Coarse Grained Modeling of Intrinsically Disordered Protein Structures and Dynamics

by

Andrei Igorevitch Vovk

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Physics
University of Toronto

Abstract

Coarse Grained Modeling of Intrinsically Disordered Protein Structures and Dynamics

Andrei Igorevitch Vovk
Doctor of Philosophy
Graduate Department of Physics
University of Toronto
2019

Contrary to natively folded proteins, which often have a single stable 3D structure, Intrinsically Disordered Protein (IDP) structures dynamically fluctuate between many conformations. This complexity impedes their study and the understanding of their functions. Experimental techniques can probe aspects of the conformations, but often different techniques or conditions offer seemingly contradictory results. In this thesis, I use coarse-grained models and simulations to investigate the structures and dynamics of IDPs.

One system I focus on is the Nuclear Pore Complex (NPC): a bidirectional selective gate for cargo traveling across the nuclear envelope of eukaryotic cells. A key functional component of the NPC is the layer of intrinsically disordered FG nucleoporins attached to the interior of the transport channel. Elucidating the mechanism of transport remains elusive due to the size and complexity of the NPC, but will contribute to the understanding of diseases associated with mutations in the NPC, viruses which enter the nucleus via the NPC, and drug delivery into the nucleus, as well as the development of NPC inspired devices for nanotechnological applications. I show how a mean field model for surface grafted FG nups reconciles several conflicting experimental observations and theories of transport.

Due to a lack of unique structure, the influence of the amino acid sequence on IDP structures and dynamics is unclear. Various evidence points to the overall importance of general properties of the sequence rather than the atomistic details of the amino acids. Using coarse-grained simulations, I investigate how the effects of several key sequence properties on the polymer dimensions and end-to-end distance dynamics of IDPs. These results have implications for interpreting experimental measurements as well as developing appropriate models of IDPs.

Contents

List of Tables

List of Figures

3.3 (a) Cohesion makes FG nup layers more compact: theoretical predictions. Layer height h/L normalized by the chain length as a function of the normalized grafting distance a/l for increasing cohesiveness. χ_{cr} varies from 0 (grey) to -1.5 (black). For any value of χ_{cr}, the curve is well approximated by the dependence $h \sim a^{-g}$. The inset shows that the exponent g increases from $\frac{2}{3}$ to 2 as the absolute value of the cohesion strength $|\chi_{cr}|$. (b) FG nup layer height depends on the grafting distance: theory vs. experiment. The dots are the experimentally measured layer heights from [159] and [160] normalized by the FG nup length. The colours represent: Nup62, Nup98, Nup153, long **Nsp1**, and short Nsp1 segments. Solid line: $h \sim a^{-\frac{2}{3}}$ is the ideal brush ($\chi_{cr} = 0$) behavior obtained from the model. Dotted line: $h \sim a^{-2}$ is the behavior of a strongly collapsed brush with $\chi_{cr} = -2.5$. All the FG Nups lie between these two regimes, indicating a significant amount of cohesion. The dashed line is for $\chi_{cr} = -0.8$. The dashed-dotted line is for $\chi_{cr} = -1.4$. To enhance the contrast, inset shows the same data with the height h normalized by the ideal brush height ($h \sim a^{-\frac{2}{3}}$). b=1.52 nm, l=1 nm. 43

3.4 **Characteristic responses of FG nup layers to the transport proteins: experimental results.** (a) Change in the layer height relative to the unperturbed layer as a function of the transport protein concentration in the outside solution. (b) Number of the transport proteins in the layer per unit length of the FG nup chain. Each line corresponds to a different run with a different initial layer height and grafting distance. Different colors correspond to different FG nups, which all exhibit qualitatively similar behavior. The colours indicate different combinations of transport proteins and FG nups: Karyopherin-β1 on Nup62, Nup98, Nup153, and **Nsp1**, and NTF2 on Nsp1. The corresponding average grafting distances are $\sim$ 2.5 nm, $\sim$ 4.2 nm, $\sim$ 4.5 nm, $\sim$ 3.75 nm, $\sim$ 3.75 nm. The data are from Refs. [159, 160] 44

3.5 **Collapse and recovery: effect of cohesion and of the grafting distance.** (a) Theoretical curves show that FG nup cohesion can convert layer collapse to swelling. The cross-linking strengths are $\chi_{cr} = 0, -0.4, -0.8, -1.1$ for $a = 5$ nm and $\chi = -550$. (b) Increasing grafting distance increases the magnitude of the layer compaction. The lines correspond to model predictions for $a = 3, 4, 5, 6$ nm for $\chi = -530$ and $\chi_{cr} = -1$. The insets show that the fraction of free space in the layer, calculated as $1 - \phi - \psi$, decreases with the addition of the transport proteins. $b = 1$ nm, $l = 0.67$ nm in both panels. 45

Chapter 1

Introduction

1.1 Intrinsically Disordered Proteins

Proteins are biopolymers whose sequences are built up out of the 20 amino acids. During the past century, the dominant view has been that a protein's function is tied to its unique 3D structure, which is encoded by its sequence. Over the past several decades, a new classification became necessary to account for the mounting evidence that many proteins are intrinsically flexible yet are biologically functional and do not conform to the classical unique structure - function paradigm. Under native conditions, instead of maintaining a single 3D structure, their structure comprises an ensemble of conformations (see Figure 1.1). These proteins fluctuate dynamically between the conformations, which gives rise to their function. Intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) became the common nomenclature used to distinguish this new class from traditional ordered proteins [1, 2].

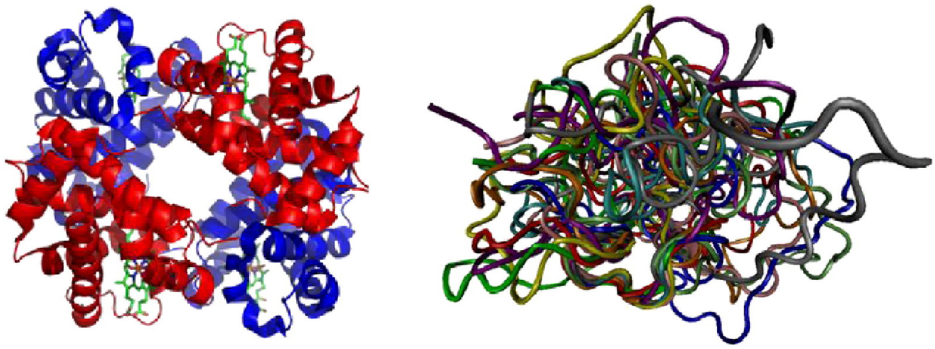


Figure 1.1: Left: a unique 3D structure of an ordered protein. Right: several conformations of an IDP, each in a different colour. Figure adapted from Uversky et al. [3].

An IDR is a region of a protein without a well defined 3D structure. The boundary separating IDPs and folded proteins is not clearly defined, as many proteins contain both ordered regions and IDRs. In humans, 44% of all protein coding genes contain IDRs of at least 30 amino acids in length. Intrinsic disorder is more prevalent in eukaryotes than in bacteria or archaea as shown in Figure 1.2. Disorder is

also more prevalent within proteins whose functions are currently unknown [1, 4].

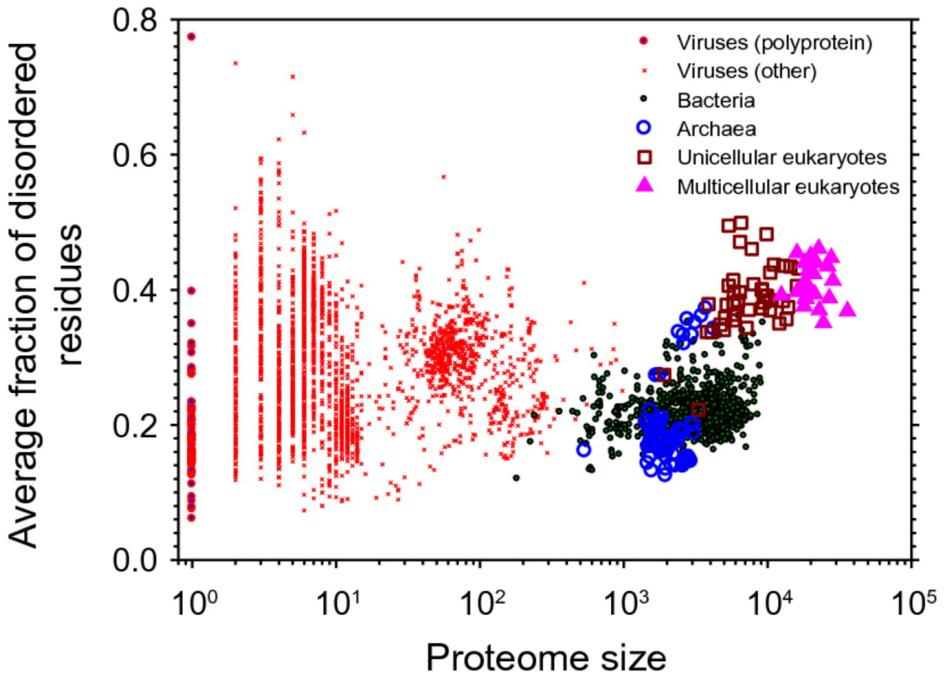


Figure 1.2: Prevalence of intrinsic disorder within the proteomes of different organisms grouped by species. The total number of proteins in the proteome of a particular species is on the x-axis. The fraction of disordered residues (evaluated by the PONDR-VSL2B neural network predictor) is averaged over all proteins for that species and shown on the y-axis. Figure adapted from Xue et al. [5].

Studying how an IDP's sequence underlies its ensemble of conformations and the dynamics of interconversion between the conformations is an important step to understanding the principles behind the functions of this class of proteins. Furthermore, a broad range of human diseases are associated with the failure of an ordered protein to adopt its native conformational state and consequently gaining some of the structural properties of an IDP [2, 6]. Proteins associated with cancer, diabetes, neurodegenrative, and cardiovascular diseases often have regions of structural disorder, making them leading targets of drug development [1, 2, 7]. In IDPs, disease mutations occur in their ordered regions more often than in the IDRs [8]. It is puzzling that structural disorder is the cause of dysfunction in one case, but both functional and resistant to mutation in another.

1.1.1 Role of Intrinsic Disorder in IDP Function

The categories of functions of IDPs are distinct from natively folded proteins, indicating that disorder allows for a diversity of functions which are not possible for ordered proteins or offers advantages over ordered proteins in these functions. Due to their extended structures, IDPs can occupy more volume per residue of the sequence than folded proteins, and therefore economize genome and protein length. Due to their relative lack of secondary and tertiary structure, they are resistant to disease-related aggregation and external stresses such as changes in solvent or temperature, compared with natively folded proteins [4]. In binding interactions, IDPs are able to adopt conformations which are difficult

for natively ordered proteins, such as protruding into or wrapping around binding partners. Further advantages include binding promiscuity: the capability of a single IDP to bind to multiple partners by adopting different structures, increased binding rates, as well as low affinity with high specificity binding due to a coupling between binding and folding [3, 9, 10].

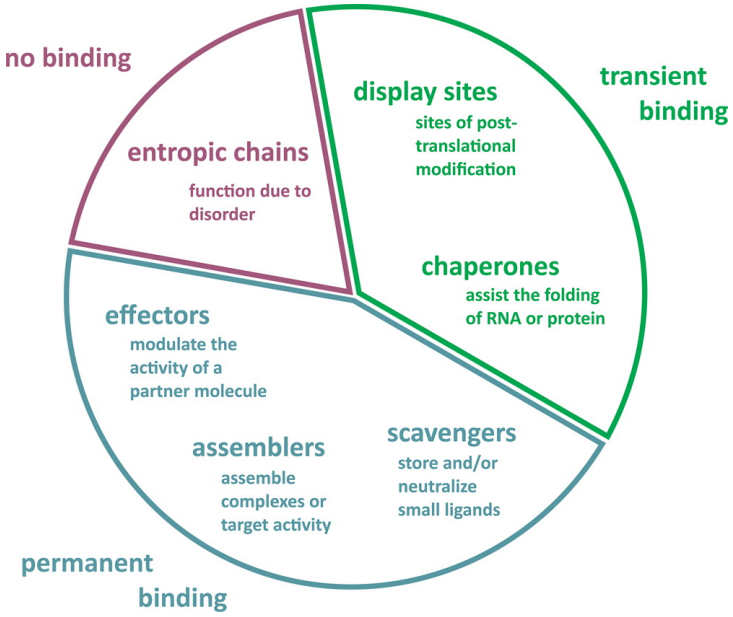


Figure 1.3: Classification of IDP functions. Figure from van der Lee et al. [4].

One of the categories of IDP functions is the "entropic chain". This function relies purely on the disorder of the conformations and does not involve any binding or folding. A few examples include: entropic bristles that exclude other molecules, flexible linkers that allow movement of two domains relative to each other, and entropic springs that can exert force [2, 4].

The IDPs of other categories function by either permanent or transient binding to other molecules such as DNA, RNA, ligands, or other proteins. Many of these functions involve a disorder-to-order transition upon binding. In others, IDPs remain disordered: they do not present a single binding site and can be thought of as a binding cloud with a distribution of weak binding sites [1]. "Effector" IDPs alter the action of the target molecule. "Scavenger" IDPs store and/or neutralize small ligands. There is a correlation between intrinsic disorder and the number of subunits of protein complexes, and thus "assembler" IDPs assemble and regulate large protein complexes such as the ribosome. In large complexes, a few proteins are termed "hub proteins" which are involved in multiple connections and organize the network. Another class of IDPs offer display sites for post-translational modifications. Finally, many RNA and protein chaperones contain IDRs. Disorder offers advantages to chaperone function as IDRs can bind multiple misfolded substrates and inhibit their aggregation [2, 4, 9, 11].

