_B T log[k_u^0(c)/k_u^0(0)]$, with $k_u^0(c)$ standing for the spontaneous dissociation rate at DMSO concentration $c$. Second, since the prefactor $A$ is inverserly proportional to solvent viscosity [83], we also used a more accurate formula that takes into account this effect, namely $\Delta\Delta G_u(c) = -k_B T log\{[k_u^0(c) * v(c)]/[k_u^0(0) * v(0)]\}$, where $v(c)$ is the viscosity of the water-DMSO mixture. We derived the viscosity values by linear interpolation of the data in [102]. The standard deviations were computed by making the simplifying assumption that for each DMSO concentration $c$, the $k_u^0(c)$ estimate comes from a normally distributed population. The $\Delta\Delta G_u(c)$ values uncorrected for viscosity, together with the viscosity-adjusted ones, are plotted in Fig. 4.5.

### 4.4.4 Force-ramp and force-clamp experiments

The force-dependent unfolding rate in the force-clamp experiments can be seen in Fig. 4.6a.

The number of unfolding events by DMSO concentration in force ramp experiments is listed in Table 4.3. The number of unfolding events and traces in the force-clamp experiments are listed in Table 4.4 and Table 4.5 respectively.

### 4.4.5 Refolding experiments

We performed refolding experiments and their analysis using in-house developed software, in conjunction with Nanoscope software v6. A scheme of the double-pulse protocol can be see in Fig. (4.7).

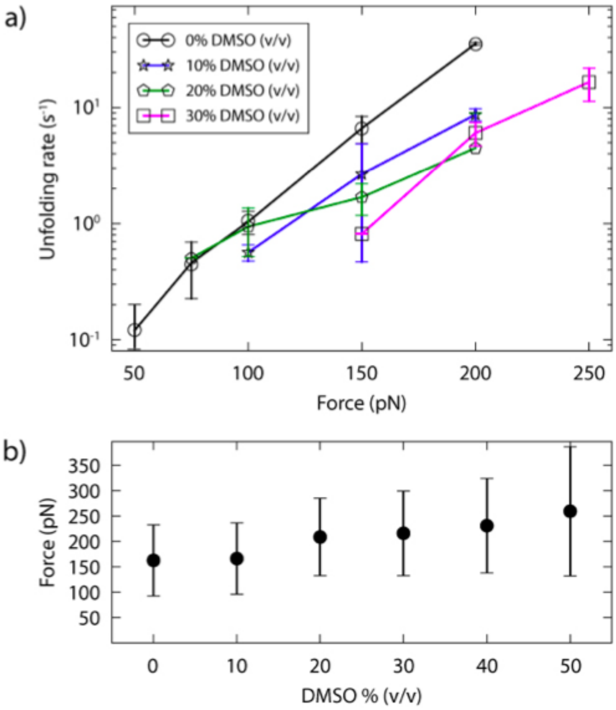


Figure 4.6: *a*) Force-dependent unfolding rate extracted from force-clamp experiments. The number of unfolding events can be found in Table 4.4. Vertical bars extend one standard deviation above and below the mean value. *b*) Average unfolding force at various DMSO concentrations in force-ramp experiments at 300 pN/nm. The number of unfolding events can be found in Table 4.3. The error bars extend one standard deviation of the mean unfolding force above and below the mean value.

Table 4.3: Number of unfolding events by DMSO concentration in force-ramp experiments.

| DMSO [v/v] | Count |
|---:|---:|
| 0% | 494 |
| 10% | 117 |
| 20% | 340 |
| 30% | 108 |
| 40% | 134 |
| 50% | 66 |

Table 4.4: Number of unfolding events by DMSO concentration and force in force-clamp experiments.

| DMSO [v/v] | Force [pN] | | | | | |
|---|---|---|---|---|---|---|
| | 50 | 75 | 100 | 150 | 200 | 250 |
| 0% | 51 | 69 | 110 | 106 | 55 | - |
| 10% | - | - | 28 | 30 | 35 | - |
| 20% | - | 10 | 34 | 56 | 21 | - |
| 30% | - | - | - | 4 | 30 | 19 |

Table 4.5: Number of traces by DMSO concentration and force in force-clamp experiments.

| DMSO [v/v] | Force [pN] | | | | | |
|---|---|---|---|---|---|---|
| | 50 | 75 | 100 | 150 | 200 | 250 |
| 0% | 7 | 9 | 17 | 18 | 9 | - |
| 10% | - | - | 5 | 5 | 4 | - |
| 20% | - | 2 | 6 | 8 | 4 | - |
| 30% | - | - | - | 2 | 5 | 3 |

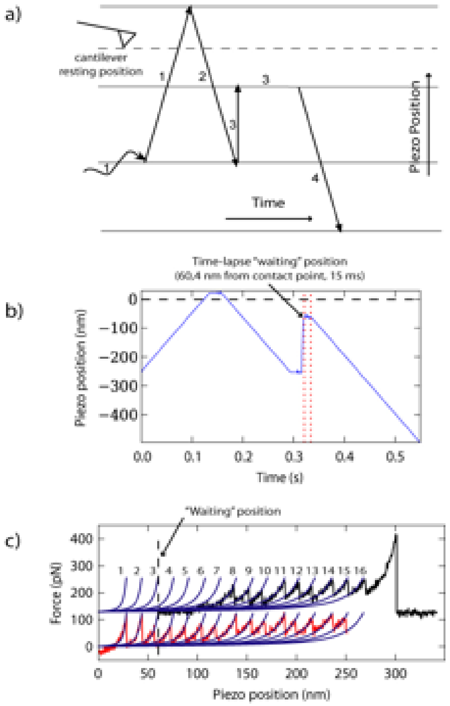


Figure 4.7: *a*) Schematic representation of the movement of the piezo positioner during one iteration of the double-pulse protocol. During Step 1, the piezo is approached to the cantilever until it pushes onto its tip for a predetermined distance. In Step 2 the protein is fetched and some of its modules are unfolded. In Step 3 the piezo is approached again to the cantilever, without touching it, maintaining a predetermined distance and then it waits there for a given amount of time. In Step 4 the piezo is retracted, unfolding those modules that did refold in during the time lapse, as well as all other modules that were not unfolded in Step 2, until complete detachment of the polyprotein. Steps 1 and 2 constitute the first phase of the double-pulse protocol, while Steps 3 and 4 constitute the second phase. *b*) Trace of the real movement performed by the piezoelectric positioner during one typical experimental iteration of the double-pulse protocol. The actual duration of the "waiting for refolding" time lapse, and the actual end-to-end distance of the protein during this interval can be measured directly for each iteration. The resulting pair of force curves from this iteration are shown in panel *c*).

### Gathered refolding data

In order to efficiently fit Bell's model one needs to find a couple of configurations with different waiting extensions (Step 3), and waiting times selected such that

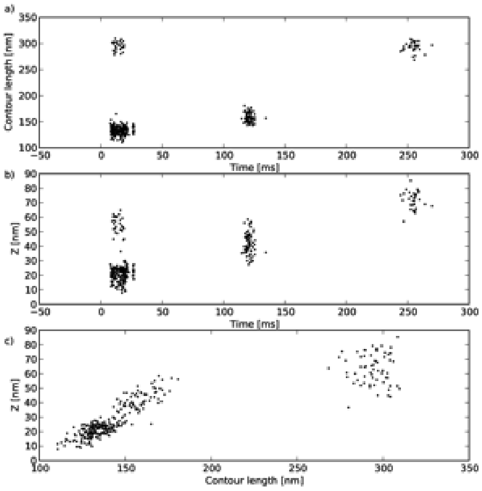


Figure 4.8: The refolding data in the absence of DMSO, summarized as one point for each double-pulse protocol execution considered. a) The contour length at the beginning of the second pulse is plotted against the waiting time. b) The waiting extension is plotted against the waiting time. c) The waiting extension is plotted against the contour length at the beginning of the second pulse.

the folding probability is neither too small, nor too high. We varied also the ramp distance of the first phase, using both higher values compatible only with the dimer $(GB1)_{16}$ and smaller values compatible with both $(GB1)_8$ and $(GB1)_{16}$. With our approach we can only roughly control these parameters a-priori, but we can accurately read the actual values a-posteriori, after each execution of the double-pulse protocol. Fig. 4.8 contains scatterplots of the waiting time for refolding to happen, the contour length at the start of the waiting period and the piezo extension during the waiting period in the absence of DMSO, while Fig. 4.9 shows similar data in the presence of DMSO 30%.

For visualisation purposes only, the data of refolding experiments has been projected onto three axes which are the amount of extension force computed for the start of the waiting period, the waiting time for refolding to happen, and the percentage of modules that refolded, resulting in Fig. 4.10 in the absence of DMSO and Fig. 4.11 in the presence of DMSO 30% v/v.

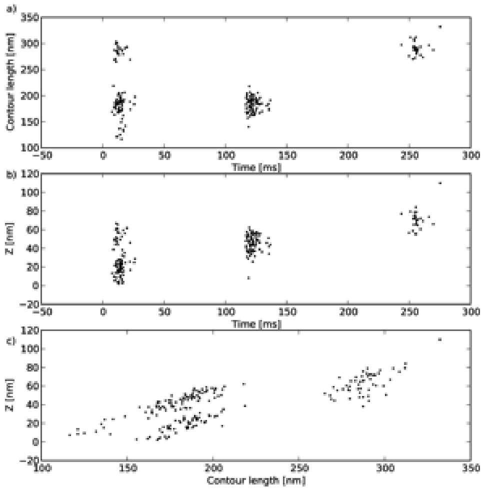


Figure 4.9: The refolding data in the presence of DMSO 30% v/v. The subplots and the axes are to be interpreted as in Fig. 4.8.

### Maximum Likelihood estimation of folding kinetic parameters

We used a Maximum Likelihood procedure for estimating the folding kinetic parameters, starting from the formula describing the force-dependent folding rate $k_f(F) = k_f^0 \exp[-F\Delta x_f/(K_bT)]$ where $k_f^0 = B\exp[-\Delta G_f/(K_bT)]$ and B is a constant prefactor inversely proportional to solvent viscosity. The procedure takes into account the fact that after each refolding event, the contour length of the construct decreases, and therefore the stretching force increases, even if the piezo is in a constant position at non-zero extension (manuscript in preparation). The likelihood function has been evaluated on a grid of values $\Delta x_f$ and $k_f^0$ with a step size of 0.1 nm for the former and 50 $s^{-1}$ for the latter, together with a local gradient search. Bootstrapping was used to compute the estimation error.

The most likely folding kinetic parameters in the two conditions, together with standard errors and the number of molecules considered, can be seen in Table 4.6.

### Further clarification of Fig. 4.2c

The lower curve represents the preliminary protein fetching phase, during which the protein is picked up by the cantilever and stretched up to a fraction of its total

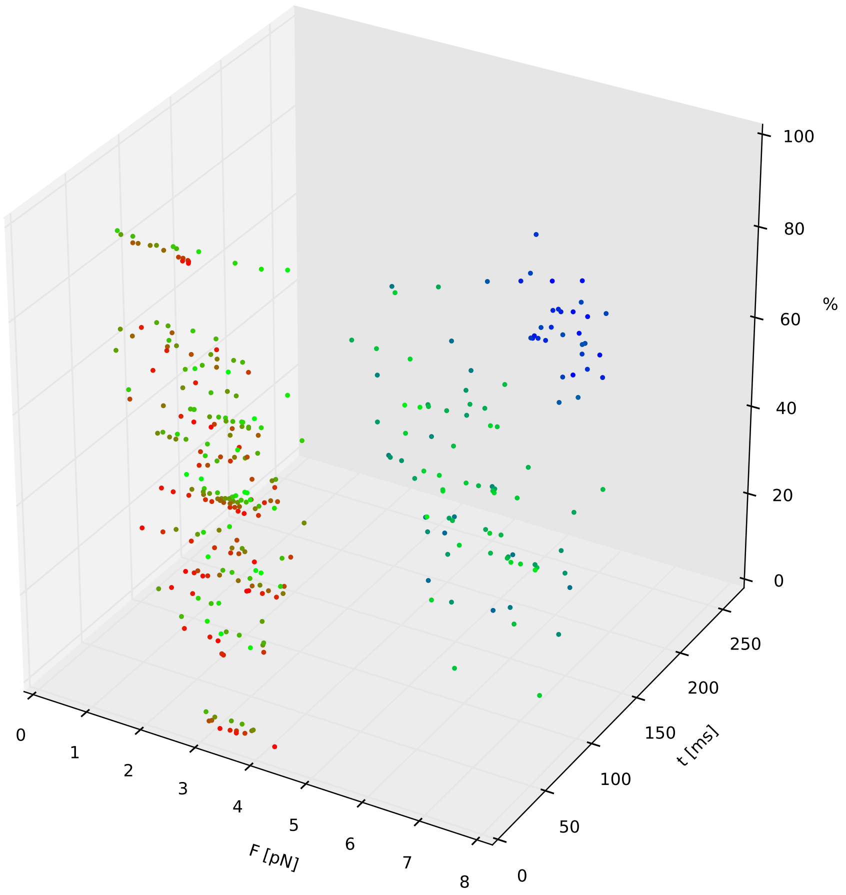


Figure 4.10: The data collected in the absence of DMSO is projected onto three axes: the computed force at the start of the waiting period, the waiting time and the percentage of refolded modules. The colors have been assigned by computing the rank of each data point in the list of points sorted by time, and then the colors were taken from a color gradient starting from red (lowest ranks) to green (middle ranks) and blue (highest ranks).

(detachment) length, resulting in the unfolding of some of its modules (15 out of 16 in this example). The higher curve (shifted by 160 pN for readability) represents the subsequent phase, during which the same molecule, after a predetermined time lapse spent at the extension marked by the vertical dashed line, is stretched until the molecule is released. There are 9 unfolding peaks in the second phase, 8 of which are in register with peaks 8-15 from the fetching curve, meaning that 8

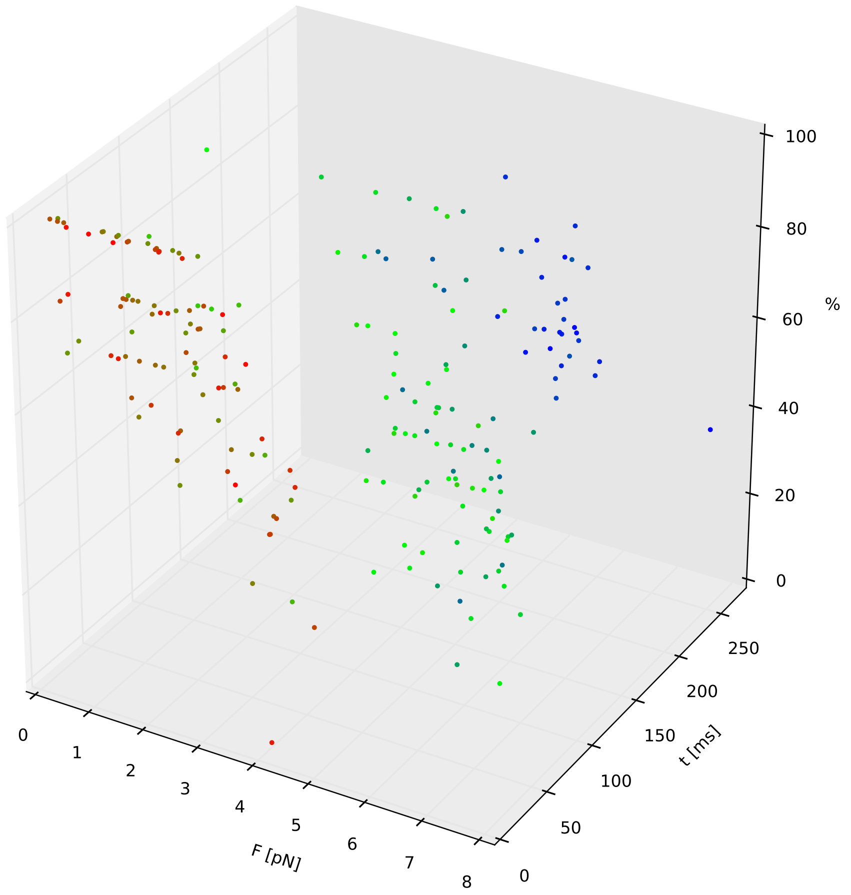


Figure 4.11: The data collected in the presence of DMSO 30% v/v, projected onto three axes as in Fig. 4.10.

modules managed to refold. The trace was obtained from a dimer of $(GB1)_8$, having 16 modules in total. Note that only 15 out of the 16 modules could have refolded, since one module was not unfolded during the fetching phase, and peak 16 corresponds either to the unfolding of the last mentioned module, or of any other module that was unfolded during the fetching phase and subsequently refolded. The vertical dashed line represents the extension to which the piezo was moved after having unfolded the 15 modules, to give them the possibility to refold for a certain amount of time (not shown). WLC fits have been performed for the 15 peaks of the fetching phase, after which one more contour length has been

Table 4.6: Folding kinetic parameters of GB1 obtained by maximizing the likelihood of the refolding data, leaving both parameters free.

| DMSO [v/v] | No Molecules | $\Delta x_f$ [nm] | $k_f^0$ [$s^{-1}$] |
|---:|---:|---:|---:|
| 0% | 331 | 2.53 ± 0.12 | 500 ± 85 |
| 30% | 208 | 2.61 ± 0.15 | 702 ± 202 |

computed by adding 18 nm to the contour length of the $15^{th}$ peak. The last mentioned contour length is the length of the construct at the start of the waiting period. Each of the 16 WLC curves is redundantly shown for a second time shifted higher, only for display purposes.

## 4.5 Acknowledgments

We are grateful to Prof. G.D. Rose for critical reading of an earlier version of the manuscript and valuable suggestions. We thank Enkelejda Miho for her help with velocity clamp experiments and Dr. Y. Cao for his help with force-clamp and force-ramp experiments. This work was supported by the Ministero dellUniversità e della RicercaFondo per gli Investimenti della Ricerca di Base (MIUR-FIRB) RBNE03PX83/001; MIURFIRB Progetto NG-lab (G.U. 29/07/05 n.175); Progetti di Ricerca di Interesse Nazionale (PRIN) 2008 (Prot. 2008KZ3E5, Prot. 2008SYP79), by the Natural Sciences and Engineering Research Council of Canada and Canada Research Chairs Program, and by Progetto di Ateneo 2008, Università di Padova.

# Chapter 5

# Observing the osmophobic effect in action at the single molecule level

Adapted with permission from Aioanei D, Tessari I, Bubacco L, Samorì B, Brucale M. "Observing the osmophobic effect in action at the single molecule level", Proteins. 2011 Jul;79(7):2214-23. Copyright (2011) Wiley-Liss, Inc.

## 5.1 Abstract

Protecting osmolytes are widespread small organic molecules able to stabilize the folded state of most proteins against various denaturing stresses in vivo. The osmophobic model explains thermodynamically their action through a preferential exclusion of the osmolyte molecules from the protein surface, thus favoring the formation of intrapeptide hydrogen bonds.

Few works addressed the influence of protecting osmolytes on the protein unfolding transition state and kinetics. Among those, previous single molecule force spectroscopy (SMFS) experiments evidenced a complexation of the protecting osmolyte molecules at the unfolding transition state of the protein, in apparent contradiction with the osmophobic nature of the protein backbone.

We present single-molecule evidence that glycerol, which is a ubiquitous protecting osmolyte, stabilizes a globular protein against mechanical unfolding with-

out binding into its unfolding transition state structure. We show experimentally that glycerol does not change the position of the unfolding transition state as projected onto the mechanical reaction coordinate.

Moreover, we compute theoretically the projection of the unfolding transition state onto two other common reaction coordinates, i.e. the number of native peptide bonds and the weighted number of native contacts. To that end, we augment an analytic Ising-like protein model with support for group-transfer free energies. Using this model, we find again that the position of the unfolding transition state does not change in the presence of glycerol, giving further support to the conclusions based on the single-molecule experiments.

## 5.2 Introduction

### 5.2.1 The osmophobic effect

Small organic solutes such as polyhydric alcohols, free amino acids and methylamines are employed by several living organisms as a means to enhance protein stability in stressed conditions of various sorts, including extreme temperatures, dehydration, high hydrostatic pressure, or the presence of denaturants. They are almost ubiquitous in all domains of cellular life, for example allowing plant seeds to remain viable for extremely long periods even when desiccated, animals to function at the extreme pressure of deep oceans or to go into anhydrobiosis or other forms of cryptobiosis, and proteins in the human kidney to resist denaturation despite high urea concentration [103, 88, 104].

These compounds are commonly referred to by several different names emphasizing specific facets of their behavior, the most frequently employed being "chemical chaperones", "nonionic kosmotropes", "folding agonists", and the one we will use throughout this paper, "protecting osmolytes". Typical examples of protecting osmolytes are trimethylamine N-oxide, betaine, sucrose, trehalose, sarcosine, sorbitol, proline, glycerol, and dimethyl sulfoxide (DMSO) [105, 13, 106].

A puzzling feature of protecting osmolytes is that their effects are interchangeable and additive [107], meaning that the stabilization of a protein can be obtained with a mixture of different osmolytes. This interchangeability, in addition to the fact that all osmolytes are able to affect the stability of highly diverse proteins in the same way, suggests that osmolytes of all sorts act in accord to a common and therefore general mechanism [88, 13, 91, 85].

The broad thermodynamic principle underlying this general stabilization mechanism is a preferential exclusion of the osmolyte from the immediate vicinity of the protein surface, and it is often referred to as the "osmophobic effect". As an example, the addition of a protecting osmolyte to water as a co-solvent mainly decreases the solvent quality for the protein backbone, disfavoring backbone-solvent hydrogen bonding and thus predominantly destabilizing structural states with a higher number of such bonds, i.e. the unfolded state (U) with respect to the native state (N) in which the backbone is less exposed. This results in an enhanced thermodynamic preference for the folded state [108, 92, 109, 110, 1].

Even though the thermodynamic model outlined above is now widely accepted, the molecular details of the mechanism by which osmolytes influence protein transition states and activation barriers still remain elusive. However, the putative mechanism implied by the osmophobic hypothesis is purely solvophobic, and as such does not require a preferential binding of the osmolyte to any specific site on the protein. To verify this hypothesis, the influence of osmolytes on the protein transition states needs to be observed.

Literature focusing on the study of the transition states associated with folding and unfolding of globular proteins in presence of osmolytes is still quite sparse, and several of the reported results are apparently in contrast with the simple thermodynamical description of osmolyte effect illustrated above. For instance, some studies reported that the osmolyte can bind to the protein in its folding/unfolding transition state [96, 76, 15] or that it can effect changes of the reaction pathway [14, 111]. Conversely, other studies reported that osmolytes stabilize the folded state of a protein without any appreciable change to the relevant transition state [99, 112, 113] and are thus in accord with the predictions of the osmophobic model.

Herein, we show experimental evidence of the osmophobic effect in action at the single molecule level. Via Single-Molecule Force Spectroscopy (SMFS) experiments, we estimate the unfolding kinetics of a small globular protein domain, finding that glycerol stabilizes it against unfolding without complexation into its unfolding transition state, as dictated by the osmophobic model. Moreover, we extend a previously reported Ising-like analytic protein model to make it applicable to proteins in aqueous osmolyte solutions, finding that its predictions are in good accord with the experimental results.

## 5.3 Materials and Methods

### 5.3.1 Preparation of protein constructs

Plasmid pQE80-$(GB1)_8$, encoding GB1 polyprotein was obtained as previously described [49]. Briefly, the $(GB1)_8$ gene was obtained by iterative cloning of the sequence on the basis of the identity of the sticky ends generated by BamHI and BglII restriction enzymes. Then, $(GB1)_8$ polyprotein was overexpressed in DH5 strain and purified by $Ni^{2+}$-affinity chromatography. The purified polyprotein sample was at a final concentration of 0.34 mg/ml and was stored at -20℃ in PBS buffer with 0.2% v/v sodium azide, diluted with 30% v/v glycerol.

### 5.3.2 Single Molecule Force Spectroscopy Experiments

Constant velocity mechanical unfolding experiments were performed with a Veeco Picoforce Atomic Force Microscope (AFM) equipped with a DI Multimode Nanoscope IIIa controller (Bruker) and gold-coated, V-shaped silicon nitride cantilevers (NPG model; Bruker) with nominal spring constant 0.06 N/m. Unfolding experiments were performed on the homomeric polyprotein (GB1)8 in two different solvent conditions. In the first condition we used Tris/HCl buffer (10 mM, pH 7.5). In the second condition we used the same buffer but with the extra presence of glycerol as a cosolvent, at a concentration of 30% v/v. All the experiments have been performed at a temperature of approximately 28 C. For the experiments, a drop of the $(GB1)_8$-containing solution (10 µL, 0.1 g/L) was deposited on a flame cleaned glass coverslip for about 20 minutes. The fluid cell was then filled with either the plain buffer or the buffer with glycerol added as a cosolute, and sealed on top of the coverslip. Thermal tuning was performed in the respective solution to determine the cantilever spring constant. Pulling experiments were then started after a few minutes of incubation.

At the constant pulling velocities of 50.1, 100, 198, 513, 969, 2180 and 4360 nm/s we recorded (*i*) 22, 35, 28, 43, 40, 112, and 334 unfolding events, respectively, from 2, 2, 2, 2, 2, and 3 different experiments, respectively in plain buffer, and (*ii*) 56, 36, 40, 414, 722, 871, and 79 unfolding events, respectively, from 4, 3, 2, 7, 5, 7, and 3 experiments, respectively with glycerol as cosolvent. Data filtering and extraction was performed using Hooke [71].

### 5.3.3 Analysis of the velocity-clamp data

The unfolding kinetics at different velocities were well described by Bells equation which expresses the dependence of the unfolding rate on the force applied $k_u(F) = k_u(0)\exp[F\Delta x_u/(k_BT)]$ as an approximation to Kramers reaction rate theory [82] where $k_u(0)$ is the spontaneous unfolding rate in absence of applied force, $\Delta x_u$ is the unfolding distance separating the folded state and the unfolding transition barrier measured along the reaction coordinate, T is the temperature in Kelvin and $k_B$ is Boltzmanns constant. Note that the spontaneous dissociation rate can be expressed as

$$k_u(0) = A\exp[-\Delta G_u/(k_BT)], \tag{5.1}$$

where A is a prefactor and $\Delta G_u$ is the height of the unfolding activation energy barrier. The prefactor A depends on the vibrational modes inside the metastable minimum and at the transition state and the friction [83].

Maximizing the likelihood of the data [17] we obtained $\Delta x_u = 0.163$ nm in the absence of glycerol and $\Delta x_u = 0.153$ nm in the presence of glycerol 30% v/v, ignoring viscosity effects (see below). Since the difference is negligible in comparison to the size difference between the glycerol and water molecules [15, 60] and within experimental error, we inferred that the presence of glycerol does not change the characteristic unfolding distance of GB1. Taking into account as well existing estimates of the unfolding distance of GB1 [17, 69] we fixed the unfolding distance to $\Delta x_u = 0.165$ nm. Using this value we obtained $k_u(0) = 0.0372 \pm 0.0016s^{-1}$ in the absence of glycerol and $k_u(0) = 0.0059 \pm 0.0001s^{-1}$ in the presence of 30% v/v glycerol.

If we consider as a first approximation that the prefactor A is not affected by the addition of glycerol, taking the ratio of the logarithm of the estimated spontaneous unfolding rates in the two solvent conditions we obtain that the presence of glycerol at a concentration of 30% v/v increases the unfolding energy activation barrier by $\Delta\Delta G_u = 1.84 \pm 0.05k_BT$.

### 5.3.4 Taking into account the viscosity of the 30% v/v glycerol solution

We computed the viscosity of the 30% v/v glycerol solution at 28°Cas 2.357 cP using a formula reported elsewhere [114].

Immersion in a liquid environment dramatically alters the thermal noise spectrum of the cantilever, due to the strong effects of fluid loading [11, 115, 116]. Problems in the determination of the spring constant arise when the resonance frequency drops to 1 kHz or below, as it happens for soft cantilevers or highly viscous media [11]. However even in the presence of 30% v/v glycerol as a cosolvent the resonance frequency of our cantilevers stays above 2 kHz, above the frequency range likely to result in viscosity-induced errors in spring constant determination that are comparable to the error intrinsic to the thermal tuning method [117].

The measured unfolding forces are underestimated in high-viscosity solutions due to the viscous drag on the AFM cantilever [118, 119, 120, 121]. A way to remove this artifact has been recently reported [118]. According to the this procedure the drag force can be computed as

$$F_{vis} = b_l v_{liquid} - b_c v_{tip}, \tag{5.2}$$

where $v_{liquid}$ is the pulling velocity, $v_{tip}$ is the velocity of the tip, $b_l$ and $b_c$ are viscous drag coefficients related to the liquid motion and the bending of the cantilever (tip motion), respectively. These two parameters depend on tip-surface separation according to the formulae

$$b_l = 6\pi\eta a_l^2/(s + h_l) \tag{5.3}$$

and

$$b_c = 6\pi\eta a_c^2/(s + h_c), \tag{5.4}$$

where s is the tip-sample separation, $\eta$ is the solvent viscosity, $a_l$, $h_l$, $a_c$ and $h_c$ being parameters to be determined. We extracted the mentioned parameters by fitting Eq. 5.3, 5.4 to the $b_l$ and $b_c$ values reported in Fig. 3 of Ref. [118], obtaining the approximate values $a_l = 24.3$ µm, $h_l = 3.5$ µm, $a_c = 19.8$ µm and

$h_c = 2.9$ µm. We then adjusted the force values collected at each sampled tip-surface separation in the experimental data taken in 30% v/v glycerol according to Eq. 5.2. The average unfolding forces at 50.1, 100, 198, 513, 969, 2180 and 4360 nm/s increased by 0.04, 0.12, 0.26, 0.36, 0.65, 1.82, and 6.22 pN, respectively. After this adjustment we obtained $k_u(0) = 0.0057 \pm 0.0001\, s^{-1}$ for $\Delta x_u = 0.165$ nm.

