The Raymond and Beverly Sackler Faculty of Exact Sciences
Tel Aviv University

THE SCHOOL OF PHYSICS AND ASTRONOMY

The role of sequence heterogeneity on the statistical structure of the intrinsically disordered Neurofilament-low tail domain

Thesis submitted in partial fulfillment of the requirements for the M.Sc. degree in the School of School of Physics and Astronomy, Tel Aviv University

Matar Kravicas

The research work for the thesis has been carried out under the supervision of

Prof. Roy Beck

February 2024

Acknowledgements

I am thankful to the faculty and staff at Tel Aviv University for providing me with the opportunity to pursue my studies and for supporting me along the way. This thesis would not have been possible without the help of present and past lab members - Sagi, Rawan, Alon, Vaishali, and Amandeep.

I would also like to extend special thanks to Omar Saleh, Hoang (Frank) Truong, and the rest of the Saleh lab group for their immense hospitality and help during my stay at the University of California, Santa Barbara, where I performed the magnetic tweezers experiments.

I especially want to thank Gil Koren, our lab manager, for teaching me how to perform biological work, conduct experiments, and much more. His willingness to patiently discuss any question I had helped me more times than I could hope to remember. I am very grateful for his significant assistance with experiments and insight into physics and biology theories, which were instrumental to this thesis and our published work.

Finally, I want to thank my supervisor, Roy Beck, for his support. His guidance helped me greatly with this work and, most importantly, developing a vast set of tools to become a better researcher. At every moment he would offer help, from bureaucracy to discussion, for which I am immensely grateful. The countless opportunities he provided me, including traveling to synchrotron facilities, visiting other research centers, funding, and much more, greatly broadened my experience and outlook on my work and research in general.

Abstract

Intrinsically disordered proteins (IDPs) are a subset of proteins that lack stable secondary structure. Given IDPs' polymeric nature, mean-field approximations have been previously used to describe their statistical structure. However, the amino-acid sequence heterogeneity and complex intermolecular interaction networks have significantly impeded the ability to get proper approximations from such theories.

In my study, I focused on the intrinsically disordered tail domain of Neurofilament low (NFLt). This polypeptide sequence comprises a 50 residue-long uncharged domain followed by a 96 residue-long negatively charged domain. I will show how sequence composition affects the complex interaction network and statistical structure of IDPs using NFLt as a model system.

Specifically, I assess the impact of the two NFLt domains in and out of equilibrium conditions using two variants of the IDP. Using synchrotron small-angle x-ray scattering, I find that in equilibrium, the uncharged domain of the NFLt induces attractive interactions that cause it to self-assemble into star-like polymer brushes. These brushes are shown to maintain their stability in increasing electrostatic screening, following Pincus' salted brush regime of polyelectrolyte brushes. On the other hand, when the uncharged domain is truncated, the remaining charged N-terminal domains remain isolated in solution with typical polyelectrolyte characteristics. Such behavior for both variants is due to the complex short- and long-range interactions within the IDP ensembles that change significantly in the presence of electrostatic screening. Lastly, I compare these equilibrium results to previous studies on NFLs and NFLts to draw further conclusions on their inner mechanisms.

Outside of equilibrium, I conducted magnetic tweezer experiments stretching and releasing the NFLt constructs attached to a support surface and a magnetic bead. This study follows up on previous experiments, which showed that the NFLt exhibits glassy dynamics upon stretching and relaxation. These experiments demonstrated that multiple independently relaxing segments within the NFLt were responsible for these dynamics. Here, the study is repeated using the two variants to attribute these independently relaxing domains to the NFLt sub-regimes.

This thesis is mostly based on my recently accepted work [1], also attached as supplementary. This work concludes with an outlook for future work and the hope of linking IDP sequences to functional, non-conventional polymers.

Contents

1 Introduction **6**

1.1 Proteins . . . 6

1.2 Intrinsically disordered proteins . . . 8

1.2.1 What biological functions do IDPs serve? . . . 8

1.2.2 How are IDPs measured and characterized? . . . 9

1.2.3 Primary structure- the IDP sequence . . . 12

1.3 Neurofilaments (NFs), and the NF-low disordered tail domain (NFLt) . . . 14

2 Polymer Physics **16**

2.1 Ideal Chains . . . 16

2.2 Real Chains . . . 19

2.3 Polyampholytes . . . 20

3 Polymer Brush Physics **28**

3.1 Neutral Brushes . . . 29

3.2 Charged Brushes . . . 31

4 Small Angle X-ray Scattering (SAXS) **34**

4.1 Basics . . . 34

4.2 Form Factor Analysis . . . 37

4.2.1 Form Factors . . . 38

4.2.2 Structural Disorder . . . 39

4.3 SAXS analysis of polymers . . . 42

4.3.1 SAXS scattering of interacting polymers . . . 43

5 Magnetic Tweezers **45**

5.1 Overview . . . 45

5.2 Force calibration . . . 46

5.3 Polymer Stretching . . . 47

5.3.1 Scaling . . . 47
5.3.2 Worm-like chain (WLC) . . . 48
5.4 NFL glassy dynamics . . . 49

6 Results **53**
6.1 Experimental setup . . . 53
6.2 The Truncated NFLt . . . 53
6.2.1 Raw data . . . 54
6.2.2 Scaling . . . 57
6.2.3 Intermolecular interactions . . . 59
6.3 WT aggregates . . . 61
6.3.1 Raw data . . . 61
6.3.2 Modeling the WT aggregation . . . 61
6.3.3 WT scaling . . . 70
6.3.4 Intermolecular interactions . . . 74
6.3.5 Additional measurements . . . 76
6.4 Comparison to segmented NFLt . . . 77
6.5 Dynamics . . . 79
6.5.1 Magnetic tweezers setup . . . 79
6.5.2 Two-step experiment . . . 79
6.5.3 Three-step experiment . . . 84

7 Experimental Methods **86**
7.1 Protein Purification . . . 86
7.2 SAXS Measurements . . . 88
7.2.1 Measurements. . . . 88
7.2.2 Analysis. . . . 89
7.3 Magnetic tweezers setup . . . 90
7.4 Tether tracking experiments . . . 92

7.5 Tether data analysis . . . 92

8 Discussion **93**

8.1 Interactions . . . 93
8.2 Scaling . . . 94
8.3 Comparisons to past studies . . . 94
8.4 Outside of equilibrium . . . 95
8.5 Research limitations . . . 95
8.6 Final remarks . . . 96

9 Future Work **97**

9.1 Link between sequence and interactions . . . 97
9.2 Theoretical models . . . 97
9.3 Dimerization of NFLt chains . . . 97
9.4 Additional magnetic tweezers experiments . . . 97
9.5 Applications of self-assembling IDP brushes . . . 98

10 Code **107**

10.1 Analyses and models . . . 107
10.2 Form factors . . . 108

11 Appendix **115**

11.3 Paper . . . 116
11.4 Paper supplementary . . . 125

1 Introduction

1.1 Proteins

Proteins play many crucial roles in biological systems, encompassing functions such as signaling [2], gene regulation [3], transmission [4], and more. This functional diversity is generally attributed to the unique composition of each protein and its interaction with the surrounding environment. Proteins are essentially linear chains of amino acids, a group of twenty molecules with distinct properties. The amino acid composition of proteins is commonly referred to as their sequence or primary structure. The well-known principle in protein research is that proteins fold into specific, stable structures known as their secondary structure [5]. Once folded, a protein interacts with other entities within the biological system by its specific function [6]. This interaction is often described as a lock-and-key mechanism due to the strong correlation between structural specificity and functionality.

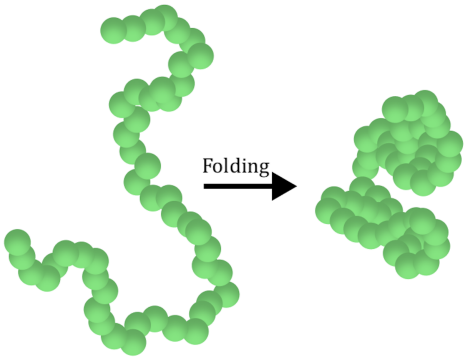

Figure 1.1: On the left: a schematic of an unfolded protein, a linear chain of amino acids (in green). Undergoing folding, the protein shifts into a unique and stable structure, as illustrated on the right.

The first work to successfully model the folded protein Myoglobin was conducted using X-ray crystallography[7]. This work paved the way for the rising field of structural biology, which significantly advanced our understanding of protein folding and function over the years [8]. Today, we recognize that folding involves various processes influenced by amino acid characteristics [8]. One of the most well-studied of these processes is within the protein's measure of hydropathy, in which the amino acids conform into structures based on their 'compatibility' with the solvent [9]. Proteins often exhibit multiple hydropathic amino acids, organizing based on hydrophobic and

hydrophilic moments [10].

The apparent link between the protein's primary and secondary structure has led to many studies trying to find theoretical models and simulations that can predict secondary structure given the sequence [11] (Fig. 1.2). More recently, artificial intelligence (AI) models with access to large databases with thousands of cataloged proteins have shown great success in predicting folding. In particular, AlphaFold is known to have been trained on over 170 thousand protein structures and can tell the structure of specific domains within the protein with high accuracy ratings [12]. Nonetheless, many proteins possess domains for which experiments and AI models fail to predict their structure. As AI models, their results are solely determined by the data they are fed. As such, this failure in predicting the structure of such domains indicates a lack of data on structures of similar protein sequences. However, the inability to experimentally resolve the unique 3D structure of those proteins remained a mystery for many years [13]. This work experimentally studies such intrinsically disordered protein (IDP), which lacks a stable 3D structure in its native form.

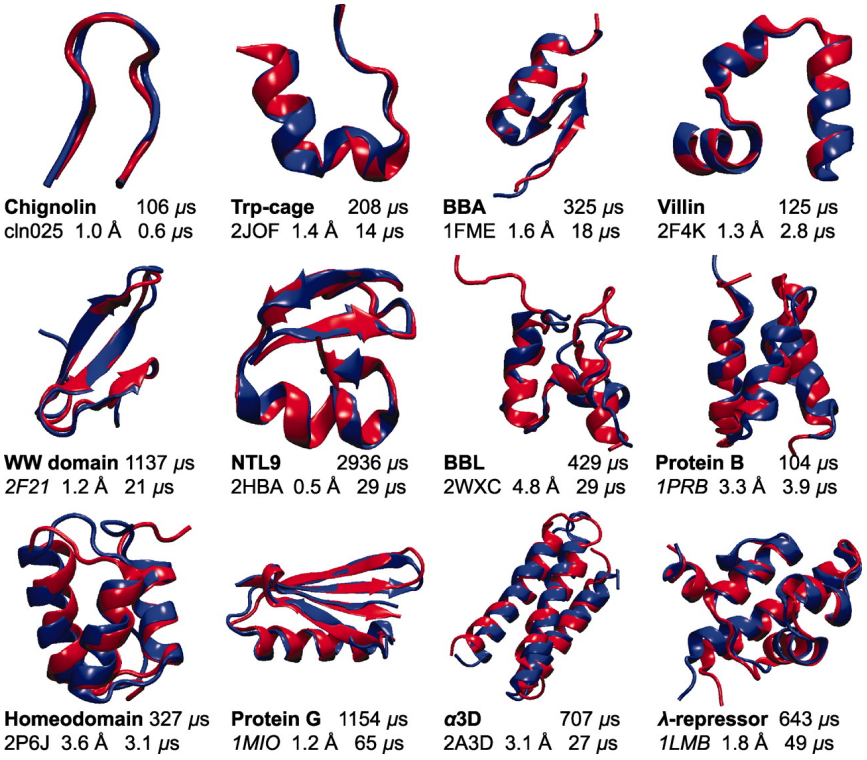

Figure 1.2: Comparisons between simulation result of protein folding (red) against measured structures (blues). Figure adapted from [11].

1.2 Intrinsically disordered proteins

Even as early as the 1950s [14], studies have implicated the notion that not all proteins, or regions of proteins, possess the ability to conform to a stable structure, regardless of environmental conditions [15]. Instead, depending on their sequences, these proteins constantly fluctuate between alternative structures within an ensemble of conformations[16]. Following the previously described structural biology dogma, the lack of secondary structure may imply that these proteins lack a function. *However, the assumption of uniquely correlating structures to functions is increasingly being proven to be false* with the discoveries of many such proteins that serve significant roles in biological systems [17, 18]. One prominent example of such proteins is the tumor suppressor P53, which was shown to participate in numerous processes within cells, even though it does not fulfill the requirement of a singular stable 3D structure in native conditions [19].

These proteins are commonly referred to as 'Intrinsically disordered proteins' (IDPs) or as proteins having 'Intrinsically disordered regions' (IDRs) if part of the chains do not uniquely fold to a stable structure. From here, we face two questions that limit our understanding of IDPs: *What biological functions do they serve? And how can we measure and characterize them?*

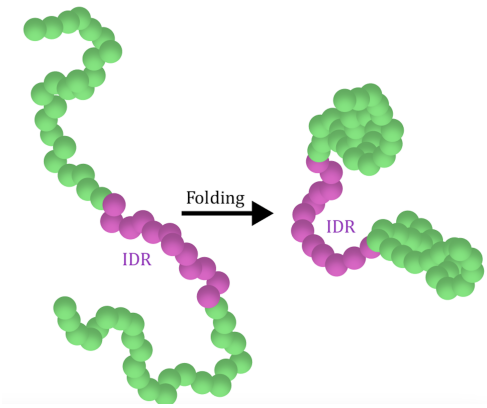

Figure 1.3: Schematic of protein folding. Green spheres represent amino acids that exhibit folding, while purple circles represent the IDR of the protein.

1.2.1 What biological functions do IDPs serve?

The increasingly apparent importance of IDPs has led to a new paradigm of protein science - the 'disorder-function' paradigm. This paradigm contrasts the well-researched 'structure-function' paradigm, as its basis lies in the ability of the IDPs to alter between multiple conformations. Nowadays, the functions of IDPs are divided into 28 groups, a number which can only grow with

future discoveries [17]. Here, I will focus on two major 'archetypes', broadly illustrating how an IDP/IDR may function.

Non-binding IDPs/IDRs: Non-binding IDPs/IDRs do not possess a binding interface to other molecules. Instead, they rely solely on their conformational freedom to perform their biological function [17]. They can either appear within the protein (Fig. 1.3) or on its carboxy and/or amid termini (Fig. 1.4A). One of their primary functions is to act as spacers, which control the distance between two protein domains or between the protein and outside elements. For example, the microtube-associated protein 2 (MAP2) projection domain repels any molecule approaching the microtubes, thus regulating the spacing between these important cytoskeleton components [20].

Binding IDPs/IDRs: Binding IDPs/IDRs rely not only on their conformational flexibility but also on their ability to bind to other proteins or molecules to perform their functions [17] (Fig. 1.4B). Binding IDPs/IDRs are further divided into two categories, relating to whether their bond is permanent or not. A prominent example of a binding IDP is the Tau protein, which was shown to bind to microtubes and promote their self-assembly [21].

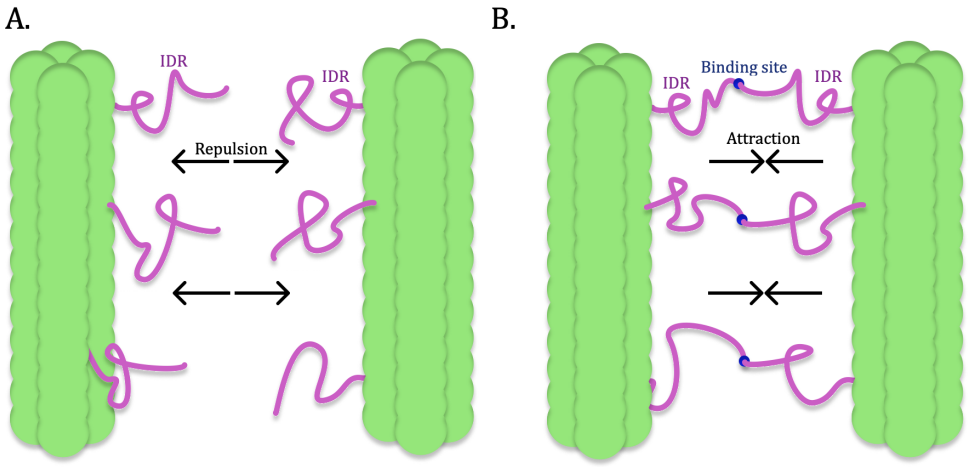

Figure 1.4: **A.** Illustration of IDRs which act as spacers without binding to other molecules. **B.** Illustration of IDRs which bind to other protein to form larger structures.

1.2.2 How are IDPs measured and characterized?

Being flexible chains, IDPs can not be characterizable using the same manners as folded proteins. Measurement techniques such as X-ray scattering are often applicable; however, due to the high flexibility of the IDPs, the amount of obtainable information becomes limited, requiring

multiple measurement techniques to characterize all aspects of a given IDP. Below, I will detail three such techniques and their advantages and disadvantages in measuring IDPs. In this work, I used two of these techniques (SAXS and magnetic tweezers), and thus, further technical details will be given in sections 4 and 5.

X-ray scattering. As one of the oldest used measurement methods, nowadays, X-ray scattering is a commonly used technique for measuring IDPs [22] (Illustration Fig. 1.5A). In general terms, X-ray scattering records the interference pattern resulting from a collimated X-ray beam and the sample. The length scales at which X-ray scattering measures then depend on the maximum angle of the scattering, as determined by the distance between the sample and the detector. For soluble polymers or proteins, the sample solution contains a large macroscopic amount of proteins, each measured at a different conformation and orientation to the incoming beam. As such, the scattering pattern can only capture the sample structural ensemble average. While this may be a limited issue for stable structures, for IDPs, this means that we can only capture the ensemble average of their conformations. This issue encompasses the most significant disadvantage of X-ray scattering measurements. However, due to recent developments in IDP research, much information can still be extracted using X-ray scattering techniques, lending it significant use in the field. Due to the relevance of this technique to this thesis, more information about X-ray scattering can be found in section 4.

Förster Resonance Energy Transfer (FRET). Using fluorescent markers attached/part of the protein sequence, FRET measures the distance between the two fluorescent molecules [23] (Illustration Fig. 1.5B). The simplicity of this method allows it to be used in numerous applications, such as protein-protein interaction and folding [24, 25]. Specifically for IDPs, this method can be used to measure the end-to-end radius between any two residues, which in conjuncture to appropriate theoretical models (see section 2) can be used to obtain important information on the statistics of their conformations [26, 27]. FRET measurements can also discern protein folding and disorder under specified environmental conditions [27, 25]. The ability of FRET to measure distances gives it a distinct advantage over X-ray scattering in its ability to study the distance between any two points, be it within the IDP by itself or with other entities in the solution. Additionally, FRET experiments require a smaller quantity of proteins and can even perform single-molecule experiments [27]. However, FRET experiments require the sample to be modified to be measured, which could impact the results compared to X-ray scattering. Additionally, FRET measurements can only measure distances up to 10 nanometres, a much smaller scale than in X-ray scattering.

Magnetic Tweezers (MT). MTs possess the unique ability to measure the mechanical response of biomolecules, in addition to their dynamics [28]. MTs can measure the distance of the IDPs from the surface to which they are attached and their response to external force [29, 30] (Illustration Fig. 1.5C). As an advantage over FRET and X-ray scattering, MTs enable us to perform specific tension-response experiments for IDPs with minimal altercation to the sequence. Additionally, as single-molecule experiments, MT experiments allow us to measure the IDP's

kinetics without interactions that might occur in a populated solution. This aspect, however, also encompasses one of their most significant disadvantages, as single-molecule experiments are prone to many external issues that make it difficult to assess the validity of the results without an extensive enough study.

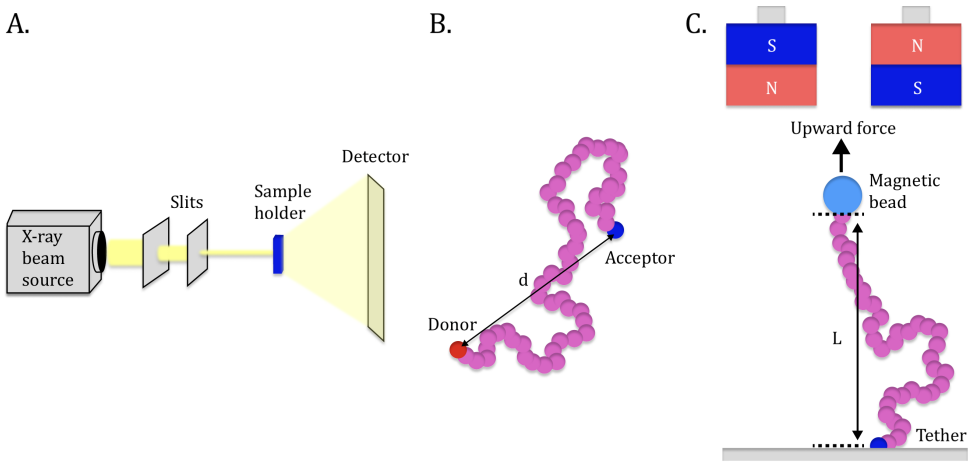

Figure 1.5: **A.** Basic illustration of an X-ray setup. X-ray beam sent from the source which is focused by two slits. The focused beam scatters off the sample, and gets recorded on the detector screen. **B.** Simple illustration of an IDP within a FRET experiment. Measurements of the fluorescent decays of the labled residues (in red and blue) are used to determine the distance between them d. **C.** Simple illustration of the magnetic tweezers (MT) setup. The IDP chain is tethered to a substrate and a magnetic bead on its open end. Two magnets exert a controllable upwards force on the bead, which in turn changes the measured bead height L.

Depending on the situation, these methods hold unique properties that make one preferable over the other. However, without a cohesive theoretical backbone, fully understanding the conformational ensemble of the IDPs could prove difficult. To that end, polymer theories are viewed as prime candidates for IDP characterization due to the IDPs' apparent similarities to polymers [31, 32, 33, 17]. Recent years have even shown a rise in models sophisticated enough to fit the properties of IDPs (charge distribution, hydropathy) and their ensemble structure behavior [34, 35, 36, 37, 38]. However, the difference between IDPs and polymers is still quite significant, as IDPs are typically heterogeneous in their sequence, while polymeric models are mostly homogeneous. By now, this difference was shown to be a key issue when using homopolymer models on some IDPs [32, 34, 26, 39, 40]. Sections 2 and 3 will focus on some of these models and their degree of success with IDPs.

1.2.3 Primary structure- the IDP sequence

With increased knowledge based on experiments and theory, the characteristics of the comprising amino acid residues are known to be key in influencing the folding process, or lack thereof. Below, I will concentrate on some of these characteristics and how they may lead to order or disorder within the protein.

Hydrophobicity. Hydrophobic interactions are typically linked to protein folding because they can arrange the protein chains in specific conformations. Hydrophobic amino acids are abundant in folded proteins and comprise the folded core of the stable protein structure. These hydrophobic cores are further stabilized by weak, short-range interactions referred to as van der Waals forces. In contrast to folded proteins, IDPs feature much fewer hydrophobic residues (Fig. 1.6), contributing to their disorder [17] and improving their solubility in water [39]. At high temperatures, however, hydrophobic interactions grow stronger, which can lead to partial folding in some IDPs.

Quantifying the strength of the hydrophobic interactions is typically done using hydrophobic scales [41]. In one such normalized scales, each amino acid is assigned a specific hydropathy value λ, which ranges between 0 and 1 [41, 35] (Table 1.2.3). Such scales can be used to predict the disorder propensity of proteins given their sequence [42] and to estimate conformational preferences of intrinsically disordered proteins [35].

Amino acid	R	H	K	D	E	S	T	N	Q	C
λ	0	0.514	0.514	0.378	0.459	0.595	0.676	0.432	0.514	0.595
Amino acid	G	P	A	I	L	M	F	W	Y	V
λ	0.649	1	0.73	0.973	0.973	0.838	1	0.946	0.865	0.892

Table 1: Table containing hydropathy value λ for all Amino acids, as provided by [41]

Electric charge. Electrostatic interactions are unique in their ability to repel/attract residues of similar/opposite charges over relatively large distances in salt-free solutions. These interactions enable residues to influence internal interactions within the proteins (intramolecular) and external with other proteins or entities (intermolecular). Within folded proteins, electrostatic forces contribute to protein stability through short-range effects such as salt bridging, where opposing charges strengthen the stability of hydrophobic bonds.

In contrast to hydrophobic residues, charged residues are much more prominent in IDPs than in folded proteins. These charged residues typically appear in repeats of similarly charged residues, contributing to the disorder of the chain by intramolecular repulsion. Within IDPs, electrostatic interactions were proven immensely influential over their interaction landscape, correlating the sequence of the charges within the protein and its conformations [37, 38]. Electrostatic effects are also highly susceptible to charge screening, which can alter the IDP conformations and incur

unfolding (denaturation) within certain folded proteins [27, 37].

Five amino acids possess a charge. Three are positively charged (K, R, H), and two are negatively charged (D, E). A useful metric to describe the average charge of an IDP is the net charge per residue (NCPR), which is defined as [43]:

$$\text{NCPR} = \mid \sum_{i}^{N} Q_i \mid, \tag{1.1}$$

where Q_i is the charge contribution of the ith residue.

In addition to these two archetypes, certain amino acids may possess additional properties. Cysteine (C) can form disulfide (S-S) bonds with other cysteines. Disulfide bonds are stronger than most other residue-residue bonds and can be broken by a reducing agent. These bonds were shown to modulate protein-protein interactions in certain proteins [44] and can induce IDP chain polymerization [29].

Proline (P) and Glycine (G) are amino acids often appearing in IDPs and IDRs. Glycine is characterized by its high degree of flexibility, favoring disorder. In contrast, Proline is much stiffer, therefore disturbing the ability of the protein to form a secondary structure [45, 46]. Many IDPs and IDRs are also characterized by a lower sequence complexity compared to folded protein [47, 48]. Additionally, as mentioned, their sequence typically comprises many repeats of charged residues, with only a few hydrophobic ones (Fig. 1.6). However, while many such correlations between the sequence and disorder can be found, their exact correlation is still largely unknown [13].

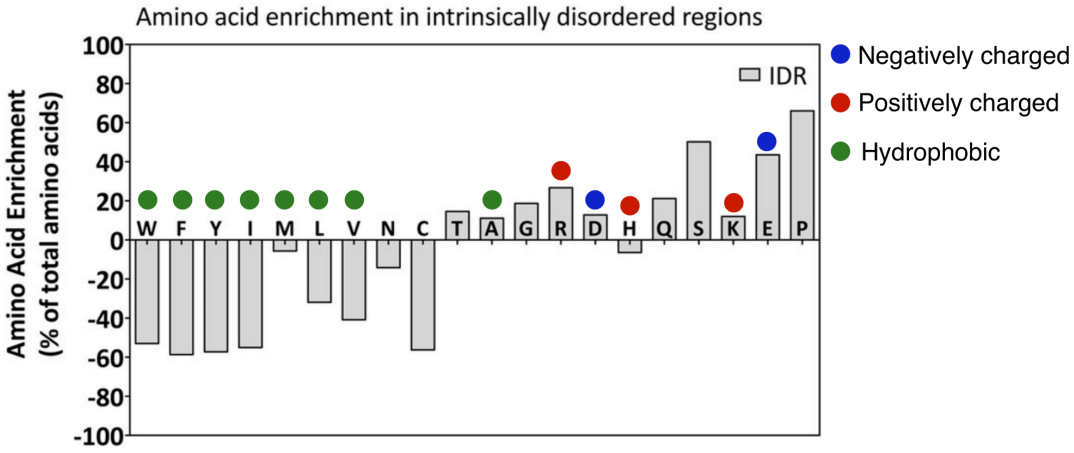

Figure 1.6: Amino acid enrichment in IDPs, adapted from Ref. [49]. Values were calculated by 100 - (% amino acids in IDR×100/% amino acids in total dataset).

1.3 Neurofilaments (NFs), and the NF-low disordered tail domain (NFLt)

The model IDP I will use in my thesis is derived from the mouse neurofilament low (NFL) protein. Below, I will briefly explain my motivation using it as a model system.

Neurofilaments (NF) are a family of neuron's specific cytoskeleton proteins. Unlike the other cytoskeleton proteins, NFs are rather static and provide the neurons their structural and mechanical support [50]. NFs were shown to aggregate into clumps together when neuroaxonal damage occurs in neurodegenerative diseases, such as Parkinson's and Alzheimer's' [4]. NF subunits consist of one of the three NF proteins: NF-low (NFL), NF-medium (NFM), and NF-high (NFH). All NF proteins possess both N- and C-termini IDR domains. The key molecular difference between the subunits is in the C-terminal IDR domains that significantly alter in length and charge distribution (Fig. 1.7). NF subunit proteins assemble into bottle-brush structures, in which their N-terminal domain makes up the folded backbone structure, while their C-terminal domain consists of a protruding IDR.

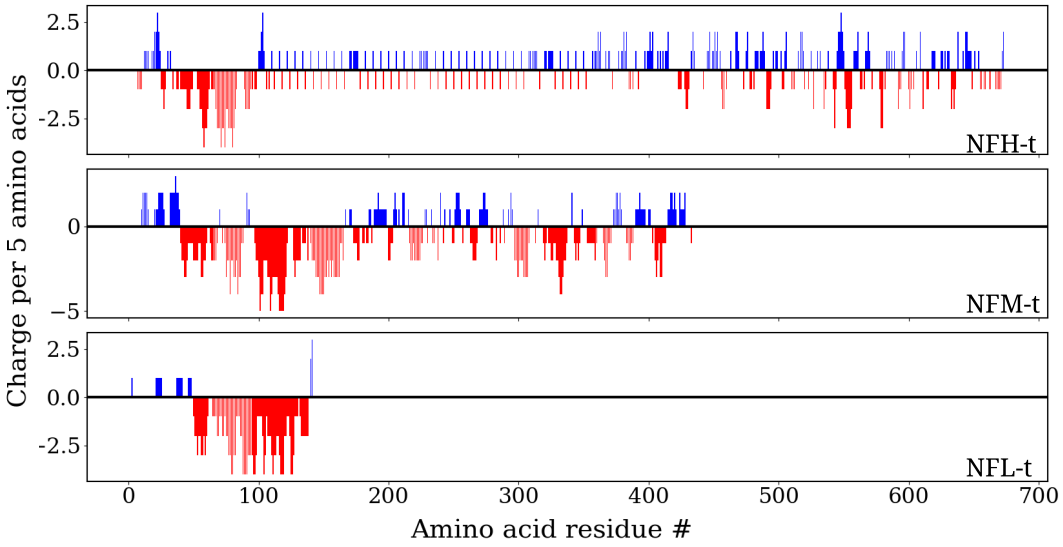

Figure 1.7: Charge distributions of neurofilament high, medium, and low from top to bottom respectively.

This thesis will focus on the NFL, specifically its disordered carboxy tail domain NFLt. The NFLts protrude from the backbone and regulate the inter-filament distance [51]. NFL is also key in regulating NF network formation under aggregation [52].

NFLt comprises 146 residues and can be roughly divided into two distinct sub-domains pertaining to their charge distribution and hydropathy (Fig. 1.8A). Starting from the N-terminal,

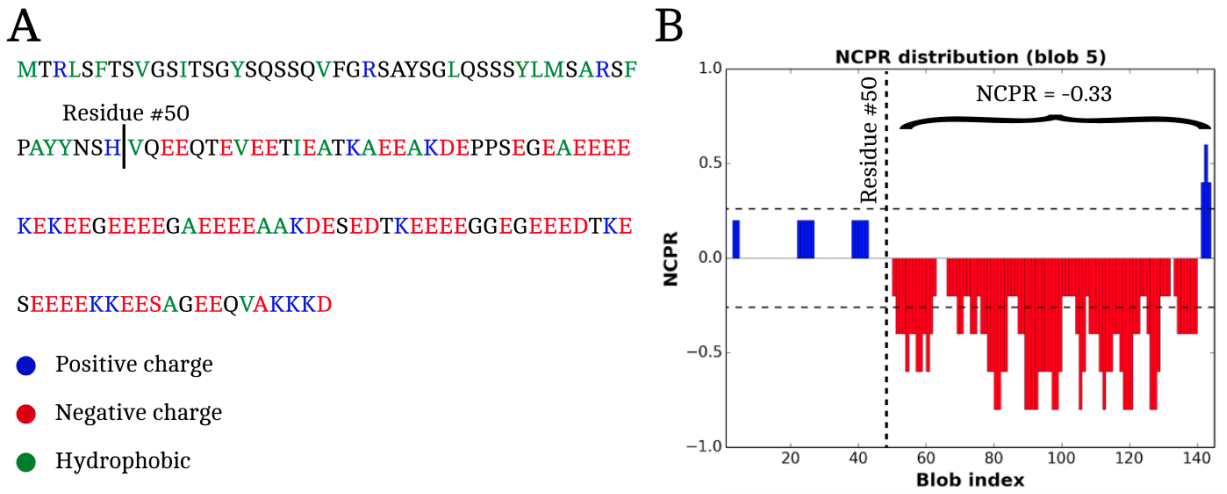

Figure 1.8: **A.** Sequence composition of the NFLt protein. Before residue # 50, the chain is mostly neutral, with 15 scattered hydrophobic sites. From residue # 50 onwards, the chain is highly negatively charged, featuring some scattered hydrophobic and positive sites. **B.** Net charge per residue (NCPR) of the NFLt sequence with a sample size of 5 residues per blob. The neutral nature of the first 50 residues contributes to a neglegible NCPR. From residue 50 onwards, the NCPR per blob is almost entirely negative, with a positive NCPR on its tail end due to the positive *KKK* residues. Graph provided by CIDER [53].

the first $\sim$ 50 residues of the NFLt are largely uncharged, with the first $\sim$ 10 residues from the N-terminal tip being prominently hydrophobic. The second domain (residues $\sim 51-146$) is largely negatively charged with a net charge per residue (NCPR) of -0.33 (Fig. 1.8B). Overall, the NFLt is described as a polyelectrolyte (negatively charged chain) with an NCPR of -0.24.

Important to this study, previous studies showed that NFLts exhibit unexpected glassy dynamics under force-induced mechanic measurements [29, 30]. I will revisit these experimental realizations in Sec. 6.5.3. Last, as I will show in section 6.3, NFLts aggregate from their hydrophobic N-terminal in isolation and form star-like brush 'structures' that still pertain to their disordered nature.

2 Polymer Physics

In their most basic form, polymers are chains of repeating monomers. For homopolymers, all subunits are equal. Polymer physics theories are prime candidates for relating the structural statistics of IDPs to well-researched and tested theoretical models for homopolymers. However, homopolymers lack the required sequence complexity to render them fully equivalent in representing IDPs. However, understanding the ensemble statistics of homopolymers can give us an estimation of how IDPs with similar characteristics will conform on average. Below, I summarize some of the fundamental theoretical building blocks in polymer physics needed for our investigations.

2.1 Ideal Chains

I start with the simple case of ideal chains. In ideal chains, any long-term interactions between the monomers are neglected, and we are left with a model that serves as the basis for the statistical analysis of polymer conformation. The following text summarizes the key results described in detail in [54].

We consider a chain of N bonded monomers with a constant bond distance l, resulting in a maximum chain length of $R_{\max} = Nl$. When determining the ensemble average end-to-end distance (R_{ee}), it is clear that the $R_{ee} = R_{\max}$ is just one (unlikely) outcome. Thus, it is easy to model the chain as a random walk of N sites, each separated by distance l. In this random walk, each site i has an equal chance to move in any direction within the given constraints. For instance, in a one-dimensional plane, site i could move from $i-1$ by $+l$ or $-l$ in an equal probability.

In three-dimensional space, the displacement vector between the $i-1$th and ith site is $\vec{r}_i$. Following, the total end-to-end length of the chain is then given by (Fig. 2.2):

$$\vec{R}_{\text{ee}} = \sum_{i=1}^{N} \vec{r}_i. \tag{2.1}$$

As the polymers are an ensemble of many different conformations, we can find their ensemble average of $\vec{R}_{\text{ee}}$, which is evaluated by its mean-square value:

$$\langle R_{\text{ee}}^2 \rangle = \sum_{i=1,\, j=1}^{N,\, N} \vec{r}_i \cdot \vec{r}_j. \tag{2.2}$$

As the bond length is equal across the entire chain, $\vec{r}_i \cdot \vec{r}_j = l^2 \cos(\theta_{i\,j})$, the $\langle R_{\text{ee}}^2 \rangle$ can be approximated as:

$$\langle R_{\text{ee}}^2 \rangle \simeq CNl^2, \tag{2.3}$$

where in the limit of $N \gg 1$:

$$C = \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} \langle \cos(\theta_{i\,j}) \rangle \right). \tag{2.4}$$

From here, we achieve the expression for the ensemble average of the end-to-end length, now simply denoted as R_{ee}:

$$R_{\mathrm{ee}} = C^{1/2} l\, N^{1/2} \tag{2.5}$$

Now, we can define an equivalent chain, for which $R_{\mathrm{ee}} = bN^{1/2}$, where b is known as the Khun length, the effective bond length of the chain. We are now left with finding the constant C, from which b can be derived. C is dependent on which model is chosen for the ideal chain, some of which are detailed below:

Freely joined chain. This model assumes a random conformation for each monomer (Fig. 2.1A), as such $\langle cos(\theta_{i\,j}) \rangle_{i \neq j} = 0$ and $\langle cos(\theta_{i\,j}) \rangle_{i=j} = 1$. From here, C in Eq. 2.4 is calculated as 1.

Freely rotating chain. In this model, the bond angle is set constant on one axis but allowed to rotate freely on the other (Fig. 2.1B). In this model, $C = (1 + \cos(\theta))/(1 - \cos(\theta))$, where θ is the chosen constant bond angle.

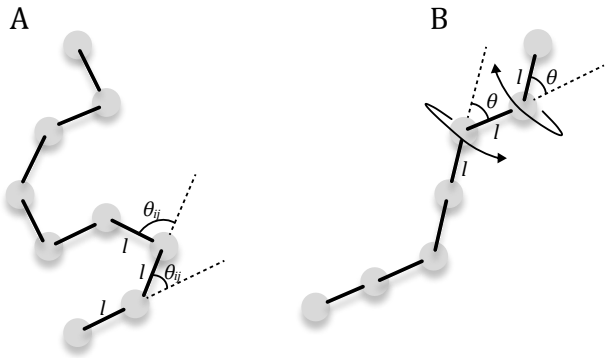

Figure 2.1: A. Freely joined chain model. B. Freely rotating chain model.

In addition to R_{ee}, another useful metric to describing the polymer ensemble is the radius of gyration, R_{G}, defined as the ensemble distance of the polymer from its center of mass (Fig. 2.2):

$$R_{\mathrm{G}}^2 = \frac{1}{N} \sum_{i}^{N} (\vec{r}_i - \vec{r}_{\mathrm{cm}})^2, \tag{2.6}$$

where

$$\vec{r}_{\mathrm{cm}} = \frac{1}{N}\sum_{i}^{N}(\vec{r}_i - \vec{r}_j). \tag{2.7}$$

Combining Eqs. 2.6 and 2.7, we can arrive to the following expression for R_{G}^2:

$$R_{\mathrm{G}}^2 = \frac{1}{N^2}\sum_{i}^{N}(\vec{r}_i - \vec{r}_j)^2. \tag{2.8}$$

We again aim to find the ensemble average, for which:

$$\langle R_{\mathrm{G}}^2 \rangle = \frac{1}{N^2}\sum_{i}^{N}\langle(\vec{r}_i - \vec{r}_j)^2\rangle. \tag{2.9}$$

In the limit of a long chain, the summations can be approximated as integrals, for which $\sqrt{\langle R_{\mathrm{G}}^2\rangle}$, now simply denoted as R_{G} is found:

$$R_{\mathrm{G}} = \frac{1}{\sqrt{6}} b N^{1/2} = \frac{1}{\sqrt{6}} R_{\mathrm{ee}}. \tag{2.10}$$

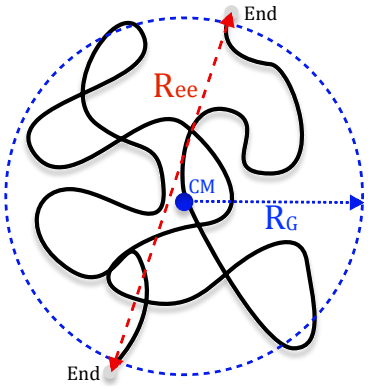

Figure 2.2: Visual schematic of the end-to-end distance R_{ee} (the distance between the two ends of the polymer) and of the radius of gyration R_{G} (the mean distance of the monomers from the center of mass (CM)).

2.2 Real Chains

Real chains take into effect the steric repulsion between the monomers, previously ignored in ideal chains. The steric effects can act as an energy boundary by a hard-core repulsion, which disallows the monomers to come close to each other within a certain distance [54]. In the discussed case of simple homopolymers, the 'repulsion volume' of the monomers, from which the monomers cannot overlap, is dictated primarily by the condition of the environment of the polymer - the solvent.

Given the solvent condition, commonly referred to as the solvent quality, the net interaction between any two monomers is summarized by the excluded volume, which is derived from the potential energy of the monomers given the solvent. Conceptually, the 'excluded volume' is positive whenever the monomers 'repel' each other and negative when they 'attract' each other. Additionally, a theoretical θ condition exists in which the attractive potential of the monomers cancels the hard-core repulsion, which results in an excluded volume of zero and near-ideal conformations.

In real chain theories, three major 'regimes' of solvent quality are defined, which dictate the obtainable range of excluded volumes. In *good solvents*, monomer-monomer attraction is preferable over monomer-solution attraction, which results in net repulsion with a positive excluded volume. In *poor solvents*, The monomer-monomer attraction is much more substantial than interactions with the solvent, resulting in a net attraction and a negative excluded volume. Lastly, in the θ solvent, the interactions are net-neutral, and the excluded volume is approximately zero.

Ideal chains assume that the polymers behave as 'random walks' - each monomer is connected to its neighbors in a random orientation constrained by predefined conditions. The same model can not be applied to real chains due to the steric interactions. Instead, a commonly applied model is the 'self-avoiding walk' (SAW), which is defined by the inability of the walk steps to coexist on the same lattice nodes. As such, SAWs are much more complex than random walks and typically require simulations to solve fully [54]. However, an effective mean-field approximation can reduce this complicated problem into a relatively simple model of energy reduction [55]. Minimizing the total free energy of the polymer in the case of a good solvent, the Flory radius R_{F}; i.e., the end-to-end distance, is found by a simple scaling law of $R_{\mathrm{F}} \propto N^{3/5}$, which is typically generalized as [54]:

$$R_{\mathrm{ee}} = aN^{\nu}, \tag{2.11}$$

where a is a parameter that depends on the excluded volume and the Khun length of the polymer, and ν is referred to as the Flory exponent or the scaling exponent. Using the same terminology, we can see that in the case of an ideal chain, $\nu = 1/2$, in contrast to $\nu = 3/5$, as obtained for a real chain in a good solvent. Using the same mean-field approximations, it can be shown that for a poor solvent, the same scaling law can be found, for which $R_{\mathrm{ee}} \propto N^{1/3}$; $\nu = 1/3$ [54].

The final parameter to be found is a, which is related to the actual size of the monomers, the polymer model used, and their excluded volume. In theta solvents, the parameter a is as

was previously described for ideal chains. Likewise, with real chains in poor and good solvents, the value of a depends on numerous parameters. For example, experimental measurements and simulations have measured the value of a for intrinsically disordered proteins within the SAW model to be $0.55nm$ [34, 56].

To complete our analogy to ideal chains, we aim to find the radius of gyration R_{G} for real chains and its relation to R_{ee}. The ratio between R_{G} and R_{ee} is approximately given by [56]:

$$\lambda = \frac{R_{\mathrm{ee}}^2}{R_{\mathrm{G}}^2} = \frac{2(\gamma + 2\nu)(\gamma + 2\nu + 1)}{\gamma(\gamma + 1)}. \tag{2.12}$$

Thus:

$$R_{\mathrm{G}} = \frac{1}{\sqrt{\lambda}} R_{\mathrm{ee}} = \sqrt{\frac{\gamma(\gamma + 1)}{2(\gamma + 2\nu)(\gamma + 2\nu + 1)}} N^{\nu}, \tag{2.13}$$

where γ is a constant given by 1.1615 ± 0.0011 [57].

In conclusion, the generalized description of chains can describe the approximate ensemble conformation of real and ideal polymers in any solvent, using only a few parameters: the monomer number N, the scaling exponent ν, and the prefactor a. Being flexible polymers, such a description could, in theory, describe the complex conformations of IDPs with only a few parameters. In particular, ν has the potential to be able to distinguish between various folding states of different proteins due to its unitless nature. For example, this usage of ν was shown to distinguish proteins undergoing unfolding in denaturation by a transition of ν from $1/3$ in the folded state to $1/2$ in the theta state and finally to $3/5$ in the random coil state [27] (Fig. 2.3). In contrast to the folded proteins, the measured IDPs maintained a ν higher than $3/5$ under all conditions.

2.3 Polyampholytes

Until this point, the presented models were centered around charge-neutral homopolymers that, at most, exerted intramolecular steric repulsion. Now, I focus on a far more complex case of polymers comprised of charged molecules, polyampholytes. As discussed in section 1.2 for proteins, charged monomers can construct a highly complex interaction network that couples every charged monomer in the system via electrostatic interactions. This coupling makes the polyampholyte problem extremely difficult to solve [58]. Regardless, efforts can be made to simplify this problem to obtain an approximation of how charge-rich polymers' regions may expand or contract.

The below discussion on the mean-field results summarises the text in Ref. [59]. Using a mean-field approximation, the free energy of the polyampholyte chain is comprised of two components:

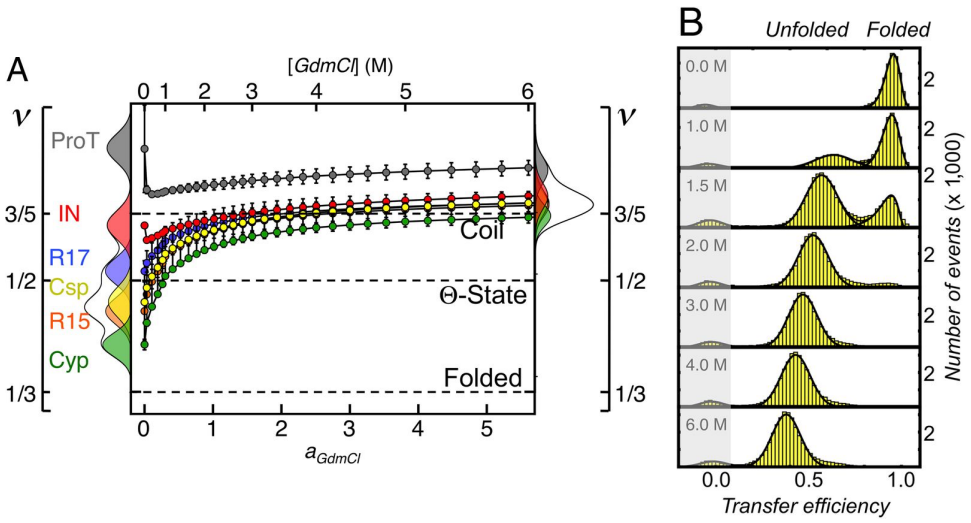

Figure 2.3: **A.** Scaling exponents ν in increasing denaturation (GdmCl concentration) for multiple proteins undergoing unfolding and an IDP (in gray). Figure adopted from Fig. 3A of [27]. **B.** Protein unfolding histogram in increasing denaturation (GdmCl concentration). Figure adapted from Fig. 1B of [27].

the entropic component and the interactions component. The entropic component favors chain conformations near the Gaussian chain size and increases upon deviation [60]. The interaction component includes short-range and electrostatic interaction, which can be approximated by the Debye Hückel approximation [61]. In generalized terms, the distinction between polyampholytes and polyelectrolytes is made by the relative amount of negative and positive charges within the chain.