One specific example of IDP function is the selective transport mechanism of the Nuclear Pore Complex (NPC) of eukaryotic cells. FG nucleoporins (IDPs with a recurring Phenylalanine-Glycine motif) are attached to the interior of a cylindrical channel via a natively folded region, while the disordered region of their sequences fills the channel. FG nucleoporins act as a barrier to most molecules but selected cargo binds to transport proteins, allowing the transport protein-cargo complex to pass through

the channel. FG nucleoporins share features with several IDP functional categories: they act both as entropic chains that exclude inert cargo, as well as possess the capability of molecular recognition and transient binding to transport proteins [2, 12]. Elucidating the mechanism of transport contributes to the understanding of diseases associated with mutations in the NPC, viruses which enter the nucleus via the NPC, and drug delivery into the nucleus, as well as the development of NPC inspired devices for nanotechnological applications [13–16]. In Chapter 3, I will discuss which properties control the structures of surface grafted FG nucleoporin layers, which resemble their organization in the NPC, and how the layers are affected by transport proteins.

Just as for natively folded proteins, there is a link between the structures of IDPs and their function. Local features such as the presence of folded motifs might be involved in mediating interactions with binding partners. Global features such as the hydrodynamic radius might be related to the exposure of an IDP's binding domains, long range communication between different parts of the IDP, or its entropic properties [2]. However, by their very nature, IDP structures cannot be experimentally characterized using techniques such as X-ray crystallography and cryogenic electron microscopy, in the same way as the unique structures of folded proteins.

1.1.2 IDP Structure and Dynamics

In their native state, IDPs can be in any one of a vast number of conformations. Furthermore, an IDP is constantly fluctuating between these conformations and the dynamics of inter-conversion between them are also important to its function. In general, a protein is identified as intrinsically disordered if it has no unique stable structure under native conditions, and can be identified using one of several biophysical techniques [2, 17, 18].

A protein's structure can be defined by its configuration at a moment in time: the relative coordinates of all of the atoms. However, no two configurations will ever be the same and even the configurations of folded protein will be innumerable. Instead we can introduce the more tractable idea of a conformation: a set of configurations grouped together using a threshold of structural similarity. This definition depends upon the choice of threshold, but generally the configurations of a folded protein will be similar to their ensemble average. This average and its variance defines the unique folded conformation and can be characterized by X-ray crystallography. On the other hand, the average of IDP configurations will be dissimilar from any particular configuration in the ensemble [19].

The difference between IDPs and natively folded proteins can be illustrated by comparing their free energy landscapes in conformation space as shown in Figure 1.4. For folded proteins, this landscape forms a funnel that has a minimum energy for a single most probable native conformation. The narrow width of the funnel represents the low variance of the configurations. On the other hand, IDPs tend to have a relatively flat free energy landscape and explore many similarly probable conformations via thermal motion. This conformational landscape is shaped by the interactions between the amino acids making up the protein, but also depends very strongly on the surrounding environment [3, 20, 21].

Protein disorder prediction is often binary, only differentiating between whether a protein has a unique 3D structure (order) or not (disorder), and hence does not describe the diversity of conformational states available to IDPs [4]. As mentioned before, many proteins are composed of both ordered and intrinsically disordered regions. Only $\sim 30\%$ of known "folded" protein structures have no disordered regions within their sequences [9]. On the other hand, many IDPs have significant amounts of transient secondary structure and for some, their structure can be thought of as a set of several folded conformations [1,

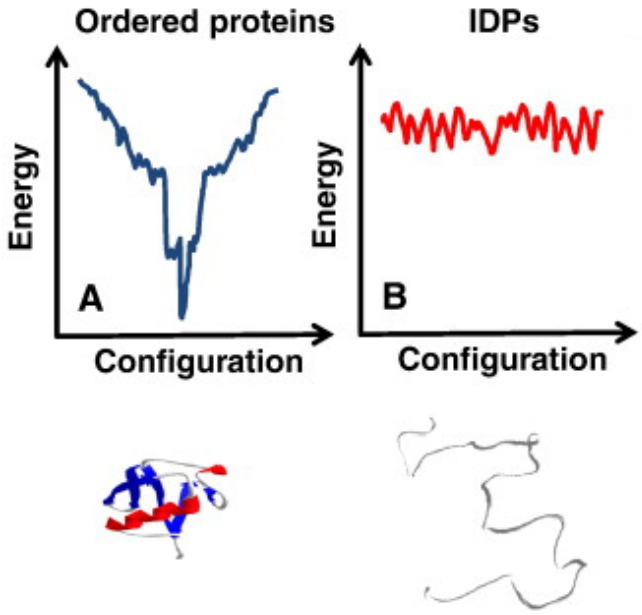


Figure 1.4: Comparison of two possible free energy landscapes of a folded protein (left) and an IDP (right). Figure adapted from van Uversky et al. [3].

4]. IDP conformations are also highly sensitive to the properties of the solution and the functions of many IDPs involve a disorder-to-order transition. In summary, the understanding of protein structure is complicated by: the presence of both ordered and disordered regions in proteins, fluctuations between order and disorder in time, disorder-to-order or order-to-disorder transitions as part of protein function, and the dependence of the degree of order and disorder on the environment [9]. The binary classification does not suffice and instead structure can be defined along a disorder-order continuum, as shown in Figure 1.5. Because a particular protein's structure is heavily modulated by it's environment, the definition of a single "native state" for ordered proteins is also called into question. Instead, a protein's environment contributes to its position on the disorder-order continuum [2–4, 21].

The disorder-order continuum can be coarsely partitioned into four types of protein structure. Folded proteins with a unique native structure are at one extreme. The molten globule is a collapsed state, with many elements of secondary structure, but lacking a single hydrophobic core, unlike folded proteins. The pre-molten globule is larger than the molten globule but still has some hydrophobic clusters and residual secondary structure. Finally, the most disordered IDPs fall into the random coil group. Experimental measures of polymer dimensions which will be discussed in Section 1.2 are capable of differentiating between these types of ensembles [1, 2, 4, 19].

In the free energy landscape picture, the disordered extreme corresponds to a relatively smooth landscape with similarly probable conformations. The ordered extreme corresponds to a single deep funnel or energy minimum. Intermediate locations on the disorder-order continuum correspond to landscapes with multiple minima of varying energies [3]. The dynamics of an IDP are determined by the ruggedness of the free energy landscape, where the heights of transition barriers between the IDP conformations correspond to the inter-conversion times.

A full characterization of an IDP will involve a description of all possible conformational states and the

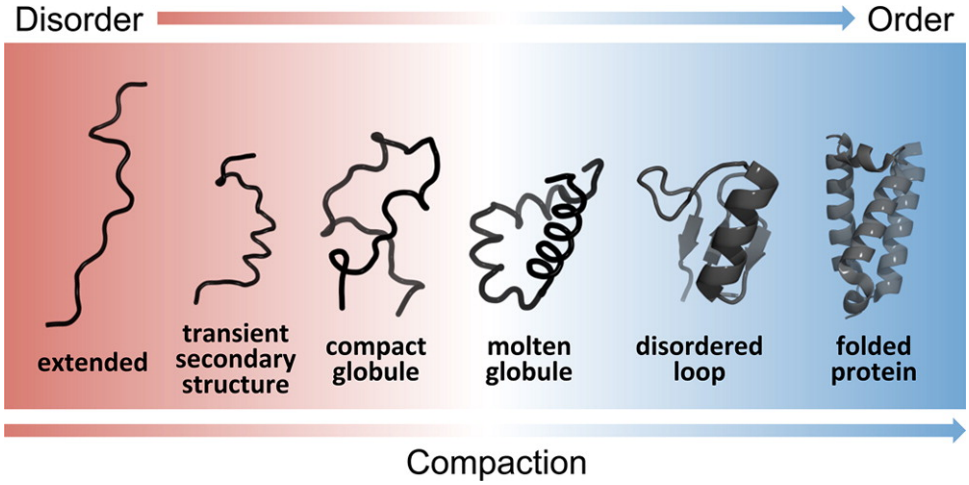


Figure 1.5: Order-disorder continuum of protein structure. Figure from van der Lee et al. [4].

rates of interconversion between them [20]. Practically, this is impossible and instead certain features of the ensembles of conformations can be probed experimentally using techniques such as nuclear magnetic resonance (NMR), small-angle x-ray scattering (SAXS), Förster resonance energy transfer (FRET), and dynamic light scattering (DLS). To understand the complete ensemble of conformations, experiments can be supplemented by simulations that reproduce the experimental data [2]. In the next section, I will describe the most common experimental observables and how they can be obtained from an ensemble of conformations.

1.2 Interpreting IDP Structure through Polymer Physics

While complete information about an IDP's state at every instant in time is impossible to obtain experimentally, several techniques can probe the conformations and reveal information about the ensembles of IDPs: NMR, fluorescence correlation spectroscopy (FCS), or DLS can measure the diffusion coefficient and hydrodynamic radius, FRET gives information about any inter-residue distance (such as the end-to-end distance), and SAXS can measure radius of gyration [2, 4, 22]. Polymer physics offers a theoretical framework for studying IDPs, linking experimental observables to conformational ensembles [2, 9, 23]. In this section I will describe a few of these observables, collectively referred to as polymer dimensions, how they are related to a polymer's ensemble of conformations, how they may be calculated from a polymer model or simulation, and the predictions of simple polymer theories for their scaling with the number of monomers and relationships between them. These relationships and scaling laws describe the universal behaviour of polymer systems independently of the details of local properties of the IDPs or solvent, and are useful in interpreting experimental results and identifying when IDPs deviate from ideal behaviour [23–27].

1.2.1 Measures of Polymer Dimensions

Consider a polymer composed of N monomers connected by $N-1$ links. A single conformation can be described by the positions $\{\vec{R}_i\}$ of it's monomers for $i \in \{1, 2, ..., N\}$. Depending on the choice of

model, the monomers of an IDP can be the individual atoms making up the polypeptide backbone, individual amino acids, or equivalent monomers corresponding to several atoms or amino acids [19, 28, 29]. A polymer ensemble comprises a large number of possible conformations, which can be described statistically in terms of probability distributions, of which moments are useful characteristics.

The end-to-end distance is measured between the first and last monomer. The ensemble average of the square end-to-end distance is:

$$\langle R_e^2 \rangle = \langle (\vec{R}_1 - \vec{R}_N)^2 \rangle \tag{1.1}$$

The radius of gyration is a more comprehensive measure of a polymer's size than the end-to-end distance, since it takes into account the positions of all the monomers rather than only two. The square radius of gyration is the average of all the square inter-monomer distances or equivalently the average square distance between each monomer and the average position of all the monomers ($\vec{R}_c = \frac{1}{N}\sum_{i=1}^{N} \vec{R}_i$). The ensemble average of the square radius of gyration is:

$$\langle R_g^2 \rangle = \frac{1}{N}\sum_{i=1}^{N} \langle (\vec{R}_i - \vec{R}_c)^2 \rangle = \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \langle (\vec{R}_i - \vec{R}_j)^2 \rangle \tag{1.2}$$

The hydrodynamic radius R_h is the radius of a solid sphere that has the same diffusion coefficient as the polymer. It depends not only on the ensemble of conformations but also the dynamics of interconversion between them. The diffusion coefficient of the polymer is inversely proportional to the hydrodynamic radius via the Stokes-Einstein relation [24, 25]:

$$D_e = \frac{k_B T}{6\pi\eta R_h} \tag{1.3}$$

where k_B is Boltzmann's constant, T is the absolute temperature, η is the solvent viscosity. D_e can be calculated by starting with the displacement of the polymer's centre of mass:

$$D(t) = \frac{\langle (\vec{R}_c(t) - \vec{R}_c(0))^2 \rangle}{6t} \tag{1.4}$$

This expression depends on t, the time at which the mean squared displacement is measured. The average is over sample paths, and the limit ensures that there are no dynamic correlations between the conformations. D_e is the long time limit of expression 1.4 [30, 31]:

$$D_e = \lim_{t\to\infty} D(t) \tag{1.5}$$

Kirkwood and Riseman [32] introduced a pre-averaging approximation for the hydrodynamic interactions between the monomers, allowing for an estimate of the hydrodynamic radius from just the ensemble of conformations. The Kirkwood approximation for the diffusion coefficient of a polymer, using the Oseen tensor for hydrodynamic interactions, is [30]:

$$D_k = \frac{k_B T}{6\pi\eta N^2}\left(\sum_{i=1}^{N}\left(\frac{1}{a_i} + \sum_{j=1, j\neq i}^{N} \langle |\vec{R}_i - \vec{R}_j|^{-1} \rangle\right)\right) \tag{1.6}$$

Here, a_i is the hydrodynamic radius of monomer i, and again the average is over the ensemble of

conformations. The Kirkwood approximation is equivalent to the short time limit ($t \to 0$) of the diffusion coefficient in Equation 1.4 [30, 31]. A derivation of this result is given in Appendix A. The inverse of the approximation to the hydrodynamic radius is defined as:

$$\langle R_k^{-1} \rangle = \frac{1}{N^2} \left(\sum_{i=1}^{N} \left(\frac{1}{a_i} + \sum_{j=1, j \neq i}^{N} \langle |\vec{R}_i - \vec{R}_j|^{-1} \rangle \right) \right) \tag{1.7}$$

For compactness of the notation in the following text, we often use a short form (without brackets representing averaging) to refer to the polymer dimensions defined above. R_e or end-to-end distance means $\sqrt{\langle R_e^2 \rangle}$, R_g or radius of gyration means $\sqrt{\langle R_g^2 \rangle}$, and R_k means $\langle R_k^{-1} \rangle^{-1}$. R_h refers to the long-time limit of the hydrodynamic radius as it is defined above. Any other meaning (such as the dimensions of a single conformation, or the average over conformations at a certain time) will be noted explicitly.

