University of California
Santa Barbara

Quantifying the conformations of sequence-defined polymers through entropic elasticity measurements

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

Doctor of Philosophy
in
Materials

by

Hoang P. Truong

Committee in charge:

Professor Omar A. Saleh, Chair
Professor Cyrus R. Safinya
Professor J. Herbert Waite
Professor Angela A. Pitenis

March 2024

The Dissertation of Hoang P. Truong is approved.

Professor Cyrus R. Safinya

Professor J. Herbert Waite

Professor Angela A. Pitenis

Professor Omar A. Saleh, Committee Chair

March 2024

Quantifying the conformations of sequence-defined polymers through entropic elasticity measurements

by

Hoang P. Truong

Acknowledgements

This work was made possible by the support of my committee members: Prof. Omar Saleh, Prof. Cyrus Safinya, Prof. Angela Pitenis, and Prof. Herbert Waite. On the NFLt study (Chapter 4), I acknowledge the contributions of my collaborators: Prof. Roy Beck, Gil Koren, Ram Avinery, and Mathar Kravikass. Especially, I would like to express my gratitude to Ian Morgan, whose prior works have helped me tremendously. On the polypeptoid study (Chapter 5), I am grateful for the support of my collaborators and their associates: Prof. Rachel Segalman, Shawn Mengel, Audra DeStefano, Beihang Yu, Amy Yang, and Zoey Meng.

I am grateful for the support and mentorship of my advisor, Prof. Omar Saleh. I will always remember the story of a mouse in a milk bucket you randomly told me in my first year. Thank you for the many valuable lessons, which have helped me grown as a scientist.

I would also like to thank my Saleh lab family: Ian Morgan, Sarah Innes-Gold, Gabrielle Abraham, Nate Conrad, Anna Nguyen, and Aria Chaderjian. Thank you for all the thoughtful conversations and helps with experiments. Thank you for the friendship outside the lab as well.

Lastly, I would like to thank my mother. Thank you for your unwavering investment in me. Thank you for realizing that the path to a better life is through good education, something you never had the chance.

Curriculum Vitæ

Hoang P. Truong

Education

2024	Ph.D. in Materials, University of California, Santa Barbara.
2018	B.A. in Chemistry, McDaniel College

Honors and Awards

2022	Connie Frank Fellowship, University of California, Santa Barbara
2017	Phi Beta Kappa, McDaniel College
2017	The Harry Clary Jones Scholarship, McDaniel College

Publications

H. P. Truong and O. A. Saleh. "Magnetic tweezers characterization of the entropic elasticity of intrinsically disordered proteins and peptoids." *Methods in Enzymology* 694, **2024**.

H. P. Truong, S. Mengel, B. Yu, R. A. Segalman, and O. A. Saleh. "Single-Molecule Polypeptoid Stretching Reveals the Effects of Charge Sequence on Chain Conformation." *Macromolecules* 56, **2023**.

H. P. Truong, I. L. Morgan, G. Koren, R. Avinery, R. Beck, and O. A. Saleh. "Pincus blob elasticity in an intrinsically disordered protein." *The European Physical Journal E* 46, **2023**.

D. Ferraris, R. Lapidus, **P. Truong**, D. Bollino, B. Carter-Cooper, M. Lee, E. Chang, M. LaRossa-Garcia, S. Dash, R. Gartenhaus, E. Y. Choi. "Pre-Clinical Activity of Amino-Alcohol Dimeric Naphthoquinones as Potential Therapeutics for Acute Myeloid Leukemia." *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)* 22, **2022**.

V. Pagonis and **P. Truong**. "Thermoluminescence due to tunneling in nanodosimetric materials: A Monte Carlo study." *Physica B: Condensed Matter* 531, **2018**.

B. A. Carter-Cooper, S. Fletcher, D. Ferraris, E. Y. Choi, D. Kronfli, S. Dash, **P. Truong**, E. A. Sausville, R. G. Lapidus, A. Emadi. "Synthesis, characterization and antineoplastic activity of bis-aziridinyl dimeric naphthoquinone–A novel class of compounds with potent activity against acute myeloid leukemia cells." *Bioorganic & Medicinal Chemistry Letters* 27, **2017**.

Abstract

Quantifying the conformations of sequence-defined polymers through entropic elasticity measurements

by

Hoang P. Truong

Intrinsically disordered proteins (IDPs) do not adopt a stable three-dimensional structure, thus posing a challenge to common characterization techniques that have found success with folded proteins. While IDP conformational characterization has been explored with small angle X-ray scattering (SAXS) and fluorescence resonance energy transfer (FRET), there is a demand for alternative experimental techniques that can provide new perspectives in this field.

Beyond biological IDPs, quantification of rapidly fluctuating conformational states is also crucial for the design and development of synthetic sequence-defined polymers. Inspired by the ability of proteins to assemble structures with high specificity, a class of polymeric materials emerged whose sequences can be tailored to achieve the same level of complexity. Among these, polypeptoids are the most promising due to their precise synthesis and chemical modifiability. The rational design of polypeptoid materials calls for a fundamental understanding of how their sequences influence their conformational behaviors.

We approach the conformational study of these two polymer systems through entropic elasticity measurements using magnetic tweezers. Force-extension measurements reveal insights into polymer structure at a broad range of length scales, from the local stiffness described by the persistence length to long-range interactions quantified by the Flory exponent. As these sequence-defined polymers are typically too short for magnetic

tweezers experiments, we also explore ligation schemes to form long tether constructs.

Here, we describe the investigation of the conformations of a model IDP, the neurofilament low molecular weight tail region (NFLt). Despite the high charge content, we find that NFLt is more compact than expected due to local hydrophobic attractions. Our results, specifically Flory exponent measurements in denaturant concentrations, are in remarkable agreement with previous results on other IDPs using SAXS or FRET. We thus demonstrate a novel framework to characterize the conformational ensembles of IDPs through entropic elasticity analysis.

We also investigate the effects of charge spacing on polypeptoid conformation through simple sequence designs and find conflicting observations on the electrostatic influence. Whereas the local flexibility is unperturbed, the solubility behavior in increasing ionic strength is dependent on the charge spacing. We attribute this unique conformational behavior to the chemical structure of polypeptoids, whose charge groups are on side chains. Our results contribute novel insights into the sequence-conformation relationship, as well as raising new questions on polyelectrolyte physics.

Through the studies on NFLt and charged polypeptoids, we establish a previously unexplored approach of using magnetic tweezers experiments and entropic elasticity analysis to quantify the conformations of sequence-defined polymers. For IDPs, magnetic tweezers can complement existing techniques (SAXS and FRET) and advance the general knowledge of protein conformational disorder. For polypeptoids, this framework can be applied for the investigation of other sequence effects, such as ampholyticity or hydrophobicity.

Contents

Chapter 1

Introduction

1.1 Motivation

Biopolymers are polymers that are synthesized in living organisms. The simplest class of biopolymers, linear polysaccharides, conforms to a strict definition of a polymer: a macromolecule composed of covalently bonded identical repeating units called monomers. The primary roles of polysaccharides in living systems are generally confined to providing structural networks in tissues [1, 2].

For more intricate functionalities that require interactions with other molecules with high specificity and precision, biopolymers must form distinct structures consist of defined and specialized domains. Homo-biopolymers, consisting of identical monomers, can only adopt random-walk configurations where the arrangements of monomers are indiscriminate, thus unable to fulfill this requirement. Instead, polypeptides and polynucleic acids, biopolymers that are characterized by their sequences, exist to serve these complex purposes (Fig. 1.1). The class of sequence-defined biopolymers follows a looser description of a polymer: in particular, polypeptides are made of repeating amino acids, which form the peptide backbone. In each amino acid, the α-carbon bears a side chain with

varying chemistry, giving each monomer a unique identity. The properties of the side chains and their arrangements within a polypeptide sequence inform its intramolecular interactions, thermodynamically driving the formation of higher-order structures. From only 20 genetically encoded amino acids, polypeptide sequences shape a diverse array of protein structures that are tailored to specific functions, from precise enzymatic machines to stiff cytoskeletal fibers. Generally, proteins submit to the sequence-structure-function paradigm, which describes how a protein sequence determines its structure, which in turn informs its function [3].

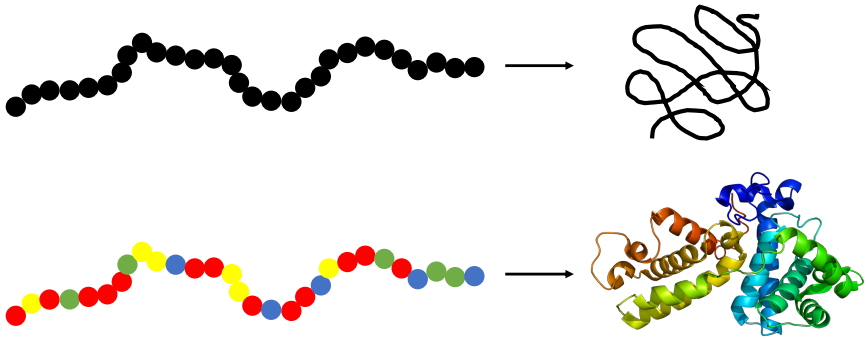

Figure 1.1: Difference between homopolymers (upper) and sequence-defined polymers (lower). The configurations of homopolymers are purely random. Meanwhile, sequence-defined polymers are able to form complex structures due to the chemical heterogeneity of their monomers. Part of this figure was reproduced from Emw - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=8763990.

Yet, there exists a ubiquitous class of proteins that do not conform to this paradigm, referred to as intrinsically disordered proteins or regions (IDPs or IDRs) [4, 5, 6]. As their name suggests, IDPs do not fold into a stable three-dimensional structure in their native state. The existence of IDPs were documented in isolated cases as far back as the 1950s [7, 8], and later estimates have demonstrated their biological prevalence [9]. In the eukaryotic proteome, 25-30% of proteins can be classified as IDPs [10], and more than 50% of proteins contain long IDRs [10, 11]. Despite not having any well-defined structures that

are typically associated with specific functions, IDPs nevertheless carry out numerous biological roles that exploit their flexible and dynamic nature [12, 13, 14, 15].

As it is believed that most proteins follow the structure-function paradigm, there has been significant research efforts in expanding the knowledge on protein structure to elucidate any unknown functions [16, 17]. The structures of proteins can be traditionally resolved through crystallography with high precision [18, 19], and recent advances in computational biophysics have facilitated structural predictions from sequences [20, 21, 22]. However, for IDPs, their unstable and rapidly fluctuating configurations, collectively referred to as conformational ensembles, present a significant challenge to conventional characterization methods [23].

Polymer physics provide an intuitive approach to the quantification of IDP conformations [24]. Because IDPs cannot be identified by specific secondary or tertiary structures, their conformational ensembles are typically quantified by size R, such as end-to-end distance R_{ee}, radius of gyration R_g, hydrodynamic radius R_h, etc. However, as these measurements are expected to scale with the number of residues N, they do not offer a meaningful comparison between IDPs of different lengths. To this end, the Flory polymer scaling law offers a simple yet effective normalization of sizes. This theory states the relationship between a polymer size and its number of monomers as $R \sim N^{\nu}$ [25], where the Flory exponent ν serves as a conformational parameter. Although the Flory scaling is formulated on the basis of homopolymers, the configurational heterogeneity of proteins can also be applied through ensemble averaging. While this concept ignores the chemistry details of individual residues, it can infer polymer-polymer and polymer-solvent interactions, referred to as solvent quality, allowing interpretation of intramolecular interactions in IDPs. The Flory polymer scaling law has been demonstrated to be a powerful approach to the quantification and comparison of IDP conformational ensembles [26, 27], and experimental studies on IDPs using small-angle X-ray scattering (SAXS) [28, 29, 30, 31] and

fluorescence resonance energy transfer (FRET) [32, 33] have found success in applying this concept.

Magnetic tweezers, a single-molecule force microscopy technique, have potential to be a novel, orthogonal approach to SAXS and FRET in the study of IDP conformations. Under a stretching force f, a polymer chain can be envisioned as segments of tensile screening length of size $\xi \sim k_B T/f$. The configuration of the polymer chain is aligned with the force at length scales above ξ, where intramolecular interactions are effectively screened; below ξ, the chain conformation is dominated by thermal fluctuations [34, 35]. By varying the applied tension, force microscopy techniques can probe polymer structures at a wide range of length scales. Notably, at very low forces (0.01-2 pN), the end-to-end extension L of the polymer chain scales with the force as a power law, $L \sim f^{1/\nu-1}$, allowing direct extraction of the Flory exponent [34]. While previous studies have applied single-molecule force microscopy techniques to investigate the conformations of IDPs, they are primarily focused on conformational dynamics when proteins are strongly stretched [36, 37, 38, 39]; measurements of slightly perturbed conformational ensembles at very low forces have largely been unexplored. Experimental measurements in this elastic regime are made feasible by the use of magnetic tweezers, which provide reliable access to the lowest stretching forces. To this end, we develop a novel framework to characterize the conformational ensembles of IDPs through magnetic tweezers measurements, an approach that is firmly rooted in polymer physics [40].

Following the success of using magnetic tweezers to probe the conformational ensembles of IDPs, we expand the framework to investigate the conformational behaviors of synthetic sequence-defined polymers. Inspired by the capabilities of proteins, a new class of polymeric materials have emerged, whose structures and functions can be engineered at the monomer level to achieve the complexity found in nature [41, 42, 43, 44]. Among those, poly(N-substituted glycines), or peptoids, have gained significant interest

due to their ability to be synthesized with precise sequences at relatively large molecular weights (up to 70 residues) [45, 46, 47]. In peptoids, on the peptide backbone, side chains are bound to the amine instead of the α-carbon. Due to their chemical similarities to peptides, peptoids exhibit good biocompatibility [48] while having excellent resistance to thermal [49] and enzymatic degradation [50]. Through sequence tuning, peptoids can self-assemble into a variety of structures, such as helices [51], nanosheets [52], nanotubes [53], and micelles [54]. To further push the boundaries of engineered peptoid materials, their rational design must be guided by fundamental principles of how the monomer sequences dictate their conformational behaviors. On this, we probe the effects of charge spacing on peptoid stiffness and solubility through simple polyelectrolyte sequences [55].

1.2 Intrinsically disordered proteins

Just as how the amino acid sequences define the folded structures of proteins, the sequences of IDPs encode disorder in their conformational ensembles [56, 11]. In an analysis of 275 folded proteins and 91 IDPs, Uversky et al. observed that a combination of mean net charge and hydrophobicity is sufficient in separating the two protein classes into two distinct regions through a linear boundary in the sequence space [56] (Fig. 1.2). Generally, IDPs contain a large fraction of charged and polar amino acids (Asp, Glu, Lys, Arg, Ser, etc.), while being deficient in non-polar and aromatic residues (Tyr, Trp, Phe, Val, etc.) [11, 57, 58] that are associated with driving hydrophobic collapse in protein folding [59, 60]. Depending on their sequence compositions, IDPs can be categorized to three groups: polar tracts, polyelectrolytes, and polyampholytes [61, 62]. Polar tracts, such as polyglutamines that are involved in neurodegenerative diseases pathways [63], are rich in polar yet uncharged residues, namely Gln, Ser, Thr, and His. IDPs that consist of predominantly one type of charge are referred to as polyelectrolyte. In contrast, IDPs

with comparable amounts of opposite charges are considered polyampholytes.

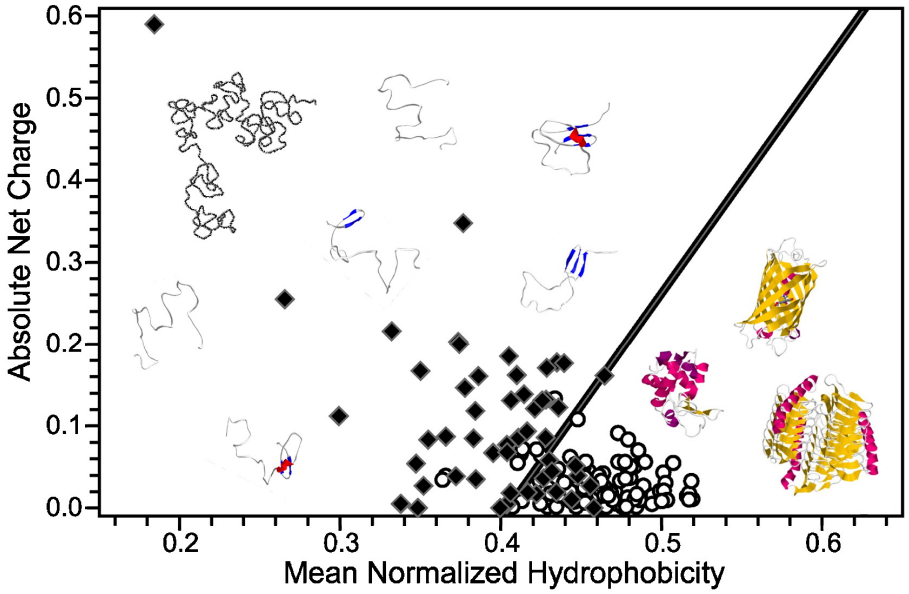

Figure 1.2: Sequence preference of IDPs [5]. Mean net charge (C) and hydrophobicity (H) combined are effective in distinguishing IDPs (black diamonds) and folded proteins (open circles). The mean hydrophobicity was computed based on the Kyte and Doolittle scale and normalized to between 0 and 1. The line $C = 2.785H - 1.151$ divides the two protein classes. This figure is reproduced from [5] with permission from Elsevier Ltd.

Despite being grouped in the same class, different IDPs vary significantly in their conformational ensembles, occupying a wide spectrum of disorder [64, 12] (Fig. 1.3A). This continuum spans fully random-walk chains, mostly disordered coils with regions of transient secondary structures, and collapsed globules. The sequence compositions of IDPs dictate where they lie on this spectrum. Polar tracts, despite the overwhelming presence of hydrophilic residues, are found to be compact globules [61] because their conformational behavior is dictated by their polypeptide backbone, which has poor solubility in water [65]. In the presence of charged residues, polyelectrolyte IDPs are typically extended coils, and their size measurements scale proportionally with mean net charge due to intramolecular electrostatic repulsion [32, 66]. For polyampholyte IDPs, the effects of sequence on their conformational ensembles are nuanced, dependent on the patterning

of opposite charges: well-mixed sequences are expanded while long charge blocks form hairpin-like conformations [67]. These sequence metrics and their expected conformational behaviors can be summarized in a phase diagram [66, 67], which can serve as an initial predictor for IDPs with unknown conformations. However, simple parameters such as net charge and hydropathy fail to capture local preferences for secondary structures that can vary among sequences with similar compositions [67, 68].

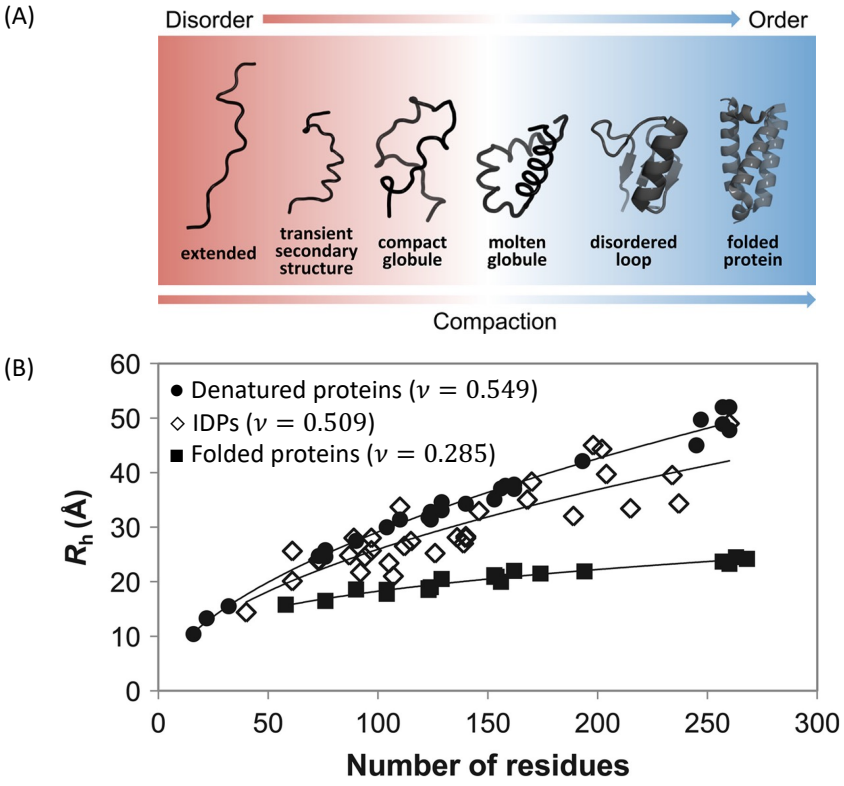

Figure 1.3: Quantification of IDP conformational ensembles. (A) Proteins exist on a structural spectrum from tightly folded to fully extended [12] Within their class, IDPs range from being entirely disordered to being more compact with regions of transient structures. This panel is reproduced from [12] with permission from American Chemical Society. (B) Flory polymer scaling concept describes IDP conformations [26] Global power laws were fitted to each protein category from previous Rh measurements. Unlike denatured and folded proteins, individual IDPs do not conform to their global fit. This demonstrates the diversity of IDP conformational behavior, as well as hinting at the potential use of the Flory exponent as a descriptive metric. This panel is reproduced from [26] with permission from Biophysical Society, published by Elsevier Inc.

Because IDPs dynamically fluctuate between various configurations on a broad and shallow energy landscape [69], their conformational ensembles cannot be defined by any specific instances. Instead, they can be characterized by statistical descriptors in polymer physics [24, 70]. Notably, the Flory polymer scaling law, $R \sim N^{\nu}$, has been found useful in describing the size and expansion of individual IDPs. Marsh and Forman-Kay explored this approach through Flory exponent analysis of three protein categories: folded, denatured, and natively disordered [26] (Fig. 1.3B). While folded proteins are compact globules and denatured proteins are extended random coils, the average conformation of IDPs rests between these two states. Further, as various folded and denatured proteins can be described by fixed Flory exponents, distinct IDPs do not follow a single value, exhibiting a diversity of conformational behaviors across the disorder spectrum. These observations demonstrate the use of the Flory scaling exponent as a simple yet effective parameter in describing and comparing conformational ensembles of IDPs.

1.3 Peptoids

Peptoids have stood out as the most promising sequence-defined polymeric material due to their much more robust and precise synthesis compared to other candidates in this class [45, 71]. Exploiting an automated solid-phased synthesis platform, the monomer identity and order can be accurately controlled. This scheme involves two steps in each iterative monomer addition cycle without the need for a protecting group on the main chain, shown in Fig. 1.4B. First, an amino resin is acylated with bromoacetic acid to form the backbone of a new monomer. The bromide is then displaced through nucleophilic substitution by a primary amine submonomer, which defines the new side chain. Due to the diversity and commercial availability of primary amines, numerous types of side chain chemistry can be incorporated to a peptoid sequence. This synthesis yields high

monodispersity in chain length of up to 70 monomers [72, 73].

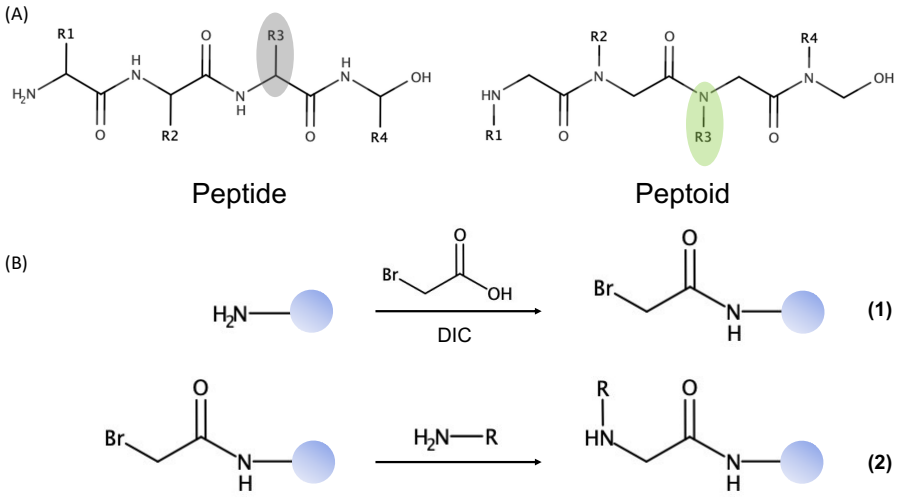

Figure 1.4: Chemistry of peptoids. (A) Peptoid side chains come from the nitrogen instead of the α-carbon. (B) Two-step solid-phase synthesis of peptoids [71] After acylation by bromoacetic acid and N,N'-diisopropylcarbodiimide (DIC), the bromide is substituted by a primary amine containing the desired side chain.

The displacement of side chains from the α-carbon to the nitrogen results in diverging conformational preferences in peptoids in comparison to peptides. The peptoid backbone does not contain hydrogen bond donors on the amides that are involved in secondary structure formation; nevertheless, peptoids can still form helices and sheets by incorporating structure-inducing side chains [74]. Most importantly, bond rotation around the tertiary amide leads to nearly equal energetics of the cis ($\omega = 0°$) and trans ($\omega = 180°$) conformations [75]. As a consequence, peptoids have a much higher preference for the cis configuration than observed in peptides, populating on average 35% of their conformational ensemble [76, 77]. Further, from experimental measurements of dihedral angles, peptoids are found to predominantly occupy two regions on the Ramachandran plot, specifically ($\phi = 90°$, $\psi = 180°$) and ($\phi = 270°$, $\psi = 180°$) [77].

1.4 Magnetic tweezers

While several methods have been explored to characterize the conformational ensembles of IDPs from a polymer physics perspective, their limitations call for an alternative experimental technique that can also tackle this question. Particularly, the approach of SAXS and FRET begins with measurements of size (R_g or R_{ee}), which can then be converted to their corresponding Flory exponent through the scaling law, $R \sim N^{\nu}$. However, for individual polymer chains (IDPs in this case), this conversion requires the exact formula, $R = AN^{\nu}$, of the scaling relationship. Determining the prefactor A requires assumptions on physical parameters of proteins, such as bond distances [78, 79, 80], which is nontrivial task that can lead to estimation inaccuracies. Further, discrepancies in measurements between SAXS and FRET obfuscate a consensus on the conformational ensembles of IDPs. Notably, with increasing denaturant concentration, while FRET demonstrated expansion due to improving solvent quality, SAXS results showed no significant changes in size [81].

Magnetic tweezers can provide new perspectives to the understanding of IDP conformational behaviors through entropic elasticity measurements. Within this approach, conformational ensembles of IDPs can be reduced to a random polymer chain with a configurational entropy Ω, where all states are equally probable. Upon applying a stretching force f, the end-to-end extension L of the polymer increases, reducing its number of available configurations. To minimize entropy loss, the polymer acts as a spring, imposing an opposing force to resist the stretching. This response, referred to as entropic elasticity, exhibits distinct characteristics at different force magnitudes and can be described by various scaling theories [35]. At high forces, as the polymer is stretched to its maximum contour length L_c, its elastic behavior follows an asymptotic elasticity model, such as the worm-like chain (WLC), which is dependent on the persistence length l_p [82].

At lower forces, when loops and self-interactions are permitted, the physical picture of the polymer can be depicted as a chain of "Pincus blobs", within whom the Flory scaling relationship applies [34]. Magnetic tweezers are particularly powerful for the latter analysis as they can reliably provide very low stretching forces (0.01-1 pN) [83] that are necessary for experimental observation of this elastic regime. Through magnetic tweezers measurements, important structural length scales and parameters can be obtained to quantify the conformational ensembles of IDPs. Further details on the scaling theories are elaborated in Chapter 2.

1.5 This dissertation

The first chapter introduces our study on the conformations of IDPs and sequence-defined synthetic peptoids, as well as our motivation for approaching this study through magnetic tweezers measurements. In Chapter 2, we lay out fundamental entropic elasticity scaling models and statistical parameters that describe the conformational ensembles of these polymers. Chapter 3 details the experimental setup and force calibration of magnetic tweezers experiments. In Chapter 4, we present our observations on the conformations of an IDP model, the neurofilament low molecular weight subunit protein tail region (NFLt), and its response to chemical denaturant. Our results on the NFLt also opens a new framework on the characterization of IDPs through entropic elasticity analysis. Chapter 5 elaborates on the investigation of the effects of sequence charge spacing on peptoid conformation. Through measurements in different solution ionic strengths and application of scaling theories at high- and low-force regimes, we uncover and reconcile seemingly conflicting results on electrostatic effects of sequence-defined polyelectrolytes. We conclude the thesis with a summary of all relevant findings and future directions in Chapter 6. The appendices contain further details of experiment protocols, supplemen-

tary data, and preliminary results on a partial construct of the NFLt and polyglycine.

Chapter 2

Polymer Physics

2.1 Statistical descriptors of polymer conformation

From a physics perspective, a polymer can be described as a freely jointed chain (FJC) with N number of identical, rigid segments of length b, referred to as Kuhn length [84]. These segments of statistical monomers are not synonymous with the chemistry definition of a monomeric unit; however, the chemistry of a polymer dictates its physical properties, which then define its Kuhn length. Due to random bending and turning between two rigid segments, a polymer chain can have numerous configurations with vastly different end-to-end extensions. Therefore, the ensemble averaged size of a given polymer chain, or extent, is commonly quantified by metrics such as the root-mean-squared end-to-end distance $\sqrt{\langle R_{ee}^2 \rangle}$ or radius of gyration $\sqrt{\langle R_g^2 \rangle}$, all of which will simply be referred to as Flory radius R. As R is expected to increase with N, their relationship is captured by the Flory polymer scaling law, $R \sim N^{\nu}$, where ν is the Flory exponent [25]. For two polymers with the same N but with incomparable chemical identities or in different solvents, their sizes are expected to diverge from one another. While the polymer physics approach is not concerned with chemistry details, it accounts for these subtleties through

assumptions of monomer interactions that are reflected by the Flory exponent (Fig. 2.1).

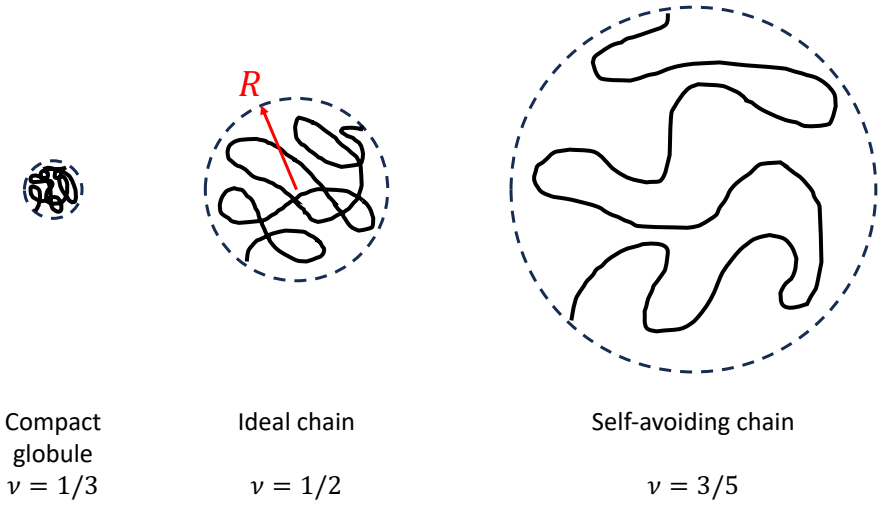

Figure 2.1: Scaling of ideal chain (middle), self-avoiding chain (right), and compact globule (left). The Flory radius R of a polymer is proportional to the number of monomers N as $R \sim N^{\nu}$, where the exponent is dependent on intra-chain interactions. Drawings reflect the approximate scaling for a polymer chain of $N = 1000$.

An ideal chain ignores any interactions between monomers. On a global scale, the configuration of an ideal chain is a random walk, where Kuhn segments can overlap and cross one another. In this physical picture, the size R of the polymer scales with N as $R \sim N^{1/2}$, giving a Flory exponent $\nu = 1/2$. For an ideal chain in solution, it can be inferred that polymer intra-chain interactions are equally favorable to polymer-solvent interactions; the solution condition that promotes a polymer to adopt an ideal chain configuration is referred to as its theta (Θ) solvent.

