University of California
Santa Barbara

Glassy Dynamics in an Intrinsically Disordered Protein Region

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Biomolecular Sciences and Engineering

by

Ian Lewis Morgan

Committee in charge:

Professor Omar A. Saleh, Chair
Professor Phil Pincus
Professor Herbert Waite
Professor Cyrus Safinya

September 2021

The Dissertation of Ian Lewis Morgan is approved.

Professor Phil Pincus

Professor Herbert Waite

Professor Cyrus Safinya

Professor Omar A. Saleh, Committee Chair

September 2021

Glassy Dynamics in an Intrinsically Disordered Protein Region

by

Ian Lewis Morgan

Acknowledgements

I would like to acknowledge the support and guidance of my committee members, Omar Saleh, Phil Pincus, Cyrus Safinya, and Herbert Waite. I would like to especially thank my advisor, Omar Saleh, for helping me grow as an independent researcher. The work in this dissertation would not have been possible without him.

I would like to acknowledge the current and former members of the Saleh Lab, including Sarah Innes-Gold, Dan Nguyen, Bob Lansdorp, David Jacobson, John Berezney, Byoung-jin Jeon, Gabrielle Abraham, Frank Truong, Anna Nguyen, Aria Chaderjian, Nate Conrad, and Sam Wilken. I benefited greatly from thoughtful discussions and helpful advice from them. In particular, I would like to thank Sarah for her consistent willingness to discuss new ideas, no matter how outlandish they were.

I would like to acknowledge my collaborators Ram Avinery, Gil Rahamim, and Roy Beck. Ram, and then Gil, were extremely helpful with the design, expression, purification, and polymerization of the proteins in this dissertation.

Finally, I would like to acknowledge my friends and family who provided moral support throughout this whole endeavor. In particular, I would like to thank my parents, my partner, and the running club.

Curriculum Vitæ

Ian Lewis Morgan

Education

2021	Ph.D. in Biomolecular Sciences and Engineering, University of California, Santa Barbara.
2014	B.A. in Physics and Biochemistry, St. Mary's College of Maryland

Publications

Ehm, T., Shinar, H., Meir, S., Sekhon, A., Sethi, V., **Morgan, I. L.**, Rahamim, G., Saleh, O. A., & Beck, R. (2021). Intrinsically disordered proteins at the nano-scale. Nano Futures, 5(2), 022501. https://doi.org/10.1088/2399-1984/abfb7c

Morgan, I. L., Avinery, R., Rahamim, G., Beck, R., Saleh, O. A. (2020). Glassy Dynamics and Memory Effects in an Intrinsically Disordered Protein Construct. Physical Review Letters, 125(5), 058001. https://doi.org/10.1103/PhysRevLett.125.058001 (Editors' Suggestion)

Innes-Gold, S. N.*, **Morgan, I. L.***, Saleh, O. A. (2017). Surface-induced effects in fluctuation-based measurements of single-polymer elasticity: A direct probe of the radius of gyration. The Journal of Chemical Physics, 148(12), 123314. https://doi.org/10.1063/1.5009049 (*Authors contributed equally to this work)

Musich, T., Demberg, T., **Morgan, I. L.**, Estes, J. D., Franchini, G., Robert-Guroff, M. (2015). Purification and functional characterization of mucosal IgA from vaccinated and SIV-infected rhesus macaques. Clinical Immunology, 158(2), 127–139. https://doi.org/10.1016/j.clim.2015.03.020

Campbell, C. T., Llewellyn, S. R., Damberg, T., **Morgan, I. L.**, Robert-Guroff, M., Gildersleeve, J. C. (2013). High-Throughput Profiling of Anti-Glycan Humoral Responses to SIV Vaccination and Challenge. PLoS ONE, 8(9). https://doi.org/10.1371/journal.pone.0075302

Babigumira, J. B., **Morgan, I.**, Levin, A. (2013). Health economics of rubella: a systematic review to assess the value of rubella vaccination. BMC Public Health, 13, 406. https://doi.org/10.1186/1471-2458-13-406

Abstract

Glassy Dynamics in an Intrinsically Disordered Protein Region

by

Ian Lewis Morgan

Conservative estimates indicate that 30% of known proteins contain long (>40 residues) intrinsically disordered regions (IDRs). In contrast to globular proteins, IDRs adopt multiple distinct conformations in their native state, similar to a random-walk polymer in a good solvent. While the sequence properties of IDRs have been extensively studied, their physical properties are still poorly understood. Characterizing these physical properties is an important step towards understanding the numerous biological functions and diseases associated with IDRs. Recently, Magnetic Tweezers (MTs) have emerged as a powerful tool for determining the structural properties of polymers in a manner that is orthogonal to other approaches, e.g., scattering experiments. With MTs, researchers use the thermodynamic effects of applied tension to study the conformations of single polymers via their end-to-end extension. This approach is particularly desirable for IDRs because many of them are found in the cellular cytoskeleton, where they play a critical structural and mechanical role.

In the first part of this dissertation, I present my work on improving the MT technique with the aim of studying IDRs. I develop a computational tool for the robust calibration of forces and their uncertainty, an important but often overlooked aspect of these experiments. I also study the effects of surfaces on the low-force entropic elastic response of polymers, showing that they can be used to extract the radius of gyration.

In the second part of this dissertation, I present the unexpected finding of glassy dynamics in a model IDR system, a polyprotein of the disordered neurofilament light

tail (NFLt) domain. The NFLt is part of a large group of IDRs in neurons that are responsible for the structure and mechanics of the axon. Glassy dynamics in globular proteins was a major finding nearly 50 years ago that emphasized the importance of protein dynamics. However, it is attributed to conformational behaviors that are missing from IDRs. Nevertheless, using MT experiments, I show that a NFLt polyprotein's extension changes, in response to a change in applied tension, with a nonexponential time dependence that is history dependent, two characteristic features of glassy systems. I show that the extension changes can be predicted using a phenomenological framework adapted from bulk glassy systems. Finally, I show that the glassy dynamics can be understood in terms of multiple, independent, and heterogeneous globules within a single NFLt. This mechanism for glassy dynamics is novel and likely to apply broadly to other IDRs.

Contents

Chapter 1

Introduction

Part of the content of this chapter is adapted from a collaborative review [1]. It is used here with permission from IOP Publishing.

1.1 Motivation

The 1977 Nobel Laureate in Physics, Phillip Anderson, described the nature of glasses as one of "the deepest and most interesting and unsolved problems in solid state theory" [2]. This problem has broad applicability because many complex systems share the properties of structural glasses, i.e., liquids that are supercooled below their freezing point. For this reason, these properties are often called 'glassy.'

One of these properties is glassy dynamics, which refers to the slow, nonexponential time dependence of a system that is driven out of equilibrium [3]. Glassy dynamics is typically attributed to competing interactions that make it difficult to transition between states, i.e., frustration [4]. As a system relaxes, it becomes trapped in long-lived metastable states and, in some cases, may never reach equilibrium on humanly observable time scales. Hence, while a simple system's state can be determined by equilibrium

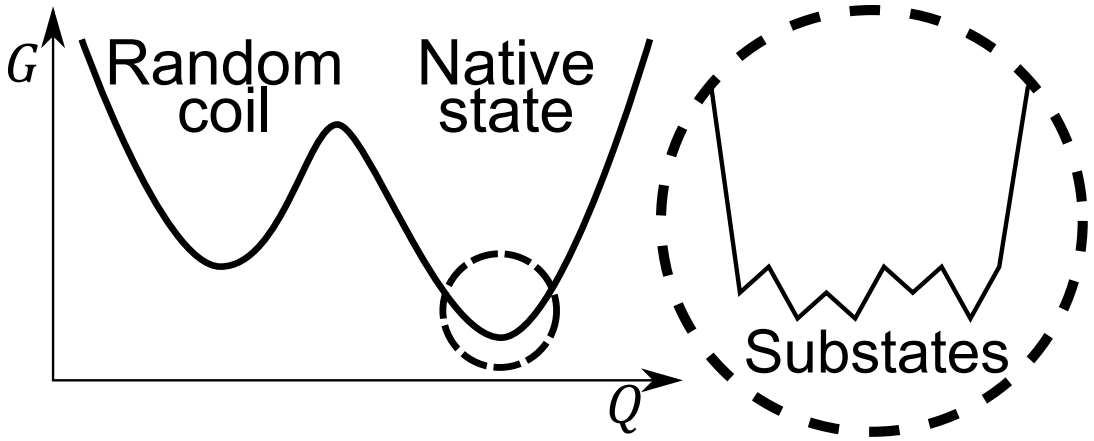

Figure 1.1: A possible free energy, G, surface for proteins with the fraction of native contacts, Q, as the reaction coordinate. The surface has a double-well structure, where the random coil state is entropically favored and the native state is energetically favored with a nucleation activation barrier separating them. The native state is the global minima; however, frustration leads to multiple local minima, corresponding to structural substates.

thermodynamics, a glassy system's state depends on its kinetics and history.

Heteropolymers are inherently frustrated systems due to the constraints imposed by a connected chain with many possible heterogeneous interactions [4]. For example, a random chain with many possible hydrophobic and hydrophilic pairs may require unfavorable contacts in order to form favorable contacts. As a result of this 'energetic frustration,' random heteropolymers often form multiple different structures with similar energies. Transition between these structures may require breaking favorable contacts and avoiding steric clashes. This 'topological' or 'geometric' frustration can lead to substantial activation barriers between structures and a rugged energy landscape. Together, these types of frustration can lead to glassy dynamics.

In 1973, Austin *et al.*[5] reported the observation of glassy dynamics in a small globular protein, myoglobin. At low temperatures, ligand rebinding experiments with this protein exhibited a nonexponential time dependence. Based on these findings, Frauenfelder *et al.* [6, 7] concluded that proteins can have multiple different substates in their native

state. Prior to this, proteins were often thought to have a rigid structure, corresponding to a single well-defined minima on its free energy surface. However, Frauenfelder *et al.* described the free energy surface as rugged with multiple local minima, corresponding to the different structural substates (Fig. 1.1). Depending on thermal energy and the activation barriers between local minima, proteins will interconvert between these different substates, which is thought to be crucial for their biological function [8].

Yet, unlike random heteropolymers, the large scale conformational changes of small single-domain proteins often follow an exponential time dependence. In 1987, Byngelson and Wolynes [9] introduced the principle of minimal frustration to explain this phenomenon. It states that evolution has selected for protein sequences that avoid kinetic traps and have a single global free energy minimum. Their work was based on earlier work by Gō which posited that a protein's native interactions are the main driving force of folding [10]. The principle of minimal frustration leads to a one-dimensional free energy surfaces with the fraction of native contacts as the reaction coordinate. Generally, the random coil state is entropically favored, the native state is energetically favored, and transitions between the two are treated as a cooperative, two-state process. (Fig. 1.1B).

Intrinsically disordered proteins and protein regions

In contrast to the principle of minimal frustration, many proteins contain disordered regions that adopt multiple distinct conformations in their native state [11, 12, 13, 14]. While it had long been known that disordered regions exist, there prevalence had not been fully appreciated [15]. In 1988, a computer algorithm revealed that more than 15,000 proteins in the Swiss Protein Database contain long (>40 residues) disordered regions [16]. Current conservative estimates indicate that 30% of known proteins contain long disordered regions, leading to considerable scientific interest in their sequence,

structure, and function [12].

Disordered regions have been called many names in the literature [12]; however, now, the field appears to have coalesced around two terms. Globally disordered proteins are called Intrinsically Disordered Proteins (IDPs), while locally disordered regions are called Intrinsically Disordered Regions (IDRs). As IDPs and IDRs share many of the same characteristics, for brevity, I will refer to both as IDRs unless specified otherwise.

Protein disorder appears to be driven by distinct sequence preferences, which computer algorithms use to identify IDRs [17, 18]. In general, IDRs have fewer hydrophobic residues and more charged and hydrophilic resides than globular proteins [19, 20, 14] (Fig. 1.2). However, their sequence preferences can be further broken up into three categories, polar tracts, polyelectrolytes, and polyampholytes, that depend on the fraction of negatively and positively charged residues [21]. Sequences rich in polar but not charged residues are called polar tracts and are predicted to form compact globules due to poor solvation by the surrounding solution [22, 23]. Sequences with a large fraction of either positively or negatively charged residues are polyelectrolyte and are predicted to be more extended in isolation due to electrostatic repulsion. Polyelectrolytes conformations also depend on their surrounding environment; they will often bind small charged molecules or undergo complex coacervation [21]. Sequences with both positively and negatively charged residues are polyampholytes and are sometimes further divided into weak and strong depending on the total fraction of charged residues in the IDR [24]. Polyampholytes with well-mixed residues are predicted to behave like random coils, while blocky polyampholytes (i.e., oppositely charged residues are segregated) form more compact configurations.

These sequence propensities drive the structural properties of IDRs. Due to their many possible conformations, the structural properties of IDRs are often characterized by their ensemble average characteristics [1]. Specifically, experiments (e.g., Small An-

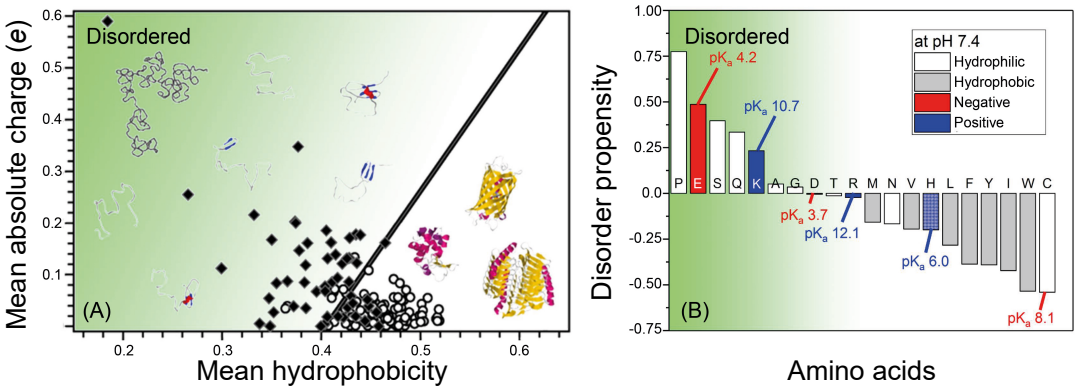

Figure 1.2: Disorder is driven by charge and hydrophilicity [1]. (A) Nearly 250 folded proteins (open circles) and 90 IDRs (filled diamonds) are empirically separated (diagonal line) by their mean absolute charge (e) and hydrophobicity (normalized Kyte–Doolittle scale [25]) [19]. (B) Certain amino acids are more common in IDRs, suggesting they promote disorder [20]. Disorder propensity is defined as the fractional difference of amino acid composition from IDRs in the DisProt database and folded proteins from the Protein Data Bank (PDB). This figure is reproduced from Ref. [1].

gle X-ray Scattering (SAXS) [26] and Förster Resonance Energy Transfer (FRET) [27]) measure the end-to-end distance (R_{ee}) or radius of gyration (R_g), which scale with the number of residues, N, and the Flory exponent, ν, e.g., $R_g \propto N^{\nu}$ (Fig. 1.3). The Flory exponent captures the balance between intra-chain and solvent-chain interactions, i.e., solvent quality. Good ($\nu \approx 0.6$), theta ($\nu = 0.5$), or poor ($\nu \approx 0.3$) solvent quality indicates intra-chain attractions that are less, equal, or greater than solvent-chain attractions, respectively. While IDRs are generally more similar to random-walk polymers in a good solvent than folded proteins, they often contain residual secondary structure or, as noted, sequence-specific interactions that affect their structural properties [28].

The structural properties of IDRs drive their biological functions [1] (Fig. 1.3). In several cases, certain stimuli trigger IDRs to undergo disorder-to-order transitions that enable them to interact strongly and specifically with a single binding partner via a 'lock and key' mechanism [29]. Other IDRs serve as 'master keys,' interacting weakly with many different binding partners, facilitating their role as regulators in cellular signalling

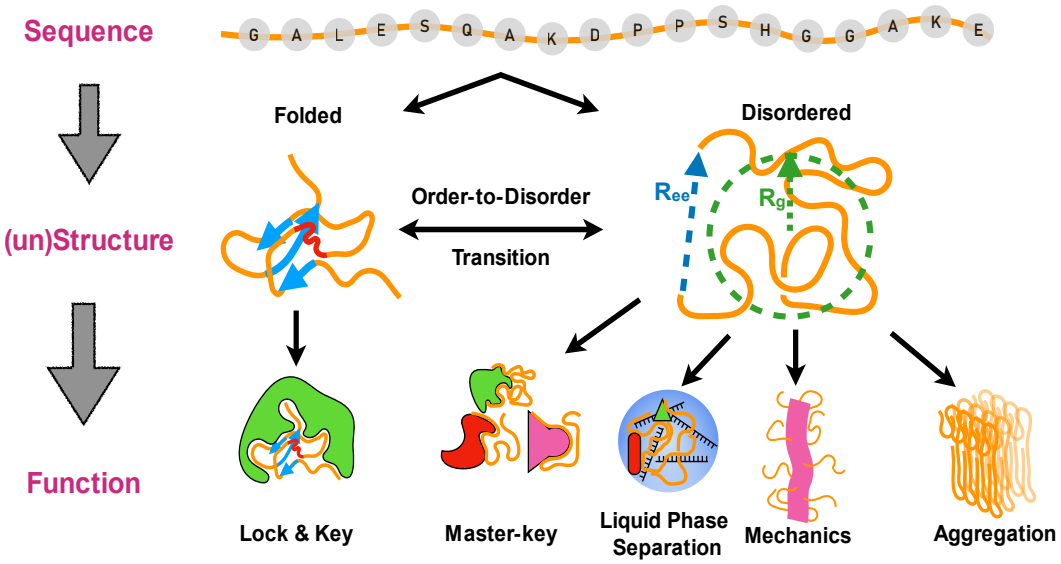

Figure 1.3: A protein's sequence of amino acids influences its structure, which, in turn, influences its biological function. Modified from Ref. [1].

pathways [15, 30, 31]. Relatedly, IDRs have received considerable attention for their ability to assemble proteins and nucleic acids into liquid-like condensates and hydrogel networks [32, 33, 34]. Many IDRs also benefit directly from their structural flexibility, particularly in the cytoskeleton, where they play an important role in the structure and mechanics of cells [35, 36].

In addition to these beneficial biological functions, IDRs are also associated with a number of diseases, including cancer, cardiovascular disease, diabetes, and neurodegeneration [12, 1]. Many of these diseases involve the aggregation of IDRs into toxic oligomers or amyloids [37]. Others involve IDRs found in pathogens, such as the recently emerged novel coronavirus (SARS-CoV-2), which encodes an IDR in its nucleocapsid that is critical for genome packaging [38]. The possibility of interrupting the action of these diseases presents an attractive target for new drugs.

While the sequence properties of IDRs have been extensively studied, their structural properties are still poorly understood. Characterizing the structural properties of IDRs is an important first step towards understanding their biological functions and associated

diseases.

Single-Molecule Force Spectroscopy

As noted, the structural properties of IDRs lend themselves to polymer physics descriptors. Single-Molecule Force Spectroscopy (SMFS) has emerged as a powerful method for characterizing the structural properties of polymers in a manner that is orthogonal to other approaches, such as SAXS and FRET. The most common SMFS instruments include Atomic Force Microscopes (AFMs), Optical Tweezers (OTs), and Magnetic Tweezers (MTs). These instruments use the thermodynamics effects of applied tension to probe polymers' conformations via their end-to-end extension, yielding valuable information about their structural length scales, free energy surface, and conformational dynamics [39]. These experiments are also particularly relevant for cytoskeletal IDRs, which play a critical role in the structure and mechanics of cells [35, 36].

The force-extension response of a polymer is well-described by scaling theories that account for the local bending stiffness and "excluded-volume" interactions via various structural length scales, including the persistence length, l_p, Kuhn length, ℓ, thermal blob size, ξ_{th}, and R_g [40]. The applied force effectively screens interactions below the tensile screening length, $\xi_f = k_B T/f$, leading to several different force regimes [41]. In the low-force regime, corresponding to forces less than $k_B T/\ell$, the extension scales as a powerlaw with the force, $\sim f^{\gamma}$. For example, when $\xi_{\text{th}} < k_B T/f < R_g$, the relevant metric is the Pincus exponent, $\gamma = 2/3$ [41]. Hence, it is possible to extract the polymer's structural length scales from the changes in γ with force.

In an SMFS experiment, the polymer extension is measured in response to an applied force. Hence, it measures a one-dimensional free energy surface, with the extension as the reaction coordinate [42]. Most reactions can be modeled as a simple two-state system with

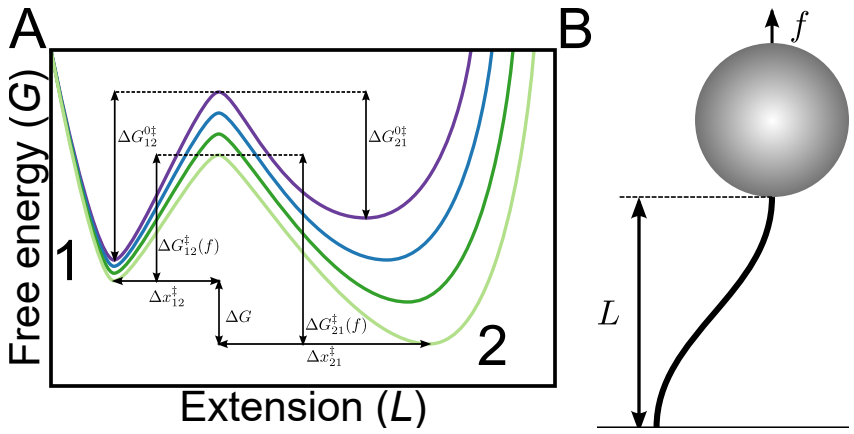

Figure 1.4: (A) Force tilts the free energy surface favoring the extended state (state 2) over the compact state (state 1), lowering the activation barrier to transition from state 1 to 2 from $\Delta G_{12}^{0\ddagger}$ to $\Delta G_{12}^{\ddagger}(f) = \Delta G_{12}^{0\ddagger} - f\Delta x_{12}^{\ddagger}$, and increasing the activation barrier to transition from state 2 to state 1 from $\Delta G_{21}^{0\ddagger}$ to $\Delta G_{21}^{\ddagger}(f) = \Delta G_{21}^{0\ddagger} - f\Delta x_{21}^{\ddagger}$, where $x_{12}^{\ddagger}$ and $\Delta x_{21}^{\ddagger}$ are the distances from state 1 or state 2 to the transition state, respectively. The free energy difference ΔG between the two states determines the equilibrium populations of each state. (B) In a MT experiment, a polymer is tethered between a static surface and micron-scale paramagnetic bead. The polymer's extension, L, is measured in response to an applied force, f.

a double-well structure. The free energy difference, ΔG between states determines the equilibrium populations of each state (Fig. 1.4). At zero force, the time to go from state 1 to state 2 is given by the Arrhenius equation: $\tau_{12} = A_1 e^{\Delta G_{12}^{0\ddagger}/k_B T}$, where A corresponds to an attempt frequency [43] and $\Delta G_{12}^{0\ddagger}$ is the height of the activation barrier. Similarly, the time to go from state 2 to state 1 is given by $\tau_{21} = A_2 e^{\Delta G_{21}^{0\ddagger}/k_B T}$.

To a first approximation, an applied force tilts the free energy surface linearly with the magnitude of the force, changing the heights of the activation barriers, such that $\Delta G_{12}^{\ddagger}(f) = \Delta G_{12}^{0\ddagger} - f\Delta x_{12}^{\ddagger}$ and $\Delta G_{21}^{\ddagger}(f) = \Delta G_{21}^{0\ddagger} - f\Delta x_{21}^{\ddagger}$, where $x_{12}^{\ddagger}$ and $\Delta x_{21}^{\ddagger}$ are the distances from state 1 or state 2 to the transition state, respectively [44] (Fig. 1.4A). As a result, the time to go from state 1 to state 2 decreases and the time to go from state 2 to state 1 increases. This effect was worked out by Zhurkov for fracture mechanics in 1965 [45] and later introduced into biophysics by Bell in 1978 [46]. For this reason, I refer to it as the Bell-Zhurkov expectation in this dissertation.

Constant-force experiments provide a straightforward means of directly measuring the free energy surface via a polymer's conformational dynamics: extension changes directly correspond to diffusion along the free energy surface. Some SMFS experiments, e.g., AFM force-clamp, require active feedback loops to maintain a constant force, which can lead to experimental artifacts and difficulties in interpreting results [42]. Other experiments, e.g., MTs, passively apply constant forces, making it simpler to carry out and interpret results. In some cases, under constant forces, biopolymers will 'hop' between the two states when the activation barrier between state 1 and 2 is of a similar height to the barrier between state 2 and 1 [44]; however, for many proteins, there is no single force at which this hopping behavior is experimentally accessible.

An alternative means of measuring the conformational dynamics of polymers is force-jump experiments. In these experiments, the force is changed rapidly to a set value, and then, the polymer extension is measured at constant force. Assuming the transition time is longer than the time required to execute the jump and the entropic elastic relaxation time (i.e., the Rouse time [47, 48]), the transition time can be directly measured from the change in polymer extension. Conformational changes are stochastic thermally induced events, but, for small-single domain proteins, the ensemble time-average of many events often follows an exponential time-dependence, corresponding to a cooperative, two-state process with the occasional addition of transient intermediates states [49, 44].

While SMFS techniques have been used extensively to study globular proteins, there have been considerably fewer studies of IDRs. Notable exceptions include AFM pulling experiments on several amyloid precursor IDRs, which showed a sawtooth pattern indicative of the mechanical unfolding of multiple different structures [50]. Similarly, optical tweezer pulling experiments on α-synuclein showed it has several marginally stable and rapidly fluctuating subsegment structures [51]. However, to my knowledge, the conformational dynamics of IDRs under constant force have not been investigated.

1.2 This dissertation

In this dissertation, I present my work on extending the capabilities of SMFS experiments and conducting high-precision nanomechanical assays on a model IDR system.

Magnetic tweezers

Much of the experimental work in this dissertation uses a custom-built MT instrument that is described in detail in Appendix A. In a MT experiment, the polymer of interest is tethered between a static surface and a micron-scale paramagnetic bead in a flow cell (Fig. 1.4B). The force on the bead (and thus, the polymer of interest) is set by the distance between the bead and magnet assembly, resulting in ultra-stable forces. This is advantageous for two reasons. First, it enables the application of low forces in force-extension measurements. Second, when measuring the conformational dynamics of polymers, the force remains constant over time.

Neurofilament light tail

The model IDR system is derived from the disordered tail domain of the neurofilament light protein (UniProt accession number P08551). The neurofilament light protein is part of a large group of neuronal-specific proteins (e.g., α-internexin, vimentin, microtubule-associated protein 2, and tau [52, 36, 53, 54]) with long IDRs that help to determine the structure, size, and mechanics of axons, with direct effects on electrical conduction [55].

The native Neurofilament light tail (NFLt) domain contains 146 residues with an average net charge per residue of -0.25 e; however, it can be divided into two subdomains with highly distinct sequence features (Fig. 1.5). Subdomain A is 47 residues in length, making up the first one-third of the NFLt. Its sequence features are consistent with a polar tract, containing relatively few charged residues and more polar residues (Fig. 1.5).

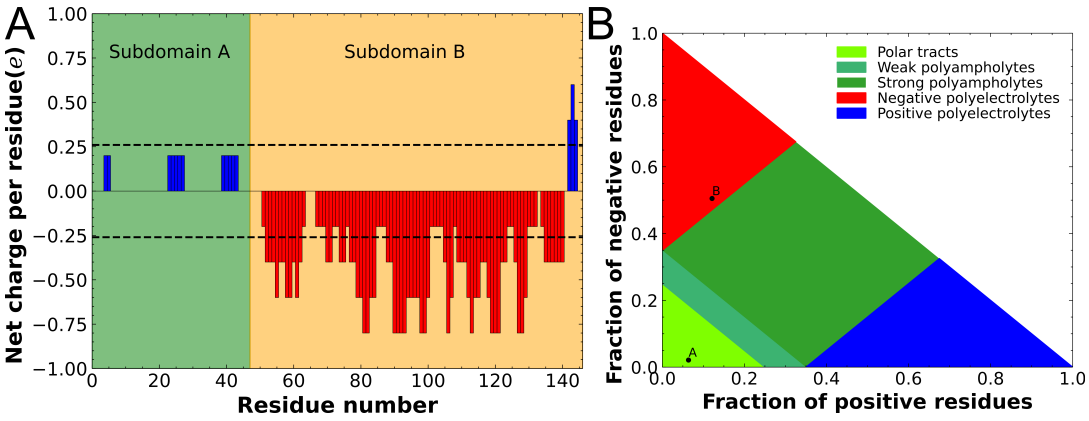

Figure 1.5: The sequence characteristics of the NFLt's two subdomains differ substantially. (A) Subdomain A (green) has mostly polar residues and very few charged residues. Subdomain B (orange) has mostly negatively charged residues with a few positively charged residues. The net charge per residue represents a five residue rolling average over the sequence. (B) Subdomain A can be classified as a polar tract, while subdomain B can be classified as a negative polyelectrolyte based on the fraction of negative and positive residues [24].

Specifically, it is rich in serine (S, 30%) and the aromatic residue tyrosine (Y, 10%). Subdomain B is 99 residues in length, making up the latter two-thirds of the NFLt. Its sequence features are consistent with a negatively charged polyelectrolyte, but it contains a few positively charged residues, particular towards the C-terminus. These positively charged residues have been shown to play an important role in assembling neurofilaments into a hydrogel network [52, 56]. Specifically, the sequence is rich in glutamic acid (E, 45%) and lysine (K, 12%). Secondary structure predictors suggest that the NFLt might contain some long alpha helices and beta sheets [36]; however, Circular Dichroism (CD) measurements are more consistent with a random coil configuration [57].

MT experiments require long polymers to prevent interactions between the bead and surface. To facilitate MT experiments with the NFLts, they are polymerized into a polyprotein as described in detail in Appendix B. In short, the NFLts are modified to contain terminal cysteines and induced to form disulfide bonds, resulting in a polydisperse sample with polyproteins of varying lengths.

A computational tool for robust force calibration

In order to interpret the results of the NFLt experiments, the applied forces need to be accurately and precisely calibrated. While robust force calibration methods exist, they can be difficult to implement computationally. As a result, many researchers opt to use alternative methods [58]. However, as has been argued elsewhere, the lack of standardized methods across SMFS experiments hinders the reproducibility of results [59]. In Chapter 2, I present the development of a software tool for robust force calibration in SMFS experiments. This tool calibrates both the magnitude of forces and its uncertainty, an important but often overlooked part of these experiments. I envision that this tool will help to standardize force calibration in SMFS experiments, leading to more accurate and reproducible results.

Surface effects on low-force powerlaw behavior

The stochastic polymerization of the NFLts results in polyproteins of varying lengths, some of which are short ($\lesssim 0.5\,\mu m$). These tethered polyproteins often exhibit a bias in the γ metric of their force-extension curves at low forces. In chapter 3, I study these γ biases using well-characterized polymers, showing that they come from interactions between the polymer and the surface. I show that these biases can be used to extract the R_g of the polymer, broadening the capabilities of MT experiments.

Glassy dynamics in an IDR

In chapter 4, I present the unexpected discovery of glassy dynamics in the NFLt polyprotein. In response to a change in applied force, I show that the polyproteins exhibit a change in extension that follows a nonexponential, often logarithmic, time dependence. Furthermore, I show that polyproteins exhibit a memory effect, indicating

that each NFLt contains multiple, independent relaxation events. Using a continuum model based on a mathematical framework devised for bulk glassy systems [60] and the Bell-Zhurkov expectation, I show that the force dependence of the logarithmic relaxations can be predicted with a single fitting parameter.

Origins of glassy dynamics in an IDR

In chapter 5, I study the origin of the glassy dynamics in the NFLt polyprotein. I consider several possible biochemical sources for the relaxations. I show that the glassy dynamics is history dependent, a characteristic feature of glassy systems and further proof that each NFLt contains multiple, independent, and heterogeneous relaxation events. Lastly, I show that the glassy dynamics can be understood in terms of a discrete model with globules of varying size that quantitatively captures the force dependence of the logarithmic relaxations without any fitting parameters. This mechanism for glassy dynamics is, to my knowledge, novel and likely to apply broadly to other IDRs.

Appendices

Following the main body of the dissertation are several appendices. Appendix A contains a description of the custom-built magnetic tweezer used to carry out most of the experiments in this dissertation. Appendix B contains basic protocols and experimental methods. Appendix C contains supporting information for the chapters, including certain derivations and control experiments. Appendix D contains a comprehensive list of abbreviations used in this dissertation.

Chapter 2

Calibrating Forces in Single-molecule Video-tracking Experiments

Using MTs to characterize the conformations of IDRs and the glassy dynamics requires the accurate and precise calibration of forces. Typically, the force is calibrated by comparing the thermal fluctuations of the bead to a model derived from the Langevin equation. Previous work from the Saleh lab had shown that the Allan variance (AV) of the paramagnetic bead fluctuations could be used to reliably calibrate forces [61]. However, this work did not address how to calculate the uncertainty or reliability of forces. Instead, to determine the uncertainty, a MT user would collect three or more independent force measurements for each magnet position on the same bead from which to calculate the average and standard deviation (a single 'bad' measurement still lead to a poor estimate of the uncertainty). When this works, it is slow and tedious. For example, a typical force-extension curve with 31 forces took 20 min; hence, three measurements required at least an hour (usually more). This was alleviated, somewhat, by the ability to measure

multiple polymers in parallel.

Around this time, it came to my attention that many researchers were using different, often less robust force calibration methods. As has been argued elsewhere, this lack of standardized methods hinders reproducibility in the SMFS field[59]. Hence, I developed a straightforward and easy-to-use computational tool that not only robustly calibrates forces but also calculates their uncertainty from a single measurement. In the interest of open science, I wrote the tool in Python, a common and freely available programming language, so it could be used regardless of access to proprietary and expensive programming languages, such as Matlab, Mathematica, and Labview.

The content of this chapter is adapted from a manuscript that, at the time of the writing of this dissertation, is currently undergoing review for publication in PLOS One.

2.1 Introduction

Single-molecule force spectroscopy (SMFS) instruments are powerful tools with a wide variety of experimental applications. They can be used to study polymer elasticity [62, 40] and dynamics [63], measure bond energies and lifetimes [64, 65], assess the activity of molecular motors [66, 67], and characterize protein and nucleic acid folding [68].

To obtain accurate and reproducible results, an essential first step in any SMFS experiment is force calibration. Typically, force calibration relies on comparing the thermal motion of a trapped bead to a model derived from the Langevin equation [69]. These methods have limitations; notably, at times, $t \lesssim 10^{-4}$ s, the standard Langevin equation does not account for certain hydrodynamic effects between the bead and the surrounding fluid [70]. Nevertheless, for longer times, these hydrodynamic effects can be ignored and the bead motion is well-described by the overdamped Langevin equation, which only depends on two parameters: the drag coefficient of the bead, γ, and the spring con-

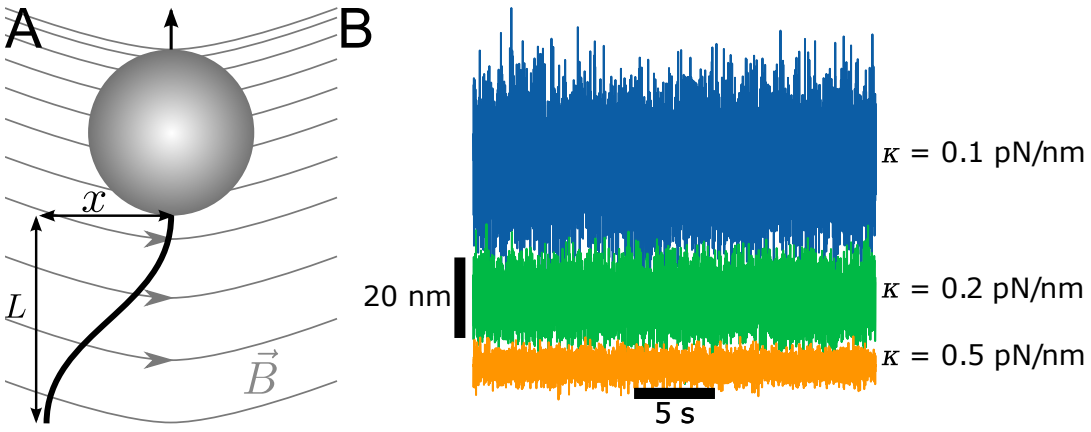

Figure 2.1: (A) Schematic of a MT experiment. The spring constant, κ, is calibrated from the bead fluctuations in the x-direction, parallel to the magnetic field, $\vec{B}$. (B) Simulated bead trajectories for $\kappa = 0.1$, 0.2, and 0.5 pN/nm.

stant of the trap, κ, from which the force can be calculated. For MT experiments, the bead-tether system is modeled as an inverted pendulum and κ is determined from bead fluctuations parallel to the magnetic field (Fig. 2.1). Applying Hooke's law and the small angle approximation, it follows that the applied force is κL.