1.2.2 Predictions of Homopolymer Models

The ensemble of conformations of an IDP is governed by the inter-monomer and monomer-solvent interactions. In polymer physics terms, the balance of the monomer and solvent interactions can be related to the solvent quality or an effective monomer cohesiveness. Simple homopolymer models can categorize the ensembles into regimes of behaviour and make predictions about the ensemble averages of polymer dimensions in each regime [9, 23, 27]. In this section, I will describe the predictions for the relative polymer dimensions, valid for large N, and the scaling with of the polymer dimensions by the power law $R \propto N^{\nu}$ in three regimes most accessible by simple polymer theories.

At one extreme lies the good solvent regime, when the monomer cohesiveness is negligible and the only interactions are short range repulsion due to excluded volume. A polymer in a good solvent can be modeled by a self avoiding walk (SAW). $\langle R_e^2 \rangle / \langle R_g^2 \rangle$ is larger than that in a θ solvent at ~ 6.3 [33]. A theoretical estimate of the ratio of $\sqrt{\langle R_g^2 \rangle} \langle R_k^{-1} \rangle$ in this regime is 1.56. The scaling exponent of the polymer dimensions is $\nu \approx 0.6$ [25]. In a solution with high denaturant concentration, both natively folded proteins and IDPs follow a scaling law close to the good solvent regime [34–36].

When cohesive and repulsive interactions between monomers are balanced, a polymer is in a θ solvent, well modeled by a random walk or a Gaussian chain. The ratio between the square end-to-end distance and square radius of gyration is $\langle R_e^2 \rangle / \langle R_g^2 \rangle = 6$ and the ratio between the radius of gyration and the Kirkwood approximation for the hydrodynamic radius is $\sqrt{\langle R_g^2 \rangle} \langle R_k^{-1} \rangle$ is approximately 1.5. All polymer dimensions scale with $\nu = 0.5$ [24, 25, 37]. When considered together, IDPs in native conditions follow roughly θ solvent scaling [36, 38]

Decreasing the solvent quality is equivalent to increasing the strength of the cohesive interactions between the monomers resulting in collapse to a compact globule. In this limit, if the polymer is approximated as a solid sphere, $\sqrt{\langle R_g^2 \rangle} / R_h \approx 0.775$, and if each end is distributed uniformly inside the sphere [39, 40], $\langle R_e^2 \rangle / \langle R_g^2 \rangle = 2$. Any measure of the polymer size scales with $\nu = \frac{1}{3}$. Natively folded proteins follow this scaling law [41].

As a simple approximation, the ensemble of conformations of IDPs and even folded protein structures fall somewhere on this disorder-to-order continuum, depending on the balance of cohesive and repulsive interactions determined by the inter-residue interactions due to their sequence and the properties of the solvent [23]. Analysis of the relative polymer dimensions as well as comparison with predictions from the scaling laws at the good and poor solvent extremes can reveal a particular protein's location on this

continuum. Deviations from the predicted relative polymer dimensions for that regime would indicate deviations from the homopolymer theories and can suggest that the conformations are heavily influenced by the heterogeneity of the sequence.

1.3 Relationships between Sequence and Structure

Analyzing the amino acid sequences of IDPs gives insight into the principles behind their structures and consequently their functions. In general, IDP sequences have lower evolutionary conservation compared to folded proteins [18, 42]. IDPs sequences also have lower information content in their sequences due to the depletion of certain amino acids relative to folded proteins and repetition [4, 9, 21].

Today, there are more than 50 disorder predictors of varying levels of complexity. Even simple predictors based on amino acid composition are capable of a high level of accuracy. For example, a disorder or order promoting score can be assigned to each amino acid based on its composition (mean fraction) in IDPs relative to natively folded proteins. Generally IDPs are depleted in hydrophobic amino acids and enriched in charged, polar, and structure-breaking amino acids, such as Glycine and Proline [42]. A predictor based only on amino acid composition is capable of 87% accuracy when predicting disorder. After reducing the sequence alphabet by assigning each amino acid to just one of 4 types, this predictor performs almost as well as when using the full 20 amino acid alphabet [43]. Even a predictor based only on two properties: the net charge per residue and and hydrophobicity per residue of a protein, can differentiate between IDPs and folded proteins [1, 2, 4, 6, 17].

Due to the low conservation and complexity of IDP sequences, as well as the success of composition-based predictors, it is possible that, compared to the unique structures of natively folded proteins, IDP structures are less sensitive to the specific amino acid sequence. Instead, the fine details of amino acid interactions can be coarse-grained out and IDPs can often be understood in terms of the basic physical properties of amino acids such as charge, hydrophobicity, flexibility of the polypeptide backbone, and how these are affected by interaction with the solvent [44].

1.3.1 IDP Amino Acid Composition

The amino acid composition as well as low sequence complexity and conservation of IDPs points to the possibility of general relationships between an IDP's amino acid composition and its function. Prior to identifying the relationships to function, it is simpler to investigate the general relationships between an IDP's amino acid composition and its structure, and identify the sequence properties which may be responsible for an IDP's location on the disorder-order continuum.

In general, mean hydrophobicity is lower and mean net charge is higher in IDP sequences than in folded proteins and therefore may contribute to the degree of disorder. Often, a lack of larger hydrophobic amino acids prevents the folding of IDPs into a stable structures with a hydrophobic core like that of natively folded proteins. A higher net charge contributes to greater repulsion between different segments of IDPs. Using coarse-grained simulations, Ashbaugh and Hatch showed that sequence hydrophobicity and net charge control the transition between expanded coil and collapsed globule states [45]. Yamada et al. [46] measured the hydrodynamic radii of several FG nucleoporins and found that the ratio of the fraction of charged amino acids to the fraction of hydrophobic was enough to distinguish between the coil and globule regimes of behaviour. Furthermore, the relative expansion of Phenylalanine to Serine mutants showed that hydrophobicity could modulate polymer dimensions. Hofmann et al. [35]

used FRET measurements of variable length segments of IDPs to extract their scaling exponents. In water, the exponents were highly sequence dependent, with the average being slightly more compacted than at the θ point. There was a positive correlation between net charge and scaling exponent as well as a negative correlation between hydrophobicity and scaling exponent. However, the variable length subsequences of a highly heterogeneous IDP sequence may not share a single scaling exponent. At high denaturant concentrations, the exponents converged to $\nu \approx 0.62 \pm 0.3$ indicative of good solvent conditions or the disordered extreme of the continuum.

Marsh and Forman-Kay [38] investigated the relationship between composition and the relative expansion of hydrodynamic radius (R_h) for several IDPs. The expansion factor of each IDP was calculated relative to $R_h^{IDP} \coloneqq R_0 N^{\nu}$ for the same number of amino acids N, and the scaling relationship was obtained by fitting to all IDPs ($\nu = 0.509$). Net charge and proline content were the best predictors of expansion of R_h while hydrophobicity was only weakly correlated with compaction. Using FRET, Muller-Spath et al. [47] also showed that expansion is associated with net charge for several IDPs. However, they also showed that the properties of the solvent are equally important. A high net charge polypeptide will collapse at high ionic concentration, while a charge balanced one will expand.

Mao et al. [48] proposed a diagram of states (shown in Figure 1.6) relating an IDP's composition described by positive and negative charge fractions as well as hydrophobicity to the polymer ensemble. Using ABSINTH simulations supported by FCS measurements, they showed differences in the polymer dimensions of proteins with identical net charge but different numbers of positively and negatively charged residues. They proposed three compositional parameters to determine the regime of protein structure: net charge per residue, the fraction of charged amino acids, and the mean hydrophobicity. These classes roughly determine whether attractive or repulsive interactions dominate and whether IDPs form compact globule or expanded coil structures [23, 27].

Milles and Lemke [49] measured the compaction of the dimensions of several similar length fragments of Nup153 relative to their denatured state using FRET. The relative dimensions of the IDPs varied greatly (between 0.68 and 0.95), but surprisingly there were only weak correlations with the amino acid compositions. The ratio of fraction of charged to fraction of hydrophobic amino acids had no correlation with the relative dimensions and there was only a weak positive correlation between the combined fraction of Asparagine and Glutamine (N and Q) and relative dimensions. However, both of these amino acids were previously found to be associated with chain compaction [38], and only Glutamine (Q) is considered a disorder promoting amino acid [42]. Although there are trends between compositional properties such as proline content, net charge, and hydrophobicity, and an IDP's dimensions, the conflicting experimental findings suggest that the amino acid composition is not the only factor determining IDP structure.

1.3.2 Sequence Heterogeneity

One way proteins are distinct from other polymers is their extreme sequence heterogeneity. Functionally, local sequences of short linear motifs in IDPs are responsible for low-affinity interactions with binding targets [4]. However, permuting the sequence while keeping amino acid composition constant, also affects the global structure of IDPs, which may have implications for functions other than local binding [27]. Similar to the way that the experimental insight into IDP structure is limited to a few polymer dimensions, the sequence space that can be feasibly explored by experiment is limited to a small fraction of the total. Simulations offer a way to systematically study the vast sequence space and the effect of sequence heterogeneity on polymer dimensions.

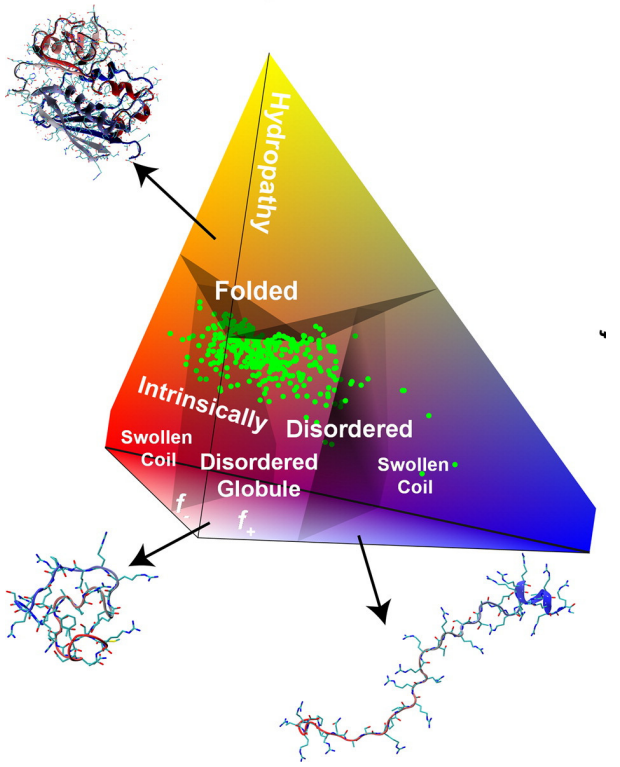


Figure 1.6: Phase diagram of IDP composition based on positively charged fraction (f_+), negatively charged fraction (f_i), and hydrophobicity. Figure from van der Lee et al. [4].

Das and Pappu [50] demonstrated the importance of sequence on IDP structure independent of composition using simulations of IDPs with equal numbers positively and negatively charged amino acids. All the sequences, had the same net charge and assuming only composition was important, should have had the same polymer dimensions. However, the radii of gyration of the sequences differed by $\sim 40\%$ between sequences of alternating positive and negative charges and sequences were each type of charge was confined to one half of the sequence. The parameter κ quantified the segregation of positive and negative charges along the sequence and was correlated with polymer dimensions. Martin et al. [51], used an analogous parameter (Ω) to describe the segregation of disorder promoting (Proline and charged residues) and order promoting (all others) amino acids in the sequence. This sequence parameter was based on the experimental results of Marsh and Forman-Kay [38] which showed that the Prolines and charged residues are correlated with expansion in R_h. The sequences they studied were permutations of the IDR of Ash1, and consequently had more a more realistic amino acid composition than those with only charged residues. Ω was similarly correlated with polymer dimensions, confirming that less segregation between the order promoting amino acids, independent of composition, results in compaction of IDPs.

On one hand, IDP sequences are not conserved across species, have low complexity, and their composition can predict disorder. However, sequence independent of composition has a great effect on polymer dimensions. Perhaps IDP sequences and structures can be represented by a model which captures the main properties without requiring the detailed interactions between all 20 amino acids. A few of these properties have been identified: the importance of electrostatic interactions, hydrophobicity, and

in general the association of certain amino acids with either expansion or compaction of the IDP. If a coarse-grained model built on these principles is sufficient for understanding the structures and dynamics of IDPs, it will be applicable to other polymer systems. Ultimately, any model must be informed by experiment. In the next section, I will review several experimental techniques for studying the structure and dynamics of IDPs, emphasizing the role of polymer theory in interpreting their results.