In real cases, polymers typically do not follow ideal chain behavior because intra-chain interactions cannot be ignored. Each monomer occupies a physical space, called excluded volume v, that cannot be accessed by other monomers. Because of this restriction, the configuration of a real (or swollen, self-avoiding) polymer chain is a random walk with self-crossings disallowed. The swelling behavior causes self-avoiding chains to

be more expanded than ideal chains, which is accounted for by a larger Flory exponent, $\nu = 3/5$, in the scaling relationship, $R \sim N^{3/5}$. A more involved theoretical calculation of the swollen chain has found a slight deviation from 3/5, specifically 0.588 [85]. Experimental measurements have also found that Flory exponent values of real polymer chains do not exactly match $\nu = 3/5$, such as 0.55-0.57 for synthetic polymers [86] or 0.57-0.598 for chemically denatured proteins [87, 88]. For a self-avoiding chain in solution, polymer-solvent interactions are more favorable than intra-chain interactions; this solution condition is referred to as a good solvent.

In the third case, when a polymer's intra-chain attractions are more favorable than polymer-solvent interactions, the solution condition is considered a poor solvent for that polymer. In a poor solvent, the configuration of a polymer is a compact or collapsed globule. By making the excluded volume negative to reflect monomer-monomer attractive interactions, the Flory exponent in this scenario is derived to be $\nu = 1/3$ [25].

2.2 Physics of stretching polymers

Under stretching, a polymer chain is aligned in the direction of the force f, where it can be statistically divided into segments of tensile screening length, $\xi = k_B T/f$, with $k_B T$ being thermal energy [34, 35]. Beyond this length scale, the configuration of the polymer chain is constrained by the force, and monomer interactions above this distance are screened out. Within ξ, unaffected by the tension, the polymer conformation is driven by thermal fluctuations where intra-chain interactions are allowed. This observation demonstrates the application of single-molecule stretching experiments: by modulating the applied force, they can probe the conformation of a polymer chain on a broad range of length scales. The elastic response of a polymer at various force regimes can be described by different scaling theories, which can reveal important structural parameters.

2.2.1 Elastic behavior at low forces

A polymer chain under very small perturbations is expected to act as an entropic spring due to its configurational entropy, where the end-to-end extension L scales linearly with the applied force f through a "spring constant" factor α, $L = \alpha F$ [89]. Accounting for ideal chain behavior, this elastic response follows [35]:

$$L = \frac{Nb^2}{3k_BT} f \tag{2.1}$$

For a real chain with a scaling relationship of $R \sim N^{3/5}$, de Gennes derived an alternative prediction to the entropic spring [90]:

$$L \approx \frac{N^{2\nu}b^2}{3k_BT} f \tag{2.2}$$

From the result above, Pincus pointed out that at intermediate forces, excluded volume interactions are weakened as monomers are stretched apart, resulting in an elastic response that instead resembles an ideal chain [34], as shown in Eq. 2.1 . The crossover between these two force regimes is reflected by the relationship between R and the tensile screening length ξ. At very low forces ($R \ll \xi$), the elastic behavior is expected to follow the de Gennes result ($L \sim N^{2\nu}$, Eq. 2.2), while at higher forces ($R \gg \xi$), the Pincus prediction holds ($L \sim N$, Eq. 2.1). Quantitatively, this elastic regime can be presented as a function Φ of R/ξ:

$$L \sim R\Phi(R/\xi) \tag{2.3}$$

where $\Phi(R/\xi) \sim (R/\xi)^p$. Φ is assumed to be a power law to accurately describe the elastic behavior at these low-to-intermediate forces, where a polymer chain is significantly stretched to an extension L beyond its Flory radius R but still far below its contour length

L_c. For very small perturbations ($R/\xi \ll 1$), $\Phi(R/\xi) \sim R/\xi$, and the de Gennes result is recovered. At higher forces ($R/\xi \gg 1$), in order to retain the $L \sim N$ scaling, the exponents must satisfy the condition of $\nu + \nu p = 1$. From this, the elastic response can be written as

$$L \sim f^{1/\nu - 1} \tag{2.4}$$

and the force-extension power-law exponent is denoted as γ, where $\gamma = 1/\nu - 1$. This force-extension scaling is also referred to as the "Pincus blobs" elastic regime: the polymer is a chain of random-walk blobs of size ξ (Fig. 2.2A). For a chain with swelling behavior, γ is 2/3 from the expected $\nu = 3/5$. An ideal chain ($\nu = 1/2$) under stretching will exhibit a γ value of 1. This demonstrates how the force-extension relationship of a polymer chain from low-force measurements can reveal the Flory scaling exponent (Fig. 2.2B).

2.2.2 Elastic behavior at high forces

At high forces, when the tensile screening length is smaller than a Kuhn segment b, the polymer chain is fully aligned in the direction of the force with minimal configurational fluctuations (Fig. 2.3A). With increasing forces, a consequence of the entropic spring picture is that the polymer chain extension will infinitely rise, which is physically impossible. To capture this finite limit, the FJC model predicts a scaling of $L \sim L_c(1 - 1/f)$, showing an asymptotic behavior to a contour length L_c with infinitely large f [91, 92].

The WLC model, derived by Marko and Siggia [82], has found the most success in describing this elastic regime [35]. Unlike the segment picture of the FJC, this model depicts a polymer chain as a continuous, semiflexible rod that can only bend at large length scales. Quantitatively, the bending direction at a point s on the polymer is exponentially correlated to that at point $s + x$, given by

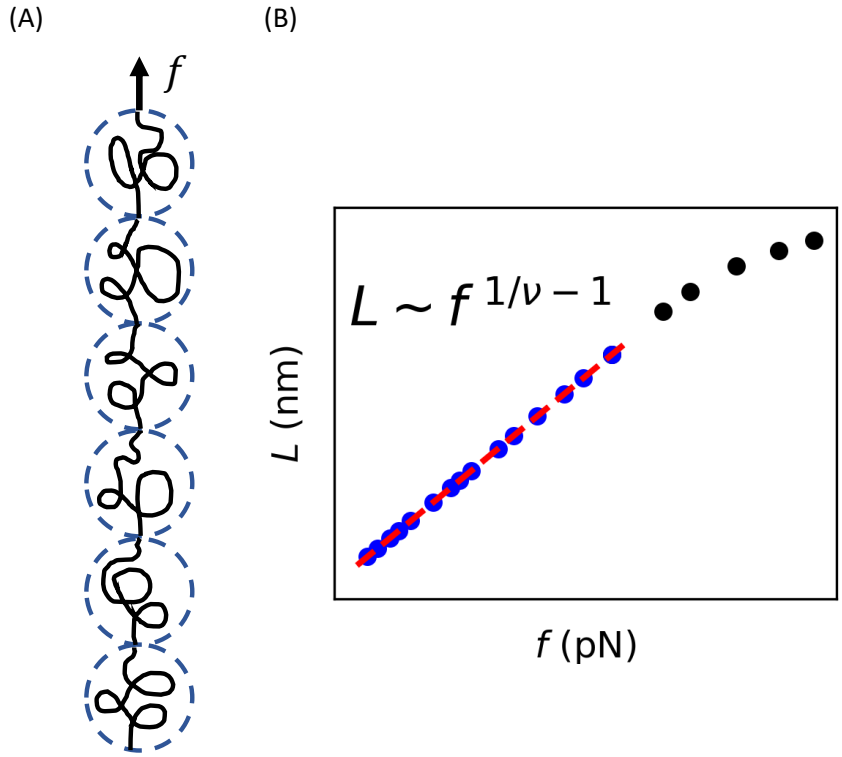

Figure 2.2: Polymer elasticity under stretching at low forces. (A) The polymer acts as a chain of Pincus blobs aligning in the direction of the force. Within each blob, the chain configuration is a random walk. (B) Representative force-extension curve of a flexible polymer showing the Pincus blob elastic regime. The plot is shown on a log-log scale. The force scales with the extension as a power law (blue circles), which can be fitted (red dashed line) to obtain the Flory exponent. On the log-log scale, the power-law relationship appears linear. Data at higher forces (black circles) do not follow this regime and are excluded from the fit.

$$\langle \hat{t}(s+x) \cdot \hat{t}(s) \rangle = \exp(-|x|/l_p) \tag{2.5}$$

where $\hat{t}(s)$ is the unit tangent vector at point s and l_p is the persistence length. Typically, l_p is equivalent to half the Kuhn length, $b = 2l_p$, in a model with jointed segments. As l_p mathematically denotes the decay rate of the backbone correlation, it physically manifests as the flexibility of the polymer: while a small l_p means a polymer can easily bend at short length scales, a polymer with a large l_p remains a rigid rod. Some example l_p

values of biopolymers are ~ 0.5 nm for single-stranded nucleic acid [93, 94, 95], 50 nm for double-stranded nucleic acid [82, 96], and 4-8 mm for microtubules [97, 98]. For a polymer under stretching, the WLC model predicts [82]:

$$L \approx L_c \left(1 - \sqrt{\frac{k_B T}{4 f l_p}}\right) \tag{2.6}$$

A numerical formula of WLC derived by Bouchiat is often used for a more accurate estimation of persistence length from force-extension results [99]. While various theoretical models have been shown to provide better predictions of this elastic regime for specific macromolecules [100, 101, 102, 103], the strength of the WLC rests in its ability to describe the behaviors of a wide range of polymers with good relative accuracy [35]. Because of this, in force-extension experiments, the WLC model is typically the first choice for analysis of measurements. From a force-extension curve, fitting the WLC model to the asymptotic regime yields l_p, providing insight into the flexibility of the measured polymer (Fig. 2.3B).

2.2.3 Applying polymer physics concepts in single-molecule force microscopy experiments

Magnetic tweezers have demonstrated success in experimentally verifying these elastic scaling theories and providing insights to polymer conformational behaviors [35]. Particularly, for flexible polymers ($l_p < 5$ nm), to apply the Pincus prediction, the forces required to probe structural length scales beyond the persistence length ($\xi > l_p$) fall in the sub-picoNewton range (below 1 pN). Compared to other single-molecule force microscopy techniques, magnetic tweezers are particularly powerful as they can reliably access this force range. Indeed, past magnetic tweezers studies have elucidated the force-extension power law regime in single-stranded nucleic acid (ssNA) [94] and hyaluronic acid (HA)

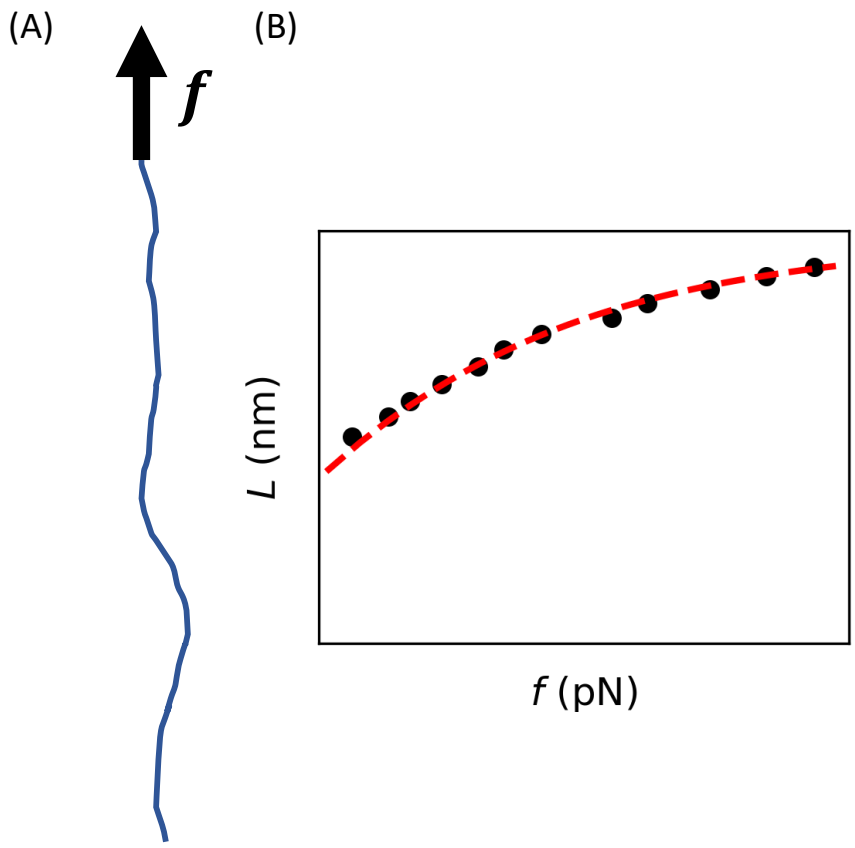

Figure 2.3: Polymer elasticity under stretching at high forces. (A) The polymer fully aligns in the direction of the force with small conformational fluctuations. (B) Representative force-extension curve of a polymer showing the asymptotic elastic regime. The plot is shown on a log-log scale. The extension approaches the contour length with increasing force (black circles). The data can be fitted to the WLC model (red dashed line) with 2 parameters: contour length L_c and persistence length l_p.

[104], showing that their γ values approach 2/3 due to intra-chain electrostatic repulsion. In a higher ionic strength, when such interactions are fully screened out, their force-extension exponents reach 1, corresponding to ideal chain behavior in a theta solvent. To our knowledge, besides the work presented in this dissertation, there has not been any studies on this elastic regime of proteins.

At higher forces where the WLC prediction applies, magnetic tweezers experiments can discern electrostatically-driven flexibility of charge-containing biopolymers. As stated

by Odijk [105], Skolnick, and Fixman [106], the flexibility of polyelectrolytes is defined by a combination of intrinsic backbone stiffness l_p^0 and electrostatic persistence length l_p^e; with increasing ionic strength, the latter component is reduced as electrostatic repulsion is screened out, causing the total persistence length to also decrease. Previous magnetic tweezers studies have validated these polyelectrolyte theories through measurements on ssNA [95] and HA [104], showing salt-dependent stiffness by varying solution salt concentration.

Various single-molecule force microscopy studies have applied the WLC model in determining the conformational dynamics of proteins [107, 108, 109, 110, 36, 111]. Under stretching, different configurations of a protein manifest as distinct extension measurements. Force ramp experiments reveal sawtooth patterns due to unraveling events of these states. These curves can then be fitted with the WLC model to obtain contour length values, which point out the identity and mechanical properties of specific domains within a protein. Although the WLC model can successfully describe protein elasticity in most cases, several studies have noted its limitations [36, 112, 100]: while the WLC prediction is based on a homopolymer, protein configuration is heterogeneous due to different side chain properties. An alternative model has been proposed, where backbone rotation angles are restricted; if the contour length is known, this model can better capture the force-extension relationship of proteins compared to the WLC prediction [100].

Chapter 3

Magnetic Tweezers Experimental Methods

3.1 Permissions and attributions

3.2 Introduction

Magnetic tweezers allow probing of a broad range of structural length scales by varying the applied tension. At intermediate to high forces (5-100 pN), the tensile screening length ξ is ∼0.8-0.04 nm, the scale at which polymers are locally rigid rods. At low forces (<1 pN), ξ is >4 nm; for the flexible polymers studied in this dissertation ($l_p \approx 1$ nm), ξ is the size of Pincus blobs in which polymer segments follow random walk configuration.

The force range of interest in this dissertation is 0.1-15 pN, which covers both elastic regimes for IDPs and polypeptoids.

A magnetic tweezers experiment involves mechanical manipulation of tethered polymer chains through magnetic force [83, 114, 115] (Fig. 3.1B). A polymer of interest is immobilized on a glass surface and attached to a paramagnetic bead, which is attracted and pulled upward by a magnetic field gradient. The three-dimensional position of the bead is traced in real time with video tracking; particularly, the bead's height, which translates to the polymer's extension, is measured through its diffraction pattern. The applied force can be precisely calibrated by analyzing the variance of bead motion. Recording the extension of the polymer in a series of forces gives a force-extension curve, which can then be examined with scaling theories described above.

The general scheme for experiments discussed in this dissertation is as follows: A microfluidic flow cell is fabricated, in which polymers are tethered. The flow cell is placed on a custom-built magnetic tweezers setup, where force-extension measurements of the polymer sample can be obtained. Through trace correction and force calibration, raw data are converted to force-extension curves, which are then fitted with entropic elasticity models at different force regimes. The structural parameters acquired from this analysis can reveal insights into the conformations of the polymers.

3.3 Magnetic tweezers

3.3.1 Instrument setup

The custom-built magnetic tweezers on which all experiments in this dissertation were performed were previously developed and described by Ribeck, Saleh, and Lansdorp, and their detailed documentation is in Refs. [115] and [116]. The instrument is essentially

an inverted microscope combined with a pair of strong magnets for force generation. A basic visualization of the setup is shown in Fig. 3.1A.

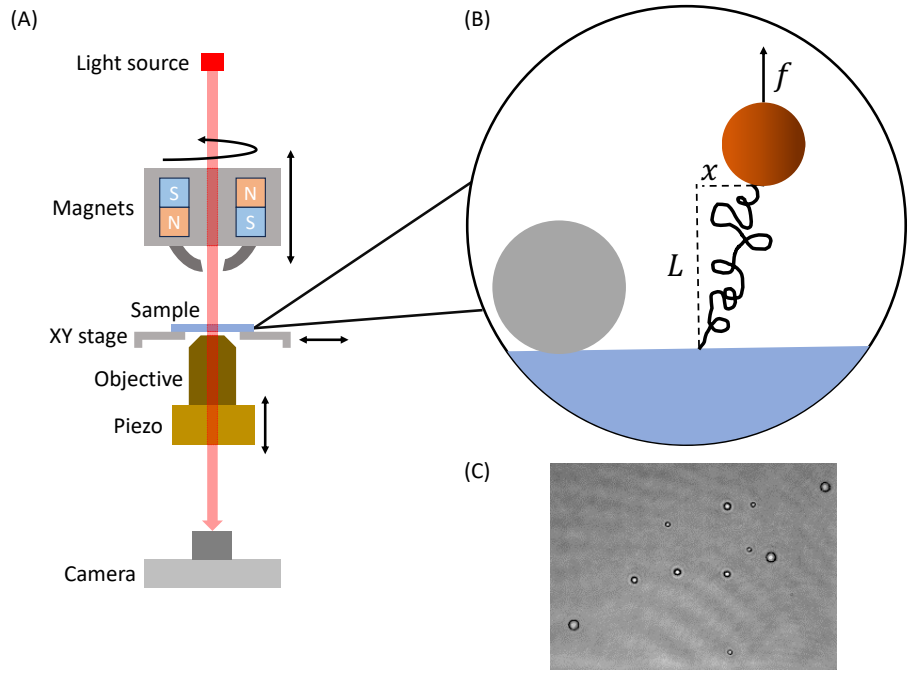

Figure 3.1: Schematics of a magnetic tweezers experiment. Drawings are not to scale. (A) Simplified scheme of the magnetic tweezers instrument setup. The arrows indicate the degrees of freedom of movable components. (B) Magnified view inside a flow cell sample. A polymer chain is tethered to a paramagnetic bead, which applies a stretching force f. The polymer extension L is measured from the glass surface to the bottom of the bead. The tether fluctuates in x and y positions. A reference bead is immobilized on the surface. (C) Example field of view in an experiment. Reference beads, tethered beads, and unbound paramagnetic beads are observed.

An XY stage (Physik Instrumente (PI) M-545.2P, operated by a PI C-867-260 controller), where a flow cell is placed, provides micron-precision movements to adjust the lateral field of view. The sample is in contact with a high numerical aperture objective (Nikon CFI Plan Fluor 60XS Oil NA 1.2) from below. This objective is mounted on a piezoelectric nanopositioning device (PI P-725.2CD), which controls its height and allows adjustments of the focal plane in increments as small as 0.5 nm. Above the stage is the magnetic piece: it consists of two pairs of 1/2-inch NdFeB cube magnets, oriented

north-north and south-south, housed in a machined aluminum casing [117] (Fig. 3.1). The magnetic field gradient is confined and amplified at the tips of two low-carbon steel pole pieces. The magnetic piece is screwed onto a set of two motors (PI M-126, C-843), which provides translational and rotational movements. The magnitude of the applied force is dependent on the position of the magnets from the sample: a higher force is achieved with a closer distance, and vice versa.

The instrument is illuminated by a red super-luminescent diode (QPhotonics QSDM-680-2) or a fiber-coupled light-emitting diode (Thorlabs M470F3). The light, aligned through the pole tips of the magnetic piece and the objective, is sent to a high-speed CMOS camera (EoSens 3CXP, Mikrotron). Images and videos taken from the camera are transferred to a computer in real time using a frame grabber (BitFlow Cyton-CXP4). This setup facilitates image capturing at frame rates as high as 4000 Hz. All data presented in this dissertation were measured at 400 Hz. Image viewing and piezo/motor controlling are accomplished through a user interface created with LabVIEW (version 2015 SP1, National Instruments).

3.3.2 Bead tracking

High-speed tracking of beads is enabled by a GPU-accelerated algorithm written in LabVIEW, which has been previously described [115]. The xy positions of beads are determined by measuring the self-convolution of pixel intensity between successive frames [118], which tracks the shift of the bead centroid with an accuracy of 1-2 nm [115]. The z position of beads can be tracked by exploiting diffraction rings around beads, which are caused by interference of incident light waves and waves scattered off from the beads [115, 114]; these diffraction patterns change with focal plane height. Tracking z-position begins with calibration, which takes images of beads at each focal plane in increments

of 0.1 μm within a 5-μm range, enabled by precise movements of the piezo. The radial intensity profile of each image is calculated and added to a lookup table. Throughout an experiment, the height of a bead at any given force can be determined in real time by referencing to this table. The typical accuracy of z-position tracking is 1-10 nm [115].

Instrumental drift adds unwanted noise to tracking data, which is removed by using reference beads. These are inert 2.8 μm-diameter polystyrene beads (Bangs Laboratories) physically immobilized onto a glass surface through partial melting (Fig 3.1B). Reference beads are simultaneously tracked with tethered beads and can capture drift in their traces, which are then subtracted off to yield noise-free data.

3.3.3 Data acquisition

A magnetic tweezers experiment begins with searching for an appropriate field of view for tracking beads (Fig. 3.1), which is done by moving the sample laterally on the motorized XY stage. A desirable screen should include tethered beads and at least 3 reference beads. A bead attached to a polymer chain can be recognized by moving the magnets away, as its diffraction pattern will significantly differ due to height change. For the tracking program to function properly, beads must be adequately separated so that their diffraction rings do not overlap. Double tethers, which are beads attached to multiple polymer molecules, can be screened out by running at least 50 magnet rotations at an intermediate force ($\sim$2 pN); this causes the chains to braid and the double tethers to collapse. Tethers are typically positioned at the center of a screen for optimal illumination and force application.

Once a field of view is selected, calibration images are taken for each selected bead [115]. This is done at the highest force (closest magnet position) where the heights of tethered beads are the most stable. Because the calibration step only provides the

relative height of a bead, the absolute extension is determined from the z-position of the glass surface. This can be found by tracking tethers at zero force, when the magnets are furthest away or entirely absent. Without any tension, beads randomly fluctuate near and frequently touch the glass surface; on a histogram plot, their motions exhibit a Gaussian distribution with a lower tail cutoff due to the physical limit at the bottom. The 1st percentile of this plot gives the z-position of the surface.

Force-extension curves are generated by recording the z-position of tethers in a series of discrete forces. Beginning with the highest force, tracking is done at fixed magnet positions in increments of 0.25 mm away from the sample (Fig. 3.2). The amount of time spent at each magnet position is increased from the previous step by an multiplicative "acquisition scaling" factor of 1.15; this is because longer traces are required for estimation of lower forces, which will be explained below. Typical force-extension experiments in prior magnetic tweezers studies would measure 31 forces for each curve [119, 120]. However, for short polymers (>1 μm), surface adhesion is common at very low forces, causing their extensions to not be properly calculated; therefore, we instead measure 20 forces per curve, which is still enough to resolve the Pincus blob elastic regime. The total tracking time for a force-extension curve containing 20 steps is 10 minutes.

Raw data from LabVIEW tracking must undergo small corrections to reflect physical measurements. The tracking program records xy-positions as numbers of pixels and z-position as uncorrected microns. The x- and y-traces are converted to nm by multiplying to the pixel length, which is found through calibration with a microscope micrometer slide; this number is dependent on the instrument setup, and it is roughly 140 nm for a 60x objective and 100 nm for a 100x one. The z-position is translated into nm by accounting for the change in refractive index: while calibration is done by moving the objective in Type A immersion oil ($n_{\text{oil}} = 1.515$), bead traces are recorded in aqueous buffer ($n_{\text{water}} = 1.333$). To correct this, the z-trace is multiplied to the ratio of the two

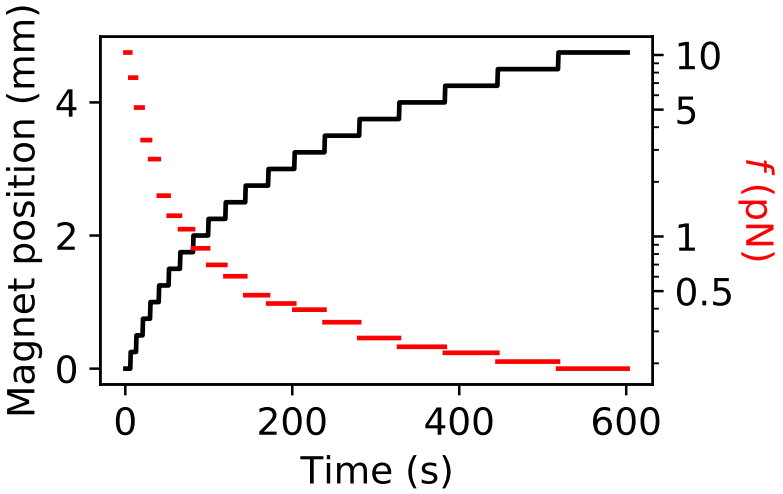

Figure 3.2: Force-extension measurement protocol. Linear steps of magnet position generates an exponential spacing of forces. Lower forces are held at increasingly longer tracking times.

refractive indices, $n_{\text{water}}/n_{\text{oil}} = 0.880$. A very high solute concentration can change the refractive index of a buffer [121, 122, 123], and this must be considered when performing raw data correction. The polymer extension of a tether is taken as the mean of its z-trace. Additionally, instrumental drift from reference bead traces is subtracted from tether measurements.

3.3.4 Force calibration

In the presence of the magnets, a paramagnetic bead aligns its axis along the direction of the magnetic field and is pulled upwards by a gradient force

$$\vec{f} = (\vec{m}(\vec{B}) \cdot \vec{\nabla})\vec{B} \tag{3.1}$$

where $\vec{B}$ is the magnetic field, $\vec{\nabla}$ is the gradient vector, and $\vec{m}(\vec{B})$ is the field-dependent magnetic moment of the bead.

To accurately estimate this force without any quantification of the magnetic field,

a bead tethered by a polymer chain is modeled as an inverted pendulum, where the upward force is analogous to gravity (Fig 3.2B). The bead is a harmonic oscillator inside a potential well with an energy of $\frac{1}{2}k_x\langle x^2\rangle$, where k_x is the force constant of the trap in the x direction and $\langle x^2\rangle$ is the variance of bead position. In addition, the equipartition theorem states that the energy of a degree of freedom is $\frac{1}{2}k_BT$. Equating the two energy estimations gives

$$k_x = \frac{k_BT}{\langle x^2\rangle} \tag{3.2}$$

demonstrating an inversely proportional relationship between the force constant and the bead position variance. This can be observed experimentally, where a bead shows larger fluctuations at lower forces as the trap is less stiff (Fig. 3.3). As the magnetic field gradient is aligned in the x-axis due to the instrument setup, the entropic spring picture of the tether dictates the applied force as

$$f = k_xL \tag{3.3}$$

This suggests that the force acting on a polymer chain in theory can be directly calculated from the variance of the bead position. However, parasitic noise in traces leads to a larger variance, resulting in a systematic underestimation of the applied force.

Instead, the Allan variance (AV), a time-domain approach, offers a more accurate estimation of the applied force in single-molecule force microscopy experiments [124]. The AV measures the variance δ^2_{AV} of the fluctuations in bead position at timescales τ, mathematically defined as

$$\delta^2_{AV}(\tau) = \frac{1}{2}\langle(\bar{x}_{\tau,j+1} - \bar{x}_{\tau,j})^2\rangle \tag{3.4}$$

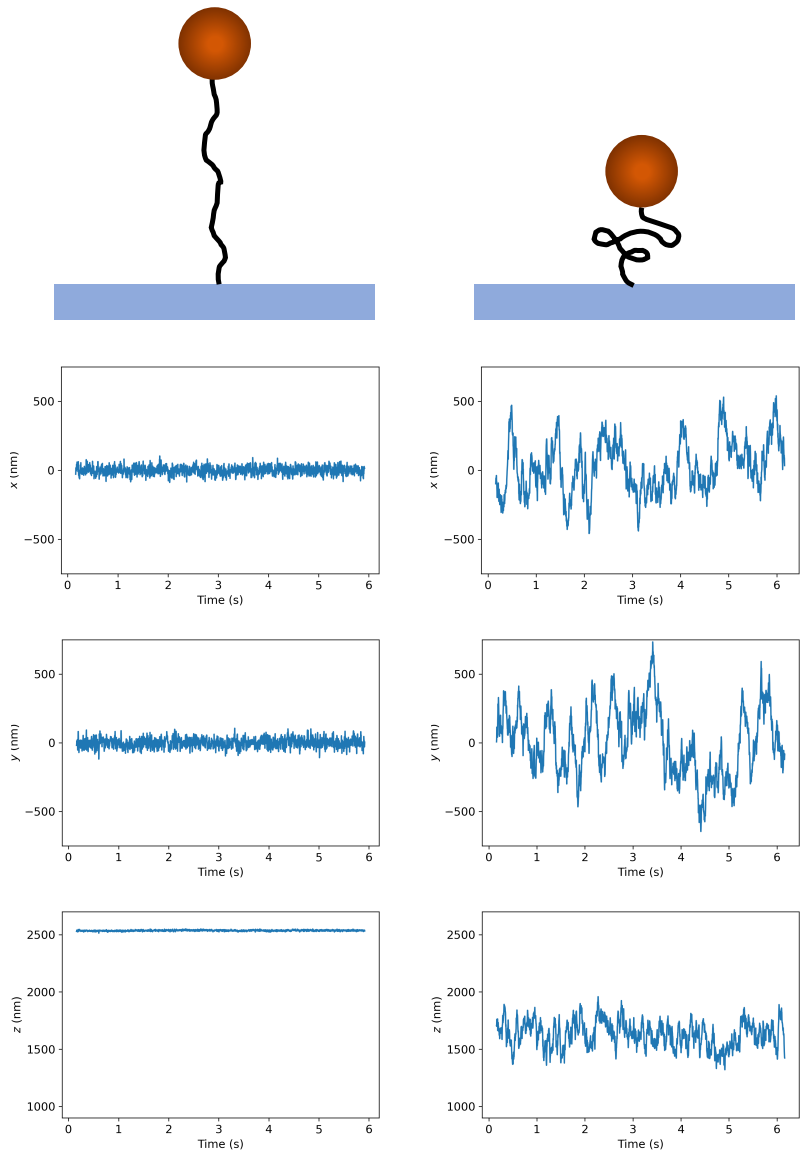

Figure 3.3: Example x, y, and z traces of a tether at high (10.3 pN, left panels) and low force (0.19 pN, right panels). Qualitatively, the bead position in all three dimensions at the lower force exhibits a much larger variance.

in which $\bar{x}_{\tau,j}$ is the mean position in the j-th window of length τ binned from the x-trace. This general expression of the AV can be transformed to specifically describe the Brownian motion of a bead in a harmonic well. [124]. Local fluctuations at short timescales are governed by hydrodynamic drag, signified by a drag coefficient α_d. For

a spherical particle with a radius R_b in water with a viscosity η, this coefficient can be determined *a priori* through $\alpha_d = 6\pi\eta R_b$, which is approximately 1e−5 pN.s/nm for a bead with $R_b = 500$ nm. On longer timescales, the bead motion is confined by the limits of the potential well, which is characterized by the spring constant k. The dependence of the AV on both free parameters α_d and k can be written as [124]:

$$\delta^2_{AV}(\tau) = \frac{2k_BT\alpha_d}{k^2\tau}\left(1 + \frac{2\alpha_d}{k\tau}e^{\frac{-k\tau}{\alpha_d}} - \frac{\alpha_d}{2k\tau}e^{\frac{-2k\tau}{\alpha_d}} - \frac{3\alpha_d}{2k\tau}\right) \tag{3.5}$$

As α_d and k dominate in opposing timescale lengths, the resulting AV function is non-monotonic (Fig. 3.4), where the maximum occurs at the characteristic timescale $\tau_c \sim \alpha_d/k$.