We also took into account viscosity effects in the calculation of the change of the unfolding barrier height effected by glycerol 30% v/v since the prefactor A of Eq. (5.1) is inversely proportional to solvent viscosity [83, 60]. Adjusting the prefactor by the inverse of the solvent viscosity we obtained our final estimate of $\Delta\Delta G_u = 0.84 \pm 0.05\, k_B T$.

### 5.3.5 A Wako-Saitô-Muñoz-Eaton Ising-like model

We endowed the previously reported WSME Ising-like protein model [122, 123, 124, 125] with support for the free energies of transfer of amino-acid residues from water to water-osmolyte mixtures, as described below.

According to the WSME model, a protein N+1 amino acids long is described as a chain of N peptide bonds. The state of the protein is captured by the vector $\vec{m}$, where the i-th peptide bond is represented by binary variable $m_i$ which has only two possible values: 1 for a native peptide bond, and 0 for a non-native peptide bond. Therefore there are exactly $2^N$ states in the WSME model. The effective free energy of the system reads as

$$H_{WSME}(\vec{m}) = 1K \times k_B \epsilon \sum_{1 \leq i < j \leq N} h_{ij} \prod_{k=i}^{j} m_k - K_b T \sum_{i=1}^{N} q_i (1 - m_i) \tag{5.5}$$

where 1K is one Kelvin, $k_B$ is Boltzmanns constant, T is the temperature, $\epsilon$ is a dimensionless enthalpic scale, $h_{ij} \leq 0$ are dimensionless numbers representing the relative strength of the contact between the i-th and j+1-th amino acids and $q_i$ ¿ 0 represents the entropic cost of ordering bond i. The contact strength hij is commonly defined as 0 if j = i+1, and $\lceil q_{ij}/5 \rceil$ where $q_{ij}$ is the number of pairs of atoms, the first atom of the pair belonging to amino acid i and the second atom of the pair belonging to amino acid j+1, that in the native state are closer than 0.4 nm and $\lceil \cdot \rceil$ is the ceiling function [126, 127]. The thermodynamics of the WSME model can be exactly computed via the transfer-matrix approach [128].

The fraction of folded molecules can be estimated as

$$p(T) = [\alpha(T) - \alpha(\infty)]/[\alpha(0) - \alpha(\infty)]\,, \tag{5.6}$$

where $\alpha(T)$ is the thermodynamic average of the number of native peptide bonds at temperature $T$. At zero temperature we have $\alpha(0) = N$, while at infinite temperature the exact expression for the mentioned thermodynamic average[129] reads as:

$$\alpha(\infty) = \sum_{i=1}^{N} [1 + \exp(q_i)]^{-1}\,. \tag{5.7}$$

Assuming the entropic costs $q_i$ are known, the parameter $\epsilon$ can be fitted by imposing the folded fraction to be equal to the experimentally determined folded fraction at a certain temperature.

For the case $q_i = \ln 2$ (see Section 5.3.8), Eq. 5.6 becomes:

$$p(T) = 3\alpha(T)/(2N) - 1/2\,. \tag{5.8}$$

The number of native peptide bonds [128, 130] and the weighted number of native contacts [131]

$$\phi(\vec{m}) = - \sum_{1 \le i < j \le N} h_{ij} \prod_{k=i}^{j} m_k \tag{5.9}$$

have been proposed as order parameters for monitoring protein folding and unfolding. Assuming equilibrium folding and unfolding dynamics, one can characterize the height and position of energy barriers on either of these two reaction coordinates.

### 5.3.6 Extension of the WSME model to study proteins in aqueous osmolyte solutions

Tanfords Transfer Model has proved very successful at predicting the osmolyte-induced energetics of protein stability [109]. In order to study the kinetics of mechanical unfolding under the effect of osmolytes, we apply the Transfer Model to all the states of the WSME model.

A native stretch is defined as a sequence of consecutive amino acids connected by native bonds and delimited by two non-native bonds. More formally, the stretch delimited by bonds i and j is said to be native if $S_{ij} = 1$ where

$$S_{ij} = (1 - m_i)(1 - m_j) \prod_{k=i+1}^{j-1} m_k \tag{5.10}$$

with $0 \leq i < j \leq N + 1$ and considering $m_0 = m_{N+1} = 0$. In the limiting case of j i = 1 the stretch reduces to one amino acid [126, 127].

Accordingly, we consider each native stretch to have the exact same structure as in the native state of the full protein, and we add to the effective free energy of Eq. 5.5 a term representing the free energy of transfer $M_{ij}$ of the native stretch delimited by peptide bonds i and j, in isolation, relative to the denatured state:

$$H(\vec{m}) = H_{WSME}(\vec{m}) + \sum_{0 \leq i \leq j-2 \leq N-1} S_{ij} M_{ij} \tag{5.11}$$

with $M_{ij}$ approximated as

$$M_{ij} = \sum_{i<k\leq j} [\Delta g^{ref}_{bb,R[k]} (A^{den}_{bb,k} - A^{ij}_{bb,k}) / A^{ref}_{bb,R[k]} + \Delta g^{ref}_{sc,R[k]} (A^{den}_{sc,k} - A^{ij}_{sc,k}) / A^{ref}_{sc,R[k]}] \, , \tag{5.12}$$

where R[k] is the amino acid at position k of the protein chain, $g^{ref}_{bb,R[k]}$ is the free energy of transfer of amino acid R[k] in the reference state (which is determined by the type of experimental data available  see Materials and Methods), $A^{den}_{bb,k}$ is the accessible surface area (ASA) of the backbone unit at position k of the protein chain in the denatured state, $A^{ij}_{bb,k}$ is the accessible surface area of the backbone unit at position k of the protein chain in the isolated native stretch from peptide bond i to peptide bond j, $A^{ref}_{bb,R[k]}$ is the accessible surface area of the backbone of amino acid R[k] in the reference state, and finally $g^{ref}_{sc,R[k]}$, $A^{den}_{sc,k}$, $A^{ij}_{sc,k}$ and $A^{ref}_{sc,R[k]}$ are defined similarly to the last four mentioned quantities, with the only difference that they refer to the side chain rather than the backbone. Note that the procedure of scaling GTFEs by the accessible surface area has previously proved accurate at interpreting the thermodynamics of the osmolyte effect [109, 132, 133].

It can be seen in Eq. 5.12 that compared to the original WSME model, the effective free energy expression in our extended version has one extra term for each native stretch. Because no terms are introduced that simultaneously involve peptide bonds from multiple native stretches, the transfer matrix approach [128] used to solve the thermodynamics of the original WSME model is equally applicable our extended model. We present next the explicit formulae to compute in polynomial time the effective free energy as a function of the number of native peptide bonds and the weighted number of native contacts.

Let us denote by $Q_{ij}$ the effective free energy of the native stretch delimited by peptide bonds i and j (up to a constant term):

$$Q_{ij}/(k_BT) = 1K/T \times \epsilon \sum\nolimits_{i<k<l<j} h_{kl} + \sum_{k=i+1}^{j-1} q_k + M_{i,j}/(k_BT)\,, \tag{5.13}$$

where $0 \leq i < j \leq N+1$.

We denote by $\vec{f}$ a two-dimensional vector such that $f_{i,j}$ is either of the two order parameters of the native stretch delimited by peptide bonds i and j with $0 \leq i < j \leq N+1$, namely the number of peptide bonds (j - i - 1) or the weighted number of contacts $(-\sum_{i<k<l<j} h_{kl})$.

We compute recursively the effective free energy $R_{ik}(f)$ (up to a term that is constant when i stays constant) of the protein subchain consisting of only the first i peptide bonds and considering only states such that the relevant order parameter of the subchain has the value p, or, more formally, $\sum_{0\leq k<l\leq i+1} S_{kl} f_{kl} = p$:

$$\begin{aligned} &\exp[-R_{ip}(\vec{f})/(k_BT)] = \delta_{p,f_{0,i+1}} \exp[-Q_{0,i+1}/(k_BT)] \\ &+ \sum_{(1\leq j\leq i)\wedge(f_{j,i+1}\leq p\leq f_{0j}+f_{j,i+1})} \exp[-R_{j-1,p-f_{j,i+1}}(\vec{f})/(k_BT) - Q_{j,i+1}/(k_BT)], \end{aligned} \tag{5.14}$$

where $0 \leq i \leq N$ and $0 \leq p \leq f_{0,i+1}$.

Finally, the quantity we look for is the effective free energy $E_p$ of the full chain considering only the states having the relevant order parameter equal to p, or, more formally, $\sum_{0\leq k<l\leq N+1} S_{kl} f_{kl} = p$. This effective free energy is then computed using the formula

$$E_p(\vec{f}) = R_{N,p}(\vec{f}) - k_BT \sum_{k=1}^{N} q_k, \tag{5.15}$$

where an entropic correction has been included such that Eq. 5.11 and Eq. 5.15 lead to the same partition function:

$$\sum\nolimits_{0\leq p\leq f_{0,N+1}} \exp[-E_p(\vec{f})/(k_BT)] = \sum\nolimits_{m_i\in\{0,1\},1\leq i\leq N} \exp[-H(\vec{m})/(k_BT)]. \tag{5.16}$$

### 5.3.7 GTFE and ASA calculations

To apply Eq. 5.11 to the glycerol 30% v/v solution we used the statistical mechanics model of Ref. [1] to estimate the transfer free energy of a backbone unit to the concentration of glycerol 30% v/v, obtaining the value of 0.42 $k_B T$. In this case the reference state was the backbone unit of the tripeptide Gly-X-Gly, with accessible surface areas taken from Table (2) in the Supplementary Information of Ref. [109]. Amino acid side-chains were ignored because of the lack of data of free energy of transfer of side-chain residues to a glycerol 30% v/v concentration.

We computed the backbone ASA of the native stretches using PyMOL with a probe radius of 0.14 nm, while the ASA of the GB1 backbone in the denatured state was estimated using ProtSA [134, 135].

### 5.3.8 Fitting the Ising-like model

We adopted the convention $q_i = \ln 2$, which is a commonly adopted value since it is intrinsic to another more recent WSME extension [127].

We extracted the native structure of protein GB1 from the Protein Data Bank [136] (PDB code 1PGA [137]). We removed the hetero-atoms from the structure after which we added the missing hydrogen atoms and flipped amino acid 8 using MolProbity [138]. Subsequently, the MolProbity optimized structure was used for all calculations involving the structure of the native state of GB1.

By digitizing the data of Fig. (3a) from Ref. [139] at low salt and pH 7.5 we obtained the GB1 unfolding temperature as 329.15 K. By fitting this unfolding temperature into the model we obtained $\epsilon = 37.24$.

## 5.4 Results and Discussion

### 5.4.1 Testing the osmophobic hypothesis at the single-molecule level

Under mechanical load the effect of mechanical work on the energy landscape is felt by the transition state structure [60] regardless of its microsecond-long lifetime [140]. On this basis, Single Molecule Force Spectroscopy (SMFS) mechanical experiments allow mapping the protein energy landscape onto an easily quantifiable, geometrically relevant folding reaction coordinate, which coincides with the vector of force application [3]. Importantly, in contrast to bulk experiments, mechanical unfolding does not require the presence of denaturants that could interact with the osmolitic analyte. Not only is the size of the activation energy of the transition state measurable by SMFS experiments, but also a direct measure of the position of the transition state along the pulling coordinate can be obtained [76]. Consequently, from such experiments it is possible to gain geometric information about the transition state and appreciate its putative movements along the reaction coordinate in response to different conditions [76, 17].

Recently, J. Fernandez and coworkers employed SMFS experiments to characterize the influence of the osmolyte glycerol on the folding and unfolding kinetics of ubiquitin [76] and of the I27 module of human cardiac titin [15]. In both cases they demonstrated that one or more glycerol molecules partake in the unfolding transition state of the protein. These observations are thus in apparent contradiction with the osmophobic nature of the protein backbone.

Compared with proteins that commonly do not bear any mechanical function under physiological conditions, mechanical proteins represent but a small fraction of the proteins inside cells [69]. However many functionally non-mechanical proteins have mechanical stability. From this category, one of the smallest model proteins that are extensively used in protein stability studies is the B1 immunoglobulin-binding domain of protein G from Streptococcus, commonly referred to as GB1. Protein GB1 has been used to study protein-related topics as diverse as electrostatic and salt-screening contributions to protein stability [139], mechanisms of beta-sheet formation [141], biomimetic materials [142, 98] and even protein aggregation [143, 144]. The existing extensive characterization of the structure, thermodynamics and kinetics of GB1 [137, 145, 146, 147, 148, 149, 150, 151, 152, 153], together with its high mechanical stability [69, 142] and the fact that it unfolds mechanically without any observable intermediates on AFM time scales make from GB1 a very compelling choice for the study of the osmophobic effect at the single molecule level.

Pulling at constant velocity a multimodular construct made up of 8 tandem repeats of GB1, which we shall call $(\mathrm{GB1})_8$, characteristic saw-tooth force-extension curves as seen in Fig. 5.1 are obtained.

The presence of glycerol at a concentration of 30% v/v consistently shifted the

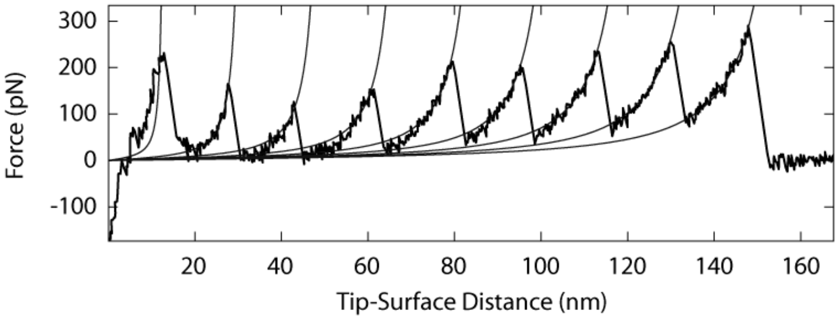


Figure 5.1: Velocity-clamp force-extension curves of $(GB1)_8$ show a characteristic saw-tooth pattern in which each peak represents one rupture event. The contour length of the stretched construct can be extracted from WLC fits before each rupture event, while the rupture forces provide information about the unfolding frequency.

unfolding force distribution of GB1 to higher forces, as seen in Fig. 5.2. Extracting the unfolding kinetics of GB1 (please refer to the Materials and Methods section) we found that glycerol reduces the spontaneous unfolding rate $k_u(0)$ of protein GB1. Moreover, the folded state of GB1 is stabilized by glycerol without changing the distance $\Delta x_u$ between the ground state and the transition state along the reaction coordinate, similarly to what was recently reported for GB1 in DMSO solutions [12]. The expected stabilization of the folded state occurs therefore without glycerol molecules bridging the critical beta strands at the transition state.

If the effect exerted by osmolytes on protein thermodynamics is indeed caused by preferential exclusion of osmolyte molecules from the immediate vicinity of the protein backbone, as proposed by the osmophobic model, it follows that (*i*) protecting osmolytes should act on all the conformations assumed by the protein, and (*ii*) the magnitude of their effect on a given conformation should be correlated to the amount of its backbone solvent hydrogen bonds (and thus anticorrelated with the amount of backbone backbone intrapeptide hydrogen bonds) [99].While some intrapeptide bonds are present in unfolding transition states of proteins [100], it is likely that the number of those bonds is higher in the native state, and lower in the unfolded state [101].

By invoking the activated-complex theory [121, 154] the above solvent-exposure considerations can be used to predict osmolyte-induced changes to the protein

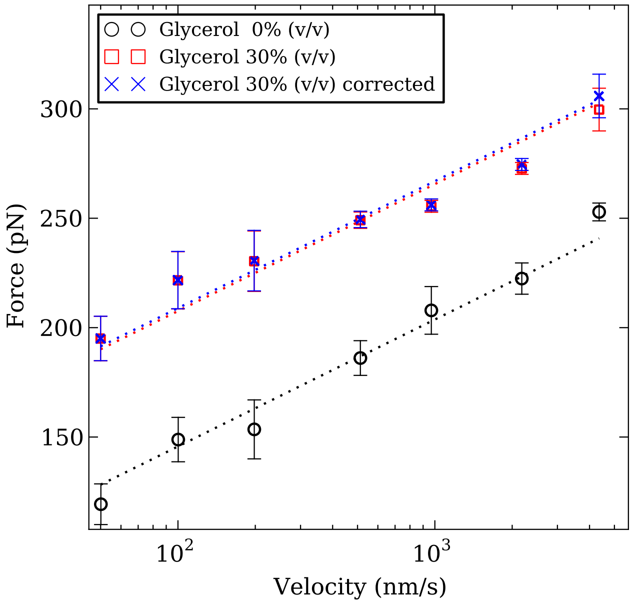


Figure 5.2: Average unfolding force of GB1 at various pulling speeds increases in the presence of glycerol 30% v/v, and increases again slightly when adjusting for the viscous drag-force on the cantilever. Only as a guide to the eye we fitted a dashed line to the average unfolding forces in each condition with a fixed slope of $k_B T/(\Delta x_u \log_{10}(e))$ where $e$ is Eulers constant and $\Delta x_u = 0.165$ nm. The formula for the eye-guide fixed slope has been inspired from the so-called "standard method" of kinetic parameter estimation [38], and it was not used for the statistical estimation of the kinetic parameters that we report, which was performed instead via Maximum Likelihood estimation (see Materials and Methods).

folding-unfolding transition state and kinetics. The main postulate of the activated-complex theory is the existence of equilibrium between reactants and the transition state [155]. From this perspective the thermodynamic stabilization effect of the native state should, as schematized in Fig. 5.3, also (*i*) be accompanied by a kinetic protection against unfolding, (*ii*) occur without binding of the osmolyte to the proteins unfolding transition state, and (*iii*) be accompanied by an augmentation of the proteins folding kinetics, as recently reported for GB1 in DMSO containing solutions [12].Therefore a protecting osmolyte should mostly destabilize the unfolded state, barely destabilize the folded state, and destabilize the transition state by an intermediate amount [101]. As outlined above, our results fully conform to this model, thus providing a single molecule level proof of the capability of glycerol to stabilize the native state of a globular protein against unfolding without active complexation into its unfolding transition state.

Previous SMFS experiments showed instead a biphasic dependence of the unfolding force of protein I27 on glycerol concentration [15]. The average unfolding force of I27 decreased in presence of concentrations of glycerol up to 30% (v/v), then increased again with an almost linear trend. This surprising behavior was explained as coming from the highly nonlinear dependency of the unfolding distance xu on glycerol concentration, which in turn is caused by glycerol molecules binding to the unfolding transition state of the protein. A structural model of the unfolding transition state of I27 was proposed that explains with remarkable accuracy the above-mentioned nonlinear behavior [15, 69]. According to this model, there are a limited number N of interaction sites that can be occupied by solvent molecules in the transition state structure. If glycerol molecules are present at one or more of the interaction sites then the unfolding distance acquires a larger value $\Delta x_u^G$, while it takes the normal value observed in water $\Delta x_u^W$ otherwise. The values $\Delta x_u^G$ and $\Delta x_u^W$ are representative of the size of the glycerol and water molecules, respectively. Then, the unfolding distance $\Delta x_u$ for any glycerol concentration is computed as an average quantity via the formula $\Delta x_u = (P_w)^N \Delta x_u^W + (1 - (P_w)^N)\Delta x_u^G$ where $P_w$ is the water occupancy probability of an interaction site, i.e., $1P_w$ is the volume fraction of glycerol in solution. Among all possible values of N, this model was found to best fit the experimental data when the number of interaction sites is set equal to the number of hydrogen bonding sites between the force-bearing beta-strands of I27. Similarly, glycerol molecules have been reported to be involved in

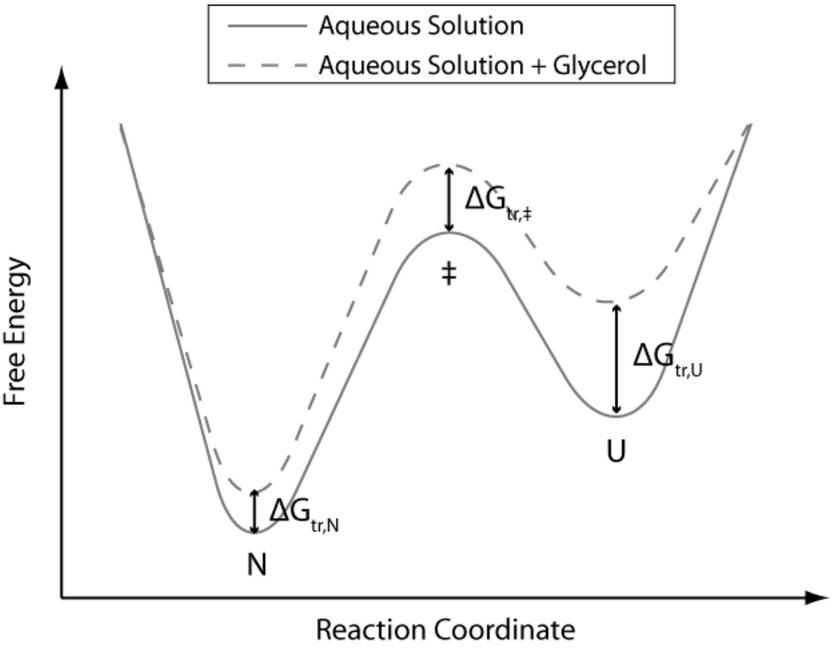


Figure 5.3: Schematic representation of the themodynamics and suggested kinetic implications of the osmophobic effect on a two-state globular protein. N represents the native state, U the unfolded state, and ‡ the transition state. In presence of a protecting osmolyte, the free energy of all states is raised proportionally to their respective amount of solvent-exposed backbone: $\Delta G_{tr,N} < \Delta G_{tr,\ddagger} < \Delta G_{tr,U}$, where $\Delta G_{tr,N}$ , $\Delta G_{tr,\ddagger}$ and $\Delta G_{tr,U}$ are the free energy of transfer from water to the water-protecting osmolyte mixture of the native state, the transition state and the unfolded state, respectively. Thermodynamically, it is known that the protecting osmolyte shifts the equilibrium between N and U towards N: $\Delta\Delta G = \Delta G_{tr,U}\Delta G_{tr,N} > 0$, where $\Delta G$ is the free energy difference between the unfolded state and the native state. Kinetically, we demonstrate that a protecting osmolyte ($i$) increases the height of the unfolding activation barrier: $\Delta\Delta G_u = \Delta G_{tr,\ddagger}\Delta G_{tr,N} > 0$, where $\Delta G_u$ is the free energy difference between the transition state and the native state and that ($ii$) it does so without any movement of ‡ along the reaction coordinate. Moreover, it is expected that protecting osmolytes also decrease the height of the folding activation barrier: $\Delta\Delta G_f = \Delta G_{tr,U} - \Delta G_{tr,\ddagger} > 0$, where $\Delta G_f$ is the free energy difference between the unfolded state and the transition state.

the mechanical unfolding transition state of the protein ubiquitin [76].

The mentioned osmolyte bridging phenomenon makes it very difficult to evaluate the magnitude of the osmophobic effect that glycerol has on proteins I27 and ubiquitin, or even to assert its presence at all. In contrast, as shown in this study, no such phenomena are induced on GB1 by the same concentration of glycerol that was previously found to actually depress the unfolding force of I27 to its minimum observed value (30% v/v). This suggests that the general, solvophobic thermodynamical force of the osmophobic model can coexist superimposed with other non standard effects such as binding of the osmolyte at specific protein sites [96, 76, 15] or modifications of the reaction pathway [14, 111]. Such effects however, are highly dependent on the specific protein-osmolyte pair and do not represent the general mechanism of protein stabilization by protecting osmolytes.

### 5.4.2 Theoretical predictions on the unfolding transition state in presence of glycerol: extending an Ising-like protein model

We also tested the osmophobic hypothesis via theoretical considerations based on an expanded version of a previously reported WSME Ising-like protein model [122, 123, 124, 125]. Our model extension (please refer to the materials and methods section) takes into account the group transfer free energies (GTFEs) of amino acid residues [109] from water to osmolyte containing solutions.

Briefly, the free energies predicted by the previously reported WSME model for each state of the protein chain were modified with an additional term accounting for the osmophobic effect. This additional term was calculated by imposing an energetic penalty proportional to the accessible surface area of the protein conformation, as described in detail in the Materials and Methods section.

This expanded model allowed us to project the energy landscape of GB1 onto two commonly studied reaction coordinates, namely (*i*) the number of native peptide bonds and (*ii*) the weighted number of native contacts.In both cases we found that adding glycerol as a cosolute at 30% v/v (*i*) lowers the free energy of the native state with respect to the unfolding transition state, and (*ii*) does not affect the position of the unfolding transition barrier along the respective reaction coordinate.

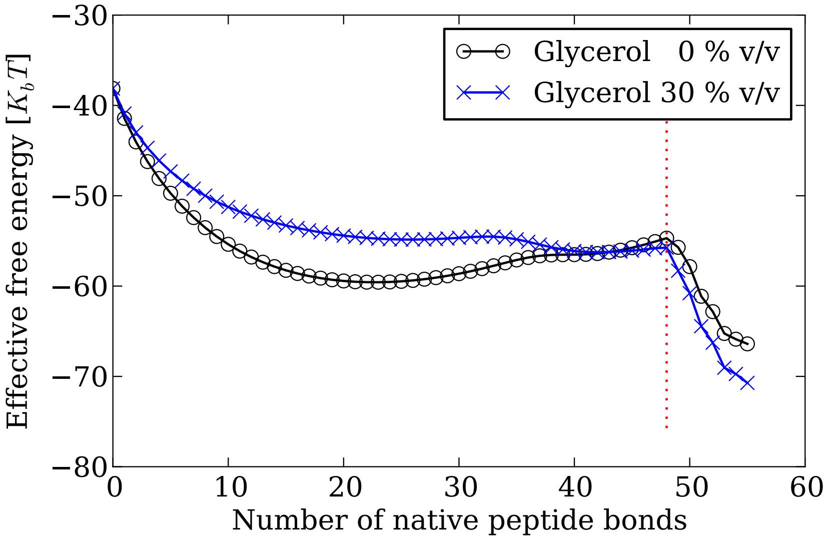


Figure 5.4: Energy landscape of protein GB1 projected onto the reaction coordinate given by the number of native peptide bonds in the absence and presence of 30% v/v glycerol, computed via Eq. 5.14, Eq. 5.15 with $f_{ij} = j - i - 1$, $0 \leq i < j \leq N + 1$. The vertical line indicates the unfolding transition barrier, which in both solvents has the same position with the number of native peptide bonds being equal to 48. The rightmost points in the graph have a number of native peptide bonds of 55 and represent the fully native state.

The energy landscape of GB1 as a function of the number of native bonds is shown in Fig. 5.4. Both in the absence and presence of 30% v/v glycerol, the main unfolding barrier stays fixed at a number of 48 native peptide bonds. By comparing the predicted effective free energies of the protein ground state (that is, the state with the highest number of native peptide bonds) and the transition state (that is, the state with 48 native peptide bonds), it is then possible to estimate the transition barrier height. The unfolding barrier increases in this case by $\Delta\Delta G_u = 3.271 k_B T$ in the presence of 30% v/v glycerol.

It has been proposed that the weighted number of native contacts is a better order parameter than the number of native bonds [131]. The energy landscape of GB1 as a function of the weighted number of native contacts is shown in Fig. 5.5. Both in the absence and presence of 30% v/v glycerol, the main unfolding barrier stays fixed at 382 native contacts. The unfolding barrier height increases in this

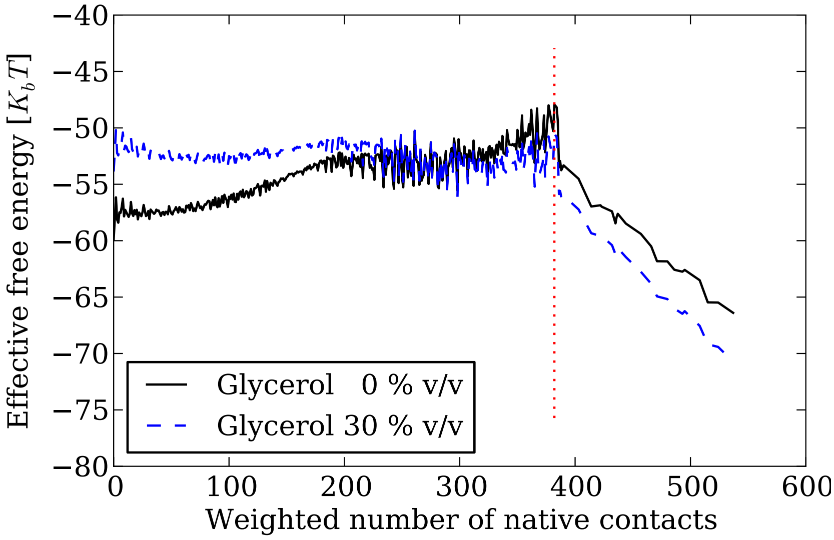


Figure 5.5: Energy landscape of protein GB1 projected onto the reaction coordinate given by the weighted number of native contacts $\theta(\vec{m})$defined in Eq. 5.9 in the absence and presence of 30% v/v glycerol, computed via Eq. 5.14, Eq. 5.15 with $f_{ij} = -\sum_{i<k<l<j} h_{kl}$, $0 \le i < j \le N+1$. The vertical line indicates the unfolding transition barrier, which in both solvents has the same position with the weighted number of native contacts being equal to 382.

case by $\Delta\Delta G_u = 1.758 k_B T$ in the presence of glycerol 30% v/v.