1.4 Experimental Methods for Probing IDP Ensembles

1.4.1 Extracting Polymer Dimensions from FRET

Single-molecule Förster resonance energy transfer (FRET) is an experimental technique which discerns between distances in the range of ∼2nm to ∼10nm, making it an appealing method of investigating IDP dimensions. FRET donor and acceptors dyes are attached to two positions, typically near the two opposing ends of an IDP sequence. Assuming rapid rotational reorientation of the the dyes, the rate of energy transfer will vary only with the distance between the two dyes. For dyes at a fixed distance R from each other, the FRET efficiency $E(R)$ is given by:

$$E(R) = \frac{1}{1 + \frac{R^6}{R_0^6}} \tag{1.8}$$

Where R_0 is the Förster radius (typically between 5 and 7 nm), which depends on the dye pair used and the refractive index of the medium between the dyes, and is independent of the IDP [22, 52–54].

IDPs typically undergo rapid fluctuations between their many conformational states and the distance between the FRET dye pair on an IDP will not remain static. The experimentally observed FRET efficiency, depends on the ensemble of conformations and dynamics of inter-conversion between them. Four timescales are important in the FRET measurement: the rotational correlation time of the dyes, the fluoresence lifetimes of the dyes, the end-to-end distance correlation time of the IDP, and the observation timescale. Typically, the rotation of the dyes ($\sim$ 100ps) is fast compared to their fluorescence lifetimes ($\sim$ 1ns), they are both faster than the end-to-end distance dynamics ($\sim$ 100ns), which is faster than the observation timescale ($\sim$ 1ms). Under these assumptions, the experimentally observed FRET efficiency E_{obs} is given by the expectation value :

$$E_{obs} \approx \langle E(R) \rangle = \int_{R_{min}}^{R_{max}} dR E(R) p(R) \tag{1.9}$$

where R_{min} and R_{max} are the closest and farthest possible separation of the dyes respectively, the latter being approximately equal to the contour length of the IDP, and $p(R)$ is the probability distribution of the dye separation. One common deviation from this behaviour, occurs due to the presence of two sub-populations in the ensemble of conformations. If the interconversion time between the two is longer than the FRET observation time, they will appear as two separate peaks on the FRET efficiency distribution. Each peak is then described by Eq. 1.9, with a different $p(R)$ for each subpopulation.

To extract chain dimensions from a FRET measurement, the conformational ensemble of the IDP, or at least the probability distribution of the dye separation, must be known. The chain dimensions are obtained by fitting the parameters of the distribution to reproduce the experimental FRET efficiency. If the dyes are placed near the IDP's ends, the approximation $p(R) \approx p(R_e)$ is made, making it simplest

to extract the end-to-end distance. The $p(R_e)$ provided by the Gaussian chain model (see Section 2.1.1) is usually sufficient in this case as the only fitting parameter is the mean square end-to-end distance $\langle R_e^2 \rangle$ and at high denaturant concentrations it gives an error that is at most $\sim 10\%$ when compared with simulations [55].

FRET is also often used to make inferences about other chain dimensions, which can be the source of discrepancies. For example, in the case of radius of gyration, the model needs to describe the probability distributions of both end-to-end distance and radius of gyration. Although the Gaussian chain model is successful in extracting end-to-end distance from FRET experiments, likely due to similarity of any typical R_e distribution to the Gaussian around its peak, attempting to extract the radius of gyration often leads predictions which contradict R_g measured by SAXS. The Gaussian chain model assumes the polymer is always in a θ solvent and gives $\frac{\langle R_e^2 \rangle}{\langle R_g^2 \rangle} = 6$ for all IDPs and solvent conditions. When compared with simulations, this leads to predictions with up to $\sim 20\%$ error in the radius of gyration [55].

A more detailed mean-field model, based on the Sanchez theory [56], includes both excluded volume and intrachain attractive interactions and can explain the coil-to-globule transition of polymers. Rather than end-to-end distance, the model only provides a probability distribution for radius of gyration $p(R_g)$ directly. As proposed by Ziv and Haran [57], the end-to-end distance distribution used in Eq. 1.9, is then constructed using:

$$p(R_e) = \int_{R_{g,min}}^{R_{g,max}} dR_g p(R_e|R_g) p(R_g) \tag{1.10}$$

where $p(R_e|R_g)$ is the probability distribution of distances between two random points in a sphere of radius $\sqrt{5}R_g$. The sphere radius is chosen to recreate the Gaussian chain behaviour at the θ point. This method has been criticized for not taking into account the asphericity of the conformations at each R_g [58]. It also suffers from the same defect as the Gaussian distribution, since $\frac{\langle R_e^2 \rangle}{\langle R_g^2 \rangle} = 6$ always, independent of the parameters and thus implicitly assumes the θ solvent relationship between the two dimensions. This assumption leads to an overestimate of R_g in good solvent conditions and an underestimate for poor solvents.

Simulations with various levels of coarse-graining can also be used to generate conformations and construct the $P(R_e)$, $P(R_g)$, and other distributions of polymer dimensions. However, care must be taken to prevent overfitting when selecting the parameters to match the experimental FRET efficiency. Another approach of extracting polymer dimensions from FRET data is based on sub-ensembles of conformations proposed by Song et al. [58, 59]. First, many conformations are generated with a particular model, such as the self-avoiding walk, and are then binned into conditional distributions $p(R_e|R_g)$. This method does not assume a distribution $p(R_g)$ but rather assumes that $\sqrt{\langle R_g^2 \rangle}$ will be approximately equal to the R_{g0} whose subensemble $p(R_e|R_{g0})$ most accurately reproduces the experimental FRET efficiency via:

$$E_{obs} \approx \int_{R_{e,min}}^{R_{e,max}} dR_e E(R_e) p(R_e|R_{g0}) \tag{1.11}$$

The sub-ensemble method illustrates the fact that there is not a consistent one-to-one relationship between different polymer dimensions such as R_g and R_e, for different polymer ensembles. Even homopolymer theories show that conformations in different regimes correspond to different ratios between their polymer dimensions, and using FRET data to derive other polymer dimensions without knowing the regime of behaviour may lead to inconsistencies.

1.4.2 Structural Characterization of IDP Conformations

Discrepancies between different methods of measuring chain dimensions called into question the methodology of dimension inference from FRET. For example, R_g of Protein L inferred using the Gaussian chain or Sanchez theory from FRET data, showed a much greater chain compaction compared to direct SAXS measurements of R_g as denaturant concentrations were reduced [60]. In an attempt to understand this discrepancy, SANS and FRET measurements were performed on the polymer PEG, which is known to be unaffected by denaturant concentrations. For varying denaturant concentrations, dye-labeled PEG exhibited changes in R_e obtained from FRET, comparable to other works studying proteins, while the R_g of unlabeled PEG obtained using SANS was unaffected [61]. Possible explanations are that denaturant affects the conformations of polymers in such a way as to affect R_e more than R_g, which may be due to interactions between the dyes [36, 62].

Borgia et al. [63] observed consistent chain expansion with increasing dentaurant, for the proteins R17 and ACTR, using multiple methods: FRET for R_e, SAXS for R_g, and FCS and DLS for R_h. In order to address previous discrepancies, ensembles were generated using the ABSINTH model and then fit to the raw experimental data of FRET and/or SAXS using Bayesian reweighting, so that the reweighted conformations differed minimally from the original ensemble. Ensembles reweighted either with FRET or SAXS data, or with both showed consistent values of R_g. The hydrodynamic radii calculated using HYDROPRO [64] from the reweighted conformations were in reasonable agreement with the FCS and DLS data, although this method assumes that each configuration remains rigid during the protein's diffusion. Zheng et al. [65] showed that explicit solvent all-atom molecular dynamics simulations of ACTR were also in accord with the experimental FRET and SAXS data and also suggested that the FRET dyes did not have a noticeable effect on the chain dimensions and that the previous inconsistencies were due to the method of extracting chain dimensions from FRET data.

Fuertes et al. [36] performed FRET and SAXS experiments of 10 chemically denatured proteins and IDPs, which further elucidated the source of the previous inconsistencies. Using SAXS measurements of both labeled and unlabeled proteins, they showed that FRET dyes do not significantly affect the scaling laws or chain dimensions and can be incorporated into a model as an extension of the amino acid chain by 5 ± 3 residues. The scaling exponent obtained from SAXS data demonstrated that, in the absence of denaturant, IDPs exhibit roughly θ solvent scaling, however the ratio of $\frac{\langle R_e^2 \rangle}{\langle R_g^2 \rangle}$ was 4.3, much less than the Gaussian chain prediction. They proposed that this disagreement was due to a decoupling of the two dimensions of end-to-end distance and radius of gyration. R_e is very sensitive to fluctuations of distances between the chain ends, while all inter-monomer distances contribute to R_g. To illustrate this, they reweighted a set of conformations generated using the ABSINTH model, using a similar procedure to Borgia et al. [63]. They showed that relative to the unweighted ensemble, reweighting using the FRET experimental data caused a greater relative change in R_e than R_g. By reweighting the ensembles to match SAXS data and different hypothetical FRET efficiencies, they showed that all but the most extreme FRET efficiencies could be observed with minimal changes to the ABSINTH force field, demonstrating that observing different R_e values with no change in R_g is not unexpected. The scaling of inter monomer distances $\langle\langle|\vec{R}_i - \vec{R}_j|\rangle\rangle$ with sequence separation $|i-j|$ of the different FRET efficiency subensembles, began to spread at half of the total sequence length. The ensembles for each of the different FRET efficiencies had different asphericities, which were only weakly coupled to the $\frac{\langle R_e^2 \rangle}{\langle R_g^2 \rangle}$ ratio. The subensemble method proposed by Song et al. [58], similarly concluded that the IDP ensembles differ from simple homopolymer models in that they are aspherical and this is responsible for

the decoupling of FRET dimensions from SAXS dimensions. Heterogeneity of the ensembles has a great effect on the relative polymer dimensions and explains the inconsistent results obtained using simple polymer models.

In summary, previous inconsistencies between FRET and SAXS experiments were due to: (i) assumptions in the polymer models used and (ii) heterogeneity in the conformational ensemble. In Chapter 4, I will systematically explore how the different polymer dimensions vary along the full disorder-order or good-bad solvent continuum and demonstrate that relative dimensions can vary even in a homopolymer model, as well as investigate the effects of sequence heterogeneity on the ensembles of conformations. With the aid of a polymer physics framework, two or more polymer dimensions obtained with different experimental techniques can reveal information about an IDP's ensemble of conformations, rather than being the source of a discrepancy.

1.4.3 End-to-end Distance Dynamics

FRET can be combined with Fluorescence Correlation Spectroscopy (FCS) to study the dynamics of IDP conformations. Fluctuations in the distance between the donor and acceptor fluorophores, usually placed at the ends of the chain, result in fluctuations of fluorescence intensity. In addition to obtaining information about the end-to-end distance with FRET, the correlation time of the fluorescence intensity can be related to the "reconfiguration time" describing the dynamics of the end-to-end distance [22].

Soranno et al. [66] investigated the end-to-end distance dynamics of the unfolded state of a cold shock protein (Csp) and two IDP domains. The unfolded state typically appears as a separate peak on the FRET intensity histogram due to long conversion times between the folded and unfolded states, and its ensemble of conformations can be studied separately. An increase in the denaturant concentration, caused a swelling of the end-to-end distance, which coincided with an initial decrease in reconfiguration time. Naively, we would expect the reconfiguration time to increase with longer end-to-end distances. The experimentally observed decrease is attributed to internal friction at lower denaturant concentrations, which is defined as the deviation from proportionality between reconfiguration time and solvent viscosity. The molecular origin of internal friction in IDPs is unclear and has been proposed to be due to dihedral angle transitions or cohesive interactions such as hydrogen bonds, but could be a combination of several intra-chain interactions [22, 67, 68]. In Chapter 5, using simulations of the coarse-grained model introduced in Section 2.3, I will investigate the end-to-end distance dynamics of polymers, and illustrate that cohesive interactions alone do not reproduce the denaturant dependence of reconfiguration time, and sufficient sequence heterogeneity is required.

1.5 Coarse-Grained Modeling of IDPs

Computational approaches allow direct sampling of the conformational states of IDPs and are therefore a great aid to experiments in studying structure and dynamics of IDPs. Challenges arise as experiments report averages which can correspond to many different ensembles of conformations. A balance needs to be struck between including detail in the models and overfitting the sparse experimental data [69].

All atom molecular dynamics (MD) simulations have been used as a tool in the structure prediction of natively folded proteins for several decades. There are several obstacles when applying these methods to IDPs. First, many MD force fields were parametrized specifically for folded proteins. And second, relative to folded proteins, computationally expensive simulations are needed to fully explore the conformational

space of an IDP [19, 20]. The predictions of molecular dynamics simulations for IDPs are very sensitive to the choice of force field [70–72]. Rauscher et al. [70] showed that eight different force fields gave divergent results for chain dimensions and secondary structure. In fact, the choice of the force field had a stronger effect on the amount and type of secondary structure present than the amino acid sequence itself.

Coarse graining of the atomistic details offers many advantages, especially in modelling the vast conformational space of IDPs [29]. Reduced system size, due to the use of bead models over all atom simulations and implicit over explicit solvents, results in lower computational complexity and the ability to sample longer times [19, 69, 73]. A reduced amino acid alphabet reduces the number of possible interactions which results in fewer parameters and is less likely to suffer from overfitting. If only structural observables are needed and information about dynamics is not required, implicit modelling of the solvent is often sufficient as it still gives a reasonable equilibrium distribution of the conformations.