Estimation of both α_d and k requires experimental traces long enough that can resolve the entire AV function. At small τ, to the left of the $\delta^2_{AV}(\tau)$ plot, fitting is only determined by the drag coefficient, whereas on the other side of the maximum, the AV is predominantly dependent on the spring constant. Further, at lower forces, as k becomes smaller, the characteristic timescale τ_c increases, shifting the AV peak to the right, shown in Fig. 3.4. Measurements at lower forces therefore require more data to estimate k, hence the acquisition scaling factor to increase the trace length. Meanwhile, fitting of the drag coefficient α_d can be affected by poor resolution of short timescales due to low tracking frame rate; for all experiments in this dissertation, 400 Hz is sufficient to overcome this issue.

Because the AV follows a Gamma distribution, fitting of Eq. 3.5 is best served by a bias-free maximum likelihood estimation (MLE) [115]. Application of this fitting algorithm can be done elegantly in Python through the open-source Tweezepy package [125]. From a time trace of bead position, the algorithm outputs the best-fit α_d and k parameters, as well as their associated errors. To account for white noise in the data

that can reduce fitting quality, the program also provides an option to add an additional noise parameter.

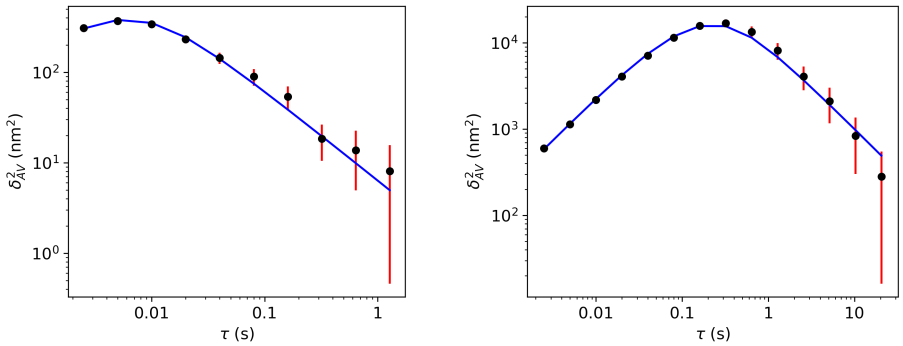

Figure 3.4: Example AV calculations (black circles) of the x-traces of a tether at 10.3 pN (left) and 0.17 pN (right), along with their respective MLE fits (blue lines). The error bars (red lines) show the 95% confidence intervals at each point. All calculations were done by the Tweezepy package [125]. At lower forces, the AV maximum is shifted to a longer timescale.

Further corrections can be taken for accurate estimation of force. At the highest forces, the AV peaks are pushed to lower timescales towards the limitation of the tracking frame rate, causing poor fitting of the drag coefficient, and consequently the force constant. As drag can be separately calculated from theory, it can be constrained in the AV fitting, leaving k as the sole free parameter. Faxén's law, which predicts the drag of a bead at a specific height from the surface, can be applied to pre-determine α_d [126]. Additionally, as the height of a bead is measured from its bottom-most point, if the polymer chain is attached to it at any other sites (referred to as "mistethering"), the extension will be underestimated, as well as the force following Eq. 3.3. Under slow magnet rotation (0.2 Hz) at an intermediate force ($\sim$2 pN), a mistethered bead produces a lateral limaçon pattern. The x- and y-traces can be fitted by a sinusoidal formula, yielding the mistethered length that can be added to the original extension measurement [127].

3.4 Sample preparation

A successful magnetic tweezers experiment requires significant considerations to the preparation of samples. This involves assembling a flow cell and optimizing attachment chemistry. We discuss below general experimental guidance to achieve high tether count and improve data throughput. Because this dissertation studies different polymer constructs that require individualized strategies, their protocols will be detailed in Appendices B and C.

A polymer chain is bound to a glass surface through a "click chemistry", which is a fast and highly specific one-step reaction. Click chemistry for biopolymers typically does not require harsh conditions and only generates harmless byproducts [128]. Common click reactions are maleimide and thiol; N-hydroxysuccinimide (NHS) and primary amine; and azide and alkyne or dibenzocyclooctyne (DBCO). A magnetic tweezers experiment design begins with selecting a click pair, where one species is labeled at a terminus of the polymer and the other is coated on the glass surface. This choice is dependent on the feasibility of having such functional group or the compatibility with the polymer. For example, the NHS-amine reaction cannot be used with a protein sample containing multiple lysines because its attachment site is rendered uncontrollable.

A functionalized glass surface can be purchased from a commercial supplier, such as Microsurfaces (now defunct) or PolyAn. These coverslips typically feature a passivation coating to prevent nonspecific surface adhesion, such as a monolayer PEG from Microsurfaces or a proprietary polymeric matrix from PolyAn. Depending on the labeling of the target polymer, the chemistry of the surface should be selected accordingly. Alternatively, functionalized coverslips with PEG coating can be fabricated in-house following established protocols [129, 130].

3.4.1 Flow cell construction

A flow cell consists of two glass coverslips sandwiching a parafilm cutout, forming a microfluidic channel in which polymer samples are bound. The assembly protocols for different coverslip sizes (22 x 22 mm for Microsurfaces, 25 x 60 mm for PolyAn) are generally identical, apart for some small modifications. Parafilm M cutouts are prepared with a laser cutter or a low-cost Cricut Maker®, which help define the features of a flow channel with high precision. Reference beads are immobilized on the functionalized bottom surface by first applying a suspension of 2.8 μm polystyrene microspheres (Bangs Laboratories, diluted to 0.01 mg/mL in 200-proof ethanol). After the ethanol has fully evaporated, the coverslip is placed on a heat plate at 100°C for 1 minute and 45 seconds, which allows the beads to be firmly fixed while still retaining their spherical shape. The top glass coverslip is cleaned with 2 M sodium hydroxide (NaOH), Milli-Q water, and then in a plasma cleaner. The upper side of this coverslip is treated with a hydrophobic Rain-X® coating, which forces fluids to enter the channel instead of spreading on top the flow cell. Upon assembly of the coverslip sandwich, the parafilm is melted with a soldering iron at ∼385°C (∼725°F on the equipment setting). The completed flow cell has a channel with a height of ∼130 μm (thickness of a parafilm sheet), along with two open outlet ends for adding and wicking off buffers.

Protocol for Microsurfaces coverslips (22 x 22 mm)

Schematics of a flow cell are shown in Fig. 3.5. The parafilm cutout for Microsurfaces coverslips is a 22 x 22 mm square with a channel whose width is 6 mm along the diagonal. The top layer is an 18 x 18 mm coverslip placed at a 45° angle rotation. Reference beads are melted on the bottom surface from 18 μL of the polystyrene microsphere suspension.

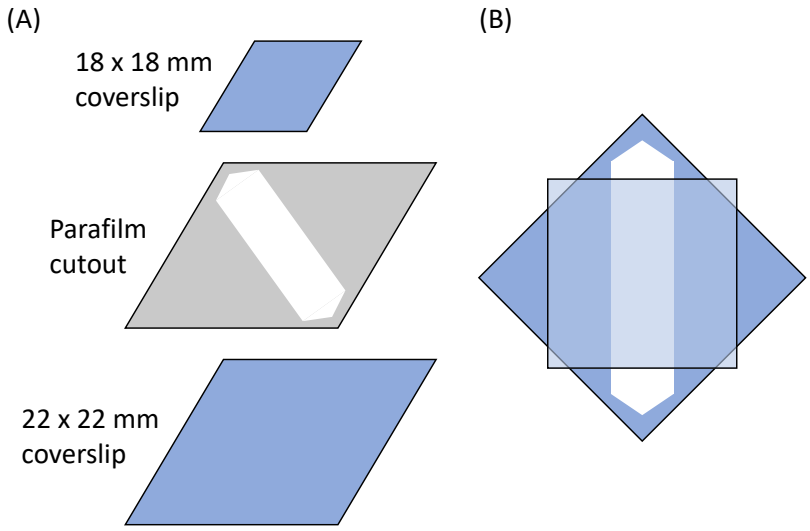

Figure 3.5: Flow cell construction for Microsurfaces coverslips. (A) The three components of a flow cell: a 22 x 22 mm functionalized coverslip, a parafilm cutout with a 6 mm channel width, and a 18 x 18 mm coverslip. (B) The completed flow cell after assembly. The gap from the parafilm cutout forms the fluid channel. The exposed ends serve as buffer outlets. Drawing B is to scale.

Protocol for PolyAn Coverslips (25 x 60 mm)

Schematics of a flow cell are shown in Fig. 3.6. The parafilm layer for this setup is a 25 x 60 mm rectangle with a 6 x 32 mm cutout centered and parallel to the length of the coverslip. While it might be tempting to increase the width of the channel to maximize surface usage, the fluid flow can become turbulent. Because of the larger size, two channels of 3.5 mm width can be fitted within the parafilm cutout, separated by 4 mm. The top coverslip is 22 x 22 mm aligned parallel to the bottom glass. A 50 μL of the polystyrene microsphere suspension is used for this flow cell size.

3.4.2 Synthesizing and tethering different polymer constructs

Polymer chains whose contour lengths are below 200 nm are typically problematic to magnetic tweezers measurements due to poor relative tracking precision and surface

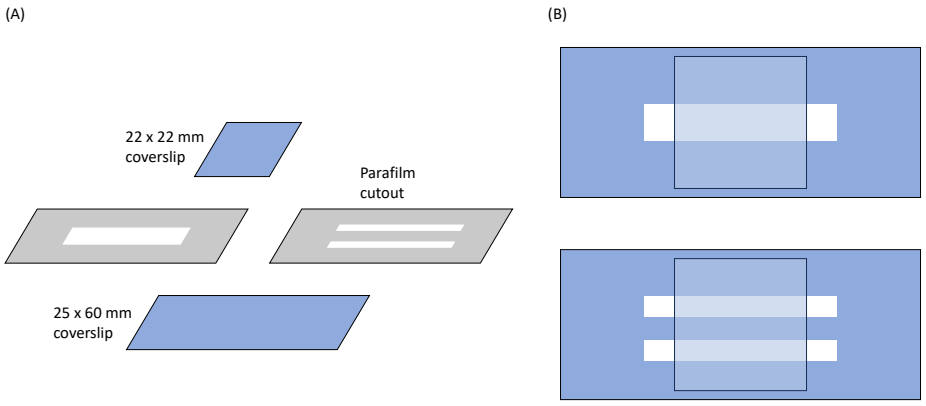

Figure 3.6: Flow cell construction for PolyAn coverslips (25 x 60 mm). (A) The three components of a flow cell. The parafilm cutout allows one or two fluid channels within a coverslip. (B) The completed flow cells after assembly, depending on the parafilm cutout choice. Drawings B are to scale.

interactions [131]. This presents a major challenge to studying IDPs and peptoids with this technique, as these polymers are inherently much shorter than 200 nm. To overcome this barrier, we engineer long polyprotein and polypeptoid constructs through end-to-end attachment of individual sequence units. For IDPs, this is done by inserting terminal cysteines during sequence design, allowing continuous disulfide linkage [132] (Fig. 3.7A). Besides this scheme, previous single-molecule force microscopy studies have explored other routes to generate polyprotein constructs [133]. Peptoids permit much more modification freedom in their synthesis, providing various options for attachment chemistry. We choose azide-alkyne terminal groups to enable end-to-end click reactions (Fig. 3.7B).

After synthesis, the long polymer constructs are labeled with terminal functional groups for tethering. Polyproteins are capped with azide and biotin: the biotin end binds to a streptavidin-coated paramagnetic bead, while the azide end attaches to a glass surface (Fig. 3.7C). Because the ascorbic acid required in azide-alkyne click chemistry is detrimental to disulfide bridges, the catalyst-free DBCO-azide reaction is preferred. However, as DBCO-coated coverslips are not offered commercially to our knowledge,

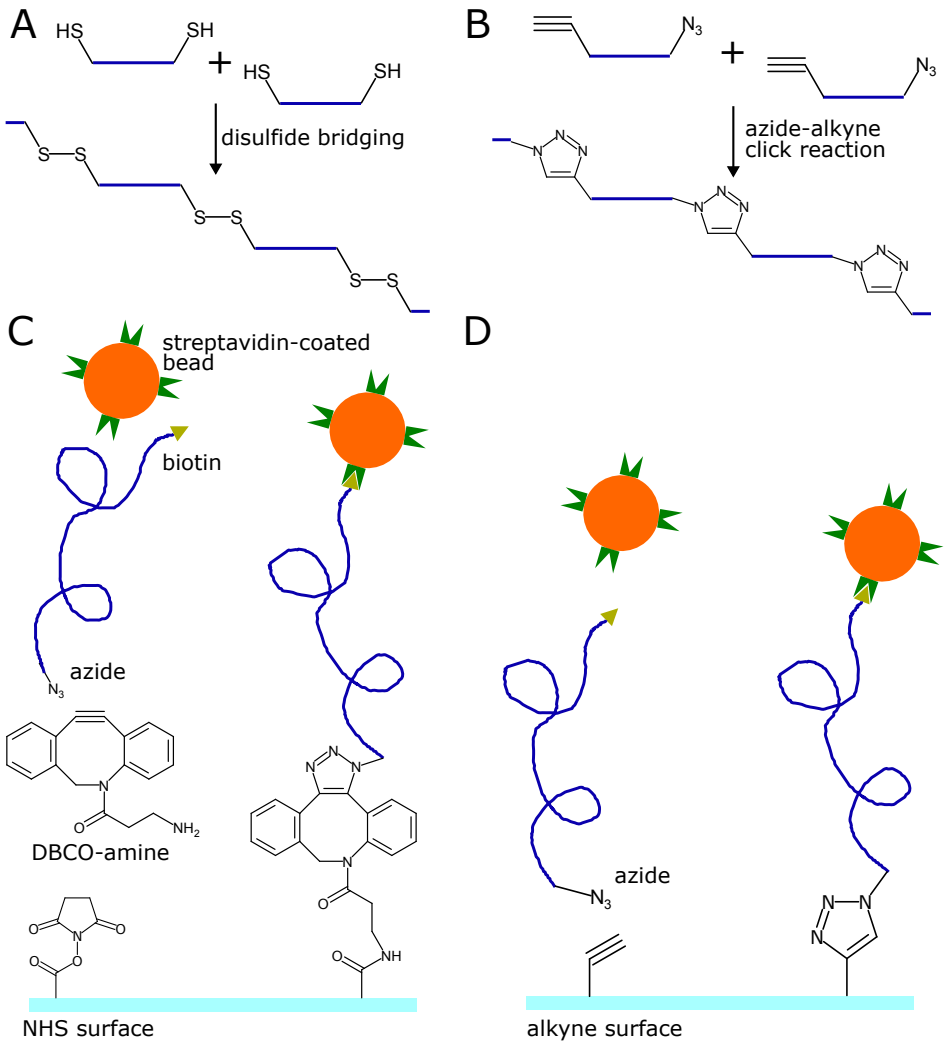

Figure 3.7: Long polyprotein and polypeptoid synthesis and attachment schemes. (A) Long polyproteins are formed through terminal disulfide bridging between individual IDP units. (B) End-to-end azide-alkyne click reactions generate long polypeptoids. (C) Tethering scheme for IDP polyproteins with azide and biotin labels. A commercially available NHS surface is treated with DBCO-amine to change the functionalization to DBCO. The azide end on the polyprotein can then bind to the surface. (D) Tethering scheme for polypeptoids. The azide end is bound to an alkyne surface.

another type of functionalized surface can be modified in one additional step to have the DBCO chemistry (Fig. 3.7C). For polypeptoids, because they already contain one click-ready terminus, we only need to label the other end with biotin (Fig. 3.7D).

Paramagnetic beads coated with streptavidin are tethered to biotin-labeled polymer

samples with a very high binding affinity. The choice of paramagnetic bead affects the applied force range. While a 1 μm-diameter bead (Dynabeads™ MyOne™ Streptavidin C1, Invitrogen) imparts smaller perturbations (0.01-15 pN), a 2.8μm-bead (Dynabeads™ M-280 Streptavidin, Invitrogen) can apply much higher tension (2-100 pN). For all experiments in this dissertation, because the Pincus blob regime is of particular interest, MyOne C1 beads were used.

Regardless of the polymer sample, a successful experiment requires adequate passivation treatment. At the single-molecule scale, nonspecific adhesions are detrimental, as they reduce data throughput and introduce surface effects that confound the analysis of measurements [134]. While functionalized glass coverslips have passivation coatings, further actions can be taken to prevent these unwanted interactions. A 0.01-0.1% Tween-20 are added to experimental buffers, which forms a surfactant layer around macromolecules and inhibits close-range attractions [135]. A solution containing 1% bovine serum albumin (BSA) can also be used after polymer immobilization to block any hydrophobic patches on the glass surface.

Chapter 4

Entropic elasticity of an IDP model

4.1 Permissions and attributions

4.2 Introduction

Intrinsically disordered proteins (IDPs) do not fold into a well-defined three-dimensional structure in their native state, yet they remain biologically active [12, 136]. The existence of IDPs challenges the classic structure-function paradigm in which the higher-order structure of a protein dictates its functionality [137]. To fully understand the role of IDPs in various biological processes, detailed knowledge of their conformations would be valuable. In the case of folded proteins, their structures can be characterized experimentally

with great precision, such as with crystallographic analyses. However, IDPs are dynamic and fluctuate across a broad conformational ensemble [138, 139], thus disallowing traditional approaches to measuring and classifying protein structure.

As an alternative, prior studies have modeled IDPs as random-walk polypeptide chains and used structural classifications derived from polymer physics scaling laws[26, 27, 33]. For example, Marsh and Forman-Kay compiled measurements of the hydrodynamic radii, R_h, of IDPs of various residue length N, and demonstrated that they followed a rough power-law scaling, $R_h \sim N^{\nu}$, with a near ideal-chain Flory exponent of $\nu \approx 1/2$ [26]. Later authors carried out direct high-precision measurements using single-molecule Forster energy transfer (smFRET) [33, 27] or small angle X-ray scattering (SAXS) [28, 140, 30], and employed a polymer physics framework to interpret the data, e.g. in terms of the scaling behavior of ensemble average quantities (such as mean-squared end-to-end distance), or of the probability distributions governing the full conformational ensemble. Generally, some caution is needed in applying simple polymer scaling models to IDPs, as such models are formulated for long, homopolymeric chains, whereas IDPs can be somewhat short and are always heteropolymeric. Thus, IDPs can have heterogeneous interactions along their contour that are not described by simple scaling theories; indeed, several prior works have observed evidence of such heterogeneity [81, 141, 142, 132, 143, 31]. Yet, the examples above demonstrate the utility of polymer scaling models in providing a framework for interpreting the dynamic structure of IDPs [144].

A hallmark of polymers is the entropic elastic response, in which the internal conformational freedom of the chain provides an entropic force that resists increases in extension. The advent of single-molecule force microscopy enabled direct measurement of the elastic response of biopolymers [82], including IDPs [36]. Such measurements are typically carried out in the high force regime where the chain is nearly aligned in the

direction of applied force, and the elastic response arises from a mix of enthalpic contributions and entropic contributions arising from small fluctuations controlled by the chain's local configurational freedom[100, 112]. However, as initially pointed out by Pincus [34], the single-chain entropic elastic response is quite rich at lower forces (i.e. forces, f, less than a few picoNewtons). At such forces, chains can exhibit featuring multiple distinct elastic regimes that are sensitive to other polymeric structural parameters. Of particular interest here is the intermediate-force regime, in which, from a scaling perspective, the chain can be modeled as a series of tensile (or Pincus) blobs of size given by the tensile screening length, k_BT/f, where k_BT is the thermal energy and f is the applied force. In this regime, the force-extension response follows a power-law scaling sensitive to the Flory exponent, $L \sim f^{1/\nu-1}$[34]. Previous single-molecule stretching studies have exploited the low-force capabilities of magnetic tweezers to observe the Pincus blob elasticity regime in various biopolymers, using it to analyze solvent-quality and polyelectrolyte effects [145, 94, 104, 35].

Here, we demonstrate the utility of measurements of entropic elasticity to quantify the structure of a disordered protein. Specifically, we use magnetic tweezers to observe Pincus blob elasticity within a model IDP construct, and, using the Flory exponent inferred from elasticity, we study the structure of the chain in various solvent conditions. The model chain is a construct formed from the disordered tail of the neurofilament low-molecular-weight (NFL) subunit protein. Neurofilaments are important constituents of the axonal cytoskeleton, and their disordered tail domains control the spacing and mechanical stability of the cytoskeletal network [148, 149, 150]. The NFL tail (NFLt) construct studied here is a 163-residue protein with a net charge per residue (NCPR) of -0.23. The sequence features a C-terminal region rich in glutamic acid and an N-terminus that is polar yet uncharged (Fig. 4.1C). Following Das and Pappu's state diagram that qualitatively predicts the conformation of an IDP based on its fraction of both charges

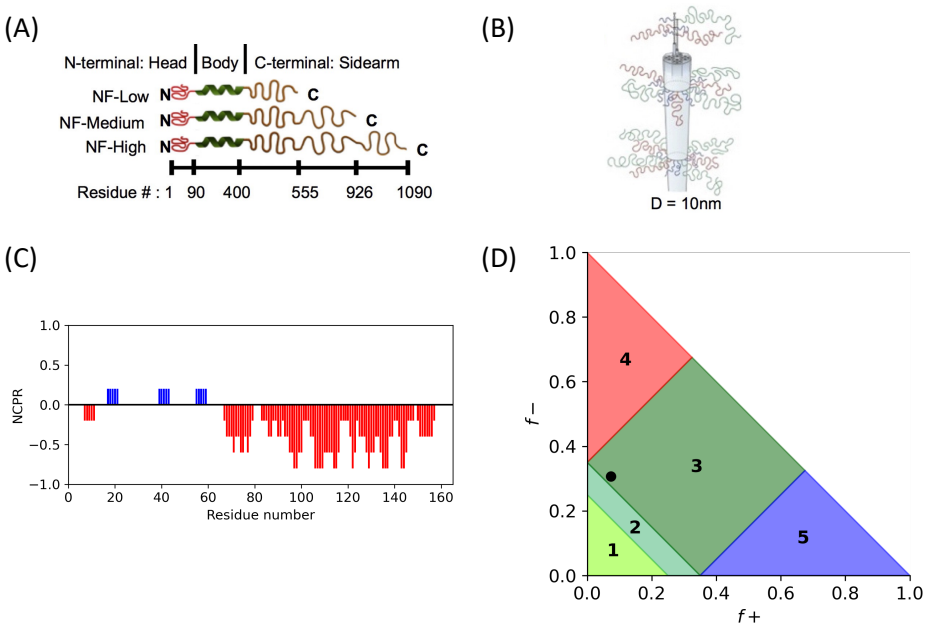

Figure 4.1: Graphical descriptions of the NFLt. (A) Neurofilament subunit proteins are categorized by their molecular weights: high, medium, and low. Their structures all feature a globular head, a helical body, and a disordered tail [146]. (B) Neurofilament subunit proteins assemble into a bottlebrush structure, whose central rod diameter is ∼10 nm [146]. (C) Charge profile of the NFLt. The average NCPR of each residue was calculated using the localCIDER algorithm with a window size of 5 [147]. (D) NFLt's position in Das and Pappu's IDP state diagram [67]. The plot provides a qualitative prediction of an IDP's conformation based on two simple metrics: the fractions of positive and negatively charged residues. Region 1 contains sequences with very few charges, which are expected to be compact globules. IDPs with larger fractions of both charges (strong polyampholytes) occupy region 3, ; depending on the amino acid arrangements, their conformations can be either coils or hairpins. Region 2 indicates a transitional space where IDPs can assume any conformations between 1 and 3. Regions 4 and 5 host sequences with predominantly one type of charge (strong polyelectrolytes), which are expected to be swollen coils. The NFLt (black dot) is located in region 3. Panels A and B were reproduced from [146] with permission from Elsevier B.V.

[67], NFLt is expected to be a semi-compact coil (Fig. 4.1D). From the native sequence, terminal cysteines are added to enable end-to-end disulfide bridging, generating long polyproteins for magnetic tweezers experiments.

Our stretching experiments show that the construct, in low-ionic-strength aqueous solvent, is surprisingly more compact than the sequence's net charge would suggest.

We find that the Flory exponent, ν, increases upon adding the denaturant guanidinium hydrochloride (GuHCl); at the highest denaturant concentration, our estimate of ν is comparable to that expected for a polymer in good solvent. This observation is consistent with other studies on disordered or denatured proteins [27, 30], as well as recent FRET and SAXS measurements on segments of the NFLt [31]. Finally, the generality of the entropic-elastic response means that the single-molecule stretching method and analysis used here could be usefully applied to probing the structure of other IDPs.

4.3 Experiments and results

A recombinant NFLt construct was expressed in *E. coli* and purified by chromatography to >95% purity. The 163-residue NFLt sequence (available in Supplementary Information) was designed to include terminal cysteine residues. This allowed the linking of individual NFLt units through end-to-end disulfide bridges; the resulting polyprotein chains typically contained 10-30 units. Given the 62 nm contour length of a single 163-residue sequence, the resulting polyprotein chains were thus long enough (> 500 nm) for magnetic tweezer analysis. The disulfide-linked polyproteins were labeled at each terminus with, respectively, azide and biotin groups, permitting specific attachments to a functionalized glass surface and 1 μm-diameter magnetic beads as needed for mechanical manipulation. Details on the synthesis and characterization of the polyprotein construct are described in a previous study [132].

Magnetic tweezer experiments were carried out by, first, constructing a glass flow cell using an NHS-functionalized, PEG-coated glass surfaces (Microsurfaces, Inc.) that were treated with 5 mg/mL dibenzocyclooctyne-amine (DBCO-amine) in 1:1 DMSO:phosphate buffered saline pH 8.3 for 1 hour. Azide-labeled polyproteins ($\sim$1 μM) in 10 mM tris buffer pH 7.4 with 0.1% Tween-20 were incubated in the flowcell for 24 hours, allow-

ing coupling of the azide on the polyprotein to the DBCO on the surface. Unbound polyproteins were washed away with the same buffer solution. Streptavidin-coated, 1 μm-diameter magnetic beads (Dynabeads) were incubated for 10 minutes for attachment to the polyproteins, thus forming stretchable tethers; excess beads were then rinsed away. Elastic measurements were then performed in a custom-built magnetic tweezers setup that applies force to the tethers through an externally-applied magnetic field. The applied force and extension were quantified using image-based bead tracking and analysis of the bead's thermal fluctuations, as previously described [115, 125]. Experiments were performed in 10 mM 2-(N-morpholino)ethanesulfonic acid (MES) buffer at pH 6.8 with 0.1% Tween-20.

We polymerized the NFLt constructs, attached their termini to a glass surface and a magnetic bead, and used magnetic tweezers to carry out elastic measurements over a range of forces. A typical force-extension measurement, carried out in low-ionic-strength aqueous solution, is shown in Fig. 4.2. As seen there, the elastic response of a single chain shows a low force (subpicoNewton) power-law regime along with a high force regime where the chain becomes less compliant. This curve shape is similar to that found previously for other biopolymers [94, 145, 104, 35], and is indicative of Pincus-blob elasticity at low forces. In this regime, the chain displays significant loop formation, and long-range monomer interactions are allowed. As the force increases beyond a few picoNewtons, the loops are pulled out, and the compliance of the chain accordingly decreases. In this regime, the chain asymptotically approaches its contour length, L_c. The crossover force, f_c between the regimes is expected to occur when the tensile screening length, $\xi \equiv k_B T/f$, is roughly equal to the chain's persistence length, l_p [34, 35]. Thus, the picoNewton scale of f_c indicates that l_p is of nanometer scale.

Estimates of l_p are found by analyzing the asymptotic regime at high force, which we find to be well-described by the Marko-Siggia worm-like chain (WLC) model [82] (dotted

line, Fig. 4.2). As Schwarzl et al. noted, the WLC model is inexact in its application to polypeptides; they suggested it is superseded by a modified freely-rotating chain model that is more accurate when the chain contour length is known [100]. However, the polydispersity that results from our synthesis scheme does not allow *a priori* knowledge of the contour length; further, we note that detailed analysis of the local stiffness is not a focus of this study. We thus proceeded to fit the force-extension data to an analytical approximation of the WLC model derived by Bouchiat [99]. Because the WLC model does not consider long-loop, excluded-volume interactions, we only fit it to the data at forces above f_c, here estimated to be 2 pN. The fitted parameters are not sensitive to small changes in the value of f_c used as a cutoff. We found the average l_p from 11 individual polyprotein molecules to be 1.07±0.07 nm (error reported is the standard deviation). This value is within the broad range of $0.4 - 2$ nm found in previous studies on polypeptides [151, 152, 153], disordered proteins [154], and denatured proteins [155, 156, 157]. Thus, the measured local flexibility is consistent with the polyprotein construct acting as an IDP without significant secondary structures. Generally, persistence length observation serves as an internal validation on the identity and integrity of a measured tether, and this result gives confidence in the subsequent Pincus blobs analysis. For data compilation of different tethers, we only select those that exhibit reasonable flexibility behavior expected of an IDP (i.e. $l_p \sim 1$ nm).

We further directly extract the Flory exponent of the polyprotein from the force-extension curve by analyzing the Pincus elasticity regime ($f < 2$ pN) (Fig. 4.2). The Pincus blob model indicates that the expected force-extension scaling relationship is $L \sim f^{1/\nu-1}$ in this regime [34]. By fitting a power-law to the $f < 2$ pN data and applying this scaling relation, we find the Flory exponent of the NFLt in low-ionic-strength aqueous solution to be 0.52±0.01. This value is distinctly lower than the value $\nu = 0.6$ expected of a swollen, self-avoiding chain, i.e. a polymer in good solvent.

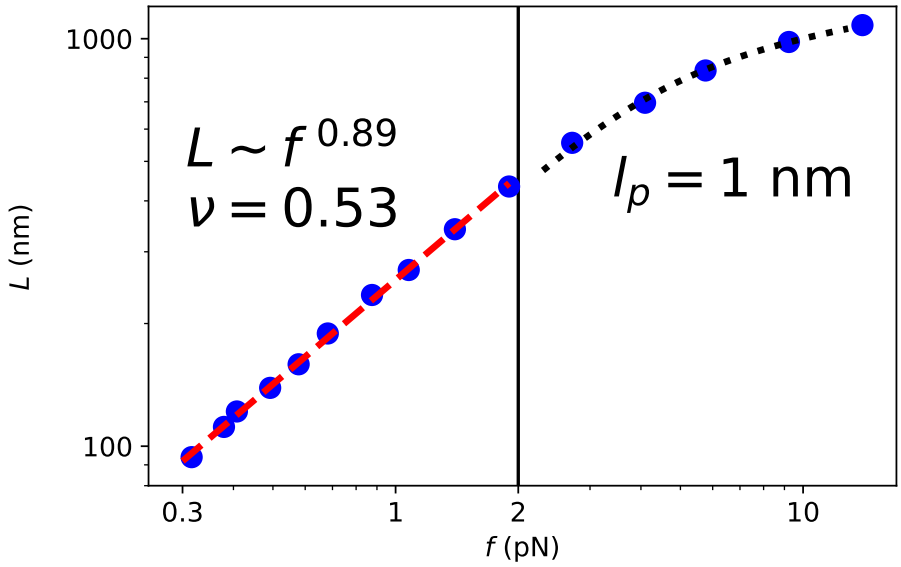

Figure 4.2: Example force-extension curve of one NFLt polyprotein tether (blue dots) [40]. The plot is shown on a log-log scale. Data at higher forces ($f > 2$ pN) are fitted to the WLC model (black dotted line). Data below the cutoff ($f < 2$ pN) are fitted to a power law (red dashed line). The horizontal line ($f = 2$ pN) shows the boundary of the two elastic regimes.