The polyampholyte chain is assumed to be random, *i.e.*, the negative and positive charge sites are randomly distributed across the chain. Here, a chain of N monomers contains N_+ positively charged monomers and N_- negatively charged monomers. Two new parameters are introduced to distinguish between the polyampholyte regimes: the variance of charge asymmetry σ and the reduced temperature t:

$$\sigma = \frac{(f_+ - f_-)^2}{f_+ + f_-}, \tag{2.14}$$

$$t = \frac{1}{(f_+ + f_-) l_B / b}. \tag{2.15}$$

Here, $f_\pm = N_\pm / N$, l_B is the Bjerrum length, and b is the Kuhn length. l_B is defined as the inter-monomer distance at which the interaction energy between two molecules is comparable

to thermal energy; *i.e.*, $l_B = e^2/(4\pi\epsilon_0\epsilon k_B T)$. Here, ϵ is the dielectric constant, and ϵ_0 is the permittivity of vacuum.

In addition, the Debye screening length $\lambda_D = \sqrt{\epsilon\epsilon_0 k_B T/2N_A e^2 I}$ is defined as the persistence length of the electrostatic interactions, where I is the ionization strength. Below, I will discuss the results of the mean-field approximation for three distinct polyampholyte regimes.

Regime I: Neutral chains. The simplest case of a polyampholyte is when the electrostatic contribution is less than the thermal energy. Here, the chain is described as in sections 2.1 and 2.2) with end-to-end radius (R_{ee}):

$$R_{\text{ee}} \sim bN^{\nu}, \tag{2.16}$$

where $\nu = 1/2$ for an ideal chain and 3/5 for a real chain. This regime applies when the electrostatic effects are weaker than the entropy contribution to the free energy. Thus, $t > t_1$, where $t_1 = \sigma N^{3/2}$ and $\sigma N > 1$.

Regime II: Polyelectrolytes. The second possible regime is that of a polyelectrolyte. The electrostatic effects are considered significant in the polyelectrolyte regime, and the effective charge is prominently negative or positive. As such, $t < t_1$ and $\sigma N > 1$ are the two conditions of this regime. Starting with the simple case of ideal polyelectrolyte chains, free energy minimization results in chain length (L) (the statistical length of the chain resulting from electrostatic interactions, not to be confused with its contour length $L_c = bN$) of:

$$L \sim bN(\sigma/t)^{1/3}. \tag{2.17}$$

A scaling argument can be made where the chain can be divided into N/g electrostatic blobs of size D which contain g monomers. D is sufficiently small so that below D entropy effects are stronger than the electrostatic interactions, and the chain behaves as a neutral chain (Fig. 2.4). From balancing the repulsion energy with the thermal energy, $g \sim (t/\sigma)^{2/3}$ and:

$$D \sim b(t/\sigma)^{1/3}. \tag{2.18}$$

In this case, the electrostatic repulsion between the blobs is strong enough so the chain is stretched into a rod of blobs:

$$L \sim D(N/g). \tag{2.19}$$

For real chains, additional terms for the free energy need to be accounted for due to the two-body repulsion. Here, chain length L is given by:

$$L \sim bN(\sigma/t)^{2/7}, \tag{2.20}$$

and blob size D by

$$D \sim b(t/\sigma)^{3/7}. \tag{2.21}$$

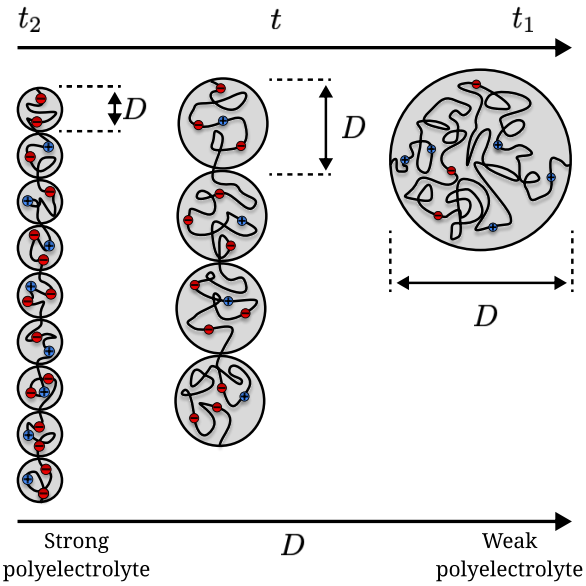

Figure 2.4: Schematic of polyelectrolytes with changing reduced temperature t. The polyelectrolyte regime is characterized as a stretched chain of electrostatic blobs of size D. Within the blobs, the chain segments behave akin to neutral chains. At the lowest limit $t = t_2$, D is minimized, and the chain length is maximized. At the highest limit $t = t_1$, D equals the chain length at its minimal value.

These repulsive interactions are prominent as long as the Debye screening length λ_D is larger than the electrostatic blob size D. This condition is applied when $t > t_2$, where $t_2 = \sigma^{1/2}$. At the limit of $t = t_2$, the electrostatic blob size reaches its minimum value and L reaches its maximum value of:(Fig. 2.4):

$$D \sim b(t/\sigma)^{1/3} \sim b\sigma^{-1/6} \text{ ; ideal chain,} \tag{2.22}$$

$$D \sim b(t/\sigma)^{3/7} \sim b\sigma^{-3/14} \text{ ; real chain,} \tag{2.23}$$

$$L \sim bN\sigma^{1/2} \text{ ; ideal chain,} \tag{2.24}$$

$$L \sim bN\sigma^{2/5} \text{ ; real chain.} \tag{2.25}$$

In the strong polyelectrolyte limit, all charges are of one type. Here, $\sigma = 1$, and D is of equivalent scale to the Kuhn length. Consequently, only one monomer is found within each electrostatic blob, and L then results in $L = bN$, *i.e.*, the length of a rod polymer.

The opposite limit of the regime is the weak polyelectrolyte limit, where $t = t_1 = \sigma N^{3/2}$.

Here, D and L converge to:

$$D \sim L \sim bN^{1/2} \text{ ; ideal chain,} \tag{2.26}$$

$$D \sim L \sim bN^{3/5} \text{ ; real chain.} \tag{2.27}$$

In this limit, D becomes sufficiently large to cover the entire chain, thus returning to the neutral chain regime (Fig. 2.4).

Regime III: Polyampholytes. The third regime is the polyampholyte regime, for which the reduced temperature condition is $t < t_2$. Here, the chain collapses from the monomer attractions and is described as a dense collection of debye blobs of radius λ_D (Fig. 2.5). In this dense state, the chain is described by two parameters: its length (L) and its width (D). Additionally, due to its new geometry, the entropy of the chain is now influenced by its surface tension. For ideal chains, balancing the electrostatic repulsion of the Debye blobs with the surface tension results in a constant number of monomers $g \sim 1/\sigma$ inside volume D^3. From here, using equations 2.19 and 2.18:

$$L \sim D(N/g) \sim Nb\sigma^{2/3}t^{1/3}. \tag{2.28}$$

For real chains, the following results can be found:

$$D \sim bt^{2/3}\sigma^{-1/3}, \tag{2.29}$$

and:

$$L \sim DN/g \sim Nb(t\sigma)^{2/3}. \tag{2.30}$$

Lastly, polyampholytes with a symmetric charge distribution ($N\sigma < 1$) begin to collapse into a globule of:

$$L \sim D \sim t^{1/3}N^{1/3}. \tag{2.31}$$

In real synthetic or natural polyampholytes, charge distribution may fluctuate between regions, leading to multiple regimes within the same chain. In sufficiently long chains, this behavior could lead to complex structures such as necklacing, where condensed polyampholyte regimes are connected by polyelectrolyte or neutral coiled regimes (Fig. 2.7A).

Using mean-field approximations, it was shown that regions with lower variance of asymmetry ($\sigma \to 0$) result in more compact chains with lower chain length (L) and width (D). Similar effects were observed in experiments comparing model synthetic polyampholytes with different charge patterning, which showed that more symmetric chains directly resulted in a higher radius of gyration (R_G) than asymmetric chains of the same length [33] (Fig. 2.7B).

As discussed in section 1.2, charged residues are abundant in IDP sequences, with most IDPs classified as polyampholytes [33]. In many cases, the charge interactions resulting from their charges are intertwined with their structure and function [37, 38, 62, 63].

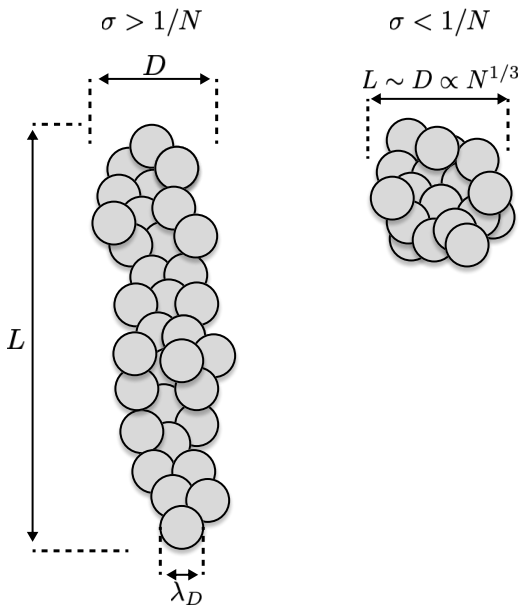

Figure 2.5: Schematic of polyampholytes with changing charge asymmetry (σ). At $\sigma > 1/N$, the polyampholyte regime is characterized as a condensate of electrostatic blobs of size λ_D, where λ_D is the Debye screening length. When $\sigma < 1/N$, the chain reduces into a near-spherical globule of scaling $L \propto N^{1/3}$.

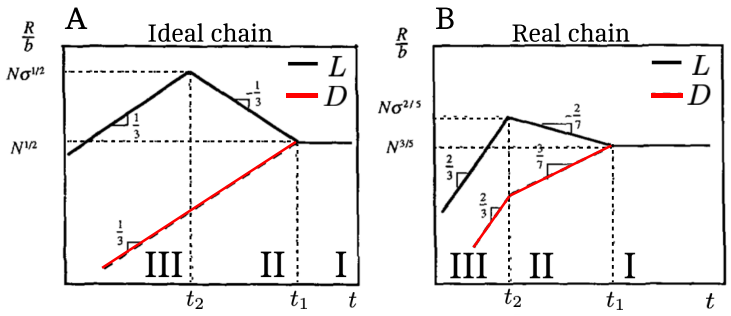

Figure 2.6: Chain length L and width D dependency on reduced temperature t. In both ideal and real chains, chain length L is maximized at $t = t_2$, where the transition between the polyelectrolyte and polyampholyte regime occurs. Figure adapted from [59].

An example of one such IDP is the Neuroscopa clock protein FREQUENCY (FRQ) [63] (Fig. 2.7C). At the start of the clock cycle, FRQ is in its closed state, stabilized by hydrophobic bonds.

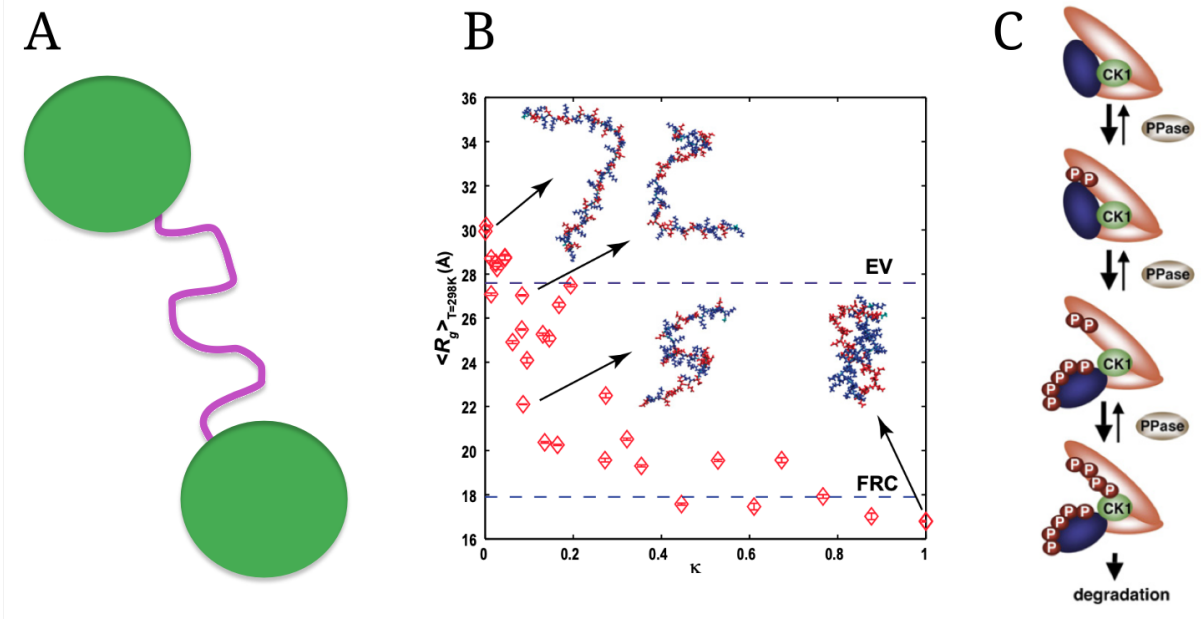

Figure 2.7: **A.** Schematic of polyampholyte necklacing, with green regions representing condensed polyampholytes and purple representing the connecting coils. **B.** Radius of gyration (R_{G}) against the charge patterning parameter (κ) for the synthetic polyampholyte $(\mathrm{Glu\text{-}Lys})_{25}$ variants of different κ. Here, lower κ, which correlates to higher charge asymmetry, is shown to result in higher R_{G} values. Figure was adopted from [33], where more information on the polyampholyte variants can be found. **C.** Neurosopa circadian clock IDP phosphorylation process schematic. The phosphorylation process causes negative charges to accumulate on the protein (P spheres) until it eventually opens. When the process is finished, a degradation process resets the cycle, and the protein closes. Figure adapted from [63].

During the day, FRQ residues accumulate a negative charge by phosphorylation. Eventually, this accumulation triggers a transition of the protein into an open state by the net-charge repulsion of the negative charges. When FRQ opens, it exposes a signal that dephosphorylates the charged residues, reverting the protein to its closed state and resetting the cycle.

Salt screening. The following discussion focuses on the case of a polyelectrolyte. Similarly to the models of the salt-free discussion, the screened polyelectrolyte model is also based on a mean-field approximation [64]. In the model, the effect of the intermolecular electrostatic repulsion is the expansion of the extended volume v to v', where [65]:

$$v' = v + \frac{4\pi\alpha^2 z_{\mathrm{p}}^2 l_{\mathrm{B}}}{\kappa^2 l^3}, \tag{2.32}$$

where $\kappa = \lambda_D^{-1}$ and thus relates to the ionization strength by $\kappa \sim \sqrt{I}$. For $R_G = R_0 R_{ee} = R_0\sqrt{N/6}\,b$, the prefactor R_0 is given by [64]:

$$R_0^5 - R_0^3 = \frac{134}{105}(\frac{3}{2\pi})^{3/2} v' \sqrt{N}. \tag{2.33}$$

On the other hand, for the high salt limit, real chain scaling is regained:

$$R_G \sim l v'^{1/5} N^{3/5}. \tag{2.34}$$

While seemingly rudimentary, as this model ignores many underlying complexities of the short- and long-range electrostatic interactions, it was shown to align with numerous polyelectrolyte studies and even with some charged IDP measurements [37, 66]. For example, Liu et al. [66] showed how the effects of salt on the Sic1 IDP correlate with the salted polyelectrolyte model up until a critical salinity, where the experimental data deviates from the model (Fig 2.8). This deviation at high salinity was deduced to be due to specific interactions between the ions, proteins, and water molecules, which are unaccounted for in the Flory screening model.

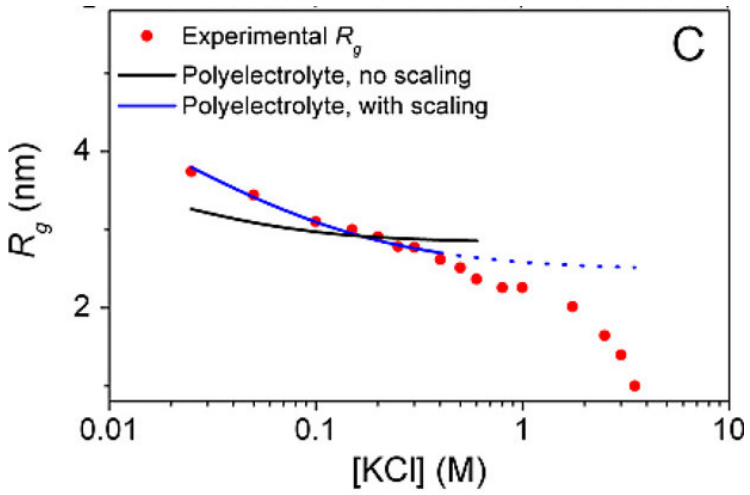

Figure 2.8: **A.** Radius of gyration R_G against increasing electrostatic screening (KCL concentration) for the SIC1 IDP. The blue line indicates an addition of a scaling parameter to equation 2.34. Figure adapted from Fig. 8 of Ref. [66].

3 Polymer Brush Physics

Polymer brushes are a subset of polymer physics dealing with polymers sprouting from surfaces. Such modified surfaces can create a repulsive or attractive potential from the surface they are attached to, which can be tuned to resist or attract molecules under specific conditions. As such, recent years have shown a rise in many theoretical and experimental investigations to explore their practical potential as surface modifiers. A common method for creating polymer brushes is the 'graft to' method, in which already polymerized molecules are grafted onto a surface (Fig. 3.1A). For example, recent studies have successfully formed high-density polymer brushes by grafting Polyethylene-Glycol (PEG) polymers with reactive end-groups in Na_2SO_4 [67].

Polymer brushes can also be made by the 'graft from' method, in which the brushes form from specific initiation sites found on the surface, from which polymerization occurs [68] (Fig. 3.1B). Along with surfaces, polymer brushes can also be created employing self-assembly, in which the polymers are designed to assemble into brush-like structures (Fig. 3.1C). For example, block copolymers designed with a hydrophobic end can self-assemble into brushes with a spherical surface [69, 70, 71].

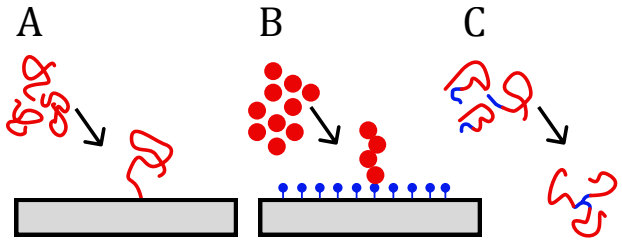

Figure 3.1: Schematics of three different methods for creating synthetic polymer brushes. **A.** 'Graft to' polymer grafting method. This method involves grafting pre-assembled polymers onto the substrate.**B.** 'Graft from' polymer grafting method. In this method, individual monomers are assembled into polymers from the substrate.**C.** Block copolymer self-assembly into polymer brushes. Hydrophobic blocks (blue) of multiple block copolymers self-assemble into spherical brushes with protruding hydrophilic (red) chains.

Equivalent to synthetic brushes, some proteins assemble in nature to form brushes. This formation can occur by several processes, such as self-assembling protein complexes with protruding IDR terminals or IDP binding into other protein structures (Fig. 3.2). An example of self-assembled protein brushes is the neurofilament proteins discussed in section 1.3. Neurofilaments assemble into bottle-brush formations with protruding C-terminal IDRs, which mediate the inter-filament distance between other neurofilament brushes [50, 52].

Binding-formed protein brushes can be exemplified by the Tau IDPs, which bind onto axonal microtubules and act as brushes that stabilize the internal skeleton of the nerve cells [21, 72]. In

neurodegenerative diseases, the Tau proteins were shown to disassociate from the microtubules and self-assemble into fibrilar aggregates with protruding disordered brushes [73].

Another example of binding-formed protein brushes is the possibility of creating substrate-bound IDP brushes, as with synthetic polymer brushes. These IDP-based polymer brushes hold many advantages over traditional synthetic polymer ones, such as designable sensitivity to external stimuli [62].

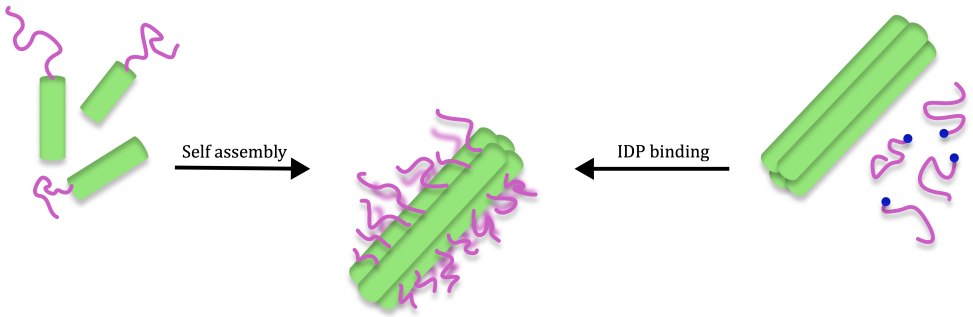

Figure 3.2: Illustration of IDP protein bottle brush formation by means of self assembly or IDP binding.

Notably, the brushes' polymeric structures differ from that of isolated polymers, thus requiring a different theoretical approach. Various models, such as ones developed by Alexander-De-Gennes, have shown success in correctly modeling experimentally observed behaviors of polymer brushes [68, 74, 75]. Here, we will discuss some of these models and how they are used in the context of real polymer brushes.

3.1 Neutral Brushes

Neutral brushes assume the tethered polymers to be uncharged homogeneous polymers of equal chain length (N) and Kuhn length (b). The simplest case within neutral brushes is polymer 'mushrooms' in a good solvent. The 'mushroom' regime of the polymer brush is applicable when the height of the brushes (L) is much smaller than the distance between the brushes (D) (Fig. 3.3).
In this case, the tethered polymers occupy the volume of around half of a sphere of Flory radius [76]:

$$R_F = bN^{3/5} \tag{3.1}$$

The other case, where $D < L$, is the overlapping brushes, or 'stretched chains' regime [77]. In this regime, the chain is subdivided into blobs of g_D monomers, such that N/g_D blobs occupy

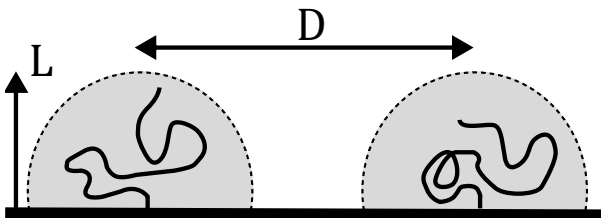

Figure 3.3: Illustration of the polymer brush 'mushroom' regime.

the chain (Fig. 3.4).

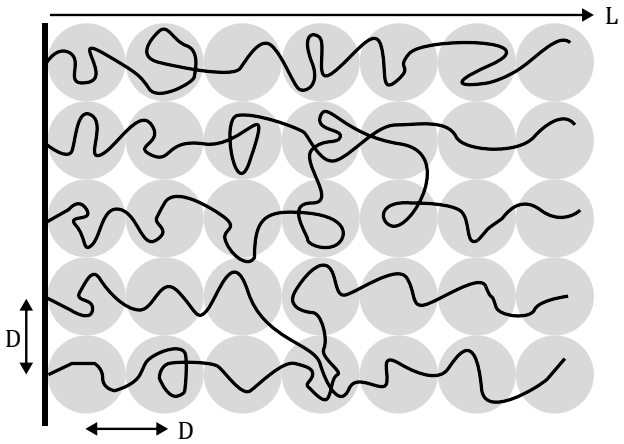

Figure 3.4: Illustration of the polymer brush stretched regime.

At small scales, the chain dimensions are dominated by excluded volume effects. As such, we expect the correlation between the blob size (g_D) and the distance between adjacent blobs (D) to follow similar scaling as in Eq. 3.1 where

$$D = b\, g_D^{3/5} \tag{3.2}$$

Thus, in the overlapping regime, each blob is considered a much smaller than the length of the chain. As such, in a blob of volume D^3, we expect the volume to be densely occupied, so that:

$$g_D / D^3 \simeq \phi_i / a^3, \tag{3.3}$$

where ϕ_i is the volume fraction of monomer i in the chain. The brush height (L) can now be derived, by considering the total volume LD^2 of a grafted chain of N monomers:

$$\phi_i / N = b^3 / LD^2. \tag{3.4}$$

Combining Eqs. 3.3 and 3.4 they've shown:

$$g_D / D^3 \simeq N / LD^2, \tag{3.5}$$

from which L was estimated as:

$$L \simeq (N/g_D)\, D \tag{3.6}$$

From this equation, we see that L and D show a linear correlation. Using Eq. 3.4 and the relation $D = b\sqrt{\sigma}$, where σ is the grafting density we find the relation:

$$L \simeq bN\sigma^{1/3} \tag{3.7}$$

The stretching coils regime then describes each polymer as a chain of $N/\, g_D$ hard spheres with volume D^3, which is roughly equivalent to the excluded volume of the monomers. An example of the brush regime transition can be seen in a past experiment of poly(acrylamide) (PAAm) polymers on a substrate [74] (Fig. 3.5).

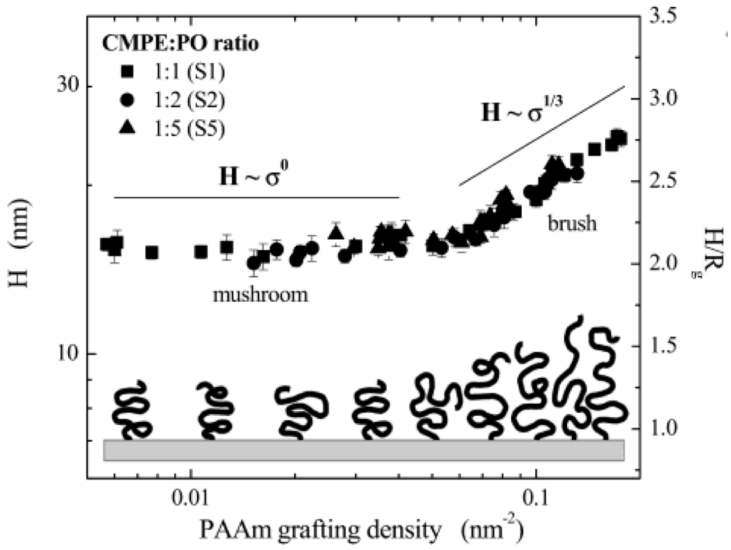

Figure 3.5: Mushroom to brush transition of PAAm polymers, adapted from Ref. [74].

3.2 Charged Brushes

Similar to the case of isolated polyampholytes, charged brushes are heavily impacted by the electrostatic interactions of the monomers and show significant response to electrostatic screening.

I again divided our discussion into salt-free and salt solutions, where I focus on the case where the polyampholyte is a polyelectrolye for bravity. Additionally, I assume the chain to be a 'strong' polyelectrolyte; the charge composition of the chain to be unchanging in ionization. Further reading on alternative cases can be found in Ref. [78]. **Salt-free solutions.** Similar to neutral brushes, multiple 'regimes' exist within charged polymer brushes, dictated by their grafting density. Here, I will discuss three such regimes furthered detailed in Ref. [79]. We start with the *charged mushroom regime*, which is applicable up to a grafting density $\sigma^* \sim L^{-2}$. In this regime,the brush height is given by:

$$L \simeq bN\,(\frac{l_B}{b}f)^{(1-\nu)/(2-\nu)}, \tag{3.8}$$

where ν depends on the solvent and f is the degree of ionization. However, due to the long-ranged electrostatic repulsion between the chains, it can be shown that above the critical grafting density $\sigma_{\text{orient}} \simeq (l_B(fN)^2\,L)^{-1}$, these effects are pronounced enough to cause side-ways orientation in the grafted chains, but not enough to affect their expansion [79].

Next is the *Pincus brush regime*. Here, the chains are close enough to each other for the inter-chain interactions to be strong enough to hold consequence over the brush dynamics. In this regime, the brush height follows:

$$H \simeq b(uf^2\sigma b^2)^{(1-\nu)/\nu}N^{(2-\nu)/\nu}. \tag{3.9}$$

Comparing the obtained brush height to the 'stretched brush' regime of the neutral brush, we notice that in good solvent ($\nu = 2/3$) the brush height scales with $\sim \sigma^{1/2}$. This regime is applicable as long as counterions are able to escape from the brush. From here, the Gouy Chapman length (λ_{GC}) is introduced, which is defined as the distance in which the attraction energy of the counterions is to the charged surface higher than that at the surface by K_BT. As such, the upper limit of the Pincus brush regime is defined as when the brush height surpasses the Gouy Chapman length. From here, the upper limit is found as $\sigma_{PB} \simeq (l_B bN^2 f^{2-\nu})^{-1}$.

The next regime, termed the *osmotic brush regime*, is applicable when $\sigma > \sigma_{PB}$. In this regime, the effects of the counter-ions dominate the properties of the brush. As such, the resulting brush height in the osmotic brush regime is independent of the grafting density, and is given by [79]:

$$H \simeq bNf^{1-\nu}. \tag{3.10}$$

Thus, in the osmotic brush regime, effectively all the counter-ions are 'trapped' inside the brush, and its structure is determined by the osmotic pressure exhorted by the counter-ions.

Salted solutions Until this point, I had discussed the behavior of polyampholyte brushes without consideration for charge screening. In the osmotic brush regime, when the osmotic pressure of the counterions dominates the brush behavior, the dependency of the brush height in grafting density is negligible. As solution ionization increases, this effect remains largely the same

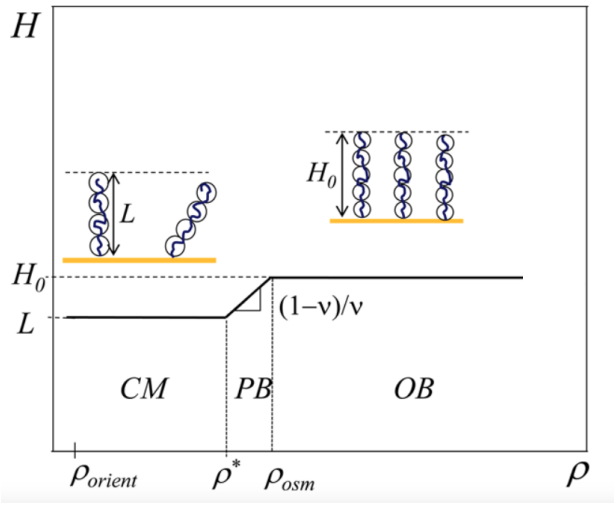

Figure 3.6: Illustration of dependence of brush height (H) on the grafting density (ρ) for the three polyelectrolyte brush regimes summarized in the text. Illustration was modified and adapted from [79].

until a critical ionization concentration (I_0), from which external ion concentrations overcome the brushes' free counterions [80]. In the presence of ionization, the osmotic pressure is derived as [80]:

$$\Pi \sim ck_BT(\kappa_0/\kappa)^2 \tag{3.11}$$

Where κ_0 is the inverse Debye length association with the counter-ions, and $\kappa^2 = \kappa_0 + \kappa_s^2$ describes the total inverse Debye length, which takes into account the external ions, described by κ_s. In the case where the ionization (I) is much larger than I_0, it follows that $\kappa_0 << \kappa_s$, and:

$$\Pi \sim c^2 k_BT/I \tag{3.12}$$

The dependency of the osmotic pressure on the square of polymer concentration (c) is of the same form as in neutral chains dominated by excluded volume repulsion [80]. As such, the 'salt brush' regime of the charged polyelectrolyte describes it as akin to a neutral brush with an enlarged excluded volume, which is proportional to $\sim I^{-1}$. From here, in the strong screening limit, the following relation between the brush height and ionization I is achieved:

$$H \sim bNI^{-1/3}\sigma^{-1/3} \tag{3.13}$$

Interestingly, we see a return in the dependency of the brush height in σ, which was shown to be lost in the osmotic brush regime.

4 Small Angle X-ray Scattering (SAXS)

4.1 Basics

The X-ray scattering technique utilizes the properties of electromagnetic waves to characterize nanostructures. The advantages of X-ray scattering include the ability to probe structures from sub-nm to 10s of nm length scales without supplement markers and its non-intrusive nature, which allows the sample to remain unaltered after the measurement. Important to protein studies, X-ray scattering techniques allow the samples to be measured in solution with an accessible way to control temperature, allowing researchers to observe their behavior under specific conditions.

In practice, x-ray setups comprise two basic components: A beam source and a detector.
Beam source. The beam source generates the X-rays as laboratory-scale apparatus or as part of a synchrotron facility. Laboratory sources mostly employ X-ray tubes, which radiate X-rays with applied voltage. These are typically small and offer limited tunability of the X-ray radiation. Synchrotron sources are generated in large facilities employing particle acceleration. Charged particles are accelerated across a large ring by a series of magnets to a speed close to the speed of light. The beam is then directed into multiple paths, which connect to different beamlines in which experiments can be conducted simultaneously (Fig. 4.1B). Further beam adjustments can be made within the beamlines, allowing the users to tune the radiation source to best fit their experimental needs. The tunability of the beam is a significant advantage over laboratory sources, which need to be configured by the manufacturer for specific applications.

Another critical advantage of synchrotron radiation is the obtainable resolution of the measurements, as the beams are with intensity larger than laboratory sources in orders of magnitude with greater focus. These properties allow for much more X-rays to scatter off the sample at any given time.

Inside the beamline, the X-rays go through a series of mirrors and slits that aim to control the beam size and its divergence (Fig. 4.1C). In addition, monochromators such as crystals enable the beamline to select a desired wavelength range for the experiment. Within laboratory setups, most of the optical alignments are done within the beam source by the manufacturer, whereas further alignments to the setup are done by motorized stages, which can direct the beam at different orientations.

Detector. The primary purpose of the detector is to collect the X-rays scattered off the sample. Two-dimensional detectors record images of the scattering positions throughout the measurement. On the other hand, one-dimensional detectors record the scattering intensity of the X-rays that come in contact with the detector. Typically, one-dimensional detectors would be used for system calibration, while two-dimensional detectors for measurements. The detector can be moved closer and further from the sample to collect different scattering angles' dynamic range. The distance between the sample and the detector allows the measurement to focus on

different length scales. Detectors in laboratories and synchrotrons operate on similar principles. However, in many cases, synchrotron detectors are much larger and designed to collect much more information at once.

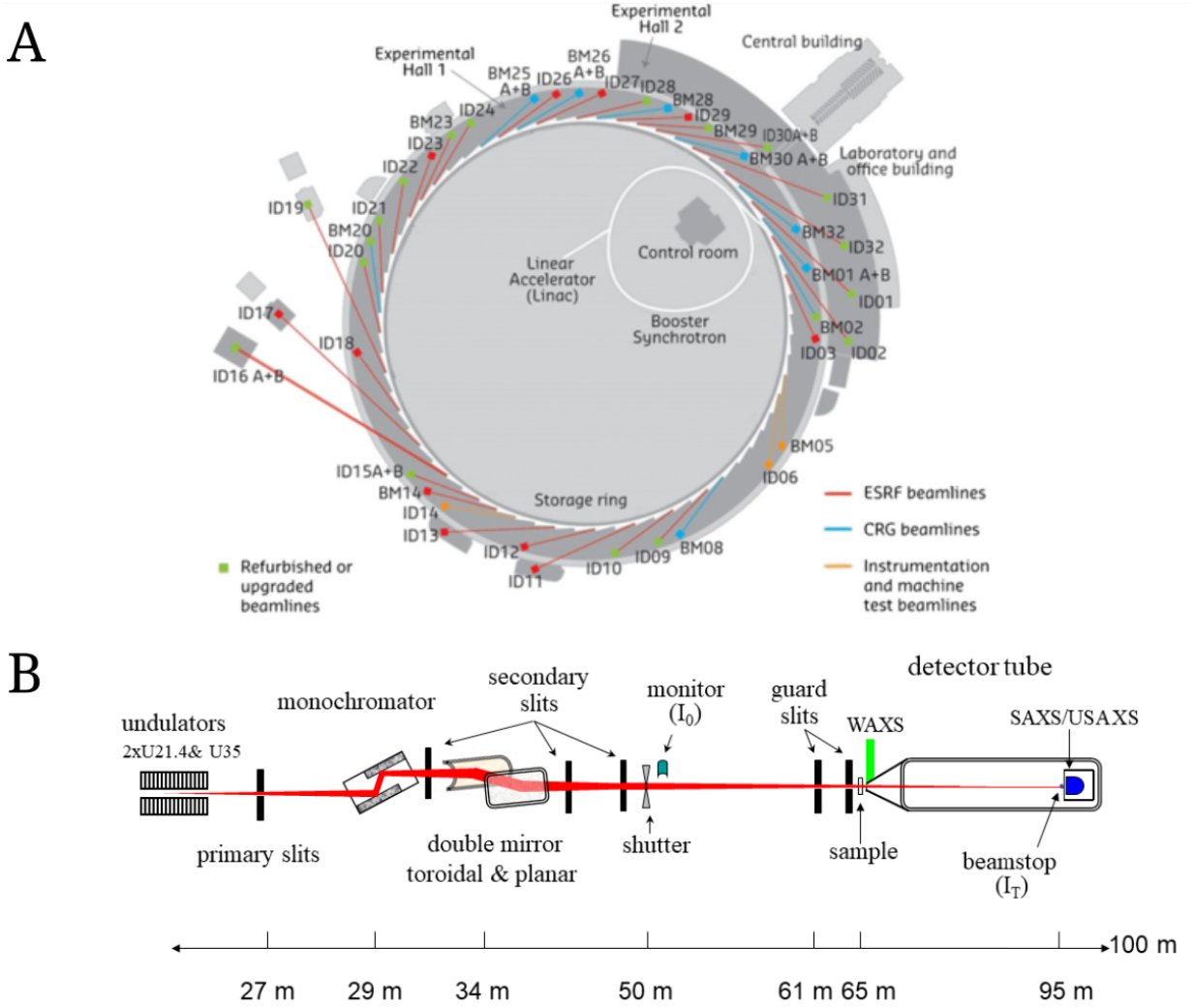

Figure 4.1: **A.** Schematic of the ESRF synchrotron facility. From the acceleration ring (storage ring), the X-rays are diverged into numerous beamlines (in lines). Image courtesy of the ESRF communication group. **B.** Schematic of the ID02 beamline of ESRF. The X-rays are first attuned to a specific wavelength through the monochromators. Following, the divergence and size of the X-ray beam is controlled by a series of slits and mirrors. Finally, the X-rays converge on the sample and are measured by the detector tube on the far end of the beamline. Figure adapted from https://www.esrf.fr/home/UsersAndScience/Experiments/CBS/ID02/BeamlineLayout.html.

In my work, I focus on a specific X-ray technique - Small Angle X-ray Scattering (SAXS). In SAXS, the detector is at a distance sufficiently large from the sample so that the scattering angles are very small [81]. Below, I will summarize the basics of SAXS, explored in more detail in the reference textbook [82].
As X-rays pass through the sample, they can scatter off the electron cloud surrounding the sample's atoms. Upon scattering, the propagation vector $\vec{k}$ of the x-rays scatters at an angle θ from its initial orientation, resulting in a final vector of $\vec{k}'$. We additionally assume the scattering to be elastic so that the magnitude of the propagation vectors does not change and is equal to $2\pi/\lambda$, where λ is the wavelength of the X-ray. From here, we can calculate the phase difference of the scattering to be:

$$\Delta\Phi = (\vec{k} - \vec{k}') \cdot \vec{r} = \vec{q} \cdot \vec{r} \tag{4.1}$$

where $\vec{q} = \vec{k} - \vec{k}'$ and is defined as the momentum transfer vector, or the scattering vector of scattering incident. As the scattering is elastic; $\mid \vec{q} \mid = 2 \mid \vec{k} \mid \sin(\theta) = 4\pi/\lambda \sin(\theta)$.

X-rays hitting the sample scatter off the electrons comprising the atoms. Considering an atom with electron cloud density $\rho(\vec{r})$, the scattering amplitude of the atom $f^0(\vec{q})$ becomes:

$$f^0(\vec{q}) = \int d\vec{r}\, \rho(r) e^{i\vec{q}\cdot\vec{r}}. \tag{4.2}$$

A molecule comprised of n atoms at distances $\vec{r}_0$, $\vec{r}_1$, $\vec{r}_2$, ..., $\vec{r}_n$ from each other is then given by:

$$f(\vec{q}) = \sum_j^n f_j^0(\vec{q})\, e^{i\vec{q}\cdot\vec{r}_j}. \tag{4.3}$$

As a sample typically comprises of a number of molecules, it then follows that its scattering amplitude is similarly described as:

$$F(\vec{q}) = \sum_j^n f_j(\vec{q})\, e^{i\vec{q}\cdot\vec{r}_j}. \tag{4.4}$$

The scattering intensity is then:

$$I(\vec{q}) = |F^2(\vec{q})| = \left| \sum_{j,l}^{n,n} f_j(\vec{q}) f_l(\vec{q})\, e^{i\vec{q}\cdot(\vec{r_j} - \vec{r_l})} \right|. \tag{4.5}$$

Assuming the sample is comprised of similar molecules, in the small scattering angles it has been shown that:

$$I(\vec{q}) = \left| \int_V dV f \rho_{\text{at}} e^{i\vec{q}\cdot\vec{r}} \right|^2, \tag{4.6}$$

where ρ_{at} is the average molecule density. Here, the form factor is defined as:

$$F(q) = \frac{1}{V} \int_V dV e^{i\vec{q}\cdot\vec{r}}, \tag{4.7}$$

and:

$$I(\vec{q}) = \rho^2 V^2 |F(q)|^2. \tag{4.8}$$

The final expression of the intensity can be divided into two terms. The $(\rho V)^2$ term is proportional to the scattering length density of the particle (ρ) squared and is related to the total observed intensity. The scattering length density is a measure of the scattering power of the sample. It is calculated by summating all atoms/molecules scattering contributions in the sample divided by volume. In this case, $\rho = \rho_{\text{sm}} - \rho_{\text{sl}}$ where ρ_{sm} is the scattering length density of the sample and ρ_{sl} is the scattering length density of the solution. From the correlation to the volume of the particle, its molecular weight can be approximated by [83]:

$$M_w = \frac{N_A I(0)}{c\Delta\rho}, \tag{4.9}$$

where N_A is the Avogadro number and $\Delta\rho$ is the effective scattering length density per mass. With good approximation, estimating the molecular weight through this method can result in values within 10% of the particle's molecular volume [84]. However, unwanted effects such as scattering from the sample holder may result in a different $I(0)$ than of the sample, rendering this analysis false.

The second term of the intensity is the form factor term $|F(q)|^2$, which is the Fourier-transformed structure of the particle and is related to the shape of the particles probed by the X-ray. When a lower symmetry structure of the building blocks is present, an additional term is added to the intensity equation: the structure factor $S(q)$. Following, the scattering intensity takes the form of $I(q) \propto |F(q)|^2 S(q)$.

4.2 Form Factor Analysis

As was shown in the previous section, the measured intensity of the sample directly relates to the Fourier transformation of its real-space structure. As such, the typical main goal of analyzing SAXS data is to identify the structural characteristics of the sample given the structural parameters extracted from its measured form factor.

4.2.1 Form Factors

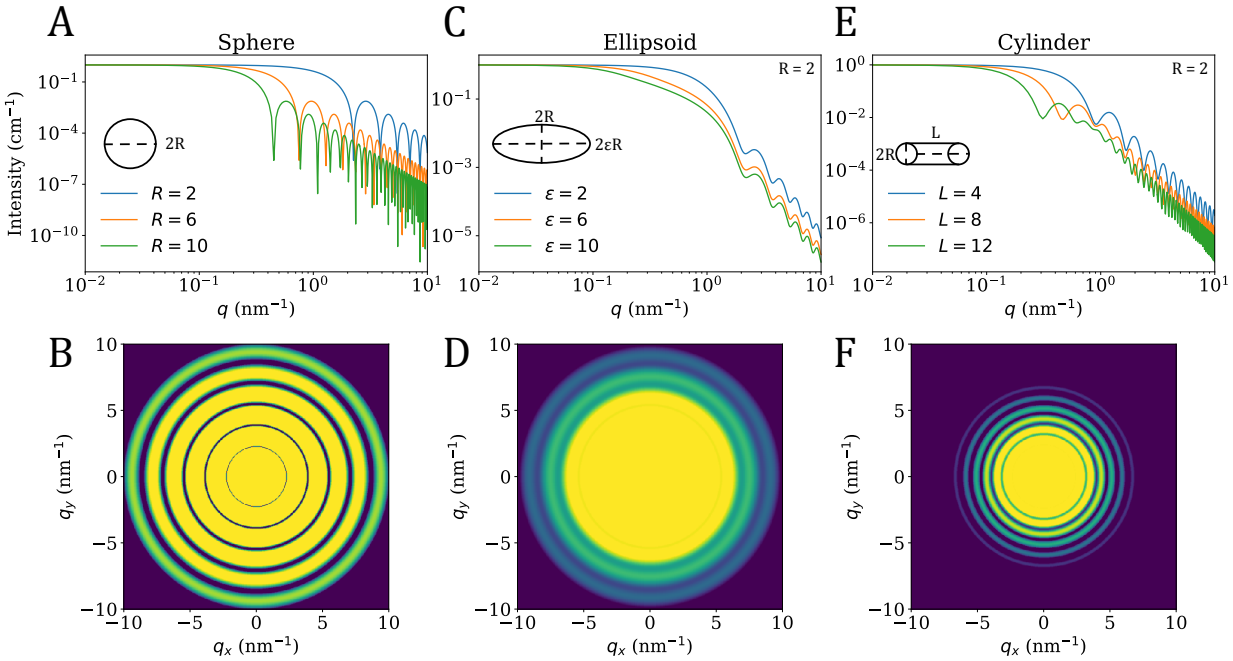

Figure 4.2: **A, B.** One-dimensional and two-dimensional scattering intensity of a sphere of radius R with varying R values. **C, D.** One-dimensional and two-dimensional scattering intensity of an ellipsoid of semi-axes radii $(R, R, \epsilon R)$ with varying ϵ values. **E, F.** One-dimensional and two-dimensional scattering intensity of a cylinder of radius R and length L with varying L values.