In practice, analyzing and fitting the bead trajectory must be done carefully. Several factors, including spectral distortions, the exposure time of the detection system (e.g., video cameras or photodiodes), parasitic noise (e.g., tracking errors and mechanical drift), and biased fitting, can all lead to inaccurate parameter estimates [71, 72, 73, 74, 61]. Robust calibration methods that account for all of these factors exist, yet they can be complex to implement computationally, leading some researchers to opt for alternative strategies [58].

Existing force-calibration software packages [75, 76, 77, 78] only account for some sources of bias; most notably they do not account for the finite exposure time of the camera in video-tracking experiments. Thus, it is often up to researchers to write with their own calibration code, of which published examples are only available in proprietary programming languages (e.g., MatLab [79] and LabView[80]), hindering easy access. As

has been argued elsewhere [81], different computational implementations, even those based on the same algorithms, can often lead to different numerical outcomes. This lack of standardized calibration methods and computational implementations hinders reproducibility and makes comparison across different research groups, instruments, and experiments difficult [59].

To help improve and standardize SMFS force calibration, we present Tweezepy: a Python package for calibrating forces in single-molecule video-tracking experiments. Tweezepy uses maximum likelihood estimation (MLE) to estimate parameters, and their uncertainties, from a user-provided bead trajectory via a thermal motion-based model of the power spectral density (PSD) or Allan variance (AV). It accounts for the most common sources of biases and parasitic noise in SMFS video-tracking experiments. Moreover, it is written in Python, a popular and freely available programming language, and is available on GitHub [82], the Python package index [83], and the Zenodo database [84], making it easy to distribute and install. It is designed for ease-of-use, with docstrings and usage examples, yet it has a versatile object-oriented framework that can be used as part of a larger scripted workflow or in a Jupyter notebook as a lab journal page [85, 86].

In this chapter, we provide a comprehensive overview of Tweezepy's force calibration scheme. In Section 2.2, we describe several common force calibration methods and motivate the use of the PSD and AV. In Section 2.3, we give closed-form expressions that account for common sources of parameter biases and parasitic noise in video-tracking experiments, such as the finite exposure time of the detection system and tracking errors. In Section 2.4, we review how to compute the experimental PSD and AV from a bead trajectory. In Section 2.5, we describe how to use MLE to reduce biased fitting and estimate parameters and their uncertainties. We note here that estimating parameter uncertainties using MLE has received relatively little attention in the SMFS literature. For experienced readers that are familiar with force calibration theory, we recommend

skipping ahead to Section 2.6 , which covers the computational implementation of the calibration methods in Tweezepy. In Section 2.7, we use Tweezepy to calibrate simulated bead trajectories, and show that it accurately estimates parameters and their uncertainties, as compared to previously published results in Ref. [61].

2.2 Background

Most force calibration methods fall into two categories: methods that calibrate against known forces, such as Stokes drag or gravitation [87], and methods that calibrate based on the thermal motion of the bead [69]. The first category generally relies on intrinsic parameters of the system (e.g. the density and viscosity of the solution) that can be difficult to measure and often vary within an experiment, leading to large uncertainties in calibrated forces [88].

In comparison, thermal motion-based calibration methods are advantageous because they only rely on the temperature of the system, which is much easier to measure and control in most experiments. Thermal calibration methods model the random, diffusive motion of the bead within the harmonic potential generated by the applied trap. By the equipartition theorem, the standard variance of the bead position, σ_x^2, can be related to the spring constant of the trap, κ:

$$\sigma_x^2 = \frac{k_B T}{\kappa}. \tag{2.1}$$

where k_B is the Boltzmann constant and T is the absolute temperature of the system [89]. When the time between measurements, τ_s, is much faster than the relaxation time of the bead, $\tau_c \equiv \gamma/\kappa$, the spring constant can be determined from Eq. 2.1. However, in practice, sources of parasitic noise always increase the variance, leading to systematic

underestimates of the apparent spring constant [73].

A better approach to thermal motion-based calibration is the PSD, which permits separation of thermal motion from parasitic noise [90, 73]. The PSD describes the distribution of the variance (i.e., total power) from different frequency components in a signal. Invariably, parasitic noise sources have different spectral signatures from the bead's thermal motion. As discussed in detail below, when using the PSD to calibrate video-tracking experiments, one needs to account for several factors, including 1) distortions from aliasing and spectral leakage [91, 71], 2) low-pass filtering from the exposure time of the camera [72], and 3) biased parameter estimates from improperly using least squares fitting routines with experimental PSD values that do not have Gaussian-distributed errors [74].

An alternative means of thermal motion-based calibration, that also distinguishes parasitic from thermal noise, is the AV. The AV measures the noise in the bead position over different observation times and was designed as a means of measuring drift in a system [92]. It was originally introduced into the SMFS literature to assess optimal measurement times [93] and low-frequency noise [94, 95]; however, it was quickly realized that the AV could be fit directly to calibrating forces [61]. As discussed in detail below, the AV is naturally suited to video-tracking experiments because it intrinsically accounts for low-pass filtering from the exposure time of the camera. As with the PSD, improperly using least-squares fitting routines on AV values that do not have Gaussian-distributed errors will lead to biased parameter estimates [61].

When identifying and accounting for various sources of parasitic noise, the PSD and AV have complementary strengths [95]. The PSD is excellent at identifying high frequency coherent noise sources, such as line frequencies from power sources, while the AV is ideal for identifying low frequency noise sources, such as mechanical drift. In combination, the PSD and AV can be used to identify most forms of parasitic noise, and under

optimal conditions, both give accurate parameter estimates and uncertainties.

2.3 Modeling thermal motion in the PSD and AV

2.3.1 Langevin dynamics

Thermal motion-based calibration methods rely on Langevin dynamics, which model the trapped bead in an SMFS experiment as randomly diffusing in a harmonic potential. Collisions between the bead and water molecules create a stochastic (Langevin) force, F_L, that obeys the fluctuation-dissipation relation, $\langle F_L(t+t')F_L(t)\rangle = 2\gamma k_B T\delta(t')$, where $\delta(t)$ is the Dirac delta function. For a micron-scale bead in water, inertial effects only become important at microsecond timescales [96, 70], well below the sampling time, τ_s, of most SMFS video-tracking instruments ($\tau_s \gtrsim 10^{-4}$ s) [39]. Thus, the bead's motion is well-described by the overdamped Langevin equation:

$$\kappa x(t) + \gamma\dot{x}(t) = F_L(t). \tag{2.2}$$

2.3.2 A closed-form expression for the PSD

For bead motion, the predicted PSD, P, at each frequency, f, follows from Fourier analysis of Eq. 2.2:

$$P(f) = \frac{k_B T}{2\pi^2\gamma\left[\left(\frac{\kappa}{2\pi\gamma}\right)^2 + f^2\right]}. \tag{2.3}$$

For frequencies above the corner frequency, $f_c \equiv \kappa/2\pi\gamma$, the bead motion is purely diffusive and the PSD can be approximated as $P(f) \approx \frac{k_B T}{2\pi^2\gamma f^2}$. For frequencies below the corner frequency, the bead is constrained by the trap, and the PSD can be approximated as $P(f) \approx \frac{2k_B T\gamma}{\kappa^2}$.

Eq. 2.3 does not account for the exposure time of the camera, τ_0, which introduces a low-pass filter to the experimental bead positions. It can be included through a correction function, I [72]:

$$P_A(f) = P(f)I(f). \tag{2.4}$$

where

$$I(f) = \frac{\sin^2(\pi f \tau_0)}{(\pi f \tau_0)^2}. \tag{2.5}$$

Eq. 2.4 also needs to be adjusted for aliasing distortions: for an instrument with a sampling rate, $f_s \equiv 1/\tau_s$, the PSD at each positive frequency, f', $(0 < f' < f_s/2)$ contains the summed power of other frequencies, nf_s, for all integers, n [71]:

$$P_{A,B}(f) = \sum_{n=-\infty}^{\infty} P_A(|f + nf_s|). \tag{2.6}$$

In the special case that $\tau_0 = \tau_s$, the sum in Eq. 2.6 can be performed analytically to give an exact, closed-form expression for the PSD that accounts for both aliasing and the exposure time of the camera [61]:

$$P_{A,B}(f) = \frac{2k_BT\gamma}{\kappa^3}\left(\kappa + \frac{2\gamma f_s \sin^2\left(\frac{\pi f}{f_s}\right)\sinh\left(\frac{\kappa}{\gamma f_s}\right)}{\cos\left(\frac{2\pi f}{f_s}\right) - \cosh\left(\frac{\kappa}{\gamma f_s}\right)}\right). \tag{2.7}$$

Most modern video cameras are designed to maximize captured light, with a dead time ($\sim 10^{-6}$ s) that is much less than the sampling time ($\tau_s \sim 10^{-1}$ s to 10^{-4} s). This ensures that the exposure time is about the same as the sampling time, $\tau_0 = \tau_s$, i.e., zero dead-time, fitting the criteria for applying Eq. 2.7.

While sources of parasitic noise will vary among different SMFS instruments, most video-tracking experiments have a frame-to-frame tracking error arising from the imprecision of the bead localization algorithm. Assuming the tracking error is Gaussian-

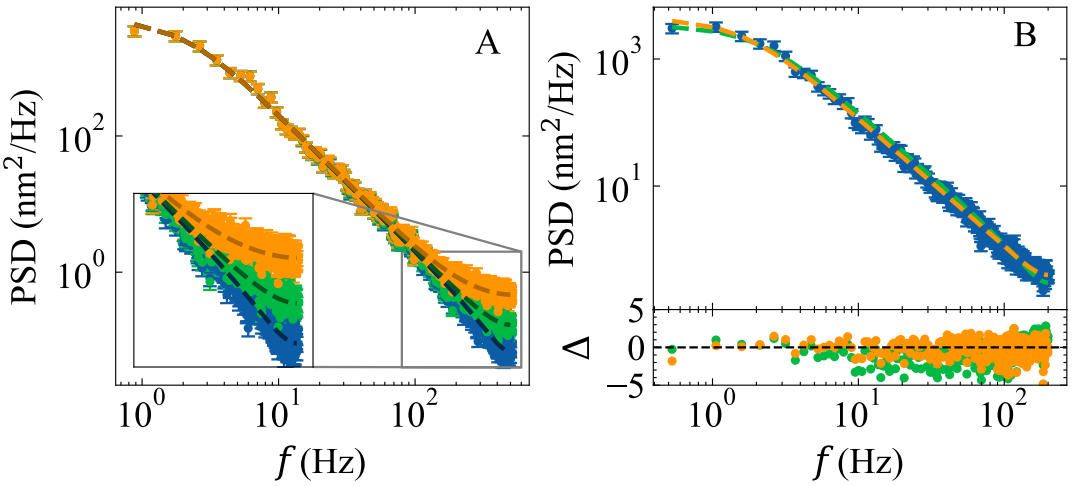

Figure 2.2: Tracking errors in simulated (A) and experimental (B) PSD values. The PSD values were computed using 35 half-overlapping bins. (A) Larger tracking errors lead to larger deviations in the simulated PSD values at higher frequencies. Simulations consisted of 20480 points carried out with parameters $f_s = 1000\,\text{Hz}$, $\gamma = 1.77 \times 10^{-5}\,\text{pNs/nm}$, $\kappa = 1.2 \times 10^{-4}\,\text{pN/nm}$, and $\epsilon = 0$, 10, and 20 nm (blue, green, orange). Dotted lines are overlays based on Eq. 2.8 with the known parameter values. (B) Experimentally derived PSD values are better described by an expression that includes tracking errors (Eq. 2.8, orange) than one that does not (Eq. 2.7, green) as judged by the normalized residuals, Δ, and Akaike Information Criterion (AIC), which balances the fit quality with the number of parameters (AIC = 334 vs. 739). The data were collected at 400 Hz on double-stranded DNA using a custom-built magnetic tweezer [98] with a video camera detection system and compared to both functions using MLE. The best fit parameters for $P_{A,B,C}$ are $\kappa = 1.8 \pm 0.2 \times 10^{-4}\,\text{pN/nm}$, $\gamma = 1.78 \pm 0.04 \times 10^{-5}\,\text{pNs/nm}$, and $\epsilon = 8.0 \pm 0.3\,\text{nm}$.

distributed with a standard deviation, ϵ, this adds a frequency-independent white noise term to $P_{A,B}$ [97]:

$$P_{A,B,C}(f) = P_{A,B}(f) + \frac{\epsilon^2}{f_s}. \tag{2.8}$$

In the PSD, the effect of tracking errors is most apparent at high frequencies, where the thermal motion is diminished (Fig. 2.2 A).

2.3.3 A closed-form expression for the AV

For bead motion, the predicted AV, σ^2_{AV}, at each observation time, τ, is similarly derived through analysis of Eq. 2.2 [61]:

$$\sigma^2_{AV,A}(\tau) = \frac{2k_BT\gamma}{\kappa^2\tau}\left(1 + \frac{2\gamma}{\kappa\tau}e^{-\frac{\kappa\tau}{\gamma}} - \frac{\gamma}{2\kappa\tau}e^{-\frac{2\kappa\tau}{\gamma}} - \frac{3\gamma}{2\kappa\tau}\right). \tag{2.9}$$

For observation times that are shorter than the bead relaxation time, $\tau \ll \tau_c \equiv \gamma/\kappa$, neighboring positions are highly correlated, and the AV increases as $\sigma^2_{AV,A} \approx 2k_BT\tau/3\gamma$. For observation times that are longer than the bead relaxation time, neighboring positions become uncorrelated, and the AV decreases as $\sigma^2_{AV,A} \approx 2k_BT\gamma/\tau\kappa^2$ [94]. The peak of the transition between the two regimes can be numerically calculated as $\tau_{\max} \approx 1.89\tau_c$ [61].

In its definition, the AV implicitly accounts for the exposure time of the camera and assumes zero dead-time, i.e., $\tau_0 = \tau_s$. As discussed in the previous section, this is usually a reasonable assumption for SMFS video-tracking systems. When this is not the case, the AV is biased and requires an additional correction function [99]. Conveniently, this bias is negligible when $\tau_s \ll \tau_c$ [94], so Eq. 2.9 can usually be applied, without further modification, to photodiode-based detection systems that include dead-time [100].

As with the PSD, tracking errors in video-tracking experiments can be accounted for by adding a white-noise term to $\sigma^2_{AV,A}$:

$$\sigma^2_{AV,A,B}(\tau) = \sigma^2_{AV,A}(\tau) + \frac{\epsilon^2\tau_s}{\tau}. \tag{2.10}$$

The effect of tracking errors is most apparent at short observation times, when the bead motion is mostly diffusive (Fig. 2.3).

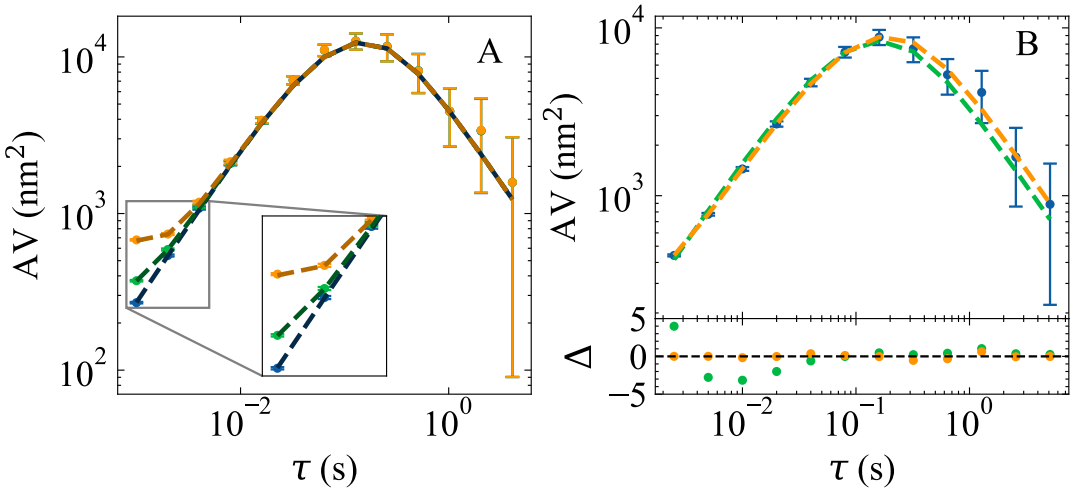

Figure 2.3: Tracking errors in (A) simulated and (B) experimental AV values. (A) Larger tracking errors lead to larger deviations in the simulated AV values at shorter observation times. The simulated data with tracking errors are the same as in Fig. 2.2, $\epsilon = 0$, 10, and 20 nm (blue, green, orange). Dotted lines are overlays based on Eq. 2.10 with the know parameter values. (B) Experimental derived AV values are better described by an expression that includes tracking errors (Eq. 2.10, orange), than one that does not (Eq. 2.9, green), as judged by the normalized residuals, Δ, and AIC (AIC = 161 vs. 198). These data are the same as in Fig. 2.2. The best fit parameters for $\sigma^2_{AV,A,B}$ are $\kappa = 1.7 \pm 0.1 \times 10^{-4}$ pN/nm, $\gamma = 1.77 \pm 0.04 \times 10^{-5}$ pNs/nm, and $\epsilon = 7.9 \pm 0.9$ nm.

2.4 Computing the PSD and AV

The SMFS experiment generates an experimental bead trajectory containing N_x points. This trajectory must be converted into a noise metric (the PSD or AV) containing N_y points, which is then fit with the expressions in Section 2.3 so as to extract parameter estimates. The conversion of the experimental trajectory to the noise metric has a few subtleties which are described here.

2.4.1 Computing the experimental PSD with Welch's method

The experimental PSD is optimally computed from the bead trajectory using Welch's method [101], i.e., by averaging periodograms. This method consists of splitting the

trajectory into half-overlapping bins, each containing m points. The total number of bins is thus $M = 2N_x/m - 1$. The PSD is calculated for frequencies $f_k = kf_s/2m$, for $k \in (1, 2, \ldots, m)$, where f_s is the sampling frequency; note the final number of calculated points is $N_y = m$. A smaller m improves the signal-to-noise ratio of the final experimental PSD, at the cost of reduced sensitivity at lower frequencies [74]. The PSD values for each bin, $\hat{P}_n$, are calculated from bead positions, $\hat{x}_j$, $j \in (0, 1, 2, \ldots, m-1)$, as

$$\hat{P}_n(f_k) = \frac{1}{mf_s} \left\| \sum_{j=0}^{m-1} w_j \hat{x}_j \exp\left(-\frac{2\pi i j k}{m}\right) \right\|^2, \tag{2.11}$$

then averaged together to give the final experimental PSD, $\hat{P}$:

$$\hat{P}(f_k) = \frac{1}{M} \sum_{n=1}^{M} \hat{P}_n(f_k). \tag{2.12}$$

The windowing function, w_j, accounts for the phenomenon of spectral leakage: the finite duration of the measurement causes power at one frequency to show up at other frequencies [102]. Most computational implementations of Welch's method use the Hann windowing function [103]:

$$w_j = \sqrt{\frac{8}{3}} \sin^2\left(\frac{\pi j}{b}\right), \tag{2.13}$$

reducing the total power of each experimental PSD value in a frequency-independent manner, which is then corrected by the leading factor of $\sqrt{8/3}$. The use of the Hann window, in conjunction with half-overlapping bins, means that data near the termini of one bin is diminished by the window, but that same data is near the center of the next bin, and thus captured by the window; this provides a reasonable trade-off between over- and under-utilizing all of the data [101].

In practice, most computer algorithms use a more computationally efficient fast

Fourier transform based on the Cooley-Tukey algorithm to calculate the experimental PSD values [104].

2.4.2 Computing the overlapping AV

The experimental AV is optimally computed from the bead trajectory by partitioning it into octave-sampled, overlapping bins. Octave sampling consists of using bin lengths, m_k, in powers of 2, i.e., $m_k = 2^k$ for $k \in (1, \ldots, N_y)$, where $N_y = \lfloor \log_2(N_x/2) \rfloor$. The bin lengths determine the number of overlapping bins, $M = N_x - 2m_k + 1$, and the observation times, $\tau = m_k \tau_s$, where τ_s is the sampling time. For each τ, the experimental AV, $\hat{\sigma}^2_{AV}$, is calculated as one-half the mean-squared difference of consecutive average bin positions:

$$\hat{\sigma}^2_{AV}(\tau) = \frac{1}{2(M-1)} \sum_{n=1}^{M-1} (\bar{x}_{n+1} - \bar{x}_n)^2 \tag{2.14}$$

where $\bar{x}_n$ is the average of bead positions, $\hat{x}_j$, $j \in (1, 2, \ldots, m_k)$:

$$\bar{x}_n = \frac{1}{m_k} \sum_{j=1}^{m_k} \hat{x}_j. \tag{2.15}$$

In practice, computing all the average bin positions for each τ can be slow, so an equivalent, but more computationally efficient, method is often used [99, 105].

2.5 Biased fitting

After computing the set of experimental AV or PSD values, $\hat{y}_k$, $k \in (1, 2, \ldots, N_y)$, they are compared to the Langevin model predictions, y_k (Eqs. 2.7-2.10), using maximum likelihood estimation (MLE), to extract the best-fit parameter estimates for γ and κ. MLE accounts for the expected probability distributions of each experimental value. For

the AV and PSD, the probability, p_k, of measuring each experimental value is given by the Gamma probability distribution function:

$$p_k(\hat{y}_k, y_k(\gamma,\kappa)) = \frac{\hat{y}_k^{\eta_k - 1} e^{-\hat{y}_k/\theta_k}}{\theta_k^{\eta_k}\Gamma(\eta_k)} \tag{2.16}$$

where η_k is termed the shape parameter, $\theta_k = y_k/\eta_k$ is termed the scale parameter, and Γ is the gamma function.

For the PSD, the shape parameter is given by the number of bins, $\eta_k = M$, which is notably the same for all values $\hat{y}_k$. For the AV, the shape parameter is generally $\eta_k = \nu_{AV,k}/2$, where $\nu_{AV,k}$ counts the degrees of freedom for each value. $\nu_{AV,k}$ depends on the number of differences used to calculate the k^{th} value, as well as the dominant type of noise at that value [99]. It is common to approximate $\nu_{AV,k}$ from the number of successive differences between non-overlapping bins of length m_k that are present in the trajectory, $\nu_{AV,k} = (N_x/m_k) - 1$ [61]; however, this is an underestimate.

For both the PSD and AV, as $\eta_k \to \infty$, the Gamma distribution approaches a normal (Gaussian) distribution, and least-squares fitting can be used. However, for moderate values of η_k, the distribution is not normal, and least-squares fitting routines lead to biased parameter estimates. While it is possible to correct for these biases analytically, in general, MLE gives more accurate parameter estimates [74].

2.5.1 Maximum likelihood estimation

MLE is based on estimating the parameters, $\hat{\gamma}$ and $\hat{\kappa}$, that maximize the likelihood function, L, which is the joint probability of all p_k:

$$L(\gamma,\kappa) = \prod_{k=1}^{N_y} p_k(\hat{y}_k, y_k(\gamma,\kappa)). \tag{2.17}$$

In practice, rather than maximizing L, it is more convenient to minimize the cost function, $\ell \equiv -\ln L$. Given Eqns. 2.16 and 2.17, the cost function is given by:

$$\ell(\gamma, \kappa) = \sum_{k=1}^{N_y} \eta_k \left[\frac{\hat{y}_k}{y_k(\gamma, \kappa)} + \ln(y_k) \right] + \text{const}, \tag{2.18}$$

where the final term is a constant with respect to the parameters. Minimizing ℓ is a straightforward optimization problem that can be solved numerically with standard algorithms (e.g., Nelder-Mead [106]).

2.5.2 Parameter uncertainties

After finding the best-fit parameters, $\hat{\gamma}$ and $\hat{\kappa}$, an estimate of their uncertainties can be found from standard approaches: In particular, the likelihood function, L, is assumed to have a Gaussian form in the vicinity of its maximum. Then, the matrix of second partial derivatives of L (i.e., the Hessian matrix) are calculated, and inverted to find the squared uncertainties (i.e., the covariance matrix). Details of this approach can be found in statistical references, e.g. Ref. [107].

The applicability and robustness of Hessian-based estimates of parameter uncertainty rests on whether L behaves as a Gaussian over a significant region near $(\hat{\gamma}, \hat{\kappa})$. This question is distinct from that of the proper distribution governing the AV or PSD estimates themselves (i.e, the values $\hat{y}_k$)– the $\hat{y}_k$ values, in certain cases, are calculated from a relatively small number of samples, and so are distributed in a highly non-Gaussian manner (Eq. 2.16), which drives the use of MLE rather than least-squares optimization methods. However, the MLE cost function is based on a relatively larger number of points (N_y), and so, by the central limit theorem, is well-modeled as Gaussian. Therefore, in practice, the Hessian approach typically results in robust estimates of parameter uncertainty.

That said, in some cases, it may not be appropriate to approximate the likelihood

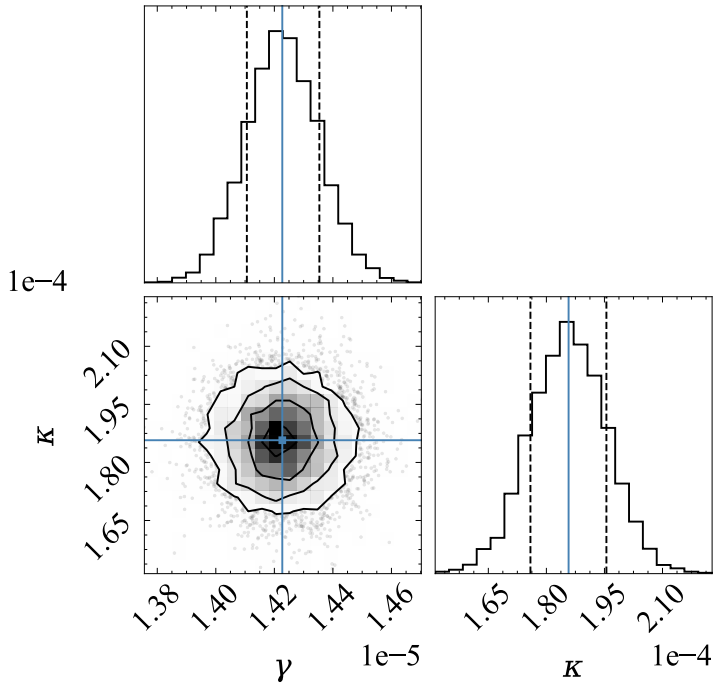

Figure 2.4: Parameter probability distributions represented as 1 and 2D histograms from Monte Carlo sampling. The distributions arise from the experimental data shown in Fig. 2.2B. On the 2D histogram, the contour (black) lines represent the 1,2, and 3 standard deviations. On the 1D histograms, the dotted (black) line represent 1 standard deviation. The blue line represents the best-fit MLE estimate.

function as a Gaussian, e.g., when there are small sample sizes, outliers, or complex parameter correlations. Such situations can be handled by an alternate, numerical approach in which a Monte Carlo algorithm is used to sample the parameter space [107]. To carry out Monte Carlo sampling, several 'walkers' are initiated around the estimated parameters. These 'walkers' take random steps in parameter space and evaluate the cost function, which determines whether each step is accepted or rejected. After a predetermined number of steps, a histogram of the accepted steps is used to generate an empirical probability distribution for the parameters (Fig. 2.4). From this distribution, the confidence intervals can be evaluated. Typically, the standard errors are estimated as half the difference between the 15.8th and 84.2nd percentiles, which corresponds to one standard deviation for a Gaussian distribution.

2.5.3 Fit quality

After fitting, the quality of the fit needs to be judged. There are several means of judging the quality of the fit, each with its own advantages and disadvantages. The simplest means is to look at the normalized residuals, Δ_k, i.e., the deviations between the experimental and predicted values:

$$\Delta_k = \frac{\hat{y}_k - y_k(\hat{\gamma}, \hat{\kappa})}{\sigma_{y,k}}, \tag{2.19}$$

where $\sigma_{y,k}$ is the standard deviation of the k^{th} experimental value. The normalized residuals can be plotted to assess systematic deviations between the data and the fit.

If the normalized residuals follow a Gaussian distribution, their variance corresponds to the reduced chi-squared value, $\chi^2_{\nu_y}$:

$$\chi^2_{\nu_y} = \frac{\chi^2}{\nu_y} = \frac{1}{\nu_y} \sum_{k=1}^{N_y} \Delta_k^2. \tag{2.20}$$

The degrees of freedom, ν_y, are estimated as $\nu_y = N_y - K$, where K is the number of parameters. A reduced chi-squared value of one is usually considered a 'good' fit [107]. A reduced chi-squared value that is greater than one is generally considered a 'poor' fit, whereas a reduced chi-squared value that is less than one is usually considered an overfit. However, the reduced chi-squared value has a variance that scales as $2/\nu_y$, so values based on small sample sizes or models with a large number of parameters can be misleading.

Instead, the cumulative distribution function of chi-squared-distributed values, F, is usually a better measure of fit quality (also termed the support for the fit) [75]:

$$F(\chi^2, \nu) = \frac{1}{\Gamma(\nu_y/2)} \int_0^{\chi^2/2} z^{\nu_y/2-1} \exp(-z) dz. \tag{2.21}$$

The support evaluates the probability that repeating the experiment will give a larger $\chi^2_{\nu_y}$ value. It is closely related to the p-value, i.e., $1 - F$. For a 'good' fit, the support is expected to be close to one.

While the support for the fit evaluates agreement between the experimental and predicted values, other statistical metrics, such as the Akaike Information Criterion (AIC), are better at comparing models with different numbers of parameters [108]. The AIC balances the quality of fit with the number of parameters. It is calculated as

$$\text{AIC} \quad = \quad 2K - 2\ln(\hat{L}). \tag{2.22}$$

Due to varying constants and sample sizes, individual AIC values are not informative. Instead, the data are considered to be best described by the model with the lowest AIC value, AIC_{min}, regardless of the number of parameters, when the difference between two models' AIC values is $\Delta_{\text{AIC}} = \text{AIC} - \text{AIC}_{\text{min}} \geq 4$ [108], as applied in Figs. 2.2 and 2.3.

2.6 Tweezepy

Tweezepy is a Python package for thermal motion-based force calibration in SMFS video-tracking experiments that estimates parameters and their uncertainties from a user-provided bead trajectory, using MLE, via the PSD or AV. For a detailed explanation of the package, including expected inputs and outputs, the reader is referred to the docstrings and usage examples. In this section, we discuss specific implementation choices and practical considerations for using the package.

To use Tweezepy, the user provides a bead trajectory and sampling frequency to either the PSD or AV class objects. Given this information, Tweezepy computes the experimental values and compares them to a user-selected predictive model using MLE.

After fitting, it reports the parameter estimates and uncertainties, as well as the fit quality. The experimental and predicted values, as well as the normalized residuals, can be visualized using the included utility plotting functions.

To compute the experimental PSD, Tweezepy uses Welch's method (Sec. 2.4.1). By default, it uses a Hann windowing function with three half-overlapping bins. The signal-to-noise ratio of the experimental PSD values can be improved by increasing the number of bins, which helps to visualize the values and slightly reduces the parameter uncertainties. However, there is a trade-off: as the number of bins increases, the low-frequency resolution decreases. For low corner frequencies, this can lead to a substantial bias in the parameter estimates (Fig. 2.5). Unfortunately, it is difficult, *a priori*, to know the optimal number of bins, so it is up to the user to choose the appropriate number of bins. This is a distinct drawback of the PSD method.

Tweezepy uses MLE to compare the experimental PSD values to, by default, Eq. 2.7, which accounts for both aliasing and the finite bandwidth of the detection system. This function assumes the exposure time is the same as the time between measurements, i.e., zero dead-time. As discussed in Section 2.3.2, this assumption is typically good for video-tracking experiments, but may not be appropriated for detection systems with dead-time (e.g., photodiodes). When dead-time is present in the measured bead trajectory, the user can also select an alternative function that uses a closed-form expression based on Eq. 2.6 that assumes a negligible exposure time [74, 61], i.e., it only accounts for aliasing. Additionally, the user can select to use a modified version of either function that includes tracking errors from video-tracking bead localization algorithms (e.g., Eq. 2.8).

To compute the experimental AV, Tweezepy uses the octave-sampled overlapping AV. It empirically determines the degrees of freedom for each value using the Greenhall algorithm [109], based on the dominant type of power-law noise for each experimental value. It estimates the dominant type of noise using the Lag1 autocorrelation algorithm [110].

This algorithm has lower precision for AV values with fewer bins, i.e., for long observation times when $N_x/m_k < 32$, so Tweezepy uses the previously estimated noise type for those values. If the algorithm fails to estimate the dominant type of noise, it warns the user and falls back on using the approximate degrees of freedom based on nonoverlapping bins, i.e., $\nu_{AV} = N_x/m_k - 1$. The user can choose to use only the approximate degrees of freedom by setting the keyword argument 'edf' to 'approx'. For visualization purposes, the user can select to plot all or decade-spaced observation times. As discussed in Section 2.4.2, the approximate degrees of freedom give nearly identical parameter estimates but underestimate the confidence for each AV value, leading to slightly larger parameter errors. After computing the experimental AV values, Tweezepy compares them to Eq. 2.9. Additionally, the user can select a predefined function that accounts for tracking errors from the video-tracking bead localization algorithms (Eq. 2.10).

In addition to its predefined functions, Tweezepy also accepts user-defined functions to compare to the experimental values. If these functions include additional fitting parameters, it is recommended that they are compared to a function without the additional parameters using the AIC to avoid overfitting (Sec. 2.5.3). Additionally, the normalized residuals can be plotted and visualized to detect deviations between the data and theoretical values. Typically, it is easier to visualize the residuals of the AV compared to the PSD because it has fewer values.

Evaluating the AIC can also be useful for determining whether one or more parameters is poorly constrained during the fit. As discussed later (Sec. 2.7), in some cases, the sampling frequency is not fast enough to resolve the purely diffusive motion of the bead, causing γ to be poorly constrained during fitting. However, κ can usually still be reliably estimated by fixing γ to a known value. Tweezepy contains keyword arguments for fixing any of the parameters for its predefined functions during the fit. Ideally, the known γ value should be estimated from the same bead at a lower force, and adjusted for

surface effects using Faxen's correction [111]. To determine whether fixing a parameter is necessary, the AIC of the fits with and without fixing are compared, and the fit with the lowest AIC value is used.

When sources of parasitic noise are present but cannot be properly described by the selected model analytically, it is recommended that the user subtract a reference spectrum or bandpass filter the measured data. Alternatively, the user can select upper and lower cutoff frequencies (observation times) to compare the function to a limited region of the data using the keyword argument 'cutoffs'.

To calculate parameter uncertainties, Tweezepy evaluates and inverts the expected Hessian (Sec. 2.5.2). To evaluate the Hessian, it uses the Autograd Python package. Autograd uses automatic differentiation to evaluate derivatives by repeatedly applying the chain rule to elementary operations. This speeds up code and reduces numerical precision errors that can occur with numerical and symbolic differentiation [112, 113]. In addition to calculating and inverting the Hessian, Tweezepy contains an optional method for robust uncertainty estimates via Monte Carlo sampling (Sec. 2.5.1). This method uses the Emcee Python package [114] to carry out Monte Carlo sampling. In our hands, this more robust, but slower (with computation time on the order of 10 s), method produces near-identical uncertainty estimates to the faster ($\approx$ 10 ms) method that inverts the Hessian.

In addition to the packages mentioned above, Tweezepy makes use of the standard python library [115], including NumPy [116], SciPy [117], and Numba [118]. All the package dependencies are noted in the requirements and setup files for easy installation.