The obvious flaw with coarse-graining is that it is only an approximation and may not include enough detail to be able to capture specific interactions. It is sometimes difficult to relate the parameters of a coarse grained model to real biophysical quantities and convert between model predictions and experimental results. Often, coarse-grained models must be parametrized based on all-atom MD simulations [74].

Proteins are polymers whose monomers are amino acids. Each amino acid is composed of an identical backbone and one of twenty variable residues, each made up of 10-26 atoms in total. There is a multitude of possible interactions between the 20 amino acids [75, 76]. The challenge of coarse-graining is to identify which of the properties and interactions of amino acids have the greatest influence on IDP structure [29, 69]. In the case of IDPs, the specific amino acids may be less important, than in natively folded proteins, since IDP sequences are less conserved and are of lower complexity [4, 9, 18, 21, 42]. The properties of low hydrophobicity, high net charge, and proline content are good predictors of disorder and expansion [1, 38, 42]. Sequence properties such as charge pattering can also tune the relative expansion of a protein [50].

1.6 Summary and Thesis Outline

At high concentrations of chemical denaturant, both IDPs and natively folded proteins obey the self-avoiding random walk scaling [34, 35]. At this extreme, the heterogeneity of the possible amino acid interactions becomes unimportant. This good solvent state can be thought of as the disordered extreme of the disorder-order continuum. At intermediate denaturant concentrations, natively folded proteins behave similarly to native IDPs, which fall somewhere on the disorder-to-order or good-to-poor solvent continuum. Both the solvent and the amino acid sequence determine the position of a protein on this continuum [3, 77].

One flaw with this picture is the view of the disorder-to-order continuum as one dimensional. It assumes that proteins falling at the same location on the continuum will share structural features such as the probability distribution of conformations, and consequently similar polymer dimensions, or follow the same scaling law. The continuum spreads out especially at the ordered extreme, where each natively folded protein will have a unique conformation and IDPs gain diverse transient secondary structure. Departures of the relative polymer dimensions from homopolymer behaviour can reveal departures of an IDP's structure from the single disorder-to-order continuum.

Incrementally constructing a coarse-grained model, by adding in levels of detail, and observing the

effects can reveal which properties of an IDP's sequence are responsible for its structure and dynamics. The simplest models - the Gaussian chain and the self-avoiding walk - can give reasonable shapes for the distributions of polymer dimensions, but do not explain why equal length IDPs have different polymer dimensions. Addition of a single cohesiveness parameter can tune the relative expansion/compaction of the IDP dimensions and determine where the IDP falls on the disorder-order continuum. This parameter integrates all of the sequence properties (both composition and patterning) as well as the properties of the surrounding solvent (temperature, salt, denaturant, pH, etc.). Finally, to explain the role of sequence in controlling IDP conformations, monomer heterogeneity must be included in the model, most simply in the form of a reduced amino acid alphabet.

In Chapter 2, I will describe several coarse-grained polymer models that can be applied to IDPs. Section 2.1 provides an overview of a few simple polymer physics theories that serve as a basis for the other models. Section 2.2 describes the mean field model used in Chapter 3 to investigate the behaviour of surface grafted FG nups and transport proteins. The model can explain the conflicting observations of *in vitro* experiments and reconcile the various theories proposed for the mechanism of transport through the Nuclear Pore Complex. The results of Chapter 3 have been partially reported in Refs. [78] and [79]. The model of Section 2.3 is used for the coarse-grained simulations of Chapters 4 and 5 and can accommodate various levels of detail such as sequence heterogeneity and hydrodynamic interactions. In Chapter 4, I report the predictions of homopolymer and heteropolymer models for relative polymer dimensions across the entire disorder-order continuum. The results of Chapter 4, provide an improvement over the simple polymer theories commonly used in the interpretation of experimental measurements of IDP dimensions, without resorting to more computationally demanding simulations. In Chapter 5, I investigate end-to-end distance dynamics of polymers and show that sequence heterogeneity is required to explain the increase in reconfiguration time with chain compaction of IDPs.

Chapter 2

Theory and Methods: Polymer Models of IDPs

In this chapter, I describe the polymer models used to study IDPs in Chapters 3, 4, and 5. In Section 2.1, I outline several simple models of single polymer chains. The models are used to make predictions about the relative polymer dimensions and the scaling of polymer dimensions with the number of monomers. The Lattice Model serves as the basis for the Polymer Brush Model of Section 2.2. The heights of layers of surface tethered FG nucleoporins of the Nuclear Pore Complex and the accumulation of transport proteins in the layers are investigated in Chapter 3 using the Polymer Brush Model. The effects of monomer sequence on the polymer dimensions and end-to-end distance dynamics of IDPs are investigated in Chapters 4 and 5 using the Coarse Grained Model of Section 2.3.

2.1 Models of an Isolated Polymer Chain

2.1.1 Gaussian Chain

Consider a chain of $N+1$ monomers. Each monomer is connected to its two nearest neighbours (one if it's on the end of the chain) by bonds of length b, for a total of N bonds in the chain, and does not interact with the solvent or non-neighbouring monomers. Ignoring non-local interactions, for large N, the probability density of the end-to-end distance is [37]:

$$p(R_e) = 4\pi R_e^2 \left(\frac{3}{2\pi N b^2}\right)^{\frac{3}{2}} e^{\left(-\frac{3R_e^2}{2Nb^2}\right)} \tag{2.1}$$

For this distribution, the mean square end-to-end distance is $\langle R_e^2 \rangle = Nb^2$, so we will define $R_0 := \sqrt{N}b$ (the scaling exponent is $\frac{1}{2}$ with N). The most probable end to end distance ($R_{e,mp}$) occurs at the maximum of $p(R_e)$: $R_{e,mp} = \sqrt{\frac{2}{3}}R_0 \approx 0.816R_0$ and the expected value of end to end distance is $\langle R_e \rangle = \sqrt{\frac{8}{3\pi}}R_0 \approx 0.921R_0$.

Using the Gaussian distribution as the probability distribution of the distance between any two monomers i and j, we can calculate the radius of gyration (R_g) and the Kirkwood approximation to the

hydrodynamic radius (R_k) of the Gaussian chain [80]:

$$\langle R_g^2 \rangle \approx \frac{1}{2N^2} \int_{i=0}^{N} di \int_{j=0}^{N} dj \langle (\vec{R}_i - \vec{R}_j)^2 \rangle = \frac{R_0^2}{6} \tag{2.2}$$

$$\langle R_k^{-1} \rangle \approx \frac{1}{N^2} \int_{i=0}^{N} di \int_{j=0}^{N} dj \langle |\vec{R}_i - \vec{R}_j|^{-1} \rangle = \sqrt{\frac{128}{3\pi}} \frac{1}{R_0} = \frac{8}{3\sqrt{\pi}} \frac{1}{\sqrt{\langle R_g^2 \rangle}} \approx \frac{1.5}{\sqrt{\langle R_g^2 \rangle}} \tag{2.3}$$

2.1.2 Collapsed Globule

For the case of a poor solvent or very strong attraction between the monomers, it is appropriate to model the polymer as a sphere of radius R_h and of uniform density [58]:

$$\rho(r) = \begin{cases} c & \text{if } r \leq R_h \quad (2.4) \\ 0 & \text{if } r > R_h \quad (2.5) \end{cases}$$

where r is the distance from the sphere's centre. The radius of gyration is then [81]:

$$\langle R_g^2 \rangle = \frac{\int_0^{R_h} r^2 \rho(r) 4\pi r^2 dr}{\int_0^{R_h} \rho(r) 4\pi r^2 dr} = \frac{3}{5} R_h^2 \tag{2.6}$$

$$\sqrt{\langle R_g^2 \rangle} \approx 0.775 R_h \tag{2.7}$$

The end to end distance may be calculated by assuming the first and last monomers of the polymer are located at two random points inside the uniform sphere of radius R_h which follows the distribution [39, 40]:

$$P\left(u := \frac{R_e}{R_h}\right) \begin{cases} \frac{3}{16}u^5 - \frac{9}{4}u^3 + 3u^2 & \text{if } u \leq 2 \quad (2.8) \\ 0 & \text{if } u > 2 \quad (2.9) \end{cases}$$

We can find the expectation values:

$$\frac{\langle R_e \rangle}{R_h} = \frac{36}{35} \text{ and } \frac{\langle R_e^2 \rangle}{R_h^2} = \frac{6}{5} \text{ which leads to: } \frac{\langle R_e^2 \rangle}{\langle R_g^2 \rangle} = 2 \tag{2.10}$$

Under the assumption that the distance between any two monomers follows the distribution of the distance between two random points inside of a sphere [58], the prediction of the Kirkwood approximation is:

$$\langle R_k^{-1} \rangle = \frac{6}{5} \frac{1}{R_h} = \frac{6\sqrt{3}}{5\sqrt{5}} \frac{1}{\sqrt{\langle R_g^2 \rangle}} \approx \frac{0.93}{\sqrt{\langle R_g^2 \rangle}} \tag{2.11}$$

In general, any measure of the polymer size scales as $R \sim N^{\frac{1}{3}}$.

2.1.3 Lattice Model

Here, we follow the theory introduced by Flory for polymer solutions [25, 82, 83], to study the effects of excluded volume and cohesive interactions on the dimensions of a single polymer. In Section 2.2, we will

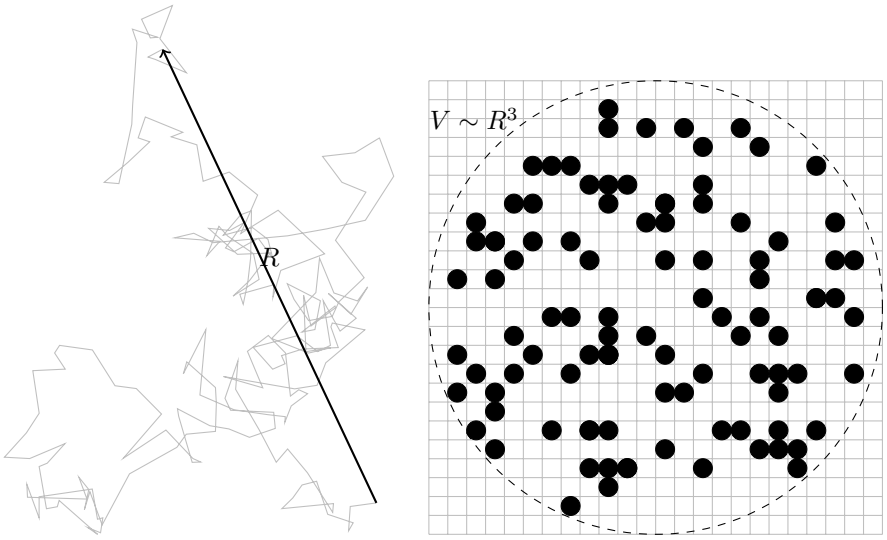


Figure 2.1: Lattice model of a single polymer. *(a)* the end-to-end distance distribution follows that of a Gaussian chain. *(b)* when counting interactions between monomers, we assume they are placed randomly on a lattice.

use this theory to construct a model for a polymer brush in a nano-particle solution. We will begin using an approximate self-avoiding walk model, by reweighing the end-to-end distance probabilities of the Gaussian chain model by the probability that each lattice cite is unoccupied as we place the monomers.

Consider a polymer on a lattice of spacing b and coordination number z. Each monomer or solvent molecule occupies a single lattice site. We will calculate the free energy of the monomer and solvent system as a function of the polymer size R. The polymer occupies a volume of roughly R^3, equivalent to $V = \frac{R^3}{v_0}$ lattice sites, where v_0 is a volume parameter reflecting the size of one monomer. We assume that the polymer and solvent particles are evenly distributed over the V lattice sites (see Figure 2.1 (b)). The total number of polymer conformations with polymer size R is equal to the total number of possible conformations (number of random walks of length N), multiplied by the probability distribution of the end-to-end distance of the Gaussian chain (probability that the conformation has size R and R roughly corresponds to the end-to-end distance, see Figure 2.1 (a)), and finally multiplied by the probability that each lattice cite is unoccupied as we place the monomers [37, 83]. We also approximate the number of monomers $N + 1 \approx N$ as equal to the number of bonds, which will not make a difference in the final result.