We further probe changes to the elasticity of NFLt in different solution conditions, specifically denaturant guanidinium hydrochloride (GuHCl). Denaturants are commonly used to unfold proteins by disrupting hydrophobic interactions, and thus they are considered good solvents to proteins. In an experiment, after measurements in buffer, denaturant solutions are flowed in, and tracking of the same tethers can be repeated. All GuHCl concentrations (0.1-6 M) contain a 10 mM MES and 0.1% Tween-20 background buffer to maintain the intended pH. In the strongest denaturant (6M GuHCl), the Flory exponent is 0.62 ± 0.01, corresponding to a fully swollen chain (Fig. 4.3). Other intermediate GuHCl concentrations also elicit a significant increase in the Flory exponent compared to that in buffer, confirming that denaturant is indeed a good solvent for the NFLt. Contour length measurements of a single tether in different denaturant solutions tend to vary noticeably likely due to solvation effects; to compare force-extension curves,

the extensions are normalized by individual L_c values from separate WLC fits. Additionally, the polyprotein persistence length does not show any appreciable change with solvent condition.

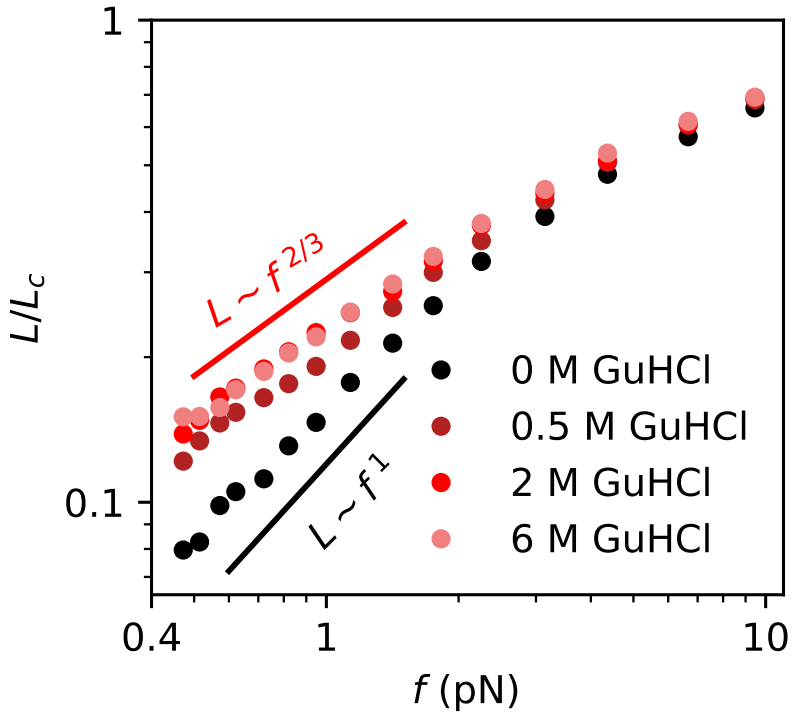

Figure 4.3: Force-extension curves of one NFLt polyprotein tether in different GuHCl concentrations. The extensions are normalized by the fitted L_c of each curve.

4.4 Discussions

Despite a relatively high NCPR (-0.23), the NFLt is quite compact, exhibiting a configuration that is close to an ideal chain rather than a swollen coil at low ionic strength. Depending on the point of reference, this result can be seen as either surprising or expected. On Das and Pappu's state diagram [67], NFLt is located in region 3 (Fig. 4.1D), where IDPs can adopt a range of coil behaviors from extended to semicompact, and the exact behavior is strongly dependent on their sequences [67]. In a study on the NFLt

characterized with SAXS and FRET, the Flory exponents of sub-sequence segments increase proportionally with their NCPR; a segment with a NCPR of -0.24 is found to have $\nu = 0.55$, roughly matching the result reported here [31]. Yet, previous experimental studies have found that other IDPs have more expanded conformations despite having lower NCPRs than the NFLt. The HIV integrase N-terminus (IN; measured with FRET; NCPR -0.07) shows $\nu = 0.58$ [27] and the pertactin N-terminus (PNt; measured with SAXS; NCPR -0.01) shows $\nu = 0.54$ [30], which are both greater than that of the NFLt. Further, prior magnetic tweezers studies on polyelectrolytes ssNA [94] and HA [104] observed swollen chain behavior ($\gamma \sim 0.66$) due to intramolecular electrostatic repulsion. While the NFLt also contains a significant number of negative charges, its Flory exponent instead indicates near-ideal chain behavior, where net intra-chain interactions are essentially zero. To counter the repulsive contribution from the charged residues, there must be a source of attractive interactions, possibly from transient secondary structures or disordered hydrophobic globules.

The presence of both attractive and repulsive interactions suggests a conformational heterogeneity. Indeed, the NFLt sequence shows two distinct domains: a C-terminus polyelectrolyte and an N-terminus polar yet largely uncharged (Fig. 4.1C). While the charged region is predicted to be swollen and extended, the polar tract domain is a compact globule (Fig. 4.4A). The combination of these two opposite domains gives rise to a global average conformation that is best described as an ideal chain. The conformational heterogeneity might also explain why IN is more expanded (larger Flory exponent) despite having a much lower NCPR. While it does not have as many charge groups, those residues are evenly distributed along its entire sequence, as opposed to being localized in one domain like the NFLt (Fig. 4.4B). Even though this likely results in weaker electrostatic repulsion, it is spread throughout the protein; thus, the overall intra-chain interaction is repulsive, and its global conformations is closer to a swollen coil.

The comparison of these proteins points to a significant limitation of IDP predictive state diagrams that are only based on simple sequence metrics. Although it is obvious that strong polyelectrolytes are swollen chains, they only account for 6% of known IDPs [12]. The vast majority of sequences falls in regions 1, 2, and 3, where their conformations exist in a broad disorder spectrum, and qualitative categorization is no longer possible without further sequence inspection. While the Flory exponent offers a very simple method to compare conformations in this continuum, because it is a single value, it cannot capture domain heterogeneity within a protein. For a complete clarification, this metric must be combined with other sequence calculations, such as charge distribution [67] or secondary structure propensity [144].

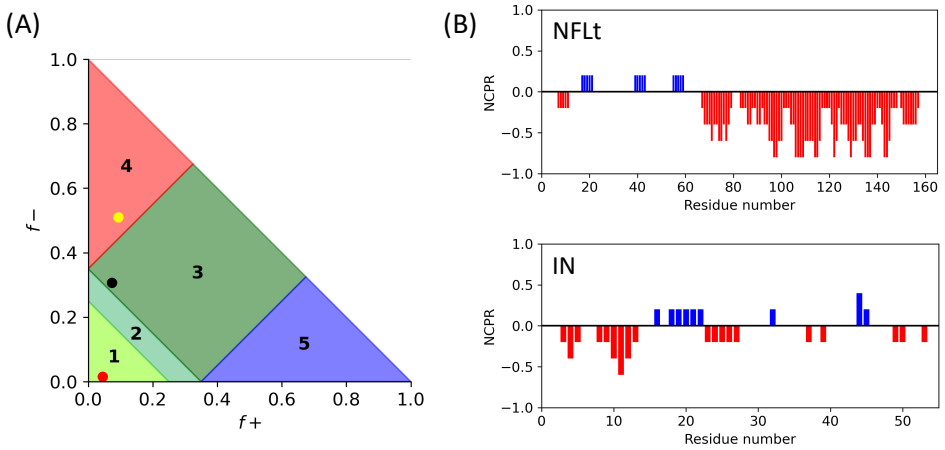

Figure 4.4: Sequence-driven conformational heterogeneity of the NFLt. (A) NFLt contains two domains with distinct conformations. The first 67 residues (red dot) is firmly in region 1 with expected globular behavior, while the remaining highly charged segment (yellow dot) is in region 4 where strong polyelectrolytes are swollen coils. (B) Charge profile comparison between NFLt and IN. The linear NCPR plot of IN is also calculated by the localCIDER package [147].

We examine the identity of NFLt intramolecular attractive interactions through its elastic response in denaturant. Unsurprisingly, NFLt polyproteins expands with GuHCl as the measured Flory exponent approaches the swollen chain expectation. This con-

firms the presence of hydrophobic globules or transient secondary structures, which are unraveled with improving solvent quality. Their existence has also been proven in a previous magnetic tweezers study on the NFLt, showing glassy dynamics arising from the stretching and collapsing of these globules [132].

We compare the expansion trajectory of the NFLt in denaturant to other IDPs and foldable proteins studied with different experimental techniques (Fig. 4.5). Overall, the results are in exceptional agreement, particularly at higher denaturant concentrations, where all proteins are essentially swollen chains in good solvent. The primary differences are at measurements in zero or low denaturant concentrations. As discussed above, despite having a relatively high NCPR, the NFLt is actually more compact than other less charged IDPs (IN, PNt) due to hydrophobic globules or transient secondary structures dominating its one domain. Instead, the NFLt's Flory exponent is comparable to that of the foldable spectrin domain R17 [27], highlighting the sizeable attractive contribution in the conformations of these proteins. At low denaturant concentrations (<0.5 M GuHCl), while two IDPs (ProTα and IN [133]) exhibit a collapse (reduction in Flory exponents), NFLt, PNt [30], and R17 [133] show a monotonic increase. ProTα is a highly charged protein (NCPR -0.39), where it expands to a rod-like conformation beyond the swollen chain description at low ionic strengths. Because GuHCl is ionic, it assumes both denaturing and electrostatic characteristics. At a low GuHCl concentration, its charge screening activity overshadows its good-solvent effects, and intramolecular electrostatic repulsion of the Protα is screened, driving compaction. Increasing GuHCl concentration causes denaturation to dominate, re-expanding the protein. While IN is less charged, its electrostatic repulsion is globally present, and the same phenomenon is observed. However, both PNt and R17 are close to being neutral (NCPR -0.01), and electrostatic interactions are not significant enough to elicit such response. While the NCPR of NFLt is only second to ProTα, it does not experience the non-monotonic expansion. This

suggests that while electrostatic repulsion exists in the NFLt, it is surprisingly not as defining as hydrophobic interactions.

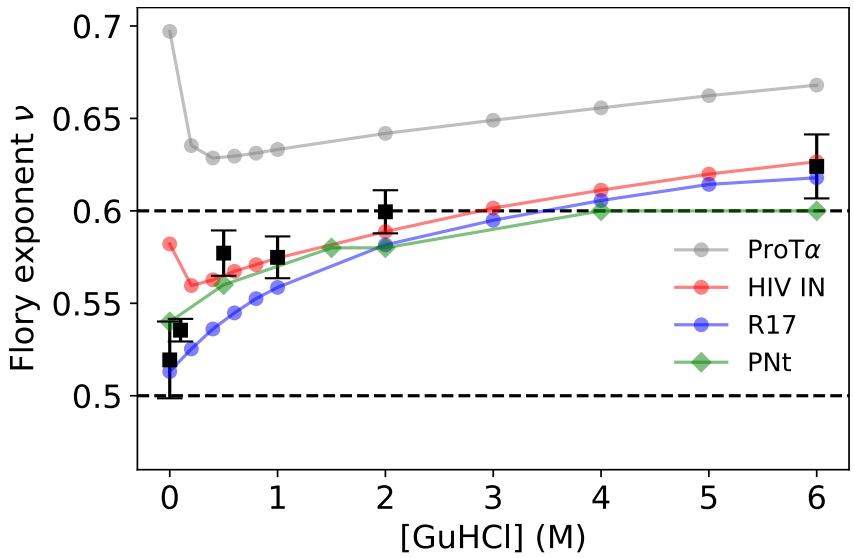

Figure 4.5: Flory exponents of the NFLt (black squares) and other IDPs (circles) in various denaturant concentrations. Each NFLt data point represents 1-11 measurements of separate tethers. The figure reuses and re-plots data from Refs. [27] and [30]. Error bars are standard deviations.

The experimental techniques and data analysis methods could explain the small discrepancies between our study and previously reported results. Notably, the extraction of Flory exponents from measurements differs among techniques. For SAXS and FRET, size measurements are applied to a model-dependent estimation of the exact scaling relationship $R = AN^{\nu}$. Magnetic tweezers experiments yield force-extension curves, which are then fitted to a power law derived from the Pincus blobs elasticity model [34, 35]. Preceding data analysis, the experimental design of each technique can lead to measurement bias. FRET uses bulky, hydrophobic fluorophores, which results in an underestimation of size [158]. For magnetic tweezers, while the Pincus blobs model assumes that the chain conformation below the tensile screening length ξ is unaffected by the applied force, a prior simulation study has shown small effects that could influence the Flory exponent estimation [159].

4.5 Conclusions and outlook

We demonstrate the use of single-molecule force experiments as a means to investigate the conformations of IDPs, specifically by allowing, from a single curve, estimates of both the Flory exponent (through analysis of the Pincus blob elastic regime at relative low force) and the persistence length (through analysis of the high-force elastic regime). Our specific results for the NFLt construct show that, despite a relatively high net charge, the chain shows a tendency towards compactness in aqueous buffer, as indicated by a Flory exponent that is comparable to that expected for an ideal chain. This compactness is removed with added denaturant, in good agreement with other studies, indicating the likely role of sequence-dependent self-attractions on chain conformation.

Comparison to different experimental techniques in prior IDP studies demonstrates the potential of magnetic tweezers as a tool for the quantification of IDP conformations. Entropic elasticity is a strategy fully orthogonal to SAXS and FRET measurements, and its robustness lies in its ability to reveal various structural parameters from well-established polymer stretching theories. This is a point where magnetic tweezers offer a unique advantage over other techniques. While the extraction of Flory exponents with SAXS and FRET involves choosing the appropriate polymer models, this process in magnetic tweezers experiments only requires a simple and model-free fitting. Beyond just reporting results on the NFLt, our study here describes a novel framework to characterize the conformational ensembles of IDPs through entropic elasticity measurements, an approach that relies only on polymer physics foundations to answer biophysics questions.

Further experiments would be helpful to fully elucidate the conformational behavior of the NFLt. To probe electrostatic repulsion in the C-terminus, force-extension measurements can be done in a series of salt concentrations (such as sodium chloride, NaCl). Alternatively, in a nonionic denaturant (urea), the elastic response is certain to be driven

explicitly by hydrophobic interactions. For meaningful conclusions about NFLt behavior in physiological conditions, results in 50-150 mM NaCl would be particularly interesting, as a previous study has reported varying NFLt gel phases with solution ionic strength [146]. For an in-depth investigation into the conformations of specific domains, one can design and measure various polyprotein constructs containing only portions of the full sequence; preliminary results in this direction are presented in Appendix D. Besides benefiting the understanding of the NFLt, these experiments could help further optimize the magnetic tweezers framework for future IDP studies.

Chapter 5

Effects of sequence charge spacing on polypeptoid conformation

5.1 Permissions and attributions

5.2 Introduction

Biopolymers, such as nucleic acids and proteins, encode their structure and functionality directly into their sequence. This has inspired the design of sequence-defined polymeric materials with engineered structural and functional complexity that approach those found in nature[160, 161, 162, 42]. The rational de novo design of such materials requires a fundamental understanding of how monomer sequence influences polymer con-

formation and structure. Specifically, electrostatic effects in polyelectrolytes have been broadly explored because they can drive structural formation and interactions with other molecules in the environment. Modulating a polyelectrolyte's charge sequence has been shown to significantly alter its conformational behavior [66, 32, 140], as well as its activity in many biophysical processes [163, 164]. For example, complex coacervates formed by chains with longer charge blocks are found to have higher critical salt concentrations [165, 166].

Single-molecule force experiments provide a unique perspective on the behavior of polymers by using applied tension to probe chain structure and interactions over a broad range of length scales [35]. At intermediate to high forces (2-100 pN), the polymer chain is stretched to near its contour length, L_c. This regime disallows loop formation, allowing intrachain interactions only among neighboring monomers; the resulting elastic behavior thus gives insight into local polymer flexibility [35]. At very low forces (< 1 pN), the polymer adopts "blob" configurations consisting of local self-avoiding random walks that are globally aligned with the applied force [34]. Prior single-molecule force experiments have exploited these low- and high-force elastic regimes to investigate the salt-dependent elasticity and structure of biological polyelectrolytes, such as single-stranded nucleic acids [95, 94] and hyaluronic acid[104]. While these studies demonstrate the insights generated through single-molecule elasticity experiments, the polymers that were studied do not allow modulation of key polyelectrolyte physical parameters, such as charge spacing.

Polypeptoids, or poly(N-substituted glycine)s, have emerged as a robust synthetic platform to explore the effect of sequence on chain conformation[45, 46] because polypeptoids with specific sequences can be synthesized with high efficiency and precision[71]. Polypeptoids are analogous to peptides, having the same peptide backbone, but with the side chains attached to the nitrogen instead of the α-carbon. This structure disallows the backbone-backbone hydrogen bonding prevalent in peptides, preventing secondary

structure formation and enabling side chains to be the primary driver of intramolecular interactions [73, 167].

Here, we use single-molecule force experiments to test the effects of charge spacing on polypeptoid conformation. To isolate electrostatic effects, we test simple polypeptoid sequences that contain only two types of monomers, either negatively charged or hydrophilic uncharged (Fig. 5.1). In addition to being testing models for electrostatic interactions in sequence-defined polymers, these polypeptoid designs can also provide new experimental insights to polyelectrolyte physics theories. Numerous studies have been done on ssNA [94, 119, 168, 169], HA [104, 169, 170], and other synthetic polymers [171, 172, 173] to verify theoretical predictions of polyelectrolyte behavior in solution. While the dependence on various external factors has been explored, these polyelectrolytes do not permit any alteration of intrinsic polymer parameters (charge spacing, density, distribution, etc.), limiting the potential scope of study. Polypeptoids open an opportunity to pose and solve new questions on polyelectrolyte physics, as their charged monomers can be arranged and modified with great precision. The polypeptoid sequences in our study serve as models for exploring how chain flexibility and solubility vary with charge spacing. For single-molecule stretching experiments, we use end-to-end click reactions to create long polypeptoid chains.

From the designed polypeptoids, we predict that a shorter charge spacing (or higher NCPR) leads to chain stiffening (larger l_p) and expansion (larger ν) due to stronger intramolecular electrostatic repulsion. Through force-extension measurements, we find conflicting evidence on electrostatic effects of the polypeptoid sequences. Our results show, surprisingly, that the polypeptoid persistence length does not vary with charge spacing or solution ionic strength. We further use low-force Pincus blob elasticity [34] to infer the solvent quality through the Flory exponent. We find a decrease in the Flory exponent with increasing solution ionic strength for all charge spacings, consistent with

an electrostatically-driven transition from good to theta solvent. Unlike the persistence length, the measured theta point is sensitive to the charge spacing of the polypeptoid sequence. We thus find two contrasting electrostatic effects, which we suggest can be reconciled considering the location of the charges on the side chains of the polymer. Similarly complex electrostatic behaviors could occur in other polyelectrolytes with a similar pendant charge structure, such as peptides.

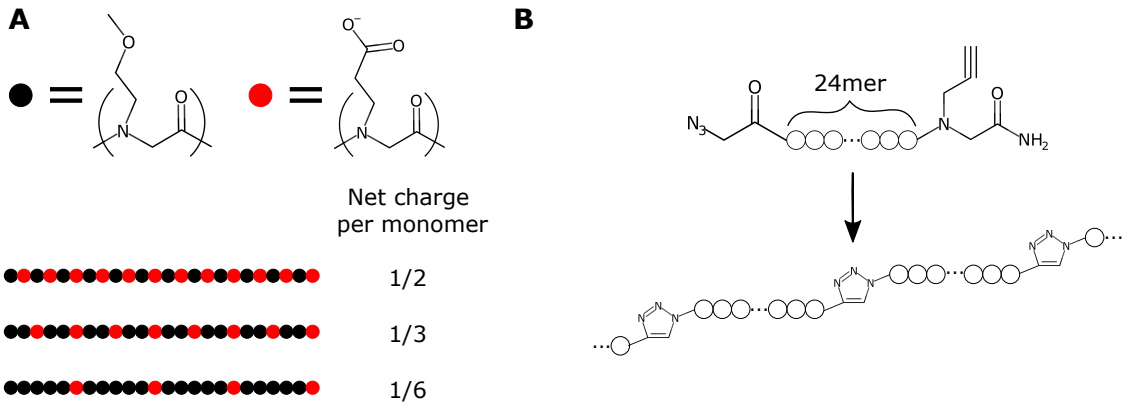

Figure 5.1: Polypeptoid design and construction. (A) Sequences are composed of two monomer identities: a negatively charged N-carboxyethyl (Nce) and neutral N-methoxyethyl (Nme) glycine. The three main sequences contain evenly spaced charges, resulting in their corresponding net charge per monomer. (B) Long polypeptoid construction scheme. Each 24mer unit includes terminal azide and alkyne groups, which enable end-to-end clicking to form long polypeptoid chains.

5.3 Experiments and results

5.3.1 Polypeptoid design and synthesis

The synthesis of long polypeptoids is discussed in detail in Appendix C. We investigated three polypeptoid sequences, each with equally-spaced charged monomers and a net charge per monomer, respectively, of 1/2, 1/3, and 1/6 (5.1). The sequences thus varied in both net charge and charge spacing, which was accomplished by inserting varying num-

bers of neutral monomers between the charged COOH groups. Both types of monomers have the same number of σ-bonds along the length of the side chains, resulting in similar steric dimensions. The hydrophilic nature of the neutral side chains is expected to only impart excluded volume repulsion to the polymer. The spacing of charges along the backbone (respectively $0.6, 0.9$ and 1.8 nm) are below the Debye screening length in low ionic strengths ($\sim$3 nm at 10 mM ionic strength), so intrachain electrostatic interactions were expected to be significant. We also synthesized a control polypeptoid sequence, consisting only of the neutral monomers, to investigate non-electrostatic conformational behaviors.

After synthesizing individual 24-mer polypeptoids, we quantified their charge state in solution through titration experiments. The measured pK_a of the carboxylic acid (COOH) groups within the 1/2, 1/3, and 1/6 sequences were, respectively, 5.51, 4.95, and 4.75. These values are similar to the reported pK_a of a glutamic acid protein residue ($\sim$4.25) [174], whose chemical structure is identical to the charged Nce peptoid side chain. Applying the Henderson-Hasselbach equation at the experimental pH of 7.4 indicates that charge groups in all sequences were $\sim$99% deprotonated.

We used a polypeptoid synthesis scheme designed to generate long polymer chains suitable for single-molecule force experiments. Magnetic tweezers experiments require chains whose contour lengths are at least a few hundred nanometers, i.e. at least 1,500 polypeptoid monomers, which far exceeds the practical limits of polypeptoid synthesis (well below 100mer) [175]. To generate long chains, we incorporated azide and alkyne terminal functional groups on 24mer polypeptoids with the desired sequences, then used end-to-end click reactions (Fig. 5.1) [73, 176]. This attachment scheme imposes unidirectionality on the sequences, ensuring even charge spacings across the entire length of the long polypeptoids. We chose 24mers for the individual polypeptoid units as a compromise between synthesis efficiency and continuous polypeptoid bonding length.

5.3.2 High-force measurements and analysis

Tethering protocol for polypeptoids is described in Appendix C. Low ionic strength measurements were done in 1-10 mM tris buffer pH 7.4 with 0.01-0.1% Tween-20. For higher ionic strengths, 10-1000 mM NaCl were included in addition to the said background buffer. We analyzed the conformations of the various polypeptoid sequences through single-molecule force-extension analysis. For all sequences and solution conditions, the measured force-extension curves showed a low-force power-law elastic regime and a high-force regime with reduced compliance (5.2). Guided by theoretical considerations and analogous results on other flexible polymers [35], we interpret the low-force regime as corresponding to the elasticity of a chain of tensile blobs, each of size k_BT/F, with looped, random-walk conformations within each blob [34]. At higher force, tensile-blob elasticity disappears when all loops have been pulled out, i.e. at a crossover force where the tensile length roughly equals the persistence length, $k_BT/F_c \sim l_p$. At higher forces, $F > k_BT/l_p$, loops are prohibited, and thermal fluctuations only modestly misalign the statistical monomers from the direction of applied force; the lack of significant misalignment means there is little extension left to gain, and the compliance decreases as the chain approaches its contour length.

The elastic behavior of the polypeptoids at high force is accurately described by the Marko-Siggia wormlike-chain (WLC) model [82]. Particularly, we fit the force-extension data in this regime to an analytical solution of the WLC model derived by Bouchiat[99] with two fitting parameters: contour length, L_c, and persistence length, l_p (5.2). Given that the average l_p is $\sim$1 nm, the crossover force is F_c $\sim$2 pN. Since the WLC model does not account for swelling (self-avoiding) interactions, we fit the model only for forces above 2 pN. This method is self-consistent, and the best-fit l_p values are not significantly sensitive to the cutoff choice. On each individual polypeptoid chain, several force-extension

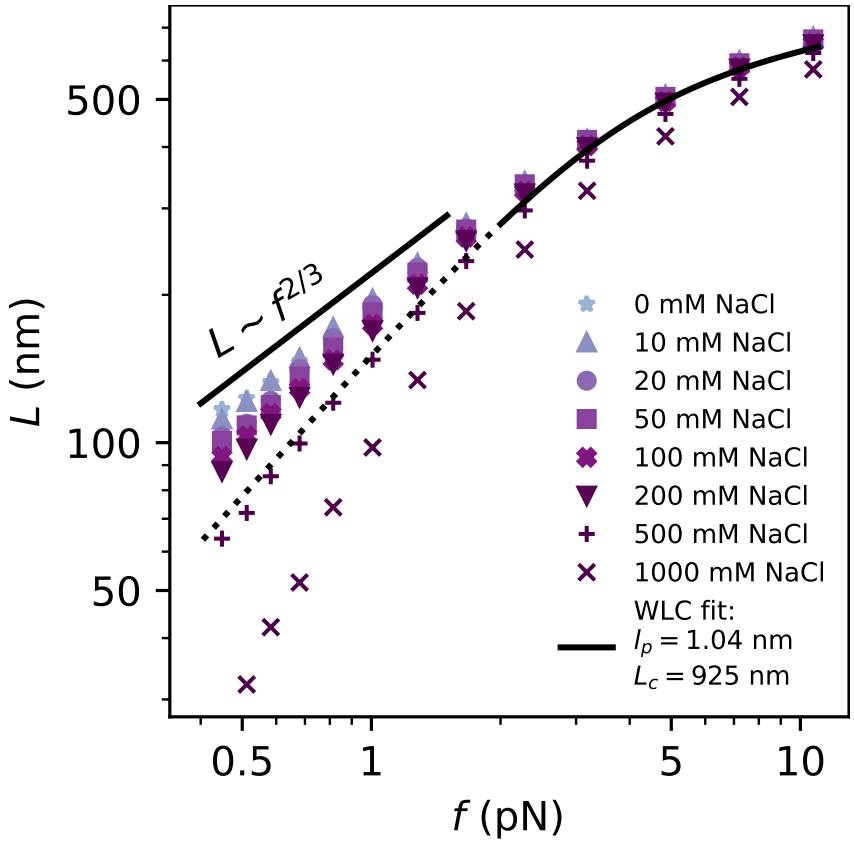

Figure 5.2: Force-extension curves of a single polypeptoid tether with a charge fraction of 1/2 in various NaCl concentrations. The WLC fit (black solid curve) is done for data above 2 pN. Low-force extrapolation of the WLC model (dotted black line) shows ideal chain prediction.

curves were acquired at different NaCl concentrations. Global WLC fitting was performed for each chain across all NaCl concentrations less than 200 mM, in which a single L_c was constrained for all such curves (since it is not expected to vary with salt), while l_p was fit individually for each curve. Goodness-of-fit metrics that account for the number of fit parameters indicate this global method is modestly better than individual fitting of both L_c and l_p to each force-extension curve; further, the resulting l_p values do not significantly deviate between methods. Elastic measurements at 200 mM and higher ionic strengths indicate a small contraction (decrease of L_c) of the chains in those conditions; such curves are excluded from the global fits, and instead fit individually. Such a contraction at high

Table 5.1: Persistence lengths, in nanometers, of polypeptoids of different charge fractions, in various solution ionic strengths, I. Each value is the average of 3-19 different chains, and the errors are the standard error of the mean. $I = 8$ mM data was obtained in 10 mM tris pH 7.4. $I > 8$ mM data was obtained in 1 mM tris pH 7.4, with NaCl added to reach the indicated ionic strength. Charge fraction 0 corresponds to the control chain containing only neutral monomers. The last row indicates results on a chain formed from 12mers; all other rows correspond to 24mers.

Charge fraction	$I = 8$ mM	11 mM	21 mM	51 mM	101 mM	201 mM	501 mM	1001 mM
1/2	1.05 ± 0.03	1.08±0.04	1.01±0.04	1.03±0.02	0.98±0.04	1.00±0.02	1.02±0.02	1.08±0.07
1/3	1.04 ± 0.02	1.10±0.03	1.06±0.06	1.10±0.03	1.02±0.02	1.03±0.04	1.08±0.01	–
1/6	1.05 ± 0.02	1.08±0.04	1.10±0.02	1.05±0.04	1.05±0.03	1.09±0.01	1.05±0.02	–
0	1.01 ± 0.03	–	–	–	–	–	–	–
1/3 (12mer)	1.06 ± 0.04	–	–	–	–	–	–	–

salt has been seen previously in other measurements [177], and has been attributed to salt-dependent water solvation effects. Measurements and fittings of various polypeptoid chains showed a range of contour lengths of 300-1000 nm, corresponding to chain sizes of 1000 to 3300 peptoid monomers, i.e. molecular weights of 125 to 410 kDa.

The persistence lengths of the polypeptoid sequences, as fit in the manner described above, are reported in Table 5.1. The measured values are similar to those observed in previous studies, which range between 0.5 and 1 nm[178, 167, 179]. However, surprisingly, our results show that l_p is independent of the chain net charge. In addition, the l_p of each sequence is insensitive to salt concentration(5.1). Further, in all conditions, the measured l_p values of the charged chains are not significantly different from that of the fully-neutral control chain (Table 5.1).

The polypeptoid chains studied here have a heterogeneous backbone chemistry, since the azide-alkyne click reactions form triazoles between every 24 polypeptoid monomers. Azide-alkyne cycloadditions have been widely explored in peptidomimetics, and the triazole product is comparable to a peptide bond in dimensionality, planarity, and polarity [180, 181]. Nonetheless, it is possible that the triazole linkers have significantly different conformational properties that could affect our measurements of chain flexibility. To examine this, we synthesized 12mer polypeptoids and assembled them into longer chains

through the same click reactions, thus creating polymers with twice the density of triazoles (Fig. 5.3). Elastic measurements showed that, with the same charge spacing, the 12mer polypeptoids do not show any significant change in l_p compared to the 24mer version (Table 5.1). This indicates that the measured elastic data are not significantly affected by the triazole linkers.

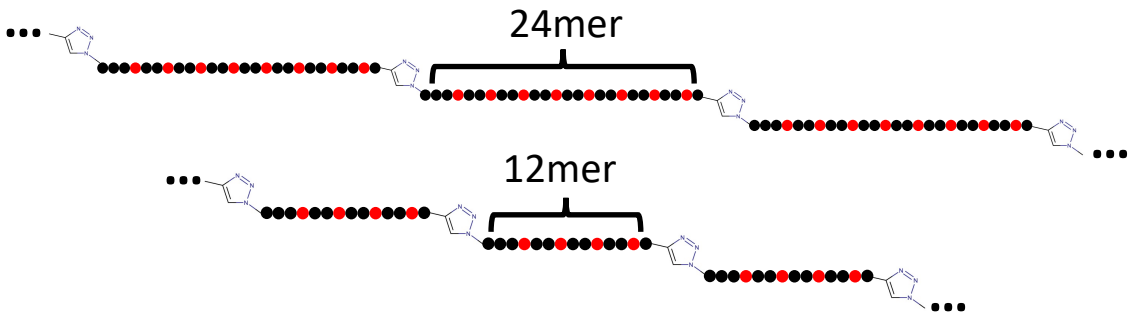

Figure 5.3: Control experiment that tests the effect of triazoles on chain flexibility. A long polypeptoid construct, made from 12mer units, contains more triazole linkers per monomer than that of a 24mer version.