## 5.5 Conclusions

In summary, using the well-established dynamic SMFS experimental strategy we found single molecule evidence that glycerol can hinder the kinetics of protein unfolding without playing a bridging effect in the unfolding transition state of a globular protein. We showed that glycerol does not change the position of the unfolding transition barrier on the mechanical reaction coordinate.

In order to study the unfolding transition barrier on other reaction coordinates, we extended an Ising-like protein model with support for the osmophobic effect. This model predicts that the presence of glycerol does not move the unfolding transition state on two widely used non mechanical reaction coordinates, further supporting the osmophobic mechanism hypothesis.

These observations give further support to the backbone based osmophobic effect, which has recently been proposed as a new dimension of the protein-folding problem [88, 91, 85]. The ability of protecting osmolytes to stabilize proteins without causing displacement of their unfolding transition state may be necessary for their leaving the functional activity of proteins unaltered.

## 5.6 Acknowledgments

We thank Prof. Hongbin Li, University of British Columbia, Vancouver, Canada for kindly providing the $(GB1)_8$ plasmid. We also thank Ahmad Al-Husamiyah for performing experimental work.

# Chapter 6

# Worm-Like Ising Model for Protein Mechanical Unfolding under the Effect of Osmolytes

Adapted from Aioanei D, Brucale M, Tessari I, Bubacco L, Samorì B. “Worm-like Ising model for protein mechanical unfolding under the effect of osmolytes”, Biophys J. 2012 Jan 18;102(2):342-50. Copyright (2012) Biophysical Society. Permission for reproduction requested but not yet granted by the completion time of this thesis.

## 6.1 Abstract

We show via single-molecule mechanical unfolding experiments that the osmolyte glycerol stabilizes the native state of the human cardiac I27 titin module against unfolding without shifting its unfolding transition state on the mechanical reaction coordinate. Taken together with similar findings on the immunoglobulin-binding domain of streptococcal protein G (GB1), these experimental results suggest that osmolytes act on proteins through a common mechanism that does not entail a shift of their unfolding transition state.

We investigate the above common mechanism via an Ising-like model for protein mechanical unfolding which adds Worm-like chain behavior to a recent generalization of the Wako-Saitô-Muñoz-Eaton (WSME) model with support for group-

transfer free energies. The thermodynamics of the model are exactly solvable, while protein kinetics under mechanical tension can be simulated via Monte-Carlo algorithms. Notably, our force-clamp and velocity-clamp simulations exhibit no shift in the position of the unfolding transition state of GB1 and I27 under the effect of various osmolytes.

The excellent agreement between experiment and simulation strongly suggests that osmolytes do not assume a structural role at the mechanical unfolding transition state of proteins, acting instead by adjusting the solvent quality for the protein chain analyte.

## 6.2 Introduction

Throughout the course of evolution, nature has successfully modulated protein stability using organic osmolytes, which are small molecules that shift the native-unfolded thermodynamic balance by changing the *solvent quality* for the protein chain. This thermodynamic description is commonly referred to as the *osmolyte effect* [85]. Even though the thermodynamic model of the osmolyte effect is now widely accepted, the molecular details of the mechanism by which osmolytes influence protein kinetics and transition states are still not completely understood.

Mechanistic information about the role of osmolytes in protein folding and unfolding processes may be obtained by projecting protein transition states onto a geometrically relevant reaction coordinate. Single-Molecule Force Spectroscopy (SMFS) has recently become the technique of choice for geometrical mapping of protein energy landscapes [3]. From protein mechanical unfolding experiments, SMFS readily provides not only the distance between the native state and the mechanical unfolding transition state, commoly referred to as the *unfolding distance* $\Delta x_u$, but also an estimate of the *spontaneous unfolding rate* $k_u(0)$ of the protein [17].

A previous SMFS study of mechanical unfolding of protein ubiquitin reported that the presence of glycerol as a cosolvent in aqueous solution leads to an increase of the protein's unfolding distance [76]. Other SMFS studies reported that glycerol [15], ethylene glycol and propylene glycol [16] increase the unfolding distance of the I27 titin module of the human muscle. Based on the ansatz that the unfolding distance of proteins may be determined by the bridging length of solvent molecules at the unfolding transition state [60, 156], the mentioned SMFS studies concluded that small osmolyte molecules bridge the critical beta strands of proteins under mechanical tension, leaving their distinct signature on their unfolding distance. In particular, since osmolyte molecules are larger than water molecules, the small osmolytes were reported to increase the unfolding distance of proteins by amounts that correlate with their molecular size. It should be noted however that larger osmolytes such as sorbitol and sucrose were instead found to leave the unfolding distance of I27 unchanged, indicating their inability to partake in solvent bridging [16].

We challenge the above view that small osmolytes increase the unfolding distance of proteins and bridge their critical beta-strands in the unfolding transition state through experimental and theoretical considerations.

1. Experimentally, we show herein that, contrary to what was reported in Ref. [15], glycerol does not change the unfolding distance of I27. It should be noted that recent studies by us and others have also shown that the small osmolytes dimethylsulfoxide (DMSO) [12], glycerol [19] and guanidinium chloride (GndCl) [63] do not change the unfolding distance of GB1. Taken together, these experimental results challenge the view that small osmolytes increase the unfolding distance of proteins.

2. Theoretically, we have recently shown that an Ising-like model with support for the osmolyte effect does not exhibit any movement of the unfolding transition state of GB1 in the presence of glycerol 30% v/v, when projected thermodynamically onto a commonly-used non-mechanical reaction coordinate [19]. Herein we show that the same lack of movement holds for various concentrations of DMSO and GndCl, indicating that osmolytes may not generally produce significant movements of the unfolding transition states of proteins. Furthermore, we expand the mentioned thermodynamic model with mechanical unfolding kinetics based on the worm-like chain (WLC) force-distance relation [58, 59], showing that our Ising-like model exhibits no significant change in the unfolding distance of GB1 and I27 under the effect of various concentrations of DMSO, glycerol and GndCl, in excellent agreement with the experimental results presented herein and those of Ref. [12, 19, 63]. Note that our Ising-like model lacks the expressive power to

account for any possible structural role of osmolytes at the unfolding transition state of proteins, while still being able to explain the mentioned experimental data. Therefore our mechanical unfolding simulations challenge the view that osmolyte-bridging may be a general phenomenon in nature, since a structural role of osmolytes at the unfolding transition state is not necessary to explain the aforementioned experimental results.

In the next two sections we give a general overview of the protein mechanical unfolding model we propose together with a short overview of how the model supports the osmolyte effect.

### 6.2.1 Ising-like protein models with exactly solvable thermodynamics

A recent model for protein mechanical unfolding [126, 127, 157], which we shall refer to as the IPZ model (based on the initials of the authors), was built as a generalization of the Ising-like Wako-Saitô-Muñoz-Eaton (WSME) protein model [122, 123, 124, 158, 125, 159]. Like the WSME model that it extends, the IPZ model has exactly solvable thermodynamics [128, 160] and it has been employed to investigate protein folding/unfolding kinetics and trajectories by simulating protein refolding under force-clamp conditions [130] and protein mechanical unfolding either in the force-clamp [131, 130, 127], force-ramp [126, 127] or velocity-clamp [131, 161] modes.

The IPZ model takes into account the entropic elasticity of the protein chain by allowing every WSME state to behave similarly to a freely-jointed chain (FJC) with the angle between consecutive segments being taken from a finite set, usually being either zero or $\pi$ radians [127]. Such an approach comes however with several limitations:

1. The experimentally observed behavior of unfolded protein chains does not show significant deviations from the ideal behavior of a worm-like chain under mechanical tension. Indeed, lock-in force spectroscopy with a resolution of 400 fN has failed to find any such deviations down to a force of 1.7 pN [162]. Therefore the WLC force-distance formula is extensively used in the analysis of velocity-clamp experiments performed with either the atomic-force

microscope (AFM) [72] or optical-tweezers [163, 164]. Rather than aiming at approximating the WLC entropic elasticity, the IPZ model [126, 127, 157] tries to approximate the FJC entropic elasticity, which differs significantly at high forces [3].

2. The IPZ model has more states than the WSME model, which may make it more difficult to obtain statistically representative sets of trajectories during Monte-Carlo simulations, or to compute exactly the kinetics of small protein domains [124, 158]. Indeed, for a protein with $N$ peptide bonds, we computed that the IPZ model has $2 \times 3^N$ states when only two possible angle values are allowed, while the WSME model has the lower number of $2^N$ states. It should be noted however that this shortcoming may be somewhat mitigated by the fact that the partition sum over the extra microscopic degrees of freedom introduced by the IPZ model (relative to the WSME model) can be computed analytically [127], and through the local equilibrium approach for kinetic simulations [157, 165].

We address both limitations by introducing an Ising model for proteins under mechanical tension that dictates that for every WSME state, the end-to-end extension of the protein chain is given by the WLC average force-distance relation (see Section 6.3.4). For any fixed stretching force, the number of states of the model is the same as that of the WSME model. Shortly, a protein state consists of a set of contiguous native stretches of the peptide chain, contributing to the contour length with the distance in the native structure of the protein between the start and end of the stretch. Our worm-like Ising protein model maintains exactly solvable thermodynamics and, in the absence of a stretching force, it reduces trivially to the WSME-like model it extends.

### 6.2.2 Incorporating the osmolyte effect into the protein model

The osmolyte effect has been dissected into groupwise free energy contributions, with the protein backbone making up most of the protein's free energy of transfer to osmolyte-containing solutions [85]. Making use of the group transfer free energies (GTFEs) of amino acid backbone units and side-chains, a recent extension of the

WSME model for the osmolyte effect enabled the thermodynamic projection of the protein energy landscape in the presence of osmolytes onto reaction coordinates commonly employed to monitor protein folding and unfolding [19].

However, a purely thermodynamic theory cannot directly explain the results of SMFS experiments in osmolyte-containing solutions, for which a kinetic theory of protein mechanical unfolding in the presence of osmolytes is needed instead. In this work we address the effect of osmolytes on GB1 (see Fig. (6.1a)) and I27 (see Fig. (6.1b)) both at the thermodynamic and kinetic levels. Thermodynamically, we extend the previous equilibrium analysis of the osmolyte effect projected onto the reaction coordinate represented by the *weighted number of native contacts* [19] to a larger range of osmolytes and osmolyte concentrations for GB1, and we also apply it to I27 in the presence of glycerol 30% v/v. Kinetically, we use our worm-like Ising model we propose in Section 6.3.4 to investigate protein mechanical unfolding in the presence of osmolytes by simulating the mechanical unfolding of GB1 and I27 under force-clamp and velocity-clamp conditions. We compare the mechanical unfolding kinetics as predicted by the worm-like Ising protein model in the presence of varying concentrations of osmolytes, namely DMSO, glycerol and GndCl, to experimental SMFS data obtained in similar solvent conditions.

# 6.3 The model

## 6.3.1 The original WSME formulation

According to the WSME model [122, 123, 124, 125, 128], a protein $N+1$ amino acids long is described as a chain of $N$ peptide bonds. The state of the protein is captured by the vector $\vec{m}$, where the $i$-th peptide bond is represented by the binary variable $m_i$ which has only two possible values: 1 for a *native* peptide bond, and 0 for a *non-native* peptide bond. The effective free energy of the system reads as

$$H_{WSME}(\vec{m}) = 1K \times k_B \epsilon \sum_{1 \le i < j \le N} h_{ij} \prod_{k=i}^{j} m_k - k_B T \sum_{i=1}^{N} q_i (1 - m_i)\,, \tag{6.1}$$

where 1K is one Kelvin, $k_B$ is Boltzmann's constant, $T$ is the temperature, $\epsilon$ is a dimensionless enthalpic scale and $h_{ij} \le 0$ are dimensionless numbers representing

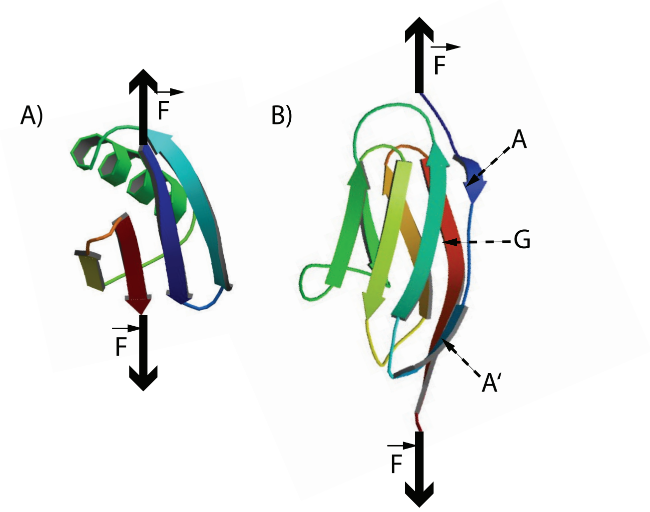


Figure 6.1: (Color online) a) The native structure of GB1 (Protein Data Bank (PDB) [136] code 1PGA [137]). b) The native structure of I27(PDB code 1TIT [166]) showing the A, A' and G strands.

the relative strength of the contact between the $i$-th and the $j+1$-th amino acids and $q_i > 0$ represents the entropic cost of ordering bond $i$. The contact strength $h_{ij}$ is commonly defined as 0 if $j = i + 1$, and $-\lceil c_{ij}/5 \rceil$ if $j > i + 1$ where $c_{ij}$ is the number of pairs of atoms, the first atom of the pair belonging to amino acid $i$ and the second atom of the pair belonging to amino acid $j + 1$, that in the native state are closer than 0.4 nm and $\lceil \cdot \rceil$ is the ceiling function [126, 127]. The thermodynamics of the WSME model can be exactly computed via the *transfer-matrix* approach [128] (see also Section 6.3.3 for zero osmolyte concentration).

The fraction of folded molecules is estimated as

$$p_f(T) = [\alpha(T) - \alpha(\infty)]/[\alpha(0) - \alpha(\infty)]\,, \tag{6.2}$$

where $\alpha(T)$ is the thermodynamic average of the number of native peptide bonds at temperature $T$. At zero temperature we have $\alpha(0) = N$, while at infinite temperature the exact expression for the mentioned thermodynamic average[129] reads as:

$$\alpha(\infty) = \sum_{i=1}^{N} [1 + \exp(q_i)]^{-1}\,. \tag{6.3}$$

Assuming the entropic costs $q_i$ are known, the parameter $\epsilon$ in Eq. (6.1) can be fitted by imposing the known value of the folded fraction $p_f(T)$ at a certain temperature.

### 6.3.2 Adding the osmolyte effect to the WSME model

Tanford's Transfer Model has proved very successful at predicting osmolyte-induced energetics of protein stability [109]. In particular, the procedure of scaling group transfer free energies (GTFEs) by the accessible surface area has previously proved accurate at interpreting the thermodynamics of the osmolyte effect [132, 133, 109]. We describe next how the osmolyte effect has been incorporated into the WSME model by applying the Transfer Model to every WSME state [19].

A *native stretch* is defined as a sequence of consecutive amino acids connected by native bonds and delimited by two non-native bonds. More formally, the stretch delimited by bonds $i$ and $j$ is said to be native if $S_{ij} = 1$ where

$$S_{ij} = (1 - m_i)(1 - m_j) \prod_{k=i+1}^{j-1} m_k \ , \tag{6.4}$$

with $0 \leq i < j \leq N + 1$ and taking $m_0 = m_{N+1} = 0$. In the limiting case of $j - i = 1$ the stretch reduces to one amino acid [126, 127].

Accordingly, we consider each native stretch to have the exact same structure as in the native state of the full protein, and we add to the effective free energy of Eq. (6.1) a term representing the free energy of transfer $M_{ij}$ of the native stretch delimited by peptide bonds $i$ and $j$, in isolation, relative to the denatured state:

$$H_O(\vec{m}) = H_{WSME} + \sum_{0 \leq i \leq j-2 \leq N-1} S_{ij} M_{ij} \ , \tag{6.5}$$

with $M_{ij}$ taken as

$$M_{ij} = \sum_{i<k\leq j} \left[ \Delta g^{ref}_{bb,R[k]} (A^{den}_{bb,k} - A^{ij}_{bb,k}) / A^{ref}_{bb,R[k]} + \Delta g^{ref}_{sc,R[k]} (A^{den}_{sc,k} - A^{ij}_{sc,k}) / A^{ref}_{sc,R[k]} \right] \ , \tag{6.6}$$

where $R[k]$ denotes the amino acid at position $k$ of the protein chain, $g^{ref}_{bb,R[k]}$ is the free energy of transfer of amino acid $R[k]$ in the *reference state* (which is determined by the type of experimental or theoretical data available - see Section 6.4.1), $A^{den}_{bb,k}$ is the accessible surface area (ASA) of the backbone unit at position $k$ of the protein chain in the denatured state, $A^{ij}_{bb,k}$ is the accessible surface area of the backbone unit at position $k$ of the protein chain in the isolated native stretch from peptide bond $i$ to peptide bond $j$, $A^{ref}_{bb,R[k]}$ is the accessible surface area of the backbone of amino acid $R[k]$ in the reference state, and finally $g^{ref}_{sc,R[k]}$, $A^{den}_{sc,k}$, $A^{ij}_{sc,k}$, $A^{ref}_{sc,R[k]}$ are defined similarly to the last four mentioned quantities, with the only difference that they refer to side chains rather than the backbone.

### 6.3.3 Solving the thermodynamics of the WSME model and its extension for the osmolyte effect

It can be seen that compared to Eq. (6.1) of the original WSME model, the effective free energy expression of Eq. (6.5) has one extra term for each native stretch. Because no terms are introduced that simultaneously involve peptide bonds from multiple native stretches, the transfer-matrix exact solution to the thermodynamics of the WSME model [128] is equally applicable when the osmolyte effect is incorporated. In this section we reiterate the equations to exactly compute, for two common order parameters, the free energy landscape of the WSME model extended with the osmolyte effect[19] described in Section 6.3.2, and of the original WSME model described in Section 6.3.1, which can be seen as the special case of zero osmolyte concentration.

Let us introduce the notation

$$Q_{ij} = 1K \times k_B\epsilon \sum_{i<k<l<j} h_{kl} + k_BT \sum_{k=i+1}^{j-1} q_k + M_{ij} \,, \tag{6.7}$$

for $0 \leq i < j \leq N+1$.

The two order parameters we consider are the number of native peptide bonds

$$\phi(\vec{m}) = \sum_{i=1}^{N} m_i \tag{6.8}$$

and the weighted number of native contacts [131]:

$$\theta(\vec{m}) = -\sum_{1\leq i<j\leq N} h_{ij} \prod_{k=i}^{j} m_k \,. \tag{6.9}$$

We denote generically by $\vec{g}$ a two-dimensional vector such that $g_{ij}$ is either of the two mentioned order parameters of the native stretch delimited by peptide bonds $i$ and $j$ with $0 \leq i < j \leq N+1$, namely the number of native peptide bonds $(j-i-1)$ or the weighted number of native contacts $(-\sum_{i<k<l<j} h_{kl})$.

Then the effective free energy $E_p$ restricted to the states having the relevant order parameter ($\phi$ or $\theta$) equal to $p$ (more formally, $\sum_{0\leq k<l\leq N+1} S_{kl}g_{kl} = p$) is computed as

$$E_p(\vec{g}) = R_{Np}(\vec{g}) - k_B T \sum_{k=1}^{N} q_k \tag{6.10}$$

for $0 \leq p \leq g_{0,N+1}$ where the quantity $R_{Np}(\vec{g})$ is computed in polynomial time via the recursion

$$\begin{aligned} exp[-R_{ip}(\vec{g})/(k_B T)] &= \delta_{p,g_{0,i+1}} exp[-Q_{0,i+1}/(k_B T)] \\ + \sum_{(1\leq j\leq i)\wedge(g_{j,i+1}\leq p\leq g_{0,j}+g_{j,i+1})} & exp[-R_{j-1,p-g_{j,i+1}}(\vec{g})/(k_B T) - Q_{j,i+1}/(k_B T)]\,, \end{aligned} \tag{6.11}$$

with $\delta_{\cdot,\cdot}$ being the Kronecker delta symbol.

The average number of native peptide bonds in the absence of force, which is needed to fit the parameter $\epsilon$, follows trivially once the free energy landscape is computed via Eq. (6.10) for the number of native peptide bonds.

### 6.3.4 The worm-like Ising model for proteins under mechanical tension

Let us represent the amino acid $k$ by its nitrogen, alpha-carbon and carbon of the carbonyl group as a three-long $N_k - C_{\alpha,k} - C_k$ sequence. For $0 \leq i < j-1 \leq N$ we define the *length* $l_{ij}$ of the native stretch delimited by peptide bonds $i$ and $j$ as the distance between the midpoint of the $C_i$ and $N_{i+1}$ atoms and the midpoint of the $C_j$ and $N_{j+1}$ atoms, making the convention that $C_0 = N_1$ and $N_{N+2} = C_{N+1}$ [126, 127]. Although the extensibility of native stretches is not taken into account[130], in order to get better agreement with the experimental contour-length increments upon unfolding of individual protein modules in characteristic velocity-clamp sawtooth patterns, we fix $l_{i,i+1} = 0.4$ nm [167, 48].

We define the *contour length* of a WSME state $\vec{m}$ as the sum of the lengths of its native stretches

$$L(\vec{m}) = \sum_{0\leq i<j\leq N+1} S_{ij} l_{ij}(\vec{m})\,. \tag{6.12}$$

In order to introduce mechanical tension into the model we add to the effective free energy $H_O$ (or $H_{WSME}$ in the absence of osmolytes) a potential energy function that depends on the end-to-end extension $x$. For a given contour length $L$ and an acting force $f$, we fix the end-to-end extension $x$ according to the following WLC force-distance interpolation relation [59, 58]:

$$F_{WLC}\left(\frac{x}{L}\right) = k_B T (4p_l)^{-1}\left[\left(1-\frac{x}{L}\right)^{-2} + 4\frac{x}{L} - 1\right] , \tag{6.13}$$

where $p_l$ is the persistence length, assumed to be constant all throughout. Therefore the end-to-end extension becomes $x(L, f) = F_{WLC}^{-1}(f)L$.

The force-clamp potential energy takes the form $V(L, f) = -\xi x(L, f)f = -\xi F_{WLC}^{-1}(f)Lf$, where $\xi$ is a dimensionless scaling factor that is computed by imposing that the folded fraction has a known value at a given temperature and force combination.

The velocity-clamp potential energy takes the form $V(L,t) = \xi\frac{\kappa}{2}[vt - x(\kappa, L, vt)]^2$, where $v$ is the constant velocity of the cantilever, $\kappa$ is the cantilever spring constant and $x(\kappa, L, z)$ is the unique root in the interval $[0, L) \cap [0, z]$ of the equation

$$F_{WLC}(x(\kappa, L, z)/L) = \kappa(z - x(\kappa, L, z))\,. \tag{6.14}$$

### 6.3.5 Solving the thermodynamics under the effect of force

In this section we present the equations that allow one to compute in polynomial time the effective free energy as a function of the number of native peptide bonds $\phi$, the weighted number of native contacts $\theta$, and contour length $L$ in the presence of a stretching force $f$. Herein we reuse the definitions of $Q_{ij}$ (Eq. (6.7)) and $\vec{g}$ from Section 6.3.3.

The key to polynomial time computation of force-dependent effective free energies lies in the discretization of the possible lengths of native stretches $l_{ij}$. As in Ref. [127], we assume that a length scale is chosen such that the size of the set of all protein subchain contour lengths grows only linearly with increasing protein size.

It is useful to consider the maximum contour length $D_i$ that the protein chain made up of the first $i$ peptide bonds can take, $0 \le i \le N$. Note that fully denaturing the protein chain is not guaranteed to lead to the maximum contour length because of two reasons. One is due to rounding of the native stretch lengths, which in general is not guaranteed to preserve the triangle inequality. The other reason is that we assign to $l_{i,i+1}$ a value not dictated by the three-dimensional structure of the protein (see Section 6.3.4), which again may, in principle, break the triangle inequality. The maximum contour length of the first $i$ peptide bonds can be computed recursively as follows:

$$D_i = \max_{0\leq j\leq i}[(1-\delta_{j,0})D_{j-1} + l_{j,i+1}]\,. \tag{6.15}$$

The effective free energy as a function of contour length $L$ and order parameter $p$ (either the number of native peptide bonds $\phi$ or the weighted number of native contacts $\theta$) in the presence of a stretching force $f$ can be written as

$$E_{L,p}(\vec{g},f) = U_{N,L,p} - k_BT\sum_{i=1}^{N} q_i - \xi F_{WLC}^{-1}(f)fL \tag{6.16}$$

for $0 \leq L \leq D_N$ and $0 \leq p \leq g_{0,N+1}$, where $U_{N,L,p}$ is independent of force and it is computed recursively:

$$\begin{aligned} exp[-U_{iLp}(\vec{g})/(k_BT)] &= \delta_{L,l_{0,i+1}}\delta_{p,g_{0,i+1}}exp[-Q_{0,i+1}/(k_BT)] \\ &+ \sum_{(1\leq j\leq i)\wedge(l_{j,i+1}\leq L\leq D_{j-1}+l_{j,i+1})\wedge(g_{j,i+1}\leq p\leq g_{0,j}+g_{j,i+1})} \\ &exp[-U_{j-1,L-l_{j,i+1},p-g_{j,i+1}}(\vec{g})/(k_BT) - Q_{j,i+1}/(k_BT)]\,. \end{aligned} \tag{6.17}$$

The average number of native peptide bonds in the presence of force, which is needed to fit the parameter $\xi$, follows trivially once the free energy landscape is computed via Eq. (6.16) for the number of native peptide bonds.

It should be noted that while Eq. (6.16) can be summed at zero force over the possible $L$ values to yield the same thermodynamic quantities like Eq. (6.10), the latter equation, specialized for the absence of force, is simply faster to evaluate in practice, when applicable.

### 6.3.6 Continuous-time Markov chain approach to kinetics

A necessary condition for Monte Carlo simulations is that the associated kinetics satisfy the *balance* condition, i.e., that they leave the Boltzmann distribution invariant [168]. Monte Carlo simulations are commonly designed to satisfy an even stronger condition called *detailed balance* [169] that implies reversibility of all transitions. We adopt the latter approach and define a continuous-time Markov chain by allowing single-bond flip transitions[158]. More formally, we allow transitions $\vec{m}^t \to \vec{m}^h$ from a tail state $\vec{m}^t$ to a head state $\vec{m}^h$ whenever the two states differ by the native status of exactly one peptide bond. By assigning transition rates according to a detail-balance preserving prescription that depends only on the difference of the energetic levels of the tail and head states expressed in $k_B T$ units, e.g. the exponential, Metropolis or Glauber transition rate expressions [170], the continuous-time Markov chain thus defined is in detailed balance with the Boltzmann distribution of states. Hereafter we denote the transition rate from tail state $\vec{m}^t$ to head state $\vec{m}^h$ by the notation

$$\begin{aligned} W(\vec{m}^t, \vec{m}^h) &= W\left([H_O(\vec{m}^h) - H_O(\vec{m}^t)]/(k_B T)\right) \\ &= \tau^{-1} W_0\left([H_O(\vec{m}^h) - H_O(\vec{m}^t)]/(k_B T)\right) \\ &= \tau^{-1} W_0(\vec{m}^t, \vec{m}^h)\,, \end{aligned} \tag{6.18}$$

where $\tau$ is the microscopic time scale and $W_0$ is a dimensionless transition rate function that does not depend on $\tau$. More details about the transition rate prescription chosen, the starting and ending states of the simulations and the architecture of the protein constructs simulated in this study can be found in Section 6.4.4 .

Since the number of outgoing transitions from each state is as small as $N$, for the simulations it is convenient to use the kinetic Monte Carlo algorithm [171]. According to this algorithm each transition requires the generation of two uniform random numbers in the interval $(0, 1]$, one used to select the time to the next transition and the other one used to select one of the outgoing transitions from the current state (see Section 6.4.3).

## 6.4 Materials and methods

### 6.4.1 GTFEs and ASAs

In order to apply Eq. (6.6) at different DMSO concentrations we linearly interpolated the experimental transfer free energies of the full amino acids Glycine, Alanine, Leucine and Tryptophan, computed from the logarithm of solubility ratios reported in Table 7 of Ref. [172]. The transfer free energies at DMSO concentrations of 50%, 40%, 30%, 20% and 10%, respectively, computed as (i) 2.81, 2.21, 1.34, 0.89 and 0.35 $k_BT$, respectively, for Glycine; (ii) 2.20, 1.75, 1.12, 0.73 and 0.30 $k_BT$, respectively, for Alanine; (iii) 1.90, 1.51, 0.97, 0.62 and 0.26 $k_BT$, respectively, for Leucine; and (iv) -0.52, -0.41, -0.38, -0.34 and -0.19 $k_BT$, respectively, for Tryptophan. Free energies of transfer for the side-chains of the above mentioned amino-acids $g^{ref}_{sc,R[k]}$ were estimated by subtracting the free energy of transfer of Glycine, which is denoted by $g^{ref}_{bb,R[k]}$ in Eq. (6.6), and set to zero for all the other sidechains. As reference state we took the full amino acids, which is compatible with the type of solubility experimental data of Ref. [172]. The amino acid structures were created and optimized using PRODRG2 [173]. On the optimized structures we computed accessible surface areas for the backbone unit $A^{ref}_{bb,R[k]}$ and side chains $A^{ref}_{sc,R[k]}$ of the amino acids using PyMOL [174] with a probe radius of 0.14 nm.