First we will count the total number of polymer conformations, ignoring excluded volume and only considering chain connectivity. Our result should be translationally invariant to the centre-of-mass of the polymer in solution, so the position of the first monomer is irrelevant. There are z configurations for the second monomer, and then $z - 1$ for every subsequent monomer since a bond may not go back on itself. For a polymer of N bonds, the total number of conformations, ignoring excluded volume is:

$$\Omega_{0,Total} = z(z-1)^{N-1} \approx (z-1)^N \tag{2.12}$$

Next, we amend this expression by considering the probability that as we place each monomer, the lattice sites are unoccupied by the already present monomers. For every monomer $i + 1$ that we place,

we multiply its possible configurations by the probability that each lattice site is unoccupied by the previous i monomers ($\frac{V-i}{V}$), assuming they are uniformly distributed within V. This is incorrect locally, as there is no way for monomer $i+1$ to be placed on the same site as i, or $i-1$ since we cannot go back along a bond. Beyond this, the probability of interfering with local monomers depends on the lattice coordination number. Nevertheless, we will see that this is a good approximation. The total number of conformations when excluded volume is considered is:

$$\Omega_{Total} \approx (z-1)^N \prod_{i=0}^{N-1} \frac{V-i}{V} = (z-1)^N \frac{V!}{(V-N)!V^N} \tag{2.13}$$

Using Stirling's approximation $\ln n! \approx n \ln n - n$, and defining $\psi := \frac{N}{V} = \frac{Nv_0}{R^3}$ as the volume fraction of monomers (and $(1-\psi)$ is the volume fraction of solvent):

$$-\ln \Omega_{Total} \approx V \left(\psi \ln \left(\frac{e}{z-1} \right) + (1-\psi)\ln(1-\psi) \right) \tag{2.14}$$

Additionally, we multiply the total number of conformations by the probability that the end-to-end distance is R:

$$\Omega(R) = \Omega_{Total} 4\pi R^2 \left(\frac{3}{2\pi N b^2} \right)^{\frac{3}{2}} e^{\left(-\frac{3R^2}{2Nb^2}\right)} \tag{2.15}$$

Since $R^3 = v_0 V$, the first term in the parentheses in Equation 2.14, which is first order in ψ, will be independent of R (when multiplied by V), and can be ignored. Along with the contribution from the end-to-end distance distribution (Equation 2.15), the free energy of the polymer, with excluded volume interaction is:

$$\beta F(R) = \frac{3R^2}{2Nb^2} - 2\ln R + \frac{R^3}{v_0}(1-\psi)\ln(1-\psi) + \text{ terms independent of } R \tag{2.16}$$

Effect of Inter-monomer Cohesiveness

We now consider the energy of the interactions among the monomer and solvent molecules. We will approximate the energy of all configurations of a polymer of size R by the same mean field approximation $\bar{E}(R)$, which assumes that the monomer and solvent molecules are uniformly mixed within the volume occupied by the polymer, so that the free energy can be written as $\beta F(R) = -\ln \Omega(R) + \beta \bar{E}(R)$.

Consider that there are W total lattice sites in solution, and each can be occupied by one solvent molecule or one monomer. In the volume occupied by the polymer, which has V lattice sites, there are N monomers, and $V-N$ solvent molecules, with volume fractions ψ and $(1-\psi)$ respectively. Outside of the V lattice sites, we have $W-V$ solvent molecules with volume fraction 1. We assume that in our lattice model, molecules interact only with their z nearest neighbours on the lattice, with interaction energies ϵ_{mm} between two non-bonded monomers, ϵ_{ms} between a monomer and a solvent molecule, and ϵ_{ss} between two solvent molecules. All except two of the monomers (at the polymer's ends) are bonded to two other monomers, so they only have $z-2$ adjacent non-bonded lattice sites, which have a probability ψ that they are occupied by a monomer and $(1-\psi)$ that they are occupied by a solvent molecule [37, 83]. The average number of neighbouring monomer-monomer (N_{mm}) and monomer-solvent (N_{ms}) pairs

on the lattice is:

$$N_{mm} = \frac{1}{2}N(z-2)\psi \tag{2.17}$$

$$N_{ms} = N(z-2)(1-\psi) \tag{2.18}$$

$$\tag{2.19}$$

The solvent molecules may interact with all z of their neighbours. The average number of neighbouring solvent-solvent pairs on the lattice is:

$$N_{ss} = \frac{1}{2}(W-V)z + \frac{1}{2}(V-N)z(1-\psi) = \frac{1}{2}z(W-2N+N\psi) \tag{2.20}$$

The energy of the system with polymer size R is:

$$\bar{E}(R) = N_{mm}\epsilon_{mm} + N_{ms}\epsilon_{ms} + N_{ss}\epsilon_{ss} \tag{2.21}$$

$$= \frac{1}{2}V\psi^2(\epsilon_{mm}(z-2) + \epsilon_{ss}z - 2\epsilon_{ms}(z-2)) + \text{ terms indepdendent of } R \tag{2.22}$$

The total interaction energy can be written in terms of the number of monomer-monomer interactions plus terms independent of R which we can again neglect. We will use the dimensionsless parameter $\chi_{cr} = \beta(\epsilon_{mm}(z-2) + \epsilon_{ss}z - 2\epsilon_{ms}(z-2))$ to represent the balance of monomer-monomer and solvent-solvent, against monomer-solvent interactions. Negative χ_{cr} corresponds to a poor solvent or high monomer cohesiveness, favouring polymer compaction, while 0 χ_{cr} corresponds to a good solvent, favouring expansion of the polymer size. χ_{cr} can be related to the second virial coefficient of monomer-monomer interactions in the more detailed model of Section 2.3. The R dependent term of the average interaction energy is:

$$\beta\bar{E}(R) = \frac{1}{2}N\psi\chi_{cr} = \frac{R^3}{v_0}\left(\frac{1}{2}\chi_{cr}\psi^2\right) \tag{2.23}$$

Combining this with Equation 2.16 gives:

$$\beta F(R) = \frac{3R^2}{2Nb^2} - 2\ln R + \frac{R^3}{v_0}\left((1-\psi)\ln(1-\psi) + \frac{1}{2}\chi_{cr}\psi^2\right) \tag{2.24}$$

Taylor expansion

For small ψ we can approximate $(1-\psi)\ln(1-\psi) \approx -\psi + \frac{1}{2}\psi^2$, we have:

$$\beta F(R) = \frac{3R^2}{2Nb^2} - 2\ln R + \frac{R^3}{v_0}\left(\frac{1}{2}\psi^2(1+\chi_{cr})\right) + \text{ terms independent of } R \tag{2.25}$$

To solve for the equilibrium polymer size, we minimize the free energy over R:

$$\frac{\beta dF(R)}{dR} \approx \frac{3R}{Nb^2} - \frac{2}{R} + \frac{3R^2}{v_0}\left(-\frac{1}{2}(1+\chi_{cr})\psi^2\right) \tag{2.26}$$

θ Solvent

When $\chi_{cr} = -1$ the repulsive and attractive interactions are balanced and we regain the Gaussian chain expression:

$$\beta F(R) = \frac{3R^2}{2Nb^2} - 2\ln R \tag{2.27}$$

$$\frac{\beta dF(R)}{dR} = \frac{3R}{Nb^2} - \frac{2}{R} \tag{2.28}$$

The minimum of the free energy is at $R_* = \sqrt{\frac{2}{3}}\sqrt{N}b$ as we found previously. This is the θ solvent or θ temperature: the transition point between a coil and a globule. A solution of polymers undergoes a separation into a mixture of dense and dilute phases.

Good Solvent

In a good solvent, there are no attractive interactions between the monomers ($\chi_{cr} = 0$). The minimization of the free energy over R (Equation 2.26) gives the equilibrium polymer size R_*. It is convenient to use our previous definition the Gaussian chain end-to-end distance $R_0 := \sqrt{N}b$. After some rearranging:

$$0 = \frac{R_*^5}{R_0^5} - \frac{2R_*^3}{3R_0^3} - \frac{v_0\sqrt{N}}{2b^3} \tag{2.29}$$

For large N, the second term on the right hand side is negligible, and we can solve for R_*:

$$\left(\frac{R_*}{R_0}\right)^5 = \frac{v_0\sqrt{N}}{2b^3} \tag{2.30}$$

$$R_* = \left(\frac{v_0 b^2}{2}\right)^{\frac{1}{5}} N^{\frac{3}{5}} \tag{2.31}$$

The scaling exponent ν is remarkably close to a more accurate estimate of 0.588 [25, 37].

Poor Solvent

For $\chi_{cr} < -1$, we expect the chain to compact and the monomer volume fraction ψ to increase. Our second order expansion for $\ln(1-\psi)$ is no longer valid, so we will include a third order term ($\ln(1-\psi) \approx -\psi - \frac{1}{2}\psi^2 - \frac{1}{3}\psi^3$):

$$\frac{\beta dF(R)}{dR} = \frac{3R}{Nb^2} - \frac{2}{R} + \frac{3R^2}{v_0}\left(-\frac{1}{2}(1+\chi_{cr})\psi^2 - \frac{1}{3}\psi^3\right) \tag{2.32}$$

Denoting R_* as the polymer size which minimizes the free energy and rearranging:

$$0 = \left(\frac{R_*}{R_0}\right)^8 - \frac{2}{3}\left(\frac{R_*}{R_0}\right)^6 - \frac{(1+\chi_{cr})v_0\sqrt{N}}{2b^3}\left(\frac{R_*}{R_0}\right)^3 - \frac{v_0^2}{3b^6} \tag{2.33}$$

For a compact chain, $R_* < R_0$ (R_0 is the equilibrium end-to-end distance of the Gaussian chain) and and the first two terms are negligible:

$$\left(\frac{R_*}{R_0}\right)^3 \approx -\frac{2v_0}{3b^3(1+\chi_{cr})\sqrt{N}} \tag{2.34}$$

The chain size scales as $R \sim N^{\frac{1}{3}}$ as in the uniform density sphere limit.

2.2 Mean Field Model of the Grafted Polymer Brush

In this section, I will extend the lattice model for a single polymer of Section 2.1.3 to describe a general model for a grafted polymer brush in a solution of nanoparticles. The aim of the model is to capture the behaviour of the physical system with a minimal number of important parameters. The steric effects are included by placing the monomers and nanoparticles on a lattice based on the Flory-Huggins theory of polymer solutions [82]. The entropic elasticity of the chains is described by the Alexander-de Gennes brush model which assumes that the monomer density is uniform inside the brush [84, 85]. Cohesive interactions among the monomers and between the monomers and nanoparticles are included via the mean field approximation. In the relevant parameter regime, the mean field model has been verified by self-consistent field theory and coarse grained Brownian dynamics simulations [86, 87].

The mean field approximation to the free energy of the system can be written in the form:

$$F = -k_B T \ln \Omega + \bar{E} \tag{2.35}$$

where Ω is the number of possible configurations of the system or multiplicity and $\bar{E}$ is the approximation to the energy of the configurations. The layer height and number of nano-particles in the layer at equilibrium are found by minimizing the free energy.

The first contribution to the multiplicity and subsequently the free energy is solely from the entropic elasticity of the polymer brush. The brush is composed of N_p polymers. Each polymer in the brush is composed of N monomers connected by bonds of length b. For a Gaussian chain (see Section 2.1.1), the probability density of the end-to-end vector is [25]:

$$P(\vec{R}) = \left(\frac{3}{2\pi N b^2}\right)^{\frac{3}{2}} \exp\left(-\frac{3\vec{R}^2}{2Nb^2}\right) \tag{2.36}$$

We will assume that one end of each polymer is grafted to a surface while the other extends to the layer height h. The elastic contribution to the free energy which resists stretching of all the polymers in the layer is [84, 85, 88]:

$$\beta F_{el} = \frac{h^2 N_p}{2Nb^2} + \text{terms independent of } h \tag{2.37}$$

In order to account for excluded volume effects and the configurational entropy of the nano-particles, we will connect the physical system to a lattice description, in which each monomer occupies one lattice site as in Section 2.1.3. The polymers are attached to a surface with a grafting distance of a and the polymer brush occupies a volume of $N_p a^2 h$. The monomers have a size l (volume l^3) and their volume fraction in the brush is $\psi = \frac{N N_p l^3}{N_p a^2 h}$. To count the lattice sites, we normalize distances by the monomer size: the dimensionless grafting distance and bond length are $\bar{a} = \frac{a}{l}$ and $\bar{b} = \frac{b}{l}$ respectively. We rewrite the monomer volume fraction as $\psi = \frac{N N_p}{V}$, where V is the number of lattice sites available to the monomers: $V = N_p \bar{a}^2 \frac{h}{l}$. Additionally, there are M nanoparticles present in the brush. The model assumes that the nanoparticles are larger than the monomers and their volume is $\bar{v} l^3$ ($\bar{v} \geq 1$) (each nanoparticle occupies $\bar{v}$ lattice sites). The volume fraction of the nanoparticles in the brush is $\phi = \frac{\bar{v} M}{V}$.