2.7 Results

To evaluate Tweezepy, we sought to benchmark its fit results against known parameter values. Following the example of Ref. [61], we simulated bead trajectories using $N_x = 4096$, $f_s = 100\,\mathrm{Hz}$, and $\gamma = 1.0 \times 10^{-5}\,\mathrm{ps/nN}$ (a typical drag coefficient for a one micron spherical bead in water), and varied the corner frequency, f_c, logarithmically from $0.2\,\mathrm{Hz}$ to $100\,\mathrm{Hz}$, giving spring constants κ that ranged from $1.4 \times 10^{-4}\,\mathrm{pN/nm}$ to $6.8 \times 10^{-3}\,\mathrm{pN/nm}$. To carry out the simulations, we recast Eq. 2.2 as an equation of probability flux and iteratively generated successive bead positions. To mimic the effects of the camera exposure time, we used a time step of $\delta t = 1000/f_s$, split the trajectory into bins of 1000 points, and took the average of each bins to generate a downsampled trajectory. For each corner frequency, we simulated 1000 trajectories and used both calibration methods in Tweezepy to estimate the parameters and their uncertainties. To estimate bias, we calculated the ratio of the median parameter estimates and true values. To estimate the error, we calculated the ratio of the median parameter uncertainties and true values.

For nearly all corner frequencies, the bias for γ and κ estimates is within $\pm 1\%$ (Fig. 2.5 A and C magenta box). There is an increase in the bias and error for κ estimates at lower corner frequencies because, for the simulated length of the trajectory, the bead motion is mostly unconstrained by the trap. As a result, the κ estimate is poorly constrained during fitting. This effect is slightly worse for the PSD because binning decreases its low frequency resolution more than the AV. In practice, this bias can usually be reduced by increasing the length of the trajectory.

At high corner frequencies, $f_c \gtrapprox f_s/8$, there is a minimum in the error, after which, the error and bias increase for both parameters (Fig. 2.5 C and D), consistent with previous findings [74, 61]. This is because the sampling frequency is not fast enough to

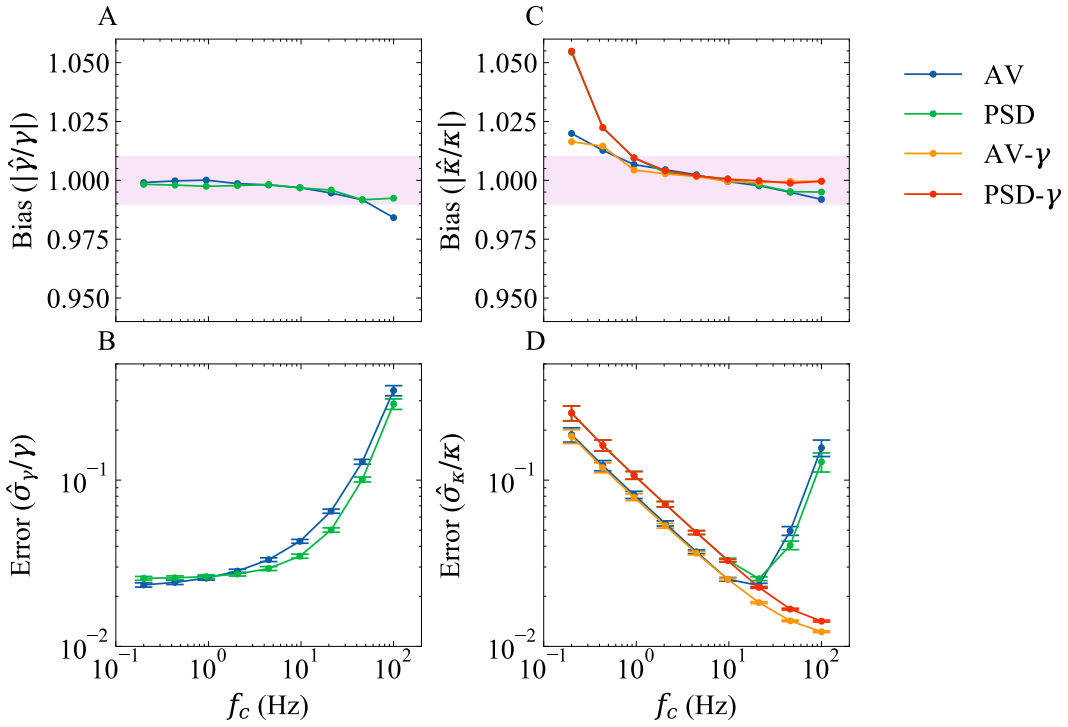

Figure 2.5: Bias and error for the AV and PSD methods in Tweezepy. Each point represents the median of 1000 simulations; each simulation contained 4096 bead positions with a constant drag coefficient, $\gamma = 1 \times 10^{-5}$ ps/N, and sampling frequency, $f_s = 100$ Hz. The corner frequency was varied logarithmically between 0.2 Hz to 100 Hz. The blue and green points represent two-parameter AV and PSD method fit results. The orange and red points represent fixed gamma AV and PSD method fit results. In the bias plots, the magenta box represents the $\pm 1\%$ bias region.

resolve the unconstrained diffusive bead motion. As a result, the γ estimate is poorly constrained during fitting, which leads to poor estimates for both parameters. In practice, for force calibration in video-tracking experiments, it is usually advantageous to collect bead trajectories at the highest available sampling frequency.

It is worth noting that the authors in Ref. [79] recommend using a low-pass-corrected standard variance calibration method [72] to avoid the small bias at high corner frequencies with the PSD and AV. However, we note that their implementation of this alternative method fixes γ to a known value during fitting. We find that fixing γ with the PSD and AV similarly removes the increased bias and error at high corner frequencies (Fig. 2.5 C and D). This suggests that, under optimal conditions, all three methods can accurately estimate parameters.

2.8 Conclusions

In this article, we have reviewed robust thermal motion-based force calibration in SMFS experiments using the PSD and AV, and discussed implementing them computationally into a Python package, Tweezepy, that is freely available on Github and the Python package index.

In designing Tweezepy, our goal was to make it as robust, versatile, and user-friendly as possible. It uses MLE to estimate parameters via the PSD or AV, and goes beyond previous computational implementations by calculating the empirical degrees of freedom for the overlapping AV and determining parameter uncertainties from MLE, either by inverting the Hessian or, optionally, via Monte Carlo sampling. It includes several predefined closed-form expressions that account for the most common biases and parasitic noise in SMFS video-tracking experiments. Yet, it also accepts user-defined functions, so it can be adapted to account for additional sources of noise or applied to other problems

that rely on fitting the PSD or AV of a bead trajectory, e.g., torque calibration [100]. Lastly, Tweezepy uses sensible default options to make it easy-to-use, only requiring a few straightforward lines of code, with computation times on the order of 10 ms. Our hope is that Tweezepy can serve as a useful tool to improve and standardize force calibration across different SMFS research groups, instruments, and experiments.

Chapter 3

Surface-induced Effects of Polymer Elasticity

3.1 Preface

My original goal was to characterize the low-force entropic elastic response of the NFLt polyproteins. During these early experiments, I often encountered short ($\lesssim 0.5\,\mu m$) polyproteins due the polydispersity of the samples. These shorter polymers often exhibited an anomalous decrease in the powerlaw scaling metric of the force-extension response at low forces. Similar behavior had been observed with other polymers. Hence, to facilitate my experiments on the NFLt polyproteins, I sought to understand these experimental biases using other well-characterized polymers.

The content of this chapter previously appeared in the Journal of Chemical Physics [119]. It is reproduced here with permission from AIP publishing.

3.2 Introduction

The mechanical manipulation of single molecules permits direct insight into polymeric elasticity and, in turn, polymer structure. This methodology has proven quite powerful and broadly applicable, having given insights into the structure of long biopolymers of every type (proteins, polysaccharides, nucleic acids), along with a variety of synthetic polymers. For example, single-molecule stretching has enabled understanding of entropic elastic effects in muscles[120] and permitted the study of solution electrostatic effects in nucleic acids[121]. Theory plays a central role in these advances, particularly as statistical mechanical methods are capable of predicting polymeric force-extension behaviors. Such approaches, in combination with experimental abilities to precisely measure and control force and polymer extension, allow direct, statistically significant comparisons between the model and data, permitting confident validation/negation of polymer structural understanding.

The majority of prior experimental and theoretical work has focused on relatively high-force elastic behavior, with less emphasis placed on understanding low-force elasticity. Low forces are those that permit the chain to loop back on itself, corresponding to forces less than k_BT/ℓ, where ℓ is the chain's Kuhn length, and k_BT is the thermal energy[40]. For most flexible biopolymers (e.g. single-stranded nucleic acids, disordered proteins, polysaccharides), ℓ ranges from one to ten nanometers, so typical values of k_BT/ℓ are 0.4 to 4 pN. Such small forces can be reliably controlled by magnetic-tweezer manipulation methods[39, 40], and are relevant to most biological and biomaterial situations where polymers are frequently under little to no tension.

In the low force regime, polymers adopt a random walk structure on short length scales, and models based on the classic scaling approaches of polymer physics become appropriate. The key physical metric defining such behavior is typically a power-law

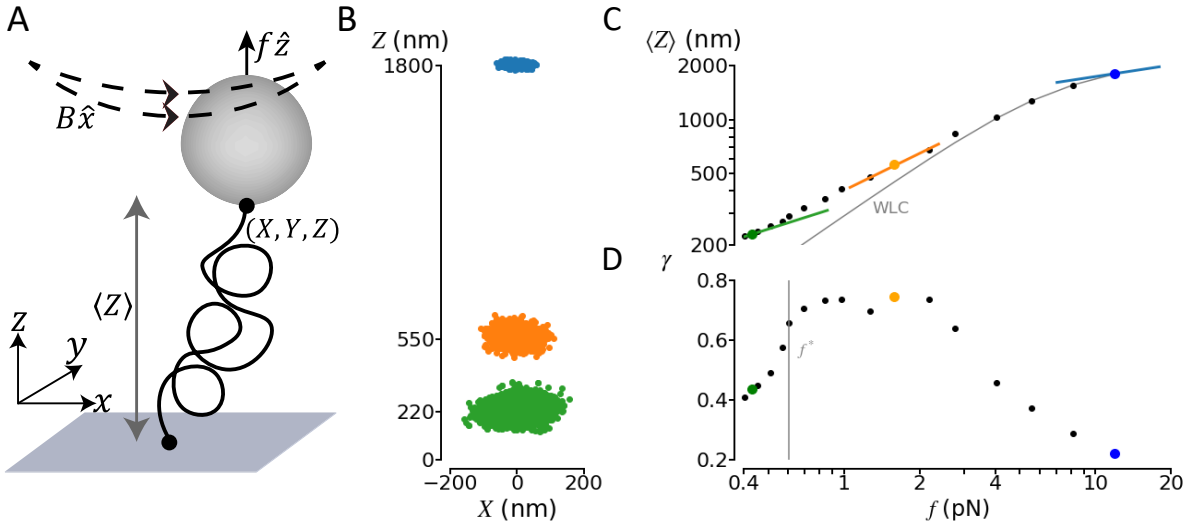

Figure 3.1: (A) Sketch of magnetic tweezer experimental geometry: A paramagnetic bead subject to a field B oriented in $\hat{x}$, but with a gradient in $\hat{z}$, feels a force $F\hat{z}$, stretching the attached polymer. (B) Typical experimental measurement of fluctuating bead position for three representative constant forces, 12 pN (blue; 895 points), 1.6 pN (orange; 2175 points) and 0.42 pN (green; 7743 points), on a tethered PEG chain. (C) Black dots: absolute extension, $\langle Z \rangle$, vs. F curve for the same PEG chain as in panel B. Colored lines indicate estimates of local power-law slope, γ, resulting from applying Eq. 3.1b to the measured fluctuations for the three representative forces. The gray line is a fit of the worm-like chain elastic function to the high force ($F > 4$ pN) data, returning $L_c = 2900$ nm and $l_p = 0.6$ nm. (D) Full curve of γ vs. F for the same PEG tether, showing a downturn for forces below $F^* = 0.6$ pN.

exponent. Here, the relevant metric is the Pincus exponent [41], $\gamma \approx 2/3$, which dictates that the polymer extension grows as f^{γ} with force, f, for a self-avoiding chain in the regime $f \lesssim k_B T/\ell$. Direct measurement of γ in a single-molecule experiment has been accomplished [122]; however, there is not a full understanding of the experimental conditions and their potential effects in biasing estimates of γ. One key issue is the effect of surfaces. Surfaces are an unavoidable component of magnetic tweezer experiments, as the polymer under study must be attached at both ends to mechanically-rigid points (typically a glass surface and paramagnetic bead; see Fig. 3.1A).

We explore the effect of surfaces in biasing experimental estimates of γ. Our focus is on a method of estimating γ from single-polymer extension fluctuations[123, 40]. We show that this method leads, at very low forces, to an unexpected decrease in the γ estimate in measurements of both Hyaluronic Acid (HA) and Polyethylene glycol (PEG) chains. To explain this decrease, we formulate a model permitting an estimate of the effect of both surfaces in restricting the available configurations of a chain under tension. The calculation replicates the observed low-force decrease in γ, and predicts that the decrease begins when the external force is $\approx 3k_B T/R_g$, where R_g is the chain's radius of gyration. This prediction is validated through comparison with the PEG and HA data. We show that estimates of R_g based on the decrease in γ are consistent with those generated from knowledge of microscopic polymer parameters, themselves gleaned from other parts of the force-extension curve.

3.3 Fluctuation estimates of polymer elasticity

Resolving power-law exponents by direct fitting to sections of extension/force curves is difficult, as it is not clear which points are fully within a given regime, and which are affected by neighboring regimes. Further, in the low-force regime, absolute extension

becomes very small, leading to sensitivity to systematic errors based on, e.g., the precise tethering position of the polymer to the probe [124, 125]. An alternate approach is based on analysis of the fluctuations in polymer extension[123, 40]. The fluctuation based method is an application of linear response theory [126], and is discussed in detail elsewhere [40]. We consider a force oriented in the $\hat{z}$ direction that leads to a mean extension $\langle Z \rangle$ of the chain (see Fig. 3.1A). Assuming a power law elastic relation $\langle Z \rangle \sim f^{\gamma}$, the effective exponent in the vicinity of force f can be estimated from

$$\gamma(f) = \frac{f \mathrm{var}(Z)_F}{k_B T \langle Z \rangle_F} \tag{3.1a}$$

$$= \frac{\mathrm{var}(Z)_F}{\mathrm{var}(X)_F} \tag{3.1b}$$

where var indicates the variance of a parameter about its mean, and Eq. 3.1b adjusts Eq. 3.1a by substituting the pendulum physics result for lateral thermal fluctuations, $\mathrm{var}(X) = k_B T \langle Z \rangle / F$, as is commonly applied for magnetic tweezer experiments [89]. We use the lateral fluctuations in the direction of the magnetic field (here, $\hat{x}$; see Fig. 3.1) to avoid the significant confounding effect of extra fluctuations due to bead rotational motion in $\hat{y}$.

Eq. 3.1b does not rely on knowledge of the mean extension of the chain. Thus, the resulting exponent estimate is independent of that found from the force-extension curve. Further, measurement of $\mathrm{var}(Z)$ and $\mathrm{var}(X)$ is robust, since it relies on sensing large changes in relative position of the bead, which is insensitive to the systematic errors in measuring $\langle Z \rangle$. Experimental application of Eq. 3.1b results in a fluctuation estimate of slope that matches the direct force-extension curve at moderate and high forces (Fig. 3.1C, D). The estimated γ is consistent with 2/3 at moderate force before decreasing as the chain extension approaches its contour length.

At lower forces, the fluctuation estimate of γ clearly decreases; however, it is difficult

to judge if this downturn agrees with the direct force-extension curve. Further, this downturn does not match simple theoretical expectations at low force, which predict γ would increase from 2/3 to 1 in the lowest force linear-elastic regime[127, 40].

3.4 Model of polymer/surface interactions

The major goal of this chapter is to explain the low force downturn in γ observed in Fig. 3.1D. We postulate it is caused by polymer/surface interactions, particularly the exclusion of the monomers from the volume occupied by the two tethering surfaces. Indeed, in this situation, at very low forces, one expects the chain extension to plateau near R_g, leading to an incompliant state that would decrease γ [128]. However, in the experimental data, the γ downturn is clear at extensions much larger than R_g. For example, for the PEG chain shown in Fig. 3.1, the downturn is clear at $\langle Z \rangle \approx 300$ nm, while we estimate this chain has a much smaller value of $R_g \approx 43$ nm. The R_g estimate is found from the best-fit contour length, $L = 2900$ nm, and using prior estimates of 0.278 nm contour length per PEG monomer [129], the monomer molecular weight of 44 Da, and the relation of R_g to PEG molecular weight M_w, $R_g = 0.0215 M_w^{0.583}$, found from light scattering measurements [130].

To test whether polymer/surface interactions can explain the γ downturn, we formulate a model of the surface's effect on the polymer configurational distribution. We use the image principle to estimate the number of allowed configurations of an ideal random-walk chain in the presence of a surface [131, 132, 133]. The experimental chains are non-ideal, as shown by the swollen-chain ($\gamma \approx 2/3$) behavior at moderate elasticity (Fig. 3.1); however, swollen chain behavior is significantly more difficult to analyze. We thus focus on a tractable ideal-chain model, with the goal of testing whether the surface effect can indeed decrease γ at relatively large chain extensions.

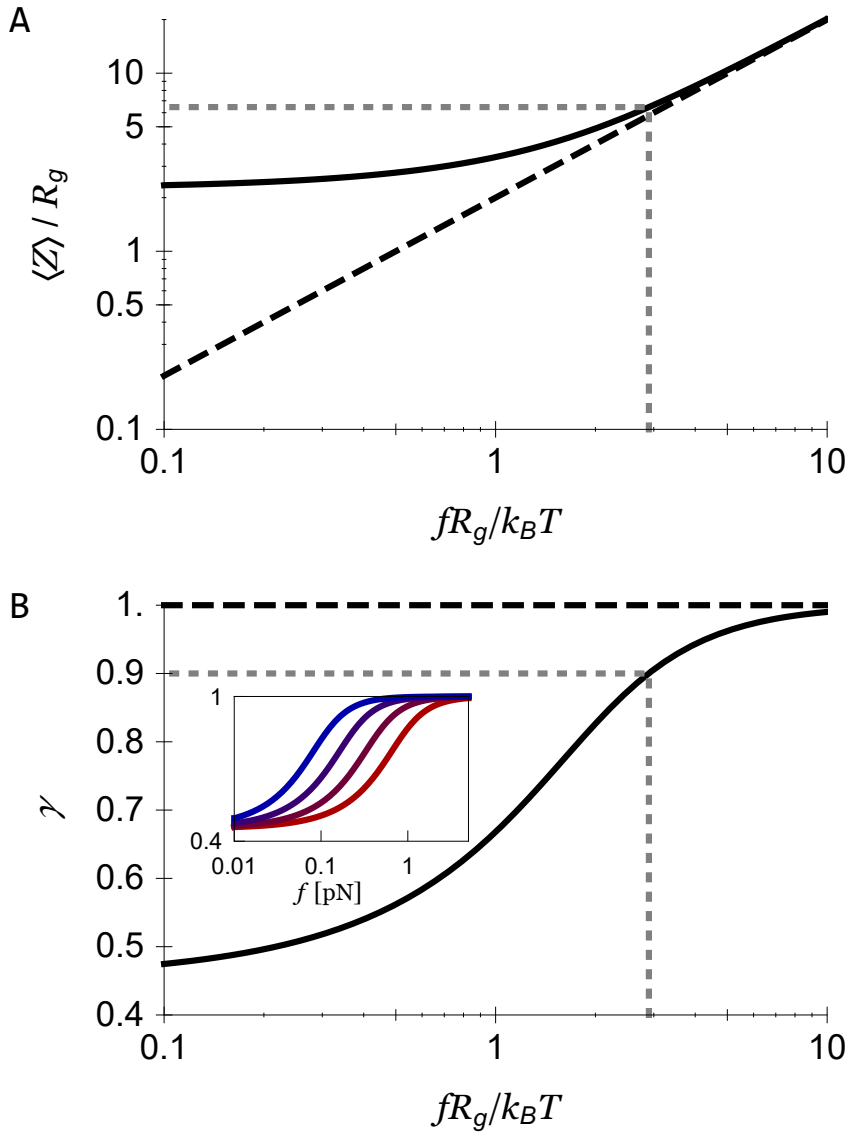

Figure 3.2: (A) Dimensionless extension vs. force relation computed from the model for an ideal chain tethered between two parallel planar surfaces (solid line), compared with that of an unconstrained entropic spring (dashed line). (B) γ vs. dimensionless force computed from the model using Eq. 3.1b (solid line), compared with the unconstrained entropic spring expectation $\gamma = 1$ (dashed line). In both panels, the coordinates at which $\gamma = 0.9$ are emphasized with gray dashed lines. Inset, panel (B): γ vs. absolute force for chains with (right to left) $R_g = 10, 20, 40$ and 80 nm.

When applied to a single excluded surface[131], the image method gives the statistical weight (proportional to the number of allowed configurations), $G_{1S}(\vec{R},\vec{R}')$, of an ideal chain tethered at position $\vec{R}$ to a planar surface (taken to be the $z=0$ plane), and whose opposite end is free and located at $\vec{R}'$:

$$G_{1S}(\vec{R},\vec{R}') = e^{-\frac{(X-X')^2}{4R_g^2}} e^{-\frac{(Y-Y')^2}{4R_g^2}} \left(e^{-\frac{(Z-Z')^2}{4R_g^2}} - e^{-\frac{(Z+Z')^2}{4R_g^2}} \right). \tag{3.2}$$

Here, the subtraction removes forbidden configurations (i.e. those that would intersect with the surface) from the ensemble of free chains (i.e. the permitted chain configurations in the absence of the surface). The Gaussian form of Eq. 3.2 enforces an infinitely-extensible polymer; this is unphysical, but consistent with our focus on low-force, low-extension behavior. It does lead to disagreement with experiment in the high-force regime, when the finite extension of the actual polymer is approached.

We rewrite Eq. 3.2 as $G_{1S}(\vec{R},\vec{R}') = G_{free}(\vec{R},\vec{R}')(1-p_{1S}(\vec{R},\vec{R}'))$, where G_{free} is the statistical weight of an unconstrained chain, and p_{1S} is the probability that a member of the unconstrained ensemble will follow a forbidden path, given by

$$p_{1S}(\vec{R},\vec{R}') = e^{-ZZ'/R_g^2}. \tag{3.3}$$

We must account for two surfaces, since, in the experiment, the chain is tethered to both a planar surface and a spherical bead. Since the bead radius (≈ 525 nm) is much larger than R_g, we treat it as a second planar surface, parallel to the first. This situation is entirely symmetric with respect to the two surfaces, which means that p_{1S} is the same for each surface individually. We then posit that $(1-p_{1S})^2$ is a good estimate for the probability that a member of the unconstrained ensemble will not intersect either surface. This estimate assumes there is no correlation in the chance of intersection between the

two surfaces. In reality, there is likely some correlation– particularly, the probabilities are likely anti-correlated, since a path that loops back to intersect one surface has less contour available to reach the other surface. However, we assume here that this is a small effect.

The statistical weight in the presence of both surfaces is then

$$G_{2S}(\vec{R},\vec{R}') = G_{Free}(\vec{R},\vec{R}')\left(1 - p_{1S}(\vec{R},\vec{R}')\right)^2. \tag{3.4}$$

To apply the image method to a tethered polymer, the terminus is fixed at a small distance, d, from the tethering surface[131]; this accounts for the unique chemical ability of the functional group to bind to the surface in a position disallowed to the monomers. So, for surface separation Z, the end monomers are located at $\vec{R}_1 = (0,0,d)$ and $\vec{R}_2 = (X,Y,Z-d)$. Further, we include the effect of force by weighting each configuration by the Boltzmann factor of the work done by the system, e^{fZ/k_BT}. Thus, the partition function for the system is

$$\mathcal{Z} = \int_{-\infty}^{\infty} dX \int_{-\infty}^{\infty} dY \int_{0}^{\infty} dZ e^{fZ/k_BT} G_{2S}(\vec{R}_1,\vec{R}_2), \tag{3.5}$$

where the bounds of the integrals reflect the bead's ability to move laterally, but its inability to pass through the glass surface. While Eq. 3.5 can be analytically integrated, the result is quite complex. In practice, we work with an approximate form of $\mathcal{Z}$, taken in the $d/R_G \ll 1$ limit, where the dependence on d drops out of the results.

We calculate ensemble average quantities of the parameters of interest, $\langle Z\rangle$, $\langle Z^2\rangle$, and $\langle X^2\rangle$, from the partition function, Eq. 3.5, in the small d limit. We set $\mathrm{var}(Z) = \langle Z^2\rangle - \langle Z\rangle^2$. Note that $\mathrm{var}(X) = \langle X^2\rangle = 2R_g^2$ for all forces. The model prediction for γ is found as the ratio of variances, per Eq. 3.1b.

3.5 Results and comparison to data

The model produces the expected entropic-spring behavior, $\langle Z \rangle = 2R_g^2 F/k_B T$, for an ideal chain in the high force limit (Fig. 3.2A). Similarly, in Fig. 3.2B, we see that the model's estimate of the variance ratio results in the expected value, $\gamma = 1$, in the high force limit. For an unconstrained ideal chain, linear behavior should persist to low forces (dashed lines in Fig. 3.2). In contrast, the model prediction for γ shows a downturn (Fig. 3.2B), in qualitative agreement with experiment (Fig. 3.1D). At a similar force, the force-extension curves show a transition to a regime where $\langle Z \rangle$ is independent of force. By approximating the partition function, we find the predicted low-force limiting values: $\lim_{F\to 0}\langle Z \rangle = 4R_g/\sqrt{\pi} \approx 2.26R_g$ and $\lim_{F\to 0}\gamma = 3 - (8/\pi) \approx 0.45$. The latter estimate is similar to the variance ratio estimated from simulations of self-avoiding chains constrained by tethering to a single surface [134].

The calculated surface-induced decrease in γ occurs at a force that scales with $k_B T/R_g$, as suggested based on inspection of the form of the partition function. The dependence on R_g is made explicit in the inset of Fig. 3.2, which shows γ versus absolute F (in pN) for a range of R_g values. As shown, the downturn moves to higher forces as R_g decreases.

While the location of the predicted γ downturn scales with $k_B T/R_g$, the actual predicted location has a greater-than-unity multiplicative prefactor in both force and length. We judge the downturn as having begun when the γ value decreases by about 10% (this is when it is experimentally visible, as shown by the placement of F^* in Fig.3.1). The model predicts that, when $\gamma = 0.9$, the force is $2.9k_B T/R_g$, and the extension is $6.5R_g$; these values are indicated by the gray lines in Fig. 3.2.

To gain more insight into the actual location of the downturn, and to confirm experimentally whether $k_B T/R_g$ is the controlling force scale for the γ downturn, we turn to a broad set of single-molecule data. In particular, we analyzed 44 force-extension curves

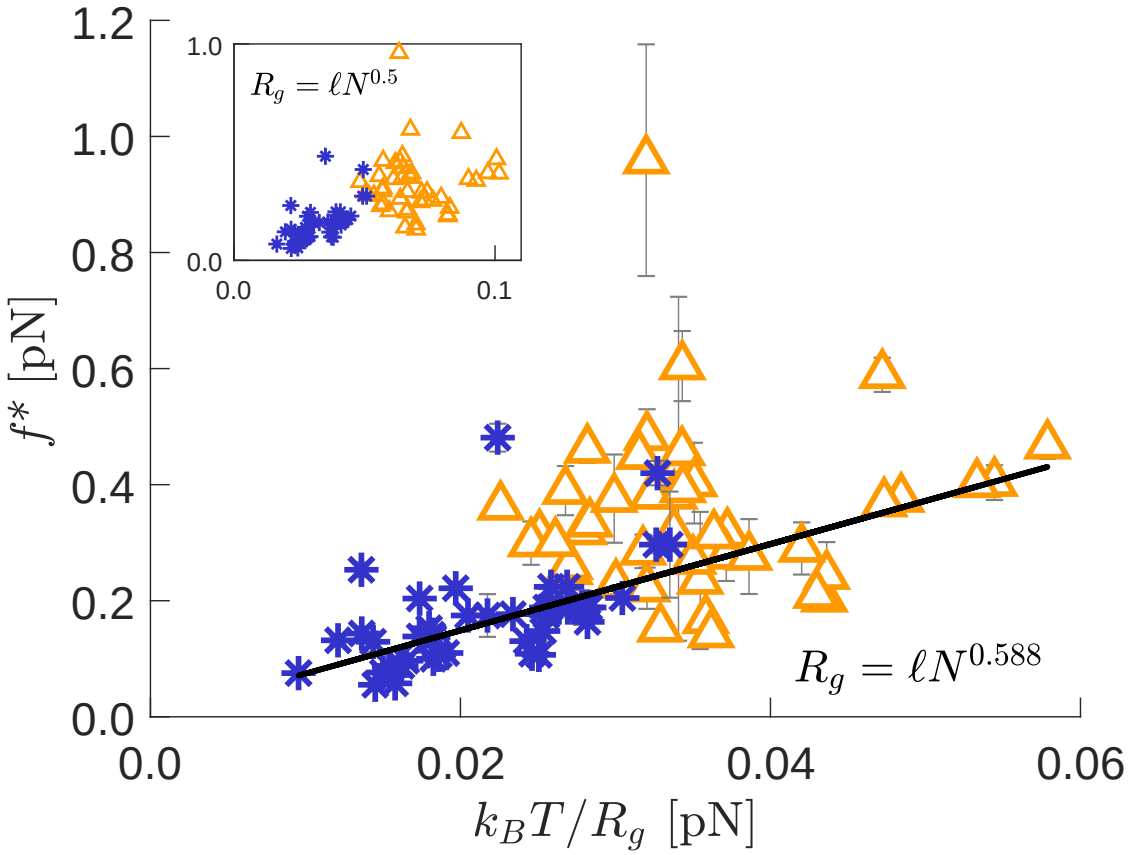

Figure 3.3: Experimental correlation of F^* with k_BT/R_g for single polymers of HA (blue stars) and PEG (orange triangles). F^* is estimated from the low-force downturn in the fluctuation estimate of γ (see Fig. 3.1D), and R_g is estimated from the parameters extracted from high-force Worm-like Chain (WLC) fits (see Fig. 3.1C), with $\ell = 2l_p$ and $N = L_0/\ell$. The HA and PEG data show better overlap when using self-avoiding estimates of R_g (main plot) rather than ideal estimates (inset). The self-avoiding data is well-fit by a line passing through the origin, with slope 7.4 ± 0.3. Error bars reflect the discrete sampling of force and/or an intrinsic 5% uncertainty in force calibration.

acquired from single chains of PEG, and 43 curves acquired from single chains of HA (details in Appendix C, and an example HA data set, Fig.). For the polyelectrolyte HA, data was acquired over a range of salt concentrations, which tunes the effective Kuhn length of the chain[135]. We apply Eq. 3.1b to the data, generating a γ vs. F curve, then estimate F^* from the observed low-force downturn (as in Fig. 3.1D).

We separately estimate the chain Kuhn length ℓ and number of Kuhn monomers $N = L_0/\ell$ by fitting the high-force portion of the $\langle Z \rangle$ vs. F curve to the WLC elasticity function[136, 137] (see Fig. 3.1C), and setting $\ell = 2l_p$, where l_p is the best-fit persistence length. Across all chains, the PEG fits gave $\langle \ell \rangle = 1.3$ nm with a standard deviation of 0.2 nm, and L_0 ranging from 940 nm to 6300 nm; the HA data had ℓ ranging from 7.6 nm to 17.8 nm, and L_0 ranging from 800 nm to 5400 nm. The observed variation of the contour length is consistent with the polydisperse samples that were used.

Both PEG and HA form higher-order structures due to hydrogen bonding between neighboring monomers, and/or between monomers and water [138, 129, 125]. High forces have been shown to disrupt these interactions [139, 129, 125], which leads to deviations from WLC behavior, and can lead to elasticity-based estimates of l_p that are smaller than l_p in the absence of force. This issue is minimized here: those prior observations are based on AFM studies at a force scale of ≈ 100 pN . Here, our l_p estimates are based on fits at a much lower force scale (1 pN to 10 pN), where force-induced structural changes are a minor effect. Thus, we expect our WLC fit parameters to be a good estimate of low-force (< 1 pN) polymer structure.

We find that, in the experimental data, k_BT/R_g indeed controls F^*, with a smooth dependence across both HA and PEG data sets seen if R_g is estimated from a scaling relation as $R_g = \ell N^{0.588}$. Using this self-avoiding approximation leads to substantial overlap between the HA and PEG data, with F^* linearly increasing with k_BT/R_g (Fig. 3.3). Using the ideal relation, $R_g = \ell N^{1/2}$, results in little overlap between the PEG and HA

data (inset, Fig. 3.3).

To more quantitatively compare data and model, we focus on PEG, for which Devanand and Selser (DS) [130] provide a relation between chain molecular weight and gyration radius. We calculate molecular weight for each PEG chain from the best-fit L_0 value, as described above, and apply their formula to find $R_{g,DS}$; the resulting values are tightly correlated with the scaling estimate (see Fig. C.2), but have the advantage of using a known numerical prefactor. Applying this gyration radius estimate, we then calculate the value of the rescaled force and length at downturn across all 44 PEG curves, finding $\text{Mean}(F^* R_{g,DS}/k_B T) = 3.3 \pm 1.3$ and $\text{Mean}(\langle Z(F^*)\rangle / R_{g,DS}) = 4.0 \pm 1.4$ (the weighted mean is used for F^*, and errors are given as the standard deviation). The large variation in each value across the population occurs because our ability to resolve F^* from the PEG data is somewhat error prone (as seen in Fig. 3.3) due to low-force noise in γ. However, the greater-than-unity value of both parameters qualititatively confirms the model predictions, and quantitatively match well in the case of F^*, where the model predicted $F^* R_g/k_B T = 2.9$. The match is less good for rescaled length, where the model predicted $\langle Z(F^*)\rangle / R_g = 6.5$. This could be due to deficiencies in the model (notably its focus on ideal, rather than more realistic swollen chains), or due to aforementioned systematic experimental issues in estimation of small absolute extensions.

3.6 Discussion

Our results confirm and illuminate the initial hypothesis: At low stretching forces, polymer extensional fluctuations become cut off by exclusion from the tethering surface, decreasing the variance of the extensional distribution, and leading to the downturn in γ seen in Fig. 3.1D. Our image-based model of the effect of the surface captures this behavior (Fig. 3.2B), and analysis of the model predicts that the force, F^*, at which γ

decreases by 10% is given by $F^* \approx 3k_BT/R_g$. From a scaling viewpoint, experimental validation is demonstrated by the linear variation of the measured force at downturn, F^*, with independent experimental estimates of k_BT/R_g for two types of chains (HA and PEG), and over a wide range of contour and Kuhn lengths (Fig. 3.3). Quantitatively, our measurements of PEG, for which R_g can be reliably independently estimated, indicate $F^*R_g/k_BT \approx 3.4$, in relatively good agreement with the model predictions.

Certain prior studies [127, 40] predict that, in the absence of a surface, the compliance of a self-avoiding chain would increase as tension decreases through k_BT/R_g. The observed γ downturn, corresponding to a decrease in chain compliance, occurs at the same scale, obscuring the predicted compliance-increase effect. Our results are consistent with the work of Neumann [128], who discussed a plateau $\langle Z \rangle \approx R_g$ as $F \to 0$, as occurs in our calculation (Fig. 3.2A). But, we do not clearly observe that plateau in experiment (Fig. 3.1C), due to the small values of $\langle Z \rangle$, and potentially due to confounding systematic errors. In contrast, the measured transition in γ is more clear (Fig. 3.1D), indicating that, in the low-force limit, variance based metrics are more sensitive probes of elasticity than absolute extension.

The transition measured here joins a variety of other elastic transitions that occur when stretching single polymer chains. These include the transition from random-walk to straight chain behavior[127, 40], at $F \sim k_BT/\ell$; the transition from WLC to FJC behavior that occurs at $f \sim k_BT/b$, where b is the bond length [140]; and the emergence of excluded volume for rod-like chains, occurring below $f \sim k_BTv/\ell^4$, where v is the monomer-level excluded volume parameter [141, 142]. A major difference of the present transition is its sensitivity to the chain contour length: the other transitions only depend on intensive, microscopic polymer parameters such as Kuhn or bond length. Thus, given two chains of identical composition, but different lengths, all transitions except the γ downturn will occur at the same forces. Indeed, this is directly demonstrated here in our

analysis of experimental data (Fig. 3.3).

A remaining puzzle is to understand the behavior of γ in the zero-force limit. The model presented here relies on both linear-response theory and the pendulum approximation to calculate γ through the application of Eq. 3.1b, and predicts a plateau in γ in the zero-force limit (Fig. 3.2B). In preliminary work, we have instead explored directly applying Eq. 3.1a, which still is a linear-response result, but does not make the pendulum approximation. In that case, we still find a downturn in γ below k_BT/R_g. However, when using Eq. 3.1a, γ decreases continuously at low forces, rather than reaching a plateau. This disagreement points to an issue with the pendulum approximation in the presence of a surface and at low forces; more work is required to understand this aspect.