For glycerol we applied the statistical mechanics model of Ref. [1] to estimate the transfer free energy of a backbone unit to the concentration of 30% v/v glycerol, obtaining the value of 0.42 $k_BT$ [19]. In this case the reference state was the backbone unit from the Gly-X-Gly tripeptide, with accessible surface areas taken from Table 2 in the Supplementary Information of Ref. [109]. Side chains were not included in the glycerol calculations.

For GndCl we applied the same statistical mechanics model of Ref. [1] to estimate the transfer free energy of the backbone unit at GndCl concentrations of 3.04, 2.25 and 1 M, obtaining the values -0.26, -0.20 and -0.10 $k_BT$, respectively. As for glycerol, the reference state was taken to be the backbone unit from the Gly-X-Gly tripeptide, and side-chains were ignored.

For urea we used again the statistical mechanics model of Ref. [1] for the backbone unit of the Gly-X-Gly tripeptide, taken as the reference state. The transfer free energy of the backbone unit in 4 M urea computed as -0.26 $k_BT$ and we ignored all sidechain contributions.

We computed the ASA of the native stretches split between backbone and sidechain atoms using PyMOL with a probe radius of 0.14 nm, while the ASA of the denatured state of GB1 and I27 was estimated using ProtSA [134, 135].

### 6.4.2 Fitting the model

We extracted the native structure of GB1 (see Fig. (6.1a)) and I27 (see Fig. (6.1b)) from the Protein Data Bank [136]. We removed the hetero-atoms from the structure after which we added the missing hydrogen atoms and flipped residue 8 of GB1 and residue 31 of I27 using MolProbity [138]. Subsequently the MolProbity optimized structures were used for all calculations involving the structure of the native stretches of GB1 and I27.

We adopted the convention that all peptide bonds have the same entropic cost for being in the native state, i.e., $q_i = q$ for all $1 \leq i \leq N$, which is a commonly adopted assumption with WSME-like models [127, 126, 19]. For any given value of $q$, the parameter $\epsilon$ can be obtained by imposing the value of the unfolding temperature onto the model.

As the unfolding temperature of GB1 we used the value 350.15 K obtained from a computational study at pH 7 [175]. Note that this unfolding temperature is higher than the one used in a previous WSME study of GB1 [19], which in turn was the experimental unfolding temperature of a less thermally stable GB1 mutant [139]. We performed a space search of $q$ to find the value that gives a folded fraction of GB1 of 1/2 in the presence of 4 M urea at a temperature of 295.15 K [98]. The simultaneous fit of $q$ and $\epsilon$ led to the approximate values $q = 0.59$ and $\epsilon = 37.19$.

From the previously characterized folding and unfolding kinetics of GB1 under mechanical tension at 301.15 K [12, 18], we estimated the unfolding force of GB1 to be $F_u = k_B T \ln[k_f(0)/k_u(0)]/(\Delta x_f + \Delta x_u) = 14.55$ pN where $k_f(0)$ is the spontaneous folding rate, $k_u(0)$ is the spontaneous unfolding rate, $\Delta x_f$ is the folding distance and $\Delta x_u$ is the unfolding distance. Imposing the above mentioned unfolding force and discretizing the native stretch contour lengths with a resolution of 0.01 nm we obtained the approximate value $\xi = 0.34$.

Similarly, we determined the parameters $q$, $\epsilon$ and $\xi$ for I27 by using the unfolding temperature of 344.95 K (obtained by digitizing Fig. (6) of Ref. [176] for pH 7.0), the denaturing concentration of 3.04 M GndCl at 298.15 K [177] and the unfolding force of 13.7 pN [178] at 298.15 K. We obtained the approximate values $q = 0.64$, $\epsilon = 41.76$ and $\xi = 0.35$.

### 6.4.3 Kinetic Monte-Carlo algorithm

We show next the formulae used to implement the kinetic Monte Carlo algorithm in the force-clamp and velocity-clamp pulling modes. In what follows, let us denote the current state by $\vec{m^t}$, the $N$ possible head states by $\vec{m^{h_1}}, \cdots, \vec{m^{h_N}}$ (the $N$ head states can be taken in any order) and let $u_1$ and $u_2$ be two uniform random numbers in the interval $(0, 1]$.

#### Force-clamp simulations

For the force-clamp case, the time $\Delta t$ before the next transition happens is given by the relation

$$\Delta t/\tau = -(\ln u_1)/\left[\sum_{i=1}^{N} W_0(\vec{m^t}, \vec{m^{h_i}})\right] , \tag{6.19}$$

and the transition is performed to head state $\vec{m^{h_h}}$ with

$$h = \max_{1\le i\le N}\left\{ i \left| \sum_{j=1}^{i-1} W_0(\vec{m^t}, \vec{m^{h_j}})/\sum_{j=1}^{N} W_0(\vec{m^t}, \vec{m^{h_j}}) < u_2 \right.\right\} . \tag{6.20}$$

It becomes clear from Eq. (6.19) and Eq. (6.20) that it is possible to perform force-clamp simulations without knowledge of the microscopic time scale $\tau$, as long as it is understood that after the simulation the transition times obtained $(\Delta t/\tau)$ need to be multiplied by $\tau$ for comparison with experimental data.

#### Velocity-clamp simulations

Let us assume that the current time is $t_s$ and that we need to compute the time to the next transition $\Delta t$, and then we need to find out which of the $N$ possible transitions is taken. In what follows we use the notation

$$\begin{aligned} W_0(\vec{m^t}, \vec{m^h}, y) = \;& W_0\Big(\{H_O(\vec{m^h}) - H_O(\vec{m^t}) + \xi\kappa/2[x(\kappa, L(\vec{m^h}), v\tau y)^2 - x(\kappa, L(\vec{m^t}), v\tau y)^2] \\ & -\xi\kappa v\tau y[x(\kappa, L(\vec{m^h}), v\tau y) - x(\kappa, L(\vec{m^t}), v\tau y)]\}/(k_BT)\Big) , \end{aligned} \tag{6.21}$$

where $x(\kappa, L, z)$ is the relevant root of Eq. (6.14) and $y \ge 0$ is a dimensionless quantity.

The time to the next transition $\Delta t$ is given by the unique solution of the equation:

$$\ln u_1 = \quad -\sum_{i=1}^{N} \int_{t_s/\tau}^{t_s/\tau+\Delta t/\tau} W_0(\vec{m^t}, \vec{m^{h_i}}, y)dy \,, \tag{6.22}$$

and the transition is performed to head state $\vec{m^{h_h}}$ with

$$h = \max_{1\leq i\leq N} \left\{ i \left| \sum_{j=1}^{i-1} W_0(\vec{m^t}, \vec{m^{h_j}}, t_s/\tau + \Delta t/\tau) \middle/ \sum_{j=1}^{N} W_0(\vec{m^t}, \vec{m^{h_j}}, t_s/\tau + \Delta t/\tau) < u_2 \right. \right\} . \tag{6.23}$$

It becomes clear from Eq. (6.22) and Eq. (6.23) that while knowledge of $\tau$ is not strictly necessary to perform velocity-clamp simulations, knowledge of the product $v\tau$ is required. Moreover, after the simulation the transition times obtained $(\Delta t/\tau)$ need to be multiplied by the microscopic time scale for comparison with experimental data.

Note that while Eq. (6.22) is exact, solving it numerically requires repeated integration of an integrand that needs to solve numerically Eq. (6.14), and therefore becomes too slow in practice for low pulling velocities. The performance issue is circumvented by observing that at least for the exponential, Metropolis and Glauber transition rate expressions, $\int W_0(ay + b)\, dy$ has an analytical form (e.g., Eq. (6.25)). Therefore we can solve Eq. (6.22) by performing the integration on small steps of size $\Delta f/(\kappa v\tau)$, with $\Delta f$ being the desired force resolution, in each step assuming that $x(\kappa, L, z)$ is constant and its value is given by solving numerically Eq. (6.14) at the beginning of the step, adding steps until $\ln u_1$ is exceeded. When that happens, the transition time is found numerically in the interval of the last added step, assuming again that $x(\kappa, L, z)$ is constant and equal to its exact value at the beginning of that last step.

### 6.4.4 Mechanical unfolding simulations

Throughout all simulations we used the exponential transition rate prescription

$$W_0(\vec{m^t}, \vec{m^h}) = W_0\left([H_O(\vec{m^h}) - H_O(\vec{m^t})]/(k_BT)\right) = \exp\left(-[H_O(\vec{m^h}) - H_O(\vec{m^t})]/(k_BT)/2\right) . \tag{6.24}$$

The exponential transition rate, when composed with a linear function (see

Section 6.4.3), integrates as

$$\int W_0(ay+b)\,dy = -2W_0(ay+b)/a\,. \tag{6.25}$$

The force-dependent unfolding rate $k_u(f)$ was estimated as the inverse of the mean unfolding time over multiple trajectories, according to the formula $\langle t(F)\rangle = \int_0^\infty t k_u(F)\exp[-tk_u(F)]\,dt = 1/k_u(F)$. Then for each solvent condition separately, in order to estimate the two unfolding kinetic parameters, namely the spontaneous unfolding rate $k_u(0)$ and unfolding distance $\Delta x_u$, we fitted Bell's equation [37, 82] on the full range for forces:

$$\ln k_u(F) = \ln k_u(0) + F\Delta x_u/(k_B T)\,. \tag{6.26}$$

Some previous mechanical unfolding with WSME-like models considered the protein to be unfolded as soon as its extension reaches either half [126, 127, 131] or two thirds [131] of its maximum value. However WSME-like models tend to be less cooperative than the real proteins they describe [129], and their low cooperativity becomes most evident in velocity-clamp simulations where it results in multiple force-distance peaks for each module (not shown). Therefore we considered a protein domain to be unfolded as soon as all contacts between the force-bearing beta strands were lost (see Ref. [161] for a similar approach).

Using the procedure described in Section 6.3.1, the following contacts were identified between the two force-bearing terminal beta strands of GB1 [179, 180, 49]: (1, 49), (3, 44), (3, 49), (3, 51), (4, 49), (4, 50), (4, 51), (5, 51), (6, 50), (6, 51), (6, 52), (6, 53), (7, 53), (8, 53), (8, 54), (8, 55), (9, 55), (10, 55).

For the GB1 force-clamp mechanical unfolding simulations we used as a starting state one GB1 domain fully in the native state ($m_i = 1$ for all $1 \le i \le N$) and we regarded a trajectory as finished as soon as the protein unfolded.

For the GB1 velocity-clamp simulations we obtained the starting state by linking together 8 (eight) GB1 modules using 7 (seven) linker peptide bonds of length 0.4 nm, construct that we shall refer to as $(GB1)_8$. The linker peptide bonds were fixed in the non-native state all throughout the simulations. The starting state was constructed by putting in the native state all GB1 (non-linker) peptide bonds. As soon as a GB1 module unfolded, all its peptide bonds were fixed into the non-native state. A velocity-clamp trajectory was ended as soon as all the GB1 modules reached the unfolded state.

Since I27 unfolds mechanically through a metastable intermediate which has strand A detached [181], we fixed the state of the peptide bond 10 to denatured all throughout the simulations so as to avoid any contacts between strands A and G. Therefore our simulations of I27 describe the unfolding of the intermediate state, the same process that is most commonly observed in AFM experiments with I27 [61]. Using the procedure described in Section 6.3.1, the following contacts were identified between the force-bearing beta strands A' and G of I27's metastable intermediate: (11, 82), (11, 83), (11, 84), (12, 83), (12, 84), (12, 86), (13, 83), (13, 84), (13, 85), (13, 86), (14, 86), (14, 87), (14, 88), (15, 85), (15, 86), (15, 87).

For both the force-clamp and velocity-clamp simulations of I27 we used as a starting state one I27 domain with all peptide bonds except the $10^{th}$ one in the native state ($m_i = 1 - \delta_{i,10}$ for all $1 \leq i \leq N$) and we regarded a trajectory as finished as soon as the protein unfolded, i.e., when all its critical contacts were lost.

The force-step size used for the performance optimization described in Section 6.4.3 for velocity-clamp simulations was $\Delta f = 1$ pN. For each unfolding event we recorded (i) the contour length just before unfolding, which was taken as an approximation to be the contour length for all the stretching time range since the previous unfolded event, if any, or otherwise since the start of the trajectory, (ii) the unfolding force, (iii) the number of modules not-yet-unfolded in the construct (for the multimodular $(GB1)_8$), and (iv) the force immediately after the previous unfolding event (for the multimodular $(GB1)_8$), or 0 pN for the first unfolding module in the trajectory. The recorded parameters were then used to perform Maximum-Likelihood (ML) estimation of the unfolding kinetic parameters [17].

For the persistence length of Eq. (6.13) we used the fixed value $p_l = 0.35$ nm [48] all throughout.

### 6.4.5 Viscosity adjustment of the spontaneous unfolding kinetics inferred from the experimental data

Kramers theory predicts that

$$k_u(0) = A/\gamma \, exp[-\Delta G_u/(k_B T)] \tag{6.27}$$

where $A$ is a constant, $\Delta G_u$ is the height of the unfolding activation barrier and $\gamma$ is the reaction friction [182]. The unfolding free energy difference in two different conditions, denoted below by superscripts $C1$ and $C2$, computes as

$$\Delta\Delta G_u = -k_B T \ln\{[\gamma^{C1} k_u^{C1}(0)]/[\gamma^{C2} k_u^{C2}(0)]\}, \tag{6.28}$$

assuming the prefactor $A$ is the same in the two conditions.

According to Kramers theory, the friction coefficient $\gamma$ is an abstraction standing for all the ways by which energy can be dissipated out of the reaction coordinate, direct dissipation into the solvent being but one such mechanism. In particular, there need not be a simple relationship between $\gamma$ and solvent viscosity [182].

The simple relationship

$$\gamma = 1/\eta\,, \tag{6.29}$$

where $\eta$ is the solvent viscosity, was found in a few instances to result in good agreement with experimental data on protein folding and unfolding [183, 184, 185, 186, 187, 188]. Eq. (6.29) has been sometimes used to account for viscosity effects in the interpretation of single-molecule mechanical folding and unfolding experiments [60, 12, 19].

However there is abundant experimental evidence that the dependency of Kramers' friction coefficient $\gamma$ on solvent viscosity $\eta$ is noticeably weaker than $1/\eta$ [189, 190, 191, 192, 193, 194, 195]. Herein we adopt another, potentially more accurate formula for the friction coefficient that also takes into account the internal friction of the protein:

$$\gamma = 1/(\sigma + \eta)\,, \tag{6.30}$$

where $\sigma = 4\,cP$ is an estimate of the protein internal viscosity [195]. Therefore when comparing to the unfolding activation barrier changes from the SMFS experiments of Ref. [12, 19], the experimentally inferred values reported in Table 6.1 have been re-adjusted for viscosity using Eq. (6.30) instead of Eq. (6.29).

Moreover, the spontaneous rate of unfolding of I27 in the presence of glycerol 30% v/v reported in the last column of Table 6.6 was also adjusted using Eq. (6.30) (see Section 6.4.9).

### 6.4.6 Polyprotein design and expression

We followed the protein construct design proposed by Prof. Julio Fernandez (Columbia University) for the study of the random coiled titin N2B segment [24]. Chimeric polyproteins were obtained starting from pAFM1-4 and pAFM5-8 vectors, gently provided by Prof. Jane Clarke (Cambridge University) and constructed according to Ref. [196]. The eight I27 module plasmid was reconstituted from pAFM1-4 and pAFM5-8, obtaining the pAFM8m vector. The plasmid was transformed into E. coli C41(DE3) cells [197] (obtained from Prof. John E. Walker (MRC - Dunn Human Nutrition Unit, Cambridge) with the agreement of the Medical Research Council centre of Cambridge). The cells were grown and the expression of proteins was induced as described in Ref. [196]. Recombinant proteins were purified by $Ni^{2+}$-affinity chromatography in 20 mM sodium phosphate buffer pH 8, 500 mM NaCl; the elution from resin was obtained with 20 mM imidazole. After purification the protein was kept at 193.15 K with glycerol 15% v/v.

### 6.4.7 Single molecule force spectroscopy experiments

Constant velocity mechanical unfolding experiments were performed with a Veeco Picoforce Atomic Force Microscope (AFM) equipped with a DI Multimode Nanoscope IIIa controller (Bruker) and gold-coated, V-shaped silicon nitride cantilevers (NPG model; Bruker) with a nominal spring constant of 0.06 N/m. Unfolding experiments were performed on the homomeric polyprotein $(I27)_8$ in two different solvent conditions. In the first condition we used standard PBS buffer at pH 7.0. In the second condition we used the same buffer but in the presence of glycerol as a cosolvent, at a final concentration of 30% v/v. All the experiments have been performed at a temperature of approximately 301.15 K. For the experiments, a drop of the $(I27)_8$-containing solution ($5\mu L$, 0.1 g/L) was deposited on a flame cleaned glass coverslip for about 20 minutes. The fluid cell was then filled with either the plain buffer or the glycerol-containing buffer, and sealed on top of the coverslip. Thermal tuning was performed in the respective solution to determine the cantilever spring constant. Pulling experiments were then started after a few minutes of incubation.

At the constant pulling velocities of 50.1, 100, 198, 513, 969, 2180 and 4360 nm/s we recorded (i) 12, 24, 32, 27, 39, 148, and 160 unfolding events, respectively, from 1, 2, 3, 4, 4, 7 and 7 different experiments, respectively in plain buffer, and (ii) 15, 24, 93, 124, 133, 151, and 362 unfolding events, respectively, from 1, 3, 5, 5, 5, 3, and 6 experiments, respectively with glycerol as cosolvent. Data filtering and extraction were performed using Hooke [71].

### 6.4.8 Adjusting forces for viscosity

We computed the viscosity of the 30% v/v glycerol solution at 301.15 K as 2.357 cP using a formula reported elsewhere [114].

Immersion in a liquid environment dramatically alters the thermal noise spectrum of the cantilever, due to the strong effects of fluid loading [11, 115, 116]. Problems in the determination of the spring constant arise when the resonance frequency drops to 1 kHz or below, as it happens for soft cantilevers or highly viscous media [11]. However in the presence of 30% v/v glycerol as a cosolvent the resonance frequency of our cantilevers stays above 2 kHz, above the frequency range likely to result in viscosity-induced errors in spring constant determination that are comparable to the error intrinsic to the thermal tuning method [117]. Indeed, by calibrating a set of three cantilevers in both conditions we obtained an error of about 1.5% between the two conditions.

The measured unfolding forces are underestimated in high-viscosity solutions due to the viscous drag on the AFM cantilever [118, 119, 120, 121]. A way to remove this artifact has been recently reported [118]. According to the this procedure the drag force can be computed as

$$F_{vis} = b_l v_{liquid} - b_c v_{tip} \,, \tag{6.31}$$

where $v_{liquid}$ is the pulling velocity, $v_{tip}$ is the velocity of the tip, $b_l$ and $b_c$ are viscous drag coefficients related to the liquid motion and the bending of the cantilever (tip motion), respectively. These two parameters depend on tip-surface separation according to the formulae

$$b_l = 6\pi\eta a_l^2/(s + h_l) \tag{6.32}$$

and

$$b_c = 6\pi\eta a_c^2/(s + h_c) \tag{6.33}$$

where $s$ is the tip-sample separation, $\eta$ is the solvent viscosity, $a_l$, $h_l$, $a_c$ and $h_c$ being parameters to be determined. We extracted the mentioned parameters by fitting Eq. (6.32), (6.33) to the $b_l$ and $b_c$ values reported in Fig. 3 of Ref.[118], obtaining the approximate values $a_l = 24.3\mu m$, $h_l = 3.5\mu m$, $a_c = 19.8\mu m$ and $h_c = 2.9\mu m$. We then adjusted the force values collected at each sampled tip-surface separation in the experimental data taken in 30% v/v glycerol according to Eq. (6.31). The average unfolding forces at 50.1, 100, 198, 513, 969, 2180 and 4360 nm/s increased by 0.01, 0.02, 0.10, 0.22, 0.44, 1.00 and 2.68 pN, respectively.

### 6.4.9 Extracting the unfolding kinetics of I27 from SMFS velocity-clamp data

The unfolding kinetics of I27 at different pulling velocities were well described by Bell's model given in Eq. (7.3). Maximizing the likelihood of the I27 SMFS data [17] we obtained $\Delta x_u = 0.25$ nm in the absence of glycerol and $\Delta x_u = 0.23$ nm in the presence of glycerol 30% v/v. Since the difference is likely within experimental error, we inferred that the presence of glycerol does not change the characteristic unfolding distance of I27. Taking into account as well existing estimates of the unfolding distance of I27 [41, 60], we fixed the unfolding distance of I27 to $\Delta x_u =$ 0.25 nm in both conditions, obtaining $k_u(0) = 0.0013 \pm 0.000063 s^{-1}$ in the absence of glycerol and $k_u(0) = 0.0003 \pm 0.000009 s^{-1}$ in the presence of glycerol 30% v/v.

If we consider as a first approximation that the prefactor $A$ of Eq. (6.27) is not affected by the addition of glycerol, taking the ratio of the logarithm of the estimated spontaneous unfolding rates in the two solvent conditions we obtained that the presence of glycerol at a concentration of 30% v/v increases the unfolding energy activation barrier of I27 by $\Delta\Delta G_u = 1.63 \pm 0.06 k_B T$. By further adjusting for the viscosity effects on the unfolding kinetics via Eq. (6.30), we obtained our final experimental estimate of $\Delta\Delta G_u = 1.36 \pm 0.06\, k_B T$.

## 6.5 Results and discussion

### 6.5.1 Our SMFS experiments show that glycerol 30% v/v does not change the unfolding distance of I27

The presence of glycerol at a concentration of 30% v/v consistently shifted the unfolding force distribution of I27 to higher forces, as seen in Fig. (6.2). Extracting the unfolding kinetics of I27 as described in Section 6.4.9 we found that glycerol reduces the spontaneous unfolding rate $k_u(0)$ of protein I27 without changing the distance $\Delta x_u = 0.25$ nm between the ground state and the transition state along the reaction coordinate. The height of the unfolding activation barrier of I27 was increased in the presence of glycerol 30% v/v by $\Delta\Delta G_u = 1.36 \pm 0.06\, k_B T$.

### 6.5.2 Thermodynamics

We projected the free energy landscape of GB1 onto the non-mechanical reaction coordinate given by the weighted number of native contacts (see Section 6.3.3), which has been proposed as a good order parameter for protein folding and unfolding [131]. We performed the projection in the absence of osmolytes and in the presence of (i) DMSO 50%, 40%, 30%, 20% and 10% v/v, (ii) glycerol 30% v/v, and finally (iii) GndCl 2.25 and 1 M, conditions for which SMFS experimental data are available [63, 19, 12]. The projection of the energy landscape in some of the mentioned solvents is shown in Fig. (6.3).

If we make the simplifying assumption that the protein diffuses on the chosen reaction coordinate, we can in principle locate the transition state and measure its height. But the obtained energy landscape has a high degree of frustration, as seen in Fig. (6.3), which interferes with the significance and correct determination of the transition state [199]. Therefore we computed a Boltzmann-weighted moving average of the energy landscape with a window size of five distinct neighboring populated reaction coordinate values (see Fig. (6.4)), and we located the transition state by finding the maximum energetic level with the reaction coordinate value greater than or equal to 324. For DMSO 50% v/v and DMSO 40 % v/v the energy landscape as projected onto the weighted number of native contacts does not show any significant minimum for the unfolded state, making it impossible to identify a transition state. For all the other conditions, the described procedure

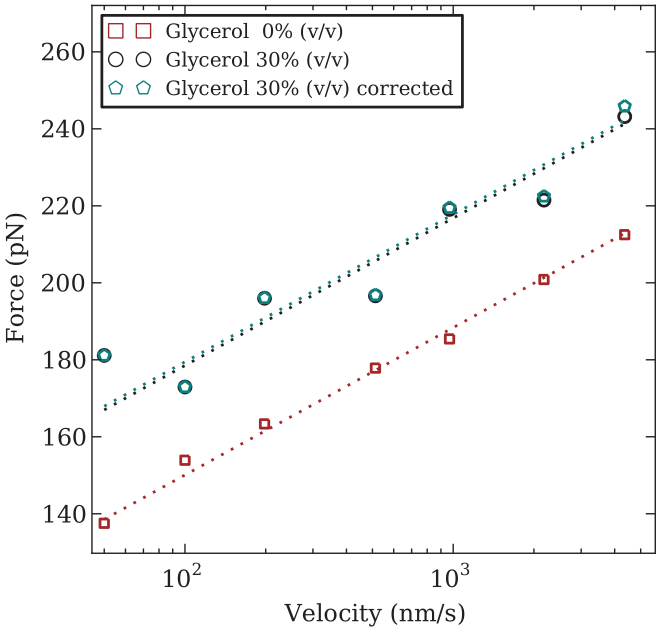


Figure 6.2: (Color online) Average unfolding force of I27 at various pulling speeds increases in the presence of glycerol 30% v/v, and increases again slightly when adjusting for the viscous drag-force on the cantilever. Only as a guide to the eye we fitted a dashed line to the average unfolding forces in each condition with a fixed slope of $k_BT/(\Delta x_u \log_{10}(e))$ where e is Eulers constant and $\Delta x_u = 0.25$ nm. The formula for the eye-guide fixed slope has been inspired from the so-called “standard method” of kinetic parameter estimation [198], and it was not used for the statistical estimation of the kinetic parameters that we report, which was performed instead via ML estimation (see Section 6.4.9).

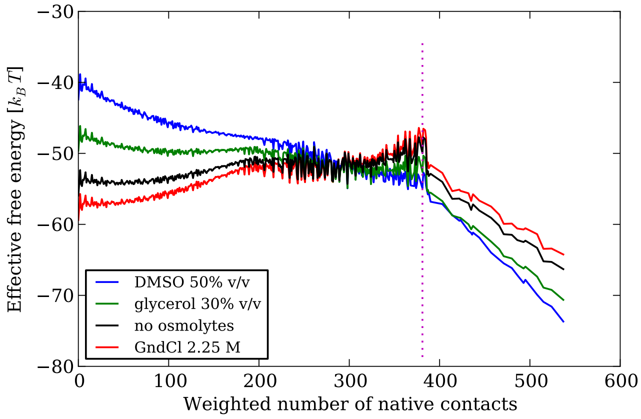


Figure 6.3: Energy landscape of protein GB1 projected onto the reaction coordinate given by the weighted number of native contacts $\phi(\vec{m})$, as defined by Eq. (6.9), in the absence of osmolytes, and in the presence of DMSO 50% v/v, glycerol 30% v/v and GndCl 2.25 M. The vertical bar at a weighted number of native contacts of 381 denotes the approximate position of the transition state.

identified the transition state at the same reaction coordinate value of 381 weighted native contacts and it measured changes in the unfolding activation energy barrier relative to the absence of osmolytes as indicated in Table 6.1. Although the activation barrier changes from in-bulk experiments are not necessarily comparable to those from SMFS experiments, the activation barrier changes estimated from the thermodynamic analysis based on the weighted number of native contacts agree qualitatively with the SMFS experimental ones.

We also projected the free energy landscape of I27 onto the non-mechanical reaction coordinate given by the weighted number of native contacts, in the absence of osmolytes and in the presence of glycerol 30% v/v. As can be seen in Fig. (6.5), the energy landscape of I27 shows little unfolding cooperativity, which is expected considering that I27 has an unfolding intermediate that is stable over AFM timescales [181]. Therefore this energy projection method does not make it possible to clearly identify an unfolding transition barrier for protein I27.

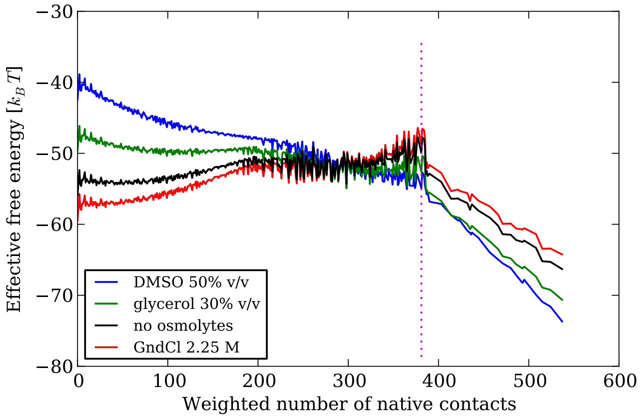


Figure 6.4: Boltzmann-weighted moving average of the effective free energy of protein GB1 as a function of the weighted number of native contacts $\phi(\vec{m})$, as defined by Eq. (6.9), in the absence of osmolytes, and in the presence of DMSO 50% v/v, glycerol 30% v/v and GndCl 2.25 M. The vertical bar at a weighted number of native contacts of 381 denotes the approximate position of the transition state.

| Solvent | $\Delta\Delta G_u[k_BT]$ | SMFS $\Delta\Delta G_u[k_BT]$ |
|---|---:|---:|
| DMSO 50% v/v | - | 1.89 |
| DMSO 40% v/v | - | 1.00 |
| DMSO 30% v/v | 1.19 | 0.66 |
| DMSO 20% v/v | 0.74 | 0.30 |
| DMSO 10% v/v | 0.28 | 0.04 |
| glycerol 30% v/v | 1.71 | 1.60 |
| no osmolytes | 0 | 0 |
| GndCl 2.25 M | -0.82 | -2.38 |
| GndCl 1 M | -0.41 | -0.94 |

Table 6.1: GB1 activation barrier height changes. The column $\Delta\Delta G_u[k_BT]$ indicates the activation barrier height change calculated from the thermodynamic analysis based on the weighted number of native contacts. The last column contains the activation barrier changes estimated from SMFS velocity-clamp experiments [12, 19, 63], adjusted for viscosity as described in Section 6.4.5.