5.3.3 Low-force measurements and analysis

At low force, the observed power-law elasticity indicates the polypeptoids can be described as a chain of tensile (Pincus) blobs, with solvent-quality-dependent random-walk behavior within each blob. Theory predicts that, in this regime, the extension scales with the force as $L \sim f^{1/\nu-1}$, where ν is the Flory exponent[34]. We find that the power law elasticity of polypeptoids is dependent on solution ionic strength (Fig. 5.2), indicating that ν (and thus the solvent quality) varies with salt. At low ionic strength, the elastic exponent approaches 2/3, consistent with good solvent behavior ($\nu = 3/5$). This result is expected for a polyelectrolyte with intramolecular electrostatic repulsion, which has been observed in previous ssNA [94] and HA [104] studies. Increasing the solution NaCl content to low-to-intermediate concentrations sees an unremarkable reduction of

the Flory exponent, suggesting that electrostatic repulsion is still dominant and the chain remains marginally swollen. At a certain NaCl concentration (500 mM for the 1/2 sequence), γ goes to 1 ($\nu = 0.5$), indicating that the polypeptoid is now an ideal chain, and this specific ionic strength is its theta solvent. This observation is corroborated by an independent WLC fitting on that force-extension curve: extrapolation of the WLC model at low force is approximately an ideal chain [82, 99], and the experimental data matches well with the prediction despite not being included in this fit (Fig. 5.2). Further increasing the NaCl concentration causes γ to extend beyond the ideal chain description, indicative of a compact globule in a poor solvent. Overall, these results are expected for a polyelectrolyte in solution, whose solubility decreases with increasing ionic strength due to electrostatic screening. The qualitative trend of γ with ionic strength observed here is very similar to that of ssNA and HA, where γ remains near 2/3 at low salt and sharply rises to 1 at their theta solvent concentration [94, 104].

We perform the same measurements for all charge sequences (1/2, 1/3, and 1/6) and compare their changes in Flory exponents to a wide range of ionic strength. However, for these experiments, most tethers do not survive all 8 solution changes. For data compilation, we include data from tethers with measurements in at least 3 different NaCl concentrations. The comparison is shown in Fig. 5.4, with each point representing 3-6 individual tethers and error bars showing standard error of the mean. At lower ionic strengths ($\leq$100 mM), while the Debye screening length is still greater than their charge spacings, all polypeptoid sequences are swollen chains due to intramolecular electrostatic repulsion. As expected, their Flory exponents then drop to $\sim$0.5 of an ideal chain and even further at higher ionic strengths due to charge screening. Particularly, the specific theta-solvent NaCl concentration is dependent on the polypeptoid charge spacing. For a sequence with a shorter distance between Nce monomers, the theta solvent is at a higher ionic strength, and vice versa.

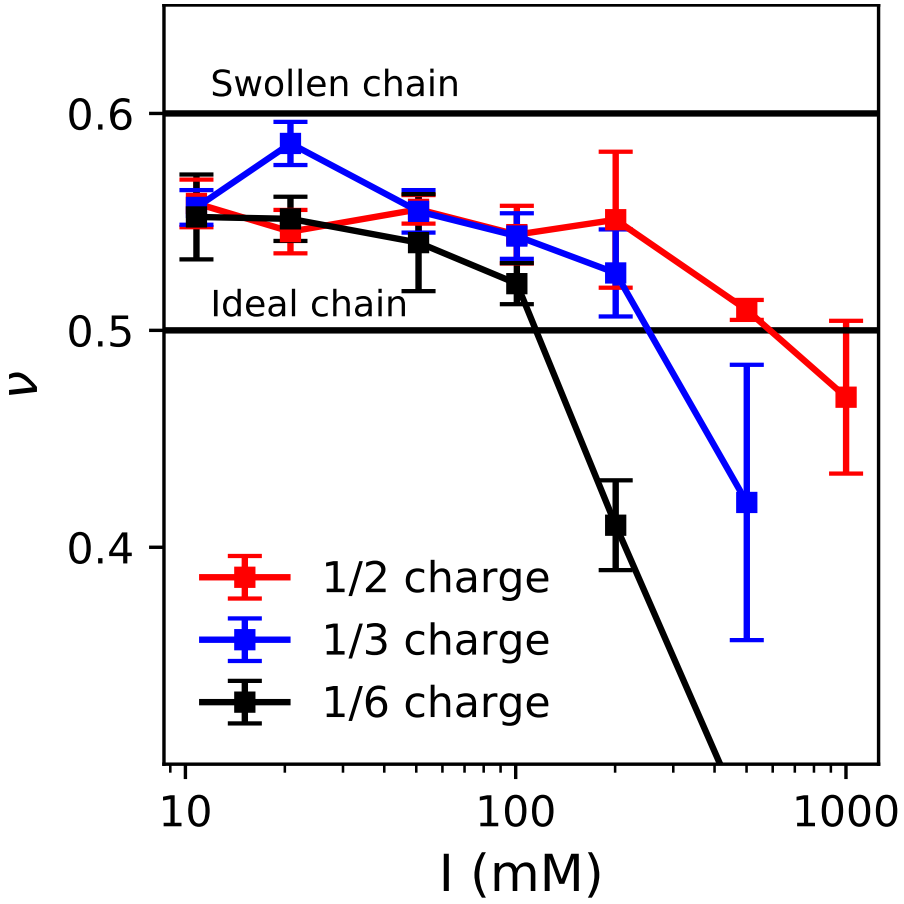

Figure 5.4: Changes in Flory exponents against increasing solution ionic strength of the three charged polypeptoid sequences. Each data point is compiled and averaged from individual tether measurements.

5.4 Discussions

Through careful sequence design and control experiments, we build a platform to investigate the effects of charge spacing on chain flexibility and conformation. Our results indicate that the effect of charge spacing in a polypeptoid chain is somewhat nuanced, displaying opposing evidence regarding the significance of electrostatic effects. On one hand, the charged groups show a clear effect in controlling solvent quality, with the collapse behavior depicted in Fig. 5.4 specifically showing an effect of intra-chain electrostatic screening. On the other hand, direct testing of conformational flexibility through

high-force elasticity measurements (Table 5.1) shows no effects of either charge pattern or ionic strength on l_p. Importantly, control measurements confirm that the chain is indeed charged in the tested conditions, and that the chemistry used to synthesize long chains does not affect the measurements; these controls rule out certain explanations that could reconcile the observations.

5.4.1 Polypeptoid flexibility

The flexibility result is unexpected because it contradicts many previous experimental findings and polyelectrolyte theories that would suggest otherwise. As predicted by Odijk [105] and Skolnick and Fixman [106] (OSF), the persistence length of a polyelectrolyte is a combination of intrinsic backbone stiffness l_p^0 and electrostatic repulsion component l_p^e, where the latter is strongly dependent on polymer charge spacing or solution ionic strength. Additionally, prior experimental studies on various polyelectrolytes using scattering techniques largely found some agreement with theoretical predictions [171, 167, 182, 172, 183]. Therefore, this discrepancy at first could be uniquely attributed to the use of magnetic tweezers, specifically the application of force on polymer chains. While the estimation of Flory exponents relies on elasticity at very low force ($f < 1$ pN), the calculation of l_p depends on moderate-to-high force ($f > 2$ pN) measurements. High force elasticity is generally a probe of shorter structural length scales in the polymer [35], and prior simulation studies of polyelectrolytes found that the structure factors at short length scales are independent of chain charge density [184, 185]. Further, a second potentially relevant influence is that the force itself can alter the short-range chain conformational flexibility [186, 187], possibly masking intrachain electrostatic effects. However, previous single-molecule studies on other polyelectrolytes, including ssNA [95, 94] and HA [104], were able to elucidate electrostatic effects in this same force range. The dif-

ference between the present data and these prior works does not rely on the specific analysis scheme used here (WLC fitting), but it is more general: specifically, single-stranded nucleic acids and hyaluronic acid showed salt-induced changes in the shape of the force-extension curves that are absent here. We can thus generally conclude: 1) elastic measurements in the moderate force regime indeed probe length scales at which charge affects chain structure, and therefore are capable of revealing electrostatic effects; and 2) that any distortion of the chain structure by the applied force is not significant enough to conceal such electrostatic effects.

We suggest that our observations on the opposing significance of electrostatic effects can be reconciled by considering the location of the charges in relation to the chain backbone. At first glance, the charge spacings of these polymers are quite similar: 0.7 nm in ssNA [188], 0.9 nm in HA [104, 189], and 0.6-1.8 nm in these polypeptoid sequences. However, the feature that stands out and separates polypeptoids from ssNA and HA is the location of charge groups relative to the polymer backbone (Fig. 5.5). In ssNA, charged phosphates are directly embedded in the sugar-phosphate chain. In HA, COOH groups are only 1 σ-bond away from the dissacharides. Meanwhile, for the charged polypeptoids in this study, COOH groups are on side chains 3 σ-bonds away from its backbone. Such minute distinction might account for the drastic difference in salt-dependent flexibility response. We refer back to the OSF theory, which states the two contributions on polyelectrolyte persistence length, l_p^0 and l_p^e. For ssNA and HA, the proximity of charge groups causes electrostatic repulsion to directly impact local flexibility in the form of l_p^e. Meanwhile, when charges are on side chains at a moderate distance away from the backbone, such as the case of polypeptoids, the surface charge density is reduced [190], and electrostatic repulsion is weakened to the point where l_p^e becomes infinitesimal compared to l_p^0. Another consequence of side chain structure is that charge spacing can no longer be simply defined on the one-dimensional axis along the backbone;

instead, the three-dimensional polymer configuration dictates the true distance between electrostatic groups. Notably, whereas the spacing of charged monomers along the backbone is 0.6 nm in the 1/2 charge sequence, the actual distance between charges could be as much as 0.9 nm, accounting for the side chain geometry. This increased distance would reduce the magnitude of repulsion between neighboring charged groups, thus minimizing the electrostatic effect on local chain flexibility (consistent with our results on l_p).

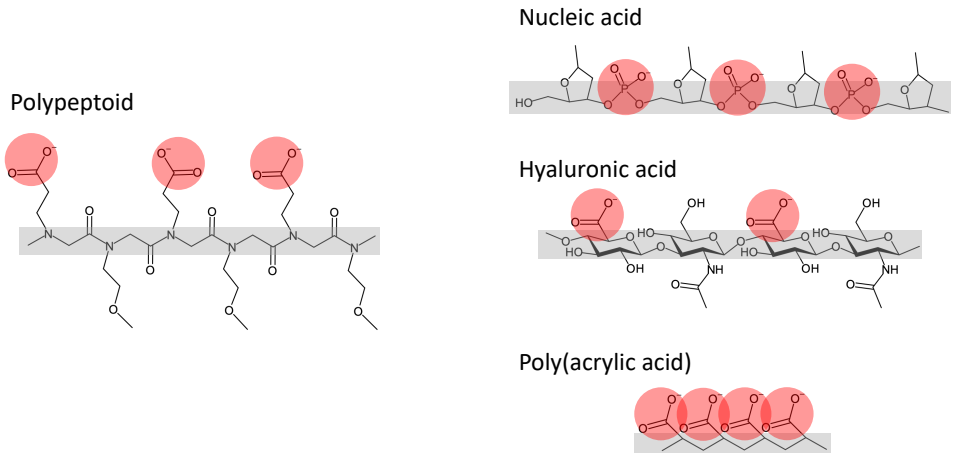

Figure 5.5: Comparison of chemical structures of various polyelectrolytes. Charge groups (red circles) are shown in relation to their polymer backbones (gray bars).

However, these arguments have yet accounted for the discrepancy between the observations here and a previously reported result on an almost identical system. Murnen et al. [167] used small angle neutron scattering (SANS) to measure the persistence length of two 36mer polypeptoid sequences: one containing only Nce monomers (NCPR of 1), and one with alternating Nce and Nme (NCPR of 0.5). The authors found that the fully charged polypeptoid has a higher l_p, and both sequences saw their flexibility reduced with increasing ionic strength. Several explanations can be made regarding our results and this study. While the chemical identities of the monomers are the same, the small differences in the sequences could be consequential. A polypeptoid where every monomer is charged might follow the expected behavior of a polyelectrolyte, but with reducing NCPR (longer

spacing) like the sequences in our work, electrostatic effects could quickly diminish until they are no longer observable. Further, results between the two studies might not be fully comparable due to the fundamental differences in experimental technique and data analysis. While our study applies WLC fitting on entropic elasticity measurements, Murnen et al. [167] relied on a polymer model to extract the persistence length from scattering functions. Their analysis resulted in relatively large errors that could confound the interpretation of any statistically significant trends related to electrostatic effects.

Prior works indeed show that the geometry of charged side chains in polyelectrolytes can influence their conformations [191, 192] and secondary structure propensities [193, 194, 195, 196, 190]. For example, one prior work noted that, when charged groups are at a further distance from the backbone, there is a weakening of the intramolecular electrostatic repulsion [190]. However, the effect of pendant charge specifically on polyelectrolyte flexibility has not been extensively explored and can only be inferred from separate studies. Poly(acrylic acid), whose COOH groups are 1 σ-bond from its backbone, shows salt-dependent persistence length as predicted in polyelectrolyte theories [171]. Yet, conjugated polyelectrolytes with pendant imidazole rings at 6 σ-bonds away exhibit no effects of charge spacing on their persistence length [197]. To our knowledge, only the computational study of Ghelichi and Eikerling [191] has systematically discussed the effect of charged side chain length on the persistence length. They found that the persistence length decreases with side chain length at lower charge fractions; yet, the effect is opposite for a fully charged polyelectrolyte. However, the choice of polymer model and simulation conditions in that study prevents direct comparison with our results. Generally, future experimental works observing polyelectrolyte flexibility as a function of side chain geometry would be highly valuable.

5.4.2 Chain conformation and solvent quality

We initially predict that a shorter charge spacing would result in a greater chain extent (i.e. a higher Flory exponent ν) due to increasing electrostatic repulsion strength. Yet, Pincus blobs analysis of force-extension measurements reveals no discernible difference in ν among the 3 charge sequences at low ionic strength. While the Debye length is still longer than the charge spacings, electrostatic repulsion remains the dominant intra-chain interaction that dictates global conformation. Additionally, the neutral spacer Nme is hydrophilic and can contribute excluded volume effects. Taken together, charge spacing does not have much influence on polymer conformation at low ionic strength, as the polypeptoids are all swollen chains in good solvent.

The primary observation that distinguishes the three polypeptoid sequences is the transition of solvent quality with increasing solution ionic strength. At higher salts, intra-chain electrostatic repulsion is screened out, inducing a compaction from swollen to ideal chain and below. Charge spacing b dictates this progression through its relation to ionic strength I, which is quantified by the Debye length κ^{-1}. This length scale denotes the exponential decay rate of electrostatic interactions; physically, it also tells the effective range in which electrostatic interactions are significant [198]. The Debye length is related to solution ionic strength inversely as $\kappa^{-1} \sim I^{-1/2}$. At low I, intra-chain electrostatic repulsion remains pronounced as long as $\kappa^{-1} > b$. With increasing salt, as κ^{-1} is reduced to below b, charge entities are essentially "invisible" to one another. A shorter charge spacing therefore requires a higher ionic strength to screen out its repulsion. Correspondingly, the theta solvent, where ν is 0.5 of an ideal chain, is at a higher salt concentration for smaller b ($\sim$500 mM NaCl for the 1/2 sequence and $\sim$100 mM NaCl for the 1/6 sequence). Beyond this analysis, this result also demonstrates that magnetic tweezers can capture nuanced electrostatic effects in the polypeptoid model,

as the absence of flexibility response casts some doubt on the effectiveness of force as a structural probe.

The dependence of solvent quality transition on charge spacing suggests a unified view of the data might result through an analysis that accounts for the relative length scales of the charge density (charge spacing b) and the screening (Debye screening length κ^{-1}). We estimate b by taking the inter-residue spacing to be 0.3 nm, which is similar to the range of reported spacings of peptoid and peptide residues (0.3-0.4 nm) [199, 200, 201, 202]. Given that peptoids have a higher preference for the cis conformation than peptides [76], and since the cis conformation shortens the residue spacing, we opted for a spacing on the low side of the range; however, our results are not significantly sensitive to this choice. After normalization, all three sequences collapse on a single curve, showing the general Flory exponent response to solution-dependent electrostatic strength. As explained above, when $\kappa^{-1} > b$, the polypeptoids are swollen chains in good solvent, and when $\kappa^{-1} > b$, they are collapsed in poor solvent. Particularly, the theta state occurs when the Debye length is roughly equal to the charge spacing ($\kappa b = 1$), setting a boundary between the two chain conformation and solvent quality states. These observations overall demonstrate that electrostatic repulsion does indeed dominate the conformational behavior of the polypeptoids, and charge spacing dictates their behaviors in solution.

5.4.3 Reconciling the two observations

The two opposing views on chain flexibility and solubility can be reconciled by examining the chemical structure of polypeptoids. Because charged COOH groups are positioned on side chains at a moderate distance from the backbone, their repulsive interactions do not have any noticeable influence on local flexibility, especially when monomers are stretched out at high forces. However, their presence can regulate solubility, which is

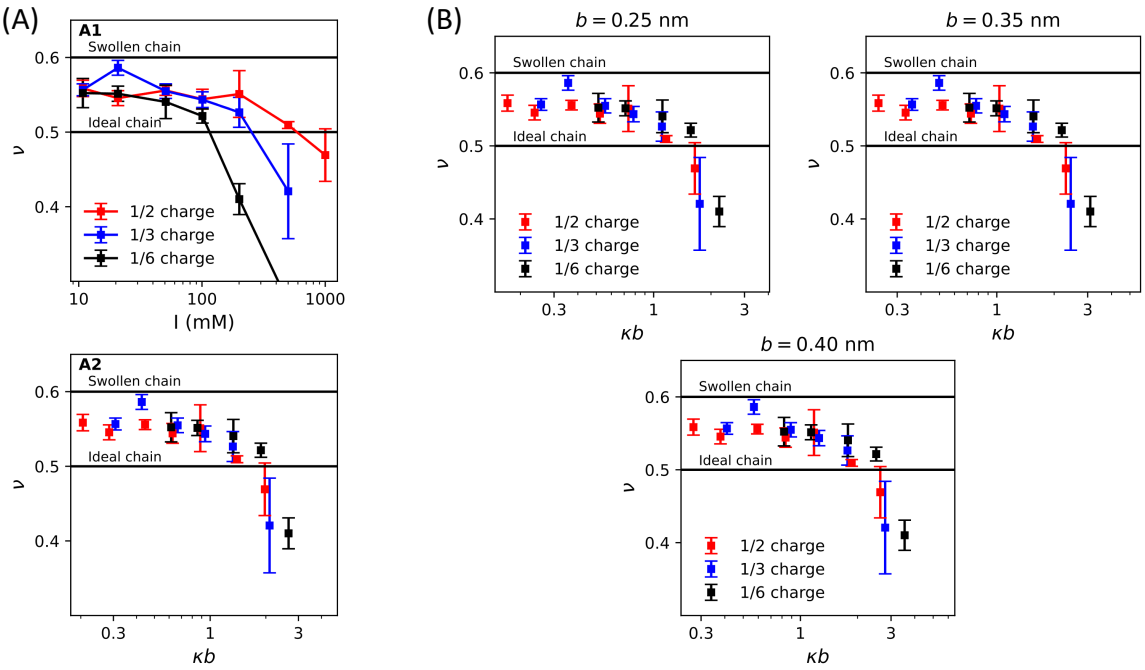

Figure 5.6: Generalization of solvent quality transition for all charge spacings b. (A) The variation of Flory exponents ν by solution ionic strength I presented in Fig. 5.4 (shown here as A1) can be normalized by the unitless ratio between charge spacing b and the Debye screening length κ^{-1} (A2). (B) Estimation choice of b does not significantly affect where the point of theta solvent ($\nu = 0.5$) falls, which is in the range of κb = 1-3.

detected by low-force measurements where long-range interactions persist. The unique electrostatic behavior reported here can also be applied to other sequence-defined polymers, such as polypeptides, as well as synthetic polyelectrolytes with pendant charges. Further, our results call for predictive models that can better describe the flexibility of polymers with this type of structure. While previous theories model polyelectrolytes as chains of charged beads, this simplistic sketch would not accurately reflect the complications of having side chains.

5.5 Conclusions and outlook

We utilize polypeptoids as a platform to investigate the effects of charge spacing on the conformational behavior of sequence-defined polymers. Through magnetic tweezers measurements on simple polypeptoid charge sequences, we find two seemingly contradictory results on the electrostatic influence. Chain flexibility, quantified by the persistence length, is surprisingly independent of both charge spacing and ionic strength. Meanwhile, the polypeptoid solubility, evidenced by the Flory exponent, shows a progression from good to theta solvent in increasing salt concentration due to electrostatic screening. This behavior can be normalized by the relationship between the Debye length and charge spacing, indicating that electrostatic repulsion plays a dominant role in the chain conformation. We connect these two observations by referring to the chemical structure of polypeptoids, where charges are located on side chains and separated from the backbone. Because of this feature, while electrostatic repulsion cannot meaningfully impact local flexibility at such distance, its long-range presence still dictates chain conformation and solubility.

Our observations with the polypeptoids raise a new question on the influence of pendant charges on polyelectrolyte flexibility. While various studies have hinted at a solution, a comprehensive experimental investigation on this effect is in need. Polypeptoids can be once again exploited for such study owing to its precise sequence synthesis, ensuring constant charge spacing along the entire polymer chain. Modulating side chain length is simply done by selecting the appropriate submonomer primary amine during solid-phase synthesis. On the theoretical side, predictive models that explicitly account for how charge spacing can fluctuate due to side chain length and backbone rotation will be particularly useful.

Our study has laid out a new framework on exploiting single-molecule entropic elas-

ticity measurements to probe the effects of monomer arrangements on sequence-defined polymers. Various polypeptoid sequences with different monomer identities can be synthesized precisely with the solid-phase platform. While the requirement of long polypeptoid constructs is a formidable chemistry challenge, it also allows robust assessment and validation of polymer physics theories, which typically assume infinitely long chains. Beyond expanding the knowledge of sequence-defined polymers, this platform can also serve any investigation on the effects of amino acid side chains on protein conformations and dynamics.

Chapter 6

Summary and outlook

6.1 Summary

This dissertation demonstrates the application of magnetic tweezers as a novel approach to the characterization of sequence-defined polymers. Typically, past studies on these systems rely on FRET or scattering techniques, where conformational analysis is done from size measurements. Magnetic tweezers offer an orthogonal strategy to this, where polymer structures on a wide range of length scales can be probed through force-extension measurements at single-molecule precision. Entropic elasticity can be described by various polymer physics theories, which reveal important structural parameters such as the persistence length (from WLC fitting) and the Flory exponent (from Pincus blobs model). We develop a framework for this approach through studies on an intrinsically disordered protein model and simple charged polypeptoid sequences, uncovering predicted and unexpected conformational behaviors in both cases.

Our study on the disordered NFLt indicates that magnetic tweezers can complement other characterization techniques in the study of IDP conformation. Elucidating the unstable, rapidly fluctuating conformations of IDPs has been a major challenge in

protein biophysics. Polymer physics offer an intuitive solution to this problem through descriptions of conformational ensembles using simple scaling theories. Commonly used techniques such as SAXS and FRET follow this approach through a model-dependent conversion of size measurements into normalized structural parameters. Magnetic tweezers instead allow a straightforward extraction of these parameters from force-extension curves, which is demonstrated through entropic elasticity measurements of the NFLt. We found that the NFLt is surprisingly compact despite its relatively high NCPR, acting like an ideal chain at low ionic strength. This behavior is attributed to the presence of local hydrophobic globules or transient secondary structures, which is proven through the observed expansion in denaturant, a good solvent. Flory exponents from magnetic tweezers measurements were found to be in good agreement those from other characterization techniques. This illustrates the potential of magnetic tweezers and entropic elasticity as a novel approach to the study of IDP conformational ensembles. While SAXS and FRET require *a priori* assumptions, such as bond lengths and distributions of conformational states in an ensemble, the polymer physics of entropic elasticity are more simple and do not need such inputs. Nevertheless, while SAXS and FRET can probe proteins as is, magnetic tweezers require long polymer constructs, which: 1) is a significant biochemistry challenge; and 2) raises the question of relating the structural parameters of polyproteins to their original unit sequences. All these techniques combined, with their advantages and drawbacks, can provide a holistic view of IDP conformation ensembles and accelerate progress in this field.

We applied the magnetic tweezers framework to investigate the sequence-conformation relationship in polypeptoids. Polypeptoids emerged as promising platform for sophisticated nanomaterials due to their precise solid-phase sequence synthesis. The rational design of such materials is aided by a thorough understanding of how the arrangements of monomers influence their conformational behaviors. To this end, we probed the ef-

fects of charge spacing on chain flexibility and solvent quality through force-extension measurements of simple polypeptoid sequences. We found that the flexibility is unexpectedly independent of both charge spacing and ionic strength, contradicting polyelectrolyte theories and previously reported experimental results. Charge spacing dictates the good-to-theta solvent quality transition in increasing ionic strength, showing that electrostatic repulsion is the dominant intra-chain interaction of the polypeptoids in solution. The seemingly opposing results on electrostatic influence can be reconciled by examining the chemical structures of polypeptoids, where charge groups are pendant. This conclusion also raises new research questions in polyelectrolyte physics regarding electrostatic interactions in sequence-defined polymers.

In summary, we perform high-precision force-extension measurements on an IDP model and polypeptoid constructs using magnetic tweezers. Through polymer physics analysis of entropic elasticity, we obtain important structural parameters, revealing both expected and surprising conformational behaviors. Our findings provide new insights and guide future directions for the study of IDPs, sequence-defined polymers, and polyelectrolytes. More generally, we lay the foundations to a novel framework of exploiting entropic elasticity measurements using magnetic tweezers to characterize the conformations of these challenging polymer systems.

6.2 Outlook

We propose several explanations for the unexpected conformational behaviors of the NFLt and the charged polypeptoid sequences. Having follow-up experiments would help verify these hypotheses.

We suggest that the global-average conformation of the NFLt in low ionic strength is an ideal chain due to local hydrophobic attractions competing against electrostatic

repulsion. Besides GuHCl, force-extension measurements can be performed in other solution conditions to quantify the strength of each type of intramolecular interaction. NaCl can screen electrostatic repulsion in the C-terminus; whether or not the global chain collapses will reveal the dominance of this domain. Urea, a nonionic denaturant, can specifically target the uncharged domain without contributing any potential electrostatic effects. Measurements in urea can resolve the cause of compaction at low GuHCl concentrations (or lack thereof) observed in other IDPs that is attributed to the dual electrostatic-denaturant effect of GuHCl [32, 27].

For a more in-depth investigation of local conformations, various polyprotein constructs containing parts of the full NFLt sequence can be designed and measured. To this end, we have designed a few variants and received a preliminary force-extension result. Details on this is discussed in Appendix D.

We interpret the lack of electrostatic influence on polypeptoid flexibility to stem from its pendant charge structure. This hypothesis can be confirmed through designing and measuring various polypeptoid constructs with differing side chain lengths. The vast commercial availability of primary amines opens up the feasibility of numerous sequence designs. As a note, the side chain length of the neutral spacer should also vary to match the steric dimension of the charged one. Possible designs are shown in Fig. 6.1.

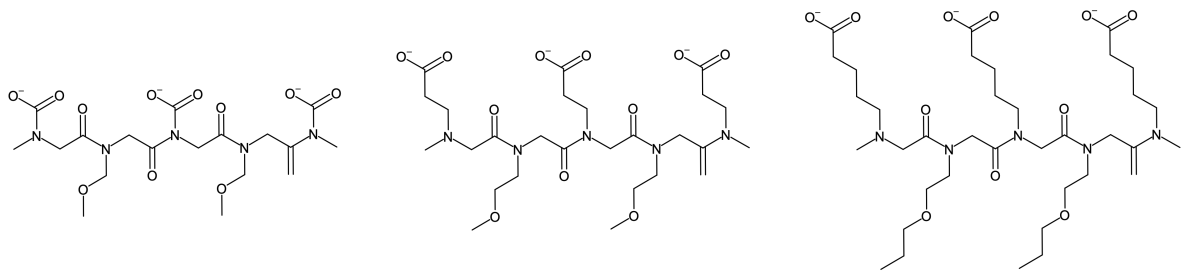

Figure 6.1: Polypeptoid constructs for testing pendant charge effects on chain flexibility. The middle structure is the 1/2 sequence in this dissertation. Primary amines to synthesize these side chains are readily available commercially.

To address the discrepancy between our observation and the results reported by Murnen et al. [167], force-extension measurements could be done on a polypeptoid design with only Nce monomers (NCPR of 1). We initially attempted to create this construct but faced issues with the solid-phase synthesis step. There are several strategies to improve the synthesis, such as increasing reagent concentration or increasing reaction time.

For both studies on the NFLt and the polypeptoids, the major bottleneck that prevents high data throughput is the synthesis of long polymer constructs. Disulfide bridging for IDPs does not have specific reaction conditions, and it is therefore very spontaneous and uncontrollable. This linking chemistry also limits which chemistry can be used for attachment and measurement, as it is susceptible to reducing agents. Polypeptoids, owing to their synthetic nature, can offer more options for linking reactions. However, building long chains of >1000mers from <50mers remains a formidable chemistry challenge, requiring many successive reactions with diminishing efficiency. For these reasons, the concentration of long polymer chains with the correct terminal labeling can be very low even for single-molecule experiment standard.

More advanced approaches involving protein biochemistry expertise might be necessary for synthesizing long polyprotein constructs. One strategy is to incorporate unnatural amino acids at the termini that can enable click chemistry [203]. Alternatively, a polyprotein can be expressed in its entirety without the need of heterogeneous linking, such as reported in atomic force microscopy studies [133].

Other click chemistry routes can be explored with polypeptoids. Thiol-alkene reaction is one option that we initially explored. Lengthening the polypeptoid unit from 24mer to 40-50mer can lower the number of click events needed to form longer chains; while solid-phase synthesis efficiency will reduce, running a larger reaction batch or multiple batches can increase the amount of raw material for the click chemistry. A longer unit also increases the space for designing sequence with higher complexity.

In another approach to increase data throughput, instead of striving for long polymer constructs, one can make improvements to magnetic tweezers experiments to allow the use of shorter polymer chains (< 100 nm in contour length). Beads can have problematic sticking interactions, which are exacerbated when they are much closer to the surface. Further, force calibration and extension measurements are susceptible to larger errors for very short tethers [204]. These issues can be addressed experimentally by enhancing the passivation chemistry of functionalized glass coverslips. On the instrument end, several studies have modified the data analysis for more accurate estimations of force and extension of short DNA tethers [204, 131], which can also be applied to proteins, peptides, and peptoids.

Our work on the NFLt establishes a platform for characterizing IDP conformational ensembles through entropic elasticity. To expand upon this framework, we can synthesize and measure other IDP models. Tau protein [205] and α-synuclein (αS) [206] are promising candidates as they have been extensively studied in a wealth of literature. Similar to the NFLt, tau and αS are heterogenous in conformation, consisting of regions with different properties [207, 208, 209]. Experimental strategies that have been applied to the NFLt, such as changing solvent conditions or isolating domains, can be applied for these other IDP models to form a complete picture of their conformational behaviors.

Besides biological proteins, experiments on simple polypeptide sequences can reveal insights into the effect of residue arrangements in a peptide backbone context. Even more rudimentary, homo-polypeptides can be a valuable model to examine polymer physics models of proteins and peptides. Long polypeptide constructs can be generated in a similar manner as polypeptoids. Polypeptide units can be purchased from a custom peptide synthesis service (Thermo Fisher Scientific, GenScript), which typically utilizes solid-phase chemistry. Terminal modification can be added, enabling end-to-end click reactions to form long chains. We describe in Appendix D a preliminary attempt on

forming long polyglycines for magnetic tweezers experiments.