3.7 Conclusion

We have shown that fluctuation-based estimates of polymer elasticity are sensitive to surface effects for forces below k_BT/R_g. An elastic transition at that scale was indeed anticipated by prior work[127, 40]; however, this work clarifies that, in the presence of surfaces, the experimental signature of this transition is a *decrease*, rather than an increase, in chain compliance. The ability to observe this transition in a magnetic tweezer experiment means it is possible to independently estimate both R_g (from F^*), as well as N and ℓ (from high-force fitting) from a single measured force-extension curve. This broadens the capabilities of single-molecule manipulation instruments.

Chapter 4

Glassy Dynamics and Memory Effects in an IDR Construct

4.1 Preface

Given that IDRs are generally thought to behave like a random walk polymer in a good solvent [12], I did not expect the NFLt polyproteins to exhibit any significant conformational changes during my experiments. Hence, my original goal was to carry out force-extension measurements on the NFLt polyproteins and characterize their low-force entropic elastic response. Collecting a force-extension curve consists of repeatedly moving the magnets away (or towards) the flow cell, pausing at a magnet position for some time, and collecting a bead trajectory from which both the extension and force is determined. Typically, at a constant magnet position, the bead-tether system is assumed to be in a stationary state - i.e., the extension is independent of time; however, as I carefully analyzed my data, I noticed that the extension was changing over time. First, I performed several control experiments on other polymers to make sure the extension changes were not due to instrumental drift. Next, I tried several different passivation strategies and

carefully analyzed the positions of the bead to ensure the polyproteins were not adhering to the the bead or the flow cell. Having convinced myself that the extension changes were not experimental artifacts, I designed a force jump protocol with better programmatic control of the magnet position to characterize these extension changes.

The conformational dynamics of small single domain proteins typically exhibit an exponential time dependence, so I tried fitting the extension changes to an exponential function, but the fits were not robust. Eventually, I tried plotting the data on semilogarithmic and log-log axis, which revealed that the extension changes followed a remarkably slow logarithmic time dependence. Looking over my data, it was as if I knocked a wrench off a table, went off to grab a coffee, and came back to find that it had yet to hit the ground.

The content of this chapter previously appeared in Physical Review Letters [143]. It is reproduced here with the permission of the American Physical Society.

4.2 Introduction

The conformational changes of globular, folded proteins can exhibit glass-like kinetics, typically measured as nonexponential relaxations [7, 144, 145, 146, 147]. This behavior is associated with the roughness of the conformational energy landscape, i.e., the presence of multiple local free energy minima that are separated by appreciable activation barriers [148]. Based on studies of random-sequence biopolymers, the heights of the barriers are usually related to either the difficulty in rearranging connected residues within the dense protein core ('topological frustration'), or to kinetic trapping by nonnative contacts ('energetic frustration'), as enabled by the nonspecific nature of the dominant hydrophobic interactions [4].

Unlike globular proteins, intrinsically disordered protein regions (IDRs) exhibit a high

degree of conformational freedom in their native state [11]. IDRs generally are enriched in hydrophilic and charged residues [11], permitting them to assume a dynamic ensemble of structures analogous to those of a random-walk polymer in good solvent, though with the typical addition of some secondary-structure formation [28] or other forms of weak attraction between residues [36, 56, 54, 149]. These structures and interactions are thought to give IDRs a rough, flat, energy landscape [51, 14]. However, because IDRs have fewer hydrophobic residues than globular proteins, and no dense core, the barriers on this landscape are small; thus, IDRs would not be expected to exhibit significant frustration, nor, in turn, glass-like kinetics.

Yet, here we show that single molecules of a model disordered protein construct, consisting of multiple repeats of the intrinsically disordered neurofilament-low (NFL) protein tail region [150], exhibit glass-like behavior in the form of slow, logarithmic relaxations in response to a one-step change in applied tension. Further, when subject to a two-step force-change protocol, the construct displays a nonmonotonic change in the polymers' extension, a glassy memory effect termed the 'Kovacs hump' [151, 152]. Based on work on other glassy systems [60, 153], we attribute these behaviors to the existence of multiple, independent local structure-forming processes with widely-varying dynamics. We corroborate this picture by showing that the force-dependence of the logarithmic relaxation is well-described by a model that couples an ensemble of parallel structuring processes to Bell-Zhurkov mechanochemistry [46, 45]. Overall, this work demonstrates glassy behavior in this IDR construct is due to a heterogeneous, distributed mechanism different from the frustration-based ones of certain globular proteins.

4.3 Methods

IDR purification and polymer synthesis are described in detail in the Appendix B. In short, single NFL IDRs, each containing 168 amino acids, were modified to carry cysteine residues at each terminus, recombinantly expressed, and purified. These IDRs

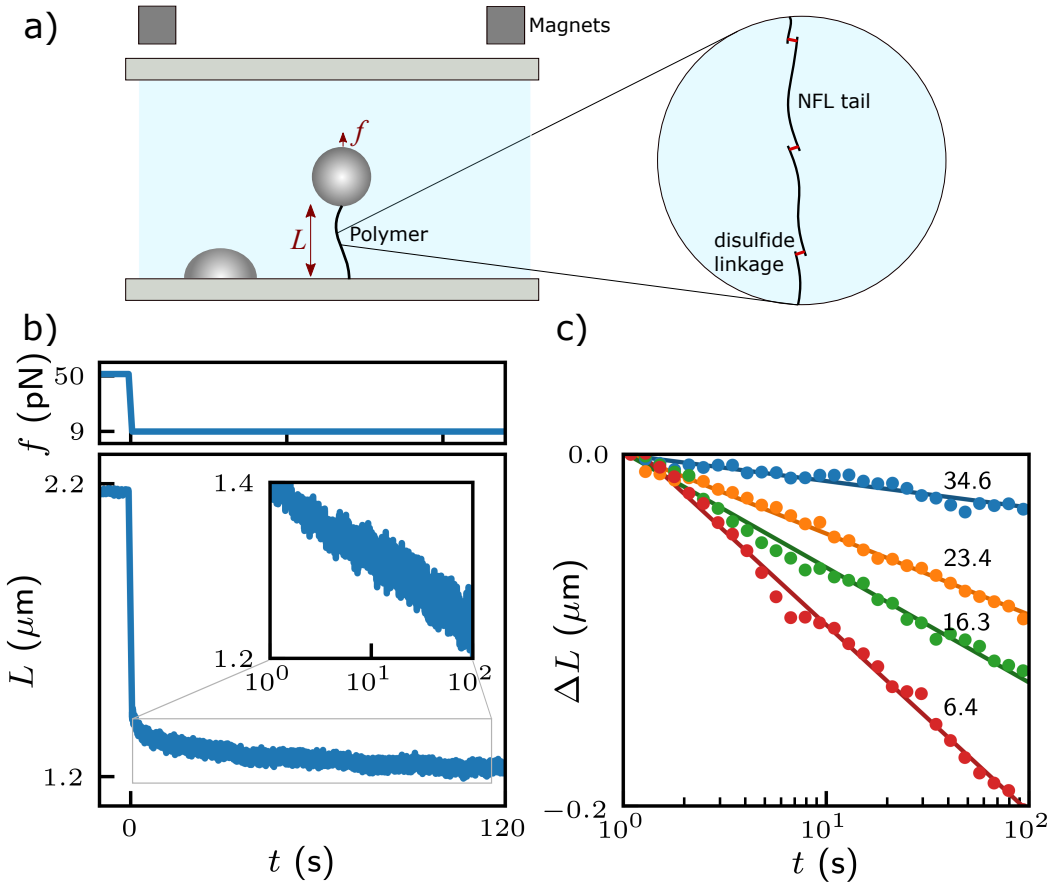

Figure 4.1: (a) Experimental setup: A polymer consisting of multiple NFL IDRs joined by disulfide bonds is stretched with a force f while its extension L is tracked. Stuck beads are tracked to remove drift. (b) Example force-quench experiment on a single polymer: at $t = 0$, the force was decreased from $f_1 = 50$ to $f_2 = 9$ pN, resulting in a rapid elastic response followed by a slow logarithmic relaxation (inset). (c) Typical force dependence of the logarithmic relaxation after quench from $f_1 = 50$ pN, plotted as the compaction, $\Delta L \equiv L(t) - L(t_0)$, after a reference time $t_0 = 1$ s. At higher f_2 (labeled, in pN), the relaxation slows due to hindering of chain shortening by tension. Data points and error bars are the mean and standard error of the mean after logarithmic binning in time (error bars are smaller than points); lines are best-fits to $b \log(t/t_0)$.

were polymerized together to form a linear polymer by inducing disulfide bonds between the cysteines. The polymers were terminally labeled with azide and biotin, respectively, allowing specific attachment between a functionalized glass surface and a 2.8-μm-diameter magnetic bead; this enabled stretching experiments [Fig. 4.1(a)].

Experiments were carried out with a custom-built magnetic tweezer setup [98, 154], at $T = 20$ °C, in a pH 7 buffer containing 20 mM 2-(N-morpholino)ethanesulfonic acid (MES), 10 mM NaCl, and 0.05% Tween-20. The stretching force was set by adjusting the distance between a pair of movable magnets and the flow cell surface [Fig. 4.1(a)]. The polymers' end-to-end extension was measured by analyzing the image of the bead [98, 155], as captured by a CMOS camera operating at 400 Hz. Instrumental drift was eliminated by simultaneously tracking reference beads stuck to the glass surface, and subtracting their height from that of the experimental beads [Fig. 4.1(a)]; the success of this procedure was demonstrated through control measurements of DNA tethers (Fig. C.4). The stretching force was estimated by analyzing lateral bead fluctuations [61], with a typical uncertainty of $\lesssim 5\%$.

For each polymer, the number of monomer tails, N, was estimated from the polymer (L_p) and monomer ($L_m \approx 64$ nm) contour length, $N = L_p/L_m \approx 2$–29, by assuming a contour length per amino acid of 0.38 nm and 168 amino acids per monomer. The polymers' contour lengths were estimated from their high force ($\gtrsim 50$ pN) extension. After accounting for N, the polymers' polydispersity did not affect our measurements.

During force-quench experiments, the force was changed from f_1 to $f_2 < f_1$ by moving the magnets away from the flow cell surface. The motion of the magnets lasted ≈ 0.25 s; we only analyzed extension changes that occur after that, particularly setting the zero of time, $t = 0$, as the point at which magnet motion stops. During the motion, the polymer extension changed rapidly due to its entropic elasticity [Fig. 4.1(b)]. The time scale of elastic relaxation is expected to be ≈ 10 ms, as judged by estimating either the Rouse

time [47] of the polymer or the relaxation time associated with the drag of the bead; thus, elastic relaxation is unrelated to the observed long timescale extension changes. The relaxation data and analysis code have been made available in a public repository.

4.4 Logarithmic relaxations

Following a force quench, the polymer extension, L, decreased logarithmically in time [Fig. 4.1(b), inset]. As shown in Fig. C.5, logarithmic relaxations have been observed to last for up to 3 decades in time, although our analysis focused on 2-decade relaxations [Fig. 4.1(c)]. During relaxation, the extension change was smooth without any detectable discrete transitions of 10 nm or larger, suggesting that the underlying individual compaction events each contribute a length change of order 1 nm.

We studied the force-dependence of the relaxation by keeping f_1 constant, typically 50–70 pN, and quenching to different values of f_2. For all polymers, the relaxation was logarithmic for all f_2, with the rate becoming faster for smaller f_2 [Fig. 4.1(c)], as expected for a structure-forming process that is hindered by an opposing force [46, 45]. All relaxations were well-fit by the relation $L(t) = b \log(t/t_0)$, with t_0 being an arbitrary reference point chosen throughout to be $t_0 = 1$ s, and b corresponding to the log-slope of the relaxation.

4.5 Heterogeneity

Recent work has suggested there are multiple classes of logarithmically-relaxing systems [60, 153]. Among these, Amir *et al.* [60] identified a mechanism in which the system is highly heterogeneous, consisting of multiple modes that relax independently and with a broad spectrum of timescales. Lahini *et al.* [153] showed that this mode structure is

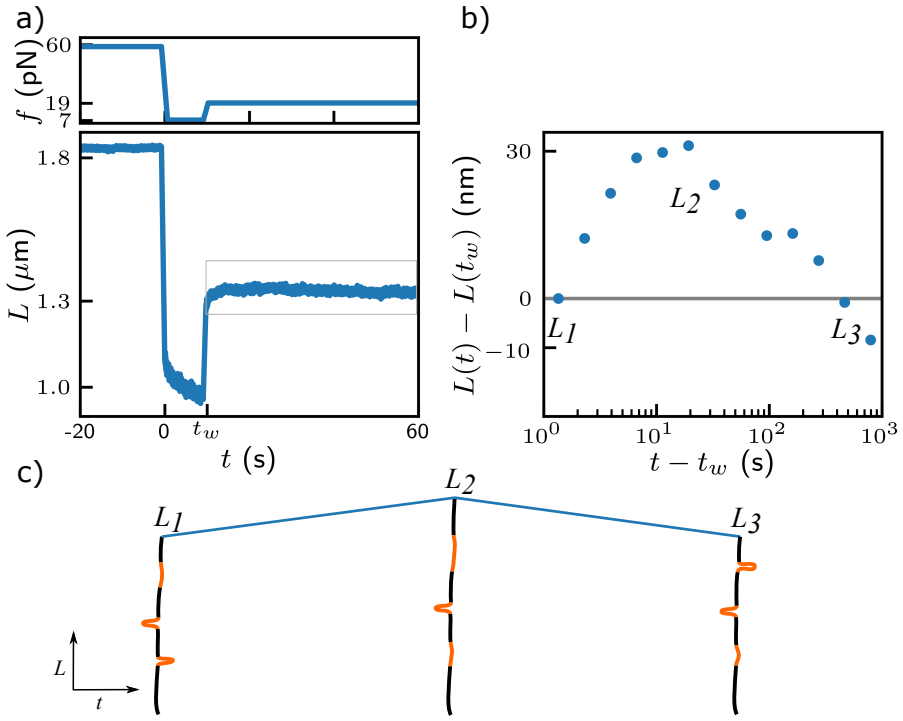

Figure 4.2: (a) Typical two-step experiment: The force was initially $f_1 = 60$ pN, then held at $f_2 = 7$ pN for $t_w = 10$ s, then increased to $f_3 = 19$ pN. (b) Detail of extension dynamics at f_3 from the data in (a), showing a nonmonotonic change in L, i.e. a Kovacs hump [152]. Data points and error bars are the mean and standard error of the mean after logarithmic binning in time (error bars are smaller than points). (c) Cartoon of heterogeneous dynamics within a single IDR domain that result in the Kovacs hump: Incubation at f_2 for t_w (left) allows folding of fast segments, but is not long enough to allow folding of slow segments. Application of the higher force f_3 causes unfolding of some fast segments, leading to the increase, $L_2 > L_1$. At long times, the slow segments finally fold, causing the slow decrease, $L_3 < L_2$.

associated with an experimentally-observable memory effect, the Kovacs hump [151, 152], which here corresponds to the prediction of a nonmonotonic change in extension with time (i.e. an increase followed by a decrease in L) after a particular two-step pattern of changes in applied force. Such a behavior unambiguously demonstrates heterogeneity: as the trajectory progresses, certain values of L are reached twice, but are followed by different system behaviors (i.e. lengthening or contraction). Thus, knowing the single parameter L is not sufficient to predict future behavior. Instead, predicting the correct

behavior requires the knowledge of more parameters, i.e. the status of the diverse modes, which store the system's memory of past force application [153, 152, 156].

To assess whether the present IDR system belongs to this 'Kovacs class', we tested for the Kovacs hump by subjecting the chain to three successive forces, f_1, f_2, f_3, such that f_2 is held for a time t_w, and the final force lies between the prior two, $f_1 > f_3 > f_2$ [Fig. 4.2(a)]. We found that single NFL IDR polymers, and not control DNA molecules Fig. C.7, consistently showed a clear Kovacs hump at f_3 ([Fig. 4.2(b)]; observations on 8 other polymers are shown in the Appendix C). The hump consisted of a slow increase in L followed by a slow decrease, and thus was not related to the fast elastic response of the polymer. We conclude that the polyIDR can be assigned to the heterogeneous, Kovacs-class of aging systems.

4.6 Microscopic view

The results of Fig. 4.2 indicate that the IDR polymers contain multiple independent relaxation modes, but does not clarify their microscopic identity. We posit that each mode corresponds to a different segment of the chain, with each segment able to independently transition from an extended coil to a compact structure. As discussed below, the different segments do not correspond to individual IDRs within the polyIDR chain, but rather to different clusters of residues within each IDR. Indeed, previous work has observed short-range structure, such as salt bridges and residual secondary-structure elements, in the NFL tail domain [36, 56, 54].

Such local structure implies the Kovacs hump occurs in the following manner (see also Fig. 4.2(c)): Incubation at the high force f_1 converts all segments to the extended coil state. After quenching to f_2 and holding for a time t_w, a fraction of the segments become structured (i.e. those with relatively fast dynamics), while the slower segments remain

unstructured. Jumping to the final force f_3 causes a transition back to the extended state for some fast segments (leading to the initial increase). After a long time, the slow segments become structured (leading to the long-term decrease).

Using the established effect of force on transition kinetics, we can develop a model that quantifies this microscopic picture, and tests it through comparison to data. We focus on the single-step force quench, and assume each IDR consists of n independent segments (and thus that the entire polymer contains Nn such segments). We then adopt the mathematical framework of Amir *et al.* [60], and take each segment to relax, on average, exponentially after the force quench, so the contraction dynamics of the j^{th} segment follows $L_j(t) = \alpha(f)e^{-t/\tau_j(f)}$, where $j = 1, 2, \ldots, n$. Both τ_j, the contraction timescale, and α, the relaxation amplitude, carry a force-dependence. Each structuring event within a single IDR is discrete and stochastic; however, the presence of multiple IDRs in the polymer ($N \gg 1$) means that the measured extension change will follow the exponential ensemble-average behavior.

The force-dependence of α accounts for segment elasticity, which is likely dominated by the flexible coil state. Thus, we take $\alpha(f) = \ell\alpha_0(f)$, where ℓ is the coil contour length, and $\alpha_0(f)$ is the relative extension, given by the worm-like chain model [136] with $l_p = 0.8$ nm, as appropriate for polypeptides [157]. Using more nuanced elastic models does not greatly affect our results, as described in the appendix C. While different segments likely have different ℓ, we expect that variation to be small ($< 10\times$) compared to the range of values of τ_j that must underlie the multi-decade dynamics, and thus take ℓ to be the same for all segments.

This mode structure predicts logarithmic relaxation if the characteristic relaxation times are distributed as $P(\tau) \propto 1/\tau$ [60]. We implement this here by noting it implies a uniform distribution of log-relaxation-time, $P(\log(\tau)) \equiv \eta$, where η is the density of mode states in log-time units. This means that the log of the time between successive

relaxation events is, on average, $1/\eta$. Thus, during a time interval t, the number of relaxation events that occurs in a single IDR is $\eta \log t$. Since each event contributes the same contraction, this leads to a logarithmic relaxation.

To compare to force-quench data, we enforce the condition that there is no relaxation if $f_2 = f_1$, reference the relaxation to the time t_0, and scale by N to get the extension change of the entire chain:

$$L(t, f_2) - L(t_0, f_2) = -N\alpha(f_2)\left[\eta(f_2) - \eta(f_1)\right]\log\left(t/t_0\right) \tag{4.1}$$

As noted, η must carry a force dependence, since the transition times τ vary with f.

The dependence on force follows from enforcing an Arrhenius dependence of τ on activation barrier, $\tau = \tau_0 \exp(\Delta G/k_B T)$, and using the Bell-Zhurkov expectation that ΔG varies as $f\Delta x$, for activation distance Δx [46, 45]. The constraint, $P(\tau) \propto 1/\tau$, corresponds to a uniform distribution of ΔG [60]; thus we take

$$\Delta G_j = G_0 + j f_2 \delta x \tag{4.2}$$

where G_0 is the barrier at zero force, and $j\delta x$ is the distance between the initial (extended) state and the j^{th} activation barrier. Eq. 4.2 implies that $P(\Delta G) = 1/f\delta x$, and thus that $\eta = k_B T/f\delta x$, which when combined with Eq. 4.1 results in a prediction for the log-slope b.

This analysis indicates useful normalized parameters for the slope and force-quench magnitude are $\bar{b} \equiv b f_1/N k_B T \alpha_0(f_2)$ and $\bar{f} \equiv f_2/f_1$. Indeed, plotting $\bar{b}$ vs. $\bar{f}$ collapses the data [Fig. 4.3], including removing the effects of polydispersity in length, as shown

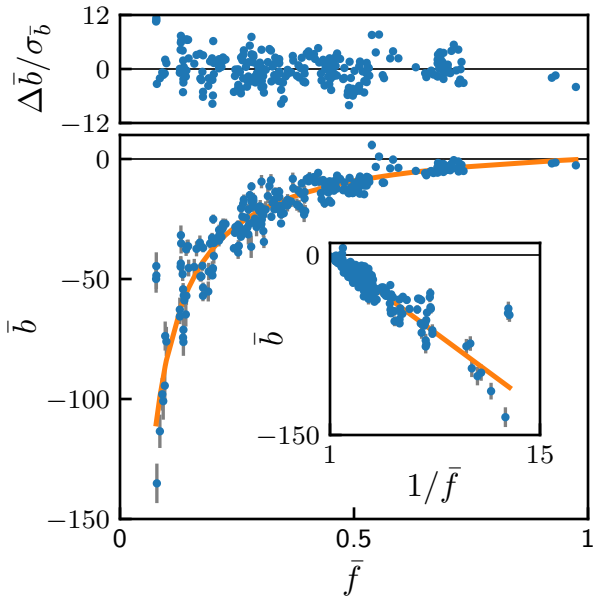

Figure 4.3: Bottom: Dependence of the normalized logarithmic relaxation rate, $\bar{b}$, on the force-quench magnitude, $\bar{f} \equiv f_2/f_1$. Each data point represents a single force-quench on a single polymer, with error estimated from the uncertainty in the measured slope. 248 data points, from 16 separate polymers, are shown. The line is a fit to Eq. 4.3 with best-fit parameter $\rho = 0.108 \pm 0.004$ (error estimated from bootstrapping, as described in the appendix C). Top: standardized residuals of the fit, $\Delta\bar{b}/\sigma_{\bar{b}}$, where $\Delta\bar{b}$ is the difference between the data and the fit, and $\sigma_{\bar{b}}$ is the error estimate of the data. Inset: The data is linearized by plotting $\bar{b}$ vs. $1/\bar{f}$.

in the appendix C. The model specifically predicts:

$$\bar{b} = \frac{1}{\rho}\left[1 - \frac{1}{\bar{f}}\right] \tag{4.3}$$

The single unknown parameter ρ represents the spacing between activation barriers relative to the coil contour length, $\rho \equiv \delta x/\ell$. For our model to be self-consistent, we expect $n\delta x < \ell$ and $\rho < 1/n < 1$.

4.7 Results and discussion

The microscopic model, Eq. 4.3, successfully describes the force-quench data, corroborating the picture of multiple independent structured segments that each follow Bell-Zhurkov mechanochemistry. We show this by fitting Eq. 4.3, using the single fitting parameter, ρ, to the results of 248 different force-quench experiments [Fig. 4.3]. The best-fit is found with $\rho = 0.108 \pm 0.004$. The fit gives a reasonable reduced-chi-squared fitting metric, $\bar{\chi}^2 = 10.7$. Note that $\bar{\chi}^2 \approx 1$ is expected for a statistically valid fit given the stochastic errors of the data; the elevated value here is due either to unknown systematic errors, or to physical effects ignored in our approximate model. The standardized residuals show no systematic deviation [Fig. 4.3, top], strongly suggesting our model captures the key features of the system.

The best fit estimate of $\rho < 1$ is consistent with the physical restriction on the activation barrier spacing, $\delta x < \ell$. The fit value of ρ implies that the upper limit on the number of structured segments per IDR is $n_{max} = 1/\rho \approx 9$. This value is also consistent with the data: A typical polymer has $N = 25$ IDRs; taking $n = 9$ structures per IDR, and with each structure contributing a compaction $\alpha \approx 1$ nm, we can estimate the total length change during relaxation, $\Delta L \approx 200$ nm. This is indeed the magnitude of the total length change seen in the force-quench experiments [Fig. 4.1(c)], further supporting our picture.

The polymeric nature of the construct allows for the possibility of inter-tail interactions [158, 159], however the measured relaxation is likely dominated by multiple intra-tail structures. Given that there are no observable discrete length changes [Fig. 4.1(b)] and the total length change is a small fraction ($< 10\%$) of the polymer's contour length, the structure-forming interactions must be quite short-range. With only $N \approx 25$ abutting neighbors, nearest-neighbor inter-tail interactions, contributing ≈ 1 nm length changes

each, could not account for the observed 200 nm length changes.

While the precise identity of the structures is as yet unclear, we can roughly estimate their free energy based on their ability to compact against a known load. Extrapolated to zero force, we find that single-segment structure stability is likely between 3 and $9k_BT$; the wide range is due to sensitivity to the choice of ℓ (see Appendix C.2.3). This range is reasonable and suggests some possible mechanisms; it encompasses prior estimates of the stability of local structures in IDRs [51], as well as estimates of attractive electrostatic interactions in the NFL tail [52, 36, 56, 54].

4.8 Conclusion

In summary, our analysis indicates that the NFL IDR has multiple independent structures with a broad distribution of relaxation times. The distribution of relaxation timescales produced a logarithmic relaxation of polymer extension that can last for three decades in time (Fig. C.5). We attribute the long timescales to the slowing of individual compactions by applied force, in analogy to the slowing caused by low temperatures in observations of nonexponential relaxations in globular proteins [7]. The heterogeneous, independent nature of the structures was confirmed by the observation of the Kovacs hump. Finally, our picture of IDR compaction dynamics is confirmed by a model that combines Bell-Zhurkov mechanochemistry with a specific distribution of independent segment relaxation times; this model successfully describes the dependence of relaxation rate with force, and produces consistent estimates of microscopic parameters.

The nonexponential relaxations reported here are similar to those observed in prior work on globular proteins [7, 146, 147] as well as those expected for random-sequence biopolymers [4], but occur for a different reason: Heterogeneity and disorder in the IDR occur due to the varying dynamics of structure formation of independent segments, and

not because of the topological/energetic frustration effects that dominate the dynamics of certain globular proteins[7, 146, 147] and random-sequence chains [4].

However, there is at least one reported mechanism for nonexponential dynamics in a globular protein that is a more apt comparison: lysozyme was shown to exhibit heterogeneous nucleation dynamics [144], where different small segments independently form local tertiary structures with varying dynamics; this is analogous to the 'foldons' proposed theoretically [160, 161]. This behavior is similar to that proposed here, with the major difference being the type of structure formed, as tertiary interactions are lacking in the IDR.

Finally, it has been suggested that many IDRs contain multiple subsegments that form local structures or interactions in an independent, noncooperative fashion [51, 14]. Our analysis indicates that the existence of such distributed, heterogeneous structures is the key feature underlying the slow, logarithmic dynamics. Thus, it is reasonable to expect that glassy relaxations and memory effects could occur broadly in other systems with IDRs.

Chapter 5

Origin of Glassy Dynamics in an IDR construct

5.1 Preface

Using a continuum model based on a general framework for bulk glassy systems [60], I was able to predict the force dependence of the glassy dynamics; however, its molecular and biochemical origins remained unclear. Furthermore, in bulk glassy systems, the spectrum of relaxations contains many closely-spaced timescales, which can then be combined and simplified by applying methods of calculus, particularly through transforming the sum of discrete, exponential relaxations into a smooth, continuous logarithmic relaxation [162]. The conceptual underpinnings of such an approach for the nanoscale NFLt polyproteins was also unclear. Hence, I sought to determine the origins of the glassy dynamics and understand it using physically realistic parameters.

5.2 Introduction

Intrinsically disordered protein regions (IDRs) are found throughout the cytoskeleton, where they play an important role in cellular mechanical processes [35, 36]. An important first step to understanding these processes, is characterizing the mechanical response of single IDRs. In contrast to globular proteins, IDRs have a greater degree of conformational freedom due to a large fraction of charged and hydrophilic residues [19, 12]. Instead, their structure is more similar to a random-walk polymer in a good solvent with the addition of both transient and stable secondary structures, electrostatic attractions, and amorphous globules [21].

The conformational changes of proteins often exhibit glass-like kinetics in their native state, typically measured as a nonexponential time dependence [5, 7]. These kinetics are associated with the roughness of the native energy landscape, i.e., multiple local free energy minima corresponding to different substates [6]. However, under force, a protein's conformational changes is often well-described by a simple two-state system with a native and unfolded state, as well as the occasional addition of transient intermediate states [44]. This system can be represented by a one-dimensional free energy landscape with two wells separated by an activation barrier along the reaction coordinate, i.e., the end-to-end extension. At constant force, transitions between states are stochastic, but their ensemble average typically follows an exponential time dependence [49].

Yet, recently, we reported force jump experiments that show a highly nonexponential time dependence in the extension of an IDR polyprotein at constant force. This polyprotein consists of multiple repeats of the disordered neurofilament light tail (NFLt) region [143], an IDR that is critical to the structure and mechanics of the axonal cytoskeleton in neurons [163]. After a one-step force jump, the extension of single NFLt polyproteins exhibited a slow, logarithmic decrease in their extension. After a two-step force jump,

they exhibited a nonmonotonic time dependence, reminiscent of the "Kovacs hump" in glassy systems [151, 153]. To explain this phenomenon, we developed a continuum model based on work on bulk glassy systems (e.g., crumpled paper balls and semiconductor materials [60]) and Bell-Zhurkov mechanics that predicted the force dependence of the logarithmic relaxations. However, the conceptual underpinnings for this model, as well as the physical and biochemical origins of the relaxations remained unclear.

Here, we revisit the NFLt polyprotein system with the goal of understanding the origins of the relaxation behavior. We present data showing the relaxations are history dependent, indicative of a system with multiple, independent, and heterogeneous relaxation events. We also show that both force increases and force decreases lead to logarithmic relaxations, consistent with structure-forming and breaking events. Based on this evidence, we develop a discrete model that is physically realistic and, without fitting, qualitatively predicts the force dependence of the logarithmic relaxations. Overall this work demonstrates that the relaxation behavior is likely caused by hidden lengths within each NFLt, which, to our knowledge, is a novel mechanism for glassy dynamics.

5.3 Materials and methods

Polymer synthesis and tweezer experiments

NFLt purification and polyprotein synthesis was carried out as previously described [143]. In short, in short single NFLts, each containing 168 amino acids, were modified with cysteine residues at each terminus. The cysteines were induced to form disulfide bonds resulting in a linear polyprotein.

The polymerization was stochastic, producing polydisperse polyproteins with a variable number of monomer tails. The number of monomer tails, N, in each polyprotein

was calculated from the ratio of its contour length, L_p, and monomer NFLt contour length, L_m: $N = L_p/L_m \approx$2–29. The polyprotein contour length was estimated from its extension at high force. The monomer contour length, $L_m \approx 64\,\text{nm}$, was estimated from the number of residues in a single NFLt assuming a residue contour length of 0.38 nm.

After polymerization, the polyproteins were terminally labeled with azide and biotin, then reacted with a DBCO-functionalized glass surface and a 2.8 µm-diameter streptavidin-coated paramagnetic bead in a flow cell, resulting in tethered polyproteins. The flow cell contained an inlet and outlet to enable solution exchange between experiments. Unless otherwise stated, the solution in the flow cell was a 20 mM 2-(N-morpholino)ethanesulfonic acid (MES) buffer, with 10 mM NaCl and 0.05% Tween-20 at pH 7 and room temperature ($\approx 20\,^\circ\text{C}$).

The polyprotein extension, L, over time, t, was measured as described in Chapter A. The applied force was calculated using the Allan variance as described in Chapter 2.

The one-step force jump protocol was carried out as illustrated in Fig. 5.1B. First, the polyprotein was kept at a force, f_1, for a long time ($> 100\,\text{s}$). Next, the force was jumped from f_1 to a second force, f_2, by moving the magnets closer or further from the flow cell, increasing or decreasing the force, respectively. The magnet motion typically lasted $\approx 0.25\,\text{s}$, during which the polyprotein extension changed rapidly due to entropic elasticity. Once the magnet motion stopped, t was set to zero and the force was constant. At constant force, the polyprotein exhibited a slow extension change that cannot be accounted for by the elastic relaxation timescale $\approx 10\,\text{ms}$, as judged by the Rouse time [47]. The relaxations are reproducible; repeating the same one-step force jump on the same polyprotein results in the same slow extension change. During relaxation, the extension changes were smooth without any detectable discrete transitions of 10 nm or more, consistent with underlying compaction or expansion events each contributing a length change of $\sim 1\,\text{nm}$.

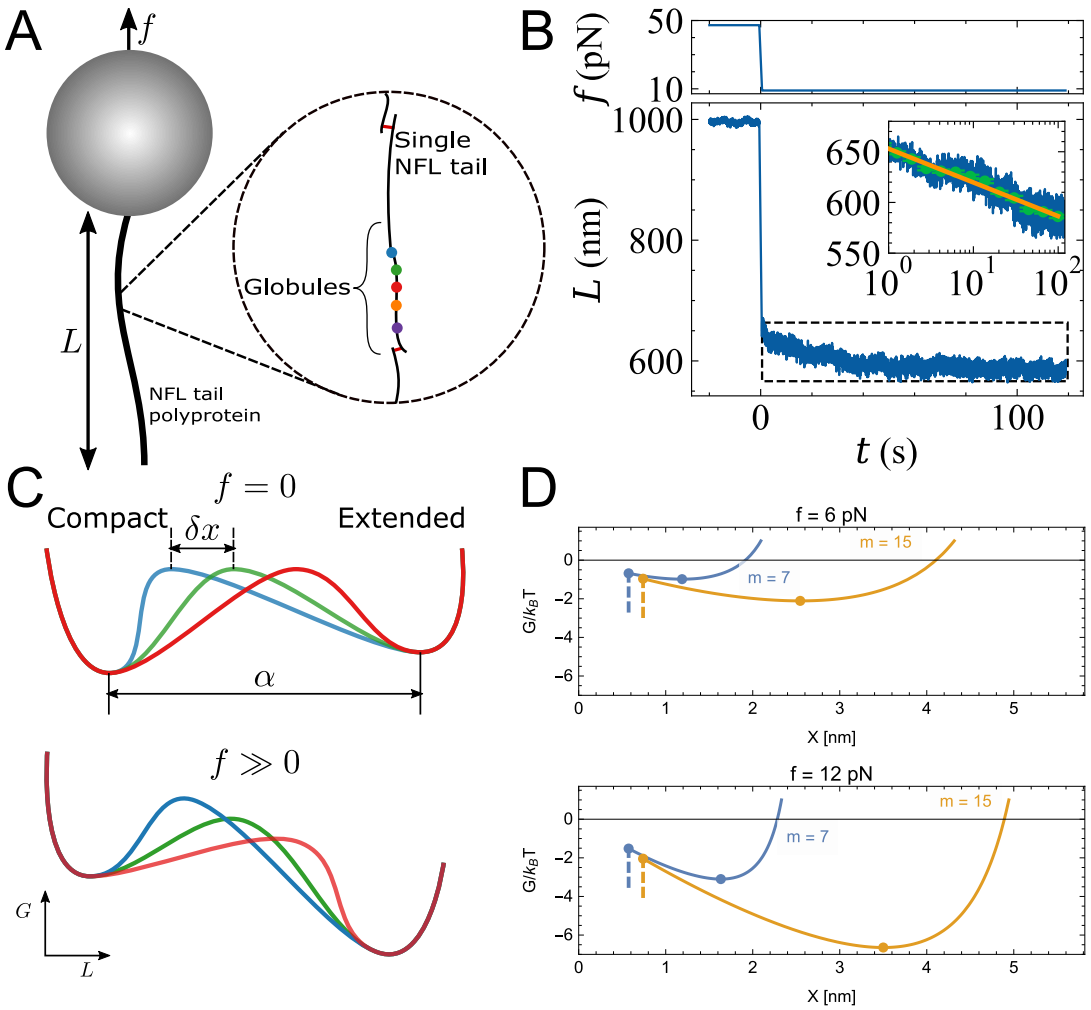

Figure 5.1: Schematic of experimental setup. (A) A polyprotein consisting of disulfide-linked NFLts is subject to an applied force, f while its extension, L is tracked. Single NFL tails contain multiple, independent globules. Typical one-step force jump data from a single polyprotein. (B, upper) The force, f, was jumped from f_1 to f_2. (B, lower) During the jump, there is a large change in extension due to the entropic elastic response. After the jump, there is a slow change in extension (dotted black box) due to compaction within each tail. (B, inset) The slow change in extension shows a logarithmic time-dependence when plotted on a logarithmic time axis. (C) In the continuum model, each globule corresponds to an independent free energy, G, landscape, which tilts with the application of force, favoring the extended state. There average spacing between activation energy barriers is δx. (D) In the discrete model, the free energy well is given by the WLC elastic model. Increasing the force deepens the well and increases length change.