Using the lattice description, we will estimate the configurational entropy of the nano-particles and

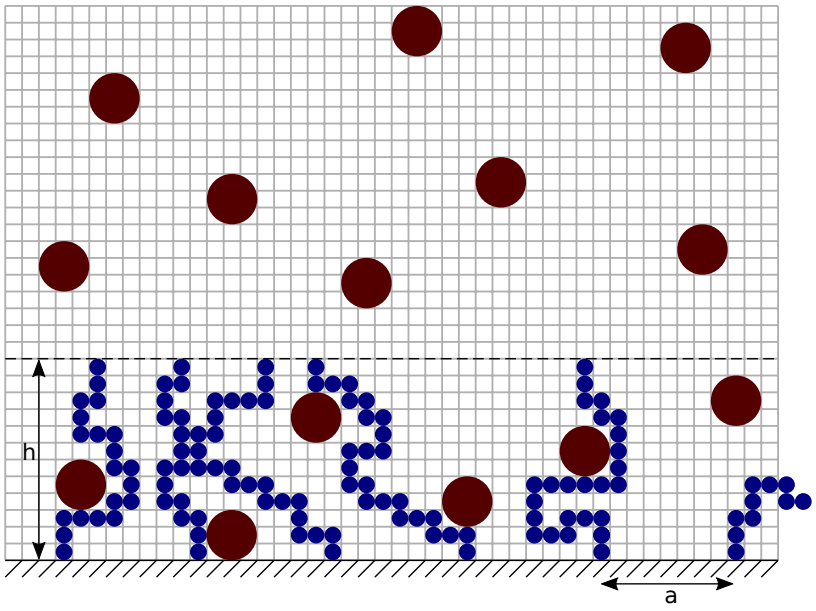


Figure 2.2: Schematic of the brush model. The monomers are in blue. The ends of the polymers are attached to a surface with grafting distance a. The other end extends to a height h defining the layer. Nanoparticles, which are both in the polymer layer and the external solution, are shown in red. Monomers occupy single sites and nanoparticles occupy multiple sites of the same lattice.

excluded volume effects of the monomers and nano-particles. Since the nanoparticles are larger than the monomers, they are placed first on the lattice to not overestimate the multiplicity [87]. The number of ways of placing M particles of volume $\bar{v}$ on V lattice sites is:

$$\Omega_{particles}(V, M) = \frac{(V/\bar{v})!}{M!(V/\bar{v} - M)!} \tag{2.38}$$

The configurational entropy of the monomers, ignoring excluded volume, is already included in the F_{el} term. To account for excluded volume interactions of the monomers, the previous term (Eq. 2.38) must be multiplied by the probability that each lattice site was unoccupied as the monomers were placed onto it. This probability is equal to the number of ways of placing NN_p distinguishable particles (monomers) on the $V - \bar{v}M$ remaining lattice sites divided by the total number of lattice sites (V) available for each monomer.

$$\Omega(V, N, N_p, M) = \frac{(\frac{V}{\bar{v}})!(V - \bar{v}M)!}{M!(\frac{V}{\bar{v}} - M)!(V - \bar{v}M - NN_p)!V^{NN_p}} \tag{2.39}$$

Using Stirling's approximation ($\ln n! \approx n \ln n - n$), we obtain the second contribution to the free energy [89]:

$$\begin{aligned}\beta F_2 = -\beta \ln \Omega =& V\left(\frac{1}{\bar{v}}\phi \ln \phi + (\frac{1}{\bar{v}} - 1)(1 - \phi)\ln(1 - \phi) + (1 - \phi - \psi)\ln(1 - \phi - \psi)\right) \\ &+ \text{ terms indepdendent of } h\end{aligned} \tag{2.40}$$

Similar to the single polymer case, we ignore the term $V\psi = NN_p$, which does not vary with layer height.

Inside the large parentheses, the first term corresponds to the translational entropy of the nano-particles, the second to the entropy reduction due to the volume difference between monomers and nanoparticles, and the third to the translational entropy of the solvent molecules.

The final contribution to the free energy is the mean field approximation of the attractive monomer-monomer and monomer-nanoparticle interactions. The average number (assuming monomers and nanoparticles are well mixed) of monomer-monomer interactions is $\frac{1}{2}V\psi^2$ and the number of monomer-nanoparticle interactions is $\frac{V}{\bar{v}}\phi\psi$. The energy of the average configuration is:

$$\beta\bar{E} = V\left(\frac{1}{2}\chi_{cr}\psi^2 + \frac{1}{\bar{v}}\chi\phi\psi\right) \tag{2.41}$$

The dimensionless parameters χ_{cr} and χ describe the strengths of the monomer-monomer and monomer-nanoparticle interactions, respectively. They roughly correspond to the attractive parts of the second virial coefficients as explained in Section 2.3.5 for the monomer-monomer case.

The three contributions combine to the following expression for the free energy:

$$\beta F = \frac{h^2 N_p}{2Nb^2} + V\left(\frac{1}{\bar{v}}\phi\ln\phi + (\frac{1}{\bar{v}}-1)(1-\phi)\ln(1-\phi) + (1-\phi-\psi)\ln(1-\phi-\psi) + \frac{1}{\bar{v}}\chi\phi\psi + \frac{1}{2}\chi_{cr}\psi^2\right) \tag{2.42}$$

Normalized by the area of the brush $N_p a^2$ and the number of monomers N, this expression becomes:

$$\frac{\beta F}{NN_p a^2} = \frac{h^2}{2N^2b^2a^2} + \frac{h}{Nl^3}f(\psi,\phi) \tag{2.43}$$

Where:

$$f(\psi,\phi) = \frac{1}{\bar{v}}\phi\ln\phi + (\frac{1}{\bar{v}}-1)(1-\phi)\ln(1-\phi) + (1-\phi-\psi)\ln(1-\phi-\psi) + \frac{1}{\bar{v}}\chi\phi\psi + \frac{1}{2}\chi_{cr}\psi^2 \tag{2.44}$$

The layer height is normalized by the polymer contour length $\tilde{h} = \frac{h}{Nb}$. The monomer volume fraction and layer height have an inverse relationship which is expressed as: $\psi = \frac{1}{\tilde{h}\bar{a}^2\bar{b}}$.

Following this normalization the free energy is:

$$\tilde{F} = \frac{\beta F}{NN_p\bar{a}^2\bar{b}} = \frac{\tilde{h}^2}{2\bar{a}^2\bar{b}} + \tilde{h}f(\psi,\phi) \tag{2.45}$$

2.2.1 Brush Without Nanoparticles

In the absence of nanoparticles ($\phi = 0$), and for a specific grafting distance, bond length, and monomer cohesiveness, the free energy (Eq. 2.45) minimized over the layer height gives an equation for the equilibrium layer height (or monomer volume fraction ψ):

$$0 = \frac{1}{\psi\bar{a}^4\bar{b}^2} + \psi + \ln(1-\psi) - \frac{\chi_{cr}}{2}\psi^2 \tag{2.46}$$

Setting $\chi_{cr} = 0$ and using an approximation for low monomer volume fraction ($\ln(1-\psi) \approx -\psi - \frac{1}{2}\psi^2$), recovers the $h \sim a^{-\frac{2}{3}}$ relationship for a polymer brush in a good solvent [85, 88]. In a poor solvent, the brush cannot be compacted beyond the total volume of the monomers: $ha^2 = Nl^3$ and therefore

$h \sim a^{-2}$.

2.2.2 Nanoparticle Solution

Nanoparticles can move between the polymer layer (which extends to height h) and the external solution. We can derive the free energy of the external nanoparticle solution using a similar lattice construction (beginning with a multiplicity similar to Equation 2.38):

$$\beta F_c = W[c \ln c + (1-c)\ln(1-c)] \tag{2.47}$$

where W is the number of lattice sites available to the nanoparticles in the external solution, and c is the volume fraction of nanoparticles in that solution. Nanoparticles will exchange between the layer and external solution, so that at equilibrium, the chemical potential of the nanoparticles and osmotic pressure of the polymer layer and external solution are equal.

For the external solution, the chemical potential of nanoparticles is:

$$\beta\mu_c = \beta\frac{\partial F_c}{\partial c} = \ln\frac{c}{1-c} \tag{2.48}$$

and the osmotic pressure is:

$$\beta\Pi_c = -\frac{\beta}{\bar{v}l^3}\frac{\partial F_c}{\partial W} = -\frac{1}{\bar{v}l^3}\ln(1-c) \tag{2.49}$$

Therefore, finding the equilibrium conditions is equivalent to minimizing the normalized grand potential $\tilde{\Phi}$, over the layer height $\tilde{h}$ and volume fraction of nanoparticles in the layer ϕ:

$$\begin{aligned}\tilde{\Phi}(\tilde{h},\phi) = &\frac{\tilde{h}^2}{2\bar{a}^2\bar{b}} \\ &+ \tilde{h}\left(\frac{1}{\bar{v}}\phi\ln\phi + (\frac{1}{\bar{v}}-1)(1-\phi)\ln(1-\phi) + (1-\phi-\psi)\ln(1-\phi-\psi) + \frac{1}{\bar{v}}\chi\phi\psi + \frac{1}{2}\chi_{cr}\psi^2\right) \\ &- \frac{\tilde{h}\phi}{\bar{v}}\ln\frac{c}{1-c} - \frac{\tilde{h}}{\bar{v}}\ln(1-c)\end{aligned} \tag{2.50}$$

where the last two terms ensure that the chemical potential of the nanoparticles and the osmotic pressure in the external solution are equal to those of the layer at equilibrium.

2.2.3 Conversion from volume fraction to experimentally measurable concentration

For M particles of volume $\bar{v}l^3$ in a solution of total volume V, their concentration is $m = \frac{M}{V}\frac{(0.1\text{m})^3}{N_A}$mol/L and their volume fraction is $\phi = \frac{l^3\bar{v}M}{V}$. In the lattice model, monomers and nano-particles completely fill the space, similar to neatly stacked cubes. In reality, the situation will be closer to that of randomly packed spheres, which occupy roughly 0.5-0.64 of the volume [90]. To convert from real concentrations to volume fractions in the model, we divide by the packing fraction. The packing fraction corrected conversion between volume fraction and concentration is:

$$m = 0.625 \times \frac{\phi}{l^3\bar{v}}\frac{(0.1\text{m})^3}{N_A}\text{M} \tag{2.51}$$

2.2.4 Effective Monomer Size and Bond Length

In our lattice model, all polymers are composed of N monomers, separated by bonds of length b. Every monomer occupies a lattice site or a volume of $\sim l^3$. The molecular details of real systems are more complex. To simplify the conversion between the lattice model and a real system, we outline a simple way by which we can redefine the "monomers" of our model. The original chain is composed of N_1 monomers, of size l_1, freely jointed by bonds of length b_1. Assuming the bonds are relatively stiff, we redefine the monomers to each include m of the original monomers, so that the chain is now composed of $N_2 = \frac{N_1}{m}$ monomers. The new monomers will have bond length $b_2 = \frac{N_1}{N_2} b_1 = m b_1$, volume $l_2^3 = \frac{N_1}{N_2} l_1^3 = m l_1^3$. The contour length $N_1 b_1 = N_2 b_2$ will be preserved.

2.2.5 Summary of Variables

N number of monomers in a polymer chain
N_p number of polymer chains in the brush
l monomer size
l^3 monomer (lattice site) volume
V number of lattice sites in the brush (normalized volume of the brush)
M number of nanoparticles in the brush
$\bar{v}$ number of lattice sites occupied by one nanoparticle (normalized nanoparticle volume)
$\bar{b} = \frac{b}{l}$ normalized bond (Kuhn) length
$\bar{a} = \frac{a}{l}$ normalized grafting distance of the polymer chains
$\tilde{h} = \frac{h}{Nb}$ brush height normalized by the contour length
$\psi = \frac{NN_p}{V} = \frac{1}{\tilde{h}\bar{a}^2\bar{b}}$ volume fraction of monomers
$\phi = \frac{\bar{v}M}{V}$ volume fraction of nanoparticles in the brush
c volume fraction of nanoparticles in solution
$n_{ads} = \frac{M}{NN_p} = \frac{\phi}{\bar{v}\psi}$ the number of nanoparticles in the brush normalized by the number of monomers
χ monomer-nanoparticle interaction strength
χ_{cr} monomer-monomer cohesiveness

2.3 Coarse Grained Polymer Chain Model

In this section, I will describe a generic model capable of multiple levels of detail, and applicable to various polymer and nanoparticle systems, including IDPs. The model can be used to verify the simple mean field model of Section 2.2 as well as extend them with more detailed interactions and sequence heterogeneity, without the complexity of all-atom molecular dynamics. The model makes two main simplifications. First, the model consists of spherically symmetric beads, which can represent groups of atoms. For example, in the case of IDPs, a single bead or monomer represents one or more amino acids. Secondly, to avoid explicitly modeling the solvent molecules, their interactions with the beads are included via a viscous (Stokes') drag force and a random force using the standard overdamped Langevin dynamics algorithm [19, 25, 73, 91]. In this Section, I will only identify the relevant parameters controlling each type of interaction. In Chapters 4 and 5, I will present the details of each specific implementation and how the model is used to represent IDPs. The computations in Chapters 4 and 5 were performed on the GPC and Niagara supercomputers at the SciNet HPC Consortium [92] and the supercomputer Mammouth Parallel

2 (MP2) from Université de Sherbrooke, managed by Calcul Québec and Compute Canada. SciNet is funded by: the Canada Foundation for Innovation; the Government of Ontario; Ontario Research Fund - Research Excellence; and the University of Toronto. The operation of MP2 is funded by the Canada Foundation for Innovation (CFI), the ministère de l'Économie, de la science et de l'innovation du Québec (MESI) and the Fonds de recherche du Québec - Nature et technologies (FRQ-NT). The computations in Appendix B were made on the MP2, Béluga, Graham, and Cedar supercomputers, enabled in part by support provided by Calcul Québec (www.calculquebec.ca), WestGrid (www.westgrid.ca), and Compute Canada (www.computecanada.ca).

2.3.1 Brownian Dynamics With Implicit Hydrodynamic Interactions

The Langevin equation for N beads in a fluid is [25, 93, 94]:

$$\bar{I}\frac{d\vec{v}}{dt} = \vec{f}_{int} + \vec{f}_H + \vec{f}_B \tag{2.52}$$

Here, all vectors have $3N$ components and all matrices are $3N \times 3N$. Specifically, $\vec{v}$ is a vector containing the velocities of all the beads:

$$\vec{v} = \begin{bmatrix} v_{1x} & v_{1y} & v_{1z} & v_{2x} & ... & v_{Nz} \end{bmatrix}^\intercal \tag{2.53}$$

$\bar{I}$ is a diagonal matrix of the masses of the beads. $\vec{f}_{int}$ are the forces on each bead due to it's interactions with the other beads or an external potential. $\vec{f}_H = -\bar{\xi}\vec{v}$ are the hydrodynamic forces exerted by the fluid on the beads. The motions of the beads cause long range disturbances in the fluid and therefore the hydrodynamic force on bead i depends on the velocities of all beads. This interaction is captured by the friction matrix $\bar{\xi}$, which depends on the bead positions. The inverse of $\bar{xi}$ is the hydrodynamic mobility matrix described in Section 2.3.7. Ignoring hydrodynamic interactions would result in a diagonal $\bar{\xi}$, whose entries would be the Stokes' drag constants of the corresponding beads. $\vec{f}_B$ are the random forces due to the collisions of the beads with the surrounding solvent. The random forces are not correlated in time and when averaged over realizations have mean $\langle\vec{f}_B\rangle = \vec{0}$ and covariance matrix $\langle\vec{f}_B(t)\vec{f}_B^T(t')\rangle = 2k_BT\bar{\xi}\delta(t'-t)$, as a consequence of the fluctuation-dissipation theorem [37, 73, 94].