Our work on polypeptoids creates an outline for exploring other sequence effects beyond just electrostatic repulsion. Polyampholytes are particularly interesting as the competition between electrostatic repulsion and attraction gives rise to rich conformational and self-assembling behaviors [210, 211]. Further, natural protein sequences are much more comparable to polyampholytes than polyelectrolytes, making studies on them more translatable to real biological systems [212]. Previous studies have demonstrated that the patterning of opposite charges influences chain conformation [67, 213], which can also be examined through the entropic elasticity approach. Another direction to pursue is on the effects of hydrophobic patterning, which can provide new insights to the influence of side chain interactions on the globule collapse in protein folding. While this has been done in a prior work using scattering [73], it would nevertheless be interesting to study this topic using magnetic tweezers on polypeptoids with varying net hydropathy and hydrophobic patching.

Appendix A

Abbreviations

AV	Allan variance
BSA	bovine serum albumin
CMOS	complementary metal–oxide–semiconductor
COOH	carboxylic acid
$CuSO_4$	copper sulfate
DBCO	dibenzocyclooctyne
DIC	N, N'-diisopropylcarbodiimide
DLS	dynamic light scattering
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
dsDNA	double-stranded deoxyribonucleic acid
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
FRET	fluorescence resonance energy transfer
GPU	graphics processing unit
GuHCl	guanidinium hydrochloride
HA	hyaluronic acid

HPLC	high-performance liquid chromatography
IC	information criterion
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
IN	HIV Integrase N-terminus
KOH	potassium hydroxide
LiCl	lithium chloride
LiI	lithium iodide
MALDI-TOF	matrix-assisted laser desorption/ionization - time of flight
MES	2-(N-morpholino)ethanesulfonic acid
MLE	maximum likelihood estimation
MW	molecular weight
NaCl	sodium chloride
NaOH	sodium hydroxide
Nce	N-carboxyethylamine
NCPR	net charge per residue
NdFeB	neodymium iron boron
NFLt	neurofilament low molecular weight subunit protein tail region
NHS	N-hydroxysuccinimide
Nme	2-methoxyethylamine
OSF	Odijk, Skolnick and Fixman
PBS	phosphate-buffered saline
PEG	poly(ethylene glycol)
PG	polyglycine
PI	Physik Instrumente

PNt	pertactin N-terminus
poly(dA)	polydeoxyadenylic acid
ProTα	prothymosin-alpha
R17	spectrin domain R17
SAXS	small-angle X-ray scattering
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
ssNA	single-stranded nucleic acid
TBTA	tris((1-benzyl-4-triazolyl)methyl)amine
TFA	trifluoroacetic acid
TIPS	triisopropylsilane
UV-Vis	ultraviolet–visible spectrophotometry
WLC	worm-like chain

Appendix B

Experimental protocols for the NFLt study

B.1 Polyprotein construct

A 163-residue NFLt construct was synthesized through recombinant protein expression in *E. coli* following the exact method described in Refs. [132, 214]. After cleavage and purification, the final sequence is:

MGCGSGASEN LYFQGASTRL SFTSVGSITS GYSQSSQVFG RSAYSGLQSS SYLMSARSFP
AYYNSHVQEE QTEVEETIEA TKAEEAKDEP PSEGEAEEEE KEKEEGEEEE GAEEEEAAKD
ESEDTKEEEE GGEGEEEDTK ESEEEEKKEE SAGEEQVGGG SCG

The terminal cysteines (C) for disulfide bridging are highlighted in red. The added glycines (G) and serines (S) around the cysteine residues (`MGCGSG` in the N-terminus and `GGGSCG` in the C-terminus) increase terminal flexibility and hydrophilicity for linking with other protein units. The protocols for polymerization and terminal labeling were designed and documented in Refs. [132, 214]. In short, in a solution of fully reduced NFLt units, maleimide-biotin and maleimide-azide were added to cap the cysteine residues. The amount of caps were set at 100x less than the concentration of NFLt units to prevent excess labeling that would inhibit disulfide bridging. The caps were also added prior to polymerization so that long chains could stay linear instead of forming circular self-links. The NFLt units with the appropriate amount of caps were allowed to polymerize for one week. The presence of long polyprotein constructs were verified through sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE), showing chains at >300 kDa in molecular weight (MW). Polymerized samples were purified through an Amicon® Ultra Centrifugal filter with a MW cutoff of 100 kDa to remove any short polyproteins and excess reagents.

B.2 Polyprotein attachments for magnetic tweezers experiments

An NFLt experiment began with an NHS-functionalized flow cell (construction protocol described in Chapter 3). A 40 μL solution of DBCO-amine (10 mg/mL, dissolved in DMSO) and PBS pH 8 in a 1:1 ratio was flowed into the flow channel and let incubate for 1 hour. Throughout all rinsing and incubation steps, the flow cell was kept in an enclosed and humidified chamber to prevent dust and liquid evaporation. The flow cell was then washed with 400 μL 10 mM tris buffer pH 7.4 with 0.1% Tween-20 and let

sit for 15 minutes to deactivate any remaining unreacted NHS sites. A 400 μL of 10 mM MES buffer pH 6.8 with 0.1% Tween-20 was used to rinse the flow cell, setting the condition for polyprotein immobilization.

From a stock solution of $\sim$2 μM NFLt polyprotein, a 1 μL aliquot was diluted in 500 μL 10 mM carbonate buffer pH 11 and rotated at room temperature for 1 hour to disrupt any aggregations. The solution was spun through a 100 kDa Amicon® Ultra Centrifugal filter and exchanged with 10 mM MES buffer pH 6.8 with 0.1% Tween-20. The recovered concentrate ($\sim$19 μL, $\sim$200 nM) was added directly into the flow cell and let incubate for 12-24 hours at 4°C.

After incubation, unbound polyproteins were rinsed out with 800 μL 10 mM MES buffer pH 6.8 with 0.1% Tween-20. A 200 μL passivating solution of 10 mM MES buffer pH 6.8 with 1% BSA and 0.1% Tween-20 was flowed in and let sit for 30 minutes. After that, the flow cell was thoroughly washed with 800 μL 10 mM MES buffer pH 6.8 with 0.1% Tween-20.

A 0.5 mg/mL solution of Dynabeads™ MyOne™ Streptavidin C1 magnetic beads (2 μL of stock diluted in a final volume of 40 μL) was prepared by thoroughly washing twice in 10 mM MES buffer pH 6.8 with 1% BSA and 0.1% Tween-20 and resuspending in 10 mM MES buffer pH 6.8 with 0.1% Tween-20. Prior to adding beads, the flow cell was placed on the magnetic tweezers. A 10 μL from the magnetic beads solution was flowed in and let incubate for 10 minutes. Beads should cover the glass surface densely to maximize tethering chance, and the coverage could be visualized with the microscope. Very small drops of buffer was added intermittently to promote bead diffusion across the entire surface. After that, excess beads were thoroughly washed out with $>$ 1000 μL 10 mM MES buffer pH 6.8 with 0.1% Tween-20.

Appendix C

Experimental protocols for the polypeptoid study

C.1 24mer polypeptoid synthesis

24mer polypeptoid units with desired sequences were synthesized in a solid-phase peptide synthesizer (Prelude, Gyros Protein Technologies) following a stepwise reaction scheme discussed in Refs. [167, 215]. For a 50 μM reaction scale, 92.59 mg of Rink amide resin (0.54 mmol/g, Novabiochem, San Diego, CA) was used in each reaction vessel. The submonomer primary amines were 2-methoxyethylamine (Nme) and β-alanine t-butyl ester HCl N-carboxyethylamine (Nce). Reagents and their respective concentrations (all in DMF solvent) were:

primary amines	1-1.5 M
4-methylpiperidine	20% v/v
bromoacetic acid	0.6 M
N,N'diisopropylcarbodiimide (DIC)	59% v/v

Prior to use, Nce was freebased with aqueous potassium hydroxide (KOH) and extracted in an ethyl acetate phase to remove the chloride counterion. The typical yield of freebasing is ∼80% accounting for the MW reduction. In one example protocol, 8 g of Nce was dissolved in 10 mL water and the solution was cooled on an ice bath. A solution of 2.5 mg KOH (1 mol equivalent) dissolved in 5 mL water was also prepared. The KOH solution was added dropwise to the chilled Nce solution while stirring. In a separatory funnel, the freebased Nce was extracted from water through 3 washes of ethyl acetate. The collected fractions were dried with anhydrous magnesium sulfate, and ethyl acetate was removed in a rotary evaporator. This protocol yielded 5 g of clear to lightly yellow liquid. If the freebased Nce was stored at -20°C a day prior to the synthesis, the liquid might be hazy due to unknown precipitation. The Nce should then be centrifuged to separate the clear liquid from the precipitation. Depending on the reaction scale, the amount of reagents here should be adjusted accordingly.

Terminal azide and alkyne groups were added to the primary sequences through additional steps. Propargylamine was used in the first sub-monomer building step for the alkyne side chain. After complete synthesis of the 24 monomers, the primary amine involving in the bromide substitution step was replaced with sodium azide, capping the polypeptoid with an azide group.

Polypeptoids were cleaved from the resin with a cleavage cocktail containing 95% trifluoroacetic acid (TFA), 2.5% water, and 2.5% triisopropylsilane (TIPS) for 30 minutes. The resin was filtered out and the TFA was dried off either through continuous nitrogen gas flow or in a centrifugal vacuum evaporator. Crude products were purified using

reverse phase preparative HPLC (Shimadzu Nexera) in acetonitrile:water fractions. The purified products were also verified with MALDI-TOF (Bruker Daltonics, Billerica, MA), which showed monodisperse MW peaks (Fig. C.1).

C.2 Long polypeptoid synthesis

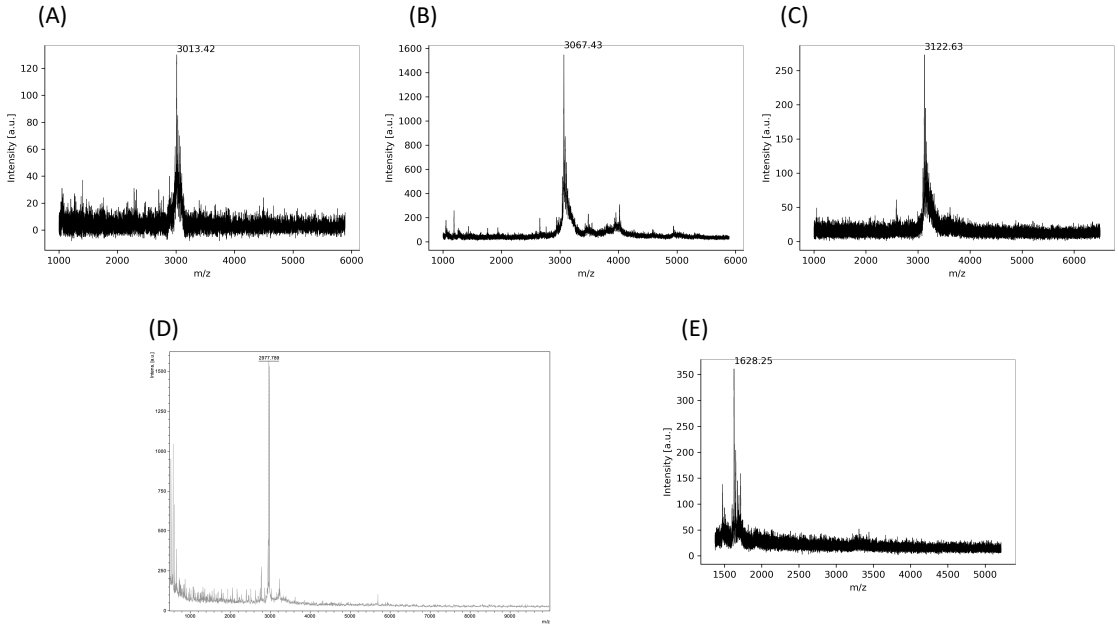

Figure C.1: MALDI spectra of all polypeptoid constructs studied in this dissertation: (A) 24mer with 1/6 charge, (B) 24mer with 1/3 charge, (C) 24mer with 1/2 charge sequence, (D) 24mer with only neutral Nme (taken by Beihang Yu), and (E) 12mer with 1/3 charge. For the charged sequences, their peaks show MW + Na^+, indicating an ionized state due to the characterization technique.

Prior to polymerization, unit polypeptoids were lyophilized for accurate weight measurement. The azide-alkyne click reaction calls for the first ionized state of copper (Cu(I)) as a catalyst, which must be generated *in situ*. Cu(I) is supplied from a more stable Cu(II) that can be reduced by ascorbic acid. Cu(II) was added in the reaction as a complex with tris((1-benzyl-4-triazolyl)methyl)amine (TBTA), which helps stabilize the transient

Cu(I) state [216]. This complex was prepared prior to the click reaction by mixing copper sulfate ($CuSO_4$) with TBTA in equal molar ratio in 1:1 water:acetonitrile. In a 3 mL glass conical reaction vial was a 200 μL aqueous mixture containing the peptoid (1 mM), copper(II)-TBTA complex (5 mM), ascorbic acid (5 mM), and triethylammonium acetate buffer pH 7 (0.2 M). The vial was capped with a rubber septum and the reaction solution was purged with nitrogen gas bubbling for 1 minute. The reaction was run at 50°C with vigorous stirring for 1-3 days. If the solution showed significant evaporation, the reaction could be run at room temperature instead.

After the reaction, the solution was filtered through a 100 kDa Amicon® Ultra Centrifugal filter to remove all shorter polypeptoids and excess reagents. In a 1.5 mL microcentrifuge tube, the long polypeptoids were mixed with biotin-azide (0.5 mM final, pre-dissolved in acetonitrile), $CuSO_4$ (0.5 mM final), and ascorbic acid (0.5 mM final). The tube was tightly wrapped with parafilm and rotated at room temperature for 24 hours. The reaction mixture was then filtered through a 100 kDa Amicon® Ultra Centrifugal filter with multiple washes to properly remove excess biotin-azide, which could bind to an alkyne surface and cause experimental issues. The concentration of the long polypeptoids was estimated by UV-Vis absorbance at 205 nm. The existence of long polypeptoids was difficult to observe with common characterization techniques, such as dynamic light scattering (DLS) or MALDI-TOF, due to low sample concentration and high polydispersity. Following a prior study's success [217], long polypeptoids were seen in SDS-PAGE, showing a broad distribution of MWs up to 500 kDa (Fig. C.2).

C.3 Titrations of 24mer polypeptoids

Titrations of polyelectrolytes are typically done at high concentrations (1-100 mM) [218], which would cost a significant amount of materials that could be used for the click

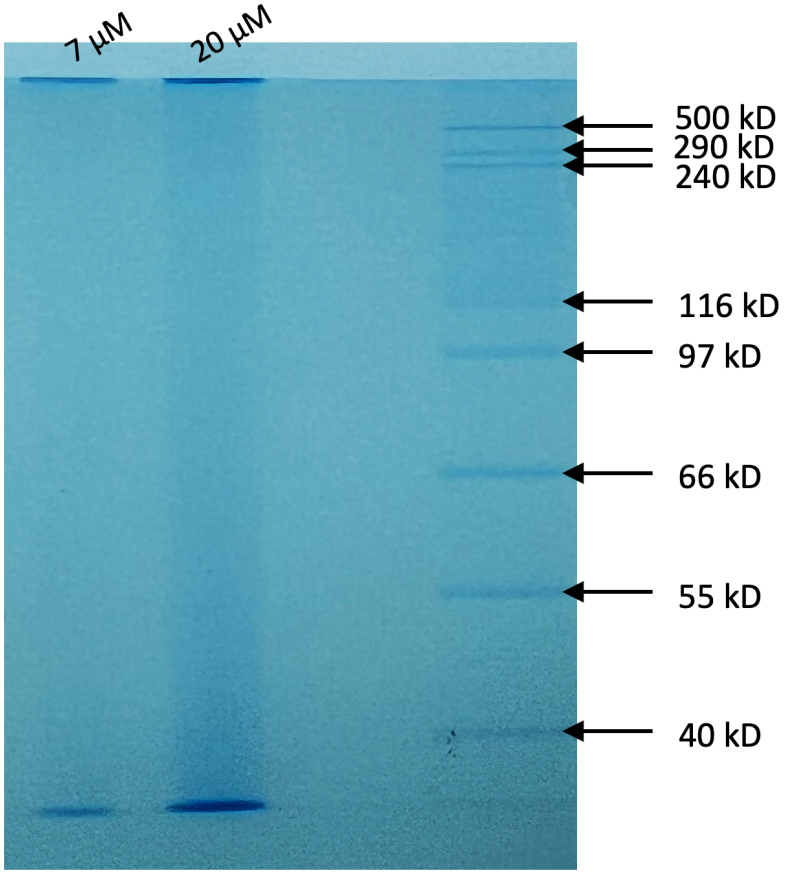

Figure C.2: SDS-PAGE of a 1/2 charge long polypeptoid sample. The gel had a 2% stacking layer and a 7.5% resolving layer. Samples (7 μM and 20 μM) were run along with the HiMark Unstained High Molecular Weight Protein Standard (Invitrogen) ladder. The gel was visualized with Coomassie Blue.

reaction. Because of this, titrations were done at low volumes, enabled by the use of a micro pH electrode. Samples of 24mer polypeptoids (1 mM in 200 μL water) were prepared. A 0.5-1.5 mol equivalent of COOH groups of HCl was also included as co-analyte. Titration was done by repetitive cycles of adding 1 μL of 50-90 mM NaOH, vortex mixing, then reading the resulting pH. The equivalence point, where NaOH had just fully neutralized all polypeptoid COOH groups, was determined as the point with the steepest slope on the curve. The polypeptoid titration starting point was calculated as the

point where NaOH had fully neutralized the HCl co-analyte. The half-equivalence point, which signifies the pK_a of the COOH groups, is the midpoint between the equivalence point and the titration starting point (Fig. C.3). The pK_a of COOH groups in the 1/6, 1/3, and 1/2 sequences are 4.75, 4.95, 5.51, respectively.

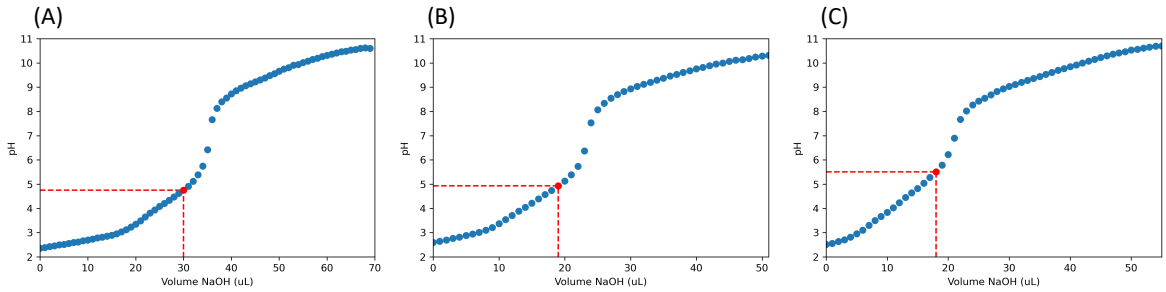

Figure C.3: Titration curves of (A) 1/6 charge, (B) 1/3 charge, and (C) 1/2 charge 24mer polypeptoid sequence. The half-equivalence points are shown in red.

C.4 Polypeptoid attachment for magnetic tweezers experiments

A flow cell with alkyne functionalization was constructed following the protocol described in Chapter 3. For all rinsing and incubation steps, the flow cell was kept in an enclosed humidified chamber. The flow cell was first rinsed with 400 μL PBS with 0.1% Tween-20. A 20 μL sample of long polypeptoids (10-100 nM final in PBS with 0.1% Tween-20) was added in the flow cell and let incubate for 1-2 hours. After that, the flow cell was then rinsed with 800 μL of 10 mM tris buffer pH 7.4 with 0.1% Tween-20 to remove any unbound polypeptoids.

A 0.5 mg/mL solution of Dynabeads™ MyOne™ Streptavidin C1 magnetic beads (2 μL of stock diluted in a final volume of 40 μL) was prepared by thoroughly washing twice in 10 mM tris buffer pH 7.4 with 0.1% Tween-20. With the flow cell on the microscope,

a 10 μL of this bead solution was added and let incubate for 10 minutes. Small drops of buffer were added intermittently to promote bead diffusion across the entire surface. Excess beads were washed away with >1000 μL 1-10 mM tris buffer pH 7.4 with 0.01-0.1% Tween-20 to begin an experiment. Very low ionic strength buffers (1 mM tris) was prepared the day of an experiment to prevent loss of buffering capacity and pH change.

Appendix D

Preliminary results

D.1 Measurements of different NFLt constructs

D.1.1 Variants of the NFLt sequence

In Chapter 4, we describe the use of entropic elasticity measurements to quantify the conformational ensemble of the NFLt. We find an ideal chain behavior likely stemming from an average between electrostatic repulsion in the charged domain and hydrophobic attraction in the N-terminus. To explore these local conformations, different variants of the original NFLt sequence, where certain regions are deleted, can be measured and compared. Roy Beck, Gil Koren, and Mathar Kravikass have provided several NFLt polyprotein constructs for magnetic tweezers experiments.

The Del0 construct has a shorter neutral domain as the first 11 residues are removed (`ASENLYFQGAS`). However, it reintroduces the last 5-residue lysine-rich region (`AKKKD`) that is associated with handshake interactions between neurofilaments [148]. This 157-residue variant has a NCPR of 0.22, similar to that of the studied NFLt construct. The sequence of Del0 is

MGCGSGTRLS FTSVGSITSG YSQSSQVFGR SAYSGLQSSS YLMSARSFPA YYTSHVQEEQ
TEVEETIEAT KAEEAKDEPP SEGEAEEEEK EKEEGEEEEG AEEEEAAKDE SEDTKEEEEG
GEGEEEDTKE SEEEEKKEES AGEEQVAKKK DGGGSCG

The Del01 construct has an even shorter neutral domain where residues 7-28 (TRLS FTSVGSITSG YSQSSQVF) from the Del0 are deleted. The rest of the Del01 is identical to the Del0. This 135-residue variant has a NCPR of -0.27, higher than that of the studied NFLt construct. The sequence of Del01 is

MGCGSGGRSA YSGLQSSSYL MSARSFPAYY TSHVQEEQTE VEETIEATKA EEAKDEPPSE
GEAEEEEKEK EEGEEEEGAE EEEAAKDESE DTKEEEEGGE GEEEDTKESE EEEKKEESAG
EEQVAKKKDG GGSCG

The last construct, Del012, has the shortest neutral domain, containing only 9 residues (PAYY TSHVQ). This 115-residue variant has a NCPR of -0.33, remarkably more charged than the studied NFLt construct and somewhat comparable to the ProTα in previous studies [32, 27]. The sequence of Del012 is

MGCGSGPAYY TSHVQEEQTE VEETIEATKA EEAKDEPPSE GEAEEEEKEK EEGEEEEGAE
EEEAAKDESE DTKEEEEGGE GEEEDTKESE EEEKKEESAG EEQVAKKKDG GGSCG

D.1.2 Del012 experiments and results

We are particularly interested in the Del012 variant as it can provide a stark contrast to the primary NFLt construct studied in this dissertation. With the uncharged, hydrophobic domain virtually eliminated, electrostatic repulsion is expected to be the only dominating intramolecular interaction of the Del012. In entropic elasticity measurements, we expect this to manifest as a swollen chain behavior ($\nu = 0.6$), where the force-extension power law γ in the Pincus blobs regime is 2/3. Further, with increasing

solution ionic strength, we predict that Del012 will show compaction (reduction of the Flory exponent) due to electrostatic screening.

We followed the NFLt protocols described in this dissertation for preliminary experiments with Del012. In addition, to account for the stronger electrostatic repulsion that could resist contact with magnetic bead and glass attachment sites, we raised the solution ionic strength with 50-100 mM NaCl. A Del012 experiment began with assembly of an NHS-functionalized flow cell. All incubation steps were done in an enclosed and humidified chamber. The surface chemistry was modified by incubating a 40 μL mixture of 1:1 DBCO-amine (10 mg/mL in DMSO) and PBS pH 8 with 0.1% Tween-20 for 1 hour. The flow cell was then rinsed with 10 mM tris buffer pH 7.4 with 0.1% Tween-20 and let sit for 15 minutes to deactivate any remaining NHS sites. After that, the flow cell was washed with 800 μL of 10 mm MES buffer pH 6.8 with 50 mM NaCl and 0.1% Tween-20.

Prior to attachment, a 1 μL aliquot of the Del012 stock sample ($\sim$2 μL) was diluted in 500 μL of 10 mM carbonate pH 11 with 500 mM NaCl and rotated for 1 hour. The solution was filtered through a 100 kDa Amicon® Ultra Centrifugal filter and exchanged to 10 mm MES buffer pH 6.8 with 50 mM NaCl and 0.1% Tween-20. The Del012 concentrate ($\sim$10 μL, $\sim$200 nM) was added into the flow cell and let incubate for 12-24 hours at 4°C. The flow cell was then rinsed with 800 μL 10 mm MES buffer pH 6.8 with 50 mM NaCl and 0.1% Tween-20 to remove any unbound polyproteins. A 200 μL passivating solution of 10 mM MES pH 6.8 with 1% BSA and 0.1% Tween-20 was flowed in and let incubate for 30 minutes. After that, the flow cell was thoroughly rinsed with 800 μL of 10 mm MES buffer pH 6.8 with 50 mM NaCl and 0.1% Tween-20.

A 0.5 mg/mL solution of Dynabeads™ MyOne™ Streptavidin C1 magnetic beads (2 μL of stock diluted in a final volume of 40 μL) was prepared by thoroughly washing twice in 10 mm MES buffer pH 6.8 with 1% BSA and 0.1% Tween-20 and resuspending in 10

mm MES buffer pH 6.8 with 50 mM NaCl and 0.1% Tween-20. A 10 μL of the magnetic beads solution was added into the flow cell and let incubate for 10 minutes. Small drops of buffers were added intermittently to promote bead diffusion. The flow cell was then thoroughly rinsed with >1000 μL of 10 mm MES buffer pH 6.8 with 0.1% Tween-20 to remove unbound beads.

Low ionic strength measurements were done in 10 mm MES buffer pH 6.8 with 0.1% Tween-20. A force-extension curve of one Del012 polyprotein tether is shown in Fig. D.1A. Similar to the NFLt construct in Chapter 4, the Del012 also exhibits two elastic regimes: an asymptotic behavior at high force and Pincus blob elasticity at low force. Data above 1.5 pN are fitted to the WLC model, giving a persistence length of 1.57 ± 0.17 nm. This is significantly higher than that of the full construct (Fig. D.1B), hinting at a potential electrostatic influence on the polyprotein flexibility. This result is somewhat sensible, as the Del012 is dominated with electrostatic repulsion, causing local stiffening. However, this observation is at odds with the conclusion from Chapter 5, which states an independence of polyelectrolyte flexibility from side chain charges. While it is true that the glutamic acid residues here also have a pendant charge structure, their proximity to one another (neighboring without spacers) likely causes electrostatic repulsion to impact local flexibility.

At low forces (< 1.5 pN), data are fitted to a power law, showing $\gamma = 0.73 \pm 0.03$ (Fig. D.1A), which also corresponds to $\nu = 0.58$. The exponents indicate swollen chain behavior due to electrostatic repulsion. Comparing to the full NFLt construct from Chapter 4, it is apparent that Del012 is more expanded, which is expected due to the almost complete removal of the neutral domain.

The preliminary result described here demonstrates the future direction of magnetic tweezers study on the NFLt and IDPs in general. By designing and measuring polyproteins of specific domains, one can gain insights to local configurations and how they

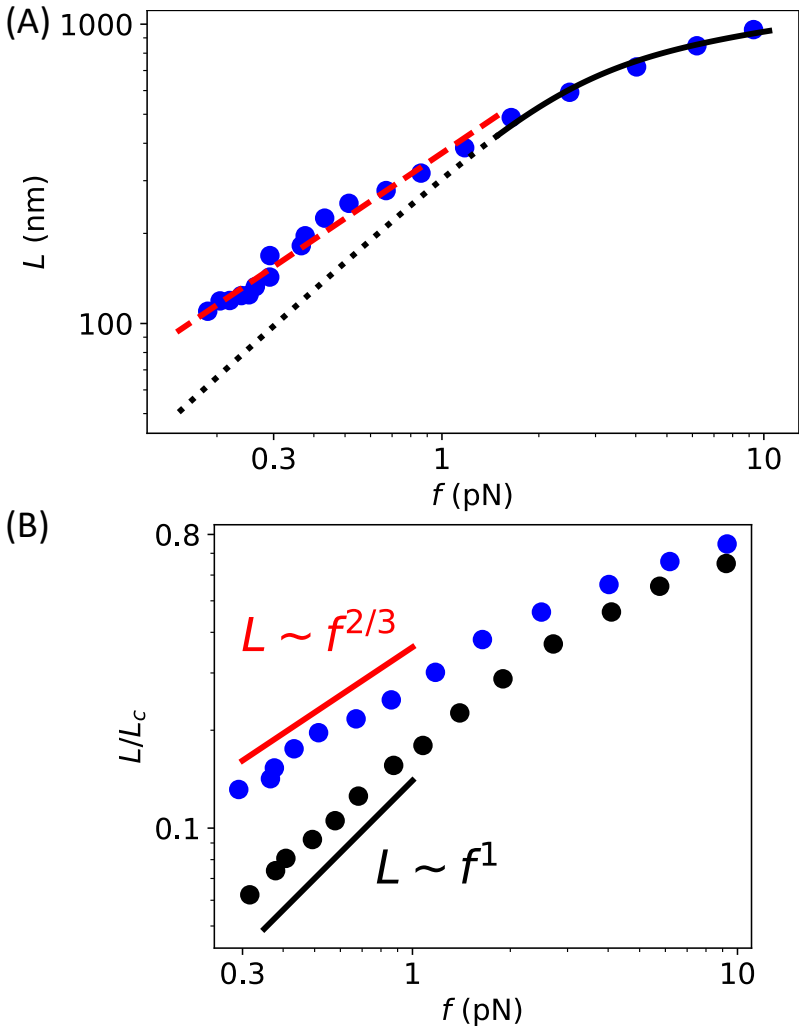

Figure D.1: (A) Force-extension curve of one Del012 polyprotein (blue circles) measured in 10 mm MES buffer pH 6.8 with 0.1% Tween-20. The plot shows WLC fitting (black solid line) and low-force power law fitting (red dashed line). Ideal-chain extrapolation of the WLC model (black dotted line) shows the deviation of Del012's low force behavior, which is better described as a swollen chain. (B) Comparison to a force-extension curve of the NFLt construct (black circles) in Chapter 4. The extensions are normalized by individual L_c values from separate WLC fittings. Force-extension curves at higher forces of the two constructs do not match up, indicating different flexibility behaviors. The low-force elasticity of the NFLt is closer to ideal chain ($\gamma = 1$, while the Del012 is similar to a swollen chain ($\gamma = 2/3$).

contribute to global conformational ensembles. For example, experiments with Del012 can reveal regional electrostatic repulsion that is not obvious from measuring the full

NFLt due to conflicting hydrophobic attractions from the other domain.

For a successful study of all NFLt variants, there are various experimental optimization and measurements to be achieved. A study involving multiple constructs requires high data throughput, which can be limited by poor polyprotein chemistry. As discussed briefly in Chapter 6, there are several methods to synthesize a polyprotein beyond disulfide bridging. To verify the preliminary result on Del012 reported here, more replicates would be necessary. Further, to test electrostatic effects, force-extension measurements must be done in different NaCl concentrations. Denaturant would also be interesting to use in order to potentially observe the initial compaction at low GuHCl concentration. Beyond just low-force power law, because the persistence length in this preliminary result shows a possible dependence on electrostatic repulsion, its measurements in NaCl would be insightful.

D.2 Designing and measuring polypeptide constructs

Synthetic polypeptides are interesting to investigate because they can serve as a model to investigate sequence effects in the context of a peptide backbone, which can exert noticeable influences on its own. Simple polypeptide designs containing 1-3 types of residues can isolate provide insights to specific features such as hydrophobic attractions or electrostatic interactions. Similar to polypeptoids, polypeptides can be synthesized on a solid-phase platform, which allows incorporation of terminal modifications that enable end-to-end click chemistry.