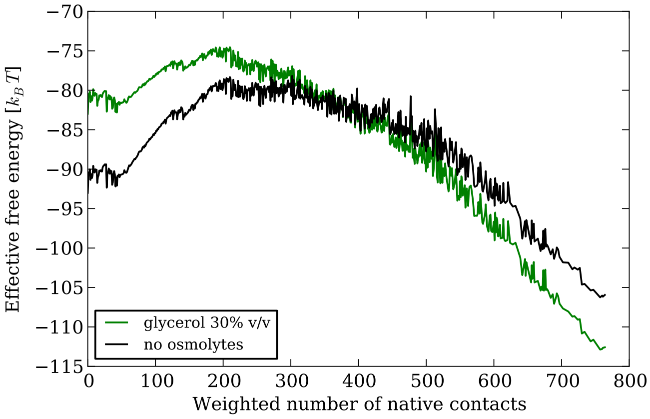


Figure 6.5: Energy landscape of protein I27 projected onto the reaction coordinate given by the weighted number of native contacts $\phi(\vec{m})$, as defined by Eq. (6.9), in the absence of osmolytes, and in the presence of glycerol 30% v/v.

### 6.5.3 Kinetics

For protecting osmolytes it holds in Eq. (6.6) that $\Delta g^{ref}_{bb,R[k]} > 0$. Therefore they disfavor microscopic states that are more exposed to solvent, thus slowing the transitions going from more native-like to more unfolded states according to Eq. (6.24) and overall slowing down the unfolding kinetics. Conversely, for denaturing osmolytes it holds that $\Delta g^{ref}_{bb,R[k]} < 0$. Therefore they favor the microscopic states that are more exposed to solvent, speeding up the transition rates going from more native-like to more unfolded states and overall speeding up the unfolding kinetics. The simulation results presented next are consistent with this view, and additionally they offer insights into the effect of osmolytes on the unfolding distance of the studied proteins.

#### Force-clamp

We performed force-clamp simulations of mechanical unfolding of GB1 with constant forces from 0 pN to 500 pN, inclusive, with a step size of 5 pN, in the absence of osmolytes and in the presence of (i) DMSO 50%, 40%, 30%, 20% and 10% v/v, (ii) glycerol 30% v/v, and (iii) GndCl 2.25 and 1 M . We found that in each solvent condition the logarithm of the force-dependent unfolding rate grows approximately linearly with force. Previous works of mechanical unfolding via WSME-like mod-

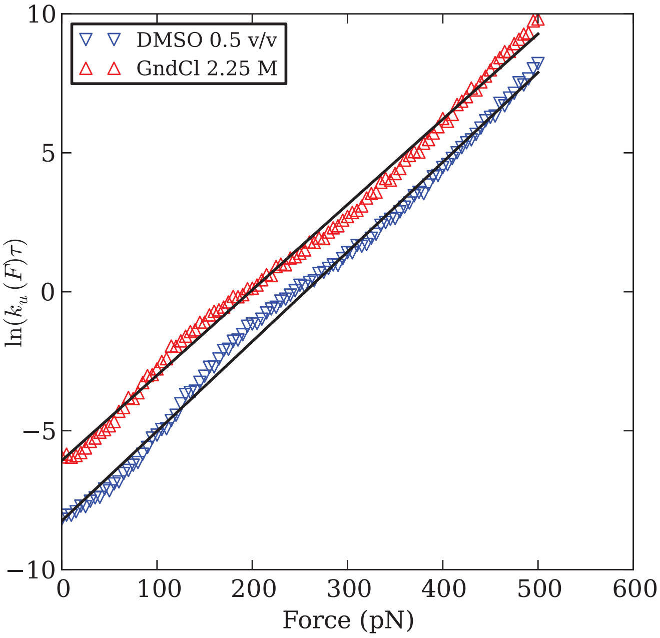


Figure 6.6: Logarithm of the unfolding rate of GB1 as a function of the stretching force in two solvent conditions, namely the most destabilizing one and the most stabilizing one, selected such as to reduce visual clutter. Each point represents the inverse of the average unfolding time from at least 125 trajectories. Lines correspond to a fit of Bell's model via Eq. (7.3).

els have fitted the Bell model only on ranges of forces selected to maximize the quality of the fit [127, 126, 130, 131]. Since we need to compare the fitted values between different solvent conditions, in order to remove any bias in the selection of the fitting force ranges, we adopted the simple approach of fitting over the full set of forces, even if that results in somewhat lower quality fits, as can be seen in Fig. (6.6).

Fitting Bell's model given by Eq. (7.3) to the force-dependent unfolding rates resulted in negligible variation of the unfolding distance $\Delta x_u$ among the different solvent conditions, as seen in Table 6.2. Moreover, the unfolding distance in all conditions was close to the experimentally measured value of $0.165 - 0.17$ nm [12, 19, 63]. The unfolding barrier height changes are also in qualitative agreement with the activation barrier changes inferred from the SMFS experiments, which are indicated in the last column of Table 6.1.

| Solvent | $\Delta x_u$[nm] | $\Delta\Delta G_u[k_BT]$ |
|---|---|---|
| DMSO 50% v/v | 0.134±0.003 | 1.717±0.279 |
| DMSO 40% v/v | 0.133±0.003 | 1.358±0.279 |
| DMSO 30% v/v | 0.131±0.003 | 0.862±0.279 |
| DMSO 20% v/v | 0.131±0.003 | 0.568±0.279 |
| DMSO 10% v/v | 0.129±0.003 | 0.203±0.279 |
| glycerol 30% v/v | 0.130±0.003 | 0.955±0.279 |
| no osmolytes | 0.128±0.003 | 0 |
| GndCl 2.25 M | 0.128±0.003 | -0.450±0.279 |
| GndCl 1 M | 0.128±0.003 | -0.192±0.279 |

Table 6.2: Mechanical unfolding kinetic parameters from force-clamp simulations. The column $\Delta\Delta G_u[k_BT]$ is to be compared with column SMFS $\Delta\Delta G_u[k_BT]$ from Table 6.1.

For the force-clamp simulations of mechanical unfolding of I27 we used constant forces from 0 pN to 200 pN, inclusive, with a step size of 5 pN, in the absence of osmolytes and in the presence of 30% glycerol v/v. We found that in either solvent condition the logarithm of the force-dependent unfolding rate shows two approximately linearly regimes with respect to force. Therefore we fitted Bell's model over each linear regime, as seen in Fig. (6.7), selecting the crossover point between the two regimes by minimizing the total squared fitting error.

For each of the two linear regimes, the unfolding distance showed only negligible variation between the two conditions, while the free energy changes are in qualitative agreement with the experimentally inferred activation barrier change, as can be seen in Table 6.3.

### Velocity-clamp

We performed velocity-clamp simulations of mechanical unfolding of $(GB1)_8$ with a cantilever of spring constant 0.06 N/m and velocities of 50.1, 100, 198, 513, 969, 2180 and 4360 nm/s, matching the spring stiffness and pulling velocities of some previous SMFS velocity-clamp experiments with $(GB1)_8$ in the presence of osmolytes [19, 12]. Since as described in Section 6.4.3, knowledge of the product $v\tau$ is required for velocity-clamp simulations, we estimated $\tau$ by confronting the average unfolding time from 1000 simulated force-clamp mechanical unfolding traces

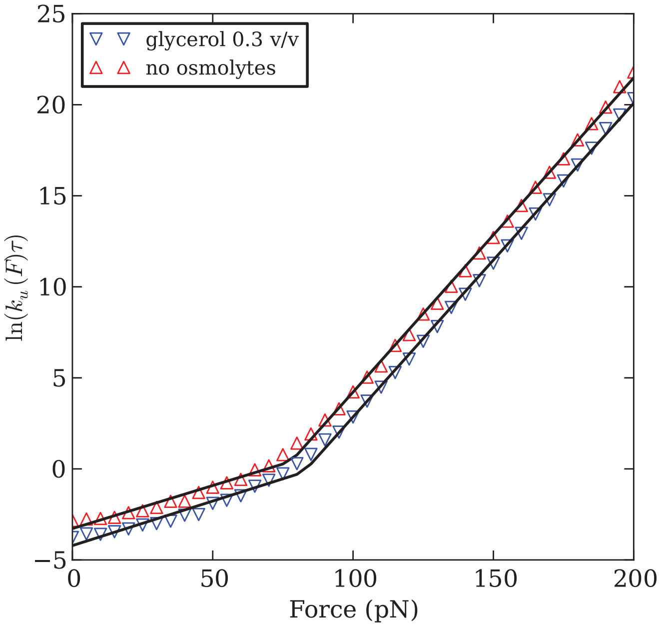


Figure 6.7: Logarithm of the unfolding rate of I27 as a function of the stretching force in two different solvent conditions. Each point represents the inverse of the average unfolding time from at least 125 trajectories. Lines correspond to a fit of Bell's model via Eq. (7.3) over two different linear regimes, selected for each solvent condition so as to minimize the total squared error. In the absence of osmolytes the crossing point was found to be 80 pN, while in the presence of glycerol 30% v/v the crossing point was found to be 85 pN.

| | no osmolytes | glycerol 30% v/v |
|---|---|---|
| I. $\Delta x_u$[nm] | 0.197±0.045 | 0.203±0.041 |
| I. $\Delta\Delta G_u[k_BT]$ | 0 | 0.933±0.666 |
| II. $\Delta x_u$[nm] | 0.718±0.023 | 0.716±0.025 |
| II. $\Delta\Delta G_u[k_BT]$ | 0 | 1.336±1.179 |

Table 6.3: Mechanical unfolding kinetic parameters of I27 from force-clamp simulations. The first three rows refer to the lower-force linear regime, while the last three rows refer to the higher-force linear regime. The estimation errors represent one standard deviation. For easy reference, the experimentally inferred $\Delta\Delta G_u$ in the presence of glycerol 30% v/v is $1.36 \pm 0.06\, k_BT$, as computed in Section 6.4.9.

| Solvent | $F^{exp}_{50.1}$ [pN] | $F^{sim}_{50.1}$ [pN] | $F^{exp}_{400}$ [pN] | $F^{sim}_{400}$ [pN] | $F^{exp}_{4360}$ [pN] | $F^{sim}_{4360}$ [pN] |
|---|---|---|---|---|---|---|
| DMSO 50% v/v | 200 | 350 | - | - | 300 | 539 |
| DMSO 40% v/v | 176 | 339 | - | - | 256 | 529 |
| DMSO 30% v/v | 172 | 318 | - | - | 244 | 515 |
| DMSO 20% v/v | 175 | 311 | - | - | 260 | 504 |
| DMSO 10% v/v | 139 | 293 | - | - | 240 | 495 |
| glycerol 30% v/v | 195 | 324 | - | - | 306 | 519 |
| no osmolytes | 119 | 278 | 178 | - | 253 | 485 |
| GndCl 2.25 M | - | 259 | 97 | - | - | 475 |
| GndCl 1 M | - | 268 | 148 [1] | - | - | 479 |

Table 6.4: Mean unfolding forces at three velocities from velocity-clamp experiments [12, 19, 63] and simulations with $(GB1)_8$.

at zero force with the experimentally inferred zero-force mean unfolding time of GB1 [12], obtaining the approximate value $\tau = 0.039$ s.

The simulated force-distance traces exhibited the characteristic WLC sawtooth pattern, as seen in Fig. (6.8). In each condition, the average unfolding forces, shown in Fig. (6.9), increased with increasing pulling velocities, as expected from Bell's model. Moreover, average unfolding forces increased with increasing DMSO concentration and they decreased with increasing GndCl concentration. The unfolding forces were however generally higher than the experimental ones (see Table 6.4), for reasons explained below.

For each solvent condition we performed Maximum-Likelihood (ML) estimation of the unfolding kinetic parameters by using all velocities simultaneously [17], with the results being summarized in Table 6.5. The unfolding activation barrier changes from velocity-clamp simulations are in good agreement with those from force-clamp experiments indicated in Table 6.2 and the experimental ones indicated in the last column of Table 6.1. The unfolding distance estimated from velocity-clamp simulations is smaller than the one estimated from force-clamp simulations and from SMFS experiments, resulting in unfolding forces much higher than the experimental SMFS ones at the same velocities. However, the unfolding distance from velocity-clamp simulations still shows only negligible variation between the different solvent conditions.

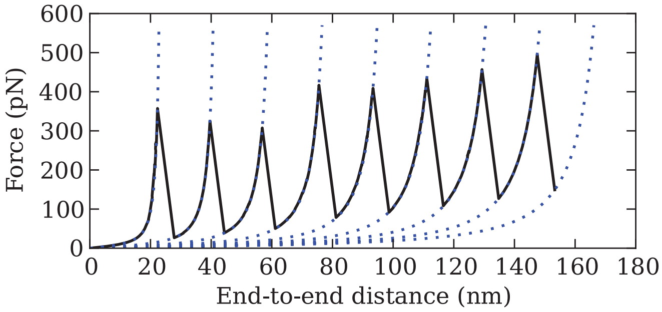


Figure 6.8: (Color online) Simulated velocity-clamp curve of $(GB1)_8$. The velocity was 969 nm/s, the cantilever spring constant was 0.06 N/m and the solvent had a GndCl concentration of 2.25 M.

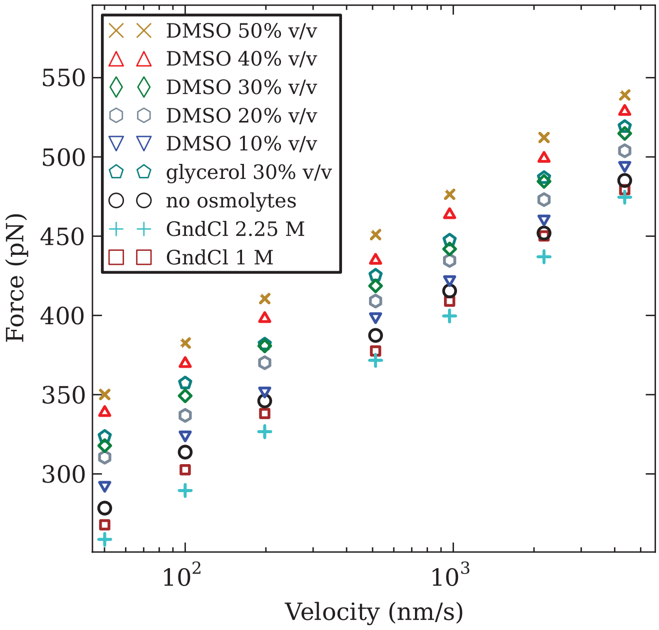


Figure 6.9: (Color online) Average unfolding force vs. pulling velocity in different solvent conditions for protein $(GB1)_8$. Each point represents the average unfolding force from at least 125 trajectories.

| Solvent | $\Delta x_u$[nm] | $\Delta\Delta G_u[k_BT]$ |
|---|---|---|
| DMSO 50% v/v | 0.086±0.0003 | 1.197±0.052 |
| DMSO 40% v/v | 0.086±0.0004 | 1.018±0.054 |
| DMSO 30% v/v | 0.085±0.0004 | 0.603±0.055 |
| DMSO 20% v/v | 0.085±0.0004 | 0.430±0.057 |
| DMSO 10% v/v | 0.085±0.0004 | 0.145±0.051 |
| glycerol 30% v/v | 0.085±0.0004 | 0.680±0.055 |
| no osmolytes | 0.085±0.0005 | 0 |
| GndCl 2.25 M | 0.084±0.0004 | -0.355±0.056 |
| GndCl 1 M | 0.084±0.0004 | -0.245±0.054 |

Table 6.5: Mechanical unfolding kinetic parameters of GB1 from velocity-clamp simulations. The estimation errors represent one standard deviation and they have been computed via a bootstrap (case resampling) procedure, where for each velocity an equal number of traces were extracted with replacement, at least 100 times. The column $\Delta\Delta G_u[k_BT]$ is to be compared with column SMFS $\Delta\Delta G_u[k_BT]$ from Table 6.1.

For the velocity-clamp mechanical unfolding simulations of I27, we used a single I27 module, in the absence and presence of glycerol 30% v/v. The simulated force-distance traces exhibited the characteristic WLC pattern, as seen in Fig. (6.10).

The I27 velocity-clamp simulations showed that the presence of glycerol 30% v/v increases the average unfolding forces of I27 (see Fig. (6.11)), without causing any significant movement of the unfolding transition state, producing but a deceleration of its spontaneous unfolding kinetics (see Table (6.6)). The unfolding forces from the simulation were also compatible with the range of unfolding forces from the SMFS experiments, as seen in Fig. (6.2) and Fig. (6.11).

## 6.6 Conclusion

We showed experimentally that contrary to what has been previously reported, glycerol does not increase the unfolding distance of I27. In light of previous similar findings with GB1 in the presence of DMSO, glycerol and GndCl, taken together these results suggest that there is a general mechanism through which osmolytes affect the mechanical stability of proteins that does not affect their unfolding

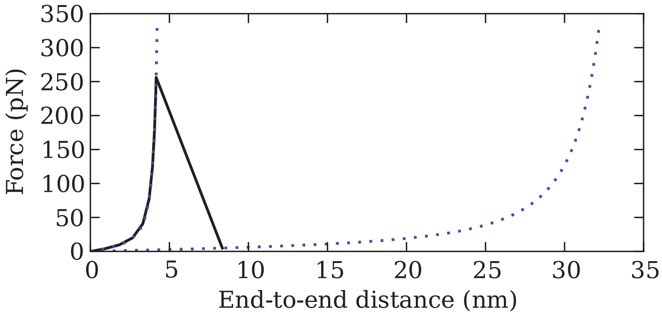


Figure 6.10: (Color online) Simulated velocity-clamp curve of I27. The velocity was 50.1 nm/s, the cantilever spring constant was 0.06 N/m and the solvent was free of osmolytes.

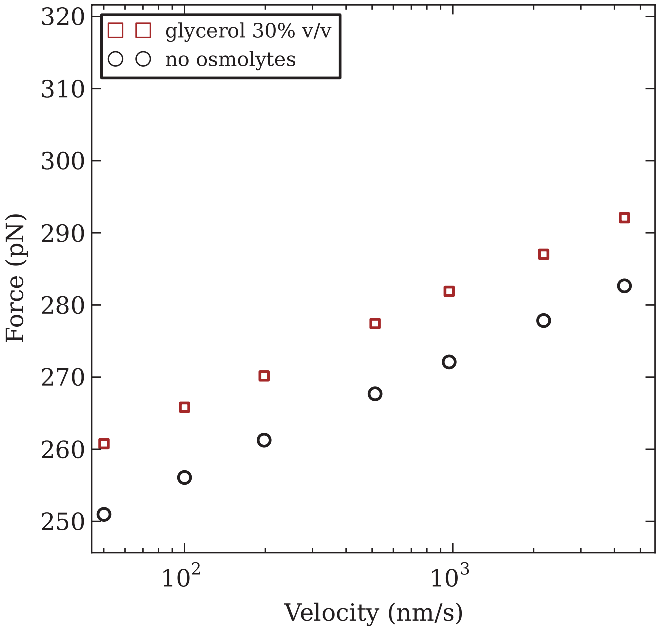


Figure 6.11: (Color online) Average unfolding force vs. pulling velocity in different solvent conditions for protein I27. Each point represents the average unfolding force from at least 1000 trajectories.

| Solvent | $\Delta x_u$[nm] | $\Delta\Delta G_u[k_BT]$ | $\Delta\Delta G_u^{exp}[k_BT]$ |
|---|---|---|---|
| glycerol 30% v/v | 0.603±0.004 | 2.233±0.351 | 1.36±0.06 |
| no osmolytes | 0.590±0.004 | 0 | 0 |

Table 6.6: Mechanical unfolding kinetic parameters of I27 from velocity-clamp simulations. The estimation errors represent one standard deviation and they have been computed via a bootstrap (case resampling) procedure, where for each velocity an equal number of traces were extracted with replacement, at least 100 times. The column $\Delta\Delta G_u[k_BT]$ is to be compared with column $\Delta\Delta G_u^{exp}[k_BT]$, the latter indicating the experimentally inferred activation barrier changes.

distance.

In order to understand if the above general mechanism is based on osmolytes adjusting the solvation quality for the protein chain, as suggested by the thermodynamic description of the osmolyte effect, we developed an Ising-like model for protein mechanical unfolding that incorporates the transfer free energy of various conformations of the protein chain. Notably, our Ising-like model is endowed with exactly solvable thermodynamics and it satisfies the WLC force-distance relation, which is a common fingerprint of mechanical unfolding experiments.

When applied to the mechanical unfolding of GB1 and I27 in osmolyte-containing solutions, our model correctly predicts that osmolytes do not change their unfolding distance. This level of agreement validates our approach for building a microscopic model of protein mechanical unfolding in the presence of osmolytes, and it strongly suggests that osmolytes may not play a structural role at the unfolding transition state of proteins, contrary to what has been previously suggested.

## 6.7 Acknowledgments

We are grateful to Prof. Julio Fernandez for his help in investigating the contrast between our results of mechanical unfolding of I27 in the presence of glycerol and those of Ref. [15].

This work was supported by Ministero dell'Università e della Ricerca-Fondo per gli Investimenti della Ricerca di Base (MIUR-FIRB) RBNE03PX83/001; MIUR-FIRB Progetto NG-lab (G.U. 29/07/05 n.175), Progetti di Ricerca di Interesse

Nazionale (PRIN) 2008 (Prot. 2008KZ3E5, Prot. 2008SYP79, Prot. 20085SYP79) and MIUR.

## 6.8 Note

After this work has been completed, Ref. [15] and Ref. [16] have been retracted [200, 201].

# Chapter 7

# Trimethylamine N-oxide has a strongly pH-dependent effect on protein mechanical stability

Unpublished (work-in-progress).

## 7.1 Abstract

We show via single-molecule force spectroscopy that the protecting osmolyte trimethylamine N-oxide (TMAO) in its zwitterionic form enhances the mechanical stability of the B1 immunoglobulin-binding domain of protein G from “Streptococcus” (GB1). Conversely, our experiments at low pH show that the cationic form of TMAO decreases the mechanical stability of GB1. By fitting an appropriate kinetic model we demonstrate that TMAO does not shift the transition state of GB1 at all pH values, indicating that any TMAO-induced differences in the height of the unfolding barrier may be attributable to non-structural effects such as protein solvation. Finally, we employ an existing statistical mechanics theory for protein backbone solvation to support the hypothesis that cationic TMAO may be preferentially excluded from the protein backbone, though to a lesser extent than zwitterionic TMAO. These results suggest that at low pH the preferential interaction with protein side-chains may dominate the energetics of TMAO-protein interactions.

## 7.2 Introduction

Across all taxa nature regulates protein stability through a class of ubiquitous organic compounds called "osmolytes". "Protecting osmolytes" (e.g., trimethylamine N-oxide (TMAO), betaine, sucrose, trehalose, sarcosine, sorbitol, proline, glycerol, dimethylsulfoxide) are known to stabilise the native states (N) of proteins with respect to their denatured states (D). Conversely, "denaturing osmolytes" (e.g., urea, guanidinium chloride (GndCl)) thermodynamically favour D over N.

According to the "osmophobic effect" theory, protecting osmolytes are excluded from the peptide backbone thus increasing the population of compact N states since they expose less backbone to the solvent than D. Although they also exert a collectively favourable interaction with protein side-chains, this opposing force is of lesser magnitude than their preferential exclusion from the backbone [202].

Among the protecting osmolytes, TMAO stands out as one of the strongest universal protein stabilisers in physiological conditions [13, 1, 203]. In fact it has been recently shown to correctly fold mutant proteins *in vivo* [204] and to counteract the effects of urea through a non protein-specific mechanism [205]. However there is also a growing number of studies linking TMAO with various deleterious effects [206, 207, 208, 209, 210, 211, 212]. For example, the fact that shallow-water teleosts accumulate less TMAO than deep-water ones even across the same species [213] has been explained through the ability of protecting osmolytes to sometimes stabilise protein conformations other than N [214].

Further adding to the elusive nature of this osmolyte [215], in-bulk experiments have shown that the effect of TMAO on the thermodynamic stability of three different proteins, namely pancreatic RNase A, hen egg white lysozyme and bovine apo-$\alpha$-actalbumin, is highly pH-dependent. More exactly, at pH above TMAO's $pK_a$ of 4.66 ± 0.10, when its neutral zwitterionic form is preferred, TMAO was found to stabilize the three mentioned proteins, while at lower pH when its cationic form is dominant, TMAO was found instead to destabilise them [21]. A later study has also shown that TMAO destabilises ovine and human prion protein at low pH, and that it does so to an even stronger extent than urea [22].

To understand to what extent the effect of TMAO on protein stability is generally pH-dependent we need to expand the set of proteins studied in TMAO-water mixtures at different pH values. Furthermore, in order to decipher the mechanism by which this apparent effect comes about we need to characterise the effect of TMAO not only on the thermodynamically-accessible N and D states, but also on the transition states of proteins.

Studies of enzyme inactivation and renaturation kinetics [216, 214, 209, 217, 218] and controversial [219, 220, 221, 222] $\phi$-value analysis [99] notwithstanding, direct observation of pH-dependent TMAO-induced effects on protein transition states has yet to be achieved. This endevour requires capturing structural characteristics and energetic levels of microsecond-long protein transition states [140] for comparison in the presence and absence of TMAO [21].

Single-Molecule Force Spectroscopy (SMFS) has recently emerged as an invaluable tool to characterise protein transition states as projected onto the direction of force application [17, 18]. The availability of such geometric information from mechanical experiments allowed us and others to extend the osmophobic effect theory to explain the speedup of the folding rate and the slowing down of the unfolding rate of proteins under the effect of protecting osmolytes such as dimethylsulfoxide [12] and glycerol [19, 20] at physiological pH.

While few intracellular proteins have mechanical function, many proteins with non-mechanical function do have mechanical stability. From the latter category, the B1 immunoglobulin-binding domain of protein G from "Streptococcus", commonly referred to as GB1, stands out due to its high mechanical stability [69] that has been extensively characterised [223, 224]. Being endowed with two-state behaviour [225] and thermodynamic stability over the large pH range from 2.7 [146] to 11.25 [226] , GB1 is an ideal model system for elucidating the effect of osmolytes at different pH values. Therefore we pulled GB1 under the Atomic Force Microscope (AFM) in the presence and absence of TMAO and at varying values of pH, measuring the unfolding force distribution in each solvent condition.

In order to check if a backbone-based effect would suffice to explain our experimental findings, we turned to a quantitative solvation model based on backbone-solvent interaction energies. Using just osmolyte polarity and interaction degeneracy, this statistical mechanics model has been previously found to predict with both correlation and slope close to 1 the experimental transfer free energies of the peptide backbone unit into osmolyte-water mixtures for a representative set of denaturing and protecting osmolytes, including the zwitterionic form of TMAO [1]. Feeding a suitable parameterisation of cationic TMAO into the mentioned

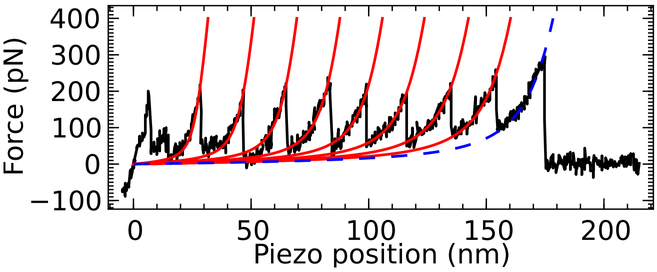


Figure 7.1: (Colour online) Force vs. extension plot obtained by stretching a complete $(GB1)_8$ molecule. The continuous red (light grey) curves and the dashed blue (medium grey) curve represent Worm-Like Chain (WLC) [58] fits of the force-response of the polypeptide chain before the unfolding of each GB1 module and before molecule detachment, respectively.

quantitative backbone solvation model we estimated the sign of the free energy of transfer of a backbone peptide unit into an aqueous TMAO solution at low pH.

Finally, by meeting the theoretical predictions of the backbone solvation theory with the experimental single-molecule experiments, we put our findings in the larger context of previous in-bulk thermodynamic studies of the pH-dependent effect of TMAO on protein stability.