As previously described in Ref. [143], the one-step force jump relaxation data appear linear when plotted on a semilogarithmic time axis, consistent with a logarithmic time dependence (Fig. 5.1B, inset). The relaxation data are binned logarithmically in time to average out noise from the bead fluctuations. After binning, the data are well-fit by the empirical logarithmic function, $L(t) = a + b\log(t/1\,\mathrm{s})$, where b is the log slope and a is the extension at $t = 1\,\mathrm{s}$.

Continuum model

The relaxation data can be interpreted in terms of a continuum model [143]. This model is based on a generic mathematical framework for bulk glassy systems by Amir *et al.* [60] and Bell-Zhurkov mechanics [46, 45].

In this model, a polyprotein consists of N tails, each of which contain a total of n independently relaxing subsegments (Fig. 5.1A). The differential length of each j^{th} ($j \in 1, 2, \ldots, n-1, n$) subsegment varies exponentially $\alpha_j(f)e^{-t/\tau_j(f)}$, over a time interval, t. As noted, both the total length change, α_j, and characteristic timescale, τ_j, carry a force dependence.

The force dependence on α_j accounts for the subsegments elasticity, which is likely dominated by the extended coil state, such that $\alpha(f) = \ell\alpha_0(f)$. The coil contour length, ℓ, and the relative extension, α_0 are estimated from the worm-like chain model [136, 137] using a persistence length, l_p, of 0.8 nm, as appropriate for unfolded polypeptides [157]. While different subsegments likely have different ℓ, the variation is likely to be small compared to range of τ_j. Hence, for simplicity, ℓ is taken to be the same for all subsegments (Fig. 5.1C).

The continuum model predicts a logarithmic relaxation if the relaxation times are distributed as $P(\tau) \propto 1/\tau$ [60], corresponding to a uniform distribution of log-relaxation-

time, $P[\log(\tau)] = n/\log(\tau_{\max}(f)/\tau_{\min}(f)) \equiv \eta(f)$, where $\tau_{\min}$ and $\tau_{\max}$ are the shortest and longest timescales in the system. The log of the time between successive relaxations is, on average, $1/\eta$. We assume the system starts in equilibrium at an initial force, f_1. After a force jump to f_2, the differential polyprotein extension, $\Delta L(t, f_2) \equiv L(t, f_2) - L(1\,\mathrm{s}, f_2)$, is

$$\Delta L(t, f_2) = N\alpha(f_2)(\eta(f_1) - \eta(f_2)) \log(t/1\,\mathrm{s}) \quad (5.1)$$

$$= b \log(t/1\,\mathrm{s}), \quad (5.2)$$

assuming $t_{\max} > t > t_{\min}$. Note that Eq. 5.2 predicts both force increases and force decreases should produce a logarithmic time dependence.

History dependence

The continuum model can be extended to multiple force jumps by accounting for the non-equilibrium state of each subsegment. After the first force jump, the system is allowed to relax at f_2 for a time interval, t_w, before jumping to a third force, f_3. At a time interval, t, later, the average extension of the j^{th} globule is $\alpha(f_3)(1 - e^{t_w/\tau_j(f_3)})e^{-t/\tau_j(f_3)}$ [60]. It then follows that the differential polyprotein extension is

$$\Delta L(t, f_3) = N\alpha(f_3) \left\{ [\eta(f_2) - \eta(f_3)] \log\left(\frac{t}{t_{\min}}\right) + [\eta(f_1) - \eta(f_2)] \log\left(\frac{t + t_w}{t_{\min}}\right) \right\}. \quad (5.3)$$

Eq. 5.3 predicts the relaxations are history-dependent, i.e., they depend on t_w. If $f_3 = f_1$, it predicts a monotonic differential extension that follows a logarithm for $t < t_w$ and a power law for $t > t_w$ [60]. If $f_1 > f_3 > f_2$, it predicts a nonmonotonic relaxation where the time to reach the peak increases with t_w [153].

Force dependence

The force dependence of the relaxations can be understood by enforcing Arrhenius kinetics, i.e., τ_j is associated with an activation barrier, G_j, such that $\tau_j = \tau_0 e^{G_j/k_BT}$, where τ_0 corresponds to the transition attempt frequency [43] (Fig. 5.1C). From the Bell-Zhurkov expectation, the activation barriers vary linearly with the force, f, and activation distance, Δx_j, i.e., $G_j(f) = G_0 + f\Delta x_j$, where G_0 is the activation barrier at zero force. The constraint, $P(\tau) \propto 1/\tau$ [60], corresponds to a uniform distribution of activation barriers. It follows that the activation distance is $\Delta x_j = j\delta x$, where δx is the average spacing between activation distances and $\eta(f) \sim k_BT/f\delta x$.

It is convenient to use normalized parameters, $\bar{b} = bf_1/Nk_BT\alpha_0$ and $\bar{f} = f_2/f_1$, to compare relaxation data across from different force jump combinations and polyprotein lengths. Combining Eq. 5.2 with the Bell-Zhurkov expectation, it follows that

$$\bar{b} = \frac{1}{\rho}\left(1 - \frac{1}{\bar{f}}\right). \tag{5.4}$$

Eq. 5.4 contains a single unknown parameter ρ that represents the average spacing between activation barriers and the coil contour length, $\rho \equiv \delta x/\ell$. For this model to be self-consistent, we would expect $n\delta x < \ell$ and $\rho < 1/n < 1$.

Discrete model

As with the continuum model, the discrete model is built from a set of n independently relaxing subsegments in the NFLt. Each subsegment is taken to collapse with an exponential time dependence, i.e. the extension of the i^{th} segment will vary as $\alpha_i \exp(-t/\tau_i)$, the sum of which gives the total extension change, as in Eq. 5.2. Thus, the relaxation of each subsegment depends on two parameters, its total length change, α_i, and timescale

τ_i. However, in contrast to the continuum model, these parameters are estimated from the physical properties of each subsegment.

To estimate $\alpha_i j$, we assume the following: immediately after a quench from a high force, each subsegment is initially in a fully extended state, acting as a stretched, unstructured polypeptide with a certain mean extension, $\langle X_i \rangle$. This mean extension can be calculated from the subsegment contour length ℓ_i and the applied force f within a WLC model, given a persistence length, which we take to be $l_p = 0.6$ nm. Thermal fluctuations will cause X_i to vary, and we assume that, once X_i shortens to a certain critical value r_i, the chain will collapse. We take r_i to scale with the number m of residues as $R_g = 0.3m^{1/3}$ nm, as appropriate for the size of a compact protein [164]. Consequently, we take the total length change of each subsegment to be

$$\alpha_i = \langle X_i \rangle - r_i = \langle X_i(f, \ell_i, l_p) \rangle - 0.3m_i^{1/3} \tag{5.5}$$

To estimate τ_i, we consider the activation barrier that governs the subsegments' thermal fluctuations. Particularly, we assume the subsegment extension explores a free energy well, $G(X_i)$, defined by the thermodynamics of a stretched WLC. As shown in the methods, $G(X_i)$ can be estimated by integrating the WLC force-extension function (which accounts for internal entropy lost by the chain as it is stretched), and subtracting an fX_i term (which accounts for the work performed by the stretching force) [44]. Notably, the energy well predicted from this procedure depends on f, ℓ_i and l_p, with the latter again taken to be 0.6 nm. An example of an energy well calculated in this way is shown in Fig. 5.1D. Given $G(X_i)$, we can then calculate the activation barrier controlling collapse as $\Delta G_i^\dagger = G(r_i) - G(\langle X_i \rangle)$. We then enforce Arrhenius kinetics, and set the mode timescale to be

$$\tau_i = \tau_0 e^{\Delta G_i^\dagger / k_B T} \tag{5.6}$$

In practice we are concerned with the relative (multiplicative) magnitude of various relaxation modes, not the absolute timescale, so we set $\tau_0 = 1$ throughout.

The discrete model gives the entire relaxation trajectory at a given force from the values of all α_i and τ_i, which in turn are fixed by the number and lengths of the globule-forming subsegments, i.e. the set of residues $\{m_i\}$ in the subsegments. The choice of $\{m_i\}$ can be made in a physically reasonable fashion by inspecting the 168-residue sequence of the construct. While NFLt contains a significant negatively-charged polyelectrolyte domain that is unlikely to collapse, there are four regions that are better candidates: three largely neutral segments enriched with somewhat hydrophobic residues (aa 1-28, 35-59, 76-82), and a region near the C terminus with multiple positive charges that has been implicated in prior work as having large electrostatic interactions, here assumed to occur with nearby negative charges (aa 152-162). This suggests using a set $\{m_i\} = 7, 11, 15, 29$. Notably, such a set has only 62 total residues, a modest fraction of the total length; this is in accord with experiment, where the chain's length changes upon relaxation are only a small amount compared to the total contour length (see Fig. 5.1B). That said, these $\{m_i\}$ values are only rough estimates; they are useful for testing whether reasonable sequence metrics can give reasonable relaxation behaviors, but should not be regarded as precise.

5.4 Results

Two-step force jumps indicate individual relaxation events are independent and heterogeneous

We sought to determine whether our logarithmic relaxation data is caused by independent relaxation events, as called for in both the continuum and discrete models. Recent work on glassy systems has suggested that multiple classes of logarithmically relaxing

system exist [60]. However, systems with independent relaxation events typically exhibit history dependent behavior [60, 153]. Based on the continuum model, if the system is composed of independent relaxation events, it should exhibit history dependent behavior when subject to a two-step force jump.

Our first two-step force jump protocol is illustrated in Fig. 5.2A. First, we allowed the protein to relax for a long time ($\gg 100\,\mathrm{s}$) at an initial force f_1. Next, we increased the force to f_2 and waited for a time, t_w, during which the polyprotein extension increased logarithmically in time. Then, we decreased the force back to f_1 and measured the resulting relaxation. After the second force jump, we found that polyprotein extension time dependence varied with t_w. Initially, t_w shows a logarithmic time dependence, but, for short t_w, it slows down and appears to approach a steady state (e.g., $t_w = 3\,\mathrm{s}$ in Fig. 5.2B). As t_w increases, the relaxation becomes more logarithmic until, at $t_w \gtrapprox 100\,\mathrm{s}$, it is indistinguishable from a one-step force jump.

Our second two-step force jump protocol is illustrated in Fig. 5.2C. As before, we first allowed the polyprotein to relax for a long time ($\gg 100\,\mathrm{s}$) at an initial force f_1. Next, we decreased the force to f_2 and waited for a time, t_w. Then, we increased the force to a new intermediate force, f_3, such that $f_1 > f_3 > f_2$. After the second force jump, we observed a variety of different relaxation behaviors depending on t_w. For short t_w (e.g., $t_w = 3\,\mathrm{s}$ in Fig. 5.2D), we found that extension largely decreases over time. For longer t_w (e.g., $t_w = 100\,\mathrm{s}$ in Fig. 5.2), we found that the ΔL largely increases over time. As previously reported[143], for intermediate t_w (e.g., $t_w = 10\,\mathrm{s}$ in Fig. 5.2D), we found that ΔL shows a nonmonotonic time dependence, first increasing until it reaches a peak, then decreasing.

Together, the two-step force jump data indicate that the NFLt polyprotein relaxation behavior can be explained by multiple, independent, and heterogeneous relaxation events. These results lend confidence to the use of the continuum and discrete models in

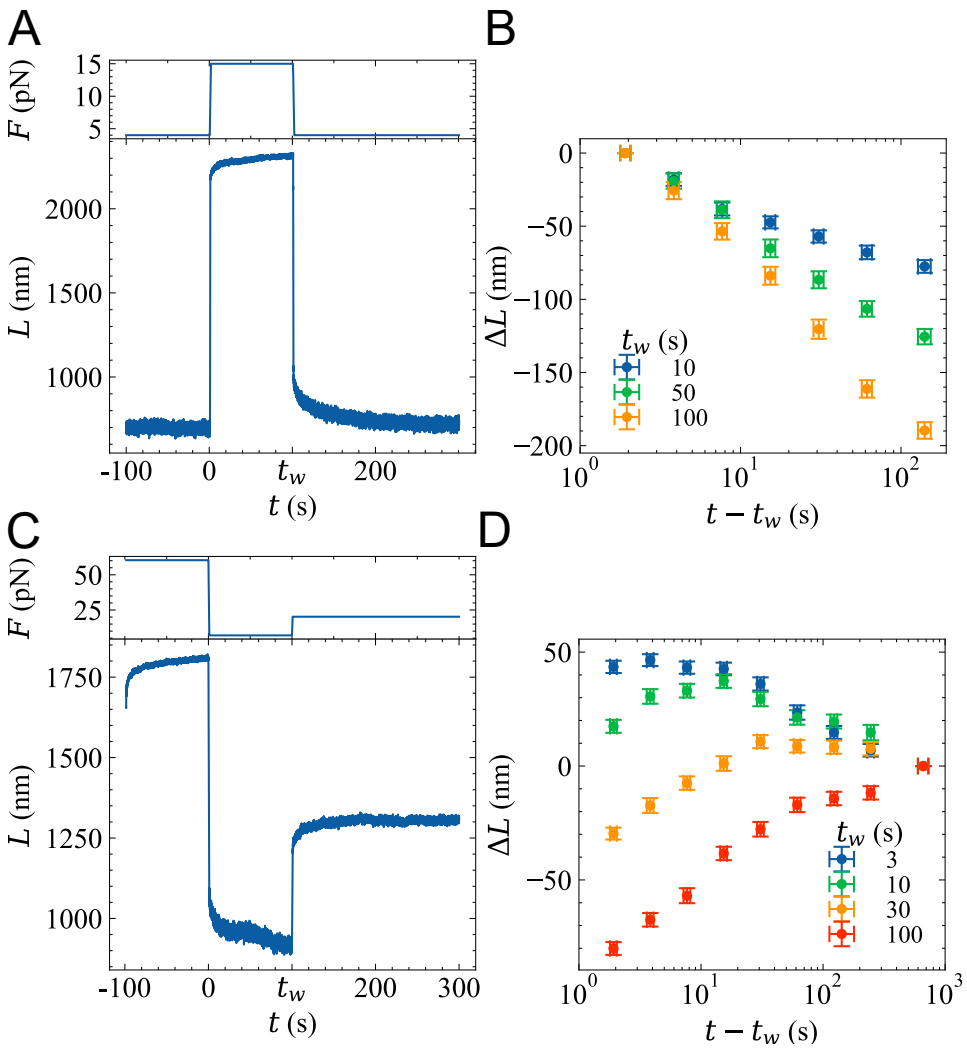

Figure 5.2: Two-step force jumps demonstrate history dependent relaxations. (A) A two-step force jump protocol where the force is first increased from f_1 to f_2. After waiting for t_w at f_2, the force is decreased back to f_1. (B) After the second force jump, the extension initially decreases with a logarithmic time dependence before diverging, depending on t_w. (C) Another two-step force jump protocol where the force is first decreased from f_1 to f_2. After waiting for t_w at f_2, the force is increased to f_3. (D) At f_3, the extension decreases, increases, or shows a nonmonotonic time dependence, depending on t_w.

interpreting our relaxation data.

Continuum model quantitatively predicts force jump decreases

To study the force dependence of the relaxations, we carried out a series of one-step force jumps between a single high force and various low forces. As predicted by the continuum model, we observed a logarithmic time dependence for both force increases and decreases, so we fit them to a logarithmic function to determine the log slope, b (Fig. 5.3A). When $f_2 > f_1$, we observed an increase in the differential extension $\Delta L = L(t) - L(t_0)$ and that b increased with f_2, as expected for a structure breaking process. When $f_2 < f_1$, we observed a decrease in ΔL and that b decreased with f_2, as expected for a structure forming process. Additionally, we found that both force jump increases and decreases depended on f_1, as predicted by the continuum model. Specifically, we found that b increased monotonically with the force jump magnitude $\bar{f} = f_2/f_1$ (Fig. 5.3B). This result indicates that the relaxations are also sensitive to the history of applied forces.

In total, we collected 450 one-step force jump relaxations across 16 polyproteins with varying $\bar{f}$ and fit each relaxation to determine b. To compare b across polyproteins with varying lengths, we used the normalized log slope, $\bar{b}$. Plotting $\bar{b}$ vs. $\bar{f}$ collapses all the data onto a single curve (Fig. 5.3C).

We compared the data to the continuum model by fitting it to Eq. 5.4, which has a single fitting parameter. First, when fitting to all the data together, we found poor agreement with the model, as judged by the reduced chi-square fitting metric, $\chi^2 = 70$, and substantial deviations in the standardized residuals. We note that a reduced chi-square fitting metric of 1 is expected for a 'good' fit with Gaussian distributed errors.

Next, we fit the $\bar{f} < 1$ and $\bar{f} > 1$ data to Eq. 5.4, separately. For $\bar{f} > 1$, we found the best fit to be $\rho = 0.4 \pm 0.1$ and a reduced chi-squared fitting metric of $\chi^2_\nu = 69.4$. Both the data and standardized residuals show systematic deviations from the fit. We attribute the poor fit to the experimental noise in the data, which is of a similar

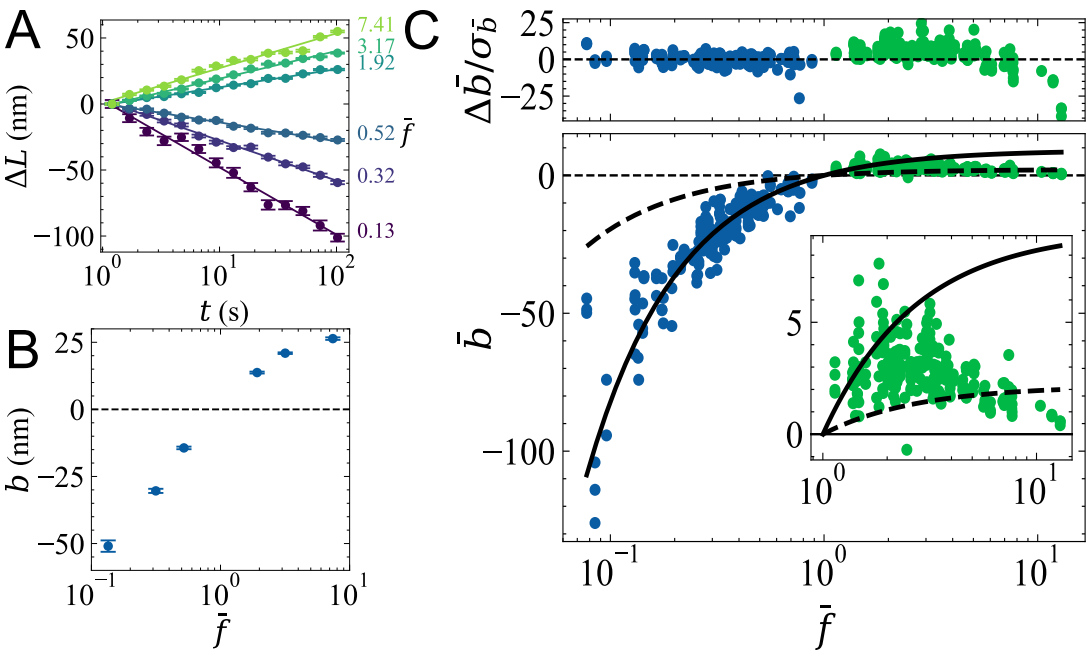

Figure 5.3: (A) Typical one-step force jump relaxation data and (B) b values from a single polyprotein using different values of $\bar{f} \equiv f_2/f_1$. (C, bottom) The normalized log slope, $\bar{b}$, as a function of the force jump magnitude, $\bar{f} \equiv f_2/f_1$ for all 450 relaxations on 16 separate polymers, where $\bar{f} < 1$ values are blue and $\bar{f} > 1$ are green. Errorbars are omitted for visual clarity. The solid and dotted lines represent fits to $\bar{f} < 1$ (blue) and $\bar{f} > 1$ (green) values with best-fit parameters $\rho = 0.4 \pm 0.1$ and $\rho = 0.108 \pm 0.004$, respectively (error estimated from bootstrapping). (C, inset) Zoomed in region of $\bar{f} > 1$ values. (C, top) Standardized residuals, $\Delta\bar{b}/\sigma_{\bar{b}}$, where $\Delta\bar{b}$ is the difference between the data and the fit, and $\sigma_{\bar{b}}$ is the error estimate of the data.

magnitude to the expected variation in $\bar{b}$ (Fig. 5.3A, inset). For $\bar{f} < 1$, we found the best fit with $\rho = 0.108 \pm 0.004$ and a reduced chi-squared fitting metric of $\chi^2_\nu = 10.7$. In contrast to the $\bar{f} > 1$ data, both the data and standardized residuals show no systematic deviations from the fit, suggesting the model captures the key features of the system in this regime. We attribute the elevated reduced chi-squared values to nongaussian or systematic underestimates of the errorbars, either due to unknown systematic errors or physical effects not captured by the model.

Sensitivity of relaxations to solution conditions

With an aim to determine the biochemical driving force behind our relaxation data, we performed one-step force jump experiments in various solution conditions. The flow cell permits solution exchange and heating between experiments, allowing us to differentiate the effects of different solution environments on individual polyproteins. We found that the relaxations are insensitive to screening by monovalent salt (NaCl) up to 1000 mM (Fig. 5.4A). In contrast, we found that the relaxations are significantly inhibited by high concentrations of an ionic denaturant, 6 M GuHCl (Fig. 5.4B). A solution with similar viscosity (20% glycerol) to the denaturant does not inhibit relaxations (Fig. 5.4C). Finally, we found that the relaxations occur faster at higher temperatures (Fig. 5.4D).

Discrete model qualitatively predicts force jump decreases

We also formulated a more physically realistic discrete model for the relaxations. Since the experimental data measured relaxations spanning two decades in time, we focused the model on that range, which meant only using forces in the range $6\,\mathrm{pN} \leq f \leq 30\,\mathrm{pN}$. For $f < 6\,\mathrm{pN}$, the model predicts relaxations that are complete in less than two decades of time, while for $f > 30\,\mathrm{pN}$, the relaxations become so widely spaced in time that only one occurs within the two decade window. Predicted trajectories in that force range are shown in Fig. 5.5A. The modeled relaxations show some waviness, i.e. they are not perfectly logarithmic; this is an unavoidable consequence of not using the continuum approximations. However, they are roughly logarithmic in shape, particularly for the lower forces. Further, the magnitude of relaxation ($\sim 5\,\mathrm{nm}$ over two decades in time) is similar to that measured in the experiment, when scaled to a single NFLt.

The resulting discrete-model predictions of $\bar{b}$ vs. $\bar{f}$ match relatively closely to the experimental data points (Fig. 5.5B), at least in the range where the model predicts two-

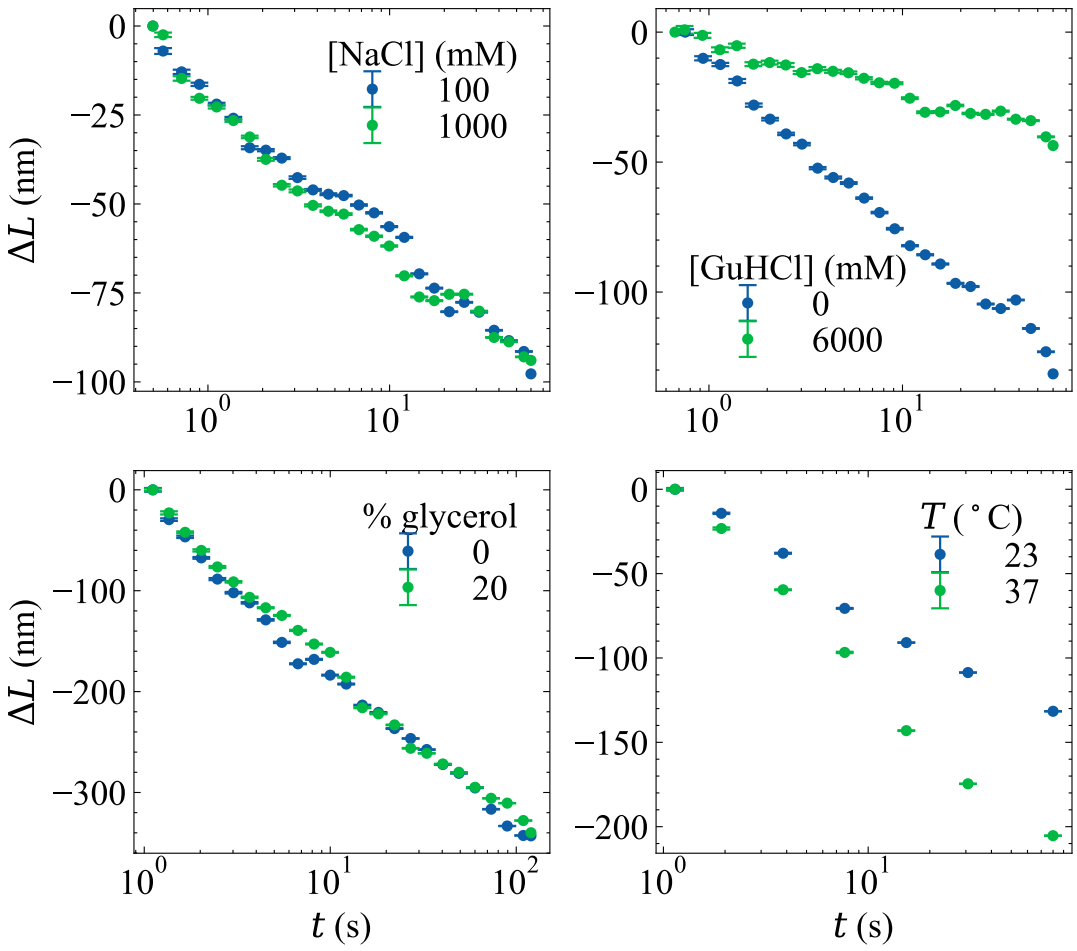

Figure 5.4: Solution dependence of logarithmic relaxations. One-step force jump on a polyprotein from (A) 68 pN to 9 pN in 10 mM to 1000 mM NaCl, (B) 85 pN to 15 pN in 10 mM NaCl and 6 M GuHCl, (C) 70 pN to 6 pN in 0 and 20% glycerol, and (C) 60 pN to 5 pN at 23 and 37 °C.

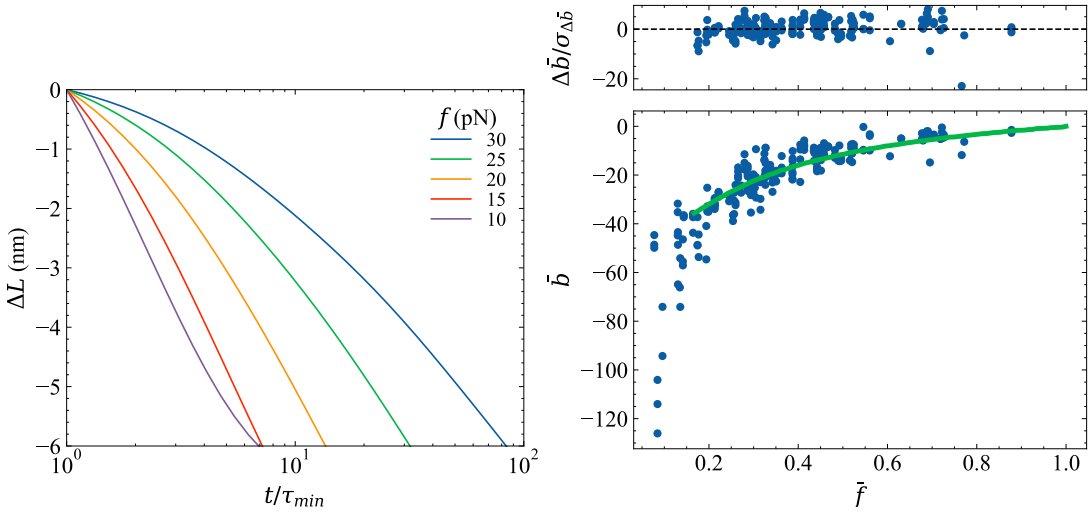

Figure 5.5: Discrete model predictions. (A)Predicted relaxations, for various forces, from the discrete model for a set of globule sizes $\{m_i\} = 7, 12, 17, 22$. Solid lines: relaxation trajectories predicted from the discrete model for forces between 6 and 30 pN, as labeled. Dashed lines: best-fit logarithms to the two decades of relaxation plotted for each trajectory. For each trajectory, the time axis is rescaled by $\tau_{min} = \tau_1/2$, where τ_1 is the relaxation time constant for the fastest (7-residue) globule. (B) Normalized log-slope of relaxation, $\bar{b}$, vs. normalized force, $\bar{f}$, for both experimental data (points) and the discrete model using globule sizes $\{m_i\} = 7, 12, 17, 22$ (line). The model is truncated at low $\bar{f}$ because it does not predict near-logarithmic relaxations for $\bar{f} < 0.2$, in disagreement with the data, in which log relaxations persist down to $\bar{f} \lesssim 0.1$. Yet in the range of validity, the agreement is relatively good; note no fitting is performed here.

decade log relaxations ($f \geq 6$ pN, corresponding to $\bar{f} \geq 0.2$). The major disagreement occurs for $\bar{f} < 0.2$, where the discrete model does not predict two-decade logarithmic relaxations, yet the experiment continues to show them.

5.5 Discussion

Taken together, our analyses suggest that the origin of the relaxation behavior is the (un)folding of multiple, independent, and heterogeneous subsegments (Fig. 5.1A). The heterogeneity in the subsegments accounts for the remarkably slow nonexponential relaxations under an applied force. The effect of the applied force varies across the dif-

ferent subsegments leading to the broad distribution of timescales. As the force increases this distribution broadens, in analogy to the slow kinetics caused by low temperatures on structural glasses [3]. The history dependence implies that at least some of the subsegments are independent, i.e., noncooperative. The nearly 10% contour length change indicates that there must be multiple subsegments in each tail (Fig. 5.1B). This picture is supported by the discrete model that quantitatively predicts the force dependence of the relaxations with physically realistic parameters.

While we attribute the relaxation to multiple subsegments in each tail, the polymeric nature of the polyproteins allows for the possibility of intertail interactions [158, 159]. However, the applied force, during the relaxations, limits the interactions to short-range interactions with nearby residues. For example, a polyprotein with an $\approx 1\,\mu m$ contour length corresponds to $N \approx 15$ NFLts (as in Fig. 5.1B). Given that there are no discernible discrete transitions, interactions of $\approx 1\,nm$ between abutting neighbors would not be able to account for the observed $\approx 100\,nm$ length changes. Hence, intratail interactions are more likely to be the dominant source of the relaxations.

The free energy of the intratail interactions can be roughly estimated by their ability to relax against a known load. Extrapolating to zero force, the free energy of these structures is likely between $3\,k_BT$ to $9\,k_BT$; the wide range is due to the sensitivity in the choice of ℓ [143]. This range encompasses prior estimates of local structures in IDRs [51].

Multiple sources of intratail interactions have been identified in the literature [21], including secondary structure, electrostatic attractions, and amorphous globules (similar to a polymer in a poor solvent). The NFLt is overall negatively charged with an average -0.25 e per residue; however, it contains some positively charged residues. The charged residues are largely segregates into the later two-thirds of the tail. The first two-thirds contain a large fraction of uncharged, mostly polar, residues, such as serine and tyrosine.

Previous work has shown that intertail electrostatic attractions play an important role in assembling reconstituted neurofilament networks [52, 56, 54]. Similar intratail attractions could presumably lead to electrostatic attractions. To test this, we sought to screen out electrostatic attractions with the addition of 1 M monovalent salt (Fig. 5.4A). The relaxations were insensitive to the addition of the monovalent salt, suggesting that electrostatic attractions are not the dominant source of the intratail interactions; however, previous measurements have shown that the NFLts can be insensitive to monovalent salt even when electrostatic interactions are present [52, 150]. Hence, it is still possible that electrostatics play a role in the relaxations.

Another source of compaction is polar regions in IDRs that form amorphous globules. The collapse of these segments into globules is attributed to poor solvation of the nonpolar backbone by the aqueous environment, which can be improved with denaturants, such as Urea and GuHCl [23]. These amorphous globules can sometimes act as molecular recognition regions and are implicated in amyloid formation [21]. The first one-third of the NFLt is relatively devoid of charged residues and rich in polar residues, including the aromatic residue tyrosine, suggesting it could form amorphous globules. Indeed, with the addition of denaturant, the relaxations are significantly inhibited, suggesting either hydrophobic interactions or hydrogen bonding plays a role in the relaxations (Fig. 5.4B). Similar experiments in 20% glycerol do not have effect on the relaxations, suggesting that the changing viscosity with denaturant is not the source of the inhibition (Fig. 5.4C). Further evidence for hydrophobic interactions is given by the increase in the rate of the relaxations with increasing temperature (Fig. 5.4D). Nevertheless, multiple sources of compaction are possible.

While IDRs are devoid of tertiary structure, many contain transient or stable secondary structure, which could lead to the observed length changes. Sequence-based secondary structure algorithms disagree on whether NFLt contains α-helices [36]; how-

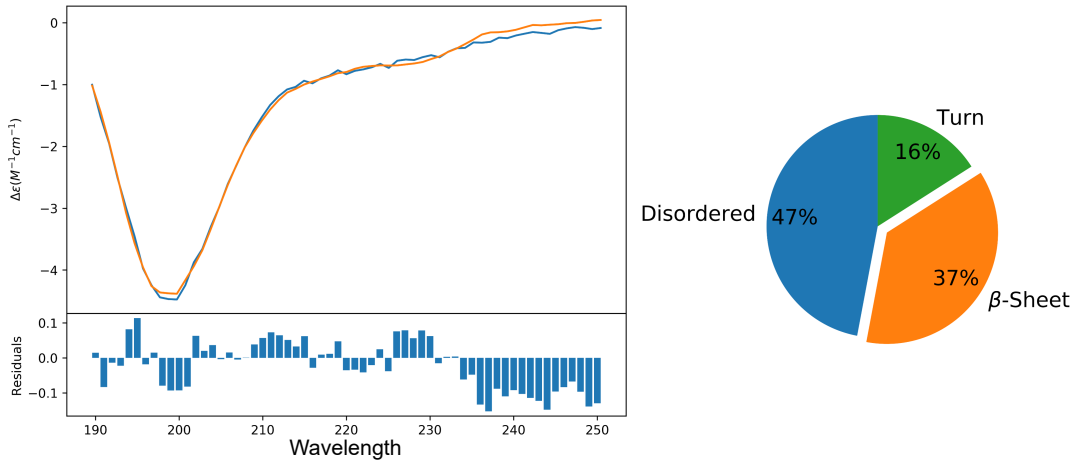

Figure 5.6: (Left) Molar ellipticity of single NFLts taken using a 1 mm path length in a 10 mM phosphate buffer at pH 7. Fits and residuals are based on the web-based BestSel fitting algorithm [165]. (Right) Fits indicate that the tails are mostly disordered with a fraction of beta-sheets and beta-turns. The BestSel algorithm can misclassify IDRs as containing beta-sheets, casting doubt on whether they are present in the tails [165].

ever, previously reported circular dichroism measurements suggest that they only adopt α-helices at pH 4 and not at pH 7 [57]. We also conducted our own circular dichroism measurements on the NFLts and fit the data using the web-based BestSel fitting algorithm [165]. Consistent with previous measurements, the circular dichroism measurements indicate that NFLts are largely disordered and devoid of alpha-helices (Fig. 5.6). The BestSel fits indicate a substantial fraction of beta-sheets could be present, but the algorithm is known to misclassify IDRs as containing beta-sheets [165], casting doubt on their presence in the tails. The algorithm also identifies a significant fraction of beta-turns in the tails. The presence of turns is consistent with previous work based on sequence metrics [57]. It is possible that the beta-turns are responsible for some of the relaxation events and their hydrogen bonding would be disrupted by the denaturant.