In this thesis, I will focus on samples having isotropic scattering. Isotropic samples possess no orientational preference, for example, due to many random orientations in the ensemble of sampled particles. The randomality of the sample orientation results in the X-ray scattering occurring at multiple relevant angles, and the form factor is to be decoupled from the angular (azimuth) angle of $\vec{q}$. Consequently, isotropic sample scattering incurs a loss of information from eliminating the secondary scattering axis, even if the basic building block is anisotropic (Fig. 4.2). However, the dependency on only one $\vec{q}$ component makes the scattering analysis easier (although not unique), as the angular component of $\vec{q}$ can now be integrated over. This integration results in a one-dimensional scattering intensity $I(q)$, which will be used in the following discussions. Below are some examples of one-dimensional form factors.

Spherical, ellipsoidal and cylindrical particles. The form factor of an ensemble of particles of a spherical radius is given by (Fig. 4.2A):

$$F(q) = \Phi^2(q, R), \tag{4.10}$$

where $\Phi^2(q,r)$ is the sphere scattering amplitude and can be found in table 2. The ellipsoid form factor is a more generalized version of the spherical variant. The ellipsoid is defined with three semi-axes with radii $(R, R, \epsilon R)$. The ellipsoid form factor is then an integration over all possible radii of the ellipse (Fig. 4.2B):

$$F(q) = \int_0^{\pi/2} d\alpha \sin(\alpha) \Phi^2(q, r(R, \epsilon, \alpha)), \tag{4.11}$$

where $r(R, \epsilon, \alpha)$ is found in table 2. The form factor of a cylinder of radius R and length L is given by [ref] (Fig. 4.2C):

$$F(q) = \int_0^{\pi/2} d\alpha \sin(\alpha) \Phi_c^2(q, r(R, \epsilon, \alpha)), \tag{4.12}$$

where $\Phi_c^2(q,r)$ is the cylindrical scattering amplitude and can be found in table 2.

Type	$\Phi(q,r)$	$r(R,\epsilon,\alpha)$
Sphere	$\frac{\sin(qr)-qr\cos(qr)}{(qr)^3}$	R
Ellipsoid	$3\frac{\sin(qr)-qr\cos(qr)}{(qr)^3}$	$R(\sin^2(\alpha)+\epsilon\cos^2(\alpha))^{1/2}$
Cylinder	$\frac{B_1(qr\cos(\alpha))\sin(qLcos(\alpha/2))}{qr\sin(\alpha)qL\cos(\alpha/2)}$	$R(\sin^2(\alpha)+\epsilon\cos^2(\alpha))^{1/2}$

Table 2: Scattering amplitudes of the spherical, ellipsoidal and cylindrical form factors. For the cylindrical form factor, B_1 is the Bessel function of the first kind.

Gaussian polymers The Gaussian polymer form factor describes the ensemble structure of Gaussian chains (section 2). The Gaussian form factor is given by (Fig. 4.3):

$$F(q) = \Psi^2(qR_G), \text{ where: } \Psi(x) = \frac{1-e^{-x}}{x^2}. \tag{4.13}$$

4.2.2 Structural Disorder

In some cases, the measured sample will have varying degrees of structural disorder. Samples such as an ensemble of homogeneous spheres possess no structural disorder, reflected in the X-ray scattering shown in Fig. 4.2. As the measured intensity profile reflects the reciprocal space of the sample's structure, the intensity curve will result in a series of repeating peaks at the high q range. This property of X-ray measurements is vital to crystallography, where the measured reciprocal

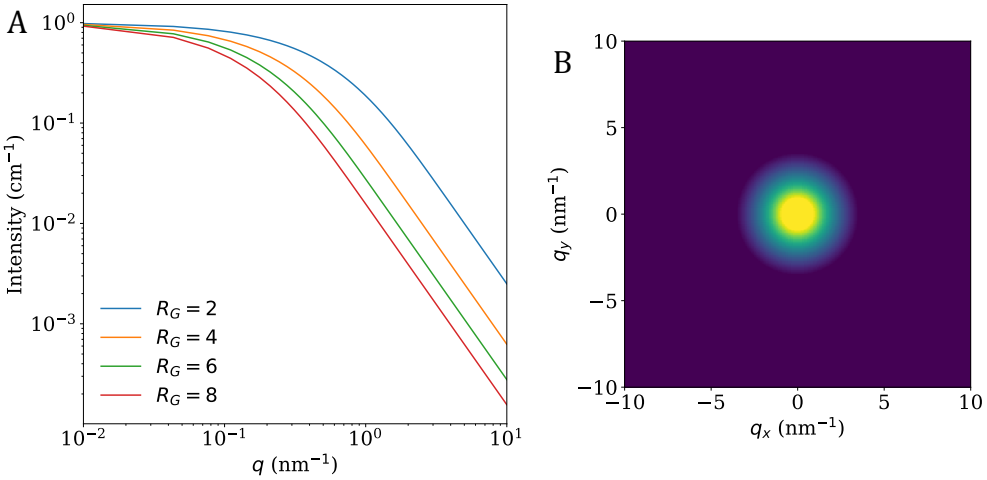

Figure 4.3: **A.** One-dimensional scattering intensity of a Gaussian chain with varying R_{G} values. **B.** Two-dimensional scattering intensity of a Gaussian chain with $R_{\mathrm{G}} = 2$.

space peaks are used to extract the lattice points of the measured crystallized sample [85]. In the opposite case, samples such as flexible polymers are entirely disordered, a characteristic that is also reflected in the resulting X-ray scattering. As the measured intensity now comprises an ensemble average on the reciprocal space of many different 'structures,' the high q range will show little to no repeating peaks (Fig. 4.3). Such scattering behavior makes it impossible to fully characterize the many possible structures of the sample, as with ordered particles or crystals. Instead, parameters such as the radius of gyration R_{G} are used to characterize the ensemble behavior of the conformations.

In addition to both extreme cases, many samples may exhibit a degree of disorder while still retaining the scattering peaks aspect of ordered structures. A commonly observed example of such scattering is in polydisperse (heterogeneous) samples. Polydispersity causes the measured peaks to exhibit 'stretching' based on the variance of the structural parameters of the sample (Fig. 4.4). In extreme cases, this stretching can cause a complete overlap of the peaks, resulting in an entirely disordered scattering profile.

The structural disorder of the sample is easily detectable by a Kratky plot (Fig. 4.5). In a Kratky plot, the intensity is multiplied by q^2 and is plotted against q. The scattering of a solid body decays at high angles at a rate of $I(q) \sim 1/q^4$ [86]. Multiplying $I(q)$ by q^2 results in a bell curve at the point where the decay of the scattering overcomes the rate at which $q^2 I(q)$ is increasing. For disordered samples, on the other hand, the high angle decay is at the rate of $I(q) \sim 1/q^2$ and will not result in a bell curve transition when multiplied by q^2 [86]. The same analysis can be done for partially disordered samples, first showing the structured peak at the

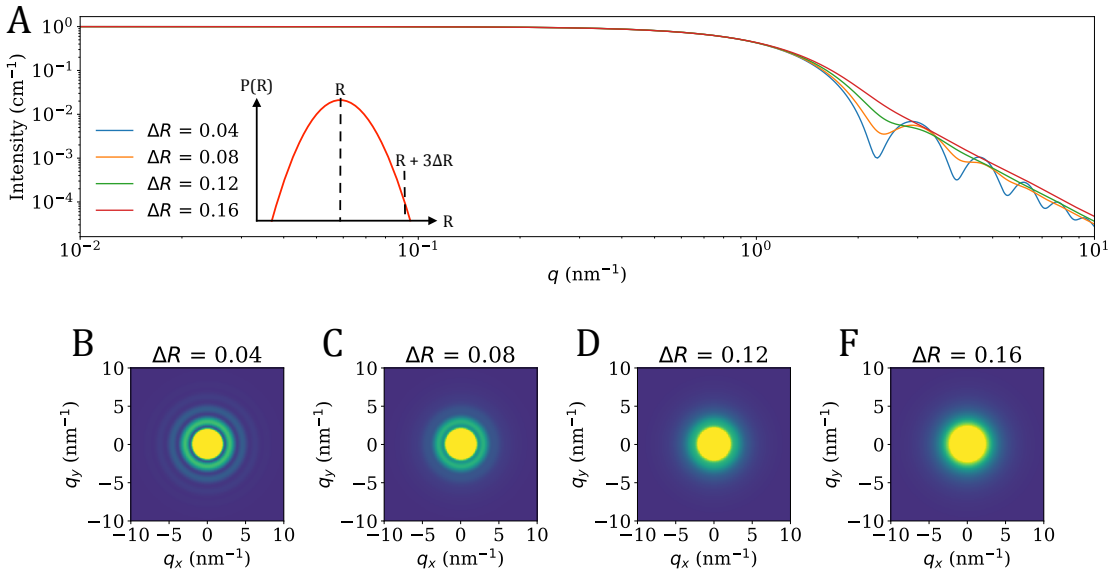

Figure 4.4: **A.** One dimensional scattering intensity of a polydisperse sphere with radius $R = 2$ and varying standard deviation ΔR. Radius polydispersity was modeled by a Gaussian distribution. **B - F.** Two dimensional scattering intensity of the polydisperse spheres .

lower q range but not exhibiting the $\sim 1/q^2$ decline at high angles.

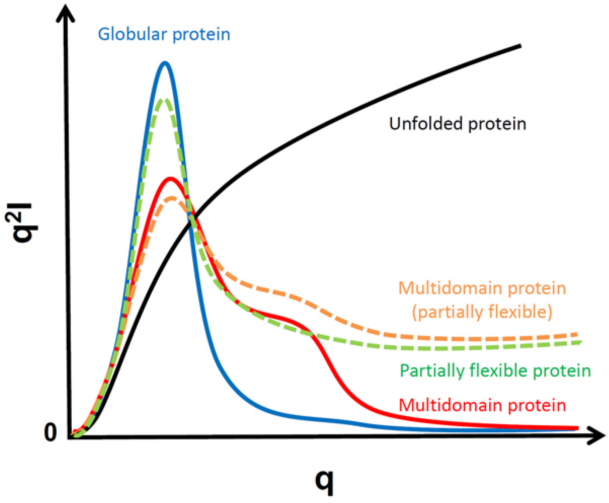

Figure 4.5: Kratky plots of proteins at multiple stages of disorder. Adapted from https://www-ssrl.slac.stanford.edu/smb-saxs/content/data-analysis-primer.

4.3 SAXS analysis of polymers

As the conformational flexibility of polymers possesses no orientational preference, their X-ray scattering is isotropic. Thus, the intensity function of measured polymers is given by the azimuthally integrated scattering intensity $I(q)$. Below, I will detail the most commonly used methods of analyzing polymer scattering:

Guinier analysis. One of the most commonly used analyses is the Guinier analysis, which can be used to estimate the particle's size [82, 5]. This estimation is done by evaluating the particle's radius of gyration R_{G}, which describes the average distance of a molecule from its center of mass (see section 4.1). In essence, the Guinier analysis is an approximation of Eq. 4.8 at the low q range, which turns to:

$$I(q) = I(q=0)e^{-\frac{q^2 R_{\mathrm{G}}^2}{3}} + O(q^4). \tag{4.14}$$

This approximation is effective up to $qR_{\mathrm{G}} \sim 1.3$. The Guinier method applies to polymers and unfolded proteins, whose scaling exponent (ν) can be estimated with Eqs. 2.9 and 2.13 for ideal and real chains, respectively. This method, however, possesses several shortcomings. First, the reliance of its upper effective bound on R_{G} means that larger chains have a smaller q range for which the analysis is applicable, resulting in larger uncertainty. Additionally, the Guinier method is known by previous experiments to mischaracterize the size of IDPs compared to more accurate methods [87].

Extended Guinier analysis. With the shortcomings of the Guinier method in mind, Zhang et al. have proposed an extended form of the Guinier analysis, which aims to be more suitable for IDP characterization [87]. Upon further expanding Eq. 4.14, the prefactor of the extended q term is dependant on ν, such that:

$$\ln\left(\frac{I(q)}{I(0)}\right) = -\frac{1}{3}(qR_{\mathrm{G}})^2 - 0.0479(\nu - 0.212)(qR_{\mathrm{G}})^4. \tag{4.15}$$

This extended form of the Guininer analysis was shown by Zheng et al. and by subsequent studies to be able to extract more accurate information on the protein's scaling [87, 26].

Gaussian chain approximation. An alternative method of characterizing the R_{G} of chains is by fitting the scattering intensity to a Gaussian chain form factor (section 4.2.1). This method was shown to work relatively well with chains close to ideal conditions, i.e., with $\nu \sim 1/2$. Outside of $\nu = 1/2$, however, deviations from the model can be expected since the chains no longer obey Gaussian statistics.

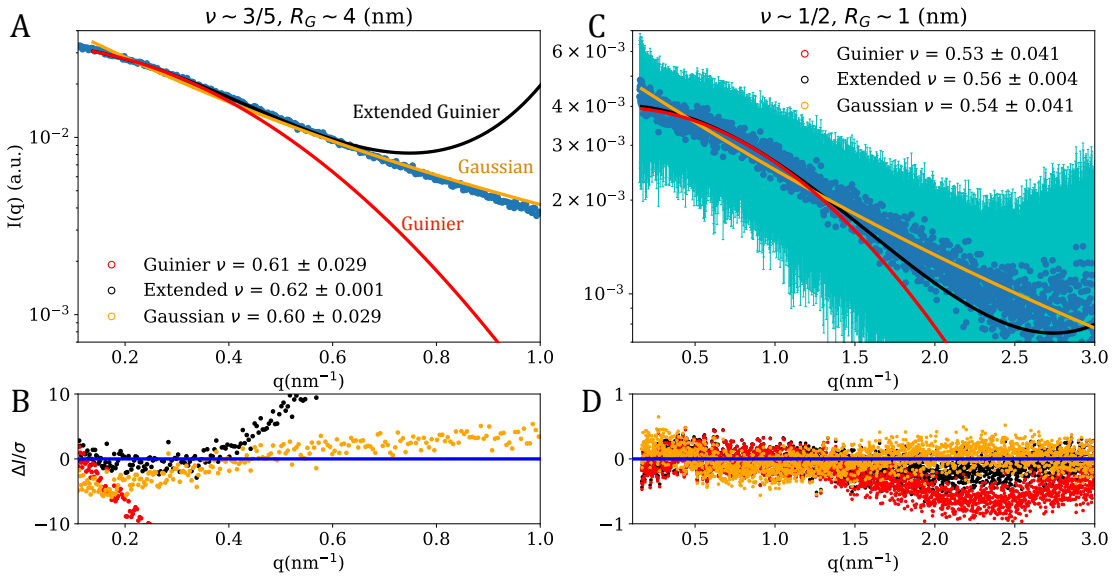

Figure 4.6: Guinier extended Guinier and Gaussian chain fittings and fitting errors on two different IDPs. **A - B.** IDP with a ν close to 3/5 and an R_G of $\sim$ 4 nm (α-Synuclein IDP at 2mg/mL). ν was calculated by Eq. 2.13. Bad fittings for the Gaussian chain model indicate that the chain is not ideal, with the resulting $\nu = 0.6$ contradicting the expected $\nu = 0.5$. The Guinier and Extended Guinier analyses exhibit a deviance from the data at very low q due to the value of R_G. **C - D.** IDP with a ν close to 1/2 and an R_G of $\sim$ 1 nm (Segment 26-45 of NFLt at 1 mg/ml as adapted from [26]). ν was calculated by Eq. 2.13. Calculated ν values were close to 0.5, thus Gaussian fitting proved better than in A. Lower R_G value than in A. resulted in a much later deviation of the Guinier and Extended Guinier analyses from the data.

4.3.1 SAXS scattering of interacting polymers

When the sample ensemble contains intermolecular interactions, its measured intensity is consequently affected through the inclusion of a non-trivial structure factor $S(q)$. These effects may occur in cases such as a system of charged polymers, where intermolecular electrostatic interactions are expected to affect the conformational ensemble.

Intermolecular effects can quantified by viral expansion, where the second virial coefficient, A_2, is defined as the pairwise deviation of a particle from the ideal gas. Positive A_2 values indicate repulsive interactions between the particles, while negative indicate attractive interactions. A_2 is related to the scattering intensity by [88, 89]:

$$\frac{M_w K c}{I(q,c)} = \frac{1}{F(q)} + 2M_w A_2 c, \tag{4.16}$$

where K is a constant dependent on the diffraction coefficient, c is the particle concentration in mg/ml, $F(q)$ is the scattering form factor (see section 4.1), and M_w is the molecular weight of the particle. Defining $\tilde{I}(q,c) \equiv I(q,c)/c$, and evaluating at $q \to 0$ where $F(q) \to 1$, Eq. 4.16 becomes:

$$\frac{M_w K}{\tilde{I}(0,c)} = 1 + 2M_w A_2 c. \tag{4.17}$$

It is convenient to normalize the $1/\tilde{I(0,c)}$ so that $\tilde{I}(0,0) = 1$, as it then follows from Eq. 4.17 that $M_w K = 1$, and finally:

$$\frac{1}{\tilde{I}_{\text{norm}}(0,c)} = 1 + 2M_w A_2 c \tag{4.18}$$

This analysis is typically presented in the form of a Zimm plot (Fig. 4.7), which plots the low q region of $1/\tilde{I}(q,c)$ at various concentrations against q^2, with an offset of $\alpha \cdot c$, where α is an arbitrary constant. The q region of the plot is sufficiently small so that the $1/\tilde{I}(q,c)$ curve is approximately linear, from which the intensity at $q \to 0$ can be approximated. Finally, the normalized $1/\tilde{I}(q,c)$ intensities for each concentration are used to extract A_2, as described above. In addition to $1/\tilde{I}(0,c)$, $1/\tilde{I}(q,0)$ can also be extracted from the Zimm plot, which describes the theoretical scattering at zero concentration - at the absence of intermolecular interactions.

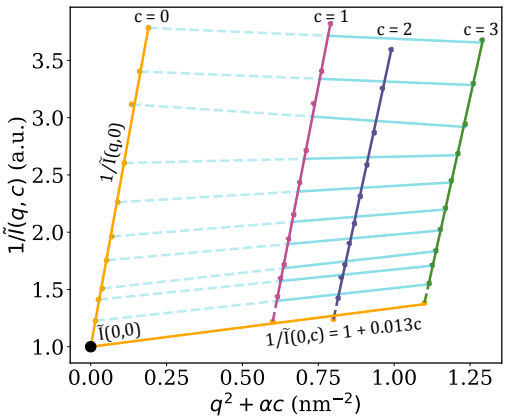

Figure 4.7: Normalized Zimm plot. Green, purple and pink lines are the $1/I(q)$ curves of the protein at concentrations $3, 2, 1$ times initial concentration $c = 0.5mg/ml$ respectively. From the $1/I(q)$ lines, normalized $1/\tilde{I}(q, c = 0)$ and $1/\tilde{I}(q = 0, c)$ lines (yellow) are obtained by linear fitting. Finally, the second virial coefficient A_2 can be obtained as described in Eq. 4.18.

5 Magnetic Tweezers

5.1 Overview

The non-equilibrium experiments of my study were conducted using a custom magnetic tweezers setup [28, 90]. Magnetic tweezers experiments begin with the preparation of the flowcell, upon which the sample chains are tethered. After the tethering process, magnetic beads are attached from the free end of the tethers. The tethers can now be stretched by the magnet assembly, which controls the magnetic field applied on the tethers by the rotation and height of the magnets. Above the magnet, a laser is focused on the sample, projecting through the flowcell to a high-speed camera. The camera is then able to record the 3D movement of the tethers using tracking software [90, 28, 29]. The method by which force calibration was achieved is described in section 5.2, and the polymer models used to describe the behavior of the polymers under strong and weak tension are described in section 5.3.

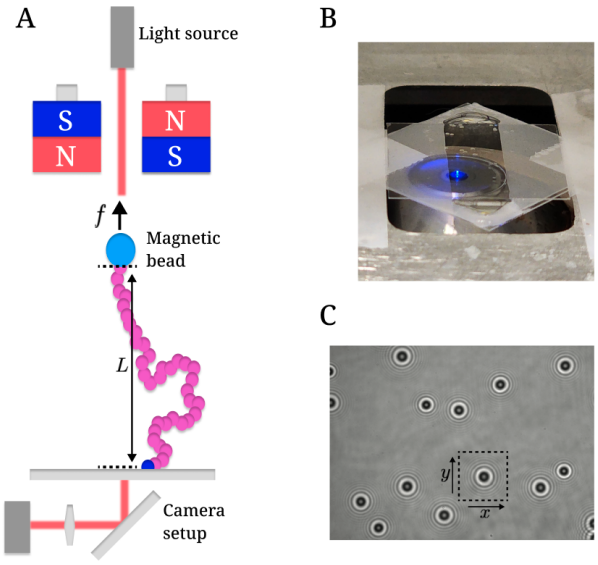

Figure 5.1: **A.** Schematic of the magnetic tweezers setup. From top to bottom: A laser light source focuses on the sample through the magnetic tweezers. The magnets stretch and rotate the sample tether (purple) with an upward stretching force f. From the bottom of the stage, the light from the laser source travels into a camera setup, from which the tethers are recorded. **B.** Photograph of the flowcell. Below the flow cell is the focus point from which the laser's light travels to the camera. The stage is motorized so that the flowcell position relative to the magnets and the laser is controllable. **C.** Resulting tether image from the captured camera, as adapted from [91].

5.2 Force calibration

A correct force calibration scheme is crucial to the success of any force-dependant single-molecule experiment. Initially, the force applied from the magnets on the bead may be calculated by the Maxwell equations:

$$\vec{f} = \left[\vec{m}(\vec{B}) \cdot \vec{\nabla}\right] \vec{B}, \tag{5.1}$$

where $\vec{B}$ is the magnetic field and $\vec{m}$ is the magnetic moment of the bead. In practice, bead-size polydispersity may result in immense deviations in the magnetic moment between beads, with a 3% polydispersity in the beads resulting in up to 9% error in calculations [92]. As such, this method is impractical for accurately reading the stretching force. Instead, each bead is calibrated individually by fitting a theoretical power spectral density (PSD) of a Brownian motion to the calculated discrete PSD of the bead. Nonetheless, measurement effects such as non-uniform frequency response, data distortion, and non-normally distributed measurement errors may prove heavily detrimental to obtaining a good correlation to the theoretical model [91]. Here, I will summarize the analysis method proposed by Lansdorp and Saleh [91], which aims to alleviate these shortcomings.

As mentioned, the theoretical PSD of the bead is obtained by assuming it complies with Brownian motion. This assumption holds due to the nature of the single-molecule experiment, which allows us to presume the particle to diffuse in isolation. Additionally, the mass-related effects of the motion are assumed to be negligible, and the tethered bead behaves as an over-damped pendulum of spring constant κ. Thus, bead position x is given by:

$$\kappa x + \alpha \dot{x} = F_L, \tag{5.2}$$

where $\alpha = 2\pi\eta r$ is the dissipation of a spherical bead of radius r in a viscosity solution η, and F_L is the Langevin force. Using a Fourier transformation, we may now calculate the PSD:

$$P(f) = \frac{k_B T}{2\pi^2 \alpha (f_c^2 + f^2)}, \tag{5.3}$$

where $f_c = \kappa / 2\pi\alpha$. From here, the positional variance $\langle x^2 \rangle$ can be found by integrating over $P(f)$, resulting in the equipartition result of $k_B T/\kappa$. In finding κ, the applied force may now be calculated as the force tension of a spring by $F = \kappa z$, where z is the tether extension.

The connection to the theoretical data is done through the bead's Allan variance (AV) [93]. AV is defined as half the ensemble-averaged variance of the difference between two consecutive time samples:

$$\sigma^2(\tau) = \frac{1}{2} \left\langle (\bar{x}_{\tau,j+1} - \bar{x}_{\tau,j})^2) \right\rangle, \tag{5.4}$$

where:

$$\bar{x}_{\tau,j} = \frac{1}{\tau} \int_{\tau(j-1/2)}^{\tau(j+1/2)} dt\, x(t). \tag{5.5}$$

$\bar{x}_{\tau,j}$ closely resembles the bead position measured by the camera, which is given by:

$$x_i = \frac{1}{\tau_s}\int_{t_i-\tau_s/2}^{t_i+\tau_s/2} dt x(t), \tag{5.6}$$

where t_i is measurement time of video frame i, and τ_s is the shutter speed of the camera. As such, $\bar{x}_{\tau,j}$ is intuitively derived from the measured data. The AV is shown to correlate to the PSD by:

$$\sigma^2(\tau) = \int_{-\infty}^{\infty} df \frac{4\sin^4(\pi f\tau)P(f)}{\pi f\tau}, \tag{5.7}$$

which, using Eq. 5.3, is solved as:

$$\sigma^2(\tau) = \frac{k_BT\alpha}{\kappa^2\tau}\left(1 + \frac{2\alpha}{\kappa\tau}e^{-\kappa\tau/\alpha} - \frac{\alpha}{2\kappa\tau}e^{-2\kappa\tau/\alpha} - \frac{3\alpha}{2\kappa\tau}\right). \tag{5.8}$$

Experimentally, AV is evaluated by binning the data into overlapping octaves, i.e., overlapping bins of increasing bean size $m = 2^n$, where $n = 1, 2, ..., N$. The containers are overlapping such that the first bin is (1,...,m), the second (2, ...,m+1), the third (2,...,m+1), and so on for m=1,2,4,8... In this case, the oversampled AV (OAV) is calculated at timescales $\tau = m\tau_s$ and is given by:

$$\sigma_m^2 = \frac{1}{2(N-2m)(m\tau_c)}\sum_{k=1}^{N=2m}(x_{k+2m} - x_{k+m} + x_k)^2. \tag{5.9}$$

Finally, from fitting the oversampled AV of the data to Eq. 5.8 κ can be obtained, from which the stretching force is easily calculated by $F = \kappa\bar{z}$, where $\bar{z}$ is the average of the measured tether extension (Fig. 5.2).

5.3 Polymer Stretching

5.3.1 Scaling

Polymers undergoing stretching may be described by scaling arguments [94]. In the scaling argument, the chain is divided into n blobs of g monomers each (Fig. 5.3), all of which scale according to:

$$\xi \approx bg^{\nu}, \tag{5.10}$$

where ν was found to be 1/2 for ideal chains and 3/5 for real chains. As such, the entire chain length is given as the product of all blobs:

$$z \approx n\xi \approx \frac{R_{ee}^{1/\nu}}{\xi^{1/\nu-1}}, \xi \approx \frac{R_{ee}^{1/(1-\nu)}}{z^{\nu/(1-\nu)}}. \tag{5.11}$$

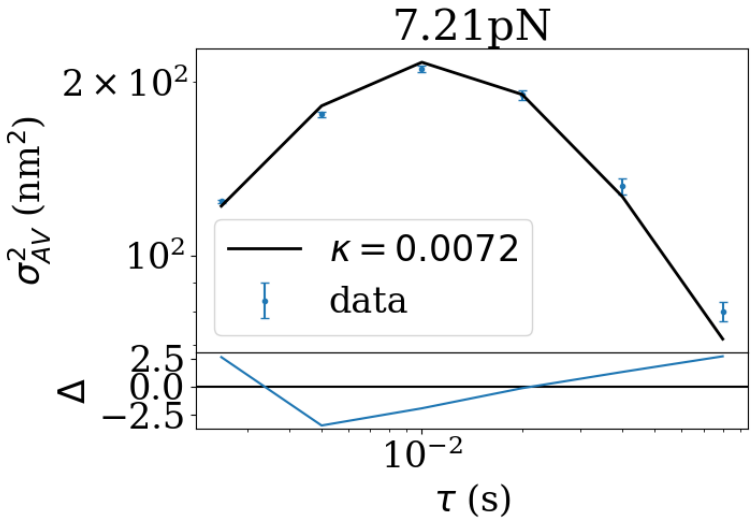

Figure 5.2: Example of Allan variance (AV) fitting. To the data (blue), the theoretical AV (Eq. 5.8) is fitted, resulting in a κ of 0.0072 and a force of , 7.21 pN.

For these arguments to hold, blob size ξ is chosen such that in length scales smaller than ξ, the chain "appears" unperturbed, and in length scales larger than ξ, the chain appears stretched. As such, each blob contributes k_BT to the free energy of the chain, and:

$$F \approx k_BTn \approx k_BT\frac{z}{\xi} \approx k_BT\left(\frac{z}{R_G}\right)^{1/(1-\nu)}. \tag{5.12}$$

From here, the stretching force f required to extend the polymer to z is given by:

$$f = \frac{\partial F}{\partial z}. \tag{5.13}$$

Finally, we obtain the following relation between f and z:

$$\frac{fb}{k_BT} \approx \left(\frac{z}{Nb}\right)^{1/(1-\nu)-1}. \tag{5.14}$$

Importantly, the relation of the force extension of the chain is valid within the limits of $R_{ee} < z \ll Nb$.

5.3.2 Worm-like chain (WLC)

As in the case of free polymers, the extension of stretched polymers may also be solved by energetic considerations. Here, this is done by the worm-like chain (WLC) model of polymer

stretching, which assumes the polymer to be near 'rod-like' under considerable stress [95] (Fig. 5.3). In this state, the polymer features numerous bends and is described by the curvature vector $\vec{r}(s)$, where s is arc length. When stretched to a rod, the maximum extension available to the length is L_c, defined as the contour length. Additionally, the tangent vector of the curve is given by $\hat{t} = \partial_s \vec{r}$ and is defined as the unit vector. The final parameter is the persistence length l_p, the length scale over which the tangent vector remains unchanged, *i.e.*, the chain is estimated as rod-like. In the context of an ideal random walk chain, this results in a Kuhn length $b = 2l_p$ and end-to-end length $R_{\text{ee}} = \sqrt{2l_p L_c}$. From here, the bending energy cost per length is given by $k_B T l_p \kappa^2$, where $\kappa = |\partial_s(\vec{r}^2)|$. In total, the effective energy of the chain is given by [95]:

$$\frac{E}{k_B T} = \int_0^{L_C} ds \, \frac{l_p \kappa(s)^2}{2} - fz, \tag{5.15}$$

where f is the force required to extend the chain to length z. Here, within strong and weak stretching limits, upon minimization of the extension z the following interpolated relation can be obtained [95]:

$$\frac{f l_p}{k_B T} \sim \frac{z}{L_c} + \frac{1}{4(1 - z/L_c)^2} - \frac{1}{4}. \tag{5.16}$$

From an analytical approximation, a more accurate form of the WLC relation can be obtained [96]:

$$\frac{f l_p}{k_B T} = \frac{1}{4(1 - z/L_c)^2} - \frac{1}{4} + \frac{z}{L_c} + \sum_{i=2}^{i \leq 7} \alpha_i \left(\frac{z}{L_c}\right)^i, \tag{5.17}$$

where $\alpha_2 = -0.5164228$, $\alpha_3 = 2.737418$, $\alpha_4 = 16.07497$, $\alpha_5 = -38.87607$, $\alpha_6 = 39.49944$, and $\alpha_7 = -14.17718$.

5.4 NFL glassy dynamics

In a previous study, the NFLt IDP was measured using a magnetic tweezers setup [29]. Due to the relevance of this study to my thesis, its results will be summarized below. In that study, our lab polymerized multiple NFLt to create long polymers suitable for magnetic tweezer experiments. Next, the poly-NFLt was measured in two- and three-step alternating force experiments.

In the two-step force experiments, The polymerized NFLt subjected to were subject to a pulling force f transitioned from a high force (f_1) to a lower one (f_2). Such a scheme resulted in a logarithmic relaxation of the polymerized chain. These relaxations were shown to fit $\Delta L = b \log(t/t_0)$, $\Delta L = L(t) - L(t_0)$, where L is the measured bead height, t_0 is an arbitrary reference time, and b is the relaxation constant (Fig. 5.4). This logarithmic mode structure is associated with an experimentally-measured structure memory effect: the Kovacs effect [97] (Fig. 5.5). To validate the Kovacs memory effect for poly-NFLt, a second set of experiments introducing a third force f_3 were conducted. Specifically, the third pulling force was intermediate in strength

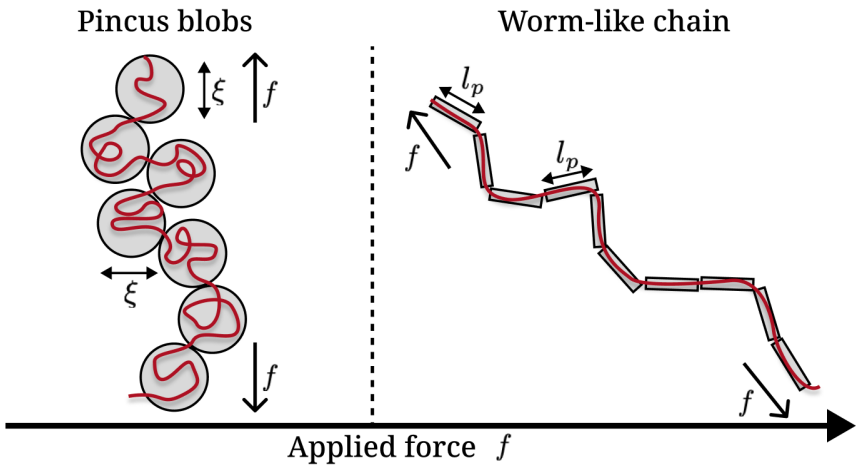

Figure 5.3: Schematic of the Pincus blob and worm-like chain (WLC) models of polymer stretching. At low stretching force f, the Pincus blob method is applied, approximating the chain into stretched spheres of size ξ. At high stretching force, the WLC regime is applied. In the WLC regime, the polymer is described as a series of rods of length l_p, where l_p is the persistence length.

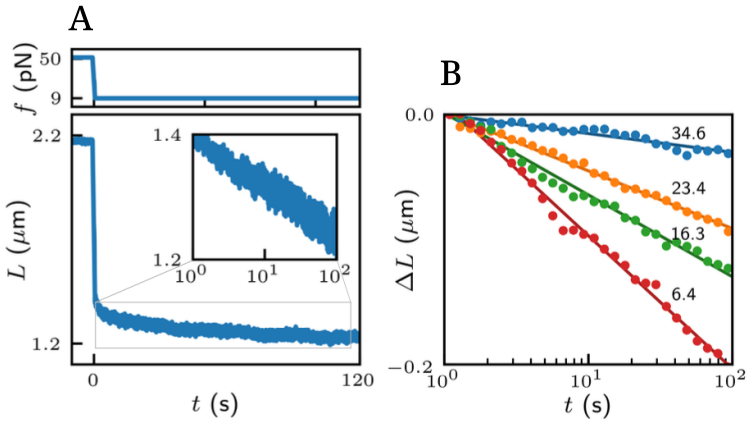

Figure 5.4: **A.** Example of a two-step experiment. At $t = 0$, the force is decreased from $f_1 = 50$pN to $f_2 = 9$pN, resulting in a logarithmic relaxation (inset). **B.** Linear fitting of tether relaxation $\Delta L = L(t) - L(t_0)$ to $\log(t/t_0)$ for different relaxation experiments of the same $f_1 = 50$ pN. Figures adapted from Ref. [29].

($f_1 > f_3 > f_2$), and the switch to f_3 occurred at a time short enough so the chain could not relax completely. Importantly, the f_3 measurements showed an increase in $L(t)$ at the start of the switch, followed by a decrease after a certain point in time - a signature of the Kovacs effect. Being

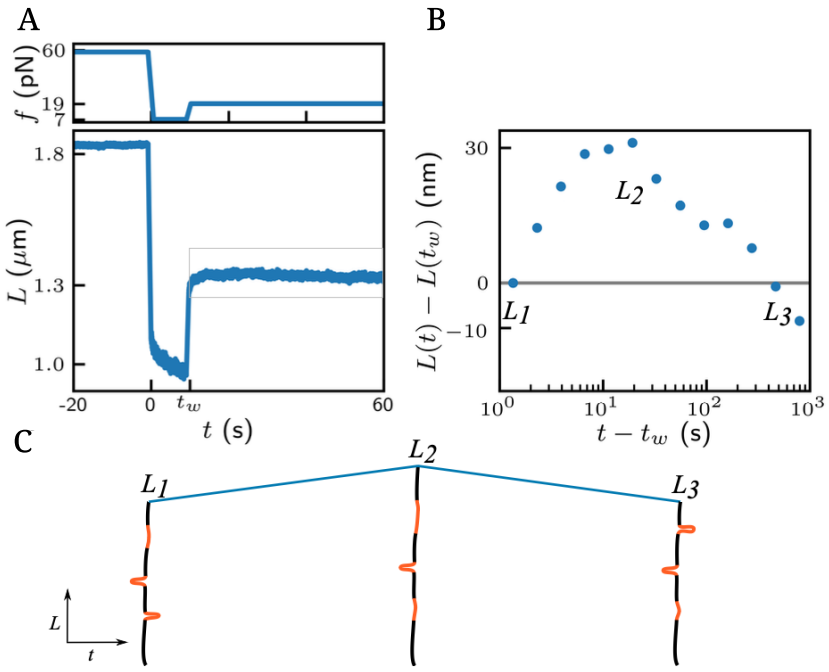

Figure 5.5: **A.** Example of a three-step experiment. The force was initially $f_1 = 60$ pN, then held at $f_2 = 7$ pN for $t_w = 10$s, then increased to $f_3 = 19$ pN. **B.** Binned extension dynamics of f_3 from A, showing a Kovacs hump [97]. **C.** Schematic of the underlying dynamics of the Kovacs hump. Incubation of f_2 for a shorter time causes most segments to fold, with the slower ones remaining expanded (left). Transitioning to f_3, the fast segments expand, leading to tether height $L_2 > L_1$. Finally, the slow tethers begin to fold, resulting in tether height $L_3 < L_2$. Figures adapted from [29].

an IDP, these heterogeneity-associated effects were suggested to be due to multiple modes within NFLt that relax independently with a broad spectrum of timescales. In correlation to the Kovacs hump, as the force transitions from f_1 to f_2, the slow modes remain extended, while the faster modes have already collapsed due to the transition. Transitioning from f_2 to the higher force f_3 before equilibrium ensures that the slow modes remain unchanged and the collapsed modes expand. This transition leads to a slow tether height expansion, which continues until the slow modes begin to collapse. Following, the height of the tether starts to decrease slowly. Using an established mathematical framework for such relaxations [98], the poly-NFLt results were fitted to a model that assumes that each of the n IDP segments relaxes on average exponentially within

a timescale. Assuming that the timescale distribution $P(\tau) \propto 1/\tau$:

$$L(t, f_2) - L(t_0, f_2) = -N\alpha(f_2)[\eta(f_2) - \eta(f_1)] \log(t/t_0), \tag{5.18}$$

where $\eta(f)$ is the density of mode states in log units, N is the degree of polymerization, and $\alpha(f)$ is the relaxation amplitude. $\alpha(f)$ was chosen by the worm-like chain model, where $\alpha(f) = \ell\alpha_0(f)$, where ℓ is the coil contour length, and $\alpha_0(f)$ is the worm-like chain relative extension. $\eta(f)$ is found by $P(\log(\tau)) \equiv \eta$. Assuming the dependence on force to be by an activation barrier ΔG, which by the Bell-Zhurkov expectation varies as $f\Delta x$ for activation distance Δx. As such, $\tau = \tau exp(\Delta G/k_B T)$, and ΔG is uniformly distributed due to the constraint on τ. From here: $\eta(f) = k_B T/f\delta x$, where δx is the distance between the activation barriers of the different segments. Finally, this expression was used together with Eq. 5.18 to obtain:

$$bf_1/Nk_B T\alpha_0(f_2) \equiv \bar{b} = \frac{1}{\rho}\left(1 - \frac{1}{\bar{f}}\right), \tag{5.19}$$

where $\bar{f} = f_2/f_1$, and $\rho = \delta x/\ell$ is the ratio between the activation distance and the contour length. ρ then signifies the relative length of each relaxation mode. This model fit well with the measured curves, with $\rho = 0.1$.

6 Results

6.1 Experimental setup

As elaborated in section 1.3, the NFLt protein can be divided into two sub-domains: an uncharged domain starting from its N-terminal and a negatively charged polyelectrolyte domain starting from residue 52 until the C terminal of the protein. To study the effects of each, two variants of NFLt were designed and measured. The first variant, WT, is the entire NFLt chain, while the second variant, ΔN42 , is a truncation of the first 42 amino acids of the protein, leaving only the charged domain (Fig. 6.1).

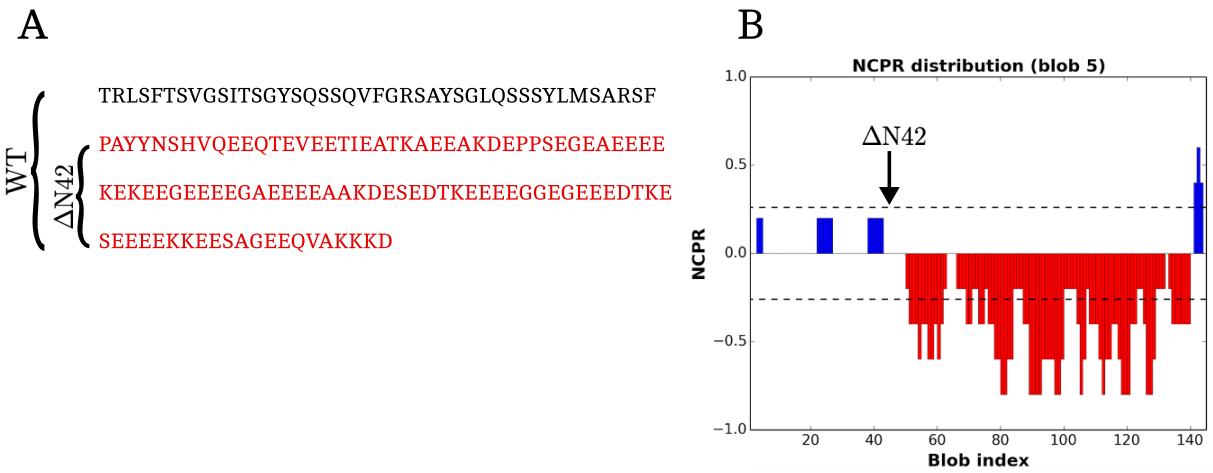

Figure 6.1: **A.** WT and ΔN42 sequence. Red indicates the ΔN42 regime, where the sequence is mostly negatively charged. **B.** Net charge per residue NCPR of NFLt. WT NCPR is of the entire chain, while ΔN42 starts from the arrow. Graph provided by CIDER with an amino acid binning of 5 residues per blob [53].

The two variants were expressed using E. coli and purified up to 96% as described in the methods. The following results are of two measurement methods: equilibrium ensemble measurements of the variants with SAXS and non-equilibrium single-molecule measurements with a magnetic tweezers setup. The SAXS results have been adapted from the paper which we have written on the results of the experiments [1].

6.2 The Truncated NFLt

The ΔN42 variant was measured at five different buffers, which differ by the amount of salt added to 20 mM Tris pH 8.0. Additionally, each buffer measurement was done in three protein

concentrations.

6.2.1 Raw data

Raw 2D data was integrated azimuthally due to the sample scattering being isotropic. All ΔN42 measurements appear disordered, with a distinct lack of ordered structural characteristics. This deduction can also be seen in the Kratky plots (Figs. 6.2B, 6.5).

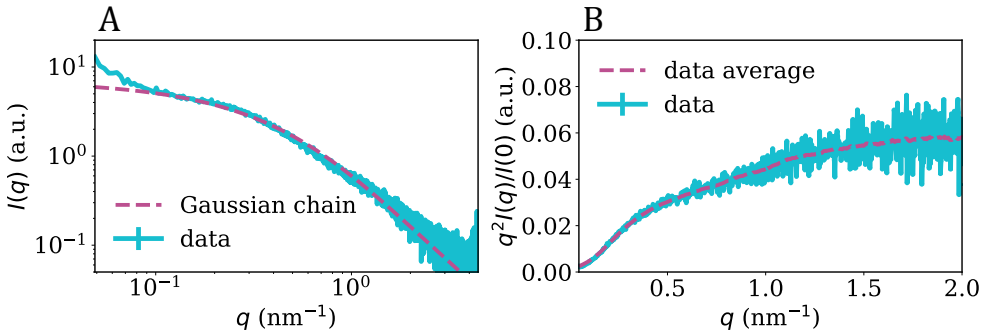

Figure 6.2: Raw data of ΔN42. **A.** Intensity ($I(q)$) against q. Scattering data resembles that of a disordered Gaussian chain (dashed line). **B.** Normalized Kratky plot ($q^2I(q)/I(0)$ vs q). Normalized Kratky plot of the same SAXS measurements. The plot displays disordered protein characteristics. Both panels represent the ΔN42 variant measured in 20 mM Tris buffer, pH 8.0. and 1 mg/ml. Figure adapted from [1].

As a disordered polymer, initial data fittings to the Gaussian model (Sec. 4.2.1) prove relatively successful, although with observable deviations from the measured intensity curve. However, as the measured IDP is not expected to be an ideal chain, deviations from the Gaussian model were expected. The most appropriate approach to analyze ΔN42 would then be the extended Guinier method (Sec. 4.3) due to its ability to extract the scaling exponent ν (Figs. 6.3, 6.4).

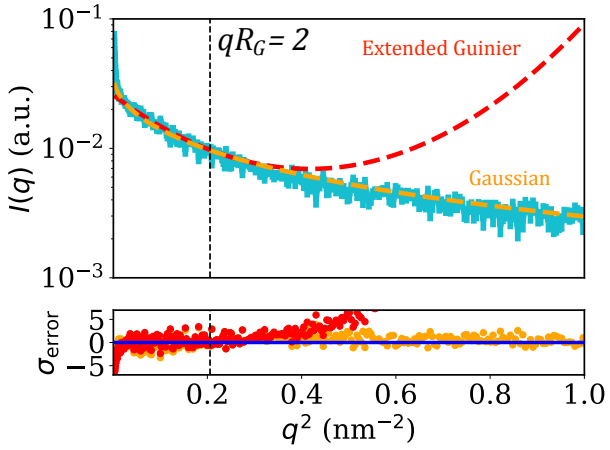

Figure 6.3: ΔN42 Intensity ($I(q)$) fittings to the extended Guinier (red) and Gaussian (yellow) models with deviation from data $\sigma_{\text{error}} = (I_{\text{fit}} - I_{\text{data}})/\sigma_{\text{data}}$. Noticeable deviation from the extended Guinier fitting arises at $qR_{\text{G}} \sim 2$. The figure represents the ΔN42 measurement in 20 mM Tris, pH 8.0, and 1 mg/ml protein concentration. The concentrations are arranged from C_3 being the highest and C_1 being the lowest measured. All concentrations and fitting results are found in table 3.