D.2.1 Polyglycine

Our first attempt in this research direction is polyglycine (PG), a homopolypeptide. Because the α-carbon side chain of glycine is simply a hydrogen, it is essentially a model

for the peptide backbone. As the peptide backbone is a shared feature among all protein classes, a thorough understanding of its conformational behavior is crucial to various protein biophysics areas. A long-existing question on chemically denatured proteins is whether they are fully random coils or if they contain residual structures, the latter of which could partially be explained by backbone hydrogen bonding [219, 220, 221]. Resolving this debate can have significant implications in determining the thermodynamics and kinetics of protein folding [222]. For disordered polar tracts, they are found to be compact globules in water despite minimal intramolecular hydrophobicity [62, 223, 224], a behavior that is hypothesized to be driven by the protein backbone [65]. In addition, PG by itself is an interesting model to investigate due to its role as a flexible linker in proteins [12, 225, 226, 227]. Knowledge on the physical and mechanical properties of PG is important to deciphering conformational dynamics of protein functionality, as well as aiding the rational design of multidomain proteins [228, 229, 230]. These reasons combined provide a compelling case for PG as a subject of study.

Literature has largely agreed that water is a poor solvent for the peptide backbone, which is attributed to intra-chain hydrogen bondings and hydration effects being more favorable than polymer-solvent attraction [65, 231, 232, 233]. Additionally, PG is found to adopt secondary structures in solution, either an antiparallel β-sheet PG type I (PG I) or a 3_1-helix PG type II (PG II) [234, 235, 236, 237], where the latter is more dominant in PG conformational ensembles [234]. While short PG segments have been examined through various experimental methods [233, 236, 234, 238], to our knowledge, there has not been any measurements on a pure, long PG construct that can provide robust polymer physics analysis of the peptide backbone. Because of that, we aim to use magnetic tweezers as a means to characterize the conformational behavior of PG through entropic elasticity measurements.

Magnetic tweezers experiments with PG face significant chemistry challenges. To

our knowledge, there has not been any reports on biochemical expression or synthesis of PG with at least 1000 residues. Reapplying the long chain construction scheme from polypeptoids to PG would result in heterogenous linkers, obscuring observations of backbone effects. Further, PG is not soluble in water, adding difficulty to experiments in aqueous solutions.

To form long PG chains, we turned to the carbodiimide-enabled crosslinking reaction between a COOH and a primary amine (Fig. D.2). Exploiting the existing N- and C-termini of an unmodified PG, this chemistry enables end-to-end attachment between individual units. One particularly powerful advantage of this reaction is the resulting amide bond, which is identical to the peptide backbone and does not introduce any heterogeneous linkages.

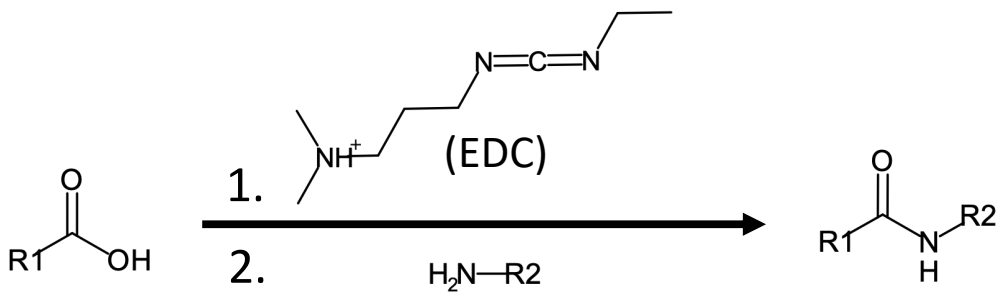

Figure D.2: Carbodiimide crosslinking chemistry. In a one-pot reaction, EDC first reacts to COOH to generate an unstable intermediate, which is followed by a nucleophilic attack by a primary amine to form an amide bond.

We then investigated the aqueous solubility of PG. While it has been reported that PG is soluble in water in the presence of TFA [236], this protocol is incompatible with the crosslinking chemistry. Following a previous work on exploring the solubility of PG in various alkali halide salts [239], we discovered that PG readily dissolves in 8M lithium iodide (LiI). Notably, through many tests, we found that this condition requirement is quite strict, as PG in bulk is not soluble in a wide range of other solvents, such as DMSO, acetonitrile, GuHCl, lithium chloride (LiCl), or $\leq$4 M LiI.

We made a preliminary attempt of the PG crosslinking reaction. Unmodified PG was purchased from Sigma-Aldrich (P8791), which features a polydispersity of MW in the range of 500-5000 Da, corresponding to 5-88 residues. Prior to the reaction, PG was dissolved in a 1 mL solution of 0.1 M MES buffer pH 5 (reaction calls for a slightly acidic condition) with 8 M LiI to make a concentration of 10 mg/mL. In a 3 mL glass conical reaction vial, a ∼100 mol equivalent of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) was weighed. The PG solution was added to the vial and the mixture was vigorously stirred for 4 days. After that, a 2x molar excess of biotin-amine and EDC were added to the solution mixture and stirred for ∼18 hours. To clean up, the solution was sufficiently filtered through a 100 kDa Amicon® Ultra Centrifugal filter and exchanged with 0.1 M 3-(N-morpholino)propanesulfonic acid (MOPS) buffer pH 8 with 8 M LiI to remove all excess reagents. The concentration of long PGs were measured with a Nanodrop using an appropriate blank, which was an aliquot of the 0.1 M MOPS buffer pH 8 with 8 M LiI filtered through an Amicon® Ultra Centrifugal filter. The resulting product is long PG chains with terminal biotin and amine. We found the concentration yield to be ∼240 nM. The stock sample was then stored at -20°C.

Magnetic tweezers experiments with PG were prepared with NHS-functionalized coverslips. After flow cell construction, the fluid channel was washed with 200 μL 1X PBS pH 8 with 0.1% Tween-20. A 20 μL sample of PG (∼3 nM final, diluted in 1X PBS pH 8 with 0.1% Tween-20) was immediately added to the flow cell and let incubate for ∼ 1 hour and 30 minutes at room temperature. The flow cell was then washed with 400 μL 10 mM tris pH 7.5 with 0.1% Tween-20 and let incubate for 5 minutes. The flow cell was then rinsed with 400 μL of 10 mM MES pH 6.8 with 0.1% Tween-20, then with 200 μL 10 mM MES pH 6.8 with 6 M GuHCl and 0.1% Tween-20. MyOne beads were prepared in MES buffer as described above and were incubated in the flow cell for ∼ 10 minutes. Bead incubation was done in 6 M GuHCl to help PG polymers expand and expose their

terminal biotins for attachment.

All "water" measurements were done in 10 mM MES pH 6.8 with 0.1% Tween-20. Upon initial examination of force-extension curves with the WLC model, we found that the persistence length of PG is greater than 1 nm (Fig. D.3). For correct WLC fitting, we assessed the appropriate cutoff force for each curve, which ranges between 0.5 and 1 pN. From 9 individual tethers, the average persistence length of PG is 2.7 ± 0.5 nm (error reported is the standard deviation), which is noticeably higher than values reported in literature (~ 0.4 nm) [240, 241]. Comparing to other homopeptides, our measured l_p is similar to poly(alanine) [240] or random-coil poly(lysine) [242] while being significantly smaller than poly(proline) [243] or α-helical poly(lysine) [242]. The unexpected stiffness can thus be attributed to PG being in the PG II conformation or a mixture of PG II and random coil, either of which are more rigid than unstructured chains but much more flexible than other conventional secondary structures. As we did not detect any tethers with a lower l_p, we conclude that PG II is the dominant configuration of PG in solution.

In another surprising observation, low-force elasticity of PG reveals that it is much more soluble in water than expected. As seen in Fig. D.3, the Pincus blobs power law regime indicates swollen chain behavior, implying that water is actually a good solvent for PG. This contradicts prior reports showing that PG is a compact globule due to intramolecular hydrogen bonding [65, 231, 232, 233]. We hypothesize that the other observation described above could help explain this discrepancy. In a helical structure such as PG II, hydrogen bonds are locally organized, and bond donors and acceptors are saturated [235]; therefore, long-range hydrogen bonding typically associated with protein globules might have been nonexistent in PG II. One analogy to this explanation is double-stranded DNA (dsDNA): while dsDNA has intramolecular attractions (base pair interactions), it behaves as a swollen coil [244] because such interactions are stable and only fulfilled locally. Nevertheless, this result is at odds with the observable insolubility

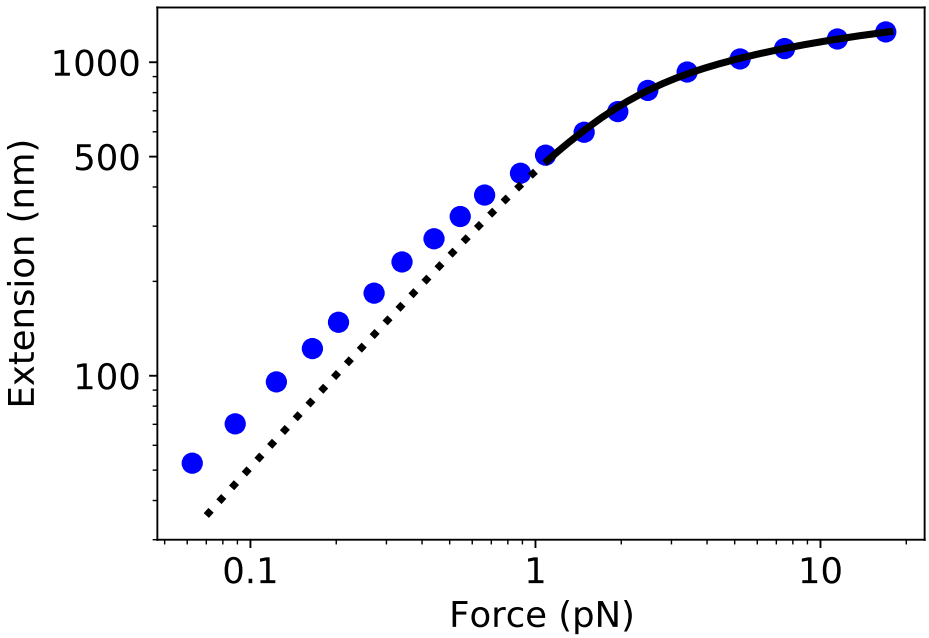

Figure D.3: Force-extension curve of a PG tether (blue circles). The WLC model is fit to forces above 1 pN (solid black line). The low-force extrapolation (dotted black line) shows ideal chain prediction.

of PG in water. It is possible that bulk and dilute solubility of PG are vastly distinguishable: while PG at high concentrations can form intermolecular hydrogen bonding causing aggregates, single-molecule PG adopts a more solvent-favorable conformation.

We predicted that the PG II configuration can be unraveled in the presence of strong denaturant, leading to an increase in flexibility. However, from our preliminary measurements of PG in 6 M GuHCl, there is no discernible change in the persistence length (Fig. D.4). This is surprising, as we expect denaturant to disrupt intrachain organized hydrogen bonding, causing PG to behave as a flexible random coil. It is possible that the PG II conformation can remain stable in a good solvent, as shown in a prior study of PG in Li^+ salts [234]. In addition, PG is unsurprisingly swollen in strong denaturant, shown by the low-force elasticity (Fig. D.4).

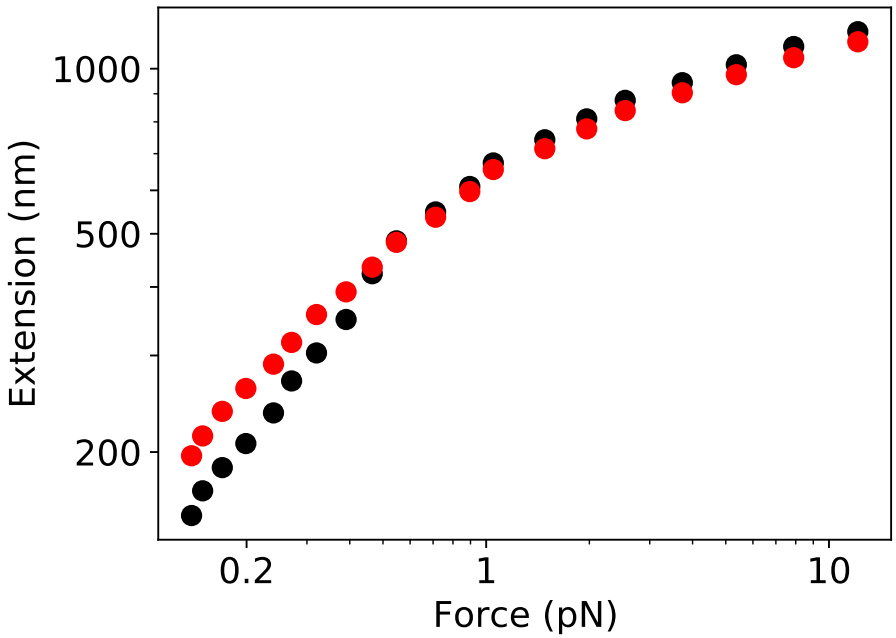

Figure D.4: Force-extension curves of a PG tether in MES buffer (black circles) and 6 M GuHCl (red circles). The denaturant solution contains 10 mM MES buffer pH 6.8 with 0.1% Tween-20. The curves overlap at higher forces, implying at the flexibility is unchanged.

From these preliminary results, we can further investigate this unexpected conformational behavior through various future experiments. By using M280 beads, we can apply very high forces that can potentially unravel the PG II helix in a similar manner to a previous study on poly(dA) [145]. Instead of a denaturant, it is possible that a kosmotropic agent, such as ammonium sulfate, could destabilize the helical conformation. Gathering a large dataset of l_p measurements can allow quantitative analysis of the prevalence of PG II in PG conformational ensemble. In addition, a complementary probing method, such as light scattering or molecular simulation, could provide other insights into PG behavior in solution.

Bibliography

[1] W. D. Comper and T. C. Laurent, *Physiological function of connective tissue polysaccharides.*, *Physiological reviews* **58** (1978), no. 1 255–315.

[2] R. Preston, *Polysaccharide conformation and cell wall function*, *Annual Review of Plant Physiology* **30** (1979), no. 1 55–78.

[3] O. C. Redfern, B. Dessailly, and C. A. Orengo, *Exploring the structure and function paradigm*, *Current opinion in structural biology* **18** (2008), no. 3 394–402.

[4] C. J. Oldfield and A. K. Dunker, *Intrinsically disordered proteins and intrinsically disordered protein regions*, *Annual review of biochemistry* **83** (2014) 553–584.

[5] V. N. Uversky, *Intrinsically disordered proteins from a to z*, *The international journal of biochemistry & cell biology* **43** (2011), no. 8 1090–1103.

[6] V. N. Uversky, *Introduction to intrinsically disordered proteins (idps)*, *Chemical reviews* **114** (2014), no. 13 6557–6560.

[7] T. McMeekin, *Milk proteins*, *Journal of Food Protection* **15** (1952), no. 2 57–61.

[8] B. Jirgensons, *Optical rotation and viscosity of native and denatured proteins. x. further studies on optical rotatory dispersion*, *Archives of biochemistry and biophysics* **74** (1958), no. 1 57–69.

[9] 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*, in *Pac Symp Biocomput*, vol. 3, pp. 437–448, 1998.

[10] C. J. Oldfield, Y. Cheng, M. S. Cortese, C. J. Brown, V. N. Uversky, and A. K. Dunker, *Comparing and combining predictors of mostly disordered proteins*, *Biochemistry* **44** (2005), no. 6 1989–2000.

[11] A. K. Dunker, P. Romero, Z. Obradovic, E. C. Garner, and C. J. Brown, *Intrinsic protein disorder in complete genomes*, *Genome informatics* **11** (2000) 161–171.

[12] 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, *et. al.*, *Classification of intrinsically disordered regions and proteins*, *Chemical reviews* **114** (2014), no. 13 6589–6631.

[13] P. E. Wright and H. J. Dyson, *Intrinsically disordered proteins in cellular signalling and regulation*, *Nature reviews Molecular cell biology* **16** (2015), no. 1 18–29.

[14] P. Tompa, E. Schad, A. Tantos, and L. Kalmar, *Intrinsically disordered proteins: emerging interaction specialists*, *Current opinion in structural biology* **35** (2015) 49–59.

[15] A. K. Dunker, C. J. Brown, and Z. Obradovic, *Identification and functions of usefully disordered proteins*, *Advances in protein chemistry* **62** (2002) 25–49.

[16] S.-H. Kim, D. H. Shin, I.-G. Choi, U. Schulze-Gahmen, S. Chen, and R. Kim, *Structure-based functional inference in structural genomics*, *Journal of structural and functional genomics* **4** (2003) 129–135.

[17] B. H. Dessailly, R. Nair, L. Jaroszewski, J. E. Fajardo, A. Kouranov, D. Lee, A. Fiser, A. Godzik, B. Rost, and C. Orengo, *Psi-2: structural genomics to cover protein domain family space*, *Structure* **17** (2009), no. 6 869–881.

[18] M. Smyth and J. Martin, *x ray crystallography*, *Molecular Pathology* **53** (2000), no. 1 8.

[19] A. Ilari and C. Savino, *Protein structure determination by x-ray crystallography*, *Bioinformatics: Data, Sequence Analysis and Evolution* (2008) 63–87.

[20] D. Baker and A. Sali, *Protein structure prediction and structural genomics*, *Science* **294** (2001), no. 5540 93–96.

[21] D. S. Marks, T. A. Hopf, and C. Sander, *Protein structure prediction from sequence variation*, *Nature biotechnology* **30** (2012), no. 11 1072–1080.

[22] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, *et. al.*, *Highly accurate protein structure prediction with alphafold*, *Nature* **596** (2021), no. 7873 583–589.

[23] 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, *et. al.*, *Intrinsically disordered protein*, *Journal of molecular graphics and modelling* **19** (2001), no. 1 26–59.

[24] V. N. Uversky, *A decade and a half of protein intrinsic disorder: biology still waits for physics*, *Protein Science* **22** (2013), no. 6 693–724.

[25] P. J. Flory, *Principles of polymer chemistry*. Cornell university press, 1953.

[26] J. A. Marsh and J. D. Forman-Kay, *Sequence determinants of compaction in intrinsically disordered proteins*, *Biophysical journal* **98** (2010), no. 10 2383–2390.

[27] H. Hofmann, A. Soranno, A. Borgia, K. Gast, D. Nettels, and B. Schuler, *Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy*, *Proceedings of the National Academy of Sciences* **109** (2012), no. 40 16155–16160.

[28] P. Bernado and D. I. Svergun, *Structural analysis of intrinsically disordered proteins by small-angle x-ray scattering*, *Molecular biosystems* **8** (2012), no. 1 151–167.

[29] A. G. Kikhney and D. I. Svergun, *A practical guide to small angle x-ray scattering (saxs) of flexible and intrinsically disordered proteins*, *FEBS letters* **589** (2015), no. 19 2570–2577.

[30] J. A. Riback, M. A. Bowman, A. M. Zmyslowski, C. R. Knoverek, J. M. Jumper, J. R. Hinshaw, E. B. Kaye, K. F. Freed, P. L. Clark, and T. R. Sosnick, *Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water*, *Science* **358** (2017), no. 6360 238–241.

[31] G. Koren, S. Meir, L. Holschuh, H. D. Mertens, T. Ehm, N. Yahalom, A. Golombek, T. Schwartz, D. I. Svergun, O. A. Saleh, *et. al.*, *Intramolecular structural heterogeneity altered by long-range contacts in an intrinsically disordered protein*, *Proceedings of the National Academy of Sciences* **120** (2023), no. 30 e2220180120.

[32] S. Müller-Späth, A. Soranno, V. Hirschfeld, H. Hofmann, S. Rüegger, 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** (2010), no. 33 14609–14614.

[33] 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** (2016) 207–231.

[34] P. Pincus, *Excluded volume effects and stretched polymer chains*, *Macromolecules* **9** (1976), no. 3 386–388.

[35] O. A. Saleh, *Perspective: Single polymer mechanics across the force regimes*, *The Journal of chemical physics* **142** (2015), no. 19.

[36] A. Solanki, K. Neupane, and M. T. Woodside, *Single-molecule force spectroscopy of rapidly fluctuating, marginally stable structures in the intrinsically disordered protein α-synuclein*, *Physical review letters* **112** (2014), no. 15 158103.

[37] 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** (2014), no. 1 e86495.

[38] J. Yu, S. Malkova, and Y. L. Lyubchenko, *α-synuclein misfolding: single molecule afm force spectroscopy study*, *Journal of molecular biology* **384** (2008), no. 4 992–1001.

[39] M. Sandal, F. Valle, I. Tessari, S. Mammi, E. Bergantino, F. Musiani, M. Brucale, L. Bubacco, and B. Samorì, *Conformational equilibria in monomeric α-synuclein at the single-molecule level*, *PLoS biology* **6** (2008), no. 1 e6.

[40] H. P. Truong, I. L. Morgan, G. Koren, R. Avinery, R. Beck, and O. A. Saleh, *Pincus blob elasticity in an intrinsically disordered protein*, *The European Physical Journal E* **46** (2023), no. 10 100.

[41] J.-F. Lutz, M. Ouchi, D. R. Liu, and M. Sawamoto, *Sequence-controlled polymers*, *Science* **341** (2013), no. 6146 1238149.

[42] J. De Neve, J. J. Haven, L. Maes, and T. Junkers, *Sequence-definition from controlled polymerization: the next generation of materials*, *Polymer Chemistry* **9** (2018), no. 38 4692–4705.

[43] S. L. Perry and C. E. Sing, *100th anniversary of macromolecular science viewpoint: Opportunities in the physics of sequence-defined polymers*, *ACS Macro Letters* **9** (2020), no. 2 216–225.

[44] A. J. DeStefano, R. A. Segalman, and E. C. Davidson, *Where biology and traditional polymers meet: The potential of associating sequence-defined polymers for materials science*, *Jacs Au* **1** (2021), no. 10 1556–1571.

[45] R. N. Zuckermann, *Peptoid origins*, *Peptide Science* **96** (2011), no. 5 545–555.

[46] J. Sun and R. N. Zuckermann, *Peptoid polymers: a highly designable bioinspired material*, *ACS nano* **7** (2013), no. 6 4715–4732.

[47] S. Xuan and R. N. Zuckermann, *Engineering the atomic structure of sequence-defined peptoid polymers and their assemblies*, *Polymer* **202** (2020) 122691.

[48] S. A. Fowler and H. E. Blackwell, *Structure–function relationships in peptoids: recent advances toward deciphering the structural requirements for biological function*, *Organic & biomolecular chemistry* **7** (2009), no. 8 1508–1524.

[49] A. M. Rosales, H. K. Murnen, R. N. Zuckermann, and R. A. Segalman, *Control of crystallization and melting behavior in sequence specific polypeptoids*, *Macromolecules* **43** (2010), no. 13 5627–5636.

[50] S. M. Miller, R. J. Simon, S. Ng, R. N. Zuckermann, J. M. Kerr, and W. H. Moos, *Proteolytic studies of homologous peptide and n-substituted glycine peptoid oligomers*, *Bioorganic & medicinal chemistry letters* **4** (1994), no. 22 2657–2662.

[51] C. W. Wu, T. J. Sanborn, K. Huang, R. N. Zuckermann, and A. E. Barron, *Peptoid oligomers with α-chiral, aromatic side chains: sequence requirements for the formation of stable peptoid helices*, *Journal of the American Chemical Society* **123** (2001), no. 28 6778–6784.

[52] E. J. Robertson, A. Battigelli, C. Proulx, R. V. Mannige, T. K. Haxton, L. Yun, S. Whitelam, and R. N. Zuckermann, *Design, synthesis, assembly, and engineering of peptoid nanosheets*, *Accounts of Chemical Research* **49** (2016), no. 3 379–389.

[53] J. Sun, X. Jiang, R. Lund, K. H. Downing, N. P. Balsara, and R. N. Zuckermann, *Self-assembly of crystalline nanotubes from monodisperse amphiphilic diblock copolypeptoid tiles*, *Proceedings of the National Academy of Sciences* **113** (2016), no. 15 3954–3959.

[54] G. L. Sternhagen, S. Gupta, Y. Zhang, V. John, G. J. Schneider, and D. Zhang, *Solution self-assemblies of sequence-defined ionic peptoid block copolymers*, *Journal of the American Chemical Society* **140** (2018), no. 11 4100–4109.

[55] H. P. Truong, S. Mengel, B. Yu, R. A. Segalman, and O. A. Saleh, *Single-molecule polypeptoid stretching reveals the effects of charge sequence on chain conformation*, *Macromolecules* **56** (2023), no. 21 8558–8564.

[56] 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.

[57] R. Williams, Z. Obradovic, V. Mathura, W. Braun, E. Garner, J. Young, S. Takayama, C. J. Brown, and A. K. Dunker, *The protein non-folding problem: amino acid determinants of intrinsic order and disorder*, in *Biocomputing 2001*, pp. 89–100. World Scientific, 2000.

[58] 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** (2007), no. 5 1439–1456.

[59] R. L. Baldwin, *How does protein folding get started?*, *Trends in biochemical sciences* **14** (1989), no. 7 291–294.

[60] G. Haran, *How, when and why proteins collapse: the relation to folding*, *Current opinion in structural biology* **22** (2012), no. 1 14–20.

[61] A. H. Mao, N. Lyle, and R. V. Pappu, *Describing sequence–ensemble relationships for intrinsically disordered proteins*, *Biochemical Journal* **449** (2013), no. 2 307–318.

[62] R. K. Das, K. M. Ruff, and R. V. Pappu, *Relating sequence encoded information to form and function of intrinsically disordered proteins*, *Current opinion in structural biology* **32** (2015) 102–112.

[63] J. Shao and M. I. Diamond, *Polyglutamine diseases: emerging concepts in pathogenesis and therapy*, *Human molecular genetics* **16** (2007), no. R2 R115–R123.

[64] H. J. Dyson and P. E. Wright, *Intrinsically unstructured proteins and their functions*, *Nature reviews Molecular cell biology* **6** (2005), no. 3 197–208.

[65] 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** (2008), no. 23 7380–7392.

[66] A. H. Mao, S. L. Crick, A. Vitalis, C. L. Chicoine, and R. V. Pappu, *Net charge per residue modulates conformational ensembles of intrinsically disordered proteins*, *Proceedings of the National Academy of Sciences* **107** (2010), no. 18 8183–8188.

[67] 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** (2013), no. 33 13392–13397.

[68] Z. Dosztányi, B. Mészáros, and I. Simon, *Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins*, *Briefings in bioinformatics* **11** (2010), no. 2 225–243.

[69] V. N. Uversky, *Unusual biophysics of intrinsically disordered proteins*, *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1834** (2013), no. 5 932–951.

[70] J.-E. Shea, R. B. Best, and J. Mittal, *Physics-based computational and theoretical approaches to intrinsically disordered proteins*, *Current opinion in structural biology* **67** (2021) 219–225.

[71] R. N. Zuckermann, J. M. Kerr, S. B. Kent, and W. H. Moos, *Efficient method for the preparation of peptoids [oligo (n-substituted glycines)] by submonomer solid-phase synthesis*, *Journal of the American Chemical Society* **114** (1992), no. 26 10646–10647.

[72] J. E. Murphy, T. Uno, J. D. Hamer, F. E. Cohen, V. Dwarki, and R. N. Zuckermann, *A combinatorial approach to the discovery of efficient cationic peptoid reagents for gene delivery*, *Proceedings of the National Academy of Sciences* **95** (1998), no. 4 1517–1522.

[73] H. K. Murnen, A. R. Khokhlov, P. G. Khalatur, R. A. Segalman, and R. N. Zuckermann, *Impact of hydrophobic sequence patterning on the coil-to-globule transition of protein-like polymers*, *Macromolecules* **45** (2012), no. 12 5229–5236.

[74] K. Kirshenbaum, A. E. Barron, R. A. Goldsmith, P. Armand, E. K. Bradley, K. T. Truong, K. A. Dill, F. E. Cohen, and R. N. Zuckermann, *Sequence-specific polypeptoids: A diverse family of heteropolymers with stable secondary structure*, *Proceedings of the National Academy of Sciences* **95** (1998), no. 8 4303–4308.

[75] J. S. Laursen, J. Engel-Andreasen, P. Fristrup, P. Harris, and C. A. Olsen, *Cis–trans amide bond rotamers in β-peptoids and peptoids: Evaluation of stereoelectronic effects in backbone and side chains*, *Journal of the American Chemical Society* **135** (2013), no. 7 2835–2844.

[76] Q. Sui, D. Borchardt, and D. L. Rabenstein, *Kinetics and equilibria of cis/trans isomerization of backbone amide bonds in peptoids*, *Journal of the American Chemical Society* **129** (2007), no. 39 12042–12048.

[77] R. K. Spencer, G. L. Butterfoss, J. R. Edison, J. R. Eastwood, S. Whitelam, K. Kirshenbaum, and R. N. Zuckermann, *Stereochemistry of polypeptoid chain configurations*, *Biopolymers* **110** (2019), no. 6 e23266.

[78] W. Zheng and R. B. Best, *An extended guinier analysis for intrinsically disordered proteins*, *Journal of molecular biology* **430** (2018), no. 16 2540–2553.

[79] W. Zheng, G. Zerze, A. Borgia, J. Mittal, B. Schuler, and R. B. Best, *Inferring properties of disordered chains from fret transfer efficiencies*, *Biophysical Journal* **114** (2018), no. 3 367a.

[80] P. Bernado and D. I. Svergun, *Analysis of intrinsically disordered proteins by small-angle x-ray scattering*, *Intrinsically Disordered Protein Analysis: Volume 2, Methods and Experimental Tools* (2012) 107–122.

[81] G. Fuertes, N. Banterle, K. M. Ruff, A. Chowdhury, D. Mercadante, C. Koehler, M. Kachala, G. Estrada Girona, S. Milles, A. Mishra, *et. al.*, *Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in saxs vs. fret measurements*, *Proceedings of the National Academy of Sciences* **114** (2017), no. 31 E6342–E6351.

[82] J. F. Marko and E. D. Siggia, *Stretching dna*, *Macromolecules* **28** (1995), no. 26 8759–8770.

[83] T. R. Strick, J.-F. Allemand, D. Bensimon, A. Bensimon, and V. Croquette, *The elasticity of a single supercoiled dna molecule*, *Science* **271** (1996), no. 5257 1835–1837.

[84] M. Rubinstein and R. Colby, *Polymer Physics*. 'Oxford University Press, New York', 2003.

[85] J. Le Guillou and J. Zinn-Justin, *Critical exponents for the n-vector model in three dimensions from field theory*, *Physical Review Letters* **39** (1977), no. 2 95.

[86] P.-G. De Gennes, *Scaling concepts in polymer physics*. Cornell university press, 1979.

[87] J. E. Kohn, I. S. Millett, J. Jacob, B. Zagrovic, T. M. Dillon, N. Cingel, R. S. Dothager, S. Seifert, P. Thiyagarajan, T. R. Sosnick, *et. al.*, *Random-coil behavior and the dimensions of chemically unfolded proteins*, *Proceedings of the National Academy of Sciences* **101** (2004), no. 34 12491–12496.

[88] D. K. Wilkins, S. B. Grimshaw, V. Receveur, C. M. Dobson, J. A. Jones, and L. J. Smith, *Hydrodynamic radii of native and denatured proteins measured by pulse field gradient nmr techniques*, *Biochemistry* **38** (1999), no. 50 16424–16431.

[89] L. D. Landau and E. M. Lifshitz, *Course of theoretical physics*. Elsevier, 2013.

[90] P. De Gennes, *Dynamics of entangled polymer solutions. i. the rouse model*, *Macromolecules* **9** (1976), no. 4 587–593.

[91] H. M. James and E. Guth, *Theory of the elastic properties of rubber*, *The Journal of Chemical Physics* **11** (1943), no. 10 455–481.

[92] W. Kuhn and F. Grün, *Beziehungen zwischen elastischen konstanten und dehnungsdoppelbrechung hochelastischer stoffe*, *Kolloid-Zeitschrift* **101** (1942) 248–271.