Regardless of the source of the relaxations, they can largely be explained by a continuum model based on a phenomenological framework developed for glassy systems [60] coupled with Bell-Zhurkov mechanochemistry [46, 45]. This model qualitatively predicts

the logarithmic time dependence of the one-step force jump data and the history dependence of the two-step force jump data. After fitting the normalized relaxation data to Eq. 5.4, it also quantitatively predicts the force dependence of b for the force jump decreases, i.e., $\bar{f} < 1$ (Fig. 5.3C). Likely due to the experimental noise in the data, the model does not capture the force jump increases, i.e., $\bar{f} > 1$. However, the different fits to the $\bar{b} < 1$ and $\bar{b} > 1$, suggest that the logarithmic increases represent activation barrier crossings from the compact to extended states, and decreases represent crossings from the extended to compact states. Furthermore, the model is self-consistent; the best fit estimate of $\rho < 1$ for both fits is consistent with the physical restriction on the activation barrier crossing $\delta x < \ell$. Assuming that $\ell \sim 1\,\mathrm{nm}$, the fit value of ρ from the $\bar{f} < 1$ data implies an upper limit on the number of independent forming structures, $n_{max} = 1/\rho \approx 9$. Together, these results support the use of the phenomenological framework.

Yet, the phenomenological framework was developed to explain the glassy relaxation dynamics of bulk systems (e.g. crumpled paper balls or semiconductor materials), containing thousands or millions of independent relaxation processes [60]. For such systems, the spectrum of relaxations contains many closely-spaced timescales, which can then be combined and simplified by applying methods of calculus, particularly through transforming the sum of discrete, exponential relaxations into a smooth, continuous logarithmic relaxation (Eq. 5.2) [162]. However, the conceptual underpinnings of this approach are unclear for the NFLt polyproteins: unlike the bulk systems for which the continuum theory was developed, the nanoscale size of the NFLt means it can only support a few independent relaxation events ($n \lesssim 10$). Thus, the mathematical criteria implicit in the discrete-to-continuous transformation are not rigorously met, leading to some uncertainty in interpreting the results of the model.

Hence, to better interpret our results, we formulated a discrete microscopic model for the NFLt relaxations that requires only a few independent relaxation events. While this

model does not result in a simple analytical function like the continuum model, it allows for more physically realistic parameters. For example, the continuum model artificially enforces that all independent relaxation modes share a common contour length, whereas the discrete model allows the contour length to vary, permitting more physically-realistic relaxation behavior.

Without any fitting parameters, the discrete model gives a quantitatively similar prediction for the normalized log slope vs. force jump magnitude (Fig. 5.5), at least in the range where the model predicts logarithmic relaxations. The disagreement at low forces could be due to larger (and thus slower) globules in the NFLt than in the model. Alternatively, it could be because of higher-order relaxations, where two globules, that each collapsed independently, coalesce with each other, leading to long-time relaxation behavior that would extend the trajectory in a way unaccounted for by the discrete model. Nevertheless, the discrete model suggests that physically realistic parameters could lead to the observed relaxation behavior.

The exponential time dependence of proteins under a constant applied force is largely based on polyproteins consisting of small folded, globular proteins [44]. While there have been previous reports of polyproteins of globular proteins exhibiting nonexponential time dependence under force [146], the deviations from an exponential time dependence are small and only apparent after averaging over many polyproteins. In contrast, our data show that single IDR polyproteins exhibit a remarkably slow, highly nonexponential time dependence over at least three decades in time [143]. Such nonexponential relaxations are reminiscent of the conformational changes of proteins in the absence of force [166, 7], but occur for a different reason: the heterogeneity comes from independently relaxing subsegments and not because of a rough native energy landscape. The NFLt kinetics are more similar to the complex kinetics of large multidomain proteins with independently folding subunits [167, 168].

5.6 Conclusion

In summary, our analysis indicates that NFLts likely contain multiple, independent, and heterogeneous hydrophobic globules. The formation of these globules under an applied force lead to a highly nonexponential time dependent mechanical response. It has been suggested that many IDRs contain subsegments that form local structures in an independent, noncooperative fashion [51, 14]. The NFLt is one of many IDRs that play a similar functional role in the cytoskeleton [35, 36]; hence, it is reasonable to expect that the rich and complex mechanical kinetics observed here could apply to other IDRs.

Chapter 6

Conclusions and Future Outlooks

6.1 Conclusions

In this dissertation, I described the improvement of SMFS techniques and the unexpected discovery of glassy dynamics in a model IDR construct.

In Chapter 2, I presented Tweezepy, a computational tool for calibrating forces in SMFS video-tracking experiments. Robust force calibration is critical for accurate results in these experiments. While robust methods have been available in the literature, they are not routinely used due to their computational complexity. This computational tool removes this barrier and provides a straightforward means of estimating the uncertainty in calculated forces, an important but often overlooked aspect in SMFS experiments. There have been calls for standardization to enhance reproducibility in the SMFS field[59]. Towards this end, Tweezepy provides a straightforward, reliable, and versatile tool for standardizing force calibration across SMFS experiments.

In Chapter 3, I presented work examining the effects of surfaces in MT experiments on low-force polymer elasticity. In the SMFS literature there has been considerably little attention give to low-force elasticity measurements, despite their relevance to most

biological and biomaterial situations. At low forces, I showed that surfaces in MT experiments can bias power-law force-extension behavior. I showed that it is possible to use this bias to independently estimate R_g, an important polymer structural length scale. This finding broadens the capability of MT experiments to study polymers.

In Chapter 4, I presented the unexpected discovery of glassy dynamics in a model IDR system, derived from disordered NFLt. In response to a change in applied force, I showed that the change in polyprotein extension exhibits an extraordinarily slow, logarithmic time dependence. I also showed that these relaxations were force-dependent and, after a two-step force jump, exhibited a nonmonotonic time dependence, indicative of memory effects. Based on these findings, I presented a continuum model that predicted the force-dependence of the relaxations.

Fundamentally, the continuum model was based on a phenomenological framework for bulk glassy systems; however, the conceptual underpinnings of this approach are for the nanoscale IDR. Bulk glassy systems have thousands to millions of relaxation events. For such systems, the spectrum of relaxations contains many closely-spaced timescales, which can then be combined and simplified by applying methods of calculus, particularly through transforming the sum of discrete, exponential relaxations into a smooth, continuous logarithmic relaxation [162].

In Chapter 5, I revisited the glassy dynamics to determine its microscopic origins. In addition to force decreases, I showed that force increases also produce a logarithmic relaxations. I addressed the biochemical origins of the relaxations and identified multiple possible sources. I also showed that the extension changes are history dependent, a characteristic feature of glassy systems that indicates multiple, independent, and heterogeneous relaxation events. To account for these findings, I presented a new discrete model based on physically realistic parameters that qualitatively accounts for the relaxation data.

In summary, I have developed new tools for SMFS and conducted high-precision nanomechanical experiments on a model IDR system. My work on SMFS instrumentation improves the usability of SMFS instruments and broadens their capabilities to measure polymer length scales. My work on NFLt reveals the rich structural and mechanical complexity of IDRs, which is likely to be relevant for its biological role and, more broadly, for many other IDRs.

6.2 Future Outlooks

My findings hint at a plethora of opportunities for improving SMFS instrumentation and characterizing the structure and function of IDRs.

Currently, SMFS instrumentation is relatively inaccessible to nonexperts. This is due, in part, to the lack of easily accessible and standardized software. Tweezepy was a first step towards making SMFS analysis more accessible. I envision a suite of standardized software tools for operating operating SMFS instruments and analyzing their data. In fact, our lab has already begun work on a suite of software tools for operating MT instruments.

While we showed that surface effects can be used to extract R_g from force-extension measurements, they remain a nuisance in most experiments. One way that SMFS techniques get around this issue is to use dsDNA handles to move proteins away from surfaces [169]. Unfortunately, when attached to a flexible polymer, dsDNA handles are too flexible to extract reliable force-extension data without prior knowledge of the persistence length. However, DNA origami is much stiffer, so it would not have the same issues. In collaboration with the Liedl lab, I worked with a summer student make a DNA tower for MT experiments using a construct from Ref. [170]. While we ran into some technical issues getting the origami to fold correctly, with a little more work, these DNA towers

could be useful for future MT experiments on short polymers.

To my knowledge, the work in this dissertation represents the first example of MT experiments on an IDR; however, there are many opportunities for further experimental work on single IDRs. With regards to the NFLt system, my findings indirectly suggest that it contains distributed structures of varying size. However, this could be more directly proven with NMR experiments. Alternatively, this could be checked by looking for glassy dynamics after removing portions of the NFLt sequence - e.g., only using subsegment A or B. Furthermore, the NFLts' low-force elasticity behavior remains to be resolved. Equilibrium elasticity measurements appear to be possible below $\approx 10\,\mathrm{pN}$. As of the writing of this dissertation, our lab is currently working on understanding its power law behavior.

Experimental data suggest that other IDRs form compact structures with noncooperative local segments [14]. Hence, it is reasonable to expect that the mechanism for glassy dynamics presented here could occur in other IDRs. However, glassy dynamics in other IDRs remains to be experimentally proven. MT experiments on other IDRs are likely to provide further insights into their structural and dynamic properties. Specifically, another IDR with a similar biological role is the tau protein. As of the writing of this dissertation, this protein is currently being studied by several labs on campus due, in part, to its prominent association with neurodegenerative disorders [171]. With the help of one of these labs, our lab has begun work on expressing and polymerizing tau protein for MT experiments.

Together with future findings, my findings are likely to lead to better SMFS instrumentation and data analysis, as well as a better understanding of the structural properties of IDRs, and thus, their associated biological functions and diseases.

Appendix A

Magnetic Tweezers

Much of the experimental work in this dissertation was carried out on a custom-built MT instrument. A schematic of this intrument is shown in Fig. A.1A. In short, the MT instrument is based on an inverted microscope with the addition of a magnet assembly above the sample. The focal plane is controlled via a piezoelectric device attached to an oil-immersion objective. A light source is focused between the magnet assembly onto the sample, allowing the sample to be imaged through the objective onto a high-speed camera.

A MT experiment begins with the preparation of a flow cell, consisting of a fluid channel in between two glass coverslips as described in Sec. B.4. In the fluid channel, the polymers of interest are tethered between the bottom glass coverslip and paramagnetic beads as described in Sec. B.3 (Fig. A.1B). The bottom surface of the flow cell contains a dense covalently bound PEG layer to prevent nonspecific adsorption of the polymer or beads. A small fraction of the PEG are functionalized to permit polymer attachment.

Once the flow cell has been prepared, it is placed on the microscope stage with the bottom coverslip touching the immersion oil on the objective. The high-speed camera is attached to a computer which displays the sample image on a monitor. The user can then

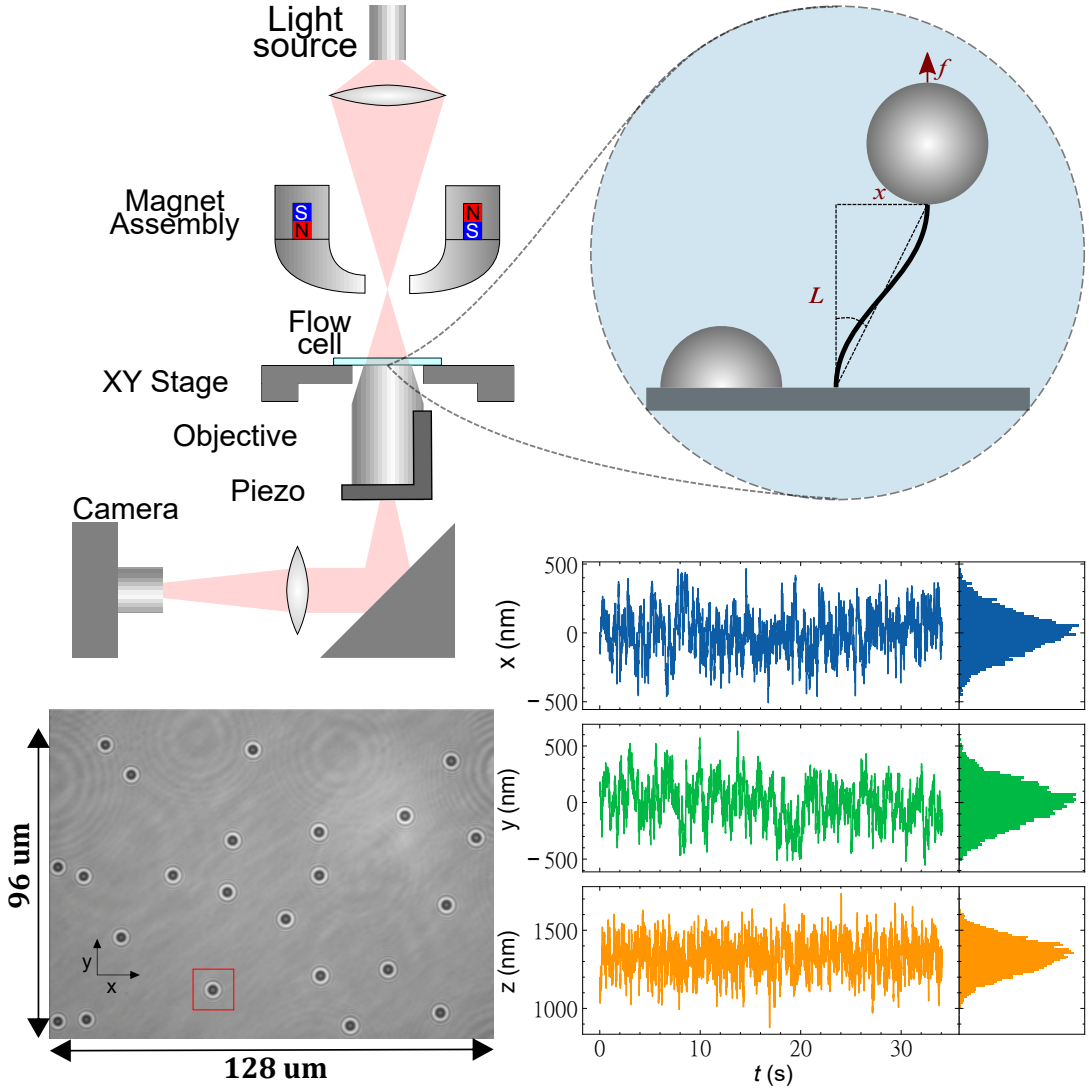

Figure A.1: Our custom-built magnetic tweezer. (A) Schematic of the magnetic tweezer. (B) The polymer of interest is tethered between a glass surface and paramagnetic bead. Tracking partially melted reference beads permits the removal of common mode noise. (C) A typical image of experimental beads. (D) A typical experimental bead trajectory with x, y, and z positions.

begin the experiment by looking for tethered polymers. A typical experiment consists of bringing the magnets close to the flow cell and translating the flow cell with the stage to find a field of view with many 'good' tethers. For my experiments, a good tether 1) only has one polymer tethered between the bead and the surface, 2) has a polymer attached near the center of the bead, and 3) is ideally long ($\gtrsim 0.5\,\mu m$). 1 and 2 can be verified by rotating the magnets and tracking the bead's position. The polymers in a doubly tethered bead will wrap around each other when rotated, resulting in a characteristic extension change. A 'mistethered' bead has a polymer attached away from its center and will produce a limacon pattern when rotated[172]. The length of a polymer can be typically be estimated by eye from the change in its diffraction pattern when the magnets are close and far away from the flow cell.

Once good tethers are found, the x and y positions of the bead in the imaging plane are determined via centroid tracking (Fig. A.1C). The z position out of the imaging plane requires a calibration between the bead height and the diffraction ring pattern from the interference between the incident light of the light source and the light scattered off the bead[173]. During this calibration, the bead is typically held at a high force to suppress Brownian fluctuations.

Once the calibration is completed, the bead height (z-position) can be tracked; however, it is susceptible to instrumental drift. To minimize the effects of drift, partially melted reference beads (or stuck beads) are adhered to the bottom coverslip to serve as reference beads. Subtracting the common mode noise of the reference beads from the experimental beads eliminates most instrumental noise. As noted in Chapter 2, tracking errors are not removed by the reference beads.

In order to determine the polymer extension, the experimental bead are tracked in the absence of force - either by moving the magnet far away or taking it off the instrument - to determine the height of the glass surface. The difference between the height of the

bead and glass surface is the end-to-end extension, L, of the polymer(Fig. A.1B). These raw extension values are then multiplied by the ratio of the index of refraction of the experimental buffer and the immersion oil to give the final measured extension, L.

For a given magnet position, the force is determined by modeling the bead-tether system as an inverted pendulum. Using the bead fluctuations parallel to the magnetic field, the spring constant, κ of the harmonic potential is determined from the bead fluctuations as described in Chapter A. Applying Hooke's law and the small angle approximation, it follows that the applied force is $f = \kappa L$.

To make a force-extension curve (Fig. 3.1C), the magnets are moved further from or closer to the flow cell. Then, at a constant magnet position, a bead trajectory is collected for a predetermined amount of time. This process is repeated for the desired number of points (typically, 31 points). From the bead trajectory, the force is calibrated and time-averaged extension is calculated.

For force-jump experiments, the magnets are moved further from or closer to the flow cell, resulting in a force decrease or increase, respectively. In contrast to the force-extension curves, the magnet position and polymer extension is tracked during and after the magnet motion. Once the magnet motion stops, the time, t, is set to zero and the polymer extension is monitored.

A.1 Instrument setup

The custom-built MT instrument (Fig. A.1) is based on designs in Refs.[98, 174]. To illuminate the beads, light is passed through a focusing lens and a magnet assembly onto a flow cell. The instrument's light source is a super-luminescent diode (SLD) (QPhotonics QSDM-680-2) or fiber-coupled light-emitting diode (LED) (Thorlabs M470F3). The SLD has similar temporal coherence as the LED, but since the SLD is single-mode, it

has better collimation preventing clipping from the magnets[173]. It also leads to sharper bead diffraction patterns; however, it increases speckle and has a longer spatial coherence, leading to diffraction patterns from dust on the surface of the flow cell.

The magnet assembly consists of a machined aluminum casing with two pairs of 1/2" NdFeB magnets oriented north-north and south-south, with two low-carbon steel pole pieces that confine the large magnetic field gradient near its tips. The magnet assembly is connected to linear translation and rotation motors (Physik Instrumente M-126, C-843). The motors control the distance between the magnet assembly tips and the top of the flow cell, as well as the orientation of the magnetic field.

The flow cell contains the polymer of interest and is assembled as described in Appendix B.4. In short, the flow cell consists of layer of parafilm wax, sandwiched between two No. 1 glass coverslips (Fig. B.3A). The bottom coverslip ($22\,\mathrm{mm}^2$) is passivated with a covalently attached PEG layer (Microsurfaces Inc.) to prevent nonspecific adsorption of the polymers and paramagnetic beads. A small fraction of the PEG layer is functionalized with a N-Hydroxysuccinimide (NHS) or maleimide functional groups to facilitate attachment of the polymer of interest. The flow cell is secured to an XY microscope stage (Physik Instrumente M-545.2P, operated by a Physik Instrumente C-867-260 controller) with Scotch tape. The stage moves the flow cell with $0.5\,\mu\mathrm{m}$ resolution and a $25\,\mathrm{mm}^2$ travel range, enabling the user to search for tethered polymers. In addition, the stage motors are self-locking at rest and hold the stage in a table position without generating heat.

After passing through the flow cell, the light is collected by a high numerical aperture oil immersion objective (Nikon CFI Plan Fluor 60XS Oil NA 1.2). The objective is mounted on a piezoelectric nanopositioning device (Physik Instrumente P-725.2CD) with $0.5\,\mathrm{nm}$ resolution and a $400\,\mu\mathrm{m}$ travel range, which allows both coarse and fine control of the image focal plane. After passing through the objective, the light is reflected by a

mirror through a 200 mm focal length lens onto a high-speed CMOS camera (Mikotron EoSens 3CXP). The high-speed camera captures up to 563 frames per second (fps) at full resolution (2,336 x 1,728 pixels) or up to 7,490 fps at reduced resolution (128 x 128 px). The camera transfers the images to the computer via a frame grabber (BitFlow Cyton-CXP4). Unless otherwise stated, all of the data in this dissertation were collected at 400 fps.

A.2 Tracking and data analysis

The computer receives and processes frames from the high-speed camera to extract the three-dimensional position of the beads. This is accomplished in real-time using a custom-written program in LabView based on a GPU-accelerated algorithm from Ref. [174].

To operate the program, the user first selects the beads of interest. Then, the program performs a calibration step, where it images the beads while changing the focal plane in 0.1 µm increments for 5 µm to 10 µm. From these images, the program computes the average radial profiles of the diffraction rings at each focal plane and inputs them into a look-up table. During an experiment, the program uses this look-up table to determine the bead's vertical (z) position with a typical accuracy of 1 nm to 10 nm [155, 98]. To determine the lateral (x, y) bead positions, the program uses a centroid tracking algorithm that has a typical accuracy of 1 nm to 2 nm [98].

To analyze the data, it needs to be converted into real units. The program collects the lateral bead positions in pixels and the vertical position in uncorrected microns. The lateral positions are converted to nm using a precalibrated conversion factor for each objective and lens configuration. The vertical positions are also converted to nm using another correction factor that accounts for the difference in the index of refraction of the immersion oil, n_{oil} and experimental buffer in the flow cell, n_{sol}. The correction

factor is the ratio of the two refractive indices, n_{sol}/n_{oil}. At 20 °C, the refractive index of water is $n_{sol} \approx 1.333$ and Type A immersion oil is $n_{\text{oil}} \approx 1.515$, so the correction factor is $n_{sol}/n_{oil} \approx 0.880$. This correction factor varies with the experimental buffer and temperatures. All of the experiments in this dissertation calculated this correction factor using refractive indices from Ref. [175].

During the experiment, the measured bead positions will drift due to instrumental noise. This common mode noise can be removed by simultaneously tracking immobilized "reference beads" and subtracting their positions from the experimental beads.

Lastly, the bead's vertical position is converted to polymer extension, L, by calculating its height above the glass surface. The height of the glass surface is chosen to be the beads' 2.5th percentile z position when the magnet assembly is at the furthest distance from the flow cell.

A.3 Determining the applied force

The applied force on each tethered bead depends on the properties of the magnet assembly and beads. The paramagnetic beads have an inherently asymmetric distribution of magnetic material causing them to align along their axis with the magnetic field of the magnets (in our setup, this is the x-direction). The magnetic gradient applies a force towards the magnets given by

$$\vec{f} = \left[\vec{m}(\vec{B}) \cdot \vec{\nabla}\right] \vec{B} \tag{A.1}$$

where $\vec{B}$ is the magnetic field and $\vec{m}(\vec{B})$ is the bead's field-dependent magnetic moment [155]. For a typical magnetic position, the magnitude and gradient of the magnetic field varies by no more than 0.02% and 0.2% over the field-of-view [98]. Hence, theoretically,

it should be possible to determine the applied force *a priori*.

In practice, it is more accurate to determine the force *a posteriori* separately for each bead and magnet position. For MT experiments, the bead-tether system is modeled as an inverted pendulum. As described in Chapter 2, the spring constant of the harmonic potential, κ, is determined from bead fluctuations parallel to the magnetic field. Applying Hooke's law and the small angle approximation, it follows that the applied force is $f = \kappa L$.

The paramagnetic bead's are polydisperse, so each bead has a different amount of magnetic material and magnetic moment. For constant volume magnetization, the magnetic moment, and thus the force, increases with its volume, R_B, $f \propto R^3$[176]. We use Dynabeads MyOne C1 ($R_B \approx 0.5\,\mu m$) or M-280 ($R_B \approx 1.4\,\mu m$) for their low polydispersity; however, they still have a size distribution that is $\approx 3\%$, leading to a $\approx 9\%$ variation in their magnetic moment and expected forces [98]. Indeed, across different beads our calculated applied forces typically vary by $\approx 10\%$.

The gradient of the magnetic field, and thus the force, depends on the distance between the magnet assembly and the beads. The force decays roughly exponentially with this distance with a characteristic length scale of $\sim$1 mm, resulting in an effective stiffness of 1×10^{-6} pN/nm [39]. This means that the applied force is extremely stable: a 10 µm change in the height of the paramagnetic bead will only change the force by 0.01 pN. This stability is one of the major advantages of the MT technique because it enables the application of low and/or constant forces over long periods of time.

Appendix B

Experimental Protocols

B.1 Neurofilament light polyproteins

The main polymer of interest in this dissertation is the model IDR system. Our model IDR system is derived from the disordered tail domain of the mouse neurofilament light (UniProt accession number P08551).

The NFLt gene was ligated into a pET vector with a PagP fusion protein with a short linker containing the sequence SRHW. The PagP improves expression of IDRs [177] by accumulating in insoluble inclusion bodies. The SRHW linker is readily cleaved via a non-enzymatic nickel-assisted peptide bond hydrolysis [178]. After purification, only the NFLt, with some minor sequence modifications, remains.

The full length NFLt and PagP fusion gene was determined via sequencing to produce the amino acid sequence:

M G C G S G A S E N L Y F Q G A S *T R L S F T S V G S I T S G Y S Q S S Q V F G R S A Y S G L Q S S S Y L M S A R S F P A Y Y* N *S H V Q E E Q T E V E E T I E A T K A E E A K D E P P S E G E A E E E E K E K E E G E E E E G A E E E E A A K D E S E D T

K E E E E G G E G E E E D T K E S E E E E K K E E S A G E E Q V A K K K D **G G G S C G** S R H W Y F L S E F M N A D E W L T T F R E N I A Q T W Q Q P E H Y D L Y I P A I T W H A R F A Y D K E K T D R Y N E R P W G G G F G L S R W D E K G N W H G L Y A I A F K D S W N K W E P I A G Y G W E S T W R P L A D E N F H L G L G F T A G V T A R D N W N Y I P L P V L L P L A S V G Y G P A T F Q A T Y I P G T Y N N G N V Y F A W I R F Q F L E H H H H H H

The purified protein sequence is bolded and the native sequence is italicized. From the 146 residues in the native sequence, a single glutamic acid (E) was deleted from the N-terminus. A single Threonine (T, boxed) was mutated to an asparagine (N) to suppress unintended cleavage. A canonical Tobacco etch virus (TEV) protease site [179] was added to the N-terminus to enable *in situ* specific cleavage. Cysteines (C) were added to the N and C termini to enable polymerization. Glycines (G) and serines (S) were added around the cysteines to facilitate its availability for polymerization. In total, our construct contains 23 additional residues, 17 at the N-terminus and 6 at the C-terminus, for a total of 168 residues.

Expression, purfication, and polymerization

Following a similar protocol to Pregent *et al.*[150], competent Escherichia coli BL21(DE3) Rosetta were transformed with the modified pET vector and plated on agar plates containing 100 µg/µL ampicillin. Colonies from these plates were picked and used to inoculate a 1 L Terrific broth containing 100 µg/µL ampicillin and 30 µg/µL chloramphenicol. Expression cultures were grown in a baffled Erlenmeyer flasks in a shaking incubator at 37 °C at 300 rpm for 3 h to 5 h until the optical density at 600 nm reached 0.7–1. Protein expression was induced by addition of Isopropyl b-D-1-thiogalactopyranoside (IPTG) at

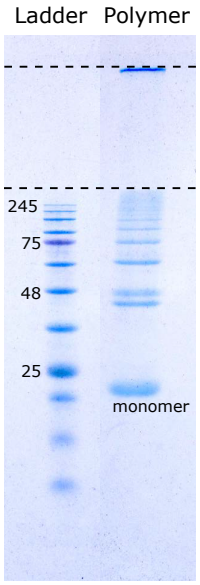

Figure B.1: Electrophoresis assay (Sodium Dodecyl Sulfate (SDS)-Polyacrylamide Gel Electrophoresis (PAGE)) of the polymerization product. Polyproteins are separated on a polyacrylamide gel with 3 regions of either 3, 6, or 15% polyacrylamide from top–to-bottom (marked by the dashed lines). (Left) Standard protein ladder with sizes in kDa. (Right) Polymer sample. The monomer and polymers that are $\lesssim 300$ kDa (≈ 6 units) have entered the 15% polyacrylamide and are clearly resolved. In certain cases, there are two bands of similar weight. These are likely closed loops and open-ended polymers. A band is clearly visible near the interface between 3% and 6% polyacrylamide. These polymers are > 300 kDa.

a final concentration of 0.5 mM. The cultures were grown for 4 h before harvesting. Cells were pelleted and stored at −80 °C for later use. For purification of proteins, cell pellets were resuspended in a ratio of 10 mL lysis buffer to 1 g pellet. The lysis buffer contained 20 mM pH 8.0, 0.1% 2-Mercaptoethanol, 1% Triton and 0.5 mg/ml lysozyme. The resuspended pellets were incubated at 25 °C for 20 min followed by addition of 10 mM MgSO4 and 1:20,000 vol/vol Benzonase nuclease (250 units/1 µL) for 20 min at 25 °C. Cell debris and NFLt-PagP were precipitated by centrifuging at 18,500 x g for 30 min. The pellets were homogenized in a resuspension buffer containing 20 mM Tris-HCl pH 8, 6 mM GuHCl, 20 mM imidazole, and 0.1% 2-Mercaptoethanol, then centrifuged as before.

After centrifugation, the supernatant was flowed through a 10 mL home-packed nickel-

affinity column at a rate of 1 mL/min. The column was pre-equilibrated and washed with 100 mL washing buffer containing 20 mL Tris pH 8.0, 6 M GuHCl,20 mM imidazole, and 0.1% 2-Mercaptoethanol. NFLt-PagP protein was eluted with elution buffer containing 20 mM Tris pH 8.0, 500 mM imidazole and 0.1% 2-Mercaptoethanol. The protein was then dialyzed overnight in a 14 kDa molecular weight cut-off (MWCO) dialysis membrane against 1 h of 20 mM MOPS pH 8.5. For cleavage, the protein was adjusted to GuHCl 6 M MOPS 20 mM NiSO4 5 mM pH 8.5 and incubated for 20 h at 50 °C [178, 177]. After incubation, 100 mM EDTA and 0.1% 2-Mercaptoethanol was added to the solution.

The solution was dialyzed to 20 mM Tris-HCl , 2 mM EDTA, and 0.1% 2-Mercaptoethanol and centrifuged as before. The pellet was discarded and 6 M GuHCl was added to the solution. Cleaved protein was run through a 100 mL size-exclusion column (HiPrep 16/60 Sephacryl S-200 HR) pre-equilibrated with 1 M GuHCl, 20 mM Tris-HCl, 0.1% 2-Mercaptoethanol, pH 8. NFLt was dialysed against a buffer containing 20 mM Tris pH 7.8 and 0.1% 2-Mercaptoethanol and purified by reverse-phase HPLC using semi-preparative Vydac C18 column. Prior to purification, the column was equilibrated with water containing 0.1% TFA. For elution of NFLt, a linear gradient of acetonitrile (from 0 to 50%) in 0.1% TFA was flowed through the column at a flow rate of 3 mL/min for 30 min. Fractions containing the eluted protein were detected by absorbance at 280 nm. Final purity was >95% as determined by SDS-page gels.

For polymerization, monomer NFLts were first reduced with Dithiothreitol (DTT) and exchanged $2-3\times$ with a 50 mM MES pH 7 and 500 mM NaCl buffer using a 10 kDa spin column. The reduced tails (100 mM) were mixed with functional caps (maleimide-biotin and maleimide-azide) at a ratio of 1:100 and allowed to polymerize under oxidative conditions for one week. The mixture was allowed to polymerize for one week and yielded polymers with a molecular weight of > 300 kDa as determined by SDS-PAGE [Fig. B.1].

Prior to incubation in the flow cell, 2 µL of 1 µM NFLt polymers were added to a

500 µL pH 11 carbonate buffer to deaggregate NFLt polymers for ≈ 1 h. Deaggregated polymers were buffer exchanged with a 100 kDa spin column into a 10 mM MES pH 6.8 buffer with 10 mM NaCl, also removing unreacted caps and monomers prior to incubation in the flow cell.

B.2 Other polymers

In some parts of this dissertation, we use PEG, HA, or Double-stranded DNA (ds-DNA). While these polymers have interesting properties in their own right, they have generally been well-studied, and thus, are primarily used in this dissertation to verify expected behaviors.

HA with a molecular weight of ≈ 2.5 MDa was purchased from Creative PEGworks with a single biotin group at the reducing end, and thiol groups incorporated at random throughout the chain with a stoichiometry of one thiol group per chain.

PEG with a molecular weight of ≈ 1 MDa was a gift from Christina G. Rodriguez and Nathaniel A. Lynd (University of Texas, Austin). Briefly, it was synthesized by modification of a commercial PEG-diol with tosylated ends. The tosyl groups were displaced with potassium thioacetate, then deprotected to reveal the thiols. Biotin-maleimide was added with a stoichiometry of one per molecule, resulting in a large fraction of molecules with a single biotin at one terminus, and a single thiol at the other.

dsDNA of ≈ 8.2 kbp, corresponding to a ≈ 2.8 µm contour length, was synthesized from λ phage DNA via Polymerase Chain Reaction (PCR) using two primers labeled on the 5' end with a biotin and thiol group. Primer 1 is called JB15 with sequence thiol-5' GGGATACGGGAAAACGTAAA 3'. Primer 2 is called JB16 with sequence biotin-5' TTCAGCATGAGAAATTGCCT 3'. For PCR, combine the following:

Reagent	Volume (µL)
Taq polymerase master mix	10
Milli-Q ultrapure water	35
Primer 1 - JB15 10 µM	1.5
Primer 2 - JB16 10 µM	1.5
λ phage dsDNA (50 ng/µL)	2

Then, use a thermocycler to run the program (BJ):

Step	Temp (°C)	Time
1	95	30 s
2	95	30 s
3	59	30 s
4	68	4 min
5	Repeat 2-4	30x
6	68	5 min
7	4	hold

and clean with a zymogen DNA cleanup kit.

Purified dsDNA was characterized using the nanodrop and agaraose gel electrophoresis. The typical purified concentration was 100 ng/µL. The agarose gel showed a clear band of ≈ 8.2 kbp, which corresponds to an ≈ 2.8 µm contour length.

B.3 Polymer attachment

Unless specified otherwise, all solutions contain 0.05% Tween-20, a non-ionic surfactant, to prevent nonspecific adsorption between the polymer, glass surface, and param-

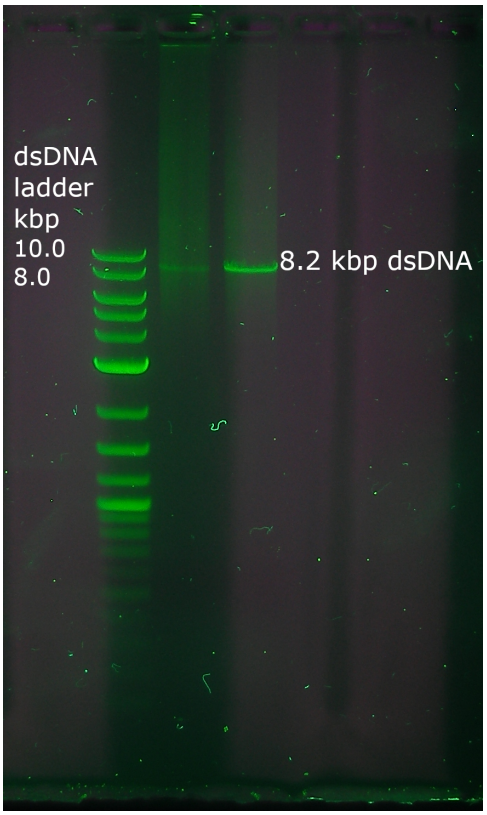

Figure B.2: A 1% agarose gel of the heterobifunctional dsDNA PCR product shows a distinct band at ≈ 8.2 kbp. From left to right, there is a DNA ladder (2-Log NEB), a sample that was stored at 4 °C for several months, and a fresh sample.

agnetic bead.

For NFL tail polyproteins, the bottom coverslip of the flow cell contained a PEG layer with low-density NHS functional groups. The flow cell was first washed with 1X PBS pH 7, then treated with a 1 mg/ml Dibenzocyclooctyne (DBCO)-Amine in a 1:1 Dimethyl sulfoxide (DMSO):PBS solution for 30 min to make a DBCO-functionalized PEG layer. The flow cell was washed with 1 M Tris pH 7 buffer, to deactivate any remaining NHS, and then with PBS. After washing, the polyprotein in PBS is incubated in the flow cell for 1 h at room temperature. The flow cell is washed again with the starting experimental buffer, typically 10 mM MES with 150 mM NaCl, to remove untethered polymers.

For PEG, HA, and dsDNA experiments, the bottom coverslip of the flow cell con-

tained a PEG layer with low-density maleimide groups. Biotin-PEG-thiol attachment was carried out in a flow cell in 1x PBS (pH 7.4) with 100 mM TCEP, and the HA-surface attachment was carried out in a 50 mM sodium phosphate buffer (pH 7.2), 50 mM NaCl and 10 mM TCEP. After attachment, excess polymer was removed by rinsing with low ($\leq 10\,\text{mM}$ ionic strength buffer (phosphate buffer or MOPS).