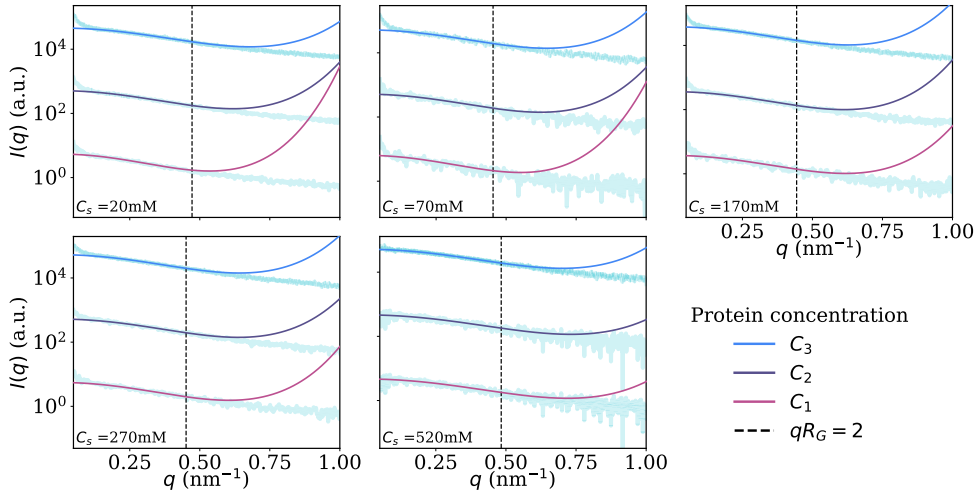

Figure 6.4: Extended Guinier fittings for all ΔN42 measurements. Dashed lines indicate the point up to the fitting was made, at $qR_{\text{G}} = 2$. Figure adapted from [1].

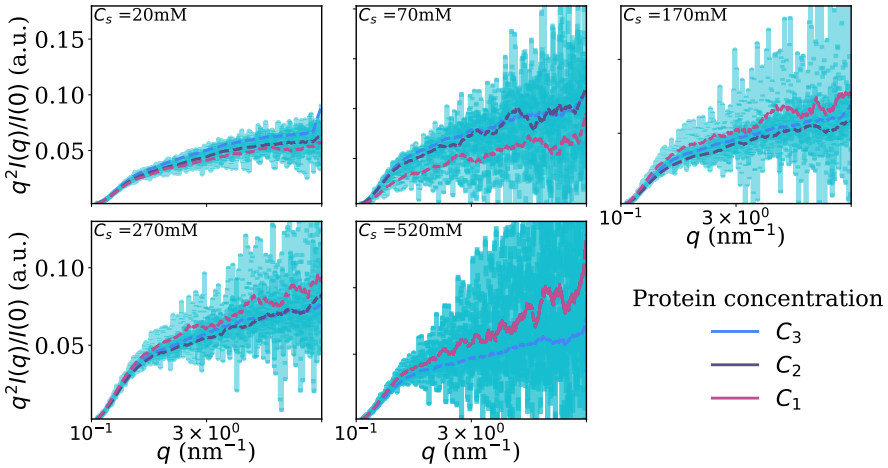

Figure 6.5: ΔN42 normalized Kratky plots ($q^2I(q)/I(0)$). Measurements show high level of disorder, with negligible difference between the alternative protein concentrations. Here, the concentrations are arranged from C_3 being the highest and C_1 being the lowest measured. All concentrations are found in table 3.

C_s (mM)	C (mg/ml)	R_G (nm)	ν	I_0 (cm^{-1})
20	1.1	4.23 ± 0.05	0.642 ± 0.002	0.0228
20	0.8	4.56 ± 0.07	0.660 ± 0.66	0.0254
20	0.6	5.11 ± 0.13	0.689 ± 0.007	0.027
70	1	4.41 ± 0.12	0.652 ± 0.007	0.0259
70	0.5	4.53 ± 0.22	0.659 ± 0.012	0.0257
70	0.3	4.99 ± 0.46	0.683 ± 0.023	0.032
170	1.5	4.51 ± 0.05	0.657 ± 0.003	0.19
170	0.8	4.59 ± 0.11	0.662 ± 0.006	0.18
170	0.3	4.56 ± 0.27	0.661 ± 0.015	0.18
270	1	4.43 ± 0.06	0.653 ± 0.003	0.026
270	0.5	4.45 ± 0.1	0.654 ± 0.006	0.026
270	0.3	4.64 ± 0.16	0.664 ± 0.009	0.028
520	1.5	4.14 ± 0.02	0.636 ± 0.001	0.026
520	0.78	4.00 ± 0.08	0.628 ± 0.005	0.023
520	0.38	4.05 ± 0.35	0.630 ± 0.02	0.023

Table 3: **ΔN42 Extended guinier analyis data**. Analysis parametres (radius of gyration R_G, scaling exponent ν, and scattering intensity at $q = 0$ (I_0)) obtained for different salt concentrations (C_s) and protein concentrations (C). Table adapted from [1].

6.2.2 Scaling

As a polyelectrolyte, it is vital to separate the discussion on ΔN42 into two parts: low salt and high salt conditions.

Low salt regime. In low salt conditions (salt concentration C_s = 20mM Tris, pH 8.0), measurements of the lowest concentration (0.6 mg/ml) resulted in $R_G = 5.11 \pm 0.13$ nm and $\nu = 0.689 \pm 0.007$. While the measured ν is higher than that of real chains, its value falls within expectations for polyelectrolytes [26]. Increasing protein concentration, we see a drop in R_G and ν, leading to $R_G = 4.23 \pm 0.05$ nm, $\nu = 0.642 \pm 0.002$ at the highest concentration of 1.1 mg/ml. This concentration-dependent shift is highly likely due to electrostatic intermolecular interactions. In order to 'eliminate' the intermolecular effects, we employ the 'zero concentration' values of R_G and ν, which depict the structural ensemble of the chain in isolation. These values are obtainable by performing the extended Guinier analysis on the Zimm $I(q, c = 0)$ curve or by extracting the $c = 0$ point on the $R_G(c)$ and $\nu(c)$ plots by linear fitting (Fig. 6.6, table 4). These result in $R_G(c = 0) = 5.76 \pm 0.31$ nm and $\nu(c = 0) = 0.729 \pm 0.015$.

C_s (mM)	R_G (nm)	ν
20	5.76 ± 0.31	0.729 ± 0.015
70	4.84 ± 0.27	0.677 ± 0.015
170	4.71 ± 0.04	0.669 ± 0.003
270	4.61 ± 0.16	0.663 ± 0.009
520	3.88 ± 0.04	0.620 ± 0.003

Table 4: **Zero concentration extended Guinier analysis data**. Analysis parameters (radius of gyration R_G and scaling exponent (ν) were extrapolated to zero protein concentration at various salt concentrations (C_s). Table adapted from [1]

High salt regime. Differences in the protein extension are expected upon increasing salt concentration (C_s). Starting with an increase of 50 mM NaCl, R_G and ν already show little response to changes in protein concentration (Fig. 6.6). This trend continues as C_s is further increased, indicating that at $C_s = 70$ mM, a transition to the high-salt regime occurs. Interestingly, at $C_s = 520$ mM, R_G and ν increase slightly with protein concentration, likely due to complex excluded volume interactions growing prominent in the apparent absence of strong electrostatic repulsion. Upon reaching $C_s = 70$mM, zero concentration R_G and ν drop to 4.84 ± 0.27 nm and 0.677 ± 0.015 respectively, a trend which continues up until $C_s = 520$ where $R_G = 3.88 \pm 0.04$ and $\nu = 0.620 \pm 0.003$ (Fig. 6.6). As suspected, the ΔN42 variant at $C_s > 70mM$ is expected to be in the high-salt polyelectrolytes regime. As such, the salted model of polyelectrolytes can be applicable (Section 2.3, [64]) where:

$$R_G = R_0 b N^{\nu}, \tag{6.1}$$

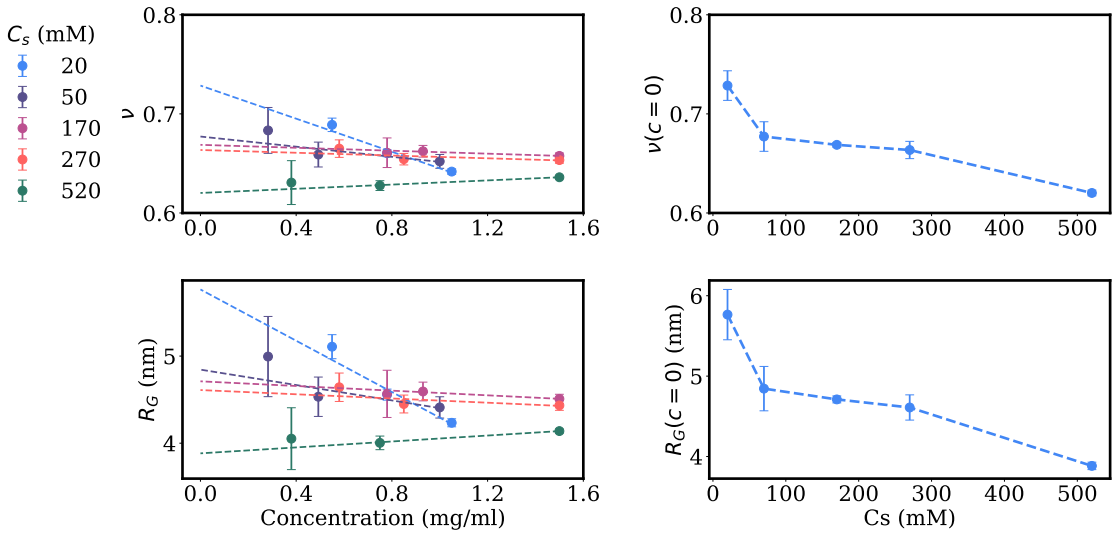

Figure 6.6: Extended Guinier analysis ν and R_G against protein concentration (C) for ΔN42. Both R_G and ν decrease with protein concentration at the lowest salinity. With increasing salinity, zero concentration $C = 0$ R_G and ν decrease monotonically. All values are found in table 3. Figure adapted from [1].

and R_0 is given by:

$$R_0^5 - R_0^3 = A\frac{134}{105}\left(\frac{3}{2\pi}\right)^{3/2} v'\sqrt{N}, \tag{6.2}$$

$$v' = v + \frac{4\pi\alpha^2 z_p^2 l_B}{\kappa^2 b^3}. \tag{6.3}$$

Here, v is the excluded volume, α is the degree of ionization, l_B is the Bjerrum length, z_p is the number of ionization groups per Kuhn segment, κ is the Debye length, and $b = 0.55$ is the Kuhn length [56]. An additional scaling factor A was added to the equation to account for discrepancies in the excluded volume. κ is directly dependant on C_s by $\sim\sqrt{C_s}$, and as such a relation between C_s and R_0 is achieved. From R_0, ν may also be extracted by:

$$R_G = \sqrt{\frac{\gamma(\gamma+1)}{2(\gamma+2\nu)(\gamma+2\nu+1)}} bN^\nu, \tag{6.4}$$

where γ is a constant given by 1.1615 ± 0.0011 [57]. Fitting the scaling parameter A to the data, the monotonic decline of ν in C_s is in good agreement with Flory's theoretical model of salted polyelectrolytes, with ν approaching that of a real chain ($\nu = 3/5$) at the high salt (Fig. 6.7).

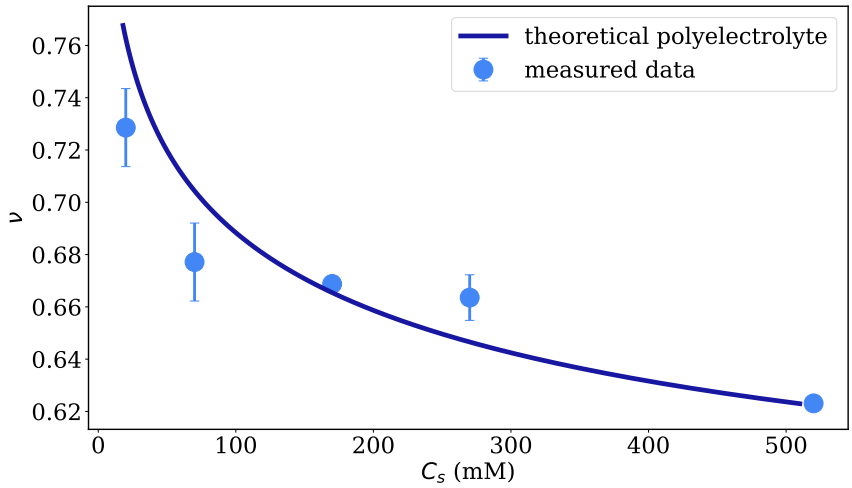

Figure 6.7: Zero concentration ν of ΔN42 against salt concentration C_s (error bars). Flory salt brush model fitting on the $\nu(c = 0)$ data is presented by the blue line(error bars). The model shows a similar contraction with salinity to the data, replicating the considerable drop in ν with salinity. All ν values are found in table 4. Figure adapted from [1]

6.2.3 Intermolecular interactions

The scaling analysis suggests strong intermolecular interactions, especially at low salinity. I quantified those intermolecular interactions by evaluating the second virial coefficient A_2 using the Zimm analysis (Section 4.3.1. In the range $C_s = 20$ mM to $C_s = 70$ mM, A_2 decreases while remaining positive, indicating a decrease in repulsive interactions. Beyond $C_s = 70$ mM, A_2 drops to zero and becomes negative, signifying attractive interactions (Fig. 6.8). These findings indicate a transition from a low-salt regime ($C_s < 70$ mM) to a high-salt regime ($C_s > 70$ mM), aligning with the scaling analysis.

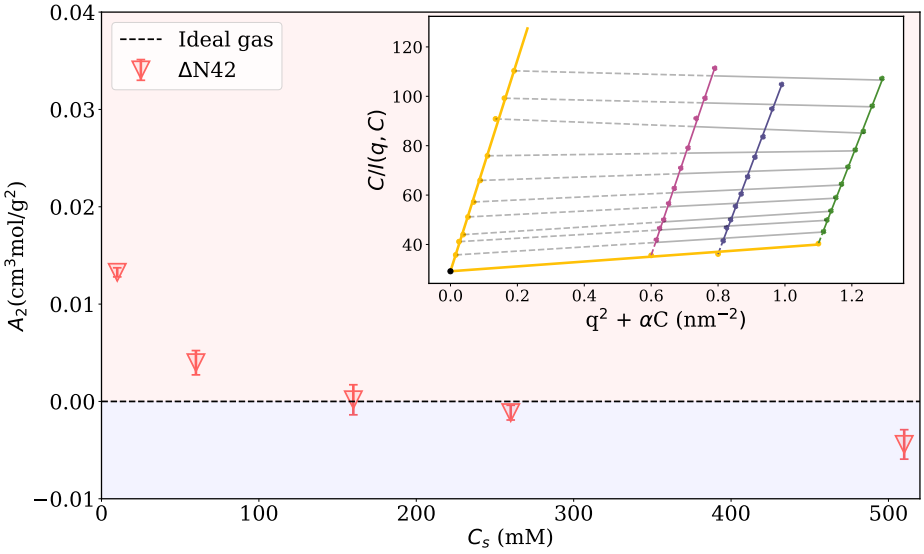

Figure 6.8: The second virial coefficient A_2 as a function of the two variants' salinity (C_s).ΔN42 intermolecular interactions transition from repulsive to attractive as Cs increases. **Inset:** The Zimm analysis used to extract A_2 from SAXS ΔN42 data measured at various protein concentrations (C). $\alpha = 0.01$ is an arbitrary constant used in this analysis. All extrapolated A_2 values can be found in table 7. Figure adapted from [1].

6.3 WT aggregates

6.3.1 Raw data

The raw data shows considerable structural deviations between the WT NFLt and ΔN42variant (Fig. 6.9A). Specifically, the scattering at WT's low q range suggests larger structures, while high q scattering retains the expected IDP lack of structure, as measured in ΔN42. Additionally, with increasing C_s, the initial low q structure grows in intensity. The kratky plot of the data again shows the same mixed-scattering behavior, with a structured protein bell curve feature at low q regime and lack of decline at higher q values (Fig. 6.9B).

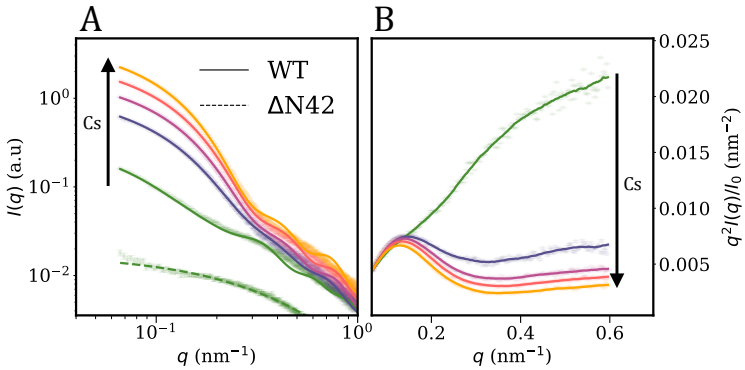

Figure 6.9: **A.** WT and ΔN42 scattering intensity $I(q)$ vs. q for the highest measured protein concentrations. Increasing salt concentration C_s results in a noticeable increase of the hump at the start of the WT scattering curve. Regardless, all C_s concentrations show a different scattering curve than the disordered ΔN42variant. **B.** Kratky plots ($q^2I(q)$) of WT in different salt concentrations C_s at the highest measured protein concentration. For the WT variants, the bell shape at low q resembles that of ordered proteins in all salt concentrations. However, an asymptotical decrease does not follow the bell-like structure, entailing that $I(q)$ at high q values do not decrease by $\sim 1/q^4$ as expected for ordered proteins. These Kratky plots then imply that WT is partially disordered. Figure adapted from [1].

6.3.2 Modeling the WT aggregation

Clear structural deviations between WT and ΔN42 are already present in initial raw data observations. These deviations present themselves in the arising peak at low q for WT, which is shown to increase with salt concentration. From equation (4.9), the molecular weight (M_w) is di-

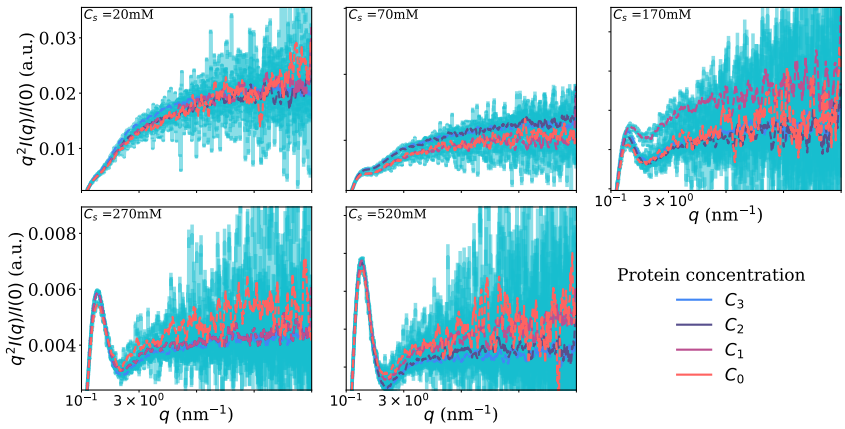

Figure 6.10: Kratky plots ($q^2I(q)$) for all WT measurements. Protein concentrations are arranged from C_4 being the highest concentration and C_1 the lowest. In all salinity cases, there is no discernible difference between the different protein concentrations. All measured concentrations can be found in table 5. Figure data adapted from [1].

rectly correlated to $I(q = 0)$ by $M_w \sim I(0)/\rho^2$. As scattering length density ρ remains unchanged for all measurements, it is then deducible that the observed increase in $I(0)$ directly correlates to measurements of samples with higher molecular mass. These results strongly indicate sample aggregation, which increases as salinity increases in the solution. Likewise, introducing strong denaturing conditions (4M GdnHCl) to the solution leads to the elimination of the aggregation, in correlation to known denaturation effects on protein unfolding and IDP expansion [27] (see section 6.3.5).

Due to the aggregation, modeling WT through a Gaussian chain model is incorrect, as it no longer represents an entirely disordered chain. Additionally, Guinier and extended Guinier analyses will only be able to estimate the size parameters of the particle as a whole rather than the individual aggregated tails.

As WT shows mixed ordered-disordered scattering, the simplest model to consider is that of polydiverse spheres, cylinders, or ellipsoids. In this case, this model may describe a system of polydiverse aggregates which fluctuate in size. Indeed, for low salt ($C_s < 170mM$), the model that fits the most was polydiverse cylinders, while for high salt ($C_s > 170mM$) it was of spheres or ellipsoids (Fig. 6.11).
These models, however, can only recapture the low q data and deviate from the high q scattering. In the small scale (*i.e.*, large q), the individual proteins are expected to behave as disordered chains. As such, an additional form factor is added to the model such that the final form factor

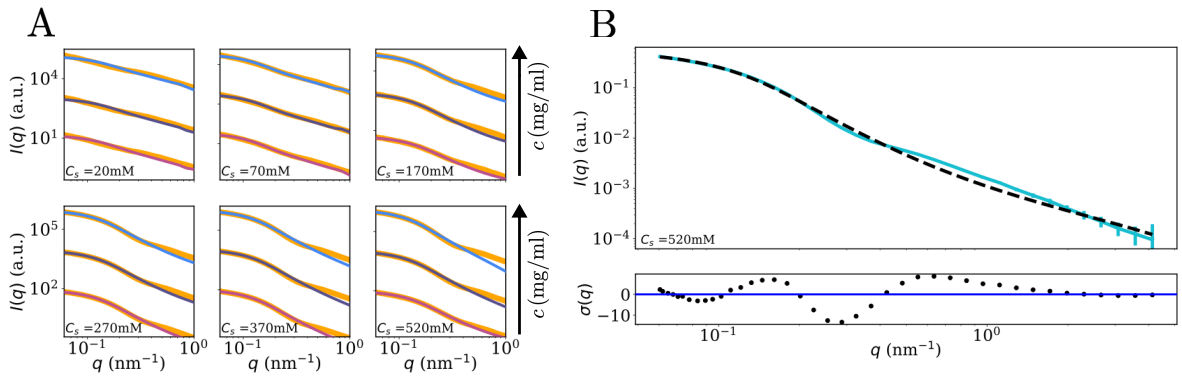

Figure 6.11: WT scattering with polydisperse form factors fittings. **A.** Polydisperse cylinder ($C_s \leq 70$ mM) and Polydisperse sphere ($C_s \geq 170$ mM) fittings for different protein concentrations. Fittings noticeably deviate from the data at high q. Here, protein concentrations were offset for clarity, with the highest curve (blue) being the highest concentration. **B.** Fitting and fitting error $\sigma_{\text{error}} = (I_{\text{fit}} - I_{\text{data}})/\sigma_{\text{data}}$ of $C_s = 520$ mM at protein concentration 2.5 mg/ml.

becomes:

$$F(q) = A \times F_{\text{pr}}(q) + B \times F_{\text{G}}(q), \tag{6.5}$$

where $F_{\text{pr}}(q)$ and $F_{\text{G}}(q)$ are the polydisperse rigid structures and the Gaussian chain form factors, respectively, and A and B are fitting parameters. With the new model, low salt measurements and some high salt measurements fit the data better than the previous (Fig. 6.12). However, the fits are still imperfect, particularly for the high salt data.

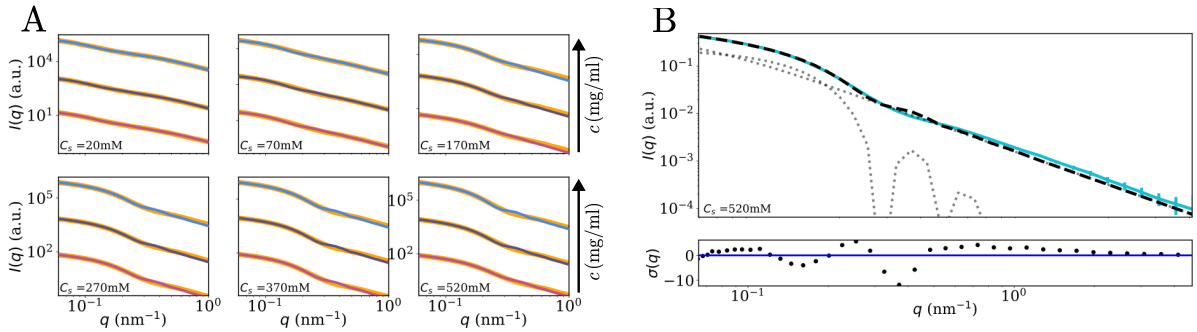

Figure 6.12: WT scattering with fitting to the mixed form factor model $F = A\,F_{\rm pr} + B\,F_{\rm G}$, where $F_{\rm pr}$ and $F_{\rm G}$ are the polydisperse rigid structure and the Gaussian form factors respectively (represented in dash lines in (B)). **A.** $F_{\rm pp}$ of polydisperse cylinder ($C_s \leq 70$ mM) and polydisperse sphere ($C_s \geq 170$ mM) fittings for different protein concentrations. Fittings agree better with the data at all q than simple polydisperse rigid structures. Here, protein concentrations were offset for clarity, with the highest curve (blue) being the highest concentration. **B.** Fitting and fitting error $\sigma_{\rm error} = (I_{\rm fit} - I_{\rm data})/\sigma_{\rm data}$ of $C_s = 520$ mM and 2.5 mg/ml protein concentration.

The next stage is to consider a more complex model, which can account for the structural relation between the disordered tails and the aggregates. As seen in ΔN42, the disordered tails are highly charged and are expected to repel each other. This property, together with the large-scale structure of spheres, could be representative of polymers that form an aggregation 'core,' from which they protrude while repelling each other, resulting in a 'spherical' structure. Theoretical models that describe this behavior are already present in literature, as they mirror the formation of block-copolymer self-assembly into decorated micelles [99, 100].

The micelle model assumes a system of polymers, of which groups of Z individual chains self-assemble into micellar structures with protruding Gaussian tails. The micelle form factor is then comprised of four terms: The core form factor term $F_{\mathrm{c}}(q)$, the Gaussian tail form factor term $F_{\mathrm{t}}(q)$, the core-tail correlation term $S_{\mathrm{ct}}(q)$, and the tail-tail correlation term $S_{\mathrm{tt}}(q)$. The core form factor may be of many different structures, such as those described in section 4.2.1. For WT, the form factors shown to fit are a cylinder for the lowest salt and a sphere of ellipsoid for the rest. The form factor of the Gaussian tails has been adjusted to account for low aggregation density [100]:

$$F_{\mathrm{RDA}}(q) = F_{\mathrm{G}}(q)/(1 + \beta F_{\mathrm{G}}(q)), \tag{6.6}$$

where $F_{\mathrm{G}}(q)$ is the Gaussian form factor given in eq. 4.13, and β is defined as $\beta = 2A_2Mc$. Here, A_2 is the second virial coefficient (obtainable from the Zimm analysis), M is the molecular mass of the individual chains, and c is their concentration. In the case of a sphere, the cross-correlation term of the tails is given by:

$$S_{\mathrm{tt}} = \Psi^2(q) \left(\frac{\sin(q(R + R_{\mathrm{G}}))}{q(R + R_{\mathrm{G}})} \right)^2, \tag{6.7}$$

where $\Psi(q)$ is as given as: $\Psi(q) = (1 - e^{-x})/x^2$. The cross-correlation term between the core and the tails is given by:

$$S_{\mathrm{ct}} = \Phi(q)\Psi(q) \left(\frac{\sin(q(R + R_{\mathrm{G}}))}{q(R + R_{\mathrm{G}})} \right)^2, \tag{6.8}$$

where $\Phi(q)$ is from table 2.

For the ellipsoid and the cylindrical cases, the cross-correlation terms now require an integration. For the ellipsoid:

$$S_{\mathrm{tt}} = \Psi^2(q) \int d\alpha \left(\frac{\sin(q(r(\alpha) + R_{\mathrm{G}}))}{q(r(\alpha) + R_{\mathrm{G}})} \right)^2, \tag{6.9}$$

$$S_{\mathrm{ct}} = \Psi(q) \int d\alpha \Phi(q, r(R, \epsilon, \alpha)) \left(\frac{\sin(q(r(R, \epsilon, \alpha) + R_{\mathrm{G}}))}{q(r(R, \epsilon, \alpha) + R_{\mathrm{G}})} \right)^2, \tag{6.10}$$

where $r(R, \epsilon, \alpha)$ and $\Phi(q)$ are given for an ellipsoid in table 2.

Similarly, for the cylinder the form-factor contributions are:

$$S_{\mathrm{tt}} = \Psi^2(q) \int d\alpha \left(\Xi(q, R + R_{\mathrm{G}}, L + 2R_{\mathrm{G}}, \alpha) \right)^2, \tag{6.11}$$

$$S_{\mathrm{ct}} = \Psi(q) \int d\alpha \Phi_C(q, R, L, \alpha) \left(\Xi(q, R + R_{\mathrm{G}}, L + 2R_{\mathrm{G}}, \alpha)\right)^2, \tag{6.12}$$

where $\Phi_C(q, R, L, \alpha)$ is the cylindrical form factor given in table 2, and Ξ is given by:

$$\Xi(q, R, L, \alpha) = \frac{R2B_1(qR\sin(\alpha))\cos(qL\cos(\alpha)/2)}{(R+L)qR\sin(\alpha)} + \frac{LB_0(qR\sin(\alpha))\sin(qL\cos(\alpha)/2)}{(R+L)qL\cos(\alpha)/2}. \tag{6.13}$$

Finally, the total form factor is written as:

$$F(q) = \frac{1}{(\beta_c + \beta_t)^2}\left(\beta_c^2 F_c(q) + \frac{\beta_t^2}{Z} F_{\mathrm{RDA}}(q) + \frac{(Z - F_{\mathrm{RDA}}(0))\beta_t^2}{Z} S_{\mathrm{tt}}(q) + 2\beta_c\beta_t S_{\mathrm{ct}}(q)\right) \tag{6.14}$$

For consistent terminology, R_G is identified as half of the brush height of the tails and is renamed to $h/2$. In the low salt case, the model with a cylindrical core was shown to have a good fit (Figs. 6.13, 6.15). In the rest of the cases, both spherical and ellipsoidal cores had good fits (Figs. 6.13, 6.14). As the resulting structure parameters were identical, the spherical core was chosen for its simplicity over the ellipsoidal core. Regrettably, measurements at $C_s = 70$ mM could not fit the model. Section 6.3.5 will further discuss these measurements. All fitted parameters are in table 5.

As the aggregation number for all measurements was relatively low (less than 10), in the context of the measurements this model is described as a 'star-like' polymer brush.

C_s (mM)	C (mg/ml)	$h/2$ (nm)	ν	Z	n	R (nm)	L (nm)	V (nm^3)	β_t (10^3nm)	β_c (10^3nm)
20	2.68	9.16±0.15	0.786±0.0033	1.60±0.03	10.11±0.711	0.89±0.028	1.19±0.09	2.26±0.027	0.227	3.826
20	1.8	8.20±0.13	0.758±0.0032	1.83±0.04	8.52±0.059	0.89	1.19	2.26	0.195	3.558
20	1	8.57±0.17	0.768±0.0041	1.87±0.05	8.12±0.060	0.89	1.19	2.26	0.195	3.558
20	0.5	8.27±0.16	0.759±0.0040	1.91±0.05	7.78±0.057	0.89	1.19	2.26	0.195	3.558
170	1.3	9.96±0.03	0.796±0.0006	3.34±0.02	2.52±0.009	0.66±0.005	X	1.18±0.007	0.039	4.014
170	0.73	10.11±0.05	0.799±0.0009	3.27±0.03	2.32±0.014	0.63±0.007	X	1.06±0.010	0.039	4.014
170	0.57	10.37±0.08	0.806±0.0015	2.13±0.02	3.10±0.037	0.60±0.006	X	0.93±0.008	0.074	3.98
170	0.24	10.78±0.10	0.815±0.0019	2.71±0.05	3.23±0.025	0.66±0.011	X	1.23±0.017	0.074	3.98
270	2	9.40±0.02	0.781±0.0003	5.37±0.02	1.89±0.007	0.70±0.004	X	1.42±0.008	0.018	4.03
270	1.5	9.43±0.02	0.782±0.0004	5.02±0.02	1.93±0.009	0.69±0.005	X	1.36±0.009	0.018	4.03
270	0.69	9.85±0.04	0.793±0.0009	4.05±0.04	2.70±0.016	0.71±0.009	X	1.53±0.016	0.039	4.014
370	2.3	9.36±0.01	0.780±0.0003	6.68±0.03	1.48±0.008	0.69±0.005	X	1.38±0.009	0.018	4.036
370	1.5	9.44±0.02	0.782±0.0003	6.43±0.03	1.65±0.009	0.71±0.006	X	1.49±0.010	0.018	4.036
370	0.96	9.53±0.02	0.785±0.0004	5.85±0.03	2.08±0.010	0.74±0.006	X	1.70±0.012	0.039	4.014
370	0.6	9.81±0.03	0.792±0.0007	5.33±0.05	2.13±0.017	0.72±0.010	X	1.59±0.019	0.039	4.014
520	2.5	9.57±0.01	0.786±0.0003	8.15±0.04	1.82±0.008	0.79±0.005	X	2.07±0.012	0.018	4.036
520	1.19	9.28±0.02	0.778±0.0004	6.66±0.04	1.57±0.010	0.70±0.006	X	1.47±0.012	0.018	4.036
520	0.45	9.82±0.03	0.792±0.0005	6.36±0.06	1.94±0.016	0.74±0.010	X	1.72±0.020	0.039	4.014

Table 5: **WT spherical and cylindrical fitting analysis data**. Analysis parameters (brush height (h), scaling exponent (ν), aggregation number (Z), core peptide length (n), core radius (R), cylindrical core length (L), core volume (V), tail scattering length (β_t) and core scattering length (β_c)) obtained for different salt concentrations (C_s) and protein concentrations (C). Cylinder length L values are only relevant to $C_s = 20$mM where a cylindrical core fit was used. For the cylindrical core, the same values of L and R were used for all concentrations to alleviate fitting errors (see Methods). Table adopted from [1].

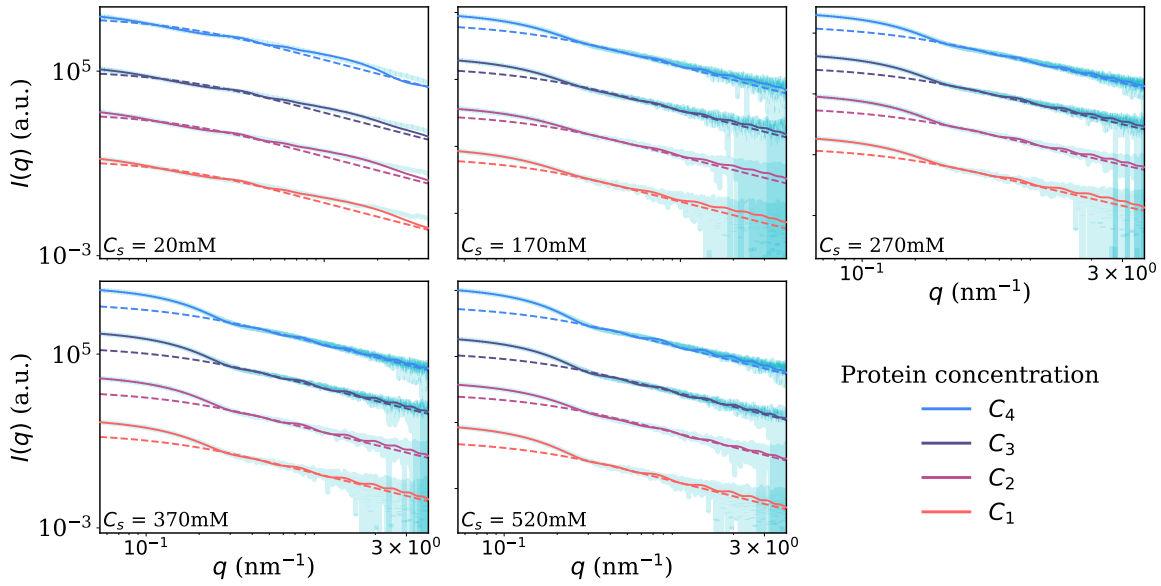

Figure 6.13: Micelle form factor fitting for all salinity concentrations (C_s). The $C_s = 20$ mM data fit is to a cylindrical core, while the rest with a spherical core. Dashed lines represent the Gaussian form factor of the structure tails. Protein concentrations were offset for clarity, with the highest (blue, C_4) being the highest concentration. All measured concentrations can be found in table 5. Figure adapted from [1].

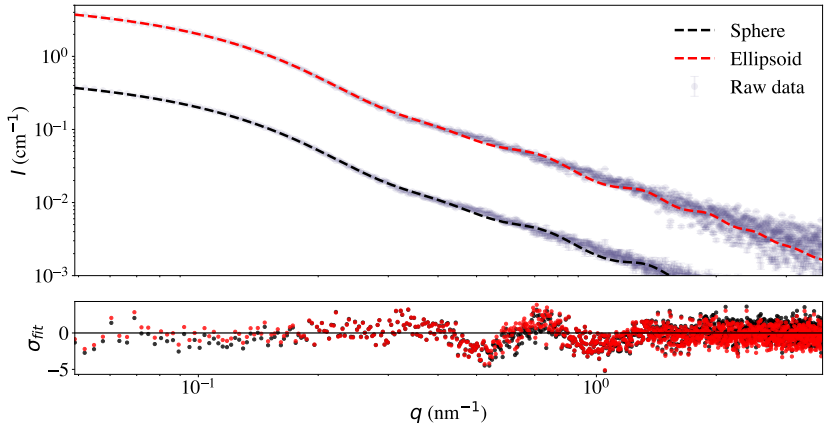

Figure 6.14: SAXS measurements of WT and its fitting to different form factors. Both form factors are of the same model but use a different core: Spherical or Ellipsoidal. Spherical core fitting yields a core radius of $R = 0.66 \pm 0.016$ nm, and the ellipsoidal core yields a core radius of $R = 1.335 \pm 0.23$ nm and a secondary radius of ϵR where $\epsilon = 0.153 \pm 0.08$. Both fittings yield close values of aggregation number Z (3.046 ± 0.04 for spherical and 3.562 ± 0.07 for ellipsoidal) and tail height $h/2$ (9.838 ± 0.04 nm for spherical and 9.584 ± 0.12 for ellipsoidal). Below: Fitting error $\sigma_{fit} = (Y_{fit} - Y_{data})/\sigma_{data}$. Both curves show similar error profiles. The spherical model proved best to describe the model due to its simplicity. Displayed data: WT in 20 mM Tris pH=8.0, and 170 mM NaCl at a concentration of 1.3 mg/ml. Figure adapted from [1].

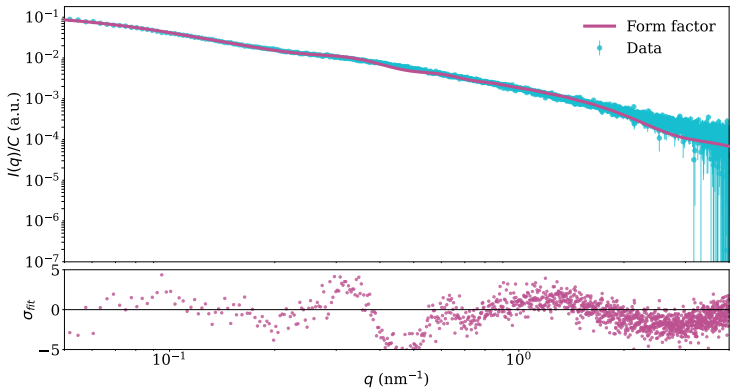

Figure 6.15: WT SAXS measurement with cylindrical fitting. Measurements at the highest concentration of $C = 2.68$ mg/ml, in a 20 mM Tris buffer at pH=8.0. To alleviate fitting inconsistencies, consequent fittings of measurements with lower protein concentrations in the same buffer were done using the obtained core parameters: Core radius $R = 0.89 \pm 0.03$ nm, and core length $L = 1.19 \pm 0.09$ nm. Figure adapted from [1].

6.3.3 WT scaling

From the fitted scattering data, the brush height $h = 2R_G$, the aggregation number Z, and the core characteristics (e.g., spherical core radius R, cylinder radius R, and length L) can be extracted. The scaling exponent ν may also be extracted using Eq. 6.4 with R_G and $b = 0.55$ nm [56]. As with ΔN42case, all parameters are fitted to their zero concentration value by a linear fit over the entire protein concentration data (Fig. 6.16, table 6).

C_s (mM)	$h/2$ (nm)	ν	Z	n	R (nm)	L (nm)	V (nm^3)
20	8.01±0.46	0.751±0.013	2.03±0.08	7.08±0.39	0.89	1.19	2.26
170	10.60±0.33	0.811±0.008	2.83±0.31	3.17±0.55	0.64±0.03	X	1.10±0.16
270	9.84±0.21	0.793±0.006	3.52±0.22	3.06±0.30	0.70±0.03	X	1.49±0.20
370	9.73±0.12	0.790±0.003	5.19±0.29	2.39±0.12	0.76±0.02	X	1.67±0.01
520	9.47±0.44	0.783±0.011	5.67±0.35	1.82±0.29	0.68±0.06	X	1.28±0.38

Table 6: **Zero concentration WT spherical and cylindrical fitting analysis data**. Analysis parametres (brush height (h), scaling exponent (ν), aggregation number (Z), core peptide length (n), core radius (R), cylindrical core length (L) and core volume (V)) were extrapolated to zero protein concentration at various salt concentrations (C_s). Cylinder length L values are only relevant to $C_s = 20$mM where a cylindrical core was used. Table adapted from [1].

For the low-salt condition of $C_s = 20$ mM, the aggregation number of WT is of a dimer $Z \sim 2$, and the core's shape is that of a cylinder ($R = 0.89$ nm, $L = 1.19$ nm) (Fig. 6.16). Comparing zero concentration R_G of ΔN42 with the equivalent WT length $h/2$ shows a considerable increase, going from 5.76 ± 0.31 nm to 8.01 ± 0.46 nm.

Advancing to higher salinities, Z increases (illustrated in Fig. 6.17A), with a maximum zero concentration value of 6 at $C_s = 520$ mM (Fig. 6.17B). Interestingly, as salinity increases, Z shows heightened sensitivity to changes in protein concentration. This change is likely due to the fading electrostatic repulsion, allowing larger aggregates to form at higher protein concentrations.

Given the small core volume ($V \sim 1 - 2$ nm^3, Fig. 6.17C), it is crucial to evaluate the 'grafting' distance between neighboring chains ρ on the core surface ($S = 4\phi R^2 = Z\rho^2$). In all cases, $h/\rho > 1$, signifying a 'brush regime' where neighboring chains repel and expand each other [101] (Fig. 6.17D). As C_s varies from 20 mM to 170 mM, $h/2$ increases to 10.60 ± 0.33 nm (Fig. 6.17E). This increase likely results from a rise in tail density, evidenced by h/ρ. Beyond 170 mM, $h/2$ slightly declines, reaching 9.47 ± 0.44. Notably, all resulting $h/2$ values exceeded R_G of ΔN42.

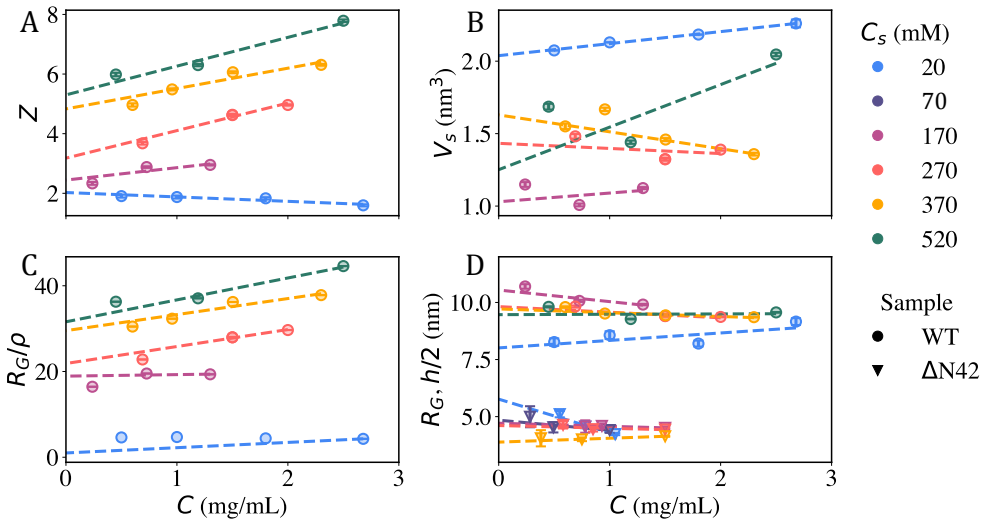

Figure 6.16: Structural parameters for WT (circles) and ΔN42 (triangles) variants extracted from fitting the SAXS data. Dashed lines demonstrate the linear fitting of the data used to obtained the zero concentration extrapolations. **A.** Aggregation number (Z) dependency on protein concentration (C) increases with increasing salt. **B.** Core volume V_s against protein concentration (C). In $C_s = 20$ mM, the V_s values are constant due to fitting constraints (see Methods). **C.** In all cases, the tail heights (h) are larger than the corresponding grafting length (ρ), indicative of a brush regime. **D.** The structurally intrinsically disordered ΔN42 variant compacts with higher C_s values and remains more compacted from the projected tails for the WT variant. For the ΔN42 variant R_G drastically changes as a function of the protein concentration (C). All values can be found in tables 3 and 5 for ΔN42 and WT respectively. Figure adapted from [1]

Comparisons between ΔN42 and WT are better made by the zero concentration extrapolation to ν. For WT, ν was calculated by the R_G equivalent length scale of $h/2$. Starting again with the lowest salinity, WT resulted in a ν of 0.751 ± 0.013 compared to 0.729 ± 0.015 for ΔN42. This difference in ν is quite minimal due to the already strong electrostatic interactions in both variants. Increasing the salinity further, however, the value of ν for WT further increases up to 0.811 ± 0.008, from which it slowly dropped up until $\nu = 0.783 \pm 0.011$ at the highest salinity of C_s. This expansion lies in grave contrast to ΔN42, which was shown to decrease monotonically for all salinities per the polyelectrolyte prediction.