[93] B. Tinland, A. Pluen, J. Sturm, and G. Weill, *Persistence length of single-stranded dna*, *Macromolecules* **30** (1997), no. 19 5763–5765.

[94] O. A. Saleh, D. McIntosh, P. Pincus, and N. Ribeck, *Nonlinear low-force elasticity of single-stranded dna molecules*, *Physical review letters* **102** (2009), no. 6 068301.

[95] 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** (2017), no. 20 5095–5100.

[96] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith, *Entropic elasticity of λ-phage dna*, *Science* **265** (1994), no. 5178 1599–1600.

[97] F. Gittes, B. Mickey, J. Nettleton, and J. Howard, *Flexural rigidity of microtubules and actin filaments measured from thermal fluctuations in shape.*, *The Journal of cell biology* **120** (1993), no. 4 923–934.

[98] M. E. Janson and M. Dogterom, *A bending mode analysis for growing microtubules: evidence for a velocity-dependent rigidity*, *Biophysical journal* **87** (2004), no. 4 2723–2736.

[99] C. Bouchiat, M. D. Wang, J.-F. Allemand, T. Strick, S. Block, and V. Croquette, *Estimating the persistence length of a worm-like chain molecule from force-extension measurements*, *Biophysical journal* **76** (1999), no. 1 409–413.

[100] R. Schwarzl, S. Liese, F. N. Brünig, F. Laudisio, and R. R. Netz, *Force response of polypeptide chains from water-explicit md simulations*, *Macromolecules* **53** (2020), no. 12 4618–4629.

[101] Y.-B. Fu, Y.-R. Liu, P.-Y. Wang, and P. Xie, *A revised worm-like chain model for elasticity of polypeptide chains*, *Journal of Polymer Science Part B: Polymer Physics* **56** (2018), no. 4 297–307.

[102] M. Hillgärtner, K. Linka, and M. Itskov, *Worm-like chain model extensions for highly stretched tropocollagen molecules*, *Journal of biomechanics* **80** (2018) 129–135.

[103] X. Liao, P. K. Purohit, and A. Gopinath, *Extensions of the worm-like-chain model to tethered active filaments under tension*, *The Journal of Chemical Physics* **153** (2020), no. 19.

[104] J. P. Berezney and O. A. Saleh, *Electrostatic effects on the conformation and elasticity of hyaluronic acid, a moderately flexible polyelectrolyte*, *Macromolecules* **50** (2017), no. 3 1085–1089.

[105] T. Odijk, *Polyelectrolytes near the rod limit*, *Journal of Polymer Science: Polymer Physics Edition* **15** (1977), no. 3 477–483.

[106] J. Skolnick and M. Fixman, *Electrostatic persistence length of a wormlike polyelectrolyte*, *Macromolecules* **10** (1977), no. 5 944–948.

[107] C. Lu, F. Wu, W. Qiu, and R. Liu, *P130cas substrate domain is intrinsically disordered as characterized by single-molecule force measurements*, *Biophysical chemistry* **180** (2013) 37–43.

[108] M. Carrion-Vazquez, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez, *Atomic force microscopy captures length phenotypes in single proteins*, *Proceedings of the National Academy of Sciences* **96** (1999), no. 20 11288–11292.

[109] H. Li, A. F. Oberhauser, S. D. Redick, M. Carrion-Vazquez, H. P. Erickson, and J. M. Fernandez, *Multiple conformations of pevk proteins detected by single-molecule techniques*, *Proceedings of the National Academy of Sciences* **98** (2001), no. 19 10682–10686.

[110] A. Löf, P. U. Walker, S. M. Sedlak, S. Gruber, T. Obser, M. A. Brehm, M. Benoit, and J. Lipfert, *Multiplexed protein force spectroscopy reveals equilibrium protein folding dynamics and the low-force response of von willebrand factor*, *Proceedings of the National Academy of Sciences* **116** (2019), no. 38 18798–18807.

[111] D. B. Ritchie and M. T. Woodside, *Probing the structural dynamics of proteins and nucleic acids with optical tweezers*, *Current opinion in structural biology* **34** (2015) 43–51.

[112] 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** (2017), no. 1 702–712.

[113] H. P. Truong and O. A. Saleh, *[5] magnetic tweezers characterization of the entropic elasticity of intrinsically disordered proteins and peptoids*, in *Methods in enzymology*, vol. 694. Elsevier, 2024.

[114] C. Gosse and V. Croquette, *Magnetic tweezers: micromanipulation and force measurement at the molecular level*, *Biophysical journal* **82** (2002), no. 6 3314–3329.

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

[116] 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** (2013), no. 4.

[117] J. Lin and M. T. Valentine, *High-force ndfeb-based magnetic tweezers device optimized for microrheology experiments*, *Review of Scientific Instruments* **83** (2012), no. 5.

[118] J. Gelles, B. J. Schnapp, and M. P. Sheetz, *Tracking kinesin-driven movements with nanometre-scale precision*, *Nature* **331** (1988), no. 6155 450–453.

[119] D. R. Jacobson, *Electrostatics of single-stranded nucleic acids: Force spectroscopy and ion counting experiments*. 2016.

[120] S. N. Innes-Gold, *Structure and interactions of hyaluronic acid, a model polyelectrolyte*. PhD thesis, University of California, Santa Barbara, 2020.

[121] J. R. Warren and J. A. Gordon, *On the refractive indices of aqueous solutions of urea*, *The Journal of Physical Chemistry* **70** (1966), no. 1 297–300.

[122] Y. Nozaki, *[3] the preparation of guanidine hydrochloride*, in *Methods in enzymology*, vol. 26, pp. 43–50. Elsevier, 1972.

[123] C.-Y. Tan and Y.-X. Huang, *Dependence of refractive index on concentration and temperature in electrolyte solution, polar solution, nonpolar solution, and protein solution*, *Journal of Chemical & Engineering Data* **60** (2015), no. 10 2827–2833.

[124] 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** (2012), no. 2.

[125] I. L. Morgan and O. A. Saleh, *Tweezepy: A python package for calibrating forces in single-molecule video-tracking experiments*, *Plos one* **16** (2021), no. 12 e0262028.

[126] E. Schäffer, S. F. Nørrelykke, and J. Howard, *Surface forces and drag coefficients of microspheres near a plane surface measured with optical tweezers*, *Langmuir* **23** (2007), no. 7 3654–3665.

[127] I. De Vlaminck, T. Henighan, M. T. Van Loenhout, D. R. Burnham, and C. Dekker, *Magnetic forces and dna mechanics in multiplexed magnetic tweezers*, .

[128] J. M. Baskin and C. R. Bertozzi, *Bioorthogonal click chemistry: Covalent labeling in living systems*, *QSAR & Combinatorial Science* **26** (2007), no. 11-12 1211–1219.

[129] J. L. Zimmermann, T. Nicolaus, G. Neuert, and K. Blank, *Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments*, *nature protocols* **5** (2010), no. 6 975–985.

[130] R. Janissen, B. A. Berghuis, D. Dulin, M. Wink, T. van Laar, and N. H. Dekker, *Invincible dna tethers: Covalent dna anchoring for enhanced temporal and force stability in magnetic tweezers experiments*, *Nucleic acids research* **42** (2014), no. 18 e137–e137.

[131] H. Chen, H. Fu, X. Zhu, P. Cong, F. Nakamura, and J. Yan, *Improved high-force magnetic tweezers for stretching and refolding of proteins and short dna*, *Biophysical journal* **100** (2011), no. 2 517–523.

[132] 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** (2020), no. 5 058001.

[133] T. Hoffmann and L. Dougan, *Single molecule force spectroscopy using polyproteins*, *Chemical Society Reviews* **41** (2012), no. 14 4781–4796.

[134] I. Rasnik, S. A. McKinney, and T. Ha, *Surfaces and orientations: much to fret about?*, *Accounts of chemical research* **38** (2005), no. 7 542–548.

[135] B. Hua, R. Zhou, H. Kim, X. Shi, A. Jain, D. Singh, V. Aggarwal, and T. Ha, *An improved surface passivation method for single-molecule studies*, *Nature Methods* **11** (1233–1236), no. 12 393a.

[136] V. N. Uversky, C. J. Oldfield, and A. K. Dunker, *Intrinsically disordered proteins in human diseases: introducing the d2 concept*, *Annu. Rev. Biophys.* **37** (2008) 215–246.

[137] P. E. Wright and H. J. Dyson, *Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm*, *Journal of molecular biology* **293** (1999), no. 2 321–331.

[138] S.-H. Chong and S. Ham, *Folding free energy landscape of ordered and intrinsically disordered proteins*, *Scientific reports* **9** (2019), no. 1 14927.

[139] G. A. Papoian, *Proteins with weakly funneled energy landscapes challenge the classical structure–function paradigm*, *Proceedings of the National Academy of Sciences* **105** (2008), no. 38 14237–14238.

[140] H. Maity, L. Baidya, and G. Reddy, *Salt-induced transitions in the conformational ensembles of intrinsically disordered proteins*, *The Journal of Physical Chemistry B* **126** (2022), no. 32 5959–5971.

[141] J. Song, G.-N. Gomes, T. Shi, C. C. Gradinaru, and H. S. Chan, *Conformational heterogeneity and fret data interpretation for dimensions of unfolded proteins*, *Biophysical journal* **113** (2017), no. 5 1012–1024.

[142] U. Baul, D. Chakraborty, M. L. Mugnai, J. E. Straub, and D. Thirumalai, *Sequence effects on size, shape, and structural heterogeneity in intrinsically disordered proteins*, *The Journal of Physical Chemistry B* **123** (2019), no. 16 3462–3474.

[143] 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** (2016), no. 14 148101.

[144] W. Zheng, G. Dignon, M. Brown, Y. C. Kim, and J. Mittal, *Hydropathy patterning complements charge patterning to describe conformational preferences of disordered proteins*, *The journal of physical chemistry letters* **11** (2020), no. 9 3408–3415.

[145] 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** (2014), no. 3 659–666.

[146] J. Deek, P. J. Chung, and C. R. Safinya, *Neurofilament networks: Salt-responsive hydrogels with sidearm-dependent phase behavior*, *Biochimica et Biophysica Acta (BBA)-General Subjects* **1860** (2016), no. 7 1560–1569.

[147] A. S. Holehouse, R. K. Das, J. N. Ahad, M. O. Richardson, and R. V. Pappu, *Cider: resources to analyze sequence-ensemble relationships of intrinsically disordered proteins*, *Biophysical journal* **112** (2017), no. 1 16–21.

[148] 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** (2010), no. 1 40–46.

[149] A. Laser-Azogui, M. Kornreich, E. Malka-Gibor, and R. Beck, *Neurofilament assembly and function during neuronal development*, *Current opinion in cell biology* **32** (2015) 92–101.

[150] 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. 19 2464–2476.

[151] M. Buscaglia, L. J. Lapidus, W. A. Eaton, and J. Hofrichter, *Effects of denaturants on the dynamics of loop formation in polypeptides*, *Biophysical Journal* **91** (2006), no. 1 276–288.

[152] A. Bustamante, J. Sotelo-Campos, D. G. Guerra, M. Floor, C. A. M. Wilson, C. Bustamante, and M. Báez, *The energy cost of polypeptide knot formation and its folding consequences*, *Nature Communications* **8** (2017), no. 1 1581.

[153] V. R. Singh and L. J. Lapidus, *The intrinsic stiffness of polyglutamine peptides*, *The Journal of Physical Chemistry B* **112** (2008), no. 42 13172–13176.

[154] A. F. Chin, D. Toptygin, W. A. Elam, T. P. Schrank, and V. J. Hilser, *Phosphorylation increases persistence length and end-to-end distance of a segment of tau protein*, *Biophysical Journal* **110** (2016), no. 2 362–371.

[155] V. R. Singh, M. Kopka, Y. Chen, W. J. Wedemeyer, and L. J. Lapidus, *Dynamic similarity of the unfolded states of proteins l and g,*, *Biochemistry* **46** (09, 2007) 10046–10054.

[156] H. Schwalbe, K. M. Fiebig, M. Buck, J. A. Jones, S. B. Grimshaw, A. Spencer, S. J. Glaser, L. J. Smith, and C. M. Dobson, *Structural and dynamical properties of a denatured protein. heteronuclear 3d nmr experiments and theoretical simulations of lysozyme in 8 m urea*, *Biochemistry* **36** (07, 1997) 8977–8991.

[157] G. Damaschun, H. Damaschun, K. Gast, C. Gernat, and D. Zirwer, *Acid denatured apo-cytochrome c is a random coil: Evidence from small-angle x-ray scattering and dynamic light scattering*, *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1078** (1991), no. 2 289–295.

[158] J. A. Riback, M. A. Bowman, A. M. Zmyslowski, K. W. Plaxco, P. L. Clark, and T. R. Sosnick, *Commonly used fret fluorophores promote collapse of an otherwise disordered protein*, *Proceedings of the National Academy of Sciences* **116** (2019), no. 18 8889–8894.

[159] G. Morrison, C. Hyeon, N. M. Toan, B.-Y. Ha, and D. Thirumalai, *Stretching homopolymers*, *Macromolecules* **40** (2007), no. 20 7343–7353.

[160] P. Nanjan and M. Porel, *Sequence-defined non-natural polymers: synthesis and applications*, *Polymer Chemistry* **10** (2019), no. 40 5406–5424.

[161] M. A. Meier and C. Barner-Kowollik, *A new class of materials: Sequence-defined macromolecules and their emerging applications*, *Advanced Materials* **31** (2019), no. 26 1806027.

[162] A. S. Knight, E. Y. Zhou, M. B. Francis, and R. N. Zuckermann, *Sequence programmable peptoid polymers for diverse materials applications*, *Advanced Materials* **27** (2015), no. 38 5665–5691.

[163] M. Rubinstein and G. A. Papoian, *Polyelectrolytes in biology and soft matter*, *Soft Matter* **8** (2012), no. 36 9265–9267.

[164] K. S. Schanze and A. H. Shelton, *Functional polyelectrolytes*, *Langmuir* **25** (2009), no. 24 13698–13702.

[165] L.-W. Chang, T. K. Lytle, M. Radhakrishna, J. J. Madinya, J. Vélez, C. E. Sing, and S. L. Perry, *Sequence and entropy-based control of complex coacervates*, *Nature communications* **8** (2017), no. 1 1–8.

[166] A. M. Rumyantsev, N. E. Jackson, B. Yu, J. M. Ting, W. Chen, M. V. Tirrell, and J. J. De Pablo, *Controlling complex coacervation via random polyelectrolyte sequences*, *ACS Macro Letters* **8** (2019), no. 10 1296–1302.

[167] H. K. Murnen, A. M. Rosales, A. V. Dobrynin, R. N. Zuckermann, and R. A. Segalman, *Persistence length of polyelectrolytes with precisely located charges*, *Soft Matter* **9** (2013), no. 1 90–98.

[168] C. F. Anderson and M. T. Record Jr, *Polyelectrolyte theories and their applications to dna*, *Annual Review of Physical Chemistry* **33** (1982), no. 1 191–222.

[169] S. N. Innes-Gold, D. R. Jacobson, P. A. Pincus, M. J. Stevens, and O. A. Saleh, *Flexible, charged biopolymers in monovalent and mixed-valence salt: Regimes of anomalous electrostatic stiffening and of salt insensitivity*, *Physical Review E* **104** (2021), no. 1 014504.

[170] E. Buhler and F. Boue, *Chain persistence length and structure in hyaluronan solutions: Ionic strength dependence for a model semirigid polyelectrolyte*, *Macromolecules* **37** (2004), no. 4 1600–1610.

[171] M. Tricot, *Comparison of experimental and theoretical persistence length of some polyelectrolytes at various ionic strengths*, *Macromolecules* **17** (1984), no. 9 1698–1704.

[172] K. S. Schmitz and J. W. Yu, *On the electrostatic contribution to the persistence length of flexible polyelectrolytes*, *Macromolecules* **21** (1988), no. 2 484–493.

[173] K. Kassapidou, W. Jesse, M. Kuil, A. Lapp, S. Egelhaaf, and J. Van der Maarel, *Structure and charge distribution in dna and poly (styrenesulfonate) aqueous solutions*, *Macromolecules* **30** (1997), no. 9 2671–2684.

[174] D. R. Lide, *CRC handbook of chemistry and physics*, vol. 85. CRC press, 2004.

[175] E. C. Davidson, A. M. Rosales, A. L. Patterson, B. Russ, B. Yu, R. N. Zuckermann, and R. A. Segalman, *Impact of helical chain shape in sequence-defined polymers on polypeptoid block copolymer self-assembly*, *Macromolecules* **51** (2018), no. 5 2089–2098.

[176] K. Luo, J. Yang, P. Kopečková, and J. Kopeček, *Biodegradable multiblock poly [n-(2-hydroxypropyl) methacrylamide] via reversible addition- fragmentation chain transfer polymerization and click chemistry*, *Macromolecules* **44** (2011), no. 8 2481–2488.

[177] D. McIntosh and O. A. Saleh, *Salt species-dependent electrostatic effects on ssdna elasticity*, *Macromolecules* **44** (2011), no. 7 2328–2333.

[178] A. M. Rosales, H. K. Murnen, S. R. Kline, R. N. Zuckermann, and R. A. Segalman, *Determination of the persistence length of helical and non-helical polypeptoids in solution*, *Soft Matter* **8** (2012), no. 13 3673–3680.

[179] B. Yu, S. P. Danielsen, K.-C. Yang, R.-M. Ho, L. M. Walker, and R. A. Segalman, *Insensitivity of sterically defined helical chain conformations to solvent quality in dilute solution*, *ACS Macro Letters* **9** (2020), no. 6 849–854.

[180] Y. L. Angell and K. Burgess, *Peptidomimetics via copper-catalyzed azide–alkyne cycloadditions*, *Chemical Society Reviews* **36** (2007), no. 10 1674–1689.

[181] J. M. Holub and K. Kirshenbaum, *Tricks with clicks: modification of peptidomimetic oligomers via copper-catalyzed azide-alkyne [3+ 2] cycloaddition*, *Chemical Society Reviews* **39** (2010), no. 4 1325–1337.

[182] S. Förster, M. Schmidt, and M. Antonietti, *Experimental and theoretical investigation of the electrostatic persistence length of flexible polyelectrolytes at various ionic strengths*, *The Journal of Physical Chemistry* **96** (1992), no. 10 4008–4014.

[183] M. Nierlich, F. Boue, A. Lapp, and R. Oberthür, *Characteristic lengths and the structure of salt free polyelectrolyte solutions. a small angle neutron scattering study*, *Colloid and Polymer Science* **263** (1985) 955–964.

[184] M. J. Stevens and K. Kremer, *Form factor of salt-free linear polyelectrolytes*, *Macromolecules* **26** (1993), no. 17 4717–4719.

[185] M. J. Stevens and K. Kremer, *The nature of flexible linear polyelectrolytes in salt free solution: A molecular dynamics study*, *The Journal of chemical physics* **103** (1995), no. 4 1669–1690.

[186] L. Livadaru, R. Netz, and H. Kreuzer, *Stretching response of discrete semiflexible polymers*, *Macromolecules* **36** (2003), no. 10 3732–3744.

[187] A. V. Dobrynin, J.-M. Y. Carrillo, and M. Rubinstein, *Chains are more flexible under tension*, *Macromolecules* **43** (2010), no. 21 9181–9190.

[188] J. C. Salamone, *Polymeric materials encyclopedia, Twelve volume set*. CRC press, 2020.

[189] A. Almond, A. Brass, and J. Sheehan, *Oligosaccharides as model systems for understanding water- biopolymer interaction: Hydrated dynamics of a hyaluronan decamer*, *The Journal of Physical Chemistry B* **104** (2000), no. 23 5634–5640.

[190] H. Lu, J. Wang, Y. Bai, J. W. Lang, S. Liu, Y. Lin, and J. Cheng, *Ionic polypeptides with unusual helical stability*, *Nature communications* **2** (2011), no. 1 206.

[191] M. Ghelichi and M. H. Eikerling, *Conformational properties of comb-like polyelectrolytes: a coarse-grained md study*, *The Journal of Physical Chemistry B* **120** (2016), no. 10 2859–2867.

[192] D. Tolmachev, N. Lukasheva, G. Mamistvalov, and M. Karttunen, *Influence of calcium binding on conformations and motions of anionic polyamino acids. effect of side chain length*, *Polymers* **12** (2020), no. 6 1279.

[193] C. N. Pace and J. M. Scholtz, *A helix propensity scale based on experimental studies of peptides and proteins*, *Biophysical journal* **75** (1998), no. 1 422–427.

[194] H.-T. Kuo, S.-L. Liu, W.-C. Chiu, C.-J. Fang, H.-C. Chang, W.-R. Wang, P.-A. Yang, J.-H. Li, S.-J. Huang, S.-L. Huang, *et. al.*, *Effect of charged amino acid side chain length on lateral cross-strand interactions between carboxylate-and guanidinium-containing residues in a β-hairpin*, *Amino Acids* **47** (2015) 885–898.

[195] L.-H. Kuo, J.-H. Li, H.-T. Kuo, C.-Y. Hung, H.-Y. Tsai, W.-C. Chiu, C.-H. Wu, W.-R. Wang, P.-A. Yang, Y.-C. Yao, *et. al.*, *Effect of charged amino acid side chain length at non-hydrogen bonded strand positions on β-hairpin stability*, *Biochemistry* **52** (2013), no. 44 7785–7797.

[196] H. Meuzelaar, J. Vreede, and S. Woutersen, *Influence of glu/arg, asp/arg, and glu/lys salt bridges on α-helical stability and folding kinetics*, *Biophysical journal* **110** (2016), no. 11 2328–2341.

[197] S. P. Danielsen, E. C. Davidson, G. H. Fredrickson, and R. A. Segalman, *Absence of electrostatic rigidity in conjugated polyelectrolytes with pendant charges*, *ACS Macro Letters* **8** (2019), no. 9 1147–1152.

[198] P. Debye and E. Hückel, *Zur theorie der elektrolyte. i. gefrierpunktserniedrigung und verwandte erscheinungen. phys*, *Zeitschrift* **24** (1923) 185–206.

[199] H. Dietz and M. Rief, *Protein structure by mechanical triangulation*, *Proceedings of the National Academy of Sciences* **103** (2006), no. 5 1244–1247.

[200] H. P. Erickson, *Reversible unfolding of fibronectin type iii and immunoglobulin domains provides the structural basis for stretch and elasticity of titin and fibronectin.*, *Proceedings of the National Academy of Sciences* **91** (1994), no. 21 10114–10118.

[201] G. Yang, C. Cecconi, W. A. Baase, I. R. Vetter, W. A. Breyer, J. A. Haack, B. W. Matthews, F. W. Dahlquist, and C. Bustamante, *Solid-state synthesis and mechanical unfolding of polymers of t4 lysozyme*, *Proceedings of the National Academy of Sciences* **97** (2000), no. 1 139–144.

[202] B. C. Hudson, A. Battigelli, M. D. Connolly, J. Edison, R. K. Spencer, S. Whitelam, R. N. Zuckermann, and A. K. Paravastu, *Evidence for cis amide bonds in peptoid nanosheets*, *The journal of physical chemistry letters* **9** (2018), no. 10 2574–2578.

[203] Y. Ryu and P. G. Schultz, *Efficient incorporation of unnatural amino acids into proteins in escherichia coli*, *Nature methods* **3** (2006), no. 4 263–265.

[204] M. J. Shon, S.-H. Rah, and T.-Y. Yoon, *Submicrometer elasticity of double-stranded dna revealed by precision force-extension measurements with magnetic tweezers*, *Science advances* **5** (2019), no. 6 eaav1697.

[205] E.-M. Mandelkow and E. Mandelkow, *Biochemistry and cell biology of tau protein in neurofibrillary degeneration*, *Cold Spring Harbor perspectives in medicine* **2** (2012), no. 7.

[206] V. N. Uversky and D. Eliezer, *Biophysics of parkinson's disease: structure and aggregation of α-synuclein*, *Current Protein and Peptide Science* **10** (2009), no. 5 483–499.

[207] E. Mylonas, A. Hascher, P. Bernado, M. Blackledge, E. Mandelkow, and D. I. Svergun, *Domain conformation of tau protein studied by solution small-angle x-ray scattering*, *Biochemistry* **47** (2008), no. 39 10345–10353.

[208] A. J. Trexler and E. Rhoades, *Single molecule characterization of α-synuclein in aggregation-prone states*, *Biophysical journal* **99** (2010), no. 9 3048–3055.

[209] M. M. Dedmon, K. Lindorff-Larsen, J. Christodoulou, M. Vendruscolo, and C. M. Dobson, *Mapping long-range interactions in α-synuclein using spin-label nmr and ensemble molecular dynamics simulations*, *Journal of the American Chemical Society* **127** (2005), no. 2 476–477.

[210] P. G. Higgs and J.-F. Joanny, *Theory of polyampholyte solutions*, *The Journal of chemical physics* **94** (1991), no. 2 1543–1554.

[211] J. J. Madinya, L.-W. Chang, S. L. Perry, and C. E. Sing, *Sequence-dependent self-coacervation in high charge-density polyampholytes*, *Molecular Systems Design & Engineering* **5** (2020), no. 3 632–644.

[212] S. E. Kudaibergenov, *Synthetic and natural polyampholytes: Structural and behavioral similarity*, *Polymers for Advanced Technologies* **32** (2021), no. 3 906–918.

[213] W. H. Shi, R. S. Adhikari, D. N. Asthagiri, and A. B. Marciel, *Influence of charge block length on conformation and solution behavior of polyampholytes*, *ACS Macro Letters* **12** (2023), no. 2 195–200.

[214] I. L. Morgan, *Glassy dynamics in an intrinsically disordered protein region.* University of California, Santa Barbara, 2021.

[215] G. M. Figliozzi, R. Goldsmith, S. C. Ng, S. C. Banville, and R. N. Zuckermann, *[25] synthesis of n-substituted glycine peptoid libraries*, in *Methods in enzymology*, vol. 267, pp. 437–447. Elsevier, 1996.

[216] L. Li and Z. Zhang, *Development and applications of the copper-catalyzed azide-alkyne cycloaddition (cuaac) as a bioorthogonal reaction*, *Molecules* **21** (2016), no. 10 1393.

[217] D. Flood, C. Proulx, E. J. Robertson, A. Battigelli, S. Wang, A. M. Schwartzberg, and R. N. Zuckermann, *Improved chemical and mechanical stability of peptoid nanosheets by photo-crosslinking the hydrophobic core*, *Chemical Communications* **52** (2016), no. 26 4753–4756.

[218] I. Borukhov, D. Andelman, R. Borrega, M. Cloitre, L. Leibler, and H. Orland, *Polyelectrolyte titration: theory and experiment*, *The Journal of Physical Chemistry B* **104** (2000), no. 47 11027–11034.

[219] E. R. McCarney, J. E. Kohn, and K. W. Plaxco, *Is there or isn't there? the case for (and against) residual structure in chemically denatured proteins*, *Critical Reviews in Biochemistry and Molecular Biology* **40** (2005), no. 4 181–189.

[220] B. E. Bowler, *Residual structure in unfolded proteins*, *Current opinion in structural biology* **22** (2012), no. 1 4–13.

[221] K. A. Dill and D. Shortle, *Denatured states of proteins*, *Annual review of biochemistry* **60** (1991), no. 1 795–825.

[222] D. Shortle, *The denatured state (the other half of the folding equation) and its role in protein stability*, *The FASEB Journal* **10** (1996), no. 1 27–34.

[223] S. L. Crick, M. Jayaraman, C. Frieden, R. Wetzel, and R. V. Pappu, *Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions*, *Proceedings of the National Academy of Sciences* **103** (2006), no. 45 16764–16769.

[224] 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** (2007), no. 8 2649–2654.

[225] N. Matsushima, H. Yoshida, Y. Kumaki, M. Kamiya, T. Tanaka, Y. Izumi, and R. H. Kretsinger, *Flexible structures and ligand interactions of tandem repeats consisting of proline, glycine, asparagine, serine, and/or threonine rich oligopeptides in proteins*, *Current Protein and Peptide Science* **9** (2008), no. 6 591–610.

[226] V. P. Reddy Chichili, V. Kumar, and J. Sivaraman, *Linkers in the structural biology of protein–protein interactions*, *Protein science* **22** (2013), no. 2 153–167.

[227] G. Sachetto-Martins, L. O. Franco, and D. E. de Oliveira, *Plant glycine-rich proteins: a family or just proteins with a common motif?*, *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression* **1492** (2000), no. 1 1–14.

[228] M. Van Rosmalen, M. Krom, and M. Merkx, *Tuning the flexibility of glycine-serine linkers to allow rational design of multidomain proteins*, *Biochemistry* **56** (2017), no. 50 6565–6574.

[229] W. Wriggers, S. Chakravarty, and P. A. Jennings, *Control of protein functional dynamics by peptide linkers*, *Peptide Science: Original Research on Biomolecules* **80** (2005), no. 6 736–746.

[230] X. Chen, J. L. Zaro, and W.-C. Shen, *Fusion protein linkers: property, design and functionality*, *Advanced drug delivery reviews* **65** (2013), no. 10 1357–1369.

[231] A. S. Holehouse, K. Garai, N. Lyle, A. Vitalis, and R. V. Pappu, *Quantitative assessments of the distinct contributions of polypeptide backbone amides versus side chain groups to chain expansion via chemical denaturation*, *Journal of the American Chemical Society* **137** (2015), no. 8 2984–2995.

[232] D. Asthagiri, D. Karandur, D. S. Tomar, and B. M. Pettitt, *Intramolecular interactions overcome hydration to drive the collapse transition of gly15*, *The Journal of Physical Chemistry B* **121** (2017), no. 34 8078–8084.

[233] D. P. Teufel, C. M. Johnson, J. K. Lum, and H. Neuweiler, *Backbone-driven collapse in unfolded protein chains*, *Journal of molecular biology* **409** (2011), no. 2 250–262.

[234] S. Bykov and S. Asher, *Raman studies of solution polyglycine conformations*, *The Journal of Physical Chemistry B* **114** (2010), no. 19 6636–6641.

[235] E. Warkentin, S. Weidenweber, K. Schühle, U. Demmer, J. Heider, and U. Ermler, *A rare polyglycine type ii-like helix motif in naturally occurring proteins*, *Proteins: Structure, Function, and Bioinformatics* **85** (2017), no. 11 2017–2023.

[236] M. Lorusso, A. Pepe, N. Ibris, and B. Bochicchio, *Molecular and supramolecular studies on polyglycine and poly-l-proline*, *Soft Matter* **7** (2011), no. 13 6327–6336.

[237] K. A. H. Wildman, E. E. Wilson, D.-K. Lee, and A. Ramamoorthy, *Determination of the conformation and stability of simple homopolypeptides using solid-state nmr*, *Solid State Nuclear Magnetic Resonance* **24** (2003), no. 2-3 94–109.

[238] R. R. Hudgins and M. F. Jarrold, *Conformations of unsolvated glycine-based peptides*, *The Journal of Physical Chemistry B* **104** (2000), no. 9 2154–2158.

[239] T. Takekiyo, N. Yamada, T. Amo, C. T. Nakazawa, A. Asano, T. Ichimura, M. Kato, and Y. Yoshimura, *Dissolution of amyloid aggregates by direct addition of alkali halides*, *Journal of Molecular Liquids* **369** (2023) 120849.

[240] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry, Part I: The Conformation of Biological Molecules*. Macmillan, 1980.

[241] F. Hanke, A. Serr, H. J. Kreuzer, and R. R. Netz, *Stretching single polypeptides: The effect of rotational constraints in the backbone*, *Europhysics Letters* **92** (2010), no. 5 53001.

[242] K. G. Wilcox, M. E. Dingle, A. Saha, M. J. Hore, and S. Morozova, *Persistence length of α-helical poly-l-lysine*, *Soft Matter* **18** (2022), no. 35 6550–6560.

[243] P. R. Schimmel and P. J. Flory, *Conformational energy and configurational statistics of poly-l-proline.*, *Proceedings of the National Academy of Sciences* **58** (1967), no. 1 52–59.

[244] D. E. Smith, T. T. Perkins, and S. Chu, *Dynamical scaling of dna diffusion coefficients*, *Macromolecules* **29** (1996), no. 4 1372–1373.