For all MT experiments, the biotin-labeled ends of the molecules were bound to $\approx 1\,\mu\text{m}$ (Dynabeads MyOne C1) or $\approx 2.8\,\mu\text{m}$ (Dynabeads M280) diameter streptavidin-coated paramagnetic beads (Invitrogen). Unbound beads were washed away by flowing through excess buffer.

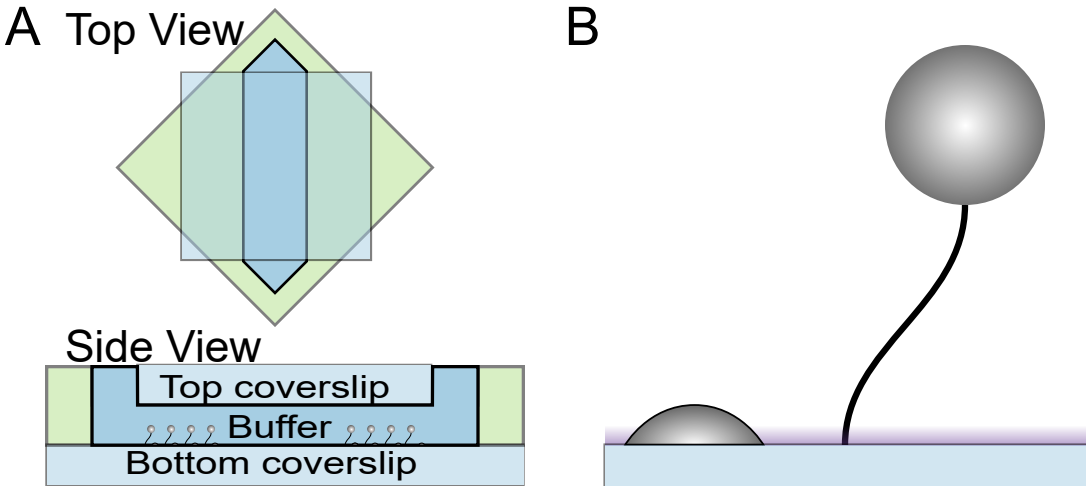

Figure B.3: Schematic of flow cell. (A) Top (upper) and Side (lower) view of the flow cell. The inlet and outlet of the flow cell allows buffer to be exchanged between experiments. (B) Inside the flow cell, heterobifunctional polymers are tethered between a functionalized paramagnetic bead and PEGylated glass coverslip. Reference beads are adhered to the glass surface and tracked to remove common mode noise during data analysis.

B.4 Flow cell construction

The flow cell is assembled as follows: 1 µL to 2 µL of nonmagnetic 2.5 µm diameter polystyrene beads in ≈ 100% isopropyl alcohol or ethanol is spread across the bottom coverslip. The solution is allowed to evaporate, depositing the beads on the coverslip. The beads are adhered to the coverslip by partially melting them for 1.5 min at 100 °C. A 22 mm^2 parafilm square with a diagonally oriented channel is cut with a CO_2 Laser Cutter (Full Spectrum Laser) or cutting machine (Cricut Maker) and placed onto the bottom coverslip (Fig. B.3). A top coverslip (18 mm^2) is plasma cleaned to make it hydrophilic, then treated on one side with a hydrophobic coating (Rain-X, ITW Global Brands). The top coverslip is placed onto the parafilm layer with the hydrophobic surface facing up, creating a central channel with approximate L×W×H dimensions 18 mm×3 mm×100 µm. The central parafilm layer is melted through the glass coverslips with a soldering iron to seal the flow cell.

Appendix C

Supplementary Information

C.1 Supplementary Information for Chapter 3

Determining downturn force

The force below which gamma decreases due to surface interactions, f^*, was determined by analysis of γ vs. force curves (see Fig. 3.3D and Fig. C.1B). In the case of PEG, γ typically reaches a relatively stable plateau value in the vicinity of 0.7, then decreases below 1 pN (Fig. 3.3D). f^* is estimated from the first force value that has $\gamma < 2/3$, and that has decreasing γ values for each of the two adjacent lower-force points. The latter condition reduces the effect of noise.

In the case of HA, a plateau in γ was not typically observed (Fig. C.1B), likely due to some combination of the larger persistence length and polyelectrolyte effects. Instead, as force decreases from a high value, γ first increases (corresponding to increasing chain compliance as the extension recedes from the contour length), then decreases (due to the surface effect). Thus, we estimate f^* simply as the force just below which γ is maximum, though again, to reduce the effect of noise, we require the next two lower-force points to

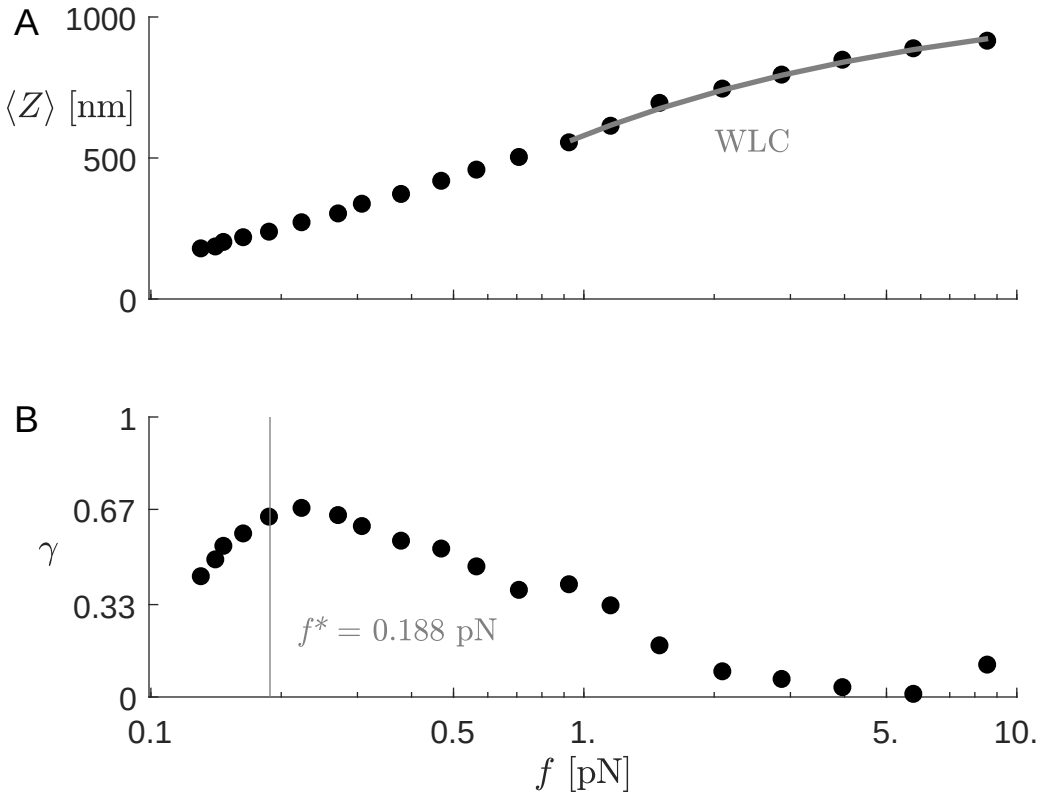

Figure C.1: Representative experimental measurement on a tethered HA chain. (A) Black dots: absolute extension, $\langle Z \rangle$, vs. f. The gray line is a fit of the WLC elastic function to the high force ($f > 0.9$ pN) data; fit shown returns $l_p = 4.6$ nm and $L_0 = 1100$ nm. (B) Full curve of γ vs. f for the same HA tether, showing a downturn for forces below $f^* = 0.188$ pN.

show a continuous decrease in γ.

C.2 Supplementary Information for Chapter 4

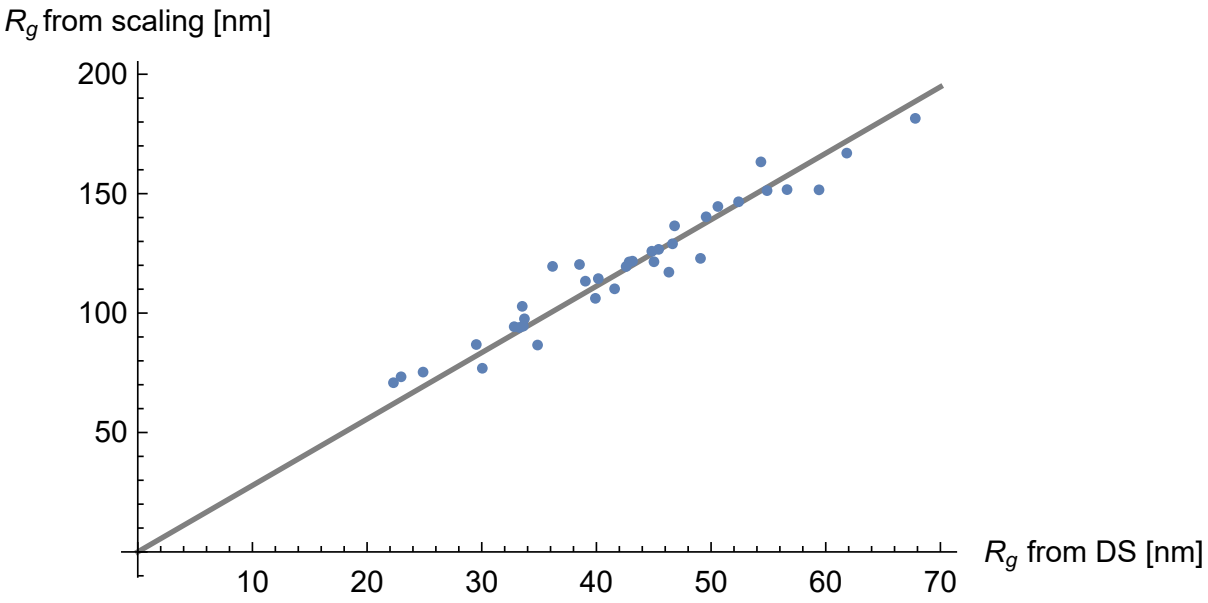

Figure C.2: Correlation of two different R_g calculations of PEG. For each PEG chain measured, WLC fits in the high-force regime give estimates of contour length (L_0) and kuhn length ($\ell = 2l_p$). The "R_g from DS" value (x-axis) plots the result of the Devanand and Selser phenomenological formula for PEG R_g [130]; this estimate inputs only L_0 (in nanometers), which is converted to molecular weight as $M_w = 44(L_0/0.278)$, then used in their formula. The "R_g from scaling" value uses both L_0 and ℓ, setting $R_g = \ell(L_0/\ell)^{0.588}$. The results are tightly correlated, and well–fit by a line that passes through the origin with slope 2.78 ± 0.03. This correlation supports our ability to estimate chain extent, up to a numerical prefactor, from the WLC fits.

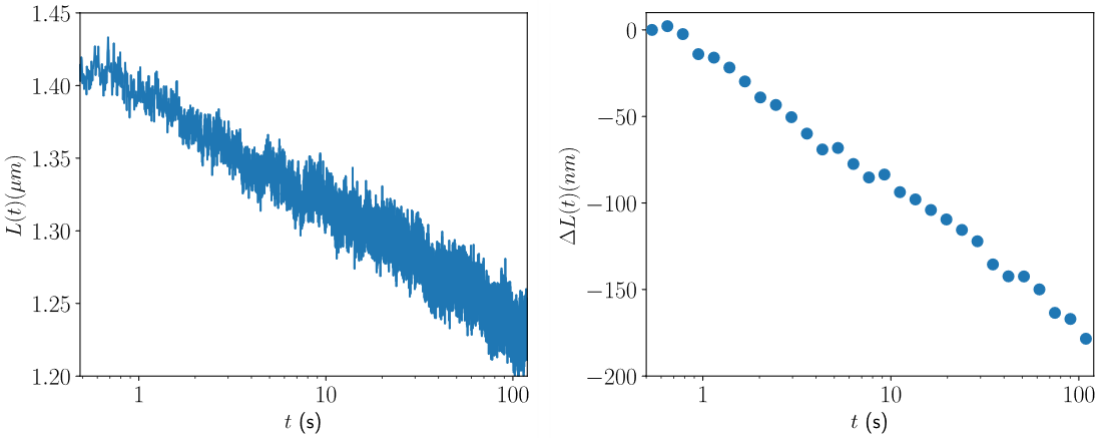

Figure C.3: Force-quench experiment on NFL polymers (left) before and (right) after logarithmic binning. Before logarithmic binning the relaxation does not show any observable discrete steps.

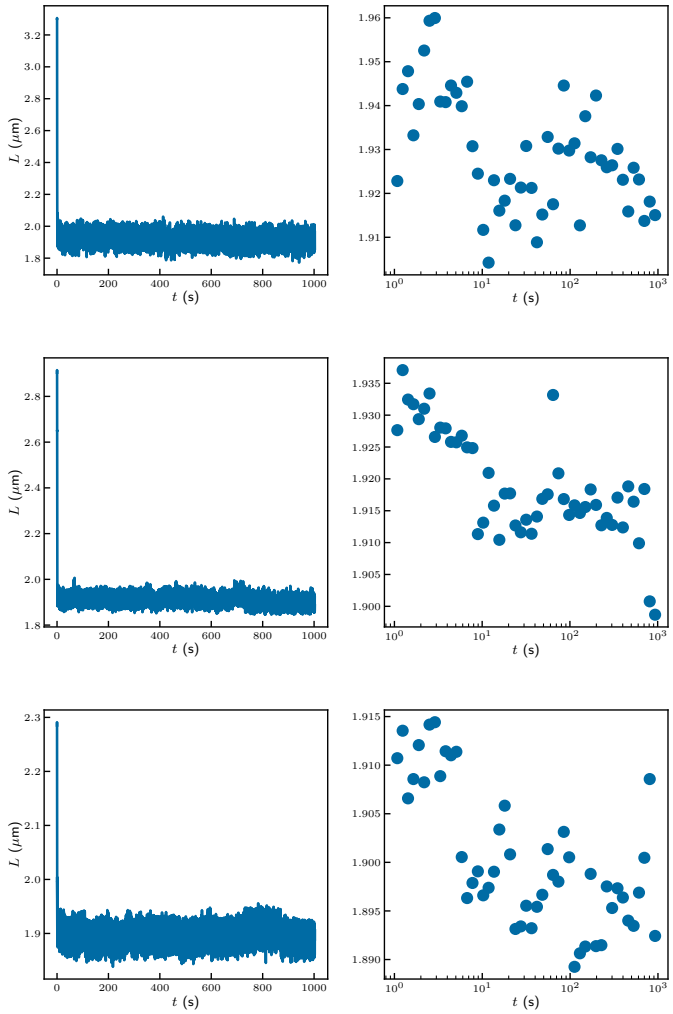

Figure C.4: Control force-quench experiments with double-stranded DNA demonstrating long-term instrument stability. The experiments were carried out at $T = 20\,^{\circ}\mathrm{C}$, in a pH 7 buffer containing 20 mM 2-(N-morpholino)ethanesulfonic acid (MES), 10 mM NaCl, and 0.05% Tween-20. After force-quench, from $f_1 = 26$, 41, and 32 pN to $f_2 = 2$, 4, and 3 pN, respectively (top to bottom), all traces show a small relaxation of $\lesssim 20\,\mathrm{nm}$ over the initial 10 s; we attribute this to re-hybridization dynamics of short AT rich regions occurring after the DNA structure is modified by the overstretching transition [62]. Note that, while the midpoint of the overstretching transition for a long DNA molecule in high salt is 65 pN, prior work has shown that this force is significantly lowered within AT-rich sequences and at low ionic strengths [180, 181]. However, after 10s, the DNA extension remained stable to within 5 nm for hundreds to thousands of seconds, confirming that the slow relaxations analyzed in the main text are due to biopolymer behavior rather than instrumental drift.

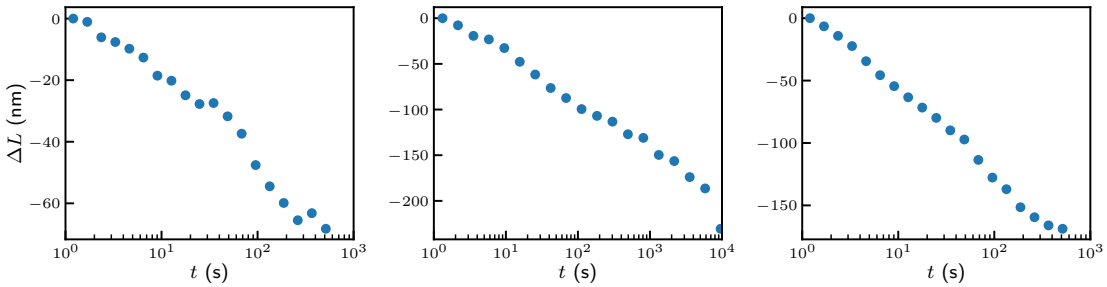

Figure C.5: Long-term logarithmic relaxations observed for three different disordered protein constructs. The relaxations proceeded for three to four decades in time. Ultimately, observing longer relaxations was limited by the tendency of the tethers to break after long times at moderately high force. From left to right, the force-quench values are $f_1 = 78$, 47, and 88 pN and $f_2 = 50$, 4.1, and 40.8 pN. Data points and error bars are the mean and standard error of the mean after logarithmic binning in time (error bars are smaller than points).

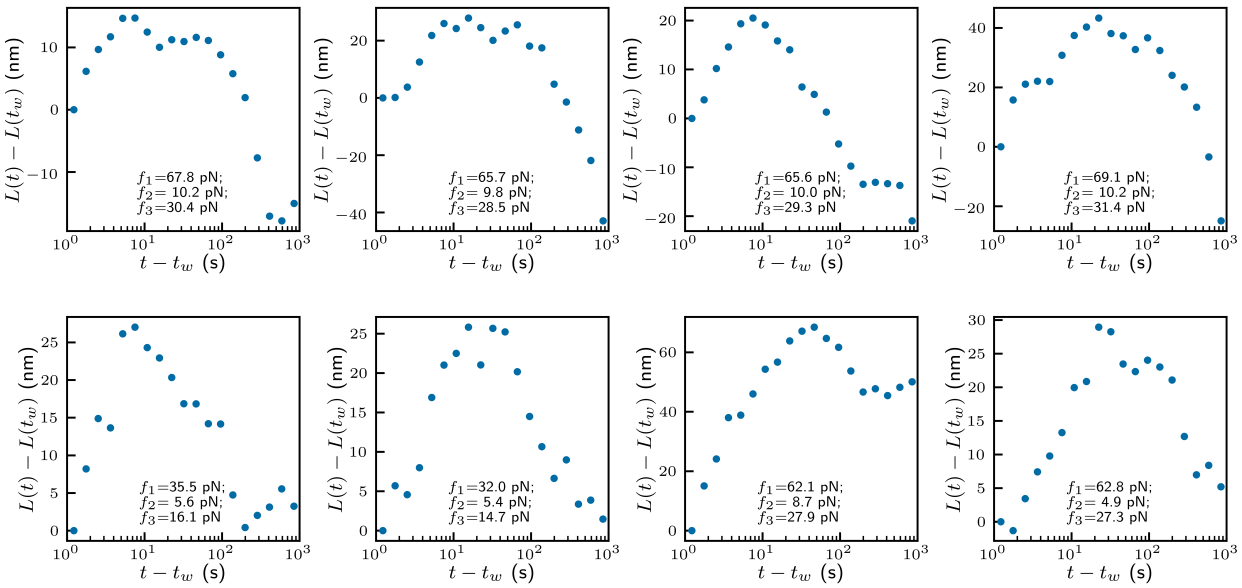

Figure C.6: Examples of the two-step experiment on eight different disordered protein constructs. All measurements used a 10 s waiting time, with force values as noted. The Kovacs hump [152] is clearly visible in all the traces. Data points and error bars are the mean and standard error of the mean after logarithmic binning in time (error bars are smaller than points).

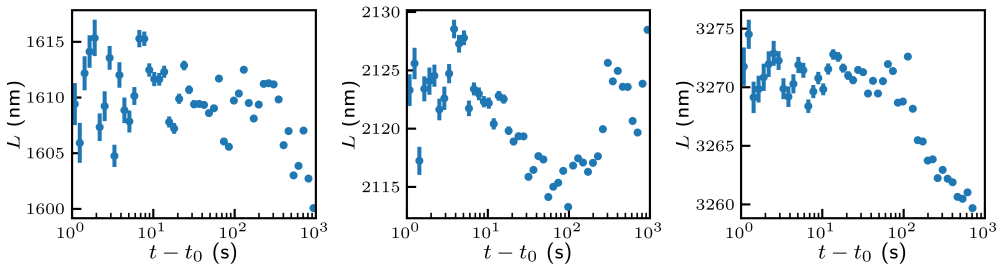

Figure C.7: Control two-step experiments on three different double-stranded DNA molecules. The forces f_1, f_2, f_3 for each molecule were (Left) $23.3, 2.9, 7.3$ pN, (Center) $27.0, 3.1, 11.3$ pN, and (Right) $46.3, 5.9, 12.4$ pN. The control DNA molecules do not show a kovacs hump. Data points and error bars are the mean and standard error of the mean after logarithmic binning in time.

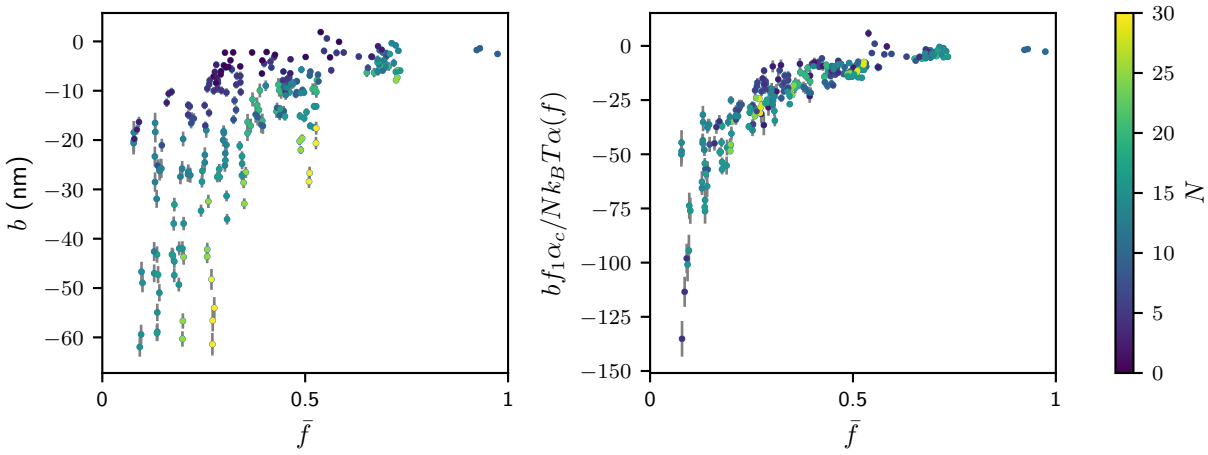

Figure C.8: Logarithmic slopes, b, from least-squares fits to 248 measured relaxations from 16 polymers (left) collapse onto a single curve after normalizing to $\bar{b}$ (right). Data points and error bars are the best-fit value and associated error. Color and color bar indicates the number of tails in each polymer.

C.2.1 Fitting

For each relaxation, we used a least-squares fit to determine the best-fit value and fitting error for the logarithmic slope, b. The raw b values without normalization are shown in Fig. C.8(a), alongside a plot of the fully-normalized $\bar{b}$ (Fig. C.8(b), which repeats the data shown in the Fig. 3 of the main text). Fig. C.8 also shows information on the contour length of the polymers, emphasizing the success of our normalization scheme in collapsing data over a wide range of contour lengths.

To determine the error in the parameter, ρ, we performed 10,000 bootstrapping iterations, in which each iteration consisted of a least-squares fit to Eq. 3 of a randomly-resampled $\bar{b}$ vs. $\bar{f}$ data-set, with the fit weighted by $1/\sigma_{\bar{b}}$. Random resampling was carried out assuming each polymer was an independent sample (not each data point from each polymer), as there could be correlated behavior between the data points from a single polymer, which would invalidate the assumptions of the bootstrapping algorithm [182]. This procedure generated 10,000 estimates of ρ; the reported value and

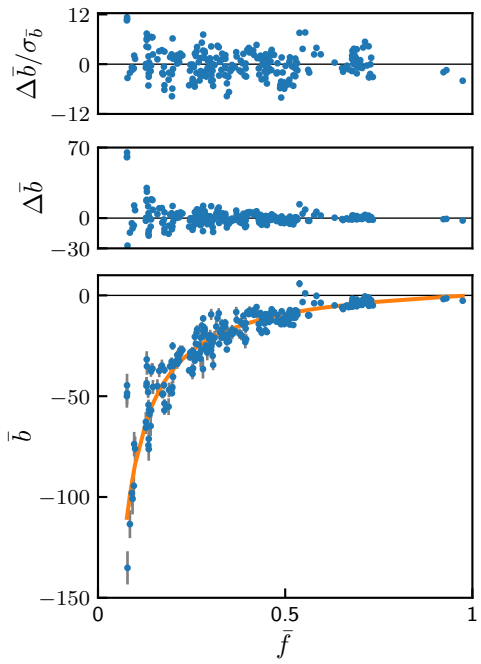

Figure C.9: Fig. 3 with (top) standardized and (center) absolute residuals. The absolute residuals show that as $\bar{f} \to 0$ (i.e. for larger force-quenches) the variance in $\bar{b}$ increases. However, the standardized residuals show no systematic deviation.

error, $\rho = 0.108 \pm 0.004$, represents the mean and standard deviations of those 10,000 estimates. We used the mean value of ρ to plot the standardized residuals in Fig. 3 and Fig. C.9.

C.2.2 Extrapolating to zero force

We wish to estimate the free energy between the structured and coiled state of a segment at zero force, $\Delta G(f = 0) = G_c(0) - G_s(0)$, from the observation that the amplitude of the extension change is around 1 nm, and that compaction proceeds against 30 pN of hindering force (in fact, we observe compaction even against slightly larger forces, but for relatively few chains; 30 pN represents a force where we have observed compaction for most of the chains).

The question of how to extrapolate the energetics of structure formation at $f > 0$ to $f = 0$ has been previously answered [183, 184], and we use those results here. Particularly, the present experiments are done under conditions of constant force ('isotensional'), in which case the effect of force on free energy of a single state can be written:

$$G(f) = G(0) + \Phi(f) \tag{C.1}$$

$$\Phi(f) = -\int_0^f x(f')df' \tag{C.2}$$

where $x(f')$ represents the elastic function (extension vs. force) for the state in question. Thus the free energy difference between coil and structured states at zero force is:

$$\Delta G(0) = \Delta G(f) + \Delta\Delta G(f) \tag{C.3}$$

$$\Delta\Delta G(f) = \int_0^f \alpha(f')df' \tag{C.4}$$

where the amplitude of length change is the difference of extensions of the two states, $\alpha(f) \equiv x_c(f) - x_s(f)$. Assuming the segments just barely form structures at $f = 30$ pN, then the free energy change at that force is not much bigger than zero, $\Delta G(f = 30\text{pN}) \gtrsim$

0, and the zero-force free energy change is

$$\Delta G(0) \gtrsim \Delta\Delta G(f) \tag{C.5}$$

Thus, we expect $\Delta G(0)$ to be not much bigger than the integral of α up to 30 pN (Eq. C.4).

To carry out this calculation, the major question is what elastic functions to use for the coil and structured states. The choice for the coil state, $x_c(f)$, is relatively clear: as discussed in the main text, and based on prior literature on unstructured polypeptides, we treat the coil state as a worm-like chain governed by the Marko-Siggia elastic function, with a persistence length $l_p = 0.8$ nm [157]. The contour length, ℓ_c, is unknown, but the results are not strongly dependent on the absolute magnitude of that parameter, so we choose a reasonable value $\ell_c = 3$ nm for all calculations that follow (Table C.1).

The choice of the structured-state elasticity function is more ambiguous, as we do not know the actual structure. However, it is sensible to assume that, because of the presence of intra-segment binding contacts, the structured state is more rigid than the coil state. Thus, as an approximation, we assume that the structured state is perfectly rigid, such that $x_s(f)$ is governed by the Langevin elasticity function. In particular, we assume the structured state acts like a single Kuhn segment, of length ℓ_s, of a freely-jointed chain, and thus the structured elastic function is

$$x_s(f) = \ell_s \left[\coth\left(\frac{f\ell_s}{k_B T}\right) - \frac{k_B T}{f\ell_s} \right] \tag{C.6}$$

We then are left with one final parameter choice: the length, ℓ_s, of the structured state. Two issues constrain this choice: First, that the extension change α should be around 1 nm. Second, we require that the extension always decreases upon structure formation

Approx. # residues	9	9	9
Coil contour length, ℓ_c [nm]	3	3	3
Structured contour length, ℓ_s [nm]	1.1	1.5	2.0
Coil extension, x_c(30 pN) [nm]	2.4	2.4	2.4
Structured extension, x_s(30 pN) [nm]	0.9	1.4	1.8
α(30 pN) [nm]	1.5	1.0	0.5
$\Delta\Delta$G(30 pN) [k_BT]	8.9	6.0	2.7
Best-fit ρ	0.112 ± 0.005	0.099 ± 0.004	0.066 ± 0.002
χ^2	11.9	11.0	11.5

Table C.1: Parameters used in estimating the elastic response of the coiled and structured states, along with the results of the integral, Eq. C.3, used to estimate zero-force stability for each parameter set. Also listed are the best-fit ρ and goodness-of-fit parameter, $\bar{\chi}^2$, found after normalizing b by elastic functions found from each parameter set (Eq. C.7), and fitting the resulting $\bar{b}$ vs. $\bar{f}$ curves (see Fig. C.10).

for the relevant range of forces– that is, $\alpha(f) = x_c(f) - x_s(f) > 0$ for $1 \leq f \leq 50$ pN. We know this must be true because in our experiments the chain always decreases in extension after the force quench, for all final forces. We enforce these constraints by appropriate choice of ℓ_s; the values used are shown in Table C.1.

Using these assumptions and estimates, we can then carry out the integration in Eq. C.4, which results in the $\Delta\Delta G(f = 30$ pN) values listed in Table C.1. Following Eq. C.5, these values act as rough estimates of $\Delta G(f = 0)$, and vary between 3 and 9 k_BT, as mentioned in the main text. These values can roughly be understood as the difference between the work against the applied force ($f\alpha \approx 7.4 k_BT$ at 30 pN, given a 1 nm amplitude), and the relative loss of entropy between the two states upon being stretched. The entropy loss is larger for the coil state, thus the final $\Delta G(f = 0)$ value is less than the work term.

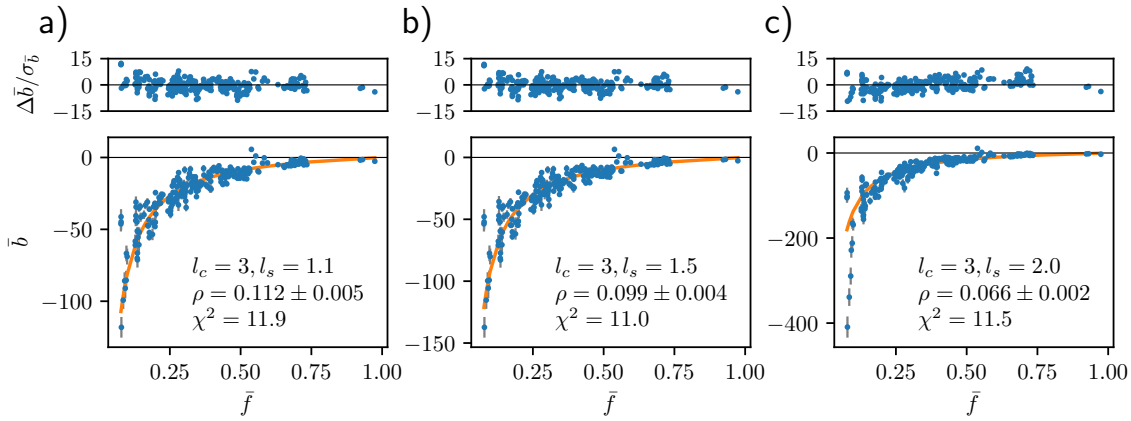

Figure C.10: $\bar{b}$ calculated by normalizing by a function that accounts for elasticity of both structured and coil states (Eq. C.6), and using the parameters of Table C.1. Results of fitting to Eq. 3 are shown. In (a,b), the value of ρ does not change significantly compared to that quoted in the main text. The parameters used in (c) lead to a slight decrease in ρ, though we note that systematic deviations in the residues also begin to appear, likely indicating this parameter set is imperfect.

C.2.3 Normalizing by more nuanced elastic models

We can use the estimates of elasticity of both structured and coiled state, as described in Table C.1, to calculate the compaction length and normalize $\bar{b}$. That is, in the main text we normalize b only by the worm-like chain relative extension function, α_0, corresponding to the dominant elastic effect of the coil state. Here, we extend this treatment to consider the elasticity of the structured state. We particularly normalize by the following relative elasticity function, as defined in the prior section:

$$\alpha_0' = \frac{x_c(f) - x_s(f)}{\ell} \tag{C.7}$$

where the denominator consists of the different in maximum extensions (i.e. the contour lengths) of the coil and structured states, i.e., $\ell = \ell_c - \ell_s$. We then recalculate $\bar{b}$, now normalizing by α_0' in place of α_0, for the different parameter choices shown in Table C.1.

The best-fit ρ, and goodness-of-fit, are shown for each parameter choice in that same table; the fits themselves are shown in [Fig. C.10]. The variation in ρ is relatively small, and all of fits have similar reduced-chi-squared metrics. This shows that our main-text conclusions regarding the maximum number of relaxation events per tail are robust against choice of elasticity parameters.

Appendix D

Abbreviations

AFM Atomic Force Microscope. 7

AV Allan variance. 14

CD Circular Dichroism. 11

DBCO Dibenzocyclooctyne. 109

DMSO Dimethyl sulfoxide. 109

dsDNA Double-stranded DNA. 107

DTT Dithiothreitol. 106

FRET Förster Resonance Energy Transfer. 5

HA Hyaluronic Acid. 42, 107

IDP Intrinsically Disordered Protein. 4

IDR Intrinsically Disordered Region. 4, 91

IPTG Isopropyl b-D-1-thiogalactopyranoside. 104

MT Magnetic Tweezer. 7, 95, 98

NFLt Neurofilament light tail. 10, 92, 103

NHS N-Hydroxysuccinimide. 99

OT Optical Tweezer. 7

PAGE Polyacrylamide Gel Electrophoresis. 105

PCR Polymerase Chain Reaction. 107

PDB Protein Data Bank. 5

PEG Polyethylene glycol. 42, 99, 107

SAXS Small Angle X-ray Scattering. 4

SDS Sodium Dodecyl Sulfate. 105

SMFS Single-Molecule Force Spectroscopy. 7, 91

WLC Worm-like Chain. 49, 50

Bibliography

[1] T. Ehm, H. Shinar, S. Meir, A. Sekhon, V. Sethi, I. L. Morgan, G. Rahamim, O. A. Saleh, and R. Beck, *Intrinsically disordered proteins at the nano-scale*, *Nano Futures* **5** (June, 2021) 022501.

[2] P. W. Anderson, *Through the Glass Lightly*, *Science* **267** (Mar., 1995) 1615–1616.

[3] H. Jensen, *Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems*. Cambridge Lecture Notes in Physics. Cambridge University Press, 1998.

[4] D. U. Ferreiro, E. A. Komives, and P. G. Wolynes, *Frustration in biomolecules*, *Quarterly Reviews of Biophysics* **47** (Nov., 2014) 285–363.

[5] R. H. Austin, K. Beeson, L. Eisenstein, H. Frauenfelder, I. C. Gunsalus, and V. P. Marshall, *Dynamics of Carbon Monoxide Binding by Heme Proteins*, *Science* **181** (Aug., 1973) 541–543.

[6] H. Frauenfelder, F. Parak, and R. D. Young, *Conformational Substates in Proteins*, *Annual Review of Biophysics and Biophysical Chemistry* **17** (1988), no. 1 451–479.

[7] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, *The energy landscapes and motions of proteins*, *Science* **254** (Dec., 1991) 1598–1603.

[8] M. Karplus, Y. Q. Gao, J. Ma, A. van der Vaart, and W. Yang, *Protein structural transitions and their functional role*, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **363** (Feb., 2005) 331–356.

[9] J. D. Bryngelson and P. G. Wolynes, *Spin glasses and the statistical mechanics of protein folding*, *Proceedings of the National Academy of Sciences* **84** (Nov., 1987) 7524–7528.