Recalling Pincus' regime of salted polyelectrolyte brushes at the high salt limit ([80], Sec. 2), the brush counterion osmotic pressure becomes:

$$\Pi \sim C^2 k_B T / C_s, \tag{6.15}$$

where C is protein concentration. Assuming the brushes behave as nonlinear osmotic brushes

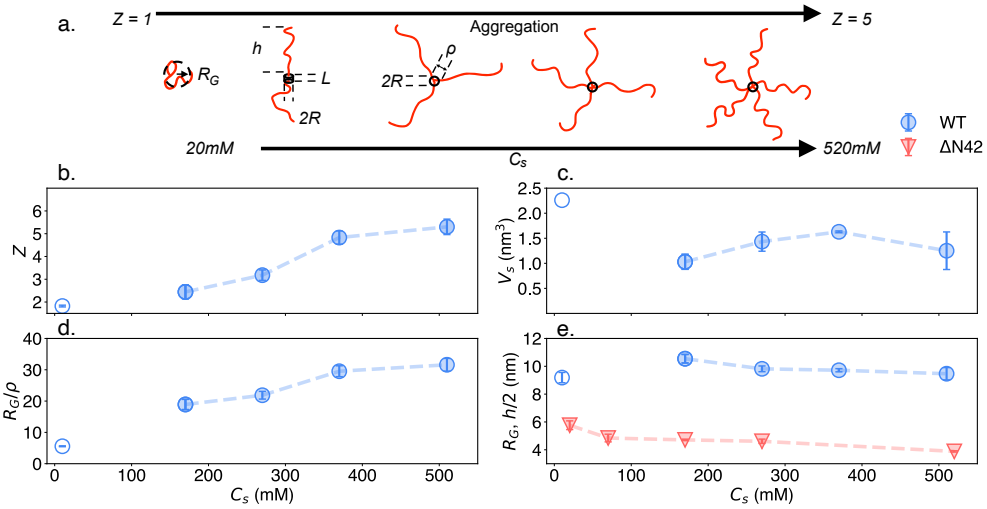

Figure 6.17: **A.** Schematic of the system's structure variation with salinity (C_s). While ΔN42 remains disordered and segregated, the WT variant aggregates to a star–like polymer with a higher aggregation number at higher C_s. **B–E.** Structural parameters for WT (blue symbols) and ΔN42 (red symbols) variants extracted from fitting the SAXS data. Full and hollow circles represent the spherical and cylindrical core fitted parameters, respectively. **D.** In all cases, the brush heights (h) are larger than the corresponding grafting length (ρ), indicative of a brush regime. **E.** The structurally intrinsically disordered ΔN42 variant compacts with higher C_s values and remains more compacted from the projected brushes for the WT variant. All values are the extrapolated 'zero concentration' fitting parameters (see Fig. 6.16), which can be found in tables 4 and 6 for ΔN42 and WT respectively. Figure adapted from [1].

[101], the brush height can be derived as [102]:

$$h \sim ANb \left(\frac{\rho_a f^2}{bC_s} \right)^{1/3}, \tag{6.16}$$

where ρ_a is the grafting density of the brushes, and f is the degree of charge. Here, ρ is calculated by $Z/4\pi R^2$. As Z and R vary in salt concentration, a linear relation between ρ_a and C_s must be estimated to obtain a continuous $h(C_s)$ to compare with the data. As ρ_a roughly rises linearly with C_s (Fig. 6.18B), this relation was obtained by a linear fitting over the data. Using the linearized $\rho_a(C_s)$, ν was then calculated using Eq. 6.4 to obtain the relation of $\nu(C_s)$.

With a fitting of $A = 0.38$, the modeled ν agrees well with the measured expansion at high salinity (Fig. 6.18A). As with ΔN42, the shift to $C_s = 170$ results in a phase shift where the brushes transition to a "salted" state.

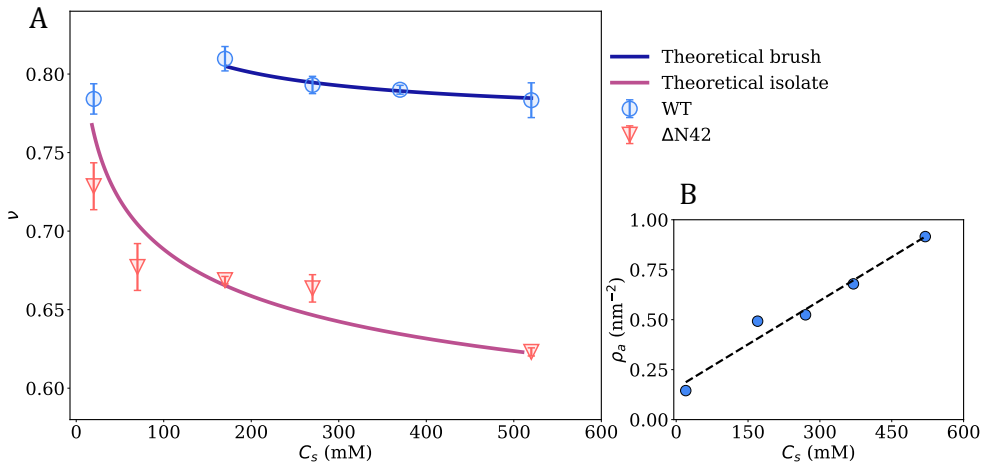

Figure 6.18: **A.** Flory exponent (ν) of WT tails and ΔN42 variants showing extended disordered scaling. The red line refers to the theoretical brush model [102], and the blue line refers to the theoretical polyelectrolyte [64]. ΔN42 shows a decrease in the protein extension due to the decline in intermolecular electrostatic repulsion (see also Fig. 6.8). WT shows an increase in the extension when shifting from a dimer to a trimer, followed by a slight decline with a further increase in salinity. **B.** Grafting density ρ_a of the WT brushes against salinity (C_s). The increase of ρ in C_s is linearly fitted (black dashed line) to obtain the continuous function of $\rho(C_s)$ for the polyelectrolyte brush model. Figure adapted from [1].

Given the results, we now take a greater interest in examining the physicality of this model. The core radius of WT is typically less than 1 nm, while brush height results in values greater than 8 nm. The "aggregated" core is then expected to have fewer residues than the radiating brushes. This number can be estimated by a simple volumetric approach, which assumes that inside a sphere volume V, a maximum of n amino acids with an average volume of $\langle\phi_{aa}\rangle$ given by tabulated values [103]. Following, n will then be given by:

$$nZ = V/\langle\phi_{aa}\rangle \tag{6.17}$$

We find that n has rather small values, with a maximum value of 7 residues at the lowest salinity (Fig. 6.19). This value drops from 7 to 3 residues as salt concentration increases to $C_s = 170$ mM and continues to drop until only two residues at $C_s = 520$ mM.

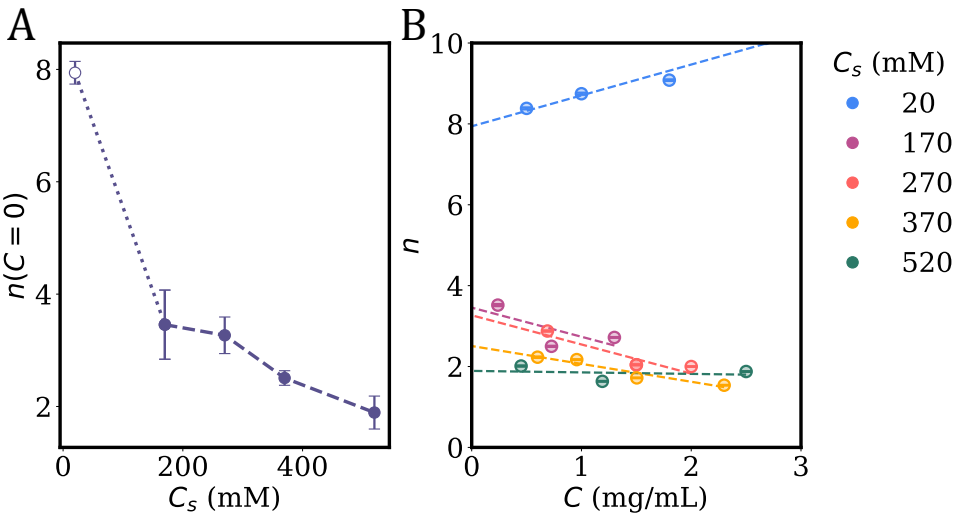

Figure 6.19: **A.** Zero protein concentration aggregation core residue number n in salt concentration (C_s). n is shown to constantly decrease with C_s. **B.** n against protein concentration C for all measured salinities. For $C_s = 20$ mM, n increase with protein concentration (C), while the rest of the salinities it show a lower n which decreases with C.

6.3.4 Intermolecular interactions

As with ΔN42, A_2 for WT may also be evaluated to gauge the variant's intermolecular interactions. However, as WT describes an ensemble of aggregates and isolated IDPs, A_2 describes the intermolecular interactions of the aggregates between themselves and the isolated IDPs. Increasing salt, A_2 changes from a nearly neutral state to mildly attractive ($A_2 < 0$). These findings correlate with the increased sensitivity of Z with protein concentration C (Fig .6.20). Interestingly, compared to the ΔN42 variant, WT does not show the same repulsive interactions at low salinity, attesting to low interactivity between different aggregates in the solution.

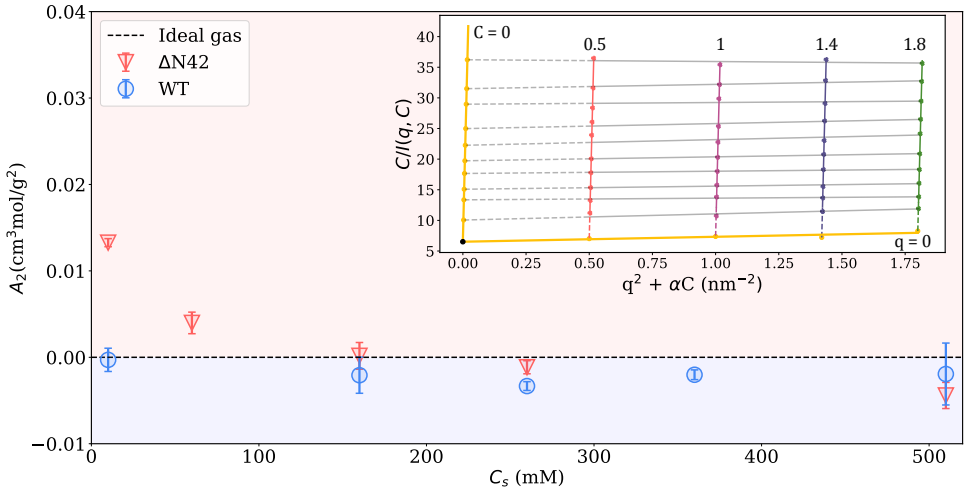

Figure 6.20: The second virial coefficient A_2 as a function salinity (c_s) of the two variants. ΔN42 intermolecular interactions transition from repulsive to attractive as C_s increases. WT intermolecular interactions transition from near-neutral to weakly attractive. **Inset:** The Zimm analysis used to extract A_2 from SAXS WT data measured at various protein concentrations (C). All concentrations are in mg/mL. $\alpha = 0.01$ is an arbitrary constant used in this analysis. Analysis presented is of WT at 20mM Tris and pH 8.0. All extrapolated A_2 values can be found in table 7.

C_s (mM)	A_2^{WT} (cm^{3}mol/g^2 × 10^3)	$A_2^{\Delta N42}$ (cm^{3}mol/g^2 × 10^3)
20	-0.295±1.346	13.264±0.466
70	X	3.978±1.248
170	-2.072±2.091	0.169±1.544
270	-3.328±0.508	-1.152±0.756
370	-2.020±0.563	X
520	-1.933±3.582	-4.417±1.514

Table 7: Second virial coefficient A_2 values for both variants in salt concentration C_s.

6.3.5 Additional measurements

In addition to the measurements presented in the paper, an additional salt concentration of $C_s = 70$ mM and a high denaturation measurement (GdnHCl 4M) were performed for WT.

As mentioned in my discussion on the aggregation, the $C_s = 70$ mM measurements were the only ones that deviated from the micelle model. These deviations were observed with all three core shapes (cylinder, sphere, ellipse), with irregular standard deviations for the fitting parameters. Moreover, increasing the fitting algorithm iterations and introducing stringent boundary conditions on the parameters did not alleviate these problems. As only one set of $C_s = 70$ mM was measured, it is hard to tell if these deviations are due to a measurement error. If that is not the case, however, I can assume these to be due to a high degree of non-uniformity in the aggregation. This postulation arises from $C_s = 70$ mM being a middle point between the low and high salt regimes, where its behavior remains unclear.

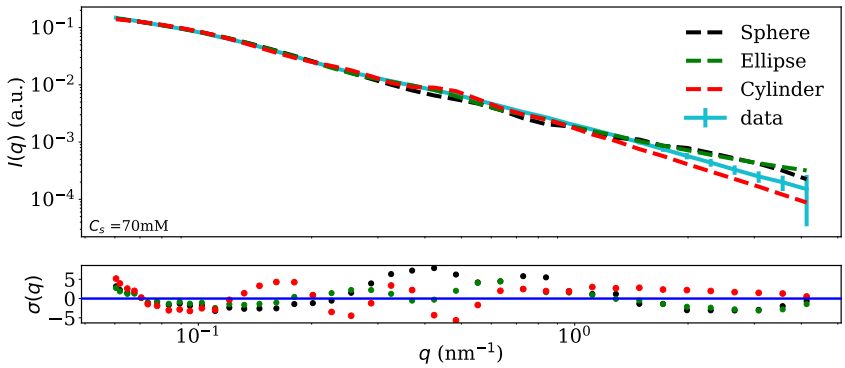

Figure 6.21: Salt concentration $C_s = 70$ mM with three model fittings. All three show deviation from the data at high q and with incalculable standard deviation by the fitting model (See methods). Below: Fitting error $\sigma = (Y_{fit} - Y_{data})/\sigma_{data}$.

In addition to measurements in NaCl, WT was also measured in strong denaturation conditions (4M GdnHCl). The variant at all protein concentrations is no longer aggregated in this condition. Additionally, extended Guinier analysis reveals ν values within the proximity of a real chain ($\nu = 3/5$), with a zero concentration value of 0.6. This result aligns with the known behavior of denatured unfolding protein, which, alongside IDPs, was previously shown to arrive at the real-chain result at sufficiently high denaturation.

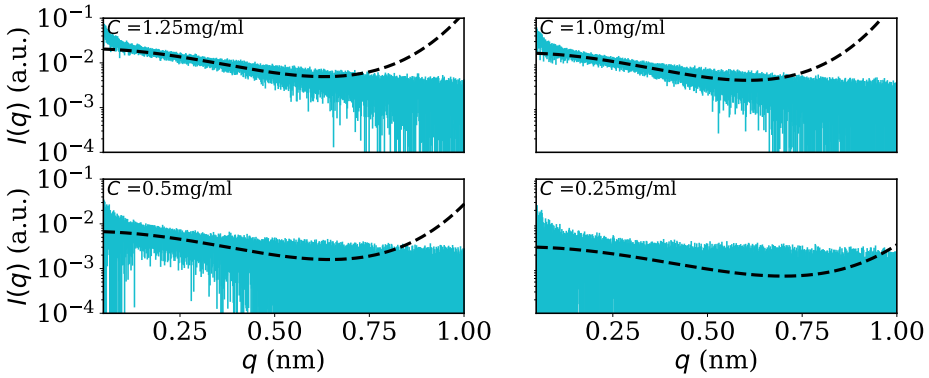

Figure 6.22: WT measurements at 4M GdnHCl in four protein concentrations. All measurements are fitted to the extended Guinier function to extract ν (Fig. 6.23).

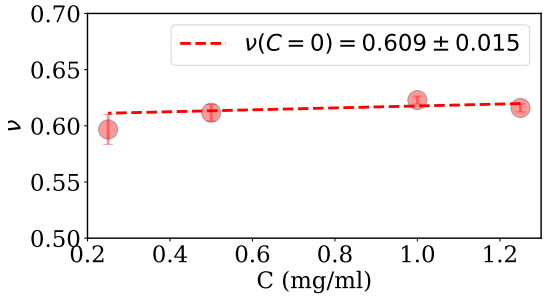

Figure 6.23: WT 4M GdnHCl measurements ν as obtained by an extended Guinier analysis (Fig. 6.22).

6.4 Comparison to segmented NFLt

In a previous study, the NFLt chain was divided into non-overlapping segments of roughly 20 residues each [26]. These segments were measured in the context of the whole NFLt using FRET (P segments) and as separate peptides using SAXS (S segments). It was shown that both P and S segments varied in scaling, with the scaling exponent ν rising with their net charge per residue (NCPR).

To better compare the segments' ν against ΔN42's with increasing salinity, I averaged over the segments found within the region of ΔN42 of NFLt. As for comparisons with WT, although the P segments were measured within the isolated NFLt chain, they likely did not aggregate. This is due to protein concentration being much lower in the FRET measurements. As such, comparing the P segments to the non-aggregating ΔN42 variant is more adequate.

Starting with the S segments, unlike ΔN42 , these did not show the same pronounced drop in expansion at $C_s < 170$ mM, measuring at lower values. As salinity increased, however, when the segments and ΔN42 arrived at $C_s = 520$, they converged at the same ν value of ~ 0.62. P segments, likewise, did not show much response to salinity, although with much higher ν values compared to both S segments and ΔN42. This discrepancy in the ν range is likely due to differences in the SAXS and FRET measurements, as explained in the paper [26].

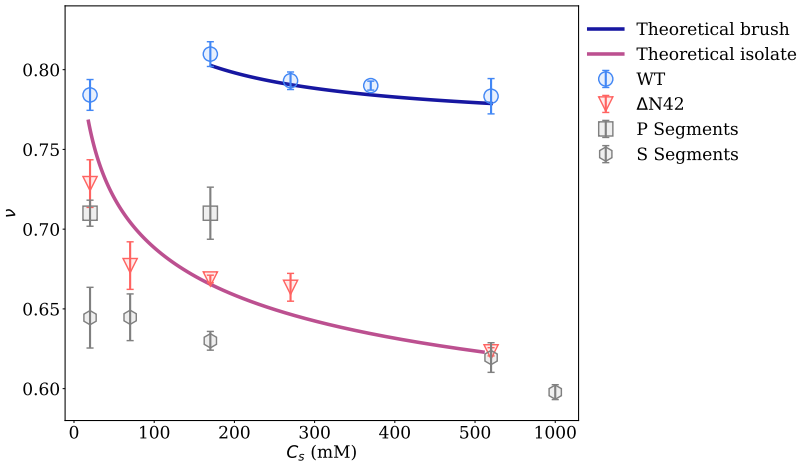

Figure 6.24: Comparisons of ν values as obtained for ΔN42 and WT to the S and P segments of NFLt from [26]. S segments show a weak reactivity to salinity at the low salt regime ($C_s < 170$ mM), in contrast to ΔN42. Comparatively, the P segments are also mostly unperturbed. At higher salinities, S segments slightly drop and converge with ΔN42 at $C_s = 520$ mM at $\nu \sim 0.62$, near the theoretical real chain value of $\nu = 0.6$. Figure adapted from [1] with additional data from [26].

6.5 Dynamics

6.5.1 Magnetic tweezers setup

Previously, magnetic tweezers (MT) experiments have shown that the NFLt IDP exhibits glassy dynamics outside of equilibrium due to its heterogeneous nature [29]. These dynamics were shown to be due to independently-relaxing modes within the NFLt. Here, using SAXS, we know that the NFLt sequence heterogeneity immensely influences its equilibrium ensemble structure. Thus, I wish to investigate whether these independently-relaxing modes could exist within specific sub-domains of the IDP.

I again employ the two variants used for the SAXS experiments, WT and ΔN42, to possibly isolate the sub-domain of the NFLt responsible for the glassy behavior. In the MT experiments, the two variants were polymerized and attached to the substrate from one end and to a magnetic bead from the other. The tethered IDPs were then subjected to two- and three-step force experiments (Sec. 6.5.2 and Sec. 6.5.3) to measure their dynamical response to tension. All measurements were conducted in low-salt conditions (10 mM MES pH 7.0 buffer).

6.5.2 Two-step experiment

In the two-step experiments, chain height $L(t)$ was measured as the applied tension on the chain was switched from a high force (f_1) to a low force (f_2). After the switch, $L(t)$ was measured for 120 seconds to observe its relaxation dynamics (Fig. 6.25A). Both variants relax logarithmically, with most measurements fitting well with the model $L(t) - L(t_0) = b \log(t/t_0)$, where $t_0 = 1$ s (Figs. 6.25B, 6.26, 6.27). To account for polydispersity in force and tether length, the relaxation constant b is normalized as $\bar{b} = bf_1/Nk_BT\alpha_0(f_2)$.

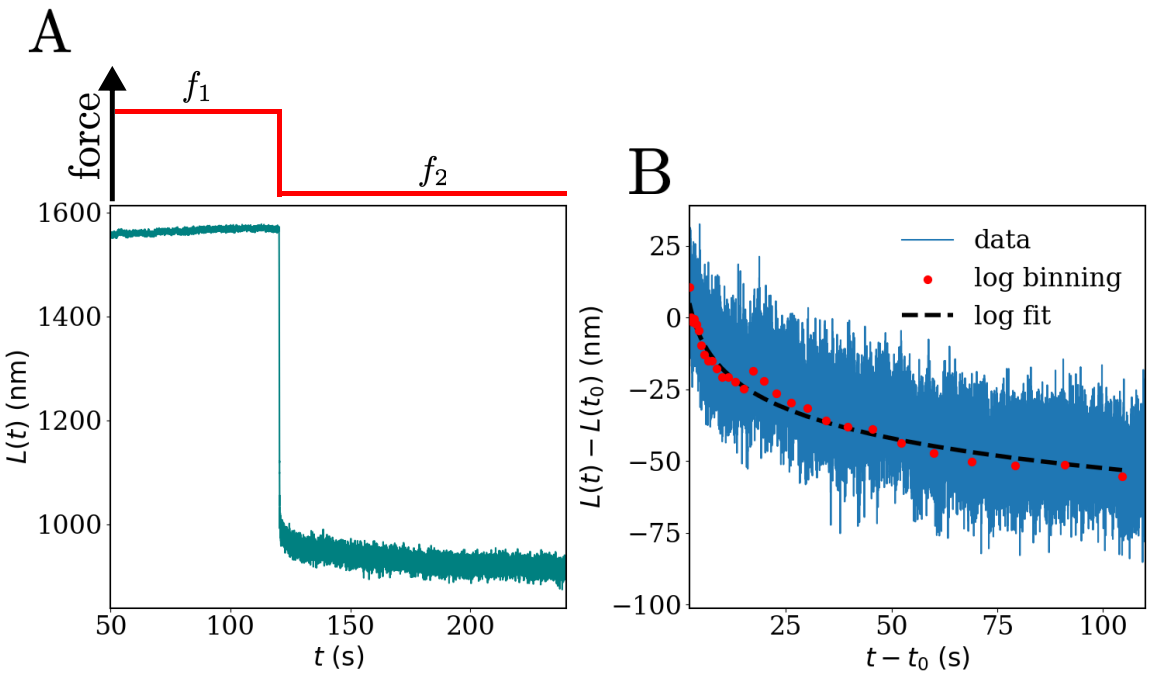

Figure 6.25: **A.** Tether height $L(t)$ against measurement time. The switch between high force f_1 to low force f_2 is noticeable in the sudden drop in $L(t)$ at $t = 120$ s. **B.** Logarithmic fitting of the relaxation curve $L(t) - L(t_0)$, where $t_0 = 1$ s. The data is initially binned into 30 logarithmic bins, which are then fitted to $L(t) - L(t_0) = b \log(t/t_0)$.

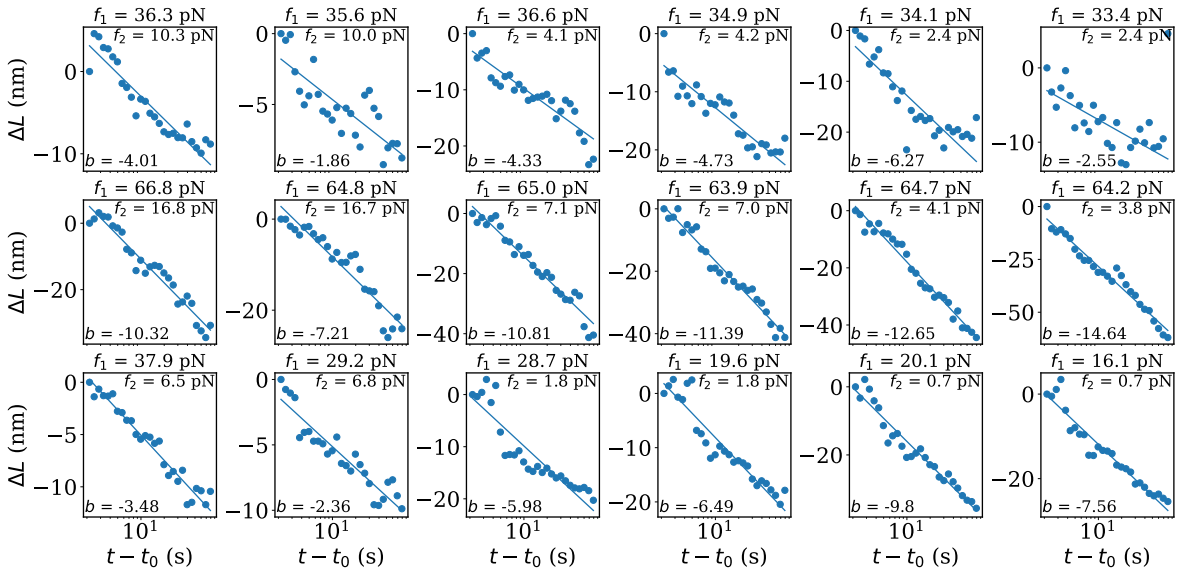

Figure 6.26: Logarithmic binning and fitting of all WT measurements.

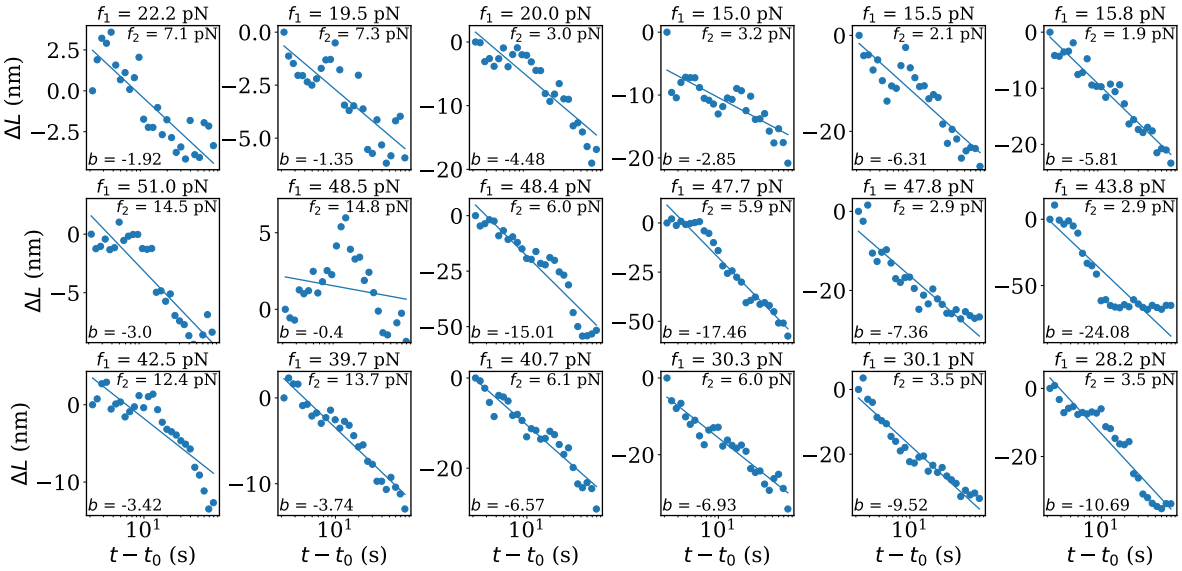

Figure 6.27: Logarithmic binning and fitting of all ΔN42 measurements. Compared to WT (Fig. 6.26), less ΔN42 could fit into the logarithmic curve.

As previously shown for NFLt [29], I use the logarithmic relaxation model for the two variant, where:

$$\bar{b} = \frac{1}{\rho}\left(1 - \frac{1}{\bar{f}}\right). \tag{6.18}$$

Here, $\bar{f} = f_2/f_1$, and $\rho = \delta x/L_c$ is the ratio between the node and contour lengths.

At first glance, the data fits well with the model, retaining the $1/\bar{f}$ relationship, with both variants resembling each other (Fig. 6.28A). However, the resulting ρ reveals unrealistic values of 0.99 and 1.7 for WT and ΔN42, respectively, compared to 0.1 observed with NFLt as discussed in the section 5.4 and [29]. To resolve the discrepancy, I filtered out measurements that did not fit well the logarithmic decay. Following, ρ of ΔN42 changes noticeably, becoming much closer to the WT variant (Fig. 6.28B). Regardless, the filtered data sets reinforce the need for additional measurements, whereas, in a larger data set, these non-conforming measurements would have a minor impact on the curve.

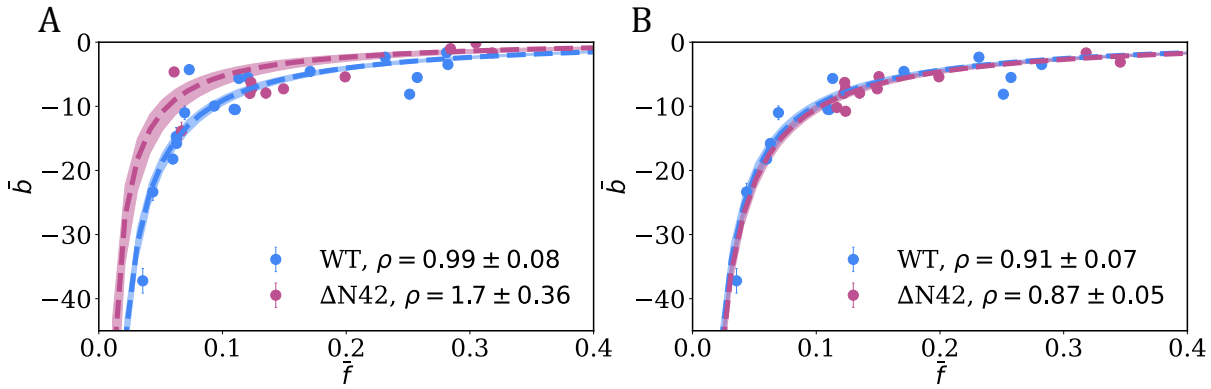

Figure 6.28: **A.** Normalized logarithmic relaxation constant ($\bar{b}$) against the ratio of low force to high force $\bar{f} = f_2/f_1$. WT shows relatively good fitting with the model, while ΔN42 lacks sufficient data points in the low $\bar{f}$ range to properly fit with the model. **B.** $\bar{b}$ against $\bar{f}$ with the poor-fitting data of Figs. 6.26 and 6.27 removed. This change results in a better fit for the ΔN42 data, although the lack of low $\bar{f}$ measurements still hurt the validity of the results.

A comparison to the SAXS data can be made with the equilibrium behavior of $L(t)$, from which ν of the variants can be extracted using the Pincus blob theory (Section 5.3). Equilibrium tether height L_{eq} is normalized by the contour length L_{c}, as calculated from the WLC model [96]. I used this normalization to account for chains of different polymerization numbers. The initial presentation of the data against the normalized WLC curve shows a good fit, most clearly with the high force data (Fig. 6.29A). The high measurement noise at low force is likely due to samples that have yet to reach equilibrium within the measured time frame.

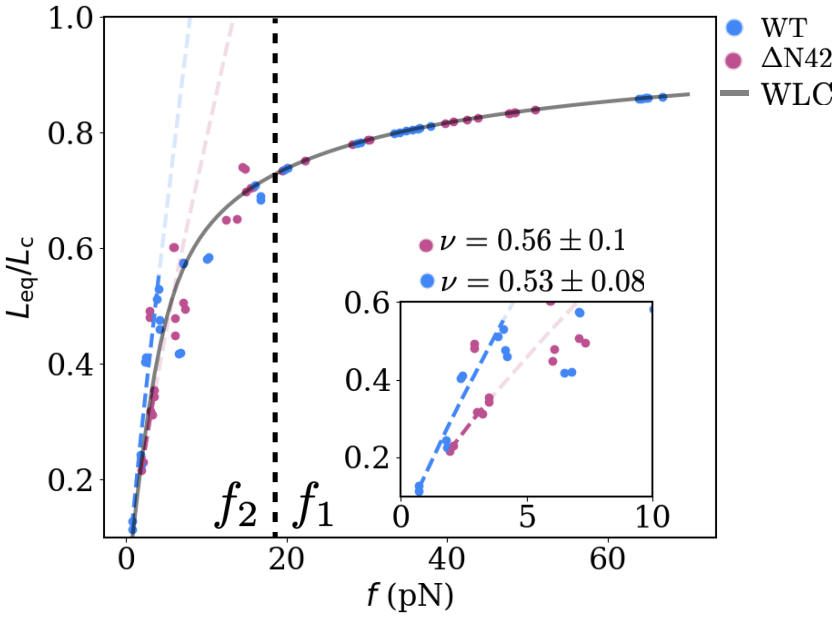

Figure 6.29: **A.** Equilibrium tether height L_{eq} normalized by the contour length L_c against applied force. With the WLC curve [96] in grey, it can be seen that the WT and ΔN42 curves agree with the WLC model. Dotted lines represent the Pincus blob relation fitting ($L \sim f^{1/\nu-1}$), with the lighter colors indicating the region outside the fitted data. **Inset** Zoomed in L_{eq}/L_c vs f graph, emphasizing the Pincus blob theory fitting for WT and ΔN42 [ref]. In both cases, the amount of points within the applicable range is too low to gain an accurate estimation of ν. **B.** Normalized bell curve error of ν for WT and ΔN42. The WLC assumption ($\nu = 1/2$) sits within the error range for both variants, while the SAXS measured values for ν sit well outside it.

As expected, for both high (f_1) and low (f_2) forces, the normalized equilibrium length (L_{eq}/L_c) is shown to rise with increase in force (Fig. 6.29A). From these data, ν can be extracted using the Pincus relation ($L_{eq} \sim f^{\frac{1}{\nu}-1}$) (See section 5.3 and [94]). The fit is valid in an appropriate force range of $f <\sim 2pN$ (Fig. 6.29A inset, [30]). Unfortunately, my data did not contain many points in the low force regime, resulting in ν values having a significantly large standard deviation, with $\nu = 0.53 \pm 0.08$ and $\nu = 0.56 \pm 0.1$ for WT and ΔN42 , respectively. Regardless, these values are much smaller than those evaluated from SAXS data, measured at $\nu = 0.751 \pm 0.013$ and $\nu = 0.729 \pm 0.0150$ for WT and ΔN42 , respectively, in 20 mM Tris pH 8.0. These deviations are likely due to the high degree of polymerization of the tethered sample, adding much more flexibility to the chains.

Furthermore, due to the large estimation error in the magnetic tweezers experiments, I can not distinguish between the ν values of both variants or whether they better fit an ideal chain

($\sim 1/2$) or a real chain ($\sim 3/5$) statistics. These results are corroborated by a recent study that showed NFLt to have a ν of 0.52 ± 0.01 in low-salt conditions [30].

In conclusion, my results of the two-step experiments show that both variants show logarithmic relaxation measurements. However, the failure to fit the data into the relaxation model shows that the currently collected data is insufficient to characterize the origin of the logarithmic relaxation and determine if they are based on independent heterogeneous modes.

6.5.3 Three-step experiment

The three-step experiments differ from the two-step experiments by adding the intermediate force (f_3), which satisfies $f_1 > f_3 > f_2$, which is applied following f_2. Additionally, f_2 is held for a much shorter timescale ($t = 10\,\text{s}$) to ensure the tethers do not reach equilibrium, after which f_3 is held for $t = 120\,\text{s}$ (Fig. 6.30). As with the two-step experiments, the final force measurement, which starts from t_ω after the start of the measurement, is logarithmically binned.

For NFLt, the logarithmic relaxation of the tethers was shown to be accompanied by a memory effect; the Kovacs hump [29, 97]. Both WT and ΔN42 largely failed to exhibit this memory effect, with only one tether of WT satisfying the Kovacs description (section 5.4, Fig. 6.31A). The lack of a measured Kovacs hump could be due to limited tether flexibility, which inhibits the ability of the segments to relax independently.

Interestingly, the only example showing the Kovacs hump is of the longest WT tether. Moreover, even a longer ΔN42 tether does not show the Kovacs hump. Regardless, as with the two-step experiment results, it is clear that more measurements are required to arrive at a conclusion.

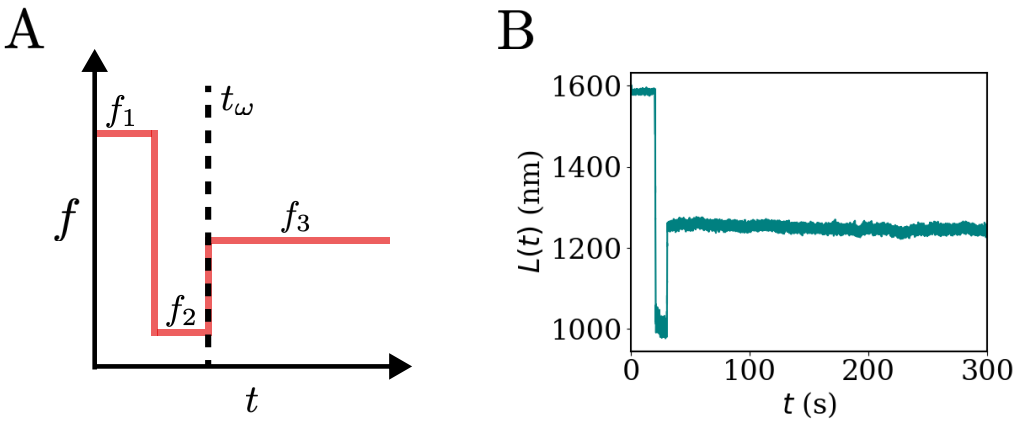

Figure 6.30: **A.** Schematic of the three-step force experiment, showing how the applied force f changes over time t. **B.** Three-step force experiment measurement of tether height ($L(t)$) over time.

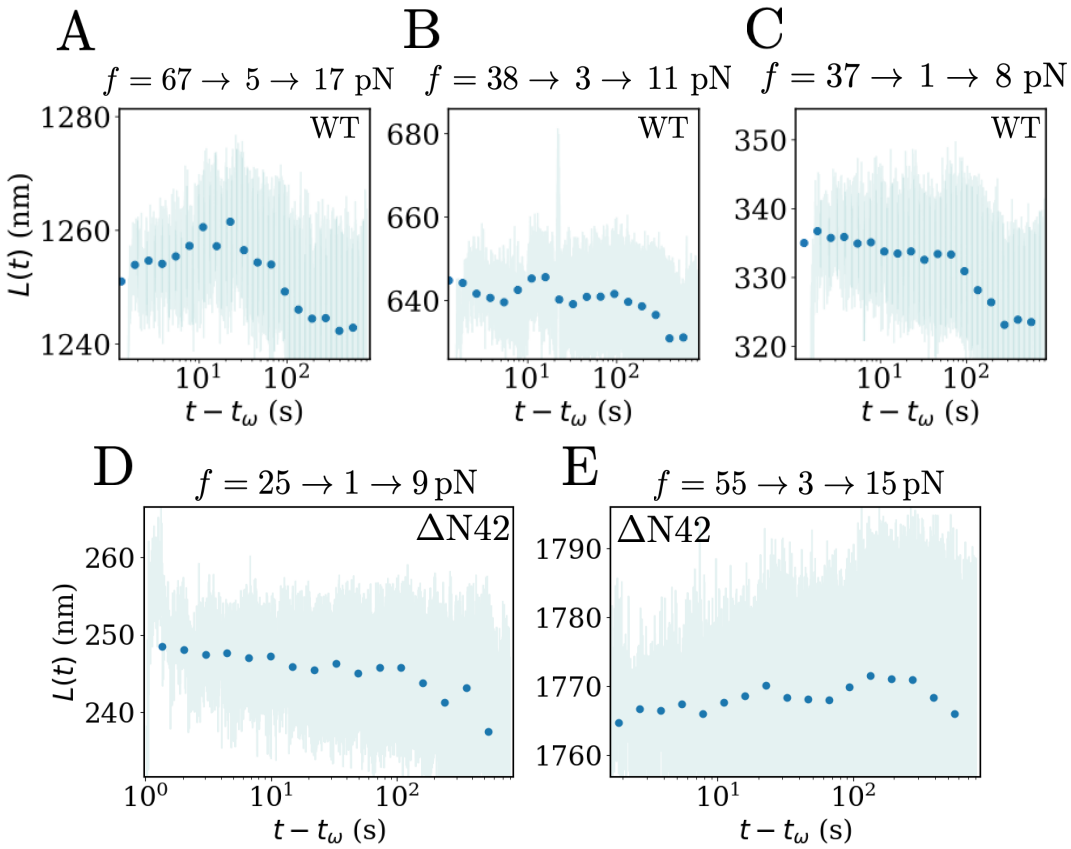

Figure 6.31: Logarithmically binned $L(t)$ for the middle force f_3 against time $t - t_\omega$ for WT (A-C) and ΔN42 (D-E). Only one measurement (A) exhibited a Kovacs hump, while the others lack the distinct features which make it so.

7 Experimental Methods

7.1 Protein Purification

Protein design. For NFLt, modified pET vectors were used. From its C terminal, PagP protein sequence is attached to the NFLt by a cleavage site. At the C-terminal end of the PagP protein, a Histag (6xH) sequence is attached for nickel binding. For ΔN42 , a modified pET vector containing NFLt with GCG sequence additions at both terminals was used as the base. Gibbson assembly was used to truncate the first 42 residues affiliated nucleotides. The same purification method as with NFLt then follows.

Purification. The protein purification process followed the protocol shown in [29]. Competent Escherichia coli BL21(DE3) Rosetta was transformed with the modified pET vector and plated on agar plates containing 100 g/ml ampicillin and 30 g/ml chlo- ramphenicol. A single colony was picked for starting cultures and grown overnight in 50 ml Terrific broth containing 100 g/ml ampicillin and 30 g/ml chloramphenicol. Cells were palletized, resuspended, and transformed in 1 L Terrific broth containing 100 g/ml ampicillin and 30 g/ml chloramphenicol. Expression cultures were grown in a baffled Erlenmeyer flask in a shaking incubator at 37°C at 280rpm for 2-4 hr until the optical density at 600 nm reached 0.7-1.0. Protein expression was induced by the addition of Isopropyl b-D-1-thiogalactopyranoside to a final concentration of 0.5 mM. The cultures were grown for 4 hr before harvesting. Cells were palletized and stored at -80°C for later use.

For the purification of proteins, cell pellets were resuspended in a 10 ml lysis buffer for each 1 g bacterial pellet. The lysis buffer contained 20 mM Tris buffer, pH 8.0, 0.1% 2-Mercaptoethanol, 1% Triton, and 0.5mg/ml Lysozyme. The solution was incubated at 25°C for 20 min. Next, the solution was added with 10 mM MgSo4 and 1k units of Benzonase nuclease for 20 min at 25°C. The solution was centrifuged at 18,500 g for 30 min at 4°C. Next, the pellet was homogenized in a washing buffer containing 20 mM Tris pH 8.0, 6 M Guanidine HCL, 20 mM imidazole, and 0.1 2-Mercaptoethanol. After centrifugation at 18,500 g for 30 min at 4°C, the supernatant was loaded on a 10 ml home-packed nickel affinity column that was equilibrated with a washing buffer at a 1 mL/min rate. After washing with 100 ml, the protein solution was eluted with an elution buffer containing 20 mM Tris buffer pH 8.0, 0.5M Imidazole, and 0.1% 2- Mercaptoethanol (Fig. 7.1). The solution was then dialyzed overnight against 1 L of 20 mM Tris pH 8.0, followed by another dialysis against 1 L of 50 mM 3-(N-morpholino) propane sulfonic acid (MOPS) buffer pH 8.5.

After dialysis, a cleavage reaction was initiated by 5 mM NiSO4 in the presence of 6 M Guanidine and incubated for 20 hr at 50°C. The cleavage reaction was stopped by 50 mM EDTA and followed by adding 0.1% 2-Mercaptoethanol. The cleaved protein was dialyzed twice overnight against 1 L 20 mM Tris pH 8.0, 2 mM EDTA, and 0.1% 2- Mercaptoethanol. The cleaved PagP precipitate was centrifuged at 18,500 g for 30 min and discarded. The protein was adjusted to 6

M Guanidine and was loaded on a 100 mL size-exclusion column (HiPrep 16/60 Sephacryl S-200 HR) at a rate of 1 mL/min pre-equilibrated with washing buffer containing 20 mM Tris pH 8.0, 1M Guanidine, 2 mM EDTA and 0.1% 2-Mercaptoethanol (Fig. 7.1). Finally, the guanidine and 2-mercaptoethanol were removed via 20 mM Tris pH 8.0 dialysis and measured for purity using SDS-PAGE gel. The final purity was over 95% (Fig. 7.2A). In addition, the identity of the produced proteins was verified by a mass-spec measurement (Fig. 7.2B,C). A growth cycle using four colonies with 4L of Terrific broth typically yielded 2 to 3 mg of protein.

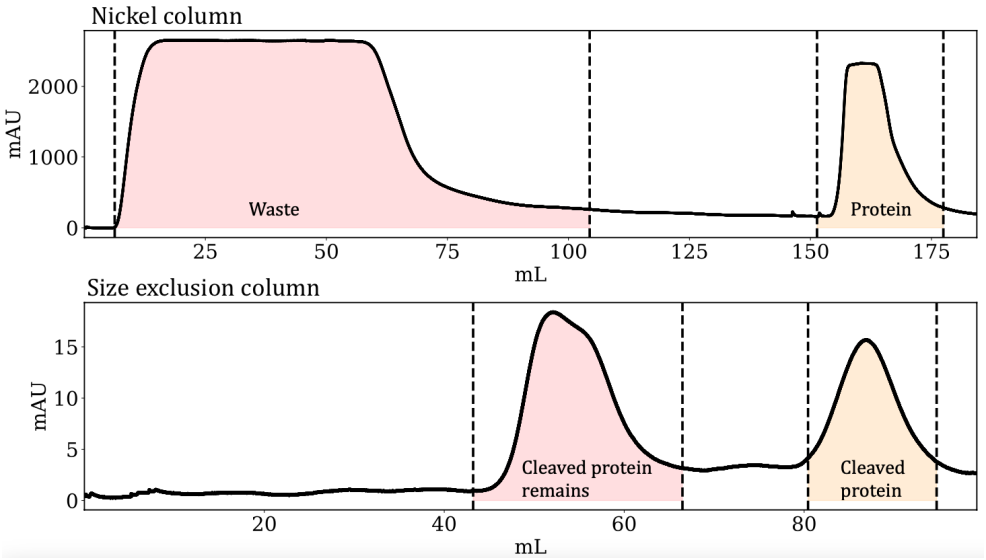

Figure 7.1: Typical Nickel and size exclusion chromatography runs. The y axis measures the absorption of the solution and the x axis pertains to the solution volume passing through the column. In both runs, only the proteins (yellow) are collected, while the waste (red) is discarded.