## 7.3 Results

We pulled at a constant velocity of 2180 nm/s homopolymer molecules consisting of 8 tandem repeats of GB1, construct we shall refer to as $(GB1)_8$ (see Section 7.5.2), obtaining characteristic sawtooth force-extension curves as seen in Fig. 7.1. The experiments were performed in sodium acetate buffer 50 mM at pH 3.5, 4, 5 and 6 and in standard phosphate buffered saline at pH 7, in the presence and absence of TMAO 2M.

In order to minimise the effect of cantilever thermal calibration errors [117] on our ability to distinguish the unfolding force distributions of GB1 in the presence and absence of TMAO we used each cantilever multiple times at the same pH by switching the buffer back and forth between TMAO 0M and TMAO 2M. For each such condition we recorded a large number of unfolding events that allowed us to

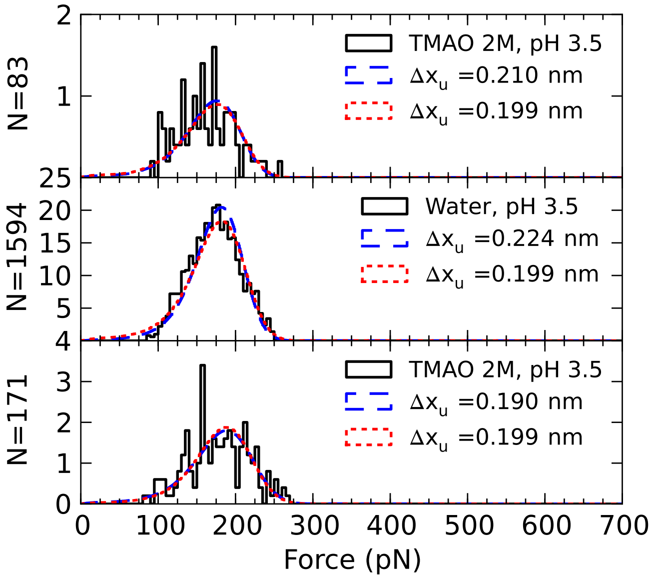


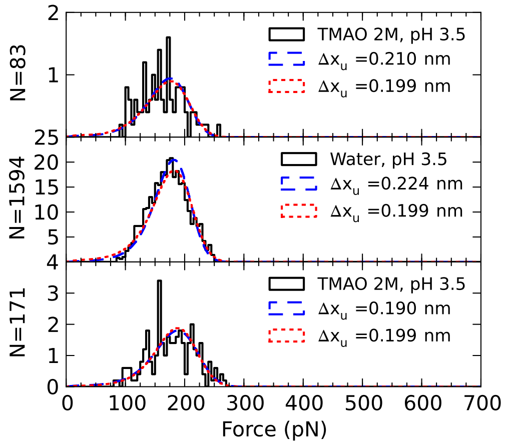
Figure 7.2: Unfolding force distributions at pH 3.5 with the same cantilever from top to bottom in chronological order. The dashed blue (medium grey) and dotted red (light grey) curves represent the expected histograms of unfolding forces computed with the Maximum Likelihood (ML) estimates of $\Delta x_u$ and $k_u(0)$ with $\Delta x_u$ that is either free or shared between all datasets at the same pH, respectively. Measurement noise was taken into account according to Eq. 7.10.

extract unfolding force histograms at pH 3.5, 4, 5, 6 and 7 as seen in Fig. 7.3, Fig. 7.4, Fig. 7.5, Fig. 7.6 and Fig. 7.7, respectively.

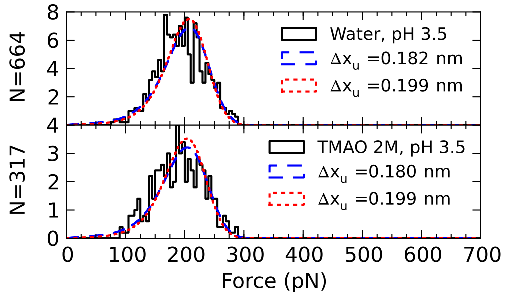
By extracting the average unfolding force at each pH value we found that while TMAO 2M increases the mean unfolding force at $pH \geq 5$, it actually lowers the mean unfolding force at $pH \leq 4$, as seen in Fig. 7.8.

For each dataset we performed Maximum Likelihood (ML) estimation [17] of $\Delta x_u$ and $k_u(0)$ with an improved procedure taking into account the measurement noise (see Section 7.5.1). By pooling together all the estimates for the same solvent we found that the presence of TMAO 2M in solution does not significantly alter GB1's $\Delta x_u$ at all pH values, as seen in Fig. 7.9.

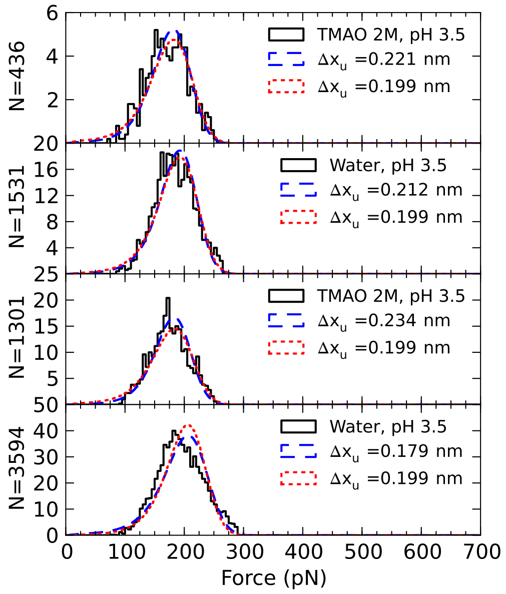
Since small errors in $\Delta x_u$ lead to large errors in $k_u(0)$, we performed ML estimation of $k_u(0)$ with shared pH-dependent $\Delta x_u$, i.e., one $\Delta x_u$ for each different pH, which does not drastically change the quality of the fits, as seen in Fig. 7.3,



(a) First cantilever



(b) Second cantilever



(c) Third cantilever

Figure 7.3: Unfolding force distributions at pH 3.5 with three cantilevers a), b) and c), for each cantilever from top to bottom in chronological order.

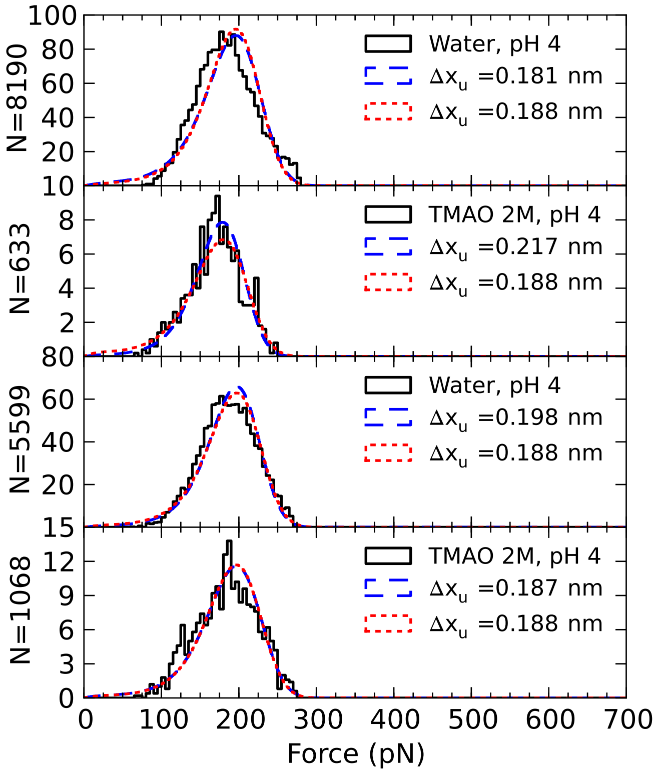


Figure 7.4: Unfolding force distributions at pH 4 with the same cantilever from top to bottom in chronological order.

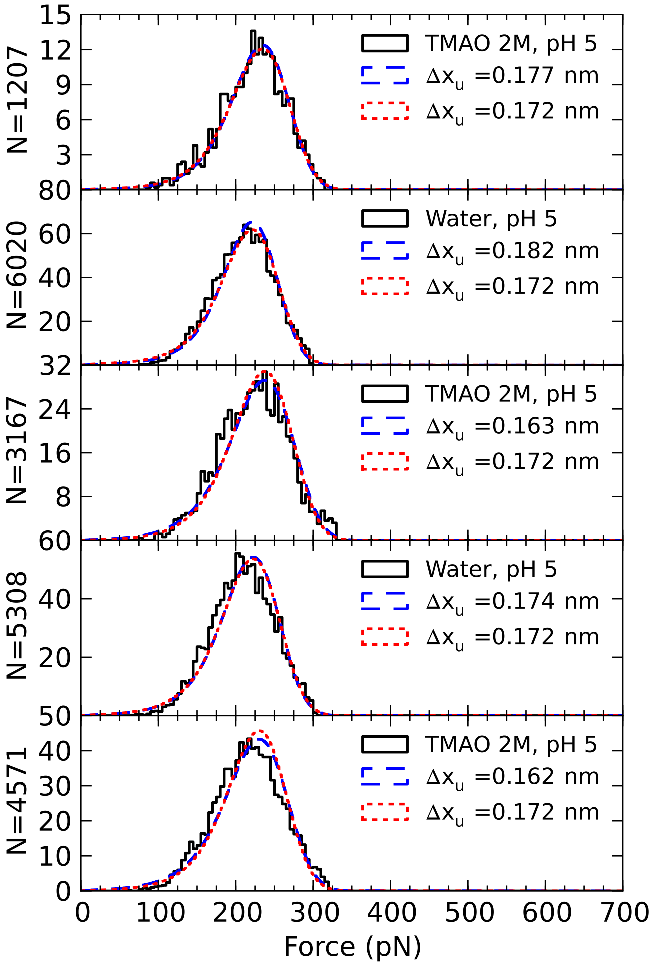


Figure 7.5: Unfolding force distributions at pH 5 with the same cantilever from top to bottom in chronological order.

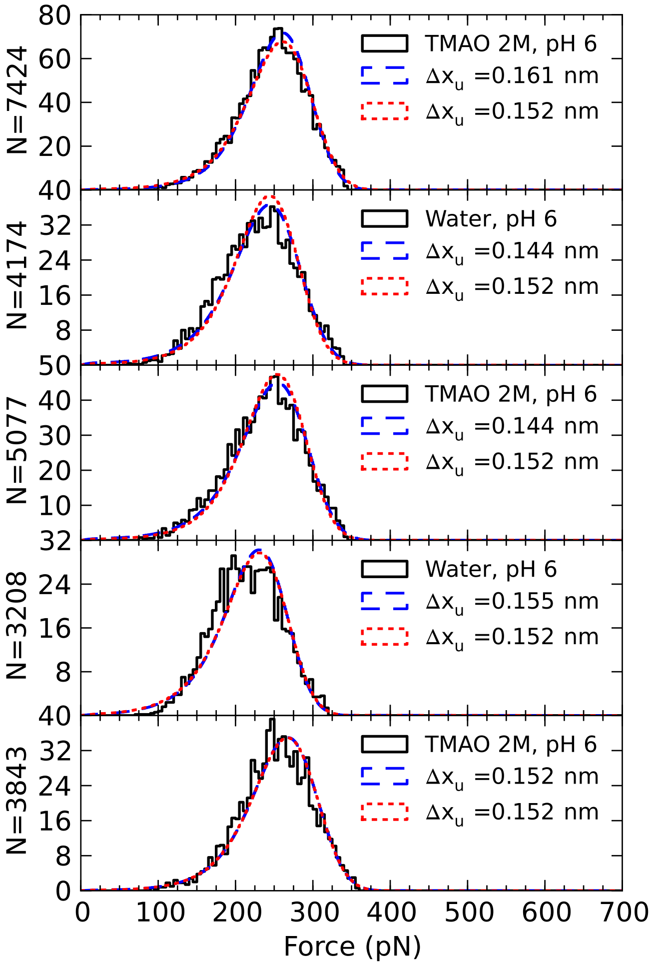


Figure 7.6: Unfolding force distributions at pH 6, measured with the same cantilever from top to bottom in chronological order.

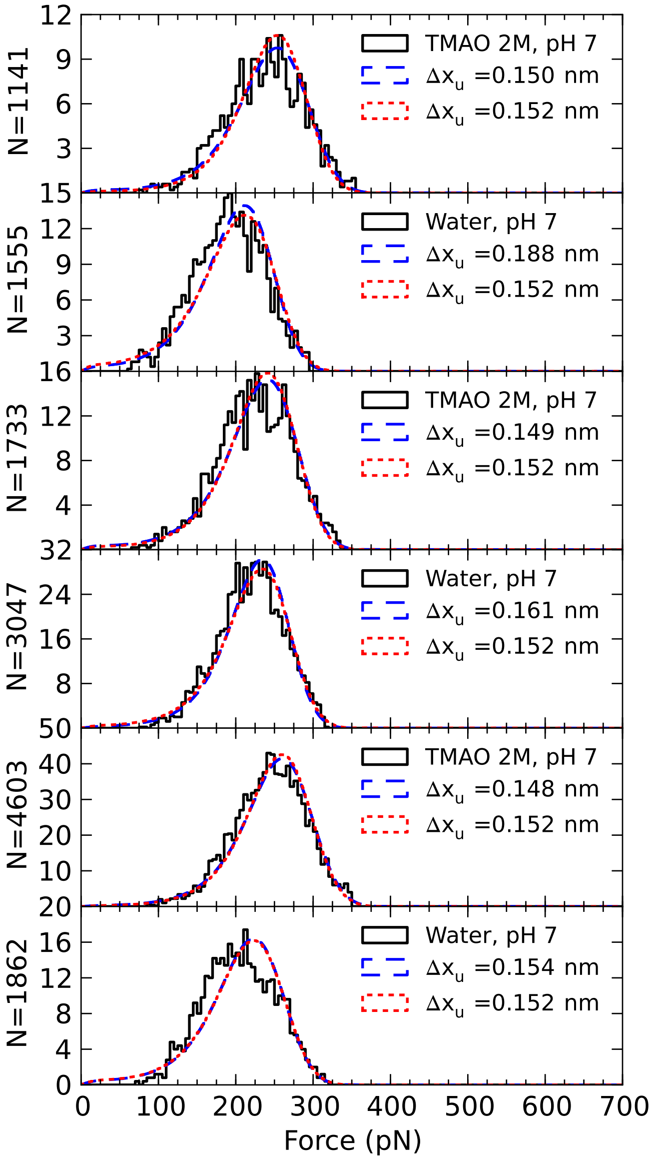


Figure 7.7: Unfolding force distributions at pH 7 with the same cantilever from top to bottom in chronological order.

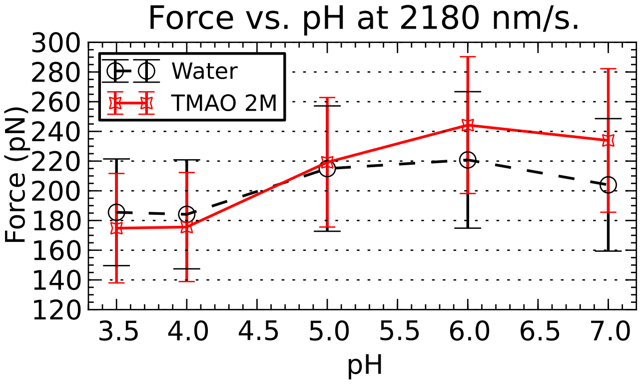


Figure 7.8: Average unfolding force vs. pH with error bars extending one standard deviation of the sample above and below the mean value. The black dashed and continuous red (light grey) lines are a guide to the eye for the average force in the absence and presence of TMAO 2M, respectively.

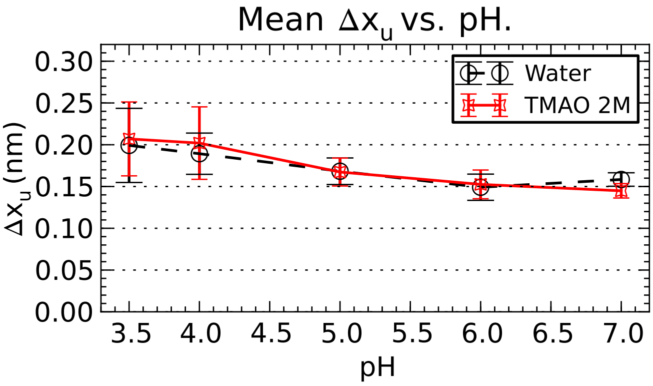


Figure 7.9: (Colour online) Average of $\Delta x_u$ estimates vs pH with error bars extending two standard deviations of the sample above and below the mean value. The black dashed and continuous red (light grey) lines are a guide to the eye for the average $\Delta x_u$ in the absence and presence of TMAO 2M, respectively.

Fig. 7.4, Fig. 7.5, Fig. 7.6 and Fig. 7.7, indicating that TMAO 2M does not significantly change $\Delta x_u$ of GB1.

In light of extensive evidence that TMAO does not perturb the native structure of proteins [21, 227, 133, 110, 228], our finding that TMAO 2M does not shift the unfolding transition state on the main reaction coordinate suggests that *a*) TMAO does not change the unfolding pathway of GB1, and *b*) the structural characteristics of GB1's unfolding transition state (TS) remain largely unchanged in the presence of TMAO.

Therefore at any given pH the difference between the height of the activation barrier in the presence and absence of the osmolyte $\Delta\Delta G_u = \Delta G_u(2M) - \Delta G_u(0M)$ can be assigned to the unequal interaction with the solvent of TS relative to N.

According to Kramers theory the spontaneous unfolding rate is described by the Arrhenius-like equation $k_u(0) = A/\gamma\, exp[-\Delta G_u/(k_B T)]$, where $A$ is a constant prefactor, $T$ is the temperature and $\gamma$ is the reaction friction that sums up direct dissipation into the solvent and all other possible means through which energy gets dissipated out of the reaction coordinate [182]. The formula $\gamma = 1/(\sigma+\eta)$ has been proposed to describe the reaction friction, where $\sigma$ is the internal friction of the protein while $\eta$ is the solvent viscosity [195]. Lacking an estimate for the internal viscosity of GB1 and assuming that the presence of TMAO does not affect the prefactor $A$, we computed $\Delta\Delta G_u$ for two extreme cases, namely $\sigma = 0$ [140, 184] or $\sigma \gg \eta$ [229], presuming that the "true" value lies somewhere in-between [195].

Both extreme estimates of the protein internal friction $\sigma$ lead to the observation that TMAO 2M lowers GB1's unfolding activation barrier when its cationic form is the dominant species, and as its zwitterionic form starts to dominate at higher pH, the effect is reversed (see Fig. 7.10).

In order to elucidate the role of TMAO-backbone chemistry in protein stability at low pH, we employed the statistical mechanics model of Ref. [1], according to which the surface area (SA) of an osmolyte is computed with a probe radius of 1.4 Å, ignoring hydrogen atoms, and assigning positive polarity to nitrogen, negative polarity to oxygen and neutral polarity to all other atoms. Notably, the model treats all nitrogens and oxygens as having the same charge magnitude, while water is considered to consist of equal positive, negative and neutral SAs of $SA_{w,+} = SA_{w,-} = SA_{w,0} = 10\,\text{Å}^2$. The peptide backbone is considered to have

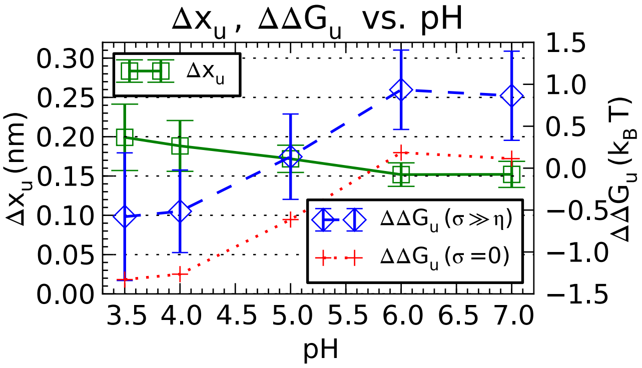


Figure 7.10: (Colour online) The continuous green (grey) line labelled $\Delta x_u$ is a guide to the eye for the shared pH-dependent unfolding distance of GB1 with error bars extending two standard deviations of the sample. The dashed blue (medium grey) line labelled $\Delta\Delta G_u(\sigma \gg \eta)$ is a guide to the eye for the change in the activation barrier height caused by the presence of TMAO 2M uncorrected for viscosity, i.e, assuming a very high protein internal friction, with error bars extending one standard deviation of the sample. The dotted red (light grey) line labelled $\Delta\Delta G_u(\sigma \gg \eta)$ is a guide to the eye for the change in the activation barrier height caused by the presence of TMAO 2M uncorrected for viscosity, i.e, assuming a very high protein internal friction.

three interaction sites, one positive (on the nitrogen atom) and two negative (on the backbone carbonyl oxygen), all of the same charge magnitude. The model has a tunable parameter which is the interaction energy $H > 0$ between like-charges, considering that opposite charges interact with energy $-H$, and that interactions involving neutral charges take place at zero energy.

We performed a least-squares fit of the parameter $H$ (see Section 7.5.4) for the 10 (ten) osmolyte descriptions in Table 1 of Ref. [1], obtaining H=1381 cal/mol, a correlation of 0.98, a regression slope of 0.99 and a small intercept of -1.94 cal/mol. Notably, for 1M zwitterionic TMAO ($SA_{o,+} = 0\,\text{Å}^2$, $SA_{o,-} = 43.2\,\text{Å}^2$, $SA_{o,0} = 168.4\,\text{Å}^2$ [1]) the fitted model predicts a transfer free energy of 88.29 cal/mol, in excellent agreement with the experimental value of 89 ± 2 cal/mol.

In order to parameterise cationic TMAO we considered that the negative SA of zwitterionic TMAO becomes positive, i.e., $SA_{o,+} = 43.2\,\text{Å}^2$, $SA_{o,-} = 0\,\text{Å}^2$, $SA_{o,0} = 168.4\,\text{Å}^2$. The model predicts in this case a transfer free energy of the peptide backbone unit in 1 M cationic TMAO of 75.26 cal/mol, suggesting that cationic TMAO is still preferentially excluded from the peptide backbone as compared to water, though to a lesser extent that zwitterionic TMAO.

## 7.4 Discussion

Existing studies of the effect of TMAO on protein stability have focused either on the thermodynamics of the N and D states or on enzyme inactivation and renaturation kinetics, without directly studying the effect of TMAO on the protein TS.

By fitting Bell's model (see Eq. 7.3) to our SMFS experimental data we found that in the pH range 3.5-7 TMAO does not shift the TS of GB1 on the mechanical reaction coordinate, suggesting that it does not change its structural characteristics or the unfolding pathway of GB1. Therefore we could compare the energetic level of TS relative to N, in the presence and absence of TMAO, ascribing any differences to non-structural effects such as protein-solvent interactions. We showed that close to physiological pH when the zwitterionic form of TMAO is dominant the energetic level of TS is raised with respect to N in the presence of TMAO, while at low pH when the cationic form is dominant, the energetic level of TS is lowered with respect to N in the presence of TMAO.

It is rather well accepted that the osmophobic effect of the peptide backbone applies to zwitterionic TMAO by raising the energetic levels of both N and D, but raising the energetic level of D more than it raises the energetic level of N. Conversely, side-chains exhibit overall a net osmophilic effect, but it is of lesser magnitude and therefore it is dominated by the osmophobicity of the peptide backbone [202]. Based on our finding that that the TS of GB1 is structurally similar in the presence and absence of TMAO, the increased mechanical stability of GB1 against mechanical unfolding close to neutral pH is also likely a consequence of the osmophobic nature of the peptide backbone.

Some previous thermodynamic studies of proteins in aqueous TMAO solutions at low pH found that TMAO destabilises them under its $pK_a$, and at least three different hypotheses have been proposed to explain this effect. *a*) One study suggested that cationic TMAO is still preferentially excluded from the peptide backbone relative to water, but that at low pH the opposing side-chain contributions start to dominate, thereby destabilising N with respect to D [21]. Conversely, another study proposed that *b*) cationic TMAO may preferentially interact with the peptide backbone, and that *c*) it may weaken electrostatic interactions which are distinctly important for the stability of some protein domains [22].

GB1 does not have any salt bridges in solution [226]. In the mechanical clamp region it has three charged amino acids, namely $\mathrm{Lys}^4$, $\mathrm{Lys}^{50}$ and $\mathrm{Glu}^{56}$, the latter with a $pK_a$ of 4.51 (see Table 3 in Ref. [226]) and therefore largely protonated at pH 3.5 and 4, suggesting that endogenous charged residues of GB1 do not play a major mechanical stabilisation role of GB1 at low pH, in agreement with the findings of Ref. [180]. In fact, a 2 M salt concentration was found to thermodynamically stabilise a structure-preserving GB1 variant at low pH (but not at high pH) [139]. On the basis of these considerations we can rule out weakening electrostatic interactions as the main reason behind TMAO-induced mechanical destabilisation of GB1 at low pH.

In order to investigate the relationship of cationic TMAO to the peptide backbone we employed a statistical mechanics solvation theory for osmolyte-backbone interactions. According to this model cationic TMAO is still preferentially excluded from the peptide backbone, although to a lesser extent than zwitterionic TMAO. It is known that protecting osmolytes stabilise proteins due to a fine balance between their preferential exclusion from the protein backbone and their preferential interaction with the protein side-chains that is tilted in favour of the former [202]. Seen in this context, the slightly weaker preferential exclusion of cationic TMAO from the protein backbone may result in a net destabilising effect on proteins, in agreement with the ansatz of Ref. [21].

However, in the absence of experimental transfer free energies of the peptide backbone unit in the presence of TMAO at low pH, we cannot rule out the hypothesis that cationic TMAO may preferentially interact, with respect to water, with the peptide backbone. In fact, one important objective of our study is strenghtning the case for such measurements to be performed.

Lastly, in the course of this study we have developed a series of technological innovations such as *a*) a method to numerically compute the expected number density of unfolding events by force for the velocity-clamp pulling protocol without the need for Monte-Carlo simulations (see Eq. 7.8), even in the presence of measurement noise, *b*) an ML method to fit Bell's model to velocity-clamp data taking into account measurement noise (see Section 7.5.1 and Section 7.5.1), and *c*) software tools for peak detection, WLC fitting, curve filtering and kinetic analysis of AFM force-spectroscopy experiments, all of which are made available with a friendly open-source licence as part of the software package Refolding [18]. Moreover, our observation that the thermodynamics of the backbone solvation theory of Ref. [1] can be computed in linear time in the number of interaction sites of both the osmolyte molecule and the peptide backbone (see Eq. 7.15) may afford more accurate future calculations of osmolyte-backbone electrostatics.

## 7.5 Methods

### 7.5.1 Expected number density of unfolding events by force

In this section we develop a simulation-free method to numerically compute the expected number density of unfolding events as a function of force and to estimate the kinetic parameters of a protein module, even in the presence of measurement noise sampled from known truncated normal distribution.

### Theory

Let us assume that we are stretching $m \geq 1$ identical protein modules in series with the force-protocol $f(t)$ starting at time instant $t_s$. Assuming that each module has the force-dependent unfolding rate $k_u(f)$, the survival probability of the construct reads as (see Eq. 3 in Ref. [17])

$$s^m_{t_s}(t) = exp\left[-m\int_{t_s}^{t} k_u(f(u))du\right]. \tag{7.1}$$

Further assuming that $f(t)$ is invertible and taking $f(t_s) = y_s$ we can easily change the domain of integration to force:

$$s^m_{y_s}(y) = \exp\left[-m\int_{y_s}^{y} k_u(z)(f^{-1})'(z)\,dz\right]. \tag{7.2}$$

Adding in the unfolding rate expression of Bell's model

$$k_u(f) = k_u(0)\,\exp[f\Delta x_u/(k_BT)]. \tag{7.3}$$

we obtain the survival probability

$$s^m_{y_s}(y) = \exp\left[-m\,k_u(0)\int_{y_s}^{y} \exp[z\Delta x_u/(k_BT)](f^{-1})'(z)\,dz\right] \tag{7.4}$$

and the density probability to observe an unfolding event becomes

$$r_{y_s}(m,y) = -ds^m_{y_s}(y)/dy = m\,k_u(0)\exp[y\Delta x_u/(k_BT)](f^{-1})'(y)s^m_{y_s}(y). \tag{7.5}$$

For the common case of a WLC polymer pulled at constant velocity $v$ with a cantilever of spring constant $k$, the following analytical approximation accurate to within 3.5% is available for the derivative of the inverse of the force-time function (see Eq. 4 in Ref. [230]):

$$(f^{-1})'(y) = [1/(kL) + h(y)]/(v/L), \tag{7.6}$$

where $k$ is the spring constant, $L$ is the contour length of the stretched construct and $h(y)$ is a function that depends on neither $k$, $v$ or $L$:

$$h(y) = 2[1 + y/(k_BT/p)]/\{(k_BT/p)[3 + 5y/(k_BT/p) + 8y^{5/2}/(k_BT/p)^{5/2}]\}, \tag{7.7}$$

with $p$ standing for the persistence length of the protein chain.