[10] N. Go, *Theoretical Studies of Protein Folding*, *Annual Review of Biophysics and Bioengineering* **12** (1983), no. 1 183–210.

[11] V. N. Uversky, *Natively unfolded proteins: A point where biology waits for physics*, *Protein Science* **11** (2002), no. 4 739–756.

[12] V. N. Uversky and A. K. Dunker, *Understanding protein non-folding*, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1804** (June, 2010) 1231–1264.

[13] V. N. Uversky, *Introduction to Intrinsically Disordered Proteins (IDPs)*, *Chemical Reviews* **114** (July, 2014) 6557–6560.

[14] V. N. Uversky, *Intrinsically Disordered Proteins and Their "Mysterious" (Meta)Physics*, *Frontiers in Physics* **7** (2019).

[15] P. E. Wright and H. J. Dyson, *Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm*, *Journal of Molecular Biology* **293** (Oct., 1999) 321–331.

[16] P. Romero, Z. Obradovic, C. R. Kissinger, J. E. Villafranca, E. Garner, S. Guilliot, and A. K. Dunker, *Thousands of proteins likely to have long disordered regions*, *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* (1998) 437–448.

[17] E. Garner, P. Cannon, P. Romero, Z. Obradovic, and A. K. Dunker, *Predicting Disordered Regions from Amino Acid Sequence*, *Genome Informatics* **9** (1998) 201–213.

[18] F. Ferron, S. Longhi, B. Canard, and D. Karlin, *A practical overview of protein disorder prediction methods*, *Proteins: Structure, Function, and Bioinformatics* **65** (2006), no. 1 1–14.

[19] V. N. Uversky, J. R. Gillespie, and A. L. Fink, *Why are "natively unfolded" proteins unstructured under physiologic conditions?*, *Proteins: Structure, Function, and Bioinformatics* **41** (2000), no. 3 415–427.

[20] P. Radivojac, L. M. Iakoucheva, C. J. Oldfield, Z. Obradovic, V. N. Uversky, and A. K. Dunker, *Intrinsic Disorder and Functional Proteomics*, *Biophysical Journal* **92** (Mar., 2007) 1439–1456.

[21] R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright, and M. M. Babu, *Classification of Intrinsically Disordered Regions and Proteins*, *Chemical Reviews* **114** (July, 2014) 6589–6631.

[22] S. Mukhopadhyay, R. Krishnan, E. A. Lemke, S. Lindquist, and A. A. Deniz, *A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures*, *Proceedings of the National Academy of Sciences* **104** (Feb., 2007) 2649–2654.

[23] H. T. Tran, A. Mao, and R. V. Pappu, *Role of Backbone-Solvent Interactions in Determining Conformational Equilibria of Intrinsically Disordered Proteins*, *Journal of the American Chemical Society* **130** (June, 2008) 7380–7392.

[24] R. K. Das and R. V. Pappu, *Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues*, *Proceedings of the National Academy of Sciences* **110** (Aug., 2013) 13392–13397.

[25] J. Kyte and R. F. Doolittle, *A simple method for displaying the hydropathic character of a protein*, *Journal of Molecular Biology* **157** (May, 1982) 105–132.

[26] T. N. Cordeiro, F. Herranz-Trillo, A. Urbanek, A. Estaña, J. Cortés, N. Sibille, and P. Bernadó, *Small-angle scattering studies of intrinsically disordered proteins and their complexes*, *Current Opinion in Structural Biology* **42** (Feb., 2017) 15–23.

[27] B. Schuler, A. Soranno, H. Hofmann, and D. Nettels, *Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins*, *Annual Review of Biophysics* **45** (July, 2016) 207–231.

[28] J. A. Marsh and J. D. Forman-Kay, *Sequence Determinants of Compaction in Intrinsically Disordered Proteins*, *Biophysical Journal* **98** (May, 2010) 2383–2390.

[29] K. Sugase, H. J. Dyson, and P. E. Wright, *Mechanism of coupled folding and binding of an intrinsically disordered protein*, *Nature* **447** (June, 2007) 1021–1025.

[30] A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, and Z. Obradovic, *Intrinsically disordered protein*, *Journal of Molecular Graphics and Modelling* **19** (Feb., 2001) 26–59.

[31] V. N. Uversky, *Intrinsically disordered proteins from A to Z*, *The International Journal of Biochemistry & Cell Biology* **43** (Aug., 2011) 1090–1103.

[32] C. R. Safinya, J. Deek, R. Beck, J. B. Jones, C. Leal, K. K. Ewert, and Y. Li, *Liquid crystal assemblies in biologically inspired systems*, *Liquid Crystals* **40** (Dec., 2013) 1748–1758.

[33] H. J. Dyson and P. E. Wright, *Intrinsically unstructured proteins and their functions*, *Nature Reviews Molecular Cell Biology* **6** (Mar., 2005) 197–208.

[34] Y. Shin and C. P. Brangwynne, *Liquid phase condensation in cell physiology and disease*, *Science* **357** (Sept., 2017).

[35] M. Guharoy, B. Szabo, S. C. Martos, S. Kosol, and P. Tompa, *Intrinsic Structural Disorder in Cytoskeletal Proteins*, *Cytoskeleton* **70** (2013), no. 10 550–571.

[36] M. Kornreich, R. Avinery, E. Malka-Gibor, A. Laser-Azogui, and R. Beck, *Order and disorder in intermediate filament proteins*, *FEBS Letters* **589** (2015), no. 19PartA 2464–2476.

[37] F. Chiti and C. M. Dobson, *Protein Misfolding, Functional Amyloid, and Human Disease*, *Annual Review of Biochemistry* **75** (June, 2006) 333–366.

[38] J. Cubuk, J. J. Alston, J. J. Incicco, S. Singh, M. D. Stuchell-Brereton, M. D. Ward, M. I. Zimmerman, N. Vithani, D. Griffith, J. A. Wagoner, G. R. Bowman, K. B. Hall, A. Soranno, and A. S. Holehouse, *The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA*, *Nature Communications* **12** (Mar., 2021) 1936.

[39] K. C. Neuman and A. Nagy, *Single-molecule force spectroscopy: Optical tweezers, magnetic tweezers and atomic force microscopy*, *Nature Methods* **5** (June, 2008) 491–505.

[40] O. A. Saleh, *Perspective: Single polymer mechanics across the force regimes*, *The Journal of Chemical Physics* **142** (May, 2015) 194902.

[41] P. Pincus, *Excluded Volume Effects and Stretched Polymer Chains*, *Macromolecules* **9** (May, 1976) 386–388.

[42] C. Bustamante, L. Alexander, K. Maciuba, and C. M. Kaiser, *Single-Molecule Studies of Protein Folding with Optical Tweezers*, *Annual Review of Biochemistry* **89** (June, 2020) 443–470.

[43] H. A. Kramers, *Brownian motion in a field of force and the diffusion model of chemical reactions*, *Physica* **7** (Apr., 1940) 284–304.

[44] C. Bustamante, Y. R. Chemla, N. R. Forde, and D. Izhaky, *Mechanical Processes in Biochemistry*, *Annual Review of Biochemistry* **73** (June, 2004) 705–748.

[45] S. N. Zhurkov, *Kinetic Concept of the Strength of Solids*, *International journal of fracture mechanics* **1** (Dec., 1965) 311–323.

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

[47] P. E. Rouse, *A Theory of the Linear Viscoelastic Properties of Dilute Solutions of Coiling Polymers*, *The Journal of Chemical Physics* **21** (July, 1953) 1272–1280.

[48] P. Pincus, *Dynamics of Stretched Polymer Chains*, *Macromolecules* **10** (Jan., 1977) 210–213.

[49] M. Schlierf, H. Li, and J. M. Fernandez, *The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques*, *Proceedings of the National Academy of Sciences* **101** (May, 2004) 7299–7304.

[50] R. Hervás, J. Oroz, A. Galera-Prat, O. Goñi, A. Valbuena, A. M. Vera, À. Gómez-Sicilia, F. Losada-Urzáiz, V. N. Uversky, M. Menéndez, D. V. Laurents, M. Bruix, and M. Carrión-Vázquez, *Common Features at the Start of the Neurodegeneration Cascade*, *PLOS Biology* **10** (May, 2012) e1001335.

[51] A. Solanki, K. Neupane, and M. T. Woodside, *Single-Molecule Force Spectroscopy of Rapidly Fluctuating, Marginally Stable Structures in the Intrinsically Disordered Protein $\ensuremath{\alpha}$-Synuclein*, *Physical Review Letters* **112** (Apr., 2014) 158103.

[52] R. Beck, J. Deek, J. B. Jones, and C. R. Safinya, *Gel-expanded to gel-condensed transition in neurofilament networks revealed by direct force measurements*, *Nature Materials* **9** (Jan., 2010) 40–46.

[53] P. J. Chung, C. Song, J. Deek, H. P. Miller, Y. Li, M. C. Choi, L. Wilson, S. C. Feinstein, and C. R. Safinya, *Tau mediates microtubule bundle architectures mimicking fascicles of microtubules found in the axon initial segment*, *Nature Communications* **7** (July, 2016) 12278.

[54] E. Malka-Gibor, M. Kornreich, A. Laser-Azogui, O. Doron, I. Zingerman-Koladko, J. Harapin, O. Medalia, and R. Beck, *Phosphorylation-Induced Mechanical Regulation of Intrinsically Disordered Neurofilament Proteins*, *Biophysical Journal* **112** (Mar., 2017) 892–900.

[55] A. Yuan, M. V. Rao, V. , and R. A. Nixon, *Neurofilaments at a glance*, *Journal of Cell Science* **125** (July, 2012) 3257–3263.

[56] M. Kornreich, E. Malka-Gibor, B. Zuker, A. Laser-Azogui, and R. Beck, *Neurofilaments Function as Shock Absorbers: Compression Response Arising from Disordered Proteins*, *Physical Review Letters* **117** (Sept., 2016) 148101.

[57] N. Geisler, E. Kaufmann, S. Fischer, U. Plessmann, and K. Weber, *Neurofilament architecture combines structural principles of intermediate filaments with carboxy-terminal extensions increasing in size between triplet proteins.*, *The EMBO Journal* **2** (Aug., 1983) 1295–1302.

[58] E. Ostrofet, F. S. Papini, and D. Dulin, *Correction-free force calibration for magnetic tweezers experiments*, *Scientific Reports* **8** (Oct., 2018) 15920.

[59] C. J. Bustamante, Y. R. Chemla, S. Liu, and M. D. Wang, *Optical tweezers in single-molecule biophysics*, *Nature Reviews Methods Primers* **1** (Mar., 2021) 1–29.

[60] A. Amir, Y. Oreg, and Y. Imry, *On relaxations and aging of various glasses*, *Proceedings of the National Academy of Sciences* **109** (Feb., 2012) 1850–1855.

[61] B. M. Lansdorp and O. A. Saleh, *Power spectrum and Allan variance methods for calibrating single-molecule video-tracking instruments*, *Review of Scientific Instruments* **83** (Feb., 2012) 025115.

[62] S. B. Smith, L. Finzi, and C. Bustamante, *Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads*, *Science* **258** (Nov., 1992) 1122–1126.

[63] T. T. Perkins, S. R. Quake, D. E. Smith, and S. Chu, *Relaxation of a Single DNA Molecule Observed by Optical Microscopy*, *Science* **264** (1994), no. 5160 822–826.

[64] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block, *Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins*, *Proceedings of the National Academy of Sciences* **103** (Apr., 2006) 6190–6195.

[65] H. Yu, X. Liu, K. Neupane, A. N. Gupta, A. M. Brigley, A. Solanki, I. Sosova, and M. T. Woodside, *Direct observation of multiple misfolding pathways in a single prion protein molecule*, *Proceedings of the National Academy of Sciences* **109** (Apr., 2012) 5283–5288.

[66] T. R. Strick, V. Croquette, and D. Bensimon, *Single-molecule analysis of DNA uncoiling by a type II topoisomerase*, *Nature* **404** (Apr., 2000) 901–904.

[67] E. A. Abbondanzieri, W. J. Greenleaf, J. W. Shaevitz, R. Landick, and S. M. Block, *Direct observation of base-pair stepping by RNA polymerase*, *Nature* **438** (Nov., 2005) 460–465.

[68] X. Zhuang and M. Rief, *Single-molecule folding*, *Current Opinion in Structural Biology* **13** (Feb., 2003) 88–97.

[69] E.-L. Florin, A. Pralle, E. Stelzer, and J. Hörber, *Photonic force microscope calibration by thermal noise analysis*, *Applied Physics A* **66** (Mar., 1998) S75–S78.

[70] B. Lukić, S. Jeney, C. Tischer, A. J. Kulik, L. Forró, and E.-L. Florin, *Direct Observation of Nondiffusive Motion of a Brownian Particle*, *Physical Review Letters* **95** (Oct., 2005) 160601.

[71] K. Berg-Sørensen and H. Flyvbjerg, *Power spectrum analysis for optical tweezers*, *Review of Scientific Instruments* **75** (Feb., 2004) 594–612.

[72] W. P. Wong and K. Halvorsen, *The effect of integration time on fluctuation measurements: Calibrating an optical trap in the presence of motion blur*, *Optics Express* **14** (Dec., 2006) 12517–12531.

[73] K. C. Neuman and S. M. Block, *Optical trapping*, *Review of Scientific Instruments* **75** (Sept., 2004) 2787–2809.

[74] S. F. Nørrelykke and H. Flyvbjerg, *Power spectrum analysis with least-squares fitting: Amplitude bias and its elimination, with application to optical tweezers and atomic force microscope cantilevers*, *Review of Scientific Instruments* **81** (July, 2010) 075103.

[75] I. M. Tolić-Nørrelykke, K. Berg-Sørensen, and H. Flyvbjerg, *MatLab program for precision calibration of optical tweezers*, *Computer Physics Communications* **159** (June, 2004) 225–240.

[76] P. M. Hansen, I. M. Tolic-Nørrelykke, H. Flyvbjerg, and K. Berg-Sørensen, *Tweezercalib 2.1: Faster version of MatLab package for precise calibration of optical tweezers*, *Computer Physics Communications* **175** (Oct., 2006) 572–573.

[77] N. Osterman, *TweezPal – Optical tweezers analysis and calibration software*, *Computer Physics Communications* **181** (Nov., 2010) 1911–1916.

[78] C. D. Taylor, T. W. Foley, A. N. Chang, S. Mowa, J. L. Burris, and B. C. Hester, *Computer-automated program for calibration of optical tweezers*, in *Optics and Photonics for Information Processing VI*, vol. 8498, p. 84980G, International Society for Optics and Photonics, Oct., 2012.

[79] Z. Yu, D. Dulin, J. Cnossen, M. Köber, M. M. van Oene, O. Ordu, B. A. Berghuis, T. Hensgens, J. Lipfert, and N. H. Dekker, *A force calibration standard for magnetic tweezers*, *Review of Scientific Instruments* **85** (Dec., 2014) 123114.

[80] P. Daldrop, H. Brutzer, A. Huhle, D. J. Kauert, and R. Seidel, *Extending the Range for Force Calibration in Magnetic Tweezers*, *Biophysical Journal* **108** (May, 2015) 2550–2561.

[81] D. C. Ince, L. Hatton, and J. Graham-Cumming, *The case for open computer programs*, *Nature* **482** (Feb., 2012) 485–488.

[82] I. Morgan, *Ianlmorgan/tweezepy*, June, 2021.

[83] I. L. Morgan, "Tweezepy: Single-molecule force spectroscopy calibration."

[84] I. Morgan, "Ianlmorgan/tweezepy: Tweezepy initial release." Zenodo, June, 2021.

[85] H. Shen, *Interactive notebooks: Sharing the code*, *Nature News* **515** (Nov., 2014) 151.

[86] T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, C. Willing, and Jupyter development team, *Jupyter Notebooks – a publishing format for reproducible computational workflows*, in *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (F. Loizides and B. Scmidt, eds.), pp. 87–90, IOS Press, 2016.

[87] H. Felgner, O. Müller, and M. Schliwa, *Calibration of light forces in optical tweezers*, *Applied Optics* **34** (Feb., 1995) 977–982.

[88] E.-L. Florin, A. Pralle, J. K. Heinrich Hörber, and E. H. K. Stelzer, *Photonic Force Microscope Based on Optical Tweezers and Two-Photon Excitation for Biological Applications*, *Journal of Structural Biology* **119** (July, 1997) 202–211.

[89] T. R. Strick, J.-F. Allemand, D. Bensimon, A. Bensimon, and V. Croquette, *The Elasticity of a Single Supercoiled DNA Molecule*, *Science* **271** (Mar., 1996) 1835–1837.

[90] J. L. Hutter and J. Bechhoefer, *Calibration of atomic-force microscope tips*, *Review of Scientific Instruments* **64** (July, 1993) 1868–1873.

[91] K. Berg-Sørensen, L. Oddershede, E.-L. Florin, and H. Flyvbjerg, *Unintended filtering in a typical photodiode detection system for optical tweezers*, *Journal of Applied Physics* **93** (Mar., 2003) 3167–3176.

[92] D. W. Allan, M. A. Weiss, and J. L. Jespersen, *A frequency-domain view of time-domain characterization of clocks and time and frequency distribution systems*, in *Proceedings of the 45th Annual Symposium on Frequency Control 1991*, pp. 667–678, May, 1991.

[93] G. M. Gibson, J. Leach, S. Keen, A. J. Wright, and M. J. Padgett, *Measuring the accuracy of particle position and force in optical tweezers using high-speed video microscopy*, *Optics Express* **16** (Sept., 2008) 14561.

[94] F. Czerwinski, A. C. Richardson, and L. B. Oddershede, *Quantifying Noise in Optical Tweezers by Allan Variance*, *Optics Express* **17** (July, 2009) 13255.

[95] M. Andersson, F. Czerwinski, and L. B. Oddershede, *Optimizing active and passive calibration of optical tweezers*, *Journal of Optics* **13** (Mar., 2011) 044020.

[96] A. Pralle, E.-L. Florin, E. Stelzer, and J. Hörber, *Local viscosity probed by photonic force microscopy*, *Applied Physics A* **66** (Mar., 1998) S71–S73.

[97] A. van der Horst and N. R. Forde, *Power spectral analysis for optical trap stiffness calibration from high-speed camera position detection with limited bandwidth*, *Optics Express* **18** (Apr., 2010) 7670.

[98] N. Ribeck and O. A. Saleh, *Multiplexed single-molecule measurements with magnetic tweezers*, *Review of Scientific Instruments* **79** (2008), no. 9 094301.

[99] W. Riley and D. A. Howe, *Handbook of Frequency Stability Analysis*, .

[100] M. M. van Oene, S. Ha, T. Jager, M. Lee, F. Pedaci, J. Lipfert, and N. H. Dekker, *Quantifying the Precision of Single-Molecule Torque and Twist Measurements Using Allan Variance*, *Biophysical Journal* **114** (Apr., 2018) 1970–1979.

[101] P. Welch, *The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms*, *IEEE Transactions on Audio and Electroacoustics* **15** (June, 1967) 70–73.

[102] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Pearson/Prentice Hall, Upper Saddle River, N.J, 2005.

[103] F. Harris, *On the use of windows for harmonic analysis with the discrete Fourier transform*, *Proceedings of the IEEE* **66** (Jan., 1978) 51–83.

[104] J. W. Cooley and J. W. Tukey, *An algorithm for the machine calculation of complex Fourier series*, *Mathematics of Computation* **19** (1965), no. 90 297–301.

[105] B. M. Lansdorp and O. A. Saleh, *Erratum: "Power spectrum and Allan variance methods for calibrating single-molecule video-tracking instruments" [Rev. Sci. Instrum. 83, 025115 (2012)]*, *Review of Scientific Instruments* **85** (Jan., 2014) 019901.

[106] J. A. Nelder and R. Mead, *A Simplex Method for Function Minimization*, *The Computer Journal* **7** (Jan., 1965) 308–313.

[107] P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, Boston, 3rd ed ed., 2003.

[108] K. P. Burnham and D. R. Anderson, *Multimodel Inference: Understanding AIC and BIC in Model Selection*, *Sociological Methods & Research* **33** (Nov., 2004) 261–304.

[109] C. A. Greenhall and W. J. Riley, *Uncertainty of stability variances based on finite differences*, .

[110] W. J. Riley and C. A. Greenhall, *Power law noise identification using the lag 1 autocorrelation*, in *2004 18th European Frequency and Time Forum (EFTF 2004)*, pp. 576–580, Apr., 2004.

[111] H. Faxén, *Der Widerstand gegen die Bewegung einer starren Kugel in einer zähen Flüssigkeit, die zwischen zwei parallelen ebenen Wänden eingeschlossen ist*, *Annalen der Physik* **373** (1922), no. 10 89–119.

[112] R. D. Neidinger, *Introduction to Automatic Differentiation and MATLAB Object-Oriented Programming*, *SIAM Review* **52** (Jan., 2010) 545–563.

[113] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, *Automatic Differentiation in Machine Learning: A Survey*, *Journal of Machine Learning Research* **18** (2018), no. 153 1–43.

[114] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, *Emcee : The MCMC Hammer*, *Publications of the Astronomical Society of the Pacific* **125** (Mar., 2013) 306–312.

[115] "The Python Standard Library — Python 3.9.5 documentation." https://docs.python.org/3/library/.

[116] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, *Array programming with NumPy*, *Nature* **585** (Sept., 2020) 357–362.

[117] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, and P. van Mulbregt, *SciPy 1.0: Fundamental algorithms for scientific computing in Python*, *Nature Methods* **17** (Mar., 2020) 261–272.

[118] S. K. Lam, A. Pitrou, and S. Seibert, *Numba: A LLVM-based Python JIT compiler*, in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, LLVM '15, (New York, NY, USA), pp. 1–6, Association for Computing Machinery, Nov., 2015.

[119] S. N. Innes-Gold, I. L. Morgan, and O. A. Saleh, *Surface-induced effects in fluctuation-based measurements of single-polymer elasticity: A direct probe of the radius of gyration*, *The Journal of Chemical Physics* **148** (Dec., 2017) 123314.

[120] M. S. Z. Kellermayer, S. B. Smith, H. L. Granzier, and C. Bustamante, *Folding-Unfolding Transitions in Single Titin Molecules Characterized with Laser Tweezers*, *Science* **276** (May, 1997) 1112–1116.

[121] D. R. Jacobson, D. B. McIntosh, M. J. Stevens, M. Rubinstein, and O. A. Saleh, *Single-stranded nucleic acid elasticity arises from internal electrostatic tension*, *Proceedings of the National Academy of Sciences* **114** (May, 2017) 5095–5100.

[122] O. A. Saleh, D. B. McIntosh, P. Pincus, and N. Ribeck, *Nonlinear Low-Force Elasticity of Single-Stranded DNA Molecules*, *Physical Review Letters* **102** (Feb., 2009) 068301.

[123] D. B. McIntosh, G. Duggan, Q. Gouil, and O. A. Saleh, *Sequence-Dependent Elasticity and Electrostatics of Single-Stranded DNA: Signatures of Base-Stacking*, *Biophysical Journal* **106** (Feb., 2014) 659–666.

[124] D. Klaue and R. Seidel, *Torsional Stiffness of Single Superparamagnetic Microspheres in an External Magnetic Field*, *Physical Review Letters* **102** (Jan., 2009) 028302.

[125] S. Liese, M. Gensler, S. Krysiak, R. Schwarzl, A. Achazi, B. Paulus, T. Hugel, J. P. Rabe, and R. R. Netz, *Hydration Effects Turn a Highly Stretched Polymer from an Entropic into an Energetic Spring*, *ACS Nano* **11** (Jan., 2017) 702–712.

[126] L. D. Landau and E. M. Lifsic, *Statistical Physics, Cours of Theoretical Physics*, vol. 5. Elsevier, Amsterdam [u.a, 1980.

[127] R. R. Netz, *Strongly Stretched Semiflexible Extensible Polyelectrolytes and DNA*, *Macromolecules* **34** (Oct., 2001) 7522–7529.

[128] R. M. Neumann, *Polymer stretching in an elongational flow*, *The Journal of Chemical Physics* **110** (Apr., 1999) 7513–7515.

[129] F. Oesterhelt, M. Rief, and H. E. Gaub, *Single molecule force spectroscopy by AFM indicates helical structure of poly(ethylene-glycol) in water*, *New Journal of Physics* **1** (Jan., 1999) 6–6.

[130] K. Devanand and J. C. Selser, *Asymptotic behavior and long-range interactions in aqueous solutions of poly(ethylene oxide)*, *Macromolecules* **24** (Oct., 1991) 5943–5947.

[131] T. Bickel, C. Jeppesen, and C. Marques, *Local entropic effects of polymers grafted to soft interfaces*, *The European Physical Journal E* **4** (Jan., 2001) 33–43.

[132] E. Eisenriegler, *Polymers near Surfaces : Conformation Properties and Relation to Critical Phenomena*. World Scientific, Singapore, 1993.

[133] J. A. Rudnick and G. D. Gaspari, *Elements of the Random Walk : An Introduction for Advanced Students and Researchers*. Cambridge University Press, Cambridge, 2004.

[134] R. Bubis, Y. Kantor, and M. Kardar, *Configurations of polymers attached to probes*, *EPL (Europhysics Letters)* **88** (Nov., 2009) 48001.

[135] J. P. Berezney and O. A. Saleh, *Electrostatic Effects on the Conformation and Elasticity of Hyaluronic Acid, a Moderately Flexible Polyelectrolyte*, *Macromolecules* **50** (Feb., 2017) 1085–1089.

[136] J. F. Marko and E. D. Siggia, *Stretching DNA*, *Macromolecules* **28** (Dec., 1995) 8759–8770.

[137] C. Bouchiat, M. D. Wang, J. F. Allemand, T. Strick, S. M. Block, and V. Croquette, *Estimating the Persistence Length of a Worm-Like Chain Molecule from Force-Extension Measurements*, *Biophysical Journal* **76** (Jan., 1999) 409–413.

[138] A. Almond, A. Brass, and J. K. Sheehan, *Oligosaccharides as Model Systems for Understanding Water-Biopolymer Interaction: Hydrated Dynamics of a Hyaluronan Decamer*, *The Journal of Physical Chemistry B* **104** (June, 2000) 5634–5640.

[139] M. I. Giannotti, M. Rinaudo, and G. J. Vancso, *Force Spectroscopy of Hyaluronan by Atomic Force Microscopy: From Hydrogen-Bonded Networks toward Single-Chain Behavior*, *Biomacromolecules* **8** (Sept., 2007) 2648–2652.

[140] A. V. Dobrynin, J.-M. Y. Carrillo, and M. Rubinstein, *Chains Are More Flexible Under Tension*, *Macromolecules* **43** (Nov., 2010) 9181–9190.

[141] A. Dittmore, D. B. McIntosh, S. Halliday, and O. A. Saleh, *Single-Molecule Elasticity Measurements of the Onset of Excluded Volume in Poly(Ethylene Glycol)*, *Physical Review Letters* **107** (Sept., 2011) 148301.

[142] D. W. Schaefer, J. F. Joanny, and P. Pincus, *Dynamics of Semiflexible Polymers in Solution*, *Macromolecules* **13** (Sept., 1980) 1280–1289.

[143] I. L. Morgan, R. Avinery, G. Rahamim, R. Beck, and O. A. Saleh, *Glassy Dynamics and Memory Effects in an Intrinsically Disordered Protein Construct*, *Physical Review Letters* **125** (July, 2020) 058001.

[144] L. A. Morozova-Roche, J. A. Jones, W. Noppe, and C. M. Dobson, *Independent Nucleation and Heterogeneous Assembly of Structure During Folding of Equine Lysozyme*, *Journal of Molecular Biology* **289** (June, 1999) 1055–1073.

[145] J. Sabelko, J. Ervin, and M. Gruebele, *Observation of strange kinetics in protein folding*, *Proceedings of the National Academy of Sciences* **96** (May, 1999) 6031–6036.

[146] J. Brujić, R. I. Hermans Z., K. A. Walther, and J. M. Fernandez, *Single-molecule force spectroscopy reveals signatures of glassy dynamics in the energy landscape of ubiquitin*, *Nature Physics* **2** (Apr., 2006) 282–286.

[147] M. Hinczewski, C. Hyeon, and D. Thirumalai, *Directly measuring single-molecule heterogeneity using force spectroscopy*, *Proceedings of the National Academy of Sciences* **113** (July, 2016) E3852–E3861.

[148] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, *THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective*, *Annual Review of Physical Chemistry* **48** (Oct., 1997) 545–600.

[149] S. Muller-Spath, A. Soranno, V. Hirschfeld, H. Hofmann, S. Ruegger, L. Reymond, D. Nettels, and B. Schuler, *Charge interactions can dominate the dimensions of intrinsically disordered proteins*, *Proceedings of the National Academy of Sciences* **107** (Aug., 2010) 14609–14614.

[150] S. Pregent, A. Lichtenstein, R. Avinery, A. Laser-Azogui, F. Patolsky, and R. Beck, *Probing the Interactions of Intrinsically Disordered Proteins Using Nanoparticle Tags*, *Nano Letters* **15** (May, 2015) 3080–3087.

[151] A. J. Kovacs, *Transition vitreuse dans les polymères amorphes. Etude phénoménologique*, in *Fortschritte Der Hochpolymeren-Forschung*, Advances in Polymer Science, (Berlin, Heidelberg), pp. 394–507, Springer, 1964.

[152] E. M. Bertin, J.-P. Bouchaud, J.-M. Drouffe, and C. Godrèche, *The Kovacs effect in model glasses*, *Journal of Physics A: Mathematical and General* **36** (Oct., 2003) 10701–10719.

[153] Y. Lahini, O. Gottesman, A. Amir, and S. M. Rubinstein, *Nonmonotonic Aging and Memory Retention in Disordered Mechanical Systems*, *Physical Review Letters* **118** (Feb., 2017) 085501.

[154] S. N. Innes-Gold, P. A. Pincus, M. J. Stevens, and O. A. Saleh, *Polyelectrolyte Conformation Controlled by a Trivalent-Rich Ion Jacket*, *Physical Review Letters* **123** (Oct., 2019) 187801.

[155] C. Gosse and V. Croquette, *Magnetic Tweezers: Micromanipulation and Force Measurement at the Molecular Level*, *Biophysical Journal* **82** (June, 2002) 3314–3329.

[156] S. Mossa and F. Sciortino, *Crossover (or Kovacs) Effect in an Aging Molecular Liquid*, *Physical Review Letters* **92** (Jan., 2004) 045504.

[157] M. Rief, J. Pascual, M. Saraste, and H. E. Gaub, *Single molecule force spectroscopy of spectrin repeats: Low unfolding forces in helix bundles11Edited by W. Baumeister*, *Journal of Molecular Biology* **286** (Feb., 1999) 553–561.

[158] J. Yu, S. Malkova, and Y. L. Lyubchenko, *α-Synuclein Misfolding: Single Molecule AFM Force Spectroscopy Study*, *Journal of Molecular Biology* **384** (Dec., 2008) 992–1001.

[159] K. Neupane, A. Solanki, I. Sosova, M. Belov, and M. T. Woodside, *Diverse Metastable Structures Formed by Small Oligomers of α-Synuclein Probed by Force Spectroscopy*, *PLOS ONE* **9** (Jan., 2014) e86495.

[160] A. R. Panchenko, Z. Luthey-Schulten, and P. G. Wolynes, *Foldons, protein structural modules, and exons*, *Proceedings of the National Academy of Sciences* **93** (Mar., 1996) 2008–2013.

[161] D. K. Klimov and D. Thirumalai, *Lattice models for proteins reveal multiple folding nuclei for nucleation-collapse mechanism*, *Journal of Molecular Biology* **282** (Sept., 1998) 471–492.

[162] A. Amir, Y. Oreg, and Y. Imry, *Slow Relaxations and Aging in the Electron Glass*, *Physical Review Letters* **103** (Sept., 2009) 126403.

[163] N. Hirokawa, M. A. Glicksman, and M. B. Willard, *Organization of mammalian neurofilament polypeptides within the neuronal cytoskeleton.*, *Journal of Cell Biology* **98** (Apr., 1984) 1523–1536.

[164] R. I. Dima and D. Thirumalai, *Asymmetry in the Shapes of Folded and Denatured States of Proteins*, *The Journal of Physical Chemistry B* **108** (May, 2004) 6564–6570.

[165] A. Micsonai, É. Bulyáki, and J. Kardos, *BeStSel: From Secondary Structure Analysis to Protein Fold Prediction by Circular Dichroism Spectroscopy*, in *Structural Genomics: General Applications* (Y. W. Chen and C.-P. B. Yiu, eds.), Methods in Molecular Biology, pp. 175–189. Springer US, New York, NY, 2021.

[166] I. E. T. Iben, D. Braunstein, W. Doster, H. Frauenfelder, M. K. Hong, J. B. Johnson, S. Luck, P. Ormos, A. Schulte, P. J. Steinbach, A. H. Xie, and R. D. Young, *Glassy behavior of a protein*, *Physical Review Letters* **62** (Apr., 1989) 1916–1919.

[167] P. Arora, G. G. Hammes, and T. G. Oas, *Folding Mechanism of a Multiple Independently-Folding Domain Protein: Double B Domain of Protein A*, *Biochemistry* **45** (Oct., 2006) 12312–12324.

[168] S. Batey and J. Clarke, *Apparent cooperativity in the folding of multidomain proteins depends on the relative rates of folding of the constituent domains*, *Proceedings of the National Academy of Sciences* **103** (Nov., 2006) 18113–18118.

[169] C. Cecconi, E. A. Shank, F. W. Dahlquist, S. Marqusee, and C. Bustamante, *Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers*, *European Biophysics Journal* **37** (July, 2008) 729–738.

[170] D. J. Kauert, T. Kurth, T. Liedl, and R. Seidel, *Direct Mechanical Measurements Reveal the Material Properties of Three-Dimensional DNA Origami*, *Nano Letters* **11** (Dec., 2011) 5558–5563.

[171] Y.-L. Gao, N. Wang, F.-R. Sun, X.-P. Cao, W. Zhang, and J.-T. Yu, *Tau in neurodegenerative disease*, *Annals of Translational Medicine* **6** (May, 2018) 7–7.

[172] I. D. Vlaminck, T. Henighan, M. T. J. van Loenhout, D. R. Burnham, and C. Dekker, *Magnetic Forces and DNA Mechanics in Multiplexed Magnetic Tweezers*, *PLOS ONE* **7** (Aug., 2012) e41432.

[173] B. Lansdorp, *Pushing the Envelope of Magnetic Tweezer Resolution*. Ph.D., University of California, Santa Barbara, United States – California, 2015.

[174] B. M. Lansdorp, S. J. Tabrizi, A. Dittmore, and O. A. Saleh, *A high-speed magnetic tweezer beyond 10,000 frames per second*, *Review of Scientific Instruments* **84** (Apr., 2013) 044301.

[175] D. R. Lide, *CRC Handbook of Chemistry and Physics*, vol. 85. CRC press, 2004.

[176] J. Lipfert, X. Hao, and N. H. Dekker, *Quantitative Modeling and Optimization of Magnetic Tweezers*, *Biophysical Journal* **96** (June, 2009) 5040–5049.

[177] S. Zahran, J. S. Pan, P. B. Liu, and P. M. Hwang, *Combining a PagP fusion protein system with nickel ion-catalyzed cleavage to produce intrinsically disordered proteins in E. coli*, *Protein Expression and Purification* **116** (Dec., 2015) 133–138.

[178] E. Kopera, A. Belczyk-Ciesielska, and W. Bal, *Application of Ni(II)-Assisted Peptide Bond Hydrolysis to Non-Enzymatic Affinity Tag Removal*, *PLOS ONE* **7** (May, 2012) e36350.

[179] R. B. Kapust, J. Tözsér, T. D. Copeland, and D. S. Waugh, *The P1′ specificity of tobacco etch virus protease*, *Biochemical and Biophysical Research Communications* **294** (June, 2002) 949–955.

[180] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, *Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM*, *Science* **276** (May, 1997) 1109–1112.

[181] J. R. Wenner, M. C. Williams, I. Rouzina, and V. A. Bloomfield, *Salt Dependence of the Elasticity and Overstretching Transition of Single DNA Molecules*, *Biophysical Journal* **82** (June, 2002) 3160–3169.

[182] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*. Chapman and Hall/CRC, zeroth ed., May, 1994.

[183] D. Keller, D. Swigon, and C. Bustamante, *Relating Single-Molecule Measurements to Thermodynamics*, *Biophysical Journal* **84** (Feb., 2003) 733–738.

[184] I. Rouzina and V. A. Bloomfield, *Force-Induced Melting of the DNA Double Helix. 2. Effect of Solution Conditions*, *Biophysical Journal* **80** (Feb., 2001) 894–900.