Variant ΔN42 included two cysteine residues at the C- and N terminals. After purification, ΔN42 variants were first reduced by 20 mM 2-Mercaptoethanol. Next, 2-Mercaptoethanol was dialysed out with 1 L of 50 mM HEPES at pH 7.2. To block the cysteine sulfhydryl group, we reacted ΔN42 variants with 2-Iodoacetamide at a molar ratio 1:20. At the reaction, the variants' concentrations were $\sim$ 2 mg/ml. The reaction solution was kept under dark and slow stirring for 5 hours and stopped by adding 50 mM 2-Mercaptoethanol, followed by overnight dialysis against 1 L of 20 mM Tris at pH 8.0 with 0.1% 2-Mercaptoethanol.

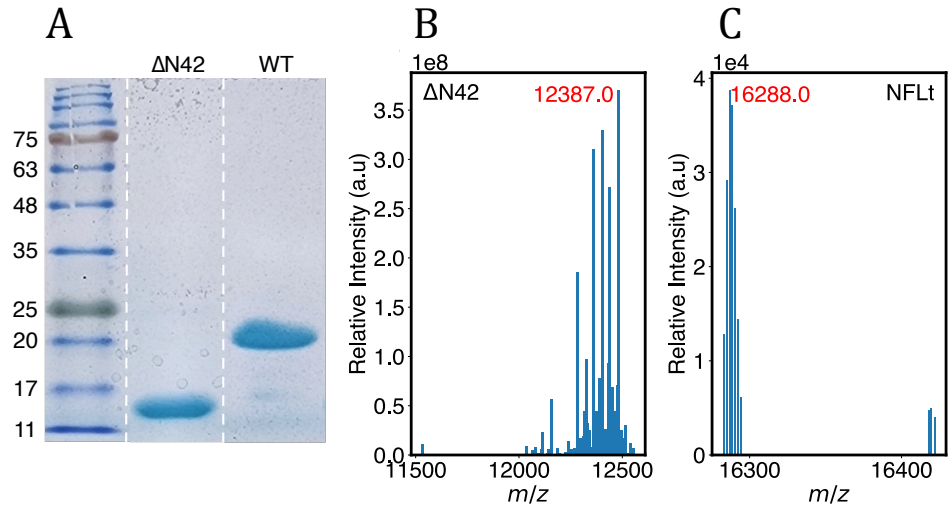

Figure 7.2: **A.** SDS-PAGE Tris-Glycine 15% of both ΔN42 and NFLt (WT), showing purity above 95%. White dashed lines indicate where image lanes were edited closer for clarity. Both show a higher molecular weight reading in the gel, which is common for IDPs. **B-C.** Deconvoluted ESI-TOF MS spectra of ΔN42 and NFLt respectively. Theoretical molecular weight values are 12423.57 and 16233.79 for ΔN42 and NFLt, respectively. Figure adapted from [1]

7.2 SAXS Measurements

7.2.1 Measurements.

Protein samples were dialyzed overnight in the appropriate solution and measured with a Nanodrop 2000 spectrophotometer (Thermo Scientific) for concentration determination. Buffers were prepared with 1 mM of TCEP to reduce radiation damage and 0.2% of Sodium Azide to impair sample infection. The samples were prepared in a final concentration of 2 mg/ml, measured in a series of 4 dilutions. Preliminary measurements were measured at Tel-Aviv University with a Xenocs GeniX Low Divergence CuKα radiation source setup with scatterless slits [104] and a Pilatus 300K detector. All samples were measured at three synchrotron facilities: beamline B21, Diamond Light Source, Didcot, UK [105], beamline P12, EMBL, DESY, Hamburg, Germany [106], and beamline BM 29 ESRF, Grenoble, France [107]. Measurements at ESRF were done using a robotic sample changer [108].

7.2.2 Analysis.

Integrated SAXS data was obtained from the beamline pipeline and 2D integration using the "pyFAI" Python library [109].

Extended Guinier analysis. Extended Guinier analyses [87] for the ΔN42 variant were done with the "curve_fit" function from the "Scipy" Python library [110], using eq (). To extract R_g and ν, extended Guinier analysis was conducted for $0.7 < qR_g < 2$. Error calculation was done from the covariance of the fitting. The data fitting code can be found in the Appendix.

Aggregate model fitting. Model fittings for the WT variant were done using the "lmfit" Python library [111] using the model described in [99, 100]. Due to the complexity of the model, cylindrical core fittings were done by binning the data in 100 logarithmic bins to reduce computation time. Core parameters (cylinder radius R and cylinder length L) were set constant within the same model to offset fitting errors. Initial values of R and L were calculated with the highest measured concentration. Physical boundary conditions were imposed on the fitting, and scattering length (SL) values were set to be unchanged by the fitting process. SL values of both the core and the tail domains were determined by tabulated values of amino acid SLD in 100% H_2O [112] (Table S3). Fitting parameter error evaluation was done by finding the covariant of the returning fitting parameters. Error calculation of the volume was done using: $\frac{dV}{V} = \sqrt{3\left(\frac{dR}{R}\right)^2}$. In addition, ν values of WT were found by a recursive search of the corresponding tail height $h/2$ over Eq. 6.4. Errors of ν were then found by assuming a simple case of $R_G \sim bN^{\nu}$, from which: $d\nu \sim \frac{\ln(1+dR_G/R_G)}{\ln N} \sim \ln(N)^{-1}\frac{dR_G}{R_G}$. All appropriate codes are found in the codes section of the thesis.

Zimm analysis Zimm analysis was performed as described in Section 4.3.1. The analysis was performed on a limited range, in which the $1/I(q,c)$ curve is roughly linear. For WT, this resulted in a q range of $0.08 < q < 0.3$ nm and $0.1 < q < 0.45$ for ΔN42. The analysis was performed on $1/I(q,c)$ binned to 10 bins within the allocated q range. Importantly, changing the amount of bins did not affect the result. Initially, the analysis was performed on the non-normalized ($1/I(q,c)$ not divided by I_{00}) data. From the Zimm plot, I_{00} was then extracted by the intersection of the $1/I(q,0)$ and $1/I(0,c)$ curves. Using I_{00}, the Zimm plot could now be normalized, from which the osmotic second virial coefficient A_2 is extracted by the slope of the normalized $1/I(0,c)$ curve.

Brush model fitting The salt-brush model, as described in section 3.2 and [102] contains a dependency on both salt concentration C_s and grafting density ρ by:

$$H \sim C(\frac{\rho}{C_s})^{1/3}. \tag{7.1}$$

Here, C is a fitting parameter used to align the model with the data. As ρ is dependant on aggregation number Z and sphere radius R by $\rho = Z/4\pi R^2$, a continuous function $\rho(C_s)$ is needed in order to find a continuous curve for $H(C_s)$. This step was achieved by a linear curve fitting

over $\rho(C_s)$, as shown in Fig. 6.18B in the results. Finally, using $\rho(C_s)$, $H(C_s)$ was estimated by fitting C of eq. 7.1, resulting in C = 0.33. With the resulting relation $H(C_s)$, ν was calculated by eq. 6.4 to obtain the final result of $\nu(C_s)$.

Polyelectrolye fitting. The polyelectrolyte model fitting was performed using the model described in Section 2.3 and Ref. [64]. Recalling the equations:

$$v' = v + \frac{4\pi\alpha^2 z_{\mathrm{p}}^2 l_{\mathrm{B}}}{\kappa^2 l^3}, \tag{7.2}$$

where $\kappa = \lambda_D^{-1} =$. For $R_{\mathrm{G}} = R_0 R_{\mathrm{ee}} = R_0\sqrt{N/6}\,b$, the prefactor R_0 is given by:

$$R_0^5 - R_0^3 \sim A\frac{134}{105}(\frac{3}{2\pi})^{3/2} v'\sqrt{N}. \tag{7.3}$$

where A is the fitting parameter. It is much easier to compute $C_s(R_0)$ rather than $R_0(C_s)$. As such, the fitting was performed on the inverse function $C_s(R_g)$. Obtaining the relation between R_{G} and C_s, ν was calculated by eq. 6.4 to obtain $\nu(C_s)$ for the model.

7.3 Magnetic tweezers setup

The magnetic tweezers setup was prepared in three steps:

1. Flowcell preparation. The flowcell consists of a top and bottom plate, with a substrate in between the plates. The preparation of the top plate starts with cleaning the glass plate in NaOH, MiliQ water, and finally, with nitrogen. After which, the plate was placed in a plasma cleaner for five minutes for additional particle removal. Finally, a hydrophobic coating was then applied to the top plate to ensure a smooth flow of liquids within the flowcell. The bottom plate was initially cleaned with nitrogen. To the plate, 18μL of reference bead solution was evenly spread on the surface. As the reference beads were suspended in Ethanol, the plate was placed in a closed container until the liquids mostly dissipated. Finally, the plate was placed on a pre-heated hotplate at 100˚C for 1 minute and 20 seconds.

With both plates now ready, the final assembly of the substrate was started. The substrate was prepared by a parafilm cut, as shown in Fig. 7.3A, which was melted on the bottom plate using a solder. The top plate was then placed on the parafilm perpendicularly to create two equal entrances to the substrate on both sides (Fig. 7.3B). Finally, the parafilm was melted with the solder on the top plate. With the prepared flowcell, 5 μL of DBCO was diluted in 20 μL of 1% PBS was loaded into the substrate. With the DBCO solution loaded, the flowcell was kept in a moisturized container for one hour.

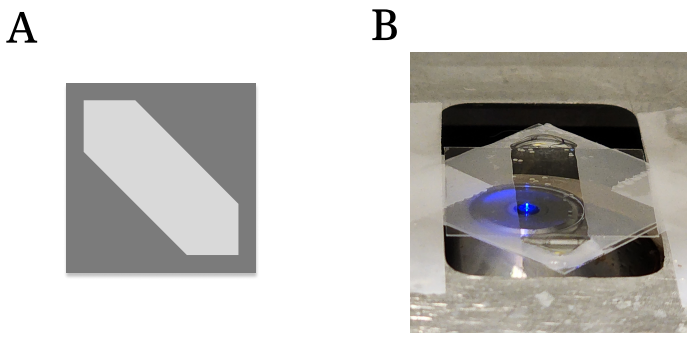

Figure 7.3: **A.** Schematic of the parafilm used to make the substrate. **B.** Overview photo of the flowcell. Liquids are dispensed into the flowcell from its front end, which can be moved across by applying pressure on the back end of the flowcell. Below the flowcell is a focal lens which is used to observe the tethers with the camera setup.

2. Sample preparation. WT and ΔN42 were synthesized with cysteines on both terminals by adding the sequence *GCG* to each to induce polymerization. The variants were terminally labeled with azide and biotin to allow specific attachment between the substrate and the magnetic beads, respectively. Polymerization was checked in 20% SDS PAGE (Fig. 7.4).

To alleviate possible aggregations, 1 μL of the variants at 0.5 mg/ml were diluted into 500 μL of 10mM carbonate pH 11 buffer and were rotated for one hour. After which, the variants were dialyzed into a solution of 500 μL 20mM Tris pH 7.4 0.1% Tween-20. This solution was finally concentrated into $\sim$ 20 μL. Returning to the flowcell, the DBCO solution is removed from the substrate. The flowcell was then cleaned using a 20 mM Tris pH 7.4 buffer. The variant solution was then added to the flowcell to allow for the chemical binding of the chains to the substrate. This step was performed at 4°C overnight.

3. Setup assembly. The prepared flowcell was again washed with a 20 mM Tris pH 7.4 buffer and placed on the mechanized stage of the magnetics tweezers setup. The magnetic bead solution was then prepared as follows: 8 μL of stock solution of 2.8 $^{-}$m diameter magnetic beads was diluted in 200 μL of the Tris buffer. The bead solution was then vortexed until the beads were separated from the liquid. Removing the liquid, 200 μL of the Tris buffer was again added, and the vortexing step was repeated. Finally, the separated beads were diluted into 40 μL of the Tris buffer.

The magnetic beads were loaded onto the flowcell slowly, using only 10 μL of the prepared solution at a time. This step was done to ensure the magnetic beads efficiently bind to the biotin terminals of the tethers. Once the beads were loaded, the flowcell was rewashed in 10 mM MES

pH 6.8 0.1% Tween-20 buffer, the measurement buffer.

7.4 Tether tracking experiments

Tracking the beads was done using a tracking program [92, 29]. For each located tether, nearby reference beads were used to track the tether's x, y, and z positions. The two-step experiments were performed by initially setting the magnet in its closest position to the flowcell. From here, the first force step was measured for 120 seconds, after which the magnet was raised by 1 mm to measure the second step for an additional 120 seconds. These two steps were repeated two times, each raising the magnet by an additional 1 mm to achieve lower forces. With these steps, a high force, typically in the 30 - 60 pN range, was achievable, with low forces up to 25% of the high force.

The three-step experiments were performed similarly. After the 120-second step of the high force, the magnet is moved 2mm away from the flowcell for 10 seconds. When 10 seconds pass, the magnet was moved closer to the flowcell by 1 mm for an additional step of 600 seconds.

7.5 Tether data analysis

From the tether tracking measurements, x, y and z positions of the tethers were extracted. With the x position, the force applied on the tethers was calibrated by an allan-curve analysis, as described in section 5.2. The fitting to the Allan curve was done using an maximum likelyhood fitting algorithm [113, 91], due to the non-normal standard deviation of the tracking experiments. Analysis code for the tracking experiments was provided by Hoang P. Truong from the University of California, Santa Barbara, and can be additionally found on the Tweezepy' python library [114].

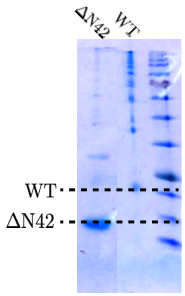

Figure 7.4: SDS-PAGE Tris-Glycine 15% of both ΔN42 and NFLt (WT) polymerizations, edited for clarity. Both variants show molecular weight readings above their expected location (dashed) in the gel, indicating that the polymerization process was successful.

8 Discussion

In my thesis, I investigated the effects of the sequence heterogeneity of the IDP model system, NFLt. From the equilibrium SAXS results, the two NFLt variants WT and ΔN42 attested to a complex interaction network that varied according to the sequence of the variant. This difference manifested in the WT variant self-assembling into star-like structures, while the ΔN42 variant remained isolated in all measured cases. In addition to the equilibrium measurements, force stretch response experiments have shown that both variants exhibit heterogeneous relaxations with a sudden drop in stretching force.

8.1 Interactions

For ΔN42, the intermolecular interactions analysis showed that the chains weakly attract each other at the high salinity regime. These attractive interactions likely arise from weak monomer-monomer interactions of possible hydrophobic origins. Such hydrophobic attractions are known to affect IDPs, as they affect the solvent condition of the chain, which in turn impacts its steric extended volume interactions [26, 42, 115]. A possible origin of these interactions is a salting out effect, where the hydrophobic attraction between the scattered hydrophobic sites (Fig. 1.8) grows stronger with salinity [116, 117].

Compared to ΔN42, the intermolecular interactions analysis of WT describes the whole aggregate rather than the individual chains. These interactions start from a near-neutral state at low sanity and transition to weakly attractive at high salinity. As such, these interactions describe the aggregate's ability to grow as an aggregate, i.e., its 'aggregation propensity.' This behavior is further observed in the aggregation numbers' (Z) response to protein concentration, which is seen to grow as screening effects are increased. Interestingly, this case is parallel to folded protein aggregation. In folded proteins, intermolecular interactions were also shown to correlate directly with their aggregation propensity [118]. As with ΔN42, the weakly attractive interactions of WT are likely of hydrophobic origin. The increased attraction and subsequent increase in aggregation numbers indicate that salting-out phenomena affect the protein interactions; however, further investigation at higher salinities is needed to confirm these effects.

The stability of the star-like core should be evaluated by the core residue number per polypeptide (n). While n values were estimated to be rather small, as seen by the SAXS scattering, the WT ensemble is dominated by aggregated structures for all salinities. The stability of the core is likely attributed to entropic hydrophobic attractions, which occur when the residue releases a bound ion or water molecule. These 'depletion' attractions are known to be prominent in numerous protein-protein interaction processes, such as IDP binding [119].

8.2 Scaling

Compared to ΔN42, WT exhibits a milder contraction in salt, which was shown to resemble the 'salted brush' regime of polyelectrolyte brushes, characterized by a $H \propto C_s^{-1/3}$ brush height (H) dependency on salt concentration (C_s). Similar salt brush behavior was previously observed in Neurofilament high tail domain (NFHt) brushes grafted on a substrate [62]. These brushes, however, transitioned to the salt-brush regime at a lower salt concentration of $C_s = 10$ mM, compared to $C_s = 170$ mM of NFLt. Recalling my discussion of the different NF sidechains, NFHt has a weaker charge composition compared to NFLt, with the majority of its sequence being a positive chain with an NCPR of +0.046 (Section 1.3). As such, this difference in domain transition is likely due to the amount of external ions required to overcome the counterions found within the brushes.

Salt brush behavior resembles that of neutral brushes, which are determined by steric interchain interactions [80]. Equating them to neutral brushes, their effective excluded volume per monomer is enlarged by the electrostatic effects and is proportional to the Debye length κ by $1/\kappa$. Consequently, it is possible that in the high salinity regime of $C_s > 170$ mM, steric effects play a crucial role in the brush ensemble of WT. These steric repulsions could be accountable for contradicting the attractive forces responsible for the aggregation, thereby preventing brush collapse at high salinities.

8.3 Comparisons to past studies

Recalling section 6.4, ΔN42 and previously measured NFLt segments were shown to be largely dissimilar from each other. These results contrast real, homogeneous chains known to retain their extension ν across any arbitrarily chosen small segment within the chain [26]. This contrast, however, is not unexpected due to the high structural heterogeneity of the segmented NFLt and the prevalence of intermolecular interactions in ΔN42 solutions. In the comparison between ΔN42 and the segments, it was seen that in the low salt regime of salt concentration $C_s < 170$ mM, while ΔN42 exhibited a significant contraction in ν, both the P and S segments remained largely unchanged. In addition, at the high salinity condition of $C_s = 520$ mM, where the intermolecular electrostatic repulsion was shown to be inconsequential, ΔN42 converged to the same ν as with the S segments. Recalling the electrostatic blob description of polyelectrolytes (section 2.3), in a sufficiently small length scale, interactions outside of the small polymer segments become negligible. In the context of the blob description, the segments are likely small enough so that the long-ranged electrostatic effects become negligible. This effect can be further seen at high salinity, where the repulsive effects are no longer present, and the expansion of the whole chain is similar to its segments.

Individual NFLs resemble bottlebrushes with protruding NFLts in the cytoskeleton. These individual NFLs are mediated by the NFLts, which create an interfilament distance D between

them [52, 50]. In a previous study on NFL formations [52], it was shown that under osmotic pressure similar to physiological conditions, the intrafilament distance suddenly drops at a salinity concentration of $C_s = 100$ mM, from which it remains relatively unchanged. This result correlates with my findings, which showed for both ΔN42 and WT that a transition from a low-salt to a high-salt regime occurs at C_s between 70 mM and 170 mM.

Furthermore, the study showed that the NFL brushes maintain their structure due to the steric repulsion of the brushes by comparing the results to a simulation model. In isolating the NFLts, this result is further confirmed experimentally, with the steric effects dominating the salted-brush expansion and preventing the WT aggregates from collapsing.

8.4 Outside of equilibrium

The equilibrium measurements showed that the aggregation of the NFLts occurred through their neutral domain. These results correlate with the non-equilibrium study on polymerized NFLts, where the N-terminal domain collapse is likely responsible for the observed slow-aging effects [29]. With the polymerized ΔN42, the lack of observed Kovacs hump compared to WT could be then due to the truncation of the N-terminal. However, as the lack of data is apparent, this argument can not be confirmed. Upon relaxing, polymerized WT and ΔN42 were identified near a real or ideal chain by their calculated scaling exponent ν. These results were consistent with a previous equilibrium study on the polymerized NFLt that showed ν to measure as 0.52 ± 0.006 at the same salinity conditions [30]. In the same study, increasing the denaturation showed swelling in the chain to the real chain value of $\nu \sim 3/5$, likely due to the reversal of the N-terminal domain collapse. My SAXS measurements on WT at high denaturation further corroborate this effect, which saw the denaturation reversing the aggregation and resulting in the real-chain value of ν.

8.5 Research limitations

The most significant limitation of my research was the lack of non-equilibrium data. With more measurements, the effects of the neutral domain on the kinetics of NFLts could be better understood. For the SAXS measurements, additional measurements featuring much higher protein and salt concentrations could have also added crucial information to test the stability of the steric effects in maintaining the aggregate's shape at extreme conditions. Such measurements, however, would have required a much larger amount of protein, which I did not manage to produce and purify.

8.6 Final remarks

In conclusion, I showed how the sequence heterogeneity of the NFLt IDP caused it to assemble into star-like brushes when isolated from the NFL backbone. These effects were used to demonstrate how the properties of these resulting brushes could be used to resist environmental conditions by the steric effects of the brushes. Additionally, I showed how these effects may explain the underlying mechanics that regulate and stabilize NFL network formations. Removing the sequence responsible for the aggregation resulted in a completely disordered chain, reinforcing the intrinsic difference between polymers and IDPs and how these differences may affect the theoretical backbone used for their study.

9 Future Work

9.1 Link between sequence and interactions

Sequence-specific effects were proven considerable on the interactions of the variants in my study. In WT, this resulted in N-terminal sequence interactions causing protein aggregation, while for ΔN42 , high electrostatic repulsion dictated the conformations of the ensemble. In future work, these effects could be further probed by inducing mutations to the NFLt sequence. For example, the highly hydrophobic residues of the N-terminal could be replaced with non-hydrophobic amino acids to assess their impact on the aggregation. These mutations could shed additional light on the link between sequence and protein-protein interactions. Specific disease-related mutations, such as those found in the Charcot-Marie-Tooth disease [120], could also be studied and compared to the WT case.

9.2 Theoretical models

For the two variants, sequence heterogeneity directly led to different theoretical models being used to characterize the IDPs. Current theoretical frameworks for IDP ensembles fail to characterize systems where protein-protein interactions completely alter the IDP structure. New models that can directly predict how sequence-derived intermolecular interactions and intramolecular interactions might affect the IDP structure could prove immensely useful for the future characterization of IDP systems and their applications.

9.3 Dimerization of NFLt chains

Due to the significant difference between the N- and C-terminal domains of the NFLt, investigating how dimerizations between different terminals could impact the equilibrium and non-equilibrium behavior of the chain. In equilibrium, different types of dimerizations have the potential to completely alter the interaction landscape of the protein due to the changing position of the N-terminal domains of the dimerized chains. These could lead to different kinds of aggregations or their elimination. Outside of equilibrium, as the neutral domain effectively 'expands' upon polymerization of dimers, which connect by their N-terminals, the slow-aging effects could grow more prominent, leading to new discoveries on the effects of the neutral domain on the kinetics of the chains.

9.4 Additional magnetic tweezers experiments

As mentioned, a severe lack of out-of-equilibrium data impeded the number of conclusions I could draw in this study. Thus, future work could include a much larger data set, with the addition

of high denaturant conditions for the variant, to test how denaturation of the hydrophobic effects could affect the glassy dynamics of the NFLts.

9.5 Applications of self-assembling IDP brushes

The ability of polymer brushes to control surface functionality enabled them to have many possible applications in recent years. Recent studies attempting to use IDP brushes as an alternative to traditionally synthetic polymer brushes have shown many benefits, such as biodegradability and enabling environmentally-sensitive surface functionality [62]. As such, IDPs in similar composition to NFLts with the ability to self-assemble to spherical brushes could see applications where environment-specific surface control might be needed. The spherical nature of these brushes enables them to easily travel to targeted areas, where they could act as barriers, apply pressure, or other surface-related effects as needed.

References

[1] Mathar Kravikass et al. "From isolated polyelectrolytes to star-like assemblies: The role of sequence heterogeneity on the statistical structure of the intrinsically disordered Neurofilament-low tail domain". In: *arXiv preprint arXiv:2311.18337* (2023).

[2] Denise Wootten et al. "Mechanisms of signalling and biased agonism in G protein-coupled receptors". In: *Nature reviews Molecular cell biology* 19.10 (2018), pp. 638–653.

[3] Christopher Buccitelli and Matthias Selbach. "mRNAs, proteins and the emerging principles of gene expression control". In: *Nature Reviews Genetics* 21.10 (2020), pp. 630–644.

[4] Chao Peng, John Q Trojanowski, and Virginia M-Y Lee. "Protein transmission in neurodegenerative disease". In: *Nature Reviews Neurology* 16.4 (2020), pp. 199–212.

[5] Christian B Anfinsen. "Principles that govern the folding of protein chains". In: *Science* 181.4096 (1973), pp. 223–230.

[6] Julie L Morrison et al. "A lock-and-key model for protein–protein interactions". In: *Bioinformatics* 22.16 (2006), pp. 2012–2019.

[7] John C Kendrew et al. "A three-dimensional model of the myoglobin molecule obtained by x-ray analysis". In: *Nature* 181.4610 (1958), pp. 662–666.

[8] Ken A Dill and Justin L MacCallum. "The protein-folding problem, 50 years on". In: *science* 338.6110 (2012), pp. 1042–1046.

[9] Lorenzo Di Rienzo et al. "Characterizing hydropathy of amino acid side chain in a protein environment by investigating the structural changes of water molecules network". In: *Frontiers in molecular biosciences* 8 (2021), p. 626837.

[10] David Eisenberg, William Wilcox, and Andrew D McLachlan. "Hydrophobicity and amphiphilicity in protein structure". In: *Journal of cellular biochemistry* 31.1 (1986), pp. 11–17.

[11] Kresten Lindorff-Larsen et al. "How fast-folding proteins fold". In: *Science* 334.6055 (2011), pp. 517–520.

[12] John Jumper et al. "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596.7873 (2021), pp. 583–589.

[13] Vladimir N Uversky. "A decade and a half of protein intrinsic disorder: biology still waits for physics". In: *Protein Science* 22.6 (2013), pp. 693–724.

[14] Fred Karush. "Heterogeneity of the binding sites of bovine serum albumin1". In: *Journal of the American Chemical Society* 72.6 (1950), pp. 2705–2713.

[15] H Jane Dyson and Peter E Wright. "Intrinsically unstructured proteins and their functions". In: *Nature reviews Molecular cell biology* 6.3 (2005), pp. 197–208.

[16] Christopher J Oldfield and A Keith Dunker. "Intrinsically disordered proteins and intrinsically disordered protein regions". In: *Annual review of biochemistry* 83 (2014), pp. 553–584.

[17] Robin Van Der Lee et al. "Classification of intrinsically disordered regions and proteins". In: *Chemical reviews* 114.13 (2014), pp. 6589–6631.

[18] Peter E Wright and H Jane Dyson. "Intrinsically disordered proteins in cellular signalling and regulation". In: *Nature reviews Molecular cell biology* 16.1 (2015), pp. 18–29.

[19] Mark Wells et al. "Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain". In: *Proceedings of the National academy of Sciences* 105.15 (2008), pp. 5762–5767.

[20] Mainak Guharoy et al. "Intrinsic structural disorder in cytoskeletal proteins". In: *Cytoskeleton* 70.10 (2013), pp. 550–571.

[21] Jesus Avila et al. "Role of tau protein in both physiological and pathological conditions". In: *Physiological reviews* (2004).

[22] Pau Bernado and Dmitri I Svergun. "Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering". In: *Molecular biosystems* 8.1 (2012), pp. 151–167.

[23] Robert M Clegg. "Förster resonance energy transfer—FRET what is it, why do it, and how it's done". In: *Laboratory techniques in biochemistry and molecular biology* 33 (2009), pp. 1–57.

[24] Kevin Truong and Mitsuhiko Ikura. "The use of FRET imaging microscopy to detect protein–protein interactions and protein conformational changes in vivo". In: *Current opinion in structural biology* 11.5 (2001), pp. 573–578.

[25] Benjamin Schuler and William A Eaton. "Protein folding studied by single-molecule FRET". In: *Current opinion in structural biology* 18.1 (2008), pp. 16–26.

[26] Gil Koren et al. "Intramolecular structural heterogeneity altered by long-range contacts in an intrinsically disordered protein". In: *Proceedings of the National Academy of Sciences* 120.30 (2023), e2220180120.

[27] Benjamin Schuler et al. "Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins". In: *Annual Review of Biophysics* 45 (2016), pp. 207–231.

[28] Kipom Kim and Omar A Saleh. "A high-resolution magnetic tweezer for single-molecule measurements". In: *Nucleic acids research* 37.20 (2009), e136–e136.

[29] Ian L Morgan et al. "Glassy dynamics and memory effects in an intrinsically disordered protein construct". In: *Physical Review Letters* 125.5 (2020), p. 058001.

[30] Hoang P Truong et al. "Pincus blob elasticity in an intrinsically disordered protein". In: *The European Physical Journal E* 46.10 (2023), p. 100.

[31] Joan-Emma Shea, Robert B Best, and Jeetain Mittal. "Physics-based computational and theoretical approaches to intrinsically disordered proteins". In: *Current opinion in structural biology* 67 (2021), pp. 219–225.

[32] Upayan Baul et al. "Sequence effects on size, shape, and structural heterogeneity in intrinsically disordered proteins". In: *The Journal of Physical Chemistry B* 123.16 (2019), pp. 3462–3474.

[33] Rahul K Das and Rohit V Pappu. "Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues". In: *Proceedings of the National Academy of Sciences* 110.33 (2013), pp. 13392–13397.

[34] Hagen Hofmann et al. "Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy". In: *Proceedings of the National Academy of Sciences* 109.40 (2012), pp. 16155–16160.

[35] Wenwei Zheng et al. "Hydropathy patterning complements charge patterning to describe conformational preferences of disordered proteins". In: *The journal of physical chemistry letters* 11.9 (2020), pp. 3408–3415.

[36] Daria Maltseva et al. "Fibril formation and ordering of disordered FUS LC driven by hydrophobic interactions". In: *Nature Chemistry* (2023), pp. 1–9.

[37] Sonja Müller-Späth et al. "Charge interactions can dominate the dimensions of intrinsically disordered proteins". In: *Proceedings of the National Academy of Sciences* 107.33 (2010), pp. 14609–14614.

[38] Sigrid Milles and Edward A Lemke. "Single Molecule Study of the Intrinsically Disordered FG-repeat Nucleoporin 153". In: *Biophysical Journal* 102.3 (2012), 10a.

[39] Joshua A Riback et al. "Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water". In: *Science* 358.6360 (2017), pp. 238–241.

[40] Xiangze Zeng, Kiersten M Ruff, and Rohit V Pappu. "Competing interactions give rise to two-state behavior and switch-like transitions in charge-rich intrinsically disordered proteins". In: *Proceedings of the National Academy of Sciences* 119.19 (2022), e2200559119.

[41] Lauren H Kapcha and Peter J Rossky. "A simple atomic-level hydrophobicity scale reveals protein interfacial structure". In: *Journal of molecular biology* 426.2 (2014), pp. 484–498.

[42] Vladimir N Uversky et al. "Natively unfolded human prothymosin α adopts partially folded collapsed conformation at acidic pH". In: *Biochemistry* 38.45 (1999), pp. 15009–15016.

[43] Albert H Mao et al. "Net charge per residue modulates conformational ensembles of intrinsically disordered proteins". In: *Proceedings of the National Academy of Sciences* 107.18 (2010), pp. 8183–8188.

[44] Aitor Manteca et al. "The influence of disulfide bonds on the mechanical stability of proteins is context dependent". In: *Journal of Biological Chemistry* 292.32 (2017), pp. 13374–13380.

[45] Vladimir N Uversky. "The alphabet of intrinsic disorder: II. Various roles of glutamic acid in ordered and intrinsically disordered proteins". In: *Intrinsically disordered proteins* 1.1 (2013), e24684.

[46] Borja Mateos et al. "The ambivalent role of proline residues in an intrinsically disordered protein: from disorder promoters to compaction facilitators". In: *Journal of molecular biology* 432.9 (2020), pp. 3093–3111.

[47] Edward A Weathers et al. "Insights into protein structure and function from disorder–complexity space". In: *Proteins: Structure, Function, and Bioinformatics* 66.1 (2007), pp. 16–28.

[48] Pedro Romero et al. "Sequence complexity of disordered protein". In: *Proteins: Structure, Function, and Bioinformatics* 42.1 (2001), pp. 38–48.

[49] Jérôme Bürgi et al. "Intrinsic disorder in transmembrane proteins: roles in signaling and topology prediction". In: *PLoS One* 11.7 (2016), e0158594.

[50] Adi Laser-Azogui et al. "Neurofilament assembly and function during neuronal development". In: *Current opinion in cell biology* 32 (2015), pp. 92–101.

[51] Roy Beck et al. "Gel-expanded to gel-condensed transition in neurofilament networks revealed by direct force measurements". In: *Nature materials* 9.1 (2010), pp. 40–46.

[52] Micha Kornreich et al. "Neurofilaments function as shock absorbers: compression response arising from disordered proteins". In: *Physical review letters* 117.14 (2016), p. 148101.

[53] Alex S Holehouse et al. "CIDER: resources to analyze sequence-ensemble relationships of intrinsically disordered proteins". In: *Biophysical journal* 112.1 (2017), pp. 16–21.

[54] Michael Rubinsten. *Polymer physics*. United States of America, 2003.

[55] Paul J Flory. "The configuration of real polymer chains". In: *The Journal of Chemical Physics* 17.3 (1949), pp. 303–310.

[56] Wenwei Zheng et al. "Inferring properties of disordered chains from FRET transfer efficiencies". In: *Biophysical Journal* 114.3 (2018), 367a.

[57] JC Le Guillou and Jean Zinn-Justin. "Critical exponents for the n-vector model in three dimensions from field theory". In: *Physical Review Letters* 39.2 (1977), p. 95.

[58] M Muthukumar. "50th anniversary perspective: A perspective on polyelectrolyte solutions". In: *Macromolecules* 50.24 (2017), pp. 9528–9560.

[59] Andrey V Dobrynin and Michael Rubinstein. "Flory theory of a polyampholyte chain". In: *Journal de Physique II* 5.5 (1995), pp. 677–695.

[60] Alexei R Khokhlov, Alexander Yu Grosberg, and Vijay S Pande. *Statistical physics of macromolecules*. Vol. 1. Springer, 1994.

[61] Lev Davidovich Landau and Evgenii Mikhailovich Lifshitz. *Statistical Physics: Volume 5*. Vol. 5. Elsevier, 2013.

[62] Nithya Srinivasan et al. “Stimuli-sensitive intrinsically disordered protein brushes”. In: *Nature communications* 5.1 (2014), p. 5145.

[63] Christina Querfurth et al. “Circadian conformational change of the Neurospora clock protein FREQUENCY triggered by clustered hyperphosphorylation of a basic domain”. In: *Molecular cell* 43.5 (2011), pp. 713–722.

[64] M Beer, M Schmidt, and M Muthukumar. “The electrostatic expansion of linear polyelectrolytes: Effects of gegenions, co-ions, and hydrophobicity”. In: *Macromolecules* 30.26 (1997), pp. 8375–8385.

[65] Paul J Flory. “Molecular configuration of polyelectrolytes”. In: *The Journal of Chemical Physics* 21.1 (1953), pp. 162–163.

[66] Baoxu Liu et al. “The effect of intrachain electrostatic repulsion on conformational disorder and dynamics of the Sic1 protein”. In: *The Journal of Physical Chemistry B* 118.15 (2014), pp. 4088–4097.

[67] Gustav Emilsson et al. “Strongly stretched protein resistant poly (ethylene glycol) brushes prepared by grafting-to”. In: *ACS applied materials & interfaces* 7.14 (2015), pp. 7505–7515.

[68] Wei-Liang Chen et al. “50th anniversary perspective: Polymer brushes: Novel surfaces for future materials”. In: *Macromolecules* 50.11 (2017), pp. 4089–4113.

[69] Matthew R Dreher et al. “Temperature triggered self-assembly of polypeptides into multivalent spherical micelles”. In: *Journal of the American Chemical Society* 130.2 (2008), pp. 687–694.

[70] Yiyong He et al. “Self-assembly of block copolymer micelles in an ionic liquid”. In: *Journal of the American Chemical Society* 128.8 (2006), pp. 2745–2750.

[71] Isamu Akiba and Kazuo Sakurai. “Characterizing block-copolymer micelles used in nanomedicines via solution static scattering techniques”. In: *Polymer Journal* 53.9 (2021), pp. 951–973.

[72] Peter J Chung et al. “Direct force measurements reveal that protein Tau confers short-range attractions and isoform-dependent steric stabilization to microtubules”. In: *Proceedings of the National Academy of Sciences* 112.47 (2015), E6416–E6425.

[73] Pascale Barbier et al. “Role of tau as a microtubule-associated protein: structural and functional aspects”. In: *Frontiers in aging neuroscience* 11 (2019), p. 204.

[74] Tao Wu, Kirill Efimenko, and Jan Genzer. “Combinatorial study of the mushroom-to-brush crossover in surface anchored polyacrylamide”. In: *Journal of the american chemical society* 124.32 (2002), pp. 9394–9395.

[75] Lionel CH Moh, Mark D Losego, and Paul V Braun. “Solvent quality effects on scaling behavior of poly (methyl methacrylate) brushes in the moderate-and high-density regimes”. In: *Langmuir* 27.7 (2011), pp. 3698–3702.

[76] JF Joanny, L Leibler, and PG De Gennes. "Effects of polymer solutions on colloid stability". In: *Journal of Polymer Science: Polymer Physics Edition* 17.6 (1979), pp. 1073–1084.

[77] S Alexander. "Adsorption of chain molecules with a polar head a scaling description". In: *Journal De Physique* 38.8 (1977), pp. 983–987.

[78] Jürgen Rühe et al. "Polyelectrolyte brushes". In: *Polyelectrolytes with Defined Molecular Architecture I* (2004), pp. 79–150.

[79] Ekaterina B Zhulina and Michael Rubinstein. "Lubrication by polyelectrolyte brushes". In: *Macromolecules* 47.16 (2014), pp. 5825–5838.

[80] Philip Pincus. "Colloid stabilization with grafted polyelectrolytes". In: *Macromolecules* 24.10 (1991), pp. 2912–2919.

[81] Lauren Boldon, Fallon Laliberte, and Li Liu. "Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application". In: *Nano reviews* 6.1 (2015), p. 25661.

[82] Jens Als-Nielsen and Des McMorrow. *Elements of modern X-ray physics*. John Wiley & Sons, 2011.

[83] P Debye. "Molecular-weight determination by light scattering." In: *The Journal of Physical Chemistry* 51.1 (1947), pp. 18–32.

[84] Efstratios Mylonas and Dmitri I Svergun. "Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering". In: *Journal of applied crystallography* 40.s1 (2007), s245–s249.

[85] MS Smyth and JHJ Martin. "x Ray crystallography". In: *Molecular Pathology* 53.1 (2000), p. 8.

[86] Alexey G Kikhney and Dmitri I Svergun. "A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins". In: *FEBS letters* 589.19 (2015), pp. 2570–2577.

[87] Wenwei Zheng and Robert B Best. "An extended Guinier analysis for intrinsically disordered proteins". In: *Journal of molecular biology* 430.16 (2018), pp. 2540–2553.

[88] Uri Raviv et al. "Insight into structural biophysics from solution X-ray scattering". In: *Journal of Structural Biology* (2023), p. 108029.

[89] Bruno H Zimm. "The scattering of light and the radial distribution function of high polymer solutions". In: *The Journal of chemical physics* 16.12 (1948), pp. 1093–1099.

[90] Charlie Gosse and Vincent Croquette. "Magnetic tweezers: micromanipulation and force measurement at the molecular level". In: *Biophysical journal* 82.6 (2002), pp. 3314–3329.

[91] Bob M Lansdorp and Omar A Saleh. "Power spectrum and Allan variance methods for calibrating single-molecule video-tracking instruments". In: *Review of Scientific Instruments* 83.2 (2012).

[92] Noah Ribeck and Omar A Saleh. "Multiplexed single-molecule measurements with magnetic tweezers". In: *Review of Scientific Instruments* 79.9 (2008).

[93] David W Allan. "Statistics of atomic frequency standards". In: *Proceedings of the IEEE* 54.2 (1966), pp. 221–230.

[94] P Pincus. "Excluded volume effects and stretched polymer chains". In: *Macromolecules* 9.3 (1976), pp. 386–388.

[95] John F Marko and Eric D Siggia. "Stretching dna". In: *Macromolecules* 28.26 (1995), pp. 8759–8770.

[96] Claude Bouchiat et al. "Estimating the persistence length of a worm-like chain molecule from force-extension measurements". In: *Biophysical journal* 76.1 (1999), pp. 409–413.

[97] Yoav Lahini et al. "Nonmonotonic aging and memory retention in disordered mechanical systems". In: *Physical review letters* 118.8 (2017), p. 085501.

[98] Ariel Amir, Yuval Oreg, and Yoseph Imry. "On relaxations and aging of various glasses". In: *Proceedings of the National Academy of Sciences* 109.6 (2012), pp. 1850–1855.

[99] Jan Skov Pedersen. "Form factors of block copolymer micelles with spherical, ellipsoidal and cylindrical cores". In: *Journal of Applied Crystallography* 33.3 (2000), pp. 637–640.

[100] Jan Skov Pedersen and Carsten Svaneborg. "Scattering from block copolymer micelles". In: *Current opinion in colloid & interface science* 7.3-4 (2002), pp. 158–166.

[101] Heiko Ahrens et al. "Nonlinear osmotic brush regime: experiments, simulations and scaling theory". In: *The Journal of Physical Chemistry B* 108.43 (2004), pp. 16870–16876.

[102] N Arun Kumar and Christian Seidel. "Polyelectrolyte brushes with added salt". In: *Macromolecules* 38.22 (2005), pp. 9341–9350.

[103] AA Zamyatnin. "Protein volume in solution". In: *Progress in biophysics and molecular biology* 24 (1972), pp. 107–123.

[104] Youli Li et al. "Scatterless hybrid metal–single-crystal slit for small-angle X-ray scattering and high-resolution X-ray diffraction". In: *Journal of Applied Crystallography* 41.6 (2008), pp. 1134–1139.

[105] Nathan P Cowieson et al. "Beamline B21: high-throughput small-angle X-ray scattering at Diamond Light Source". In: *Journal of Synchrotron Radiation* 27.5 (2020), pp. 1438–1446.

[106] Clement E Blanchet et al. "Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY)". In: *Journal of applied crystallography* 48.2 (2015), pp. 431–443.

[107] Petra Pernot et al. “Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution”. In: *Journal of synchrotron radiation* 20.4 (2013), pp. 660–664.

[108] Adam Round et al. “BioSAXS Sample Changer: a robotic sample changer for rapid and reliable high-throughput X-ray solution scattering experiments”. In: *Acta Crystallographica Section D: Biological Crystallography* 71.1 (2015), pp. 67–75.

[109] J Kieffer et al. “New tools for calibrating diffraction setups”. In: *Journal of synchrotron radiation* 27.2 (2020), pp. 558–566.

[110] Pauli Virtanen et al. “SciPy 1.0: fundamental algorithms for scientific computing in Python”. In: *Nature methods* 17.3 (2020), pp. 261–272.

[111] Matthew Newville et al. “LMFIT: Non-linear least-square minimization and curve-fitting for Python”. In: *Astrophysics Source Code Library* (2016), ascl–1606.

[112] B Jacrot. “The study of biological structures by neutron scattering from solution”. In: *Reports on progress in physics* 39.10 (1976), p. 911.

[113] Simon F Nørrelykke and Henrik Flyvbjerg. “Power spectrum analysis with least-squares fitting: Amplitude bias and its elimination, with application to optical tweezers and atomic force microscope cantilevers”. In: *Review of Scientific Instruments* 81.7 (2010).

[114] Ian L Morgan and Omar A Saleh. “Tweezepy: A Python package for calibrating forces in single-molecule video-tracking experiments”. In: *Plos one* 16.12 (2021), e0262028.

[115] Andreas Möglich, Karin Joder, and Thomas Kiefhaber. “End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation”. In: *Proceedings of the National Academy of Sciences* 103.33 (2006), pp. 12394–12399.

[116] Pradip K Nandi and Dwight R Robinson. “Effects of salts on the free energy of the peptide group”. In: *Journal of the American Chemical Society* 94.4 (1972), pp. 1299–1308.

[117] Pradip K Nandi and Dwight R Robinson. “Effects of salts on the free energies of nonpolar groups in model peptides”. In: *Journal of the American Chemical Society* 94.4 (1972), pp. 1308–1315.

[118] A Quigley and DR Williams. “The second virial coefficient as a predictor of protein aggregation propensity: a self-interaction chromatography study”. In: *European Journal of Pharmaceutics and Biopharmaceutics* 96 (2015), pp. 282–290.

[119] Franziska Zosel et al. “Depletion interactions modulate the binding between disordered proteins in crowded environments”. In: *Proceedings of the National Academy of Sciences* 117.24 (2020), pp. 13480–13489.

[120] Janet Brownlees et al. “Charcot–Marie–Tooth disease neurofilament mutations disrupt neurofilament assembly and axonal transport”. In: *Human molecular genetics* 11.23 (2002), pp. 2837–2844.

10 Code

All the code below was written in Python. Form factors, models and analysis were all fitted to the appropriate data using the lmfit Python library [111]. All the appropriate parameters required for the fittings are in the tables of the appendix.