Now let's consider $n \geq 1$ such events indexed by the subscript variable $1 \leq i \leq n$. Then the expected number density of unfolding events by force (i.e., the ideal histogram) is

$$\mathcal{C}(y) = \sum_{i=1}^{n} r_{y_s}(m, y). \tag{7.8}$$

For compiling expected histograms of unfolding forces under velocity-clamp conditions, Monte-Carlo simulation has been so far the main technique spanning from the first AFM studies [40] to the more recent ones [231]. Taken together, Eq. 7.5, Eq. 7.6, Eq. 7.7 and Eq. 7.8 constitute a simulation-free method to numerically compute the ideal histogram. Together with ML estimation based on the likelihood function (see Eq. 4 in Ref. [17])

$$\mathcal{L}(\Delta x_u, k_u(0)) = \prod_{i=1}^{n} r_{y_{si}}(m_i, y_i) \tag{7.9}$$

where the parameters $p_i$, $k_i$, $v_i$ and $L_i$ are implicitly considered, the above equations make up a complete solution for simulation-free fitting of the kinetic parameters $\Delta x_u$ and $k_u(0)$. Section 7.5.1 presents a way to greatly speed-up the computation of Eq. 7.5 with a piecewise linear approximation of Eq. 7.7.

Now let us consider that the force measurement of an unfolding event may be affected by error sampled from an appropriately truncated normal distribution $\mathcal{N}(0, q^2)$. Then the probability density of Eq. 7.5 must be convoluted as follows:

$$c_{y_s}(m, w) = \int_{y_s}^{\infty} r_{y_s}(m, y) g_{w-y_s}(w - y)\, dy, \tag{7.10}$$

where $g_b(\cdot)$ is the density function of the normal distribution $\mathcal{N}(0, q^2)$ truncated at the upper bound $b$:

$$g_b(x) = \phi(x/q)/[q\, \Phi(b/q)]. \tag{7.11}$$

Here, $\phi(\xi) = \exp(-\xi^2/2)/\sqrt{2\pi}$ is the probability density function of the standard normal distribution and $\Phi(\cdot)$ is its cumulative distribution function.

The expected number density of unfolding events by force and the likelihood function in this case are the sum respectively the product of the convoluted functions.

### Discrete approximation

Let us consider the definite integral of Eq. 7.5 on an interval $[z_1, z_2]$:

$$I(z_1, z_2, h) = \int_{z_1}^{z_2} \exp[y\Delta x_u/(k_B T)][1/(kL) + h(y)]/(v/L)\, dy. \tag{7.12}$$

We can discretise all the forces onto a finite-size set and then precompute the integral of Eq. 7.12 on the interval between each two consecutive points. Being independent of $k$, $v$ or $L$, these integrals can then be reused for every unfolding event for which the stretching starts at a (discrete) force $y_s \leq z_1$ and the unfolding happens at a (discrete) force $y \geq z_2$.

A piecewise linear interpolation of $h(\cdot)$ on the finite-size set of forces takes the following form on each interval between two consecutive points $[z_1, z_2]$ in the finite-size set of forces:

$$\tilde{h}(z) = [h(z_1)z_2 - h(z_2)z_1]/(z_2 - z_1) + [h(z_2) - h(z_1)]/(z_2 - z_1)z,\ z_1 \leq z \leq z_2. \tag{7.13}$$

For each linear segment $[z_1, z_2]$ of Eq. 7.13 the integral of Eq. 7.12 has the following analytical form considering $\tilde{h}(\cdot)$ to substitute $h(\cdot)$:

$$\begin{aligned}
&I(z_1, z_2, \tilde{h})(v/L)\Delta x_u/(k_B T) = \\
&\{\exp[z_2\Delta x_u/(k_B T)] - \exp[z_1\Delta x_u/(k_B T)]\}/(kL) \\
&+ \{\exp[z_2\Delta x_u/(k_B T)] - \exp[z_1\Delta x_u/(k_B T)]\}[h(z_1)z_2 - h(z_2)z_1]/(z_2 - z_1) \\
&+ \{\exp[z_2\Delta x_u/(k_B T)](z_2 - k_B T/\Delta x_u) - \exp[z_1\Delta x_u/(k_B T)](z_1 - k_B T/\Delta x_u)\} \\
&\times [h(z_2) - h(z_1)]/(z_2 - z_1).
\end{aligned} \tag{7.14}$$

For taking into account the experimental noise, it is of course convenient to discretise the truncated distribution of Eq. 7.11 on the same finite-size set of forces.

## 7.5.2 Protein constructs

GB1-polyprotein-encoding plasmid pQE80-pQE80-$(\mathrm{GB1})_8$ was obtained as previously described [49]. Briefly, the $(\mathrm{GB1})_8$ gene was obtained by iterative cloning, overexpressed in the DH5$\alpha$ strain of E. coli and purified by $\mathrm{Ni}^{2+}$ chromatography.

### 7.5.3 AFM experiments and analysis parameters

Velocity-clamp experiments were performed with a Picoforce AFM equipped with a DI Multimode Nanoscope IIIa controller (Bruker) and gold-coated V-shaped silicon nitride cantilevers (NPG model; Bruker) with a nominal spring constant of 0.06 N/m. The temperature inside the cell was about 301.15 K. Cantilevers were calibrated via thermal tuning according to the equipartition theorem [117].

In order to avoid the effects of both intrinsic [117] and viscosity-induced [11, 115, 116] thermal tuning errors that may hamper the ability to distinguish small shifts in unfolding force distributions in different solvent conditions, we used each cantilever at one single pH value, changing the buffer back-and-forth between TMAO 0M and TMAO 2M, each time performing thermal tuning calibration to check that the cantilever properties did not suffer drastic changes in time, but using just one thermal tuning result per cantilever during data analysis.

For the viscosity adjustments in Fig. 7.10 we considered viscosity to be independent of pH and took $\eta(0M) = 0.890$ and $\eta(2M) = 1.884$, obtained by digitising Fig. 4 from Ref. [232].

Since the zero-force baseline noise is likely to also contribute to the force measurement error, we estimated the standard deviation of the experimental noise $q$ (see Eq. 7.11) as the standard deviation of the force measured after molecule detachment in the same force-curve. For the discrete approximation of Section 7.5.1 we used force-multiples of 1 pN.

### 7.5.4 Using the statistical mechanics model of Ref. [1]

Eq. 6 of Ref. [1] computed the average energy as a function of the osmolyte molar concentration $Y$ by enumerating all possible states of the osmolyte-peptide backbone unit system. Since the three interaction sites on the peptide backbone unit, namely positive N and negative $O_1$ and $O_2$, are independent, the same average energy can also be computed simply as the sum of the average interaction energy at each of the three sites on the protein backbone:

$$
\begin{aligned}
\langle E_N(Y)\rangle &= \frac{SA_+(Y)\exp[-H/(k_BT)]H - SA_-(Y)\exp[H/(k_BT)]H}{SA_+(Y)\exp[-H/(k_BT)] + SA_-(Y)\exp[H/(k_BT)] + SA_0(Y)} \\
\langle E_{O_1}(Y)\rangle &= \frac{-SA_+(Y)\exp[H/(k_BT)]H + SA_-(Y)\exp[-H/(k_BT)]H}{SA_+(Y)\exp[H/(k_BT)] + SA_-(Y)\exp[-H/(k_BT)] + SA_0(Y)} \\
\langle E_{O_2}(Y)\rangle &= \langle E_{O_1}(Y)\rangle \\
\langle E(Y)\rangle &= \langle E_N(Y)\rangle + \langle E_{O_1}(Y)\rangle + \langle E_{O_2}(Y)\rangle \qquad (7.15)
\end{aligned}
$$

where $SA_\bullet$ is defined as in Eq. 2, Eq. 3 and Eq. 4 of Ref. [1]:

$$SA_\bullet(Y) = 55.5\, SA_{w,\bullet} + Y\, SA_{o,\bullet}. \qquad (7.16)$$

The free energy of transfer of a peptide backbone unit into an 1 M osmolyte solution is simply the difference $\Delta g_{tr} = \langle E(1)\rangle - \langle E(0)\rangle$. It should also be noted that Eq. 7.15 is trivially extendable with linear time complexity in both the number of interaction sites on the peptide backbone unit and on the osmolyte molecule.

## 7.6 Acknowledgments

The authors thank Prof. Hongbin Li, University of British Columbia, Vancouver, Canada for kindly providing the $(\mathrm{GB1})_8$ plasmid.

# Chapter 8

# Conclusions

Mechanical processes play central roles in nearly every single aspect of life such as chromosomal dynamics during the cell cycle, protein translocation, cell movements, assisted protein folding and unfolding and enzyme catalysis. Hoping to elucidate the underlying molecular mechanisms, SMFS has become the technique of choice to study the effect of mechanical tension on these processes when applied according to predefined protocols. The measured effect of mechanical force on a biochemical process can reveal valuable information on the reaction free energy surface via suitable theoretical models. The mechanical stability of macromolecules in particular cannot generally be predicted from their thermodynamic stability alone, and therefore it is inaccessible to bulk experiments and it must be characterised by direct single-molecule mechanical experiments [74].

We applied such measurements to study the effect of ubiquitous osmolytes on protein energy landscapes, bringing manifold contributions to the biophysics field. First, we have tackled the technical difficulties of performing and analysing SMFS experiments by developing an improved method to fit and compute theoretical unfolding force distributions and refolding probabilities, and implemented such codes in open-source software [17, 18]. Second, we experimentally demonstrated that protecting osmolytes do not bridge the separating beta strands of a protein under mechanical tension, explaining instead our experimental results by applying the osmophobic effect applied to the native, transition and denatured protein states [12]. Third, we incorporated the osmophobic effect into a protein Ising model that affords exact thermodynamic calculations [19] and fast kinetic simulations [20].

Fourth, we found that an effect other than the osmophobic one dominates the protein energetics in the presence of an ubiquitous protecting osmolyte at low pH.

By combining mathematical, computational and experimental techniques this work has greatly enhanced our understanding of the underlying molecular details of osmophobic stabilisation and has opened up new avenues for future studies of more specific osmolyte-protein interactions.

# Bibliography

[1] Timothy O. Street, D. Wayne Bolen, and George D. Rose. A molecular mechanism for osmolyte-induced protein stability. *Proc. Natl. Acad. Sci. USA.*, 103(38):13997–4002, 2006.

[2] Y. Roiter and S. Minko. AFM single molecule experiments at the solid-liquid interface: in situ conformation of adsorbed flexible polyelectrolyte chains. *J. Am. Chem. Soc.*, 127:15688–15689, Nov 2005.

[3] Sanjay Kumar and Mai Suan Li. Biomolecules under mechanical force. *Phys. Rep.*, 486(1-2):1 – 74, 2010.

[4] F. Oesterhelt, D. Oesterhelt, M. Pfeiffer, A. Engel, H. E. Gaub, and D. J. Muller. Unfolding pathways of individual bacteriorhodopsins. *Science*, 288(5463):143–146, 2000.

[5] Massimo Sandal, Fabio Grandi, and Bruno Samorì. Single molecule force spectroscopy discovers mechanochemical switches in biology: The case of the disulfide bond. *Polymer*, 47(7):2571–2579, 2006.

[6] Rafael D. Gonzalez-Cruz, Vera C. Fonseca, and Eric M. Darling. Cellular mechanical properties reflect the differentiation potential of adipose-derived mesenchymal stem cells. *Proc Natl Acad Sci U S A*, 109(24):E1523–1529, 2012.

[7] Jeffrey L. Hutter and John Bechhoefer. Calibration of atomic-force microscope tips. *Review of Scientific Instruments*, 64(7):1868–1873, 1993.

[8] Florin E.-L., Rief M., Lehmann H., Ludwig M., Dornmair C., Moy V.T., and Gaub H.E. Sensing specific molecular interactions with the atomic force microscope. *Biosens. Bioelectron.*, 10(9):895–901(7), 1995.

[9] John E. Sader, James W. M. Chon, and Paul Mulvaney. Calibration of rectangular atomic force microscope cantilevers. *Review of Scientific Instruments*, 70(10):3967–3969, 1999.

[10] Christopher T Gibson, Gregory S Watson, and Sverre Myhra. Determination of the spring constants of probes for force microscopy/spectroscopy. *Nanotechnology*, 7(3):259, 1996.

[11] Tobias Pirzer and Thorsten Hugel. Atomic force microscopy spring constant determination in viscous liquids. *Rev. Sci. Instrum.*, 80(3):035110, 2009.

[12] Daniel Aioanei, Shanshan Lv, Isabella Tessari, Aldo Rampioni, Luigi Bubacco, Hongbin Li, Bruno Samorì, and Marco Brucale. Single-molecule-level evidence for the osmophobic effect. *Angew. Chem., Int. Ed.*, 50(19):4394–4397, 2011.

[13] D. W. Bolen and I. V. Baskakov. The osmophobic effect: natural selection of a thermodynamic force in protein folding. *J Mol Biol*, 310(5):955–963, 2001.

[14] Lovy Pradeep and Jayant B. Udgaonkar. Osmolytes induce structure in an early intermediate on the folding pathway of barstar. *J Biol Chem*, 279(39):40303–13, 2004.

[15] Lorna Dougan, Gang Feng, Hui Lu, and Julio M. Fernandez. Solvent molecules bridge the mechanical unfolding transition state of a protein. *Proc. Natl. Acad. Sci. USA*, 105(9):3185–3190, 2008.

[16] Lorna Dougan, Georgi Z. Genchev, Hui Lu, and Julio M. Fernandez. Probing osmolyte participation in the unfolding transition state of a protein. *Proc. Natl. Acad. Sci. USA*, 108(24):9759–9764, 2011.

[17] Daniel Aioanei, Bruno Samorì, and Marco Brucale. Maximum likelihood estimation of protein kinetic parameters under weak assumptions from un-

folding force spectroscopy experiments. *Phys. Rev. E*, 80(6):061916, Dec 2009.

[18] Daniel Aioanei, Marco Brucale, and Bruno Samorì. Open source platform for the execution and analysis of mechanical refolding experiments. *Bioinformatics*, 27(3):423–425, 2011.

[19] Daniel Aioanei, Isabella Tessari, Luigi Bubacco, Bruno Samorì, and Marco Brucale. Observing the osmophobic effect in action at the single molecule level. *Proteins: Struct., Funct., Bioinf.*, 79(7):2214–2223, 2011.

[20] Daniel Aioanei, Marco Brucale, Isabella Tessari, Luigi Bubacco, and Bruno Samorì. Worm-like Ising model for protein mechanical unfolding under the effect of osmolytes. *Biophys J*, 102(2):342–350, 2012.

[21] Rajendrakumar Singh, Inamul Haque, and Faizan Ahmad. Counteracting osmolyte trimethylamine N-oxide destabilizes proteins at pH below its pKa. Measurements of thermodynamic parameters of proteins in the presence and absence of trimethylamine N-oxide. *J Biol Chem*, 280(12):11035–42, 2005.

[22] Vincenzo Granata, Pasquale Palladino, Barbara Tizzano, Alessandro Negro, Rita Berisio, and Adriana Zagari. The effect of the osmolyte trimethylamine N-oxide on the stability of the prion protein at low pH. *Biopolymers*, 82(3):234–240, 2006.

[23] M. Rief, J. M. Fernandez, and H. E. Gaub. Elastically Coupled Two-Level Systems as a Model for Biopolymer Extensibility. *Phys. Rev. Lett.*, 81:4764–4767, November 1998.

[24] M. Carrion-Vazquez, A. F. Oberhauser, T. E. Fisher, P. E. Marszalek, H. Li, and J. M. Fernandez. Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Prog. Biophys. Mol. Biol.*, 74(1-2):63–91, 2000.

[25] T. E. Fisher, P. E. Marszalek, A. F. Oberhauser, M. Carrion-Vazquez, and J. M. Fernandez. The micro-mechanics of single molecules studied with atomic force microscopy. *J. Physiol.*, 520 Pt 1:5–14, 1999.

[26] Wolfgang A. Linke and Anika Grutzner. Pulling single molecules of titin by AFM–recent advances and physiological implications. *Pflugers Archiv : Eur. J. Physiol.*, 456(1):101–105, 2008.

[27] T. E. Fisher, A. F. Oberhauser, M. Carrion-Vazquez, P. E. Marszalek, and J. M. Fernandez. The study of protein mechanics with the atomic force microscope. *Trends Biochem. Sci*, 24(10):379–384, 1999.

[28] K. Wang, J. G. Forbes, and A. J. Jin. Single molecule measurements of titin elasticity. *Prog. Biophys. Mol. Biol.*, 77(1):1–44, 2001.

[29] Andres F. Oberhauser and Mariano Carrion-Vazquez. Mechanical biochemistry of proteins one molecule at a time. *J. Biol. Chem.*, 283(11):6617–6621, 2008.

[30] T. E. Fisher, P. E. Marszalek, and J. M. Fernandez. Stretching single molecules into novel conformations using the atomic force microscope. *Nat. Struct. Biol.*, 7(9):719–724, 2000.

[31] Ignacio Tinoco Jr., Pan T. X. Li, and Carlos Bustamante. Determination of thermodynamics and kinetics of RNA reactions by force. *Q. Rev. Biophys.*, 39(4):325–330, 2006.

[32] Ignacio Tinoco Jr. Force as a useful variable in reactions: unfolding RNA. *Annu. Rev. Biophys. Biomol. Struct.*, 33:363–365, 2004.

[33] T. E. Fisher, M. Carrion-Vazquez, A. F. Oberhauser, H. Li, P. E. Marszalek, and J. M. Fernandez. Single molecular force spectroscopy of modular proteins in the nervous system. *Neuron*, 27(3):435–436, 2000.

[34] T. R. Strick, J. F. Allemand, D. Bensimon, and V. Croquette. Stress-induced structural transitions in DNA and proteins. *Annu. Rev. Biophys. Biomol. Struct.*, 29:523–523, 2000.

[35] A. Janshoff, M. Neitzert, Y. Oberdorfer, and H. Fuchs. Force Spectroscopy of Molecular Systems-Single Molecule Spectroscopy of Polymers and Biomolecules. *Angew. Chem.*, 39(18):3212–3237, 2000.

[36] E. Evans and K. Ritchie. Dynamic strength of molecular adhesion bonds. *Biophys. J.*, 72(4):1541–1545, 1997.

[37] G. I. Bell. Models for the specific adhesion of cells to cells. *Science*, 200(4342):618–627, 1978.

[38] M. Raible, M. Evstigneev, F. W. Bartels, R. Eckel, M. Nguyen-Duong, R. Merkel, R. Ros, D. Anselmetti, and P. Reimann. Theoretical analysis of single-molecule force spectroscopy experiments: heterogeneity of chemical bonds. *Biophys. J.*, 90(11):3851–3854, 2006.

[39] Mykhaylo Evstigneev and Peter Reimann. Dynamic force spectroscopy: Optimized data analysis. *Phys. Rev. E*, 68(4):045103(R), Oct 2003.

[40] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 276(5315):1109–1112, 1997.

[41] M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez. Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl. Acad. Sci. U.S.A.*, 96(7):3694–3699, 1999.

[42] Robert B. Best and Jane Clarke. What can atomic force microscopy tell us about protein folding? *Chem. Commun. (Cambridge, U. K.)*, (3):183–192, 2002.

[43] Ross Rounsevell, Julia R. Forman, and Jane Clarke. Atomic force microscopy: mechanical unfolding of proteins. *Methods*, 34(1):100 – 111, 2004.

[44] Robert B. Best, Susan B. Fowler, Jose L. Toca-Herrera, and Jane Clarke. A simple method for probing the mechanical unfolding pathway of proteins in detail. *Proc. Natl. Acad. Sci. U.S.A.*, 99(19):12143–8, 2002.

[45] Andres F. Oberhauser, Piotr E. Marszalek, Harold P. Erickson, and Julio M. Fernandez. The molecular elasticity of the extracellular matrix protein tenascin. *Nature*, 393(6681):181–185, 1998.

[46] Sebastian Getfert and Peter Reimann. Optimal evaluation of single-molecule force spectroscopy experiments. *Phys. Rev. E*, 76(5 Pt 1):052901, 2007.

[47] Alexander Fuhrmann, Dario Anselmetti, Robert Ros, Sebastian Getfert, and Peter Reimann. Refined procedure of evaluating experimental single-molecule force spectroscopy data. *Phys. Rev. E*, 77(3 Pt 1):031912, 2008.

[48] M. Carrion-Vazquez, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez. Atomic force microscopy captures length phenotypes in single proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 96(20):11288–92, 1999.

[49] Yi Cao and Hongbin Li. Polyprotein of GB1 is an ideal artificial elastomeric protein. *Nat. Mater.*, 6(2):109–114, 2007.

[50] Thorsten Hugel, Matthias Rief, Markus Seitz, Hermann E. Gaub, and Roland R. Netz. Highly stretched single polymers: Atomic-force-microscope experiments versus ab-initio theory. *Phys. Rev. Lett.*, 94(4):048301, Jan 2005.

[51] Elias M. Puchner, Gereon Franzen, Mathias Gautel, and Hermann E. Gaub. Comparing proteins by their unfolding pattern. *Biophys. J.*, 95(1):426–434, 2008.

[52] R. Merkel. Force spectroscopy on single passive biomolecules and single biomolecular bonds. *Phys. Rep.*, 346:343–385, jun 2001.

[53] Shuxun Cui, You Yu, and Zhangbi Lin. Modeling single chain elasticity of single-stranded dna: A comparison of three models. *Polymer*, 50(3):930 – 935, 2009.

[54] C. Bouchiat, M. D. Wang, J. Allemand, T. Strick, S. M. Block, and V. Croquette. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys. J.*, 76(1 Pt 1):409–413, 1999.

[55] Ray W. Ogden, Giuseppe Saccomandi, and Ivonne Sgura. Computational aspects of worm-like-chain interpolation formulas. *Comput. Math. Appl.*, 53(2):276–286, 2007.

[56] Robert B. Best, David J. Brockwell, Jos L. Toca-Herrera, Anthony W. Blake, D. Alastair Smith, Sheena E. Radford, and Jane Clarke. Force mode atomic force microscopy as a tool for protein folding studies. *Anal. Chim. Acta*, 479(1):87 – 105, 2003.

[57] Paul Garthwaite, Ian Jolliffe, and Byron Jones. *Statistical inference*. Oxford University Press, USA, second edition, 2002.

[58] John F. Marko and Eric D. Siggia. Stretching DNA. *Macromolecules*, 28(26):8759–8770, 1995.

[59] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith. Entropic elasticity of lambda-phage DNA. *Science*, 265(5178):1599–1600, 1994.

[60] Lorna Dougan, Ainavarapu Sri Rama Koti, Georgi Genchev, Hui Lu, and Julio M. Fernandez. A single-molecule perspective on the role of solvent hydrogen bonds in protein folding and chemical reactions. *Chemphyschem*, 9(18):2836–2837, 2008.

[61] Robert B. Best, Susan B. Fowler, Jose L. Toca Herrera, Annette Steward, Emanuele Paci, and Jane Clarke. Mechanical unfolding of a titin Ig domain: structure of transition state revealed by combining atomic force microscopy, protein engineering and molecular dynamics simulations. *J. Mol. Biol.*, 330(4):867–867, 2003.

[62] Sean P. Ng, Ross W. S. Rounsevell, Annette Steward, Christian D. Geierhaas, Philip M. Williams, Emanuele Paci, and Jane Clarke. Mechanical unfolding of TNfn3: the unfolding pathway of a fnIII domain probed by protein engineering, AFM and MD simulation. *J. Mol. Biol.*, 350(4):776–779, 2005.

[63] Yi Cao and Hongbin Li. How do chemical denaturants affect the mechanical folding and unfolding of proteins? *J. Mol. Biol.*, 375(1):316–324, 2008.

[64] Dan Gusfield. *Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology*, chapter 12. Cambridge University Press, Cambridge, UK, first edition, 1997.

[65] Michael L Fredman. On computing the length of longest increasing subsequences. *Discrete Math*, 11:29 – 35, 1975.

[66] Anirban DasGupta. *Asymptotic Theory of Statistics and Probability*. Springer Science+Business Media, LLC, New York, NY, USA, first edition, 2008.

[67] G. A. Young and R. L. Smith. *Essentials of Statistical Inference*. Cambridge University Press, Cambridge, UK, first edition, 2005.

[68] Bruce Hoadley. Asymptotic Properties of Maximum Likelihood Estimators for the Independent Not Identically Distributed Case. *Ann. Math. Statist.*, 42(6):1977–1991, dec 1971.

[69] Yi Cao, Canaan Lam, Meijia Wang, and Hongbin Li. Nonmechanical protein can have significant mechanical stability. *Angew. Chem.*, 45(4):642–645, 2006.

[70] Hongbin Li. Engineering proteins with tailored nanomechanical properties: a single molecule approach. *Org. Biomol. Chem.*, 5(21):3399–3406, 2007.

[71] Massimo Sandal, Fabrizio Benedetti, Marco Brucale, Alberto Gomez-Casado, and Bruno Samorì. Hooke: an open software platform for force spectroscopy. *Bioinformatics*, 25(11):1428–1430, 2009.

[72] Mariano Carrion-Vazquez, Andres F. Oberhauser, Hector Diez, Ruben Hervas, Javier Oroz, Jesus Fernandez, and David Martinez-Martin. Protein nanomechanics - as studied by afm single-molecule force spectroscopy. In *Advanced Techniques in Biophysics*, pages 163–245. Springer-Verlag Berlin Heidelberg, Berlin, 2006.

[73] RWD Nickalls. A new approach to solving the cubic: Cardan's solution revealed. *The Mathematical Gazette*, 77:354–359, 1993.

[74] Carlos Bustamante, Yann R. Chemla, Nancy R. Forde, and David Izhaky. Mechanical processes in biochemistry. *Annual Review of Biochemistry*, 73(1):705–748, 2004.

[75] Ashlee Jollymore and Hongbin Li. Measuring "unmeasurable" folding kinetics of proteins by single-molecule force spectroscopy. *Journal of molecular biology*, 402(3):610–617, 2010.

[76] Sergi Garcia-Manyes, Lorna Dougan, and Julio M. Fernandez. Osmolyte-induced separation of the mechanical folding phases of ubiquitin. *Proc. Natl. Acad. Sci. USA.*, 106(26):10540–5, 2009.

[77] Sergi Garcia-Manyes, Jasna Brujic, Carmen L. Badilla, and Julio M. Fernandez. Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of I27 and ubiquitin. *Biophysical journal*, 93(7):2436–2436, 2007.

[78] Julio M. Fernandez and Hongbin Li. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science*, 303(5664):1674–1678, 2004.

[79] Sergi Garcia-Manyes, Lorna Dougan, Carmen L. Badilla, Jasna Brujic, and Julio M. Fernandez. Direct observation of an ensemble of stable collapsed states in the mechanical folding of ubiquitin. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26):10534–9, 2009.

[80] Belinda Bullard, Tzintzuni Garcia, Vladimir Benes, Mark C. Leake, Wolfgang A. Linke, and Andres F. Oberhauser. The molecular elasticity of the insect flight muscle proteins projectin and kettin. *Proceedings of the National Academy of Sciences of the United States of America*, 103(12):4451–4456, 2006.

[81] A. F. Oberhauser, P. K. Hansma, M. Carrion-Vazquez, and J. M. Fernandez. Stepwise unfolding of titin under force-clamp atomic force microscopy. *Proc. Natl. Acad. Sci. U.S.A.*, 98(2):468–472, 2001.

[82] Sam Walcott. The load dependence of rate constants. *J. Chem. Phys.*, 128(21):215101, 2008.

[83] Peter Hänggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: fifty years after kramers. *Rev. Mod. Phys.*, 62:251–341, Apr 1990.

[84] Erich Leo Lehmann and George Casella. *Theory of Point Estimation.* Springer-Verlag, New York, NY, USA, second edition, 1998.

[85] D. Wayne Bolen and George D. Rose. Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu. Rev. Biochem.*, 77:339–342, 2008.

[86] E. Papp and P. Csermely. Chemical chaperones: mechanisms of action and potential use. *Handb Exp Pharmacol*, (172):405–406, 2006.

[87] J. P. Morello, U. E. Petaja-Repo, D. G. Bichet, and M. Bouvier. Pharmacological chaperones: a new twist on receptor folding. *Trends Pharmacol Sci*, 21(12):466–469, 2000.

[88] George D. Rose, Patrick J. Fleming, Jayanth R. Banavar, and Amos Maritan. A backbone-based theory of protein folding. *Proc. Natl. Acad. Sci. USA.*, 103(45):16623–33, 2006.

[89] T. Y. Lin and S. N. Timasheff. Why do some organisms use a urea-methylamine mixture as osmolyte? Thermodynamic compensation of urea and trimethylamine N-oxide interactions with protein. *Biochemistry*, 33(42):12695–701, 1994.

[90] Matthew Auton and D. Wayne Bolen. Additive transfer free energies of the peptide backbone unit that are independent of the model compound and the choice of concentration scale. *Biochemistry*, 43(5):1329–1332, 2004.

[91] Jayanth R. Banavar and Amos Maritan. Physics of proteins. *Annu Rev Biophys Biomol Struct*, 36:261–270, 2007.

[92] Matthew Auton and D. Wayne Bolen. Application of the transfer model to understand how naturally occurring osmolytes affect protein stability. *Methods Enzymol*, 428:397–418, 2007.

[93] Sandeep K. Sharma, Philipp Christen, and Pierre Goloubinoff. Disaggregating chaperones: an unfolding story. *Curr Protein Pept Sci*, 10(5):432–436, 2009.

[94] Matthias P. Mayer. Gymnastics of molecular chaperones. *Mol Cell*, 39(3):321–321, 2010.

[95] Fu-Feng Liu, Xiao-Yan Dong, and Yan Sun. Molecular mechanism for the effects of trehalose on beta-hairpin folding revealed by molecular dynamics simulation. *J Mol Graph Model*, 27(4):421–429, 2008.