10.1 Analyses and models

Extended Guinier function

```
def exGuinlim(x, a, c, NN):
    """
    Input:  x - scattering vector q
            a - Zero concentration intensity I0
            c - Scaling exponent nu
            NN - Chain monomer number
    Output: Estimated scattering intensity by the extended Guinier approximation.
    """
    g = 1.1615
    b = np.sqrt(g * (g + 1) / (2 * (g + 2 * c) * (g + 2 * c + 1))) * 0.55 * NN ** c
    return np.log(a) - (1 / 3) * (b ** 2) * x + 0.0479 * (c - 0.212) * (b ** 4) * (x ** 2)
```

Salted polyelectrolyte model [64]

```
def k_model(x, C, N, b, nu, lb_):
    """
    Debye length (kappa) as a function of the radius of gyration Rg of the salted polyeelct
    Input:  x - Radius of gyration Rg
            C - Arbitrary fitting constant
            N - Chain monomer number
            b - Kuhn length
            nu - Extended volume
            lb_ - Bjerrum length
    Output: kappa(Rg) of the salted polyeelctrolyte model
    """
    Rg = x / (C * np.sqrt(N / 6) * b)
    k = 4 * np.pi * lb_ / ((Rg ** 5 - Rg ** 3) / (0.44 * np.sqrt(N)) - nu)
    return np.sqrt(k / b ** 3)
```

10.2 Form factors

Gaussian form factor

```
def _fa_coil(qrg):
    """
    Auxiliary function
    Input:  qrg - scattering vector q times radius
            of gyration Rg
    Output: Coil scattering amplitude function fa_coil: square root of the gaussian
            form factor
    """
    fa = np.ones(qrg.shape)
    fa[qrg != 0] = (1 - np.exp(-qrg[qrg > 0])) / (qrg[qrg > 0])
    qRg
    return fa
def gaussian_ff(x, scale, Rg):
    """
    Input:  x - Scattering vector q
            scale - Intensity scale
            Rg - Radius of gyration
    Output: scaled Gaussian form factor
    """
    q = np.atleast_1d(x)
    Q = np.where(q == 0, q * 0 + 1e-10, q)
    fa_RDA = _fa_coil(Rg * Q)
    return scale * fa_RDA ** 2
```

Spherical form factors

```
def _fa_sphere(qr):
    """
    Auxiliary function
    Input:  qrg - scattering vector q times sphere
            radius R
    Output: Spherical scattering amplitude
    """
    fa = np.ones_like(qr)
    qr1 = qr[qr > 0]
    fa[qr > 0] = 3 / qr1 ** 3 * (np.sin(qr1) - qr1 * np.cos(qr1))
```

```
return fa
def sphere_ff(x, scale, R):
    """
    Input:  x - Scattering vector q
            scale - Intensity scale
            R - Sphere radius R
    Output: Spherical form factor
    """
    q = np.atleast_1d(x)
    Q = np.where(q == 0, q * 0 + 1e-10, q)
    return _fa_sphere(Q * R) ** 2 * scale

def sphere_ff_poly(x, scale, R, FWHM, iterations):
    """
    Input:  x - Scattering vector q
            scale - Intensity scale
            R - Sphere radius R
            FWHM - Radius polydispersity
            iterations - Integration iterations
    Output: Polydisperse spherical form factor
    """
    q = np.atleast_1d(x)
    Q = np.where(q == 0, q * 0 + 1e-10, q)
    res = np.zeros(len(Q))
    dR = 1 / iterations
    Rn = np.linspace(R - 3 * FWHM, R + 3 * FWHM, iterations)
    for i in range(iterations):
        res += dR / (2 * np.pi * FWHM) * sphere_ff(Q, scale, Rn[i]) * np.exp(-0.5 * ((Rn[i]
return res * scale / res[0]
```

Ellipsoidal form factors

```
def ellipsoid(x, scale, R, epsilon):
    """
    Ellipsoid of semiaxis (R,R,epsilon*R)
    Input:  x - Scattering vector q
            scale - Intensity scale
            R - Semiaxis radius R
            epsilon - Secondary semiaxis scaling
            parameter
```

```
    Output: Scaled ellipsoid form factor
    """
    q = np.atleast_1d(x)
    Q = np.where(q == 0, q * 0 + 1e-10, q)
    int_it = 10
    res = np.zeros(len(Q))
    for i in range(len(Q)):
        res[i] = integration(lambda a: np.sin(a) * _fa_sphere(Q[i] * r_alpha(R, epsilon, a)
return res * scale
def epsiloid_ff_poly(x, L_R, scale, R, L, FWHM, iterations):
    """
    Input:  x - Scattering vector q
            L_R - Parameter polydispersity boolean.
            0: Polydispersity of R
            1: Polydispersity of epsilon
            scale - Intensity scale
            R - Semiaxis radius R
            L - Secondary semiaxis scaling
            parameter
            FWHM - Polydispersity
            iterations - Integration iterations
    Output: Polydisperse ellipsodial form factor
    """
    q = np.atleast_1d(x)
    Q = np.where(q == 0, q * 0 + 1e-10, q)
    res = np.zeros(len(Q))
    dR = 1 / iterations
    if L_R == 0:
        Ln = np.linspace(L - 3 * FWHM, L + 3 * FWHM, iterations)
        for i in range(iterations):
            res += dR / (2 * np.pi * FWHM) * ellipsoid_ff(Q, scale, R, Ln[i]) * np.exp(-0.5
    if L_R == 1:
        Rn = np.linspace(R - 3 * FWHM, R + 3 * FWHM, iterations)
        for i in range(iterations):
            res += dR / (2 * np.pi * FWHM) * ellipsoid_ff(Q, scale, Rn[i], L) * np.exp(-0.5
    return res * scale / res[0]
```

Cylindrical form factors

```
def cylinder_ff(x, scale, R, L):
```

```
    """
    Cylinder of radius R and length L form factor
    Input:  x - Scattering vector q
            scale - Intensity scale
            R - Cylinder radius R
            L - Cylinder length L
    Output: Scaled Cylindrical form factor
    """
    q = np.atleast_1d(x)
    Q = np.where(q == 0, q * 0 + 1e-10, q)
    int_it = 100
    res = np.zeros(len(Q))
    for i in range(len(Q)):
        res[i] = integration(lambda a: np.sin(a) * cylinder_(a, Q[i], R, L) ** 2, 0.001, np
                             int_it)
    return res * scale
def cylinder_ff_poly(x, L_R, scale, R, L, FWHM, iterations):
    """
    Input:  x - Scattering vector q
            L_R - Parameter polydispersity boolean.
            0: Polydispersity of R
            1: Polydispersity of L
            scale - Intensity scale
            R - Cylinder radius R
            L - Cylinder length L
            FWHM - Polydispersity
            iterations - Integration iterations
    Output: Polydisperse cylindrical form factor
    """
    q = np.atleast_1d(x)
    Q = np.where(q == 0, q * 0 + 1e-10, q)
    res = np.zeros(len(Q))
    dR = 1 / iterations
    if L_R == 0:
        Ln = np.linspace(L - 3 * FWHM, L + 3 * FWHM, iterations)
        for i in range(iterations):
            res += dR / (2 * np.pi * FWHM) * cylinder_ff(Q, scale, R, Ln[i]) * np.exp(-0.5
    if L_R == 1:
        Rn = np.linspace(R - 3 * FWHM, R + 3 * FWHM, iterations)
        for i in range(iterations):
```

```
            res += dR / (2 * np.pi * FWHM) * cylinder_ff(Q, scale, Rn[i], L) * np.exp(-0.5 *
    return res * scale / res[0]
```

Micelle form factors [99, 100]

```
def PG_spherical_gaussian(x, scale, beta, R, h, z, pos, solventSLD, d=1):
    """
    Spherical micelle form factor
    Input:  x - Scattering vector q
            scale - Intensity scale
            beta - beta parameter
            R - Sphere radius R
            h - brush height h
            z - aggregation number z
            pos - number of core residues per tail
            solventSLD - Scattering length density of the solvent
            d - Tail to core distance (default 1: tails directly stem from the core shell)
    Output: Spherical micelle form factor
    """
    sphere_SL, sphereSLD, coil_SL, coilSLD = SL_calc(pos)
    volu = coil_SL / coilSLD
    coilequR = (volu / (4 / 3. * np.pi)) ** (1 / 3.)
    q = np.atleast_1d(x)
    Q = np.where(q == 0, q * 0 + 1e-10, q)
    cg = coilSLD - solventSLD
    cs = sphereSLD - solventSLD
    coilVolume = (4 / 3. * np.pi * coilequR ** 3)
    fa_RDA = (_fa_coil(h * Q) ** 2 / (1 + beta * _fa_coil(h * Q) ** 2))
    fa_coil = fa_RDA
    res = (fa_coil / z) * (cg * coilVolume * z) ** 2
    f0 = (4 / 3. * np.pi * R ** 3 * cs)
    s_term = _fa_sphere(Q * R) ** 2
    sc_term = _fa_sphere(Q * R) * (np.sin(Q * (R + d * h)) / (Q * (R + d * h)))
    cc_term = (np.sin(Q * (R + d * h)) / (Q * (R + d * h))) ** 2
    res += s_term * f0 ** 2
    res += _fa_coil(h * Q) * sc_term * 2 * f0 * (cg * coilVolume * z)
    res += ((z - fa_RDA[0]) / z) * (fa_coil * cc_term) * (cg * coilVolume * z) ** 2
    return res * scale / (((cg * coilVolume * z) + f0) ** 2)

def PG_ellipsoid_corona(x, scale, R, epsilon, h, z, pos, solventSLD, d=1):
```

```
"""
Ellipsodial micelle form factor
Input:  x - Scattering vector q
        scale - Intensity scale
        R - Semiaxis radius R
        epsilon - Secondary semiaxis scaling
        parameter
        h - brush height h
        z - aggregation number z
        pos - number of core residues per tail
        solventSLD - Scattering length density of the solvent
        d - Tail to core distance (default 1: tails directly stem from the core shell)
Output: Ellipsodial micelle form factor
"""
sphere_SL, sphereSLD, coil_SL, coilSLD = SL_calc(pos)
volu = coil_SL / coilSLD
coilequR = (volu / (4 / 3. * np.pi)) ** (1 / 3.)
q = np.atleast_1d(x)
Q = np.where(q == 0, q * 0 + 1e-10, q)
cg = coilSLD - solventSLD
cs = sphereSLD - solventSLD
coilVolume = (4 / 3. * np.pi * coilequR ** 3)
fa_RDA = _fa_coil(h * Q)
fa_coil = fa_RDA
res = fa_coil ** 2 / z * (cg * coilVolume * z) ** 2
f0 = (4 / 3. * np.pi * epsilon * R ** 3 * cs)  # forward scattering Q=0
s_term = np.zeros(len(Q))
sc_term = np.zeros(len(Q))
cc_term = np.zeros(len(Q))
int_it = 10
for i in range(len(Q)):
    s_term[i] = integration(lambda a: np.sin(a) * _fa_sphere(Q[i] * r_alpha(R, epsilon,
                            int_it)
    sc_term[i] = integration(
        lambda a: np.sin(a) * _fa_sphere(Q[i] * r_alpha(R, epsilon, a)) *
                  np.sin(Q[i] * (r_alpha(R, epsilon, a) + d * h)) / (Q[i] * (r_alpha(R,
        np.pi / 2, int_it)
    cc_term[i] = integration(lambda a: np.sin(a) * (np.sin(Q[i] * (r_alpha(R, epsilon,
res += s_term * f0 ** 2
res += fa_coil * sc_term * 2 * f0 * (cg * coilVolume * z)
```

```
    res += (z - 1) / z * (fa_coil ** 2 * cc_term) * (cg * coilVolume * z) ** 2

    return scale * res * 1 / (((cg * coilVolume * z) + f0) ** 2)

def PG_cylindrical_corona(x, scale, R, L, beta, z, h, pos, solventSLD, d=1):
    """
    Cylindrical micelle form factor
    Input:  x - Scattering vector q
            scale - Intensity scale
            R - Cylinder radius R
            L - Cylinder length L
            h - brush height h
            z - aggregation number z
            pos - number of core residues per tail
            solventSLD - Scattering length density of the solvent
            d - Tail to core distance (default 1: tails directly stem from the core shell)
    Output: Cylindrical micelle form factor
    """
    pass1, sphereSLD, coil_SL, coilSLD = SL_calc(pos)
    volu = coil_SL / coilSLD
    coilequR = (volu / (4 / 3. * np.pi)) ** (1 / 3.)
    q = np.atleast_1d(x)
    Q = np.where(q == 0, q * 0 + 1e-10, q)
    cg = coilSLD - solventSLD
    coilVolume = (4 / 3. * np.pi * coilequR ** 3)
    fa_RDA = (_fa_coil(h * Q) / (1 + beta * _fa_coil(h * Q)))
    fa_coil = coilVolume * cg * fa_RDA
    cs = sphereSLD - solventSLD
    f0 = (np.pi * L * R ** 2 * cs)  # forward scattering Q=0
    s_term = np.zeros(len(Q))
    sc_term = np.zeros(len(Q))
    cc_term = np.zeros(len(Q))
    int_it = 10
    for i in range(len(Q)):
        s_term[i] = f0 ** 2 * integration(lambda a: np.sin(a) * cylinder_(a, Q[i], R, L) **
        sc_term[i] = f0 * integration(
            lambda a: np.sin(a) * cylinder_(a, Q[i], R, L) * c_s_ff(a, Q[i], R + d * h, L +
            np.pi / 2, int_it)
        cc_term[i] = integration(
            lambda a: np.sin(a) * c_s_ff(a, Q[i], R + d * h, L + 2 * d * h) ** 2, 0.001,
```

```
        np.pi / 2, int_it)

res = z * fa_coil ** 2
res += s_term
res += 2 * z * fa_coil * sc_term
res += z * (z - fa_RDA[0]) * (fa_coil ** 2 * cc_term)
return res * scale
```

Eur. Phys. J. E (2024)47:13
https://doi.org/10.1140/epje/s10189-024-00409-8

THE EUROPEAN PHYSICAL JOURNAL E

Regular Article - Soft Matter

From isolated polyelectrolytes to star-like assemblies: the role of sequence heterogeneity on the statistical structure of the intrinsically disordered neurofilament-low tail domain

Mathar Kravikass[1,2], Gil Koren[1,2], Omar A. Saleh[3,4], and Roy Beck[1,2,a]

[1] School of Physics and Astronomy, The Center for Nanoscience and Nanotechnology, Tel Aviv University, Tel Aviv, Israel
[2] The Center of Physics and Chemistry of Living Systems, Tel Aviv University, Tel Aviv, Israel
[3] Materials Department, Biomolecular Sciences and Engineering Program, University of California, Santa Barbara, USA
[4] Physics Department, University of California, Santa Barbara, USA

Received 26 November 2023 / Accepted 2 February 2024

Abstract Intrinsically disordered proteins (IDPs) are a subset of proteins that lack stable secondary structure. Given their polymeric nature, previous mean-field approximations have been used to describe the statistical structure of IDPs. However, the amino-acid sequence heterogeneity and complex intermolecular interaction network have significantly impeded the ability to get proper approximations. One such case is the intrinsically disordered tail domain of neurofilament low (NFLt), which comprises a 50 residue-long uncharged domain followed by a 96 residue-long negatively charged domain. Here, we measure two NFLt variants to identify the impact of the NFLt two main subdomains on its complex interactions and statistical structure. Using synchrotron small-angle x-ray scattering, we find that the uncharged domain of the NFLt induces attractive interactions that cause it to self-assemble into star-like polymer brushes. On the other hand, when the uncharged domain is truncated, the remaining charged N-terminal domains remain isolated in solution with typical polyelectrolyte characteristics. We further discuss how competing long- and short-ranged interactions within the polymer brushes dominate their ensemble structure and, in turn, their implications on previously observed phenomena in NFL native and diseased states.

Introduction

Intrinsically disordered proteins (IDPs) are a subset of proteins that, instead of forming a rigid singular structure, fluctuate between different conformations in their native form [1,2]. Nonetheless, IDPs serve significant biological functions and account for about 44% of the human genome [3]. The lack of fixed structure provides IDPs many advantages in regulatory systems in which they often play a crucial role in mediating protein interaction [4,5]. These roles often come into play from intrinsically disordered regions (IDRs) of folded proteins interacting with other IDRs. For example, in the neurofilament proteins, tails emanating from the self-assembled filament backbone domains bind together and form a network of filaments [6–10].

The ensemble statistics of IDPs stem from their sequence composition and the surrounding solution [2]. For example, previous studies showed that IDPs comprising mostly negatively charged amino acids (polyelectrolytes) are locally stretched due to electrostatic repulsion between the monomers [11]. Moreover, different properties, such as hydrophobicity, were shown to be linked with local IDP domain collapse [12]. The complex interactions that arise from sequence heterogeneity allow IDPs to form specific complexes without losing their disordered properties [13]. For example, Khatun et al. recently showed how, under limited conditions, the human amylin protein self-assembles into fractal structures [14].

As IDPs are disordered chains, polymer theories are prime candidates to relate the measured structural statistics to known models, which can help link the sequence composition of the IDP to its conformations [15–18]. Specifically, polymer scaling theories allow us to derive the statistical structure of IDPs given sequence-derived parameters, such as charge density and hydrophobicity [11,12,19–21]. However, due to the heterogeneity of the IDP primary structure (i.e., the amino-acid sequence), some systems showed contradictions with the behavior theorized by standard heterogeneous polymer physics [17,19,22–24].

The unique biological properties of IDPs have given rise to numerous attempts to use them as building

[a] e-mail: roy@tauex.tau.ac.il (corresponding author)

Published online: 15 February 2024

blocks for self-assembled structures [25]. For example, IDPs were proposed as brush-like surface modifiers, due to their enhanced structural plasticity to environmental conditions [26,27]. Another example of an IDP brush system is the neurofilament (NF) protein system [6,28,29], described as interacting bottle-brushes. NF subunit proteins form mature filaments with protruding disordered C-terminus IDR known as 'tails.' NF tails were shown to mediate NF network formation and act as shock absorbents in high-stress conditions [29]. Moreover, NF aggregations are known to accumulate alongside other proteins in several neurodegenerative diseases, such as Alzheimer's and Parkinson's. [30].

The NF-low disordered tail domain (NFLt) sequence can be divided into two unique regions: an uncharged region (residues 1–50) starting from its N-terminal and a negatively charged region (residues 51–146). The NFLt can be described as a polyelectrolyte with a net charge per residue (NCPR) of −0.24. Furthermore, the statistical structures of segments within the NFLt are influenced by the amount, type, and disperse of the charged amino acid within a segment [22]. Nonetheless, other structural constraints, particularly long-range contacts, impact the local statistical structures. Additionally, NFLt was shown to have glassy dynamics with the response to tension [31]. Such dynamics were associated with multiple weakly interacting domains and structural heterogeneity.

In this paper, we revisit NFLt as a model system for charged IDP and focus on the contribution of its neutral and hydrophobic N-terminal domain. We will show that increased salt concentration causes NFLt to form star-like brushes with increased aggregation number (Z). Here, we are motivated by theoretical models, in particular the Pincus' model for salted polyelectrolytes [32], that capture key physical properties of IDPs, including the model system presented here [26,29,33]. We will further quantify the competition between hydrophobic attraction and electrostatic and steric repulsion in the formation of the structures of NFLt.

Results

To study the N-terminal domain contribution to the structure of NFLt, we designed two variants and measured them at various buffer conditions. The first construct is the entire 146 residues of the NFLt chain, which we term as WT (NCPR = −0.24), and the second is isolating the 104 negatively charged residues from the C-terminal of NFLt (NCPR = −0.33), termed as ΔN42. We expressed the variants in *E-coli* and purified it up to 96% (see methods).

We assessed the variants in solution using small-angle X-ray scattering (SAXS), a technique extensively used to characterize the statistical structures of IDPs [34]. From the raw SAXS data, measured at various salinities, we can already find high structural differences between the two variants (Fig. 1a). Dominantly at the low wave-vector (q) region, the WT variant scattering

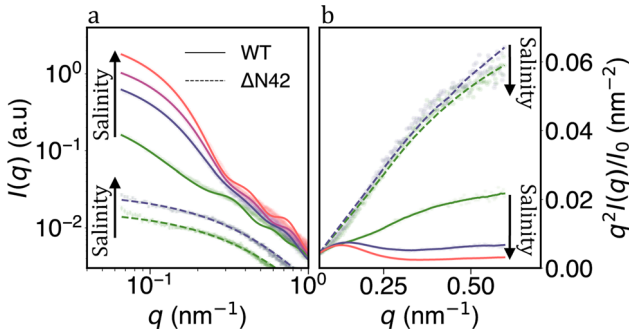

Fig. 1 SAXS measurements of WT and ΔN42 at different salinity (C_s). **a** For increasing C_s, the WT variant shows increased small-angle scattering, a signature for aggregation. In contrast, ΔN42 remains structurally intrinsically disordered as C_s vary. Data points are shifted for clarity. Lines are form-factor fittings, as described in the text. **b** Normalized Kratky plot of the same SAXS measurements. The ΔN42 variant remains disordered and unchanged with salinity, while the WT variant shows a hump at low q, typical for a collapse region. With increasing C_s, the hump at the lower q range becomes a sharper peak accompanied by a scattering rise at the higher q range. Such behavior indicates that the aggregation coexists with the WT variant's highly dynamic and disordered regions. Both variants shown are at the highest measured concentration (Table S1, S3). WT measurements are in 20 mM Tris pH 8.0 with 0, 150, 250, and 500 mM added NaCl (from bottom to top). Likewise, for ΔN42, measurements are in 20 mM Tris pH 8.0 with 0 and 150 mM added NaCl (bottom to top)

(I) rises with added NaCl salt. Such an increase at low q implies high molecular mass particles due to aggregation of the WT variant.

In contrast, ΔN42 shows a separated Gaussian polymer profile (Figs. 1a, S1), nearly insensitive to total salinity (C_s = 20 − 520 mM). Similarly, the data presented in Kratky format (qI^2 vs. q, Fig. 1a) show the ΔN42 has the signature of a disordered polymer. In contrast, the WT variant, in particular at high salinity, has a combination of a collapse domain (the peaks from below $q = 0.25\text{nm}^{-1}$) and a disordered polymeric structure (the scattering rise at higher q Fig. 1a).

Being completely disordered, ΔN42 lacks a stable structure and can be described using a statistical ensemble of polymeric conformations [35] were:

$$I(q) = I_0 \exp\{-\frac{1}{3}(qR_G)^2 + 0.0479(\nu - 0.212)(qR_G)^4\}. \quad (1)$$

Here, I_0 is the scattering at $q = 0$, ν is Flory scaling exponent, and R_G is the radius of gyration defined by:

$$R_G = \sqrt{\frac{\gamma(\gamma+1)}{2(\gamma+2\nu)(\gamma+2\nu+1)}} bN^{\nu}, \quad (2)$$

where $\gamma = 1.615$ and $b = 0.55$nm (see [35]) and the analysis is viable up to $qR_G \sim 2$ (Fig. S2, S3). In all

ΔN42 cases, the scattering profile fits Eq. 1 and with ν ranging between 0.63–0.69 depending on the buffer salinity (Table S1). In 'infinite dilution' conditions (zero polymer concentration), we find ν to decrease monotonically from 0.73 to 0.62 with added salt (Table S2).

Given the noticeable aggregation for the WT variant, alternative form factors were considered to match the scattering profiles (lines in Fig. 1). The absence of structural motifs at high q values ($q > 0.3$ nm^{-1}) indicates a disordered nature for WT at shorter length scales. Conversely, in the lower q region ($q < 0.3$ nm^{-1}), the scattering suggests stable structural motifs or a larger molecular weight particles. Such SAXS resembles that of self-assembled decorated spherical micelles [36]. Variations of micelle models are shown to fit the data (Figs. 1, S4–S6). Sufficiently low aggregation number and core size distill the description of the spherical micelle into a 'star-like' brush. Alternative attempts to fit the scattering profiles to other form factors models, including vesicles and lamellar, were unsuccessful.

For the star-like model, the aggregated variants form a small spherical core of volume V_{core} made out of $n \cdot Z$ monomers (comparison with different cores described in [37] and in Fig. S4), where n denotes the peptide length per polypeptide within the core, and Z is the aggregation number, i.e. the number of polypeptides per 'star.' The remainder of the WT variant then protrudes from the core as the star polymer brush (Figs. 2a, S4–S6).

The star-like scattering form factor is described as a combination of four terms [36]: the self-correlation term of the core F_{c}, the self-correlation term of the tails F_{t}, the cross-correlation term of the core and the tails S_{ct} and the cross-correlation term of the tails S_{tt}:

$$F_{\mathrm{total}}(q) = Z^2\beta_{\mathrm{c}}^2 F_{\mathrm{c}}(q) + Z\beta_{\mathrm{t}}^2 F_{\mathrm{t}}(q) + 2Z^2\beta_{\mathrm{c}}\beta_{\mathrm{t}} S_{\mathrm{ct}}(q) + Z(Z-1)\beta_{\mathrm{t}}^2 S_{\mathrm{tt}}(q). \quad (3)$$

Here, β_{c} and β_{t} are the excess scattering length of the core and the tails, respectively. From fitting the scattering data, we extracted the height of the tails $h = 2R_{\mathrm{G}}$, the aggregation number Z, and the relevant core's parameters (e.g., core radius R for a spherical core, cylinder radius R and length L for a cylindrical core [37]), schematically illustrated in Fig. 2a. All fitting parameters are found in Table S3.

To avoid misinterpretation and to minimize intermolecular interaction effects, we present the fitting results at the 'infinitely diluted regime' by extrapolating the relevant parameters measured at various protein concentrations to that at zero protein concentration (Fig. S7, Table S4). The parameters are mostly independent of the concentration unless explicitly mentioned.

At low salinity (20 mM), the aggregation number for the WT variant is of a dimer ($Z \approx 2$), and the core's shape is that of a cylinder (with a radius $R = 0.89$ nm and length $L = 1.19$ nm). At higher salt conditions (170–520 mM), the form factor fits spherical core aggregates with increasingly higher Z's (Fig. 2a).

Given the relatively small core volume ($V_{\mathrm{core}} \approx 1 - 2\mathrm{nm}^3$, Fig. 2c), it is crucial to evaluate the 'grafting' distance between neighboring chains, ρ, on the core surface ($S = 4\pi R^2 = Z\rho^2$) and the brush extension, h, outside the core. As shown in Fig. 2d, in all cases, $h/\rho \gg 1$ indicates a 'brush regime' where neighboring chains repel each other while extending the tail's height [38].

The repulsion between the grafted tail is further emphasized when comparing $h/2$ for WT to the equivalent ΔN42 length-scale (R_{G}), showing a significant extension for WT (Fig. 2e). We notice that the WT tail's length (h) increases at low salt (during the transitions from a dimer to a trimer), followed by a steady mild decrease as the C_s, and following Z increase. Similar compactness with increasing C_s is shown for ΔN42 and is expected for polyelectrolyte due to the reduction in electrostatic repulsion [39]. To better compare the statistical structure of two variants of disordered regions, we followed the polymeric scaling notation ν that quantifies the compactness of the chain. For ΔN42, we extracted ν from Eqs. 1 and 2 and found a significant decrease in its value as 50 mM of NaCl is added to the 20 mM Tris buffer (Fig. 3a). The following monotonic decline is in line with polyelectrolytic models and electrostatic screening effects [40], shown in a solid red line in Fig. 3a. Interestingly, previous measurements of segments within the NFLt charged domain were shown to have similar ν values as in ΔN42 . However, the same decline in salinity was not observed (Fig. 3a) [22].

For the WT variant, the scaling factor (ν) of the 'star-like polymer' brushes is extracted from Eq. 2. Here, we use $R_{\mathrm{G}} = h/2$, where h is obtained from Eq. 3. For $C_s = 20$ mM, we find that ν is of similar scale as for ΔN42 . This similarity can be attributed to the nature of the dimer, where the intramolecular electrostatic interactions dominate the expansion of each of the two tails. As C_s increases by 150 mM, ν exhibits a considerable increase, presumably due to neighboring tail repulsion. Above $C_s = 170$ mM, ν shows a weak decrease. We attribute this weak decline to the salt-brush regime of polyelectrolyte brushes [41] shown in solid blue in Fig. 3a. In this regime, $h \propto C_s^{-1/3}$, and subsequently $\nu \propto -\frac{1}{3} log(C_s)$.

We note that the cores of the star-like polymers are relatively small and that each polypeptide aggregates through only a few, most likely hydrophobic, amino acids. From the tabulated amino-acid partial volume, $\langle\phi_{aa}\rangle$ [42], we estimate the comprising amino acids as spheres of volume $\langle\phi_{aa}\rangle$. From here, the average number of amino acids per polypeptide inside the core is estimated by the number of spheres that can fit within the core volume, divided by the aggregation number: $n = V_{\mathrm{core}}/(\langle\phi_{aa}\rangle \cdot Z)$. Noticeably, our fit results with small n values, ranging between $\sim 7 - 2$ residues on average within the aggregate ensemble and depending on the buffer salinity. Attempting to 'fix' n to a larger constant residue per tail number results in a poorer fitting (Fig. S9). In Fig. 3a, we indeed see that the most significant change occurs at the low salt regime,

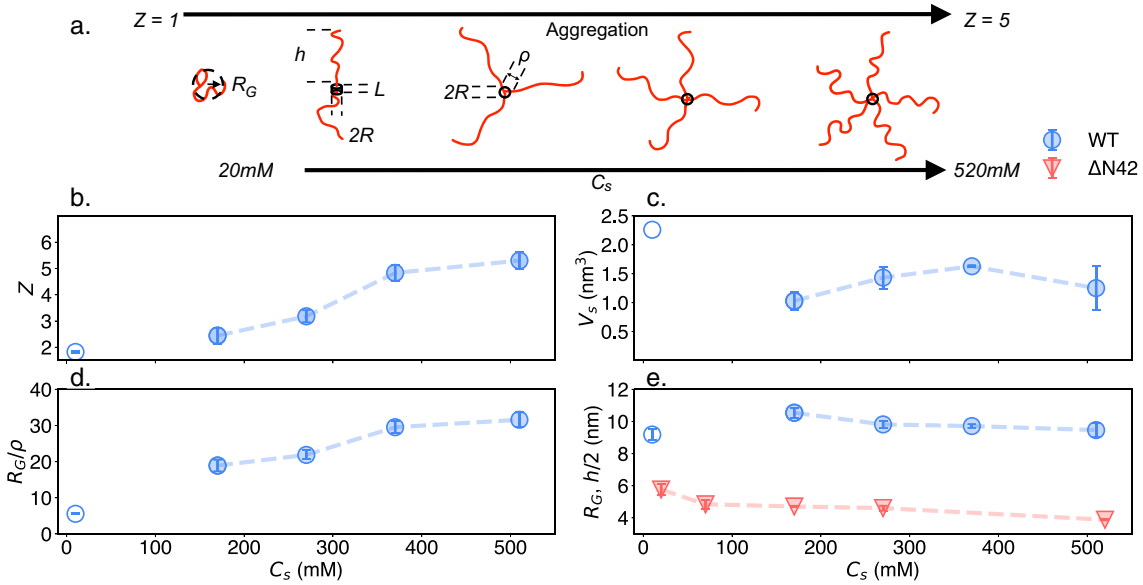

Fig. 2 **a** Schematic of the system's structure variation with salinity (C_s). While ΔN42 remains disordered and segregated, the WT variant aggregates to a star-like polymer with a higher aggregation number at higher C_s. **b**–**e** Structural parameters for WT (blue symbols) and ΔN42 (red symbols) variants extracted from fitting the SAXS data. Full and hollow circles represent the spherical and cylindrical core fitted parameters, respectively. **d** In all cases, the brush heights (h) are much larger than the corresponding grafting length (ρ), indicative of a brush regime. **e** The structurally intrinsically disordered ΔN42 variant compacts with higher C_s values and remains more compacted from the projected brushes for the WT variant. All values are the extrapolated 'zero concentration' fitting parameters (see Fig. S7)

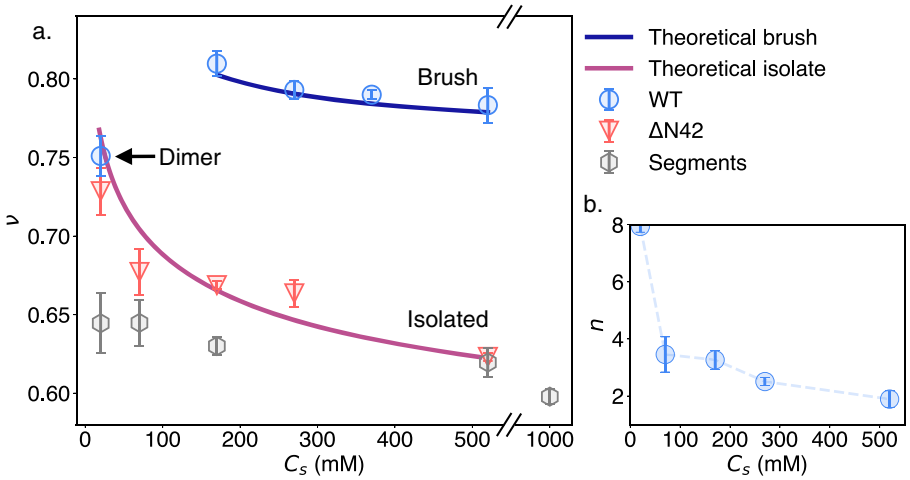

Fig. 3 Deduced structural parameters from the SAXS data fitting. **a** Flory exponent (ν) of WT tails and ΔN42 variants showing extended disordered scaling. The red line refers to the theoretical brush model [41], and the blue line refers to the theoretical polyelectrolyte [40]. ΔN42 shows a decrease in the protein extension due to the decline in intermolecular electrostatic repulsion (see also Fig. 4). WT shows an increase in the extension when shifting from a dimer to a trimer, followed by a slight decline with a further increase in salinity. In gray, average ν is obtained from measuring separate NFLt segments with an NCPR of −0.3 to −0.6 [22]. **b** The core (aggregated) peptide length per polypeptide as a function of salinity. At high salinity, each polypeptide aggregates via 2–3 amino acids that form the star-like polymer core. Both panels' values are the extrapolated 'zero concentration' parameters (supplementary Fig. S8)

where n drops from an average of 7 to 3 amino acids ($C_s = 20,170$ mM, respectively). Such behavior is known to occur within globular proteins [43] and was recently alluded to impact IDPs [44]. The following trend is a further decrease in n, albeit much weaker, which results in a final average n of about two as the salinity reaches $C_s = 520$ mM.

Last, in Fig. 4, we quantify the intermolecular interactions by evaluating the second virial coefficient, A_2, using a Zimm analysis [45] (Table S5). Here, A_2 describes the deviation of the statistical ensemble from an ideal gas. In agreement with our previous data, we find that the inter-molecular interactions of ΔN42 change from repulsive ($A_2 > 0$) to weakly attractive ($A_2 \leq 0$) as the salinity increases. In contrast, for WT, A_2 changes from a nearly neutral state of intermolecular interactions (i.e., ideal gas regime) to mildly attractive ($A_2 < 0$). These findings are reflected in the dependency of the variant Flory coefficient ν in concentration. While at the lowest salinity, ΔN42 is shown to expand as protein concentration is decreased, for higher salinities and for the WT measurements, ν remain primarily unchanged (Fig. S8a).

Combining our results for both variants, we find an exemplary role of long-range electrostatic interactions tuning the statistical structure of IDPs. Without the uncharged N-terminal domain, the NFLt exhibited significant change as the electrostatic interactions were screened, causing them to condense further. In contrast, the presence of the uncharged domain incurred aggregation of the proteins, bringing the tails much closer to each other. The increase in proximity was reflected in a significant increase in the expansion compared to the truncated variant, which exhibited a much weaker contraction with salinity.

Discussion and Conclusions

We investigated the effects of sequence heterogeneity on the interactions and structures of NFLt, an IDP model system. For NFLt, the N-terminal region consisting of the first ∼ 50 residues is hydrophobic and charge neutral, while the remaining chain is highly charged. We found that the sequence heterogeneity differentiates between the structures of the entire WT NFLt and a variant lacking the N-terminal domain. In particular, the WT variant self-assembles into star-like structures, while the ΔN42 one remains isolated in all measured cases.

Since ΔN42 can be attributed as a charged polymer, weakly attractive interactions take center stage as the electrostatic repulsion diminishes with charge screening (Fig. 4). These interactions could be attributed to monomer–monomer attractions that arise from the sequence heterogeneity of the IDP, such as weak hydrophobic attraction from scattered hydropathic sites [22,28,29,46–48].

For the WT variant, the intermolecular interactions started from a near-neutral state and transitioned to weakly attractive. However, as the WT measurements describe self-assembling complexes, the interpretation of these results differs from ΔN42. As such, we interpret the intermolecular interactions as the 'aggregation propensity,' the protein complex's growing ability. The aggregation propensity grows as the attractivity between the complex and the other polypeptides in the solution increases. This behavior can be observed when examining the responsiveness of the aggregation number Z to protein concentration C (Fig. S7). In the lowest measured screening, Z dependency on protein concentration was minimal. As we increase the screening effects, this dependency becomes more substantial. This characterization is also found in folded proteins, where intermolecular interactions were shown to indicate aggregation propensity [49]. The increased intermolecular attraction induced at increasing salinity is indicative of a salting-out phenomenon [50,51], although further investigation at higher salinity is needed.

The stability of the star-like polymer core should be evaluated by the participating residues per polypeptide (n). Indeed, while our fittings result with rather small n values, the SAXS signal at low q is dominated with aggregated structures under all salinity conditions. Within the occurring hydrophobic interactions, the release of bound water molecules and ions from the polypeptides is likely to contribute to the core's stability. Such entropic-based effects have been observed in similar processes such as protein flocculation [52,53] and in temperature-specific IDP binding modulation [54].

In our previous study [22], Flory exponents (ν) of shorter segments from the same NFLt were measured independently and in the context of the whole NFLt using SAXS and time-resolved Förster resonance energy transfer (trFRET). There, regardless of the peptide sequence, in the context of the entire NFLt, the segments' structural statistics were more expanded (*i.e.,* with larger ν values) than when measured independently. Similarly, these short segments measured with SAXS have smaller ν values (i.e., with a compacted statistical structure) than those of measured here for ΔN42 in all salt conditions (Fig. 3a, gray symbols).

The expansion of segments in the context of a longer chain corroborates that long-range contacts contribute to the overall disordered ensemble [22]. Interestingly, at $C_s = 520$ mM salinity, we found similar ν values of the ΔN42 and the previous short segment measurements, indicating a comparable expansion. We suggest that at higher salinities, the significance of electrostatic long-range contacts diminishes, aligning the expansion 'scaling laws' regardless of the chain length. Importantly, comparisons between our ΔN42 variant results (and not to the WT variant) and the previous segments' measurements are more suitable as the chains did not aggregate in those cases.

Compared to ΔN42, WT exhibits a mild contraction in salt, resembling the behavior of the 'salt-brush' regime observed in polyelectrolyte brushes, as demonstrated in Fig. 3. Similar salt-brush behavior was

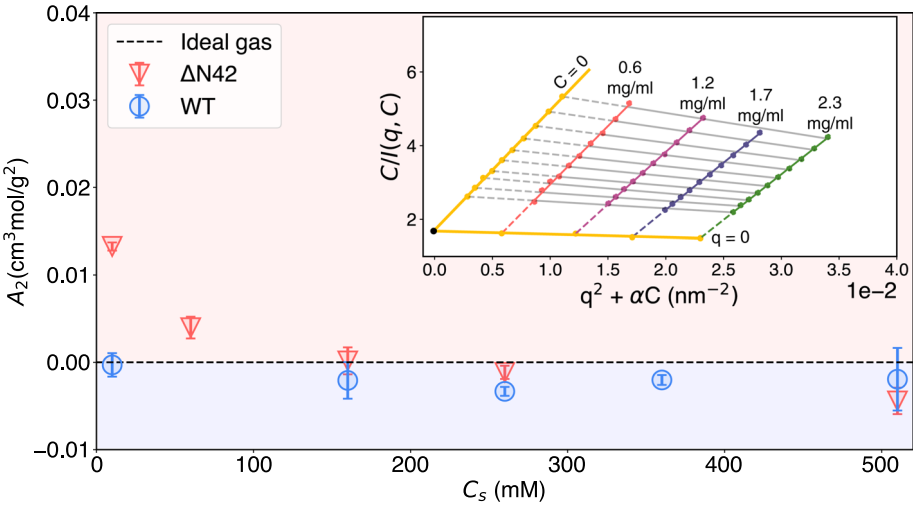

Fig. 4 The osmotic second virial coefficient A_2 as a function of the two variants' salinity (C_s). ΔN42 intermolecular interactions transition from repulsive to attractive as C_s increases. WT changes from a nearly neutral state of intermolecular interactions to attractive. Inset: A demonstration (WT variant, 20 mM Tris and 500 mM NaCl pH 8.0) for the Zimm analysis used to extract A_2 from SAXS data measured at various protein concentrations (C). Values shown in the graph are in mg/ml units. The dashed lines show the extrapolation from the measured data (colored lines) to the fitted $q \to 0$ and $C \to 0$ yellow lines, where $\alpha = 0.01$ is an arbitrary constant used in this analysis

previously observed in neurofilament-high tail domain brushes grafted onto a substrate [26], and in a recent polyelectrolytic brush scaling theory [55]. In the salt-brush regime, Pincus showed that brush mechanics resemble neutral brushes, determined by steric interchain interactions [32]. In this interpretation, the effective excluded volume per monomer enlarges and is proportional to $1/\kappa_s^2$, where κ_s is the Debye length attributed to the added salt. Consequently, we suggest that the heightened charge screening in the WT solution allows steric interactions between brushes to play a more significant role in determining the brush ensemble. Additionally, we deduce that the increased prevalence of steric repulsion counteracts the attractive forces responsible for aggregation, thereby preventing brush collapse.

The NFLt contraction aligns with previous studies of native NFL hydrogel networks [28,29]. At high osmotic pressure, the NFL network showed weak responsiveness to salinity higher than $C_s = 100$ mM, in agreement with theory [55]. With the observed salt-brush behavior for WT, we suggest that weak salt response in NFL hydrogels coincides with the increase in steric repulsion shown for the star-like structures (Fig. 3a, blue line).

Additionally, our measurements show that the hydrophobic N-terminal regime of the NFLt domain aggregates. This result is consistent with the findings of Morgan et al. [31], where single-molecule pulling experiments were performed on WT NFLt, and slow aging effects were observed, likely due to collapse (and potential aggregation) of the neutral domain. Indeed, follow-up studies by Truong et al. [56] used single-molecule stretching to show that added denaturant led to a swelling of the chain (increased ν), demonstrating that the WT chain has hydrophobic aggregation that can be disrupted by the denaturant. These observations suggest that at higher salt, the loss of repulsion may lead to attractive hydrophobic interactions growing more prominent in the NFL network. However, the steric repulsion from the remaining NFL tail may shield such an unwanted effect. Nonetheless, such effects may grow more prominent as the native filament assembly is disrupted.

In summary, we showed how the sequence composition of the NFLt IDP caused structural deviation from a disordered polyelectrolyte to a self-assembled star-like polymer brush. Together with the self-regulatory properties of the brushes, such behavior can be exploited to design structures that can resist specific environmental conditions. Additionally, our results showed possible implications on NFL aggregates that could shed light on the underlying correlations between the complex structure and the conditions driving it. While IDPs resemble polymers in many aspects, as we showed here, it is critical to assess their sequence to distinguish where and how to use the appropriate theoretical arguments to describe their statistical properties and structure.

Methods

Protein purification

Protein purification followed Koren et al. [22]. Variant ΔN42, included two cysteine residues at the C- and N terminals. After purification, ΔN42 variants were first reduced by 20 mM 2-Mercaptoethanol. Next, 2-Mercaptoethanol was dialyzed out with 1 L of 50 mM HEPES at pH 7.2. To block the cysteine sulfhydryl

group, we reacted ΔN42 variants with 2-Iodoacetamide at a molar ratio of 1:20. At the reaction, the variants' concentrations were ∼2 mg/ml. The reaction solution was kept under dark and slow stirring for 5 hr and stopped by adding 50 mM 2-Mercaptoethanol followed by overnight dialysis against 1 L of 20 mM Tris at pH 8.0 with 0.1% 2-Mercaptoethanol. Final purity was >95% as determined by SDS-PAGE (Fig. S10).

SAXS measurement and analysis

Protein samples were dialyzed overnight in the appropriate solution and measured with a Nanodrop 2000 spectrophotometer (Thermo Scientific) for concentration determination. Buffers were prepared with 1 mM of TCEP to reduce radiation damage and 0.2% of Sodium Azide to impair sample infection. The samples were prepared in a final concentration of 2 mg/ml, measured in a series of 4 dilutions. Preliminary measurements were measured at Tel-Aviv University with a Xenocs GeniX Low Divergence CuKα radiation source setup with scatterless slits [57] and a Pilatus 300K detector. All samples were measured at three synchrotron facilities: beamline B21, Diamond Light Source, Didcot, UK [58], beamline P12, EMBL, DESY, Hamburg, Germany [59], and beamline BM 29 ESRF, Grenoble, France [60]. Measurements at ESRF were taken using a robotic sample changer [61].

Integrated SAXS data were obtained from the beamline pipeline and 2D integration using the "pyFAI" Python library [62]. Extended Guinier analyses for the ΔN42 variant were done with the "curve_fit" function from the "Scipy" Python library [63]. To extract R_g and ν, extended Guinier analysis was conducted for $0.7 < qR_g < 2$. Error calculation was done from the covariance of the fitting.

Model fittings for the WT variant were done using the "lmfit" Python library [64] using the model described in [36,37]. Due to the complexity of the model, cylindrical core fittings were done by binning the data in 100 logarithmic bins to reduce computation time. Within the same model, core parameters (cylinder radius R and cylinder length L) were set constant, to offset fitting errors. Initial values of R and L were calculated with the highest measured concentration. Physical boundary conditions were imposed on the fitting, and scattering length (SL) values were set to be unchanged by the fitting process. SL values of both the core and the tail domains were determined by tabulated values of amino-acid SLD in 100% H_2O [65] (Table S3). Fitting parameter error evaluation was done by finding the covariant of the returning fitting parameters. Error calculation of the volume was done using: $\frac{dV}{V} = \sqrt{3\left(\frac{dR}{R}\right)^2}$. In addition, ν values of WT were found by a recursive search of the corresponding tail height $h/2$ over Eq. 2. Errors of ν were then found by assuming a simple case of $R_g = bN^{\nu}$, from which $d\nu \sim \frac{\ln(1+dR/R)}{\ln N} \sim \ln(N)^{-1}\frac{dR}{R}$

Zimm analysis

Zimm analysis was performed as described in [45]. Data normalization was done by first determining I_0 by fitting a linear curve over the Guinier plot ($\ln I(q)$ vs q^2). Normalized $1/I(q)$ linear fitting was done starting with the earliest possible data point until a deviation from the linear behavior occurs. Data points were then binned for visual clarity without impacting the result.

Brush model fitting

Brush height model as described in [41] was fitted with a prefactor $c = 0.33$ to match data. Resulting heights were converted to ν by $h = bN^{\nu}$ where $b = 0.38$ nm and $N = 146$. To accommodate for the change in grafting density, a linear curve was fitted to the grafting density's change in salinity and was used to obtain a continuous plot.

Polyelectrolyte fitting

The fitting model was used as described in [40] with a pre-factor $c = 1.24$ to match data.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1140/epje/s10189-024-00409-8.

Acknowledgements R.B. and O.A.S. dedicate this article to Fyl Pincus, for his continuous leadership and friendship over the years. His past works on charged polymer brushes, and polymers' scaling laws, inspired much research in the field, including this work. The synchrotron SAXS data were collected at beamline P12, operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany), at beamline B21, operated by Diamond Light Source (Didcot, UK), and at beamline BM29, operated by ESRF (Grenoble, France). We would like to thank Cy M. Jefferies (DESY), Katsuaki Inoue (DLS), and Mark Tully (ESRF) for their assistance in using the beamlines. This work has been supported by the NSF (MCB-2113302), the NSF-BSF program (2020787), the Israel Science Foundation (1454/20), and by iNEXT-Discovery (15410), funded by the Horizon 2020 program of the European Commission. We also acknowledge the fruitful discussion and help from Yacov Kantor, Uri Raviv, and Sagi Meir.

Funding Information Open access funding provided by Tel Aviv University.

Author contribution statement

M.K., G.K., and R.B. designed the project. M.K. conducted experiments and analysis with G.K.'s and O.A.S.'s assistance. M.K., G.K., R.B., and O.A.S. wrote the paper.

Data availability statement The raw SAXS data is available in the Small-Angle Scattering Biological Data Bank (SASBDB) at: https://www.sasbdb.org/project/2190/.

Declarations

Conflict of interest The authors claim no conflicting interests.

References

1. A.S. Holehouse, B.B. Kragelund, The molecular basis for cellular function of intrinsically disordered protein regions. Nat. Rev. Mol. Cell Biol. **2023**, 1–25 (2023)
2. A. Chowdhury, D. Nettels, B. Schuler, Interaction dynamics of intrinsically disordered proteins from single-molecule spectroscopy. Annu. Rev. Biophys. **52**, 433–462 (2023)
3. B. Xue, A.K. Dunker, V.N. Uversky, Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. J. Biomol. Struct. Dyn. **30**, 137–149 (2012)
4. V.N. Uversky, Intrinsic disorder-based protein interactions and their modulators. Curr. Pharm. Des. **19**, 4191–4213 (2013)
5. T. Ehm et al., Self-assembly of tunable intrinsically disordered peptide amphiphiles. Biomacromolecules **24**, 98–108 (2022)
6. A. Laser-Azogui, M. Kornreich, E. Malka-Gibor, R. Beck, Neurofilament assembly and function during neuronal development. Curr. Opin. Cell Biol. **32**, 92–101 (2015)
7. A.A. Chernyatina, S. Nicolet, U. Aebi, H. Herrmann, S.V. Strelkov, Atomic structure of the vimentin central α-helical domain and its implications for intermediate filament assembly. Proc. Natl. Acad. Sci. **109**, 13620–13625 (2012)
8. E. Malka-Gibor et al., Phosphorylation-induced mechanical regulation of intrinsically disordered neurofilament proteins. Biophys. J . **112**, 892–900 (2017)
9. N. Hirokawa, M.A. Glicksman, M.B. Willard, Organization of mammalian neurofilament polypeptides within the neuronal cytoskeleton. J. Cell Biol. **98**, 1523–1536 (1984)
10. C.R. Safinya, J. Deek, R. Beck, J.B. Jones, Y. Li, Assembly of biological nanostructures: isotropic and liquid crystalline phases of neurofilament hydrogels. Annu. Rev. Condens. Matter Phys. **6**, 113–136 (2015)
11. S. Müller-Späth et al., Charge interactions can dominate the dimensions of intrinsically disordered proteins. Proc. Natl. Acad. Sci. **107**, 14609–14614 (2010)
12. S. Milles, E.A. Lemke, Single molecule study of the intrinsically disordered fg-repeat nucleoporin 153. Biophys. J . **102**, 10a (2012)
13. N. Sekiyama, R. Kobayashi, T.S. Kodama, Toward a high-resolution mechanism of intrinsically disordered protein self-assembly. J. Biochem. **174**, 391–398 (2023)
14. S. Khatun et al., Fractal self-assembly and aggregation of human amylin. Soft Matter **16**, 3143–3153 (2020)
15. J.-E. Shea, R.B. Best, J. Mittal, Physics-based computational and theoretical approaches to intrinsically disordered proteins. Curr. Opin. Struct. Biol. **67**, 219–225 (2021)
16. R. Van Der Lee et al., Classification of intrinsically disordered regions and proteins. Chem. Rev. **114**, 6589–6631 (2014)
17. U. Baul, D. Chakraborty, M.L. Mugnai, J.E. Straub, D. Thirumalai, Sequence effects on size, shape, and structural heterogeneity in intrinsically disordered proteins. J. Phys. Chem. B **123**, 3462–3474 (2019)
18. R.K. Das, R.V. Pappu, Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc. Natl. Acad. Sci. **110**, 13392–13397 (2013)
19. H. Hofmann et al., Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. Proc. Natl. Acad. Sci. **109**, 16155–16160 (2012)
20. W. Zheng, G. Dignon, M. Brown, Y.C. Kim, J. Mittal, Hydropathy patterning complements charge patterning to describe conformational preferences of disordered proteins. J. Phys. Chem. Lett. **11**, 3408–3415 (2020)
21. D. Maltseva et al., Fibril formation and ordering of disordered fus lc driven by hydrophobic interactions. Nat. Chem. **15**, 1–9 (2023)
22. G. Koren et al., Intramolecular structural heterogeneity altered by long-range contacts in an intrinsically disordered protein. Proc. Natl. Acad. Sci. **120**, e2220180120 (2023)
23. J.A. Riback et al., Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. Science **358**, 238–241 (2017)
24. X. Zeng, K.M. Ruff, R.V. Pappu, Competing interactions give rise to two-state behavior and switch-like transitions in charge-rich intrinsically disordered proteins. Proc. Natl. Acad. Sci. **119**, e2200559119 (2022)
25. P.G. Argudo, J.J. Giner-Casares, Folding and self-assembly of short intrinsically disordered peptides and protein regions. Nanoscale Adv. **3**, 1789–1812 (2021)
26. N. Srinivasan, M. Bhagawati, B. Ananthanarayanan, S. Kumar, Stimuli-sensitive intrinsically disordered protein brushes. Nat. Commun. **5**, 5145 (2014)
27. S. Pregent et al., Probing the interactions of intrinsically disordered proteins using nanoparticle tags. Nano Lett. **15**, 3080–3087 (2015)

28. R. Beck, J. Deek, J.B. Jones, C.R. Safinya, Gel-expanded to gel-condensed transition in neurofilament networks revealed by direct force measurements. Nat. Mater. **9**, 40–46 (2010)
29. M. Kornreich, E. Malka-Gibor, B. Zuker, A. Laser-Azogui, R. Beck, Neurofilaments function as shock absorbers: compression response arising from disordered proteins. Phys. Rev. Lett. **117**, 148101 (2016)
30. A. Didonna, P. Opal, The role of neurofilament aggregation in neurodegeneration: lessons from rare inherited neurological disorders. Mol. Neurodegener. **14**, 1–10 (2019)
31. I.L. Morgan, R. Avinery, G. Rahamim, R. Beck, O.A. Saleh, Glassy dynamics and memory effects in an intrinsically disordered protein construct. Phys. Rev. Lett. **125**, 058001 (2020)
32. P. Pincus, Colloid stabilization with grafted polyelectrolytes. Macromolecules **24**, 2912–2919 (1991)
33. E. Zhulina, F. Leermakers, On the polyelectrolyte brush model of neurofilaments. Soft Matter **5**, 2836–2840 (2009)
34. G. Tria, H.D. Mertens, M. Kachala, D.I. Svergun, Advanced ensemble modelling of flexible macromolecules using x-ray solution scattering. IUCrJ **2**, 207–217 (2015)
35. W. Zheng, R.B. Best, An extended guinier analysis for intrinsically disordered proteins. J. Mol. Biol. **430**, 2540–2553 (2018)
36. J.S. Pedersen, C. Svaneborg, Scattering from block copolymer micelles. Curr. Opin. Colloid Interface Sci. **7**, 158–166 (2002)
37. J.S. Pedersen, Form factors of block copolymer micelles with spherical, ellipsoidal and cylindrical cores. J. Appl. Crystallogr. **33**, 637–640 (2000)
38. W.-L. Chen, R. Cordero, H. Tran, C.K. Ober, 50th anniversary perspective: Polymer brushes: Novel surfaces for future materials. Macromolecules **50**, 4089–4113 (2017)
39. C.-H. Wang, M.-B. Luo, X. Xu, C. Wang, L.-Z. Sun, Effects of salt concentration on the polyelectrolyte translocation through a cylinder nanopore. Eur. Polym. J. **121**, 109332 (2019)
40. B.-Y. Ha, D. Thirumalai, Conformations of a polyelectrolyte chain. Phys. Rev. A **46**, R3012 (1992)
41. N.A. Kumar, C. Seidel, Polyelectrolyte brushes with added salt. Macromolecules **38**, 9341–9350 (2005)
42. A. Zamyatnin, Protein volume in solution. Prog. Biophys. Mol. Biol. **24**, 107–123 (1972)
43. H.I. Okur et al., Beyond the hofmeister series: Ion-specific effects on proteins and their biological functions. J. Phys. Chem. B **121**, 1997–2014 (2017)
44. S. Wohl, M. Jakubowski, W. Zheng, Salt-dependent conformational changes of intrinsically disordered proteins. J. Phys. Chem. Lett. **12**, 6684–6691 (2021)
45. B.H. Zimm, The scattering of light and the radial distribution function of high polymer solutions. J. Chem. Phys. **16**, 1093–1099 (1948)
46. V.N. Uversky et al., Natively unfolded human prothymosin α adopts partially folded collapsed conformation at acidic ph. Biochemistry **38**, 15009–15016 (1999)
47. A. Möglich, K. Joder, T. Kiefhaber, End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. Proc. Natl. Acad. Sci. **103**, 12394–12399 (2006)
48. R.V. Pappu, R. Srinivasan, G.D. Rose, The flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. Proc. Natl. Acad. Sci. **97**, 12565–12570 (2000)
49. A. Quigley, D. Williams, The second virial coefficient as a predictor of protein aggregation propensity: a self-interaction chromatography study. Eur. J. Pharm. Biopharm. **96**, 282–290 (2015)
50. P.K. Nandi, D.R. Robinson, Effects of salts on the free energies of nonpolar groups in model peptides. J. Am. Chem. Soc. **94**, 1308–1315 (1972)
51. P.K. Nandi, D.R. Robinson, Effects of salts on the free energy of the peptide group. J. Am. Chem. Soc. **94**, 1299–1308 (1972)
52. E. Dickinson, Strategies to control and inhibit the flocculation of protein-stabilized oil-in-water emulsions. Food Hydrocoll. **96**, 209–223 (2019)
53. T.B. Blijdenstein, C. Veerman, E. van der Linden, Depletion- flocculation in oil-in-water emulsions using fibrillar protein assemblies. Langmuir **20**, 4881–4884 (2004)
54. F. Zosel, A. Soranno, K.J. Buholzer, D. Nettels, B. Schuler, Depletion interactions modulate the binding between disordered proteins in crowded environments. Proc. Natl. Acad. Sci. **117**, 13480–13489 (2020)
55. E.B. Zhulina, O.V. Borisov, Cylindrical brushes with ionized side chains: scaling theory revisited. Soft Matter **19**, 8440–8452 (2023)
56. H.P. Truong et al., Pincus blob elasticity in an intrinsically disordered protein. Eur. Phys. J. E **46**, 100 (2023)
57. Y. Li, R. Beck, T. Huang, M.C. Choi, M. Divinagracia, Scatterless hybrid metal-single-crystal slit for small-angle x-ray scattering and high-resolution x-ray diffraction. J. Appl. Crystallogr. **41**, 1134–1139 (2008)
58. N.P. Cowieson et al., Beamline b21: high-throughput small-angle x-ray scattering at diamond light source. J. Synchrotron Radiat. **27**, 1438–1446 (2020)
59. C.E. Blanchet et al., Versatile sample environments and automation for biological solution x-ray scattering experiments at the p12 beamline (petra iii, desy). J. Appl. Crystallogr. **48**, 431–443 (2015)
60. P. Pernot et al., Upgraded esrf bm29 beamline for saxs on macromolecules in solution. J. Synchrotron Radiat. **20**, 660–664 (2013)
61. A. Round et al., Biosaxs sample changer: a robotic sample changer for rapid and reliable high-throughput x-ray solution scattering experiments. Acta Crystallogr. D Biol. Crystallogr. **71**, 67–75 (2015)
62. J. Kieffer, V. Valls, N. Blanc, C. Hennig, New tools for calibrating diffraction setups. J. Synchrotron Radiat. **27**, 558–566 (2020)
63. P. Virtanen et al., Scipy 1.0 fundamental algorithms for scientific computing in python. Nat. Methods **17**, 261–272 (2020)
64. M. Newville, et al. Lmfit: Non-linear least-square minimization and curve-fitting for python. Astrophys. Source Code Libr. ascl-1606 (2016)
65. B. Jacrot, The study of biological structures by neutron scattering from solution. Rep. Prog. Phys. **39**, 911 (1976)

Supplementary Information

From isolated polyelectrolyte to star-like assemblies: the role of sequence heterogeneity on the statistical structure of the intrinsically disordered Neurofilament-low tail domain

Mathar Kravikass, Gil Koren, Omar Saleh, Roy Beck
Corresponding E-mail:roy@tauex.tau.ac.il

Contents:
Table S1-S5,
Figure S1-S10

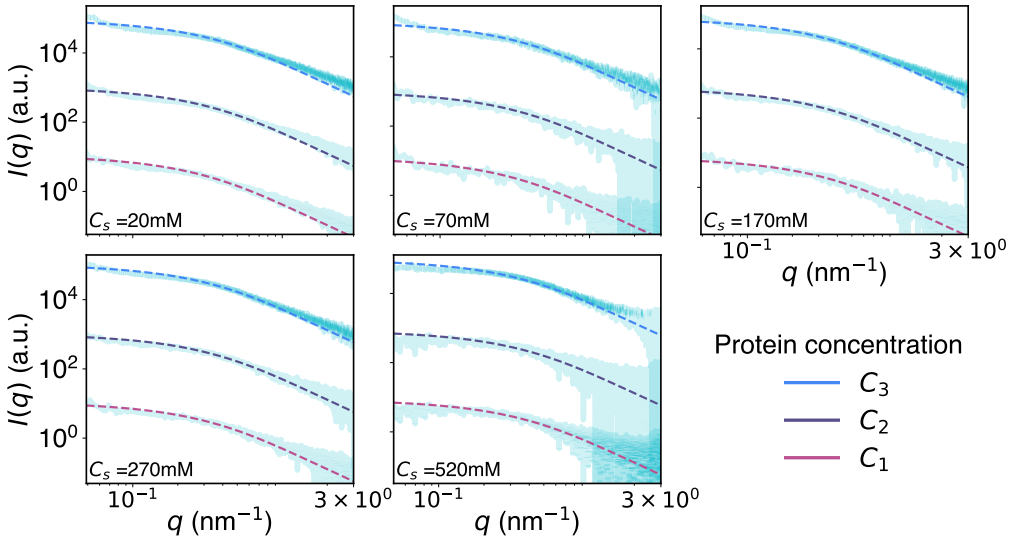

Fig. S1 ΔN42 SAXS measurements Gaussian form factor fitting for all salinity concentrations C_s. R_G used for the Gaussian form factor is as obtained by the Extended Guinier analysis (see table S1).

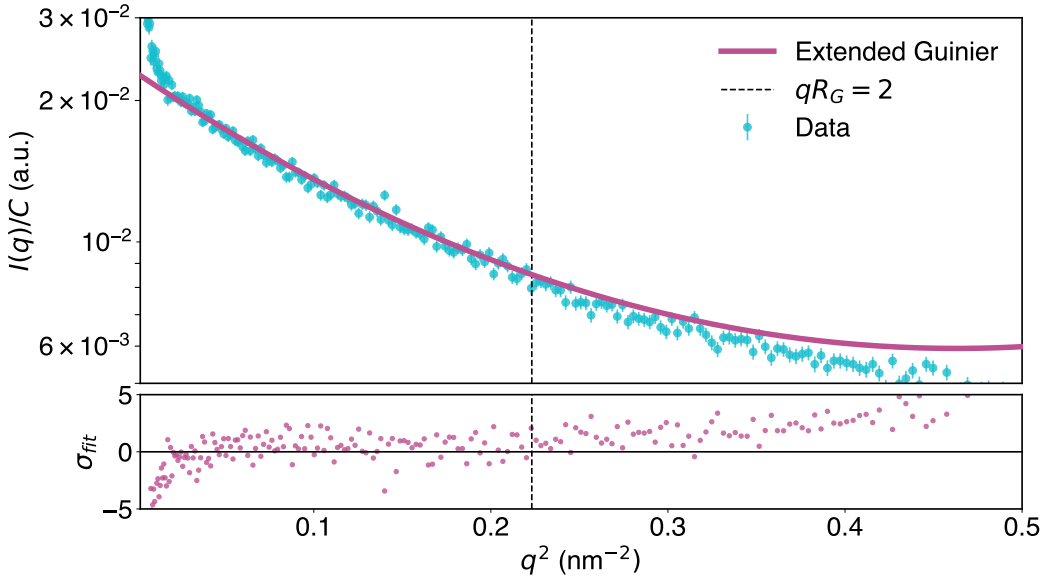

Fig. S2 ΔN42 SAXS measurement with corresponding extended Guinier curve. Bottom: Deviation from fit $\sigma_{fit} = (Y_{data} - Y_{fit})/\sigma_{data}$ Dashed line represents the maximum analysis point $qR_G = 2$ from which deviation starts. Displayed data: 20mM Tris pH8.0 in 1.1mg/ml.

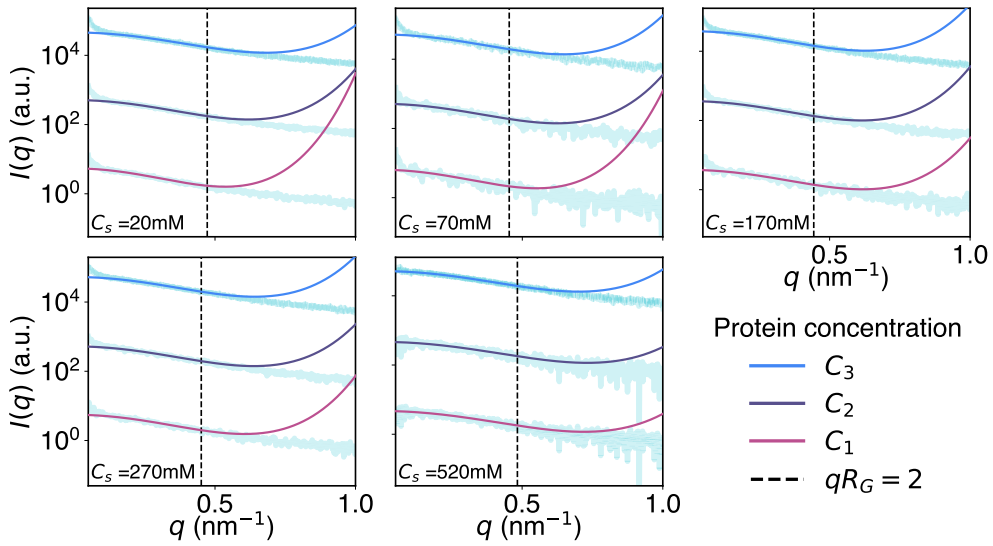

Fig. S3 ΔN42 SAXS measurements with corresponding extended Guinier curves. Dashed lines represent the maximum analysis point $qR_G = 2$ from which deviation starts. Protein concentrations were offset for clarity, with the lowest (blue) being of the highest concentration.

C_s (mM)	C (mg/ml)	R_G (nm)	ν	I_0 (cm^{-1})
20	1.1	4.23 ± 0.05	0.642 ± 0.002	0.0228
20	0.8	4.56 ± 0.07	0.660 ± 0.66	0.0254
20	0.6	5.11 ± 0.13	0.689 ± 0.007	0.027
70	1	4.41 ± 0.12	0.652 ± 0.007	0.0259
70	0.5	4.53 ± 0.22	0.659 ± 0.012	0.0257
70	0.3	4.99 ± 0.46	0.683 ± 0.023	0.032
170	1.5	4.51 ± 0.05	0.657 ± 0.003	0.19
170	0.8	4.59 ± 0.11	0.662 ± 0.006	0.18
170	0.3	4.56 ± 0.27	0.661 ± 0.015	0.18
270	1	4.43 ± 0.06	0.653 ± 0.003	0.026
270	0.5	4.45 ± 0.1	0.654 ± 0.006	0.026
270	0.3	4.64 ± 0.16	0.664 ± 0.009	0.028
520	1.5	4.14 ± 0.02	0.636 ± 0.001	0.026
520	0.78	4.00 ± 0.08	0.628 ± 0.005	0.023
520	0.38	4.05 ± 0.35	0.630 ± 0.02	0.023

Table S1 ΔN42 Extended guinier analyis data. Analysis parametres (radius of gyration R_G, scaling exponent ν, and scattering intensity at $q = 0$ (I_0)) obtained for different salt concentrations (C_s) and protein concentrations (C).

C_s (mM)	R_G (nm)	ν
20	5.76 ± 0.31	0.729 ± 0.015
70	4.84 ± 0.27	0.677 ± 0.015
170	4.71 ± 0.04	0.669 ± 0.003
270	4.61 ± 0.16	0.663 ± 0.009
520	3.88 ± 0.04	0.620 ± 0.003

Table S2 Zero concentration extended Guinier analysis data. Analysis parameters (radius of gyration R_G and scaling exponent ν) were extrapolated to zero protein concentration at various salt concentrations (C_s).

C_s (mM)	C (mg/ml)	$h/2$ (nm)	ν	Z	n	R (nm)	L (nm)	V (nm^3)	β_t (10^3nm)	β_c (10^3nm)
20	2.68	9.16±0.15	0.786±0.0033	1.60±0.03	10.11±0.711	0.89±0.028	1.19±0.09	2.26±0.027	0.227	3.826
20	1.8	8.20±0.13	0.758±0.0032	1.83±0.04	8.52±0.059	0.89	1.19	2.26	0.195	3.558
20	1	8.57±0.17	0.768±0.0041	1.87±0.05	8.12±0.060	0.89	1.19	2.26	0.195	3.558
20	0.5	8.27±0.16	0.759±0.0040	1.91±0.05	7.78±0.057	0.89	1.19	2.26	0.195	3.558
170	1.3	9.96±0.03	0.796±0.0006	3.34±0.02	2.52±0.009	0.66±0.005	X	1.18±0.007	0.039	4.014
170	0.73	10.11±0.05	0.799±0.0009	3.27±0.03	2.32±0.014	0.63±0.007	X	1.06±0.010	0.039	4.014
170	0.57	10.37±0.08	0.806±0.0015	2.13±0.02	3.10±0.037	0.60±0.006	X	0.93±0.008	0.074	3.98
170	0.24	10.78±0.10	0.815±0.0019	2.71±0.05	3.23±0.025	0.66±0.011	X	1.23±0.017	0.074	3.98
270	2	9.40±0.02	0.781±0.0003	5.37±0.02	1.89±0.007	0.70±0.004	X	1.42±0.008	0.018	4.03
270	1.5	9.43±0.02	0.782±0.0004	5.02±0.02	1.93±0.009	0.69±0.005	X	1.36±0.009	0.018	4.03
270	0.69	9.85±0.04	0.793±0.0009	4.05±0.04	2.70±0.016	0.71±0.009	X	1.53±0.016	0.039	4.014
370	2.3	9.36±0.01	0.780±0.0003	6.68±0.03	1.48±0.008	0.69±0.005	X	1.38±0.009	0.018	4.036
370	1.5	9.44±0.02	0.782±0.0003	6.43±0.03	1.65±0.009	0.71±0.006	X	1.49±0.010	0.018	4.036
370	0.96	9.53±0.02	0.785±0.0004	5.85±0.03	2.08±0.010	0.74±0.006	X	1.70±0.012	0.039	4.014
370	0.6	9.81±0.03	0.792±0.0007	5.33±0.05	2.13±0.017	0.72±0.010	X	1.59±0.019	0.039	4.014
520	2.5	9.57±0.01	0.786±0.0003	8.15±0.04	1.82±0.008	0.79±0.005	X	2.07±0.012	0.018	4.036
520	1.19	9.28±0.02	0.778±0.0004	6.66±0.04	1.57±0.010	0.70±0.006	X	1.47±0.012	0.018	4.036
520	0.45	9.82±0.03	0.792±0.0005	6.36±0.06	1.94±0.016	0.74±0.010	X	1.72±0.020	0.039	4.014

Table S3 WT spherical and cylindrical fitting analysis data. Analysis parameters (brush height (h), scaling exponent (ν), aggregation number (Z), core peptide length (n), core radius (R), cylindrical core length (L), core volume (V), tail scattering length (β_t) and core scattering length (β_c)) obtained for different salt concentrations (C_s) and protein concentrations (C). Cylinder length L values are only relevant to $C_s = 20$mM where a cylindrical core fit was used. For the cylindrical core, the same values of L and R were used for all concentrations to alleviate fitting errors (see Methods).

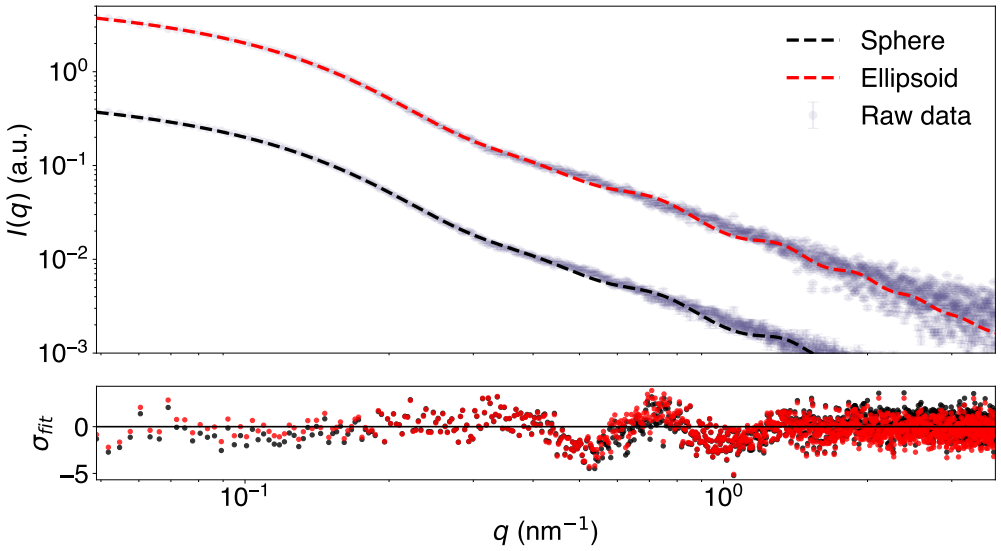

Fig. S4 SAXS measurements of WT and its fitting to different form factors. Both form factors are of the same model but use a different core: Spherical or Ellipsoidal. Spherical core fitting yields a core radius of $R = 0.66 \pm 0.016$ nm, and the ellipsoidal core yields a core radius of $R = 1.335 \pm 0.23$ nm and a secondary radius of ϵR where $\epsilon = 0.153 \pm 0.08$. Both fittings yield close values of aggregation number Z (3.046 ± 0.04 for spherical and 3.562 ± 0.07 for ellipsoidal) and tail height $h/2$ (9.838 ± 0.04 nm for spherical and 9.584 ± 0.12 for ellipsoidal). Below: Fitting error $\sigma_{fit} = (Y_{fit} - Y_{data})/\sigma_{data}$. Both curves show similar error profiles. The spherical model proved best to describe the model due to its simplicity. Displayed data: WT in 20 mM Tris pH=8.0, and 170 mM NaCl at a concentration of 1.3 mg/ml.

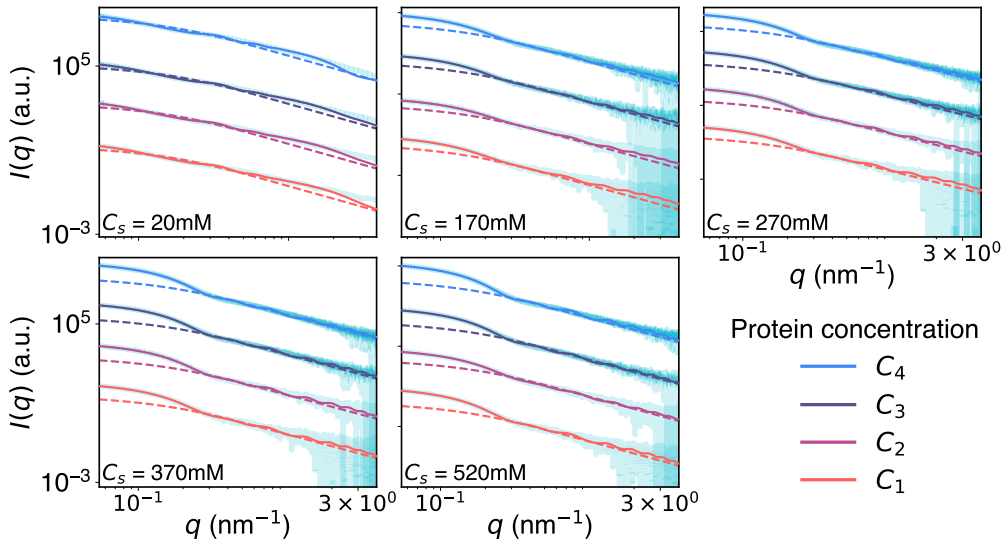

Fig. S5 SAXS measurements and spherical form factor fitting for all salinity concentrations (C_s). The $C_s = 20$ mM data fit is to a cylindrical core. Dashed lines represent the Gaussian form factor of the structure tails. Protein concentrations were offset for clarity, with the lowest (blue) being the highest concentration.

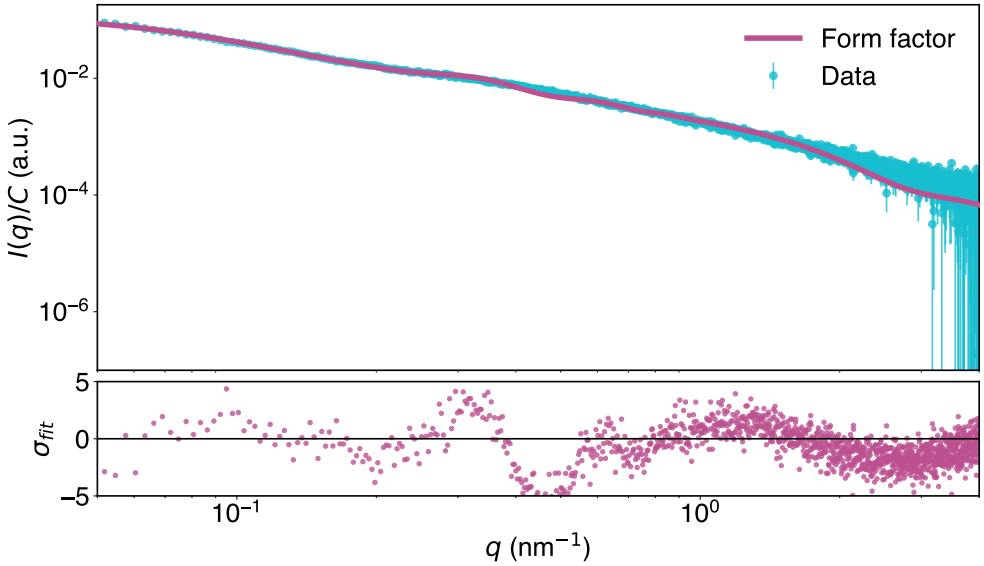

Fig. S6 WT SAXS measurement with cylindrical fitting. Measurements at the highest concentration of $C = 2.68$ mg/ml, in a 20 mM Tris buffer at pH=8.0. To alleviate fitting inconsistencies, consequent fittings of measurements with lower protein concentrations in the same buffer were done using the obtained core parameters: Core radius $R = 0.89 \pm 0.03$ nm, and core length $L = 1.19 \pm 0.09$ nm.

C_s (mM)	$h/2$ (nm)	ν	Z	n	R (nm)	L (nm)	V (nm^3)
20	8.01±0.46	0.751±0.013	2.03±0.08	7.08±0.39	0.89	1.19	2.26
170	10.60±0.33	0.811±0.008	2.83±0.31	3.17±0.55	0.64±0.03	X	1.10±0.16
270	9.84±0.21	0.793±0.006	3.52±0.22	3.06±0.30	0.70±0.03	X	1.49±0.20
370	9.73±0.12	0.790±0.003	5.19±0.29	2.39±0.12	0.76±0.02	X	1.67±0.01
520	9.47±0.44	0.783±0.011	5.67±0.35	1.82±0.29	0.68±0.06	X	1.28±0.38

Table S4 Zero concentration WT spherical and cylindrical fitting analysis data. Analysis parametres (brush height (h), scaling exponent (ν), aggregation number (Z), core peptide length (n), core radius (R), cylindrical core length (L) and core volume (V)) were extrapolated to zero protein concentration at various salt concentrations (C_s). Cylinder length L values are only relevant to $C_s = 20$mM where a cylindrical core was used.

C_s (mM)	A_2^{WT} (cm^{3}mol/g^2 × 10^3)	$A_2^{\Delta N42}$ (cm^{3}mol/g^2 × 10^3)
20	-0.295±1.346	13.264±0.466
70	X	3.978±1.248
170	-2.072±2.091	0.169±1.544
270	-3.328±0.508	-1.152±0.756
370	-2.020±0.563	X
520	-1.933±3.582	-4.417±1.514

Table S5 Second virial coefficient A_2 values for both variants in salt concentration C_s.

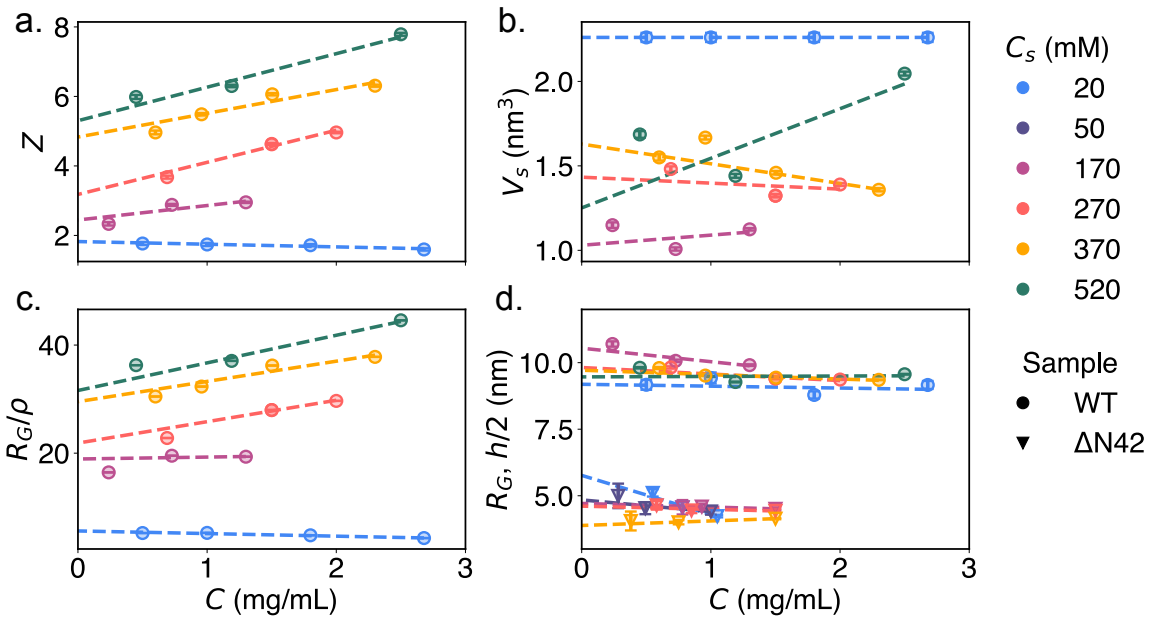

Fig. S7 Structural parameters for WT (circles) and ΔN42 (triangles) variants extracted from fitting the SAXS data. Dashed lines demonstrate the linear fitting of the data used to obtained the zero concentration extrapolations. **a.** Aggregation number (Z) dependency on protein concentration (C) increases with increasing salt. **b.** Core volume V_s against protein concentration (C). In $C_s = 20$ mM, the V_s values are constant due to fitting constraints (see Methods). **c.** In all cases, the tail heights (h) are much larger than the corresponding grafting length (ρ), indicative of a brush regime. **d.** The structurally intrinsically disordered ΔN42 variant compacts with higher C_s values and remains more compacted from the projected tails for the WT variant. For the ΔN42 variant R_G drastically changes as a function of the protein concentration (C).

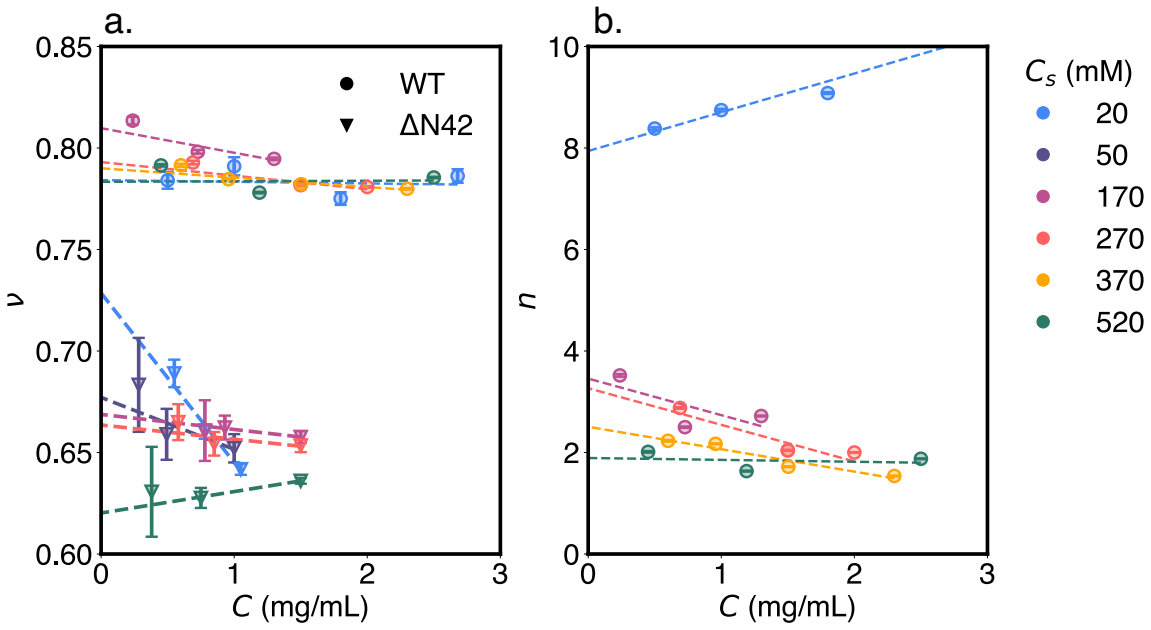

Fig. S8 **a.** Flory exponent (ν) of WT tails and ΔN42 variants as a function of the concentration. ΔN42 shows to change radically as a function of the concentration at the lowest salinities. This effect is reduced as salinity concentration Cs reaches 170mM. WT and the rest of ΔN42 ν data shows little change as a function of the protein concentration.**b.** The core (aggregated) peptide length per polypeptide as a function of the concentrations. The large drop observed from $C_s = 20$mM to $C_s = 170$mM can be attributed to the shift from a dimer to a trimer. Core peptide length difference diminishes with increasing salinity, however the value still remain largely similar.

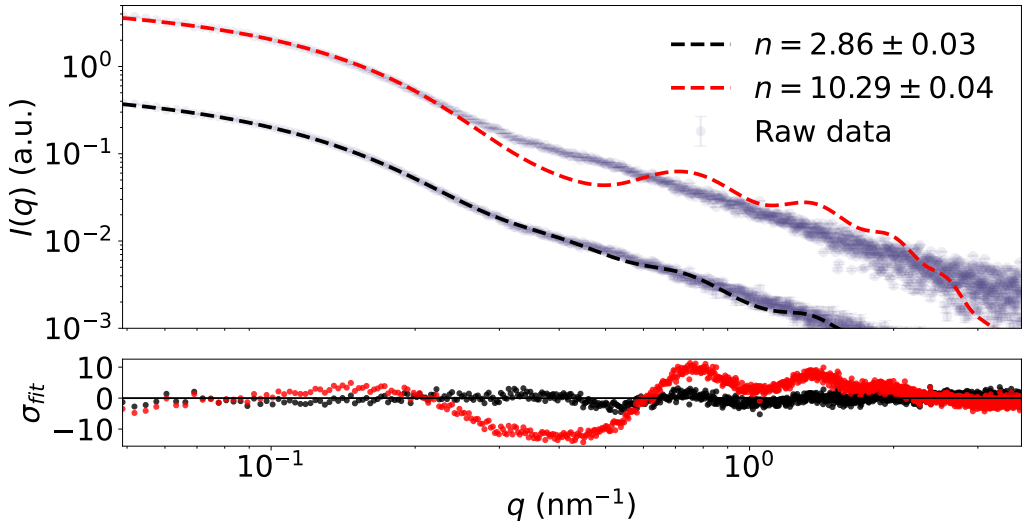

Fig. S9 SAXS measurement of WT and its fitting with different core residue number n. In fixing the core residue number to a constant value of 10 (in red), the fitting becomes noticeably worse than when n is allowed to vary (in black). Displayed data: WT in 20 mM Tris pH=8.0, and 170 mM NaCl at a concentration of 1.3 mg/ml.

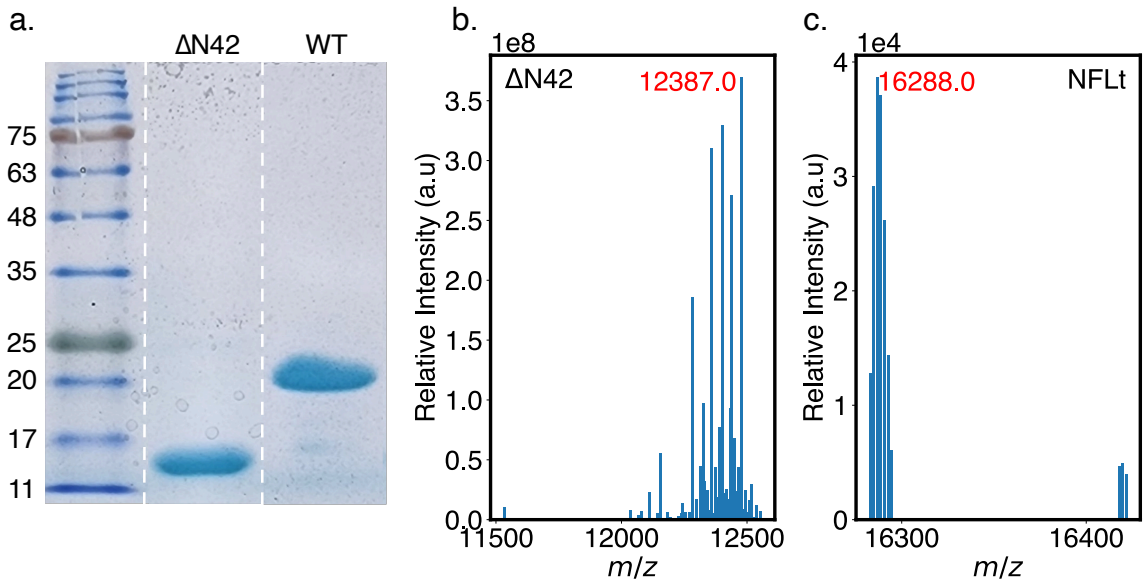

Fig. S10 **a.** SDS-PAGE Tris-Glycine 15% of both ΔN42 and NFLt (WT), showing purity above 95%. White dashed lines indicate where image lanes were edited closer for clarity. Both show a higher molecular weight reading in the gel, which is common for IDPs.**b-c.** Deconvoluted ESI-TOF MS spectra of ΔN42 and NFLt respectively. Theoretical molecular weight values are 12423.57 and 16233.79 for ΔN42 and NFLt, respectively

תקציר

חלק גדול מן החלבונים בטבע אינם מסוגלים להתקפל למבנים קבועים. בכך שחלבונים אלו שומרים את הטבע הפולימרי שלהם, בעבר הם תוארו לפי המבנה הסטטיסטי שלהם באמצעות תאוריות שדה ממוצע (mean field). אף על פי כן, ההטרוגניות של רצף החלבון יכולה להוביל לסטיות נכרות ממודלים אלו עקב רשת האינטראקציות המסובכת הנוצרת בין המונומרים בחלבון.

בתזה זאת אני התמקדתי בזנב הלא מסודר של Neurofilament-low. חלבון זה מורכב מכחמישים חומצות אמינו שרובן חסרי מטען והידרופוביות, וכתשעים ושישה חומצות אמינו טעונות שלילית ברובן. בעבודה זו אני הדגמתי איך רצף חלבון-המודל משפיע על הרשת אינטראקציות ועל המבנה הסטטיסטי שלו.

השפעת הרצף נבדקה על ידי מדידות בשיווי משקל ומחוץ לשיווי משקל באמצעות מדידות על שני גרסאות של החלבון. מדידות שיווי משקל נעשו באמצעות קרינת קרני אקס בזויות קטנות (Small-angle X-ray scattering). במדידות אלו התגלה כי התחום הלא טעון של החלבון גורם הצטברות ויצירת על-מבנים דמוי כוכבים בתמיסה. לעומת זאת , כאשר התחום הלא טעון של החלבון מוסר, החלבונים נשארים מבודדים בתמיסה, ומתנהגים בדומה לפוליאלקטרוליטים כפי שמותרים בתיאוריה. במהלך התזה אדון כיצד האינטראקציות ארוכות וקצרות טווח גורמות לתופעות אלו. בנוסף, יצרתי קשר בין המדידות הנוכחיות שלי להתנהגויות שפורסמו בעבר של Neurofilament-low.

מחוץ לשיווי משקל, ערכתי ניסויי מתיחת שרשראות החלבון באמצעות מערכת צוותות מגנטית (magnetic tweezers). ניסויים אלו נערכו בהעזר משטח אליו החלבונים חוברו, וכדורים מגנטים שנקשרו לקצה החופשי של החלבונים. ניסויים אלו נערכו בהמשך לניסויי עבר בהם התגלה כי הזנב של חלבון המודל מראה דינמיות זכוכיות (glassy dynamics) במתיכה ושחרור. ניסויים אלו קשרו את מקור הדינמיקה הזאת לקיומם של מספר תחומים בחלבון בעלי דינמיקה בלתי תלוייה לתחומים אחרים. בעבודתי חזרתי על הבדיקות באמצעות שתי הגרסאות של החלבון במטרה לקשר את התחומים הבלתי תלויים לחלק ההידרופובי של החלבון.

רוב תוצאות תזה זאת פורסמו לאחרונה במאמר שעבר ביקורת עמיתים ומצורף כנספח.

ביה"ס לפיזיקה ואסטרונומיה

החוג לפיזיקה של חומר מעובה

השפעת הטרוגניות הרצף על המבנה הסטטיסטי של הזנב הלא מסודר של החלבון Neurofilament-low

מטר קרביקס

תזה זו מוגשת תחת השלמה חלקית של הדרישות לתואר מוסמך אוניברסיטה
מדעי בבית הספר לפיזיקה ואסטרונומיה, אוניברסיטת תל אביב

תחת הנחייתו של פרופ׳ רועי בק

פברואר 2024, תשפ"ד