[96] C. P. Yiu, M. G. Mateu, and A. R. Fersht. Protein folding transition states: elicitation of Hammond effects by 2,2,2-trifluoroethanol. *Chembiochem*, 1(1):49–55, 2000.

[97] Hongbin Li, Hui-Chuan Wang, Yi Cao, Deepak Sharma, and Meijia Wang. Configurational entropy modulates the mechanical stability of protein GB1. *J Mol Biol*, 379(4):871–880, 2008.

[98] Shanshan Lv, Daniel M. Dudek, Yi Cao, M. M. Balamurali, John Gosline, and Hongbin Li. Designed biomaterials to mimic the mechanical properties of muscles. *Nature*, 465(7294):69–73, 2010.

[99] Sung Lun Lin, Arash Zarrine-Afsar, and Alan R. Davidson. The osmolyte trimethylamine-N-oxide stabilizes the Fyn SH3 domain without altering the structure of its folding transition state. *Protein Sci*, 18(3):526–526, 2009.

[100] Emanuele Paci, Jane Clarke, Annette Steward, Michele Vendruscolo, and Martin Karplus. Self-consistent determination of the transition state for protein folding: application to a fibronectin type III domain. *Proc Natl Acad Sci U S A*, 100(2):394–399, 2003.

[101] E. R. Main and S. E. Jackson. Does trifluoroethanol affect folding pathways and can it be used as a probe of structure in transition states? *Nat Struct Biol*, 6(9):831–835, 1999.

[102] J.M.G. Cowie and P.M. Toporowski. Association in binary liquid system dimethyl sulphoxidewater. *Can. J. Chem.*, 39:22402243, 1961.

[103] Jos Via. Biochemical adaptation: Mechanism and process in physiological evolution. *Biochemistry and Molecular Biology Education*, 30(3):215–216, 2002.

[104] D. Wayne Bolen. Protein stabilization by naturally occurring osmolytes. 168, April 2001.

[105] P. H. Yancey, M. E. Clark, S. C. Hand, R. D. Bowlus, and G. N. Somero. Living with water stress: evolution of osmolyte systems. *Science*, 217(4566):1214–1222, 1982.

[106] Raj Kumar. Role of naturally occurring osmolytes in protein folding and stability. *Arch Biochem Biophys*, 491(1-2):1–6, 2009.

[107] Luis Marcelo F. Holthauzen and D. Wayne Bolen. Mixed osmolytes: the degree to which one osmolyte affects the protein stabilizing ability of another. *Protein Sci*, 16(2):293–298, 2007.

[108] Peng Wu and D. W. Bolen. Osmolyte-induced protein folding free energy changes. *Proteins*, 63(2):290–296, 2006.

[109] Matthew Auton and D. Wayne Bolen. Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc. Natl. Acad. Sci. USA.*, 102(42):15065–8, 2005.

[110] I. Baskakov and D. W. Bolen. Forcing thermodynamically unfolded proteins to fold. *J Biol Chem*, 273(9):4831–4834, 1998.

[111] Maria Silow and Mikael Oliveberg. High concentrations of viscogens decrease the protein folding rate constant by prematurely collapsing the coil. *J Mol Biol*, 326(1):263–271, 2003.

[112] Atsushi Mukaiyama, Yuichi Koga, Kazufumi Takano, and Shigenori Kanaya. Osmolyte effect on the stability and folding of a hyperthermophilic protein. *Proteins*, 71(1):110–118, 2008.

[113] Andrew T. Russo, Jorg Rosgen, and D. W. Bolen. Osmolyte effects on kinetics of FKBP12 C22A folding coupled with prolyl isomerization. *J Mol Biol*, 330(4):851–856, 2003.

[114] Nian-Sheng Cheng. Formula for the viscosity of a glycerol-water mixture. *Ind. Eng. Chem. Res.*, 47(9):3285–3288, 2008.

[115] C. P. Green and J. E. Sader. Frequency response of cantilever beams immersed in viscous fluids near a solid surface with applications to the atomic force microscope. *J. Appl. Phys.*, 98(11):114913–+, December 2005.

[116] Christian Bergaud and Liviu Nicu. Viscosity measurements based on experimental investigations of composite cantilever beam eigenfrequencies in viscous media. *Rev. Sci. Instrum.*, 71(6):2487, 2000.

[117] Christopher T. Gibson, Daniel J. Johnson, Christopher Anderson, Chris Abell, and Trevor Rayment. Method to determine the spring constant of atomic force microscope cantilevers. *Rev. Sci. Instrum.*, 75(2):565–567, 2004.

[118] Runcong Liu, Marisa Roman, and Guoliang Yang. Correction of the viscous drag induced errors in macromolecular manipulation experiments using atomic force microscope. *Rev. Sci. Instrum.*, 81(6):063703, 2010.

[119] Olga I. Vinogradova, Hans-Jurgen Butt, Gleb E. Yakubov, and Francois Feuillebois. Dynamic effects on force measurements. I. Viscous drag on the atomic force microscope cantilever. *Rev. Sci. Instrum.*, 72(5):2330–2339, 2001.

[120] Harald Janovjak, Jens Struckmeier, and Daniel J. Muller. Hydrodynamic effects in fast AFM single-molecule force measurements. *Eur. Biophys. J.*, 34(1):91–96, 2005.

[121] S. Glasstone, K.J. Laidler, and H. Eyring. *The theory of rate processes: the kinetics of chemical reactions, viscosity, diffusion and electrochemical phenomena.* International chemical series. McGraw-Hill Book Company, inc., 1941.

[122] Hiroshi Wako and Nobuhiko Saitô. Statistical mechanical theory of the protein conformation. i. general considerations and the application to homopolymers. *J. Phys. Soc. Jpn.*, 44(6):1931–1938, 1978.

[123] Hiroshi Wako and Nobuhiko Saitô. Statistical mechanical theory of the protein conformation. ii. folding pathway for protein. *J. Phys. Soc. Jpn.*, 44(6):1939–1945, 1978.

[124] V. Munoz, P. A. Thompson, J. Hofrichter, and W. A. Eaton. Folding dynamics and mechanism of beta-hairpin formation. *Nature*, 390(6656):196–199, 1997.

[125] V. Munoz and W. A. Eaton. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl. Acad. Sci. USA.*, 96(20):11311–6, 1999.

[126] A. Imparato, A. Pelizzola, and M. Zamparo. Ising-like model for protein mechanical unfolding. *Phys. Rev. Lett.*, 98(14):148102, 2007.

[127] A. Imparato, A. Pelizzola, and M. Zamparo. Protein mechanical unfolding: a model with binary variables. *J. Chem. Phys.*, 127(14):145105, 2007.

[128] Pierpaolo Bruscolini and Alessandro Pelizzola. Exact solution of the Munoz-Eaton model for protein folding. *Phys. Rev. Lett.*, 88(25 Pt 1):258101, 2002.

[129] Pierpaolo Bruscolini, Alessandro Pelizzola, and Marco Zamparo. Downhill versus two-state protein folding in a statistical mechanical model. *J. Chem. Phys.*, 126(21):215103, 2007.

[130] A. Imparato and A. Pelizzola. Mechanical unfolding and refolding pathways of ubiquitin. *Phys. Rev. Lett.*, 100(15):158104, 2008.

[131] M. Caraglio, A. Imparato, and A. Pelizzola. Pathways of mechanical unfolding of FnIII(10): low force intermediates. *J. Chem. Phys.*, 133(6):065101, 2010.

[132] A. Wang and D. W. Bolen. A naturally occurring protective system in urea-rich cells: mechanism of osmolyte protection of proteins against urea denaturation. *Biochemistry*, 36(30):9101–9108, 1997.

[133] Y. Qu, C. L. Bolen, and D. W. Bolen. Osmolyte-driven contraction of a random coil protein. *Proc. Natl. Acad. Sci. USA.*, 95(16):9268–9273, 1998.

[134] Pau Bernado, Martin Blackledge, and Javier Sancho. Sequence-specific solvent accessibilities of protein residues in unfolded protein ensembles. *Biophys. J.*, 91(12):4536–4543, 2006.

[135] Jorge Estrada, Pau Bernado, Martin Blackledge, and Javier Sancho. ProtSA: a web application for calculating sequence specific protein solvent accessibilities in the unfolded ensemble. *BMC Bioinformatics*, 10:104, 2009.

[136] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The Protein Data Bank. *Nucleic Acids Res.*, 28(1):235–242, 2000.

[137] T. Gallagher, P. Alexander, P. Bryan, and G. L. Gilliland. Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry*, 33(15):4721–4729, 1994.

[138] Vincent B. Chen, W. Bryan 3rd Arendall, Jeffrey J. Headd, Daniel A. Keedy, Robert M. Immormino, Gary J. Kapral, Laura W. Murray, Jane S. Richardson, and David C. Richardson. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta. Crystallogr. D.*, 66(Pt 1):12–21, 2010.

[139] Stina Lindman, Wei-Feng Xue, Olga Szczepankiewicz, Mikael C. Bauer, Hanna Nilsson, and Sara Linse. Salting the charged surface: pH and salt dependence of protein G B1 stability. *Biophys. J.*, 90(8):2911–2911, 2006.

[140] Hoi Sung Chung, Kevin McHale, John M. Louis, and William A. Eaton. Single-molecule fluorescence experiments determine protein folding transition path times. *Science*, 335(6071):981–984, 2012.

[141] B. Odaert, F. Jean, C. Boutillon, E. Buisine, O. Melnyk, A. Tartar, and G. Lippens. Synthesis, folding, and structure of the beta-turn mimic modified B1 domain of streptococcal protein G. *Protein Sci*, 8(12):2773–2773, 1999.

[142] Hongbin Li and Yi Cao. Protein mechanics: from single molecules to functional biomaterials. *Acc Chem Res*, 43(10):1331–1331, 2010.

[143] Marina Ramirez-Alvarado, Melanie J. Cocco, and Lynne Regan. Mutations in the B1 domain of protein G that delay the onset of amyloid fibril formation in vitro. *Protein Sci*, 12(3):567–576, 2003.

[144] M. Ramirez-Alvarado, J. S. Merkel, and L. Regan. A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro. *Proc Natl Acad Sci U S A*, 97(16):8979–8984, 2000.

[145] A. M. Gronenborn, D. R. Filpula, N. Z. Essig, A. Achari, M. Whitlow, P. T. Wingfield, and G. M. Clore. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science*, 253(5020):657–661, 1991.

[146] P. Alexander, S. Fahnestock, T. Lee, J. Orban, and P. Bryan. Thermodynamic analysis of the folding of the streptococcal protein G IgG-binding domains B1 and B2: why small proteins tend to have high denaturation temperatures. *Biochemistry*, 31(14):3597–3603, 1992.

[147] P. Alexander, J. Orban, and P. Bryan. Kinetic analysis of folding and unfolding the 56 amino acid IgG-binding domain of streptococcal protein G. *Biochemistry*, 31(32):7243–7248, 1992.

[148] C. K. Smith, J. M. Withka, and L. Regan. A thermodynamic scale for the beta-sheet forming tendencies of the amino acids. *Biochemistry*, 33(18):5510–5517, 1994.

[149] C. K. Smith and L. Regan. Guidelines for protein design: the energetics of beta sheet side chain interactions. *Science*, 270(5238):980–982, 1995.

[150] S. H. Park, K. T. O'Neil, and H. Roder. An early intermediate in the folding reaction of the B1 domain of protein G contains a native-like core. *Biochemistry*, 36(47):14277–83, 1997.

[151] S. H. Park, M. C. Shastry, and H. Roder. Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing. *Nat Struct Biol*, 6(10):943–947, 1999.

[152] J. S. Merkel and L. Regan. Aromatic rescue of glycine in beta sheets. *Fold Des*, 3(6):449–455, 1998.

[153] J. S. Merkel, J. M. Sturtevant, and L. Regan. Sidechain interactions in parallel beta sheets: the energetics of cross-strand pairings. *Structure*, 7(11):1333–1333, 1999.

[154] F.H. Johnson. *The kinetic basis of molecular biology.* Wiley & Sons, 1954.

[155] Fan-Guo Meng, Yuan-Kai Hong, Hua-Wei He, Arkadii E. Lyubarev, Boris I. Kurganov, Yong-Bin Yan, and Hai-Meng Zhou. Osmophobic effect of glycerol on irreversible thermal denaturation of rabbit creatine kinase. *Biophys. J.*, 87(4):2247–2254, 2004.

[156] Jingyuan Li, Julio M. Fernandez, and B. J. Berne. Water's role in the force-induced unfolding of ubiquitin. *Proc. Natl. Acad. Sci. USA*, 107(45):19284–9, 2010.

[157] Marco Zamparo and Alessandro Pelizzola. Kinetics of the Wako-Saito-Munoz-Eaton model of protein folding. *Phys. Rev. Lett.*, 97(6):068106, 2006.

[158] V. Munoz, E. R. Henry, J. Hofrichter, and W. A. Eaton. A statistical mechanical model for beta-hairpin kinetics. *Proc. Natl. Acad. Sci. USA.*, 95(11):5872–5879, 1998.

[159] Pierpaolo Bruscolini and Athi N. Naganathan. Quantitative prediction of protein folding behaviors from a simple statistical model. *J. Am. Chem. Soc.*, 133(14):5372–5379, 2011.

[160] V I Tokar and H Dreyss. Transfer matrix solution of the wako-saitô-muñoz-eaton model augmented by arbitrary short range interactions. *J. Stat. Mech.: Theory Exp.*, 2010(08):P08028, 2010.

[161] M. Caraglio, A. Imparato, and A. Pelizzola. Direction-dependent mechanical unfolding and green fluorescent protein as a force sensor. *Phys. Rev. E*, 84(2-1):021918, 2011.

[162] Michael Schlierf, Felix Berkemeier, and Matthias Rief. Direct observation of active protein folding using lock-in force spectroscopy. *Biophys. J.*, 93(11):3989–3998, 2007.

[163] Elizabeth A. Shank, Ciro Cecconi, Jesse W. Dill, Susan Marqusee, and Carlos Bustamante. The folding cooperativity of a protein is controlled by its chain topology. *Nature*, 465(7298):637–640, 2010.

[164] Philipp Bechtluft, Ruud G. H. van Leeuwen, Matthew Tyreman, Danuta Tomkiewicz, Nico Nouwen, Harald L. Tepper, Arnold J. M. Driessen, and Sander J. Tans. Direct observation of chaperone-induced changes in a protein folding pathway. *Science*, 318(5855):1458–1461, 2007.

[165] Alessandro Pelizzola. Exactness of the cluster variation method and factorization of the equilibrium probability for the wako-saitô-muñoz-eaton model of protein folding. *J. Stat. Mech.: Theory Exp.*, 2005(11):P11010, 2005.

[166] S. Improta, A. S. Politou, and A. Pastore. Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure*, 4(3):323–327, 1996.

[167] S. R. Ainavarapu, J. Brujic, H. H. Huang, A. P. Wiita, H. Lu, L. Li, K. A. Walther, M. Carrion-Vazquez, H. Li, and J. M. Fernandez. Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophys. J.*, 92(1):225–233, 2007.

[168] Vasilios I. Manousiouthakis and Michael W. Deem. Strict detailed balance is unnecessary in Monte Carlo simulation. *J. Chem. Phys.*, 110(6):2753, 1999.

[169] Gavin E. Crooks. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E*, 61(3):2361–2366, Mar 2000.

[170] A. M. Mariz, F. D. Nobre, and C. Tsallis. Generalized single-spin-flip dynamics for the Ising model and thermodynamic properties. *Phys. Rev. B*, 49(5):3576–3579, 1994.

[171] Malvin H. Kalos and Paula A. Whitlock. *Monte Carlo Methods*. Wiley-VCH Verlag GmbH and Co. KGaA Weinheim, second edition, 2008.

[172] Tsutomu Arakawa, Yoshiko Kita, and Serge N. Timasheff. Protein precipitation and denaturation by dimethyl sulfoxide. *Biophys. Chem.*, 131(1-3):62 – 70, 2007.

[173] Alexander W. Schuttelkopf and Daan M. F. van Aalten. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta. Crystallogr. D.*, 60(Pt 8):1355–1363, 2004.

[174] W. L. Delano. The PyMOL Molecular Graphics System. 2002.

[175] Jihua Wang, Zhiyong Zhang, Haiyan Liu, and Yunyu Shi. Quasiequilibrium unfolding thermodynamics of a small protein studied by molecular dynamics simulation with an explicit water model. *Phys. Rev. E*, 67(6 Pt 1):061903, 2003.

[176] A. S. Politou, D. J. Thomas, and A. Pastore. The folding and stability of titin immunoglobulin-like modules, with implications for the mechanism of elasticity. *Biophys. J.*, 69(6):2601–2610, 1995.

[177] S. B. Fowler and J. Clarke. Mapping the folding pathway of an immunoglobulin domain: structural detail from Phi value analysis and movement of the transition state. *Structure*, 9(5):355–356, 2001.

[178] Hongbin Li, Wolfgang A. Linke, Andres F. Oberhauser, Mariano Carrion-Vazquez, Jason G. Kerkvliet, Hui Lu, Piotr E. Marszalek, and Julio M. Fernandez. Reverse engineering of the giant muscle protein titin. *Nature*, 418(6901):998–1002, 2002.

[179] Yi Cao, Yongnan Devin Li, and Hongbin Li. Enhancing the Mechanical Stability of Proteins through a Cocktail Approach. *Biophys. J.*, 100(7):1794–1799, 2011.

[180] Peng Zheng, Yi Cao, Tianjia Bu, Suzana K. Straus, and Hongbin Li. Single molecule force spectroscopy reveals that electrostatic interactions affect the mechanical stability of proteins. *Biophys. J.*, 100(6):1534–1541, 2011.

[181] Yukinori Taniguchi, David J. Brockwell, and Masaru Kawakami. The effect of temperature on mechanical resistance of the native and intermediate states of I27. *Biophys. J.*, 95(11):5296–5305, 2008.

[182] Stephen J. Hagen. Solvent viscosity and friction in protein folding dynamics. *Curr. Protein Pept. Sci.*, 11(5):385–385, 2010.

[183] B. A. Chrunyk and C. R. Matthews. Role of diffusion in the folding of the alpha subunit of tryptophan synthase from Escherichia coli. *Biochemistry*, 29(8):2149–2154, 1990.

[184] M. Jacob, M. Geeves, G. Holtermann, and F. X. Schmid. Diffusional barrier crossing in a two-state protein folding reaction. *Nat. Struct. Biol.*, 6(10):923–926, 1999.

[185] K. W. Plaxco and D. Baker. Limited internal friction in the rate-limiting step of a two-state protein folding reaction. *Proc. Natl. Acad. Sci. USA.*, 95(23):13591–6, 1998.

[186] Carlos H. I. Ramos, Sebastien Weisbuch, and Marc Jamin. Diffusive motions control the folding and unfolding kinetics of the apomyoglobin pH 4 molten globule intermediate. *Biochemistry*, 46(14):4379–4379, 2007.

[187] Dieter Perl, Maik Jacob, Mikulas Bano, Marek Stupak, Marian Antalik, and Franz X. Schmid. Thermodynamics of a diffusional protein folding reaction. *Biophys. Chem.*, 96(2-3):173–180, 2002.

[188] C. D. Waldburger, T. Jonsson, and R. T. Sauer. Barriers to protein folding: formation of buried polar interactions is a slow step in acquisition of structure. *Proc. Natl. Acad. Sci. USA.*, 93(7):2629–2634, 1996.

[189] Suzette A. Pabit, Heinrich Roder, and Stephen J. Hagen. Internal friction controls the speed of protein folding from a compact configuration. *Biochemistry*, 43(39):12532–8, 2004.

[190] Linlin Qiu and Stephen J. Hagen. A limiting speed for protein folding at low solvent viscosity. *J. Am. Chem. Soc.*, 126(11):3398–3399, 2004.

[191] R. F. Grote and J. T. Hynes. The stable states picture of chemical reactions. II. Rate constants for condensed and gas phase reaction models. *J. Chem. Phys.*, 73:2715–2732, September 1980.

[192] J. Schlitter. Viscosity dependence of intramolecular activated processes. *Chem. Phys.*, 120:187–197, February 1988.

[193] Gouri S. Jas, William A. Eaton, and James Hofrichter. Effect of viscosity on the kinetics of alpha-helix and beta-hairpin formation. *J. Phys. Chem. B*, 105(1):261–272, 2001.

[194] D. Beece, L. Eisenstein, H. Frauenfelder, D. Good, M. C. Marden, L. Reinisch, A. H. Reynolds, L. B. Sorensen, and K. T. Yue. Solvent viscosity and protein dynamics. *Biochemistry*, 19(23):5147–5147, 1980.

[195] A. Ansari, C. M. Jones, E. R. Henry, J. Hofrichter, and W. A. Eaton. The role of solvent viscosity in the dynamics of protein conformational changes. *Science*, 256(5065):1796–1798, 1992.

[196] Annette Steward, Jose Luis Toca-Herrera, and Jane Clarke. Versatile cloning system for construction of multimeric proteins for use in atomic force microscopy. *Protein Sci.*, 11(9):2179–2183, 2002.

[197] B. Miroux and J. E. Walker. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.*, 260(3):289–298, 1996.

[198] M. Raible, M. Evstigneev, P. Reimann, F. W. Bartels, and R. Ros. Theoretical analysis of dynamic force spectroscopy experiments on ligand-receptor complexes. *J. Biotechnol.*, 112(1-2):13–23, 2004.

[199] H. Nymeyer, N. D. Socci, and J. N. Onuchic. Landscape approaches for determining the ensemble of folding transition states: success and failure hinge on the degree of frustration. *Proc. Natl. Acad. Sci. USA.*, 97(2):634–639, 2000.

[200] Retraction for Dougan et al., Solvent molecules bridge the mechanical unfolding transition state of a protein. *Proc. Natl. Acad. Sci. U.S.A.*, 108(51):20850, 2011.

[201] Retraction for Dougan et al., Probing osmolyte participation in the unfolding transition state of a protein. *Proc. Natl. Acad. Sci. U.S.A.*, 108(51):20850, 2011.

[202] Matthew Auton, Jorg Rosgen, Mikhail Sinev, Luis Marcelo F. Holthauzen, and D. Wayne Bolen. Osmolyte effects on protein stability and solubility: a balancing act between backbone and side-chains. *Biophys Chem*, 159(1):90–99, 2011.

[203] Swarita Gopal and Jagdish C. Ahluwalia. Effect of osmoregulatory solutes on the stability of proteins. *J. Chem. Soc., Faraday Trans.*, 89:2769–2774, 1993.

[204] Anannya Bandyopadhyay, Kanika Saxena, Neha Kasturia, Vijit Dalal, Niraj Bhatt, Asher Rajkumar, Shuvadeep Maity, Shantanu Sengupta, and Kausik Chakraborty. Chemical chaperones assist intracellular folding to buffer mutational variations. *Nat Chem Biol*, 8(3):238–245, 2012.

[205] Allan Chris M. Ferreon, Mahdi Muhammad Moosa, Yann Gambin, and Ashok A. Deniz. Counteracting chemical chaperone effects on the single-molecule alpha-synuclein structural landscape. *Proc Natl Acad Sci U S A*, 2012.

[206] Francesca Macchi, Maike Eisenkolb, Hans Kiefer, and Daniel E. Otzen. The effect of osmolytes on protein fibrillation. *Int J Mol Sci*, 13(3):3801–3809, 2012.

[207] Joseph Loscalzo. Lipid metabolism by gut microbes and atherosclerosis. *Circ Res*, 109(2):127–129, 2011.

[208] Zeneng Wang, Elizabeth Klipfell, Brian J. Bennett, Robert Koeth, Bruce S. Levison, Brandon Dugar, Ariel E. Feldstein, Earl B. Britt, Xiaoming Fu, Yoon-Mi Chung, Yuping Wu, Phil Schauer, Jonathan D. Smith, Hooman Allayee, W. H. Wilson Tang, Joseph A. DiDonato, Aldons J. Lusis, and Stanley L. Hazen. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*, 472(7341):57–63, 2011.

[209] Laishram R. Singh, Xulin Chen, Viktor Kozich, and Warren D. Kruger. Chemical chaperone rescue of mutant human cystathionine beta-synthase. *Mol Genet Metab*, 91(4):335–342, 2007.

[210] Brian J. Bennion, Mari L. DeMarco, and Valerie Daggett. Preventing misfolding of the prion protein by trimethylamine N-oxide. *Biochemistry*, 43(41):12955–63, 2004.

[211] V. N. Uversky, J. Li, and A. L. Fink. Trimethylamine-N-oxide-induced folding of alpha-synuclein. *FEBS Lett*, 509(1):31–35, 2001.

[212] D. S. Yang, C. M. Yip, T. H. Huang, A. Chakrabartty, and P. E. Fraser. Manipulating the amyloid-beta aggregation pathway with chemical chaperones. *J Biol Chem*, 274(46):32970–4, 1999.

[213] Athena L. Samerotte, Jeffrey C. Drazen, Garth L. Brand, Brad A. Seibel, and Paul H. Yancey. Correlation of trimethylamine oxide and habitat depth within and among species of teleost fish: an analysis of causation. *Physiol Biochem Zool*, 80(2):197–208, 2007.

[214] Oscar P. Chilson and Anne E. Chilson. Perturbation of folding and reassociation of lactate dehydrogenase by proline and trimethylamine oxide. *Eur J Biochem*, 270(24):4823–4824, 2003.

[215] Deepak R. Canchi and Angel E. Garcia. Cosolvent Effects on Protein Stability. *Annu Rev Phys Chem*, 2013.

[216] Raj Kumar, Justin M. Serrette, and E. Brad Thompson. Osmolyte-induced folding enhances tryptic enzyme activity. *Arch Biochem Biophys*, 436(1):78–82, 2005.

[217] Shazia Jamal, Nitesh K. Poddar, Laishram R. Singh, Tanveer A. Dar, Vikas Rishi, and Faizan Ahmad. Relationship between functional activity and protein stability in the presence of all classes of stabilizing osmolytes. *FEBS J*, 276(20):6024–6032, 2009.

[218] Jonathan E. Nuss, Laura M. Wanner, Lyal E. Tressler, and Sina Bavari. The osmolyte trimethylamine N-oxide (TMAO) increases the proteolytic activity of botulinum neurotoxin light chains A, B, and E: implications for enhancing analytical assay sensitivity. *J Biomol Screen*, 15(8):928–936, 2010.

[219] S. B. Ozkan, I. Bahar, and K. A. Dill. Transition states and the meaning of Phi-values in protein folding kinetics. *Nat. Struct. Biol.*, 8(9):765–769, 2001.

[220] Iksoo Chang, Marek Cieplak, Jayanth R. Banavar, and Amos Maritan. What can one learn from experiments about the elusive transition state? *Protein Sci.*, 13(9):2446–2447, 2004.

[221] M. Vendruscolo, E. Paci, C. M. Dobson, and M. Karplus. Three key residues form a critical contact network in a protein folding transition state. *Nature*, 409(6820):641–645, 2001.

[222] Thomas R. Weikl. Transition States in Protein Folding. *Commun. Comput. Phys.*, 7(2):283–300, 2010.

[223] Tao Shen, Yi Cao, Shulin Zhuang, and Hongbin Li. Engineered bi-histidine metal chelation sites map the structure of the mechanical unfolding transition state of an elastomeric protein domain GB1. *Biophys J*, 103(4):807–816, 2012.

[224] Yongnan Devin Li, Guillaume Lamour, Jörg Gsponer, Peng Zheng, and Hongbin Li. The molecular mechanism underlying mechanical anisotropy of the protein gb1. *Biophys J*, 103(11):2361–2368, Dec 2012.

[225] Bryan A. Krantz, Leland Mayne, Jon Rumbley, S. Walter Englander, and Tobin R. Sosnick. Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J Mol Biol*, 324(2):359–361, 2002.

[226] Jennifer H. Tomlinson, Saif Ullah, Poul Erik Hansen, and Mike P. Williamson. Characterization of salt bridges to lysines in the protein G B1 domain. *J Am Chem Soc*, 131(13):4674–4674, 2009.

[227] G. S. Ratnaparkhi and R. Varadarajan. Osmolytes stabilize ribonuclease S by stabilizing its fragments S protein and S peptide to compact folding-competent states. *J Biol Chem*, 276(31):28789–98, 2001.

[228] I. Baskakov and D. W. Bolen. Time-dependent effects of trimethylamine-N-oxide/urea on lactate dehydrogenase activity: an unexplored dimension of the adaptation paradigm. *Biophys J*, 74(5):2658–2665, 1998.

[229] Jian-Min Yuan, Chia-Lin Chyan, Huan-Xiang Zhou, Tse-Yu Chung, Haibo Peng, Guanghui Ping, and Guoliang Yang. The effects of macromolecular crowding on the mechanical stability of protein molecules. *Protein Sci.*, 17(12):2156–2156, 2008.

[230] Olga K. Dudko, Gerhard Hummer, and Attila Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc Natl Acad Sci U S A*, 105(41):15755–60, 2008.

[231] William T. King, Meihong Su, and Guoliang Yang. Monte Carlo simulation of mechanical unfolding of proteins based on a simple two-state model. *Int J Biol Macromol*, 46(2):159–166, 2010.

[232] Kamila Mazur, Ismael A. Heisler, and Stephen R. Meech. THz spectra and dynamics of aqueous solutions studied by the ultrafast optical Kerr effect. *J Phys Chem B*, 115(11):2563–2563, 2011.