Assuming the velocity decays much faster than the timescale over which the interaction forces $\vec{f}_{int}$ change, which is roughly the time taken by a bead to diffuse it's own size, we can neglect the inertial term in Equation 2.52. For spherical beads with the mass and volume of an average amino acid, the timescale of velocity decay is $\frac{m}{\xi} = \frac{m}{6\pi\eta a} \approx 4.6 \times 10^{-14}$s, while the timescale for a bead to diffuse its own size is $\frac{a^2}{D} = \frac{6\pi\eta a^3}{k_BT} \approx 1.1 \times 10^{-10}$s, so the approximation is valid [95]. In these expressions a is the bead radius and D its diffusion coefficient. In this regime, known as "overdamped", the Langevin equation for the bead positions ($\vec{x}$) is [37, 73, 91]:

$$\frac{d\vec{x}}{dt} = \bar{\xi}^{-1}\vec{f}_{int} + \bar{\xi}^{-1}\vec{f}_B \tag{2.54}$$

A rigorous interpretation in the Itô sense is the stochastic differential equation [96, 97]:

$$\Delta\vec{x} = \bar{\xi}^{-1}\vec{f}_{int}\Delta t + \bar{h}\Delta\vec{w} \tag{2.55}$$

Where the random forces have been decoupled into $3N$ Weiner processes ($\vec{w}$), which are independent with mean 0 and variance t, and $\bar{h}$ is a matrix preserving the correlations between the random forces, satisfying $\bar{h}\bar{h}^\intercal = 2k_BT\bar{\xi}^{-1}$. In a simulation, the positions of the displacements of the particles $\Delta\vec{x}$ are updated each time step Δt according to the algorithm of Ermak and McCammon [98]. The components Δw_i of $\Delta\vec{w}$ are independent random variables with Gaussian distributions such that $\langle \Delta w_i \rangle = 0$ and $\langle \Delta w_i(t) \Delta w_j(t') \rangle = \Delta t \delta(t' - t)\delta_{ij}$.

In the following sections I will describe the deterministic interaction potentials that give $\vec{F}_{int}$ and how to calculate the mobility matrix $\bar{\mu} = \bar{\xi}^{-1}$ (and $\bar{h}$).

2.3.2 Dimensionless Units and Notation

Defining the dimensionless notation: $\vec{X} = \frac{\vec{x}}{x_c}$, $\Delta T = \frac{\Delta t}{t_c}$, $\vec{F}_{int} = \frac{\vec{f}_{int}}{f_c}$, $\bar{M} = \xi_0 \bar{\xi}^{-1}$, and $\bar{H} = \frac{\bar{h}}{\sqrt{\frac{2k_BT}{\xi_0}}}$, so that $\bar{H}\bar{H}^\intercal = \bar{M}$. x_c, t_c, and f_c have the dimensions of length, time, and force respectively. $\xi_0 = 6\pi\eta a_0$ is the Stokes drag coefficient for a bead with hydrodynamic radius a_0. Equation 2.55 becomes:

$$\Delta\vec{X} = \frac{t_c f_c}{\xi_0 x_c} \bar{M}\vec{F}_{int}\Delta T + \sqrt{\frac{2k_BTt_c}{\xi_0 x_c^2}} \bar{H}\Delta\vec{W} \tag{2.56}$$

The components ΔW_i of $\Delta\vec{W}$ are independent random variables with Gaussian distributions such that $\langle \Delta W_i \rangle = 0$ and $\langle \Delta W_i(T) \Delta W_j(T') \rangle = \Delta T \delta(T' - T)\delta_{ij}$.

We choose $f_c = \frac{2k_BT}{x_c}$, $x_c = \sqrt{\frac{2k_BT}{k}} = \sqrt{\frac{2}{3}} b_0$ (k and b_0 are defined in the following Sections 2.3.3 and 2.3.4), and $t_c = \frac{\xi_0}{k}$, so that Equation 2.56 simplifies to:

$$\Delta\vec{X} = \bar{M}\vec{F}_{int}\Delta T + \bar{H}\Delta\vec{W} \tag{2.57}$$

In the following sections, we describe the interactions that contribute to $\vec{F}_{int}$. For each type of interaction, we give the potential $U(r_{ij})$ and $\vec{F}(\vec{r}_{ij})$, the force on bead i due to an interaction with bead j. In these cases the force and the vector $\vec{r}_{ij}$ contain 3 components, e.g.:

$$\vec{r}_{ij} = \begin{pmatrix} x_{3i} \\ x_{3i+1} \\ x_{3i+2} \end{pmatrix} - \begin{pmatrix} x_{3j} \\ x_{3j+1} \\ x_{3j+2} \end{pmatrix}, \quad r_{ij} = ||\vec{r}_{ij}||$$

In simulations, we sum the forces on i over interaction partners j and all interaction types. The $3N$ component vector $\vec{F}_{int}$ is obtained by concatenating the 3 component vectors of the total forces on each of the N beads.

In the following sections, we omit the subscripts i and j and give the force on bead i due to it's interaction with j. Furthermore, lower case letters represent un-normalized lengths, and the corresponding capital letters are those lengths normalized by x_c.

2.3.3 Bonded

The bonded potentials between adjacent monomers along a polymer chain preserve the polymer connectivity. Bonded beads interact via the finitely extensible non-linear elastic potential (FENE) [99]:

$$U_{FENE} = -\frac{1}{2}kl_{max}^2 \ln(1 - \left(\frac{r}{l_{max}}\right)^2) \tag{2.58}$$

This potential is harmonic close to $r = 0$, but diverges asymptotically as the bond length approaches l_{max}. Because the FENE potential is purely attractive, a minimum bond length is maintained by the excluded volume potential described in the next section.

The dimensionless force is:

$$\vec{F}_{FENE} = -\frac{\vec{R}}{1 - \left(\frac{R}{L_{max}}\right)^2} \tag{2.59}$$

where $\vec{R} = \frac{\vec{r}}{x_c}$ and $L_{max} = \frac{l_{max}}{x_c}$.

2.3.4 Excluded Volume

The repulsive term of the 12-6 Lennard-Jones potential is often used to approximate the steric repulsion between two atoms [75, 76]. If each bead were to precisely represent multiple atoms in a particular geometry (e.g. an amino acid), the repulsive force between beads would depend on their relative rotational orientation in addition to their separation. Instead, in our model we approximate the beads as spherically symmetric and use a phenomenological potential, a less steep 8-6 Lennard-Jones (LJ) potential which accommodates possible closer distances between the beads for certain orientations of the atoms [100], to describe the repulsive interactions between them.

All bead pairs interact via the 8-6 LJ potential:

$$U_{EV} = \begin{cases} \epsilon_{LJ}\left[\left(\frac{b}{r}\right)^8 - \frac{4}{3}\left(\frac{b}{r}\right)^6\right] + \frac{1}{3}(\epsilon_{LJ} - \epsilon) & \text{if } r < b \\ 0 & \text{if } r > b \end{cases} \tag{2.60}$$

$$b = \frac{b_i + b_j}{2}$$

ϵ_{LJ} is the strength of the interaction, and b is the bead separation at which the force goes to 0 (the sum of the radii of the two interacting beads). An exception to this rule occurs if the two interacting beads are bonded monomers of a polymer, then $b = b_0$, which reflects the bond length rather than the bead radius. In this case, the repulsive potential is considered to be part of the bond interaction: the only interactions between two bonded monomers are the FENE (attractive) and the modified 8-6 LJ (repulsive, with $b = b_0$) potentials. The potential is cut off at b. The potential is shifted by $\frac{1}{3}(\epsilon_{LJ} - \epsilon)$ in order to maintain continuity at $r = b$ with the attractive potential described in the following section. This does not affect the force or the simulations.

The dimensionless force is:

$$\vec{F}_{EV} = \begin{cases} \frac{4E_{LJ}}{R^2}\left[\left(\frac{B}{R}\right)^8 - \left(\frac{B}{R}\right)^6\right]\vec{R} & \text{if } R < B \\ 0 & \text{if } R > B \end{cases} \qquad B = \frac{B_i + B_j}{2} \tag{2.61}$$

where $B = \frac{b}{x_c}$, $\vec{R} = \frac{\vec{r}}{x_c}$, $E_{LJ} = \frac{\epsilon_{LJ}}{kT}$.

2.3.5 Cohesive Interactions

To capture all of the physical effects that can lead to the aggregation of molecules in solution or the compaction of a polymer, we introduce a short-range attractive force between "cohesive" beads. These attractive interactions can represent van der Waals forces, the hydrophobic effect [101], and any other interactions between order promoting amino acids in an IDP, or even the poor quality of a particular solvent.

Two non-bonded cohesive beads interact through the attractive portion of the 8-6 LJ potential:

$$U_C = \begin{cases} \epsilon\left[\left(\frac{b}{r}\right)^8 - \frac{4}{3}\left(\frac{b}{r}\right)^6\right] & \text{if } b < r < 4b \\ 0 & \text{if } r \le b \text{ or } r \ge 4b \end{cases} \qquad b = \frac{b_i + b_j}{2}, \epsilon = \sqrt{\epsilon_i \epsilon_j} \tag{2.62}$$

The parameter ϵ controls the strength of the attractive forces. The particles may also have different cohesive strengths ϵ_i and ϵ_j, in this case the square root of the product of the two strengths is taken. The overall strength of the cohesive interaction may also be adjusted to reflect solvent quality. The sum of the radii of the two beads (b) is the same as in the repulsive force described previously. The attractive potential begins where the repulsive ends: at separations larger than b. To reduce computational complexity, the potential is also cut off beyond $4b$, where it is $\sim 0.1\%$ of it's minimum depth.

The dimensionsless force is:

$$\vec{F}_C = \begin{cases} \frac{4E}{R^2}\left[\left(\frac{B}{R}\right)^8 - \left(\frac{B}{R}\right)^6\right]\vec{R} & \text{if } B < R < 4B \\ 0 & \text{if } R \le B \text{ or } R \ge 4B \end{cases} \qquad B = \frac{B_i + B_j}{2}, E = \sqrt{E_i E_j} \tag{2.63}$$

$E = \frac{\epsilon}{kT}$ is the dimensionless interaction strength.

Connection Between Mean Field Theory and Simulations

Notice that in the lattice construction of Section 2.1.3, for monomer-monomer interactions, we treated the monomers as uniformly distributed over the lattice sites. This is equivalent to a weakly interacting gas, so we can compare the 2nd order interaction term of the lattice model to the 2nd virial coefficient

$B(\beta)$ of the combined excluded volume and cohesive interactions of the coarse grained model [102]:

$$\frac{N^2}{R^3}B(\beta) = \frac{R^3}{v_0}\left(\frac{1}{2}\psi^2(1+\chi_{cr})\right) \tag{2.64}$$

$$B(\beta) = \frac{v_0}{2}(1+\chi_{cr}) \tag{2.65}$$

The excluded volume and cohesive interactions are described by the 8-6 Lennard Jones potential (Equations 2.60 and 2.62 respectively), with distance parameters b, repuslive strength ϵ_{LJ} and cohesive strength ϵ. We set $\beta\epsilon_{LJ} = 1$. For $\beta\epsilon = 0$, there is only repulsion and so $\chi_{cr} = 0$. Using numerical integration, we can find:

$$B(\beta) = 2\pi\int_0^\infty r^2(1-e^{-\beta U(r)})dr \tag{2.66}$$

$$B(\beta) \approx 1.231b^3 \approx \frac{v_0}{2} \tag{2.67}$$

$$v_0 \approx 2.462b^3 \tag{2.68}$$

$$\tag{2.69}$$

We also find that $B(\beta) = 0$ at $\beta\epsilon \approx 0.64$, and this should correspond to $\chi_{cr} = -1$. $B(\beta)$ is roughly linear with ϵ, so we can write:

$$B(\beta) \approx 1.231b^3\left(1-\frac{\beta\epsilon}{0.64}\right) \tag{2.70}$$

$$\chi_{cr} \approx -\frac{25}{16}\beta\epsilon \tag{2.71}$$

2.3.6 Ionic Interactions

Ionic interactions between net charged monomers may be may be included in the model explicitly. In this case, two non-bonded, charged monomers interact via the screened coulomb potential:

$$U_Q = \frac{q_1 q_2}{4\pi\epsilon r}e^{-\frac{r}{l_D}} \tag{2.72}$$

where q_1 and q_2 are the charges of the beads, and ε is the absolute permittivity of the solution. The Debye length l_D describes the implicit presence of ions in solution in addition to the charged beads. The dimensionless force is:

$$\vec{F}_Q = \pm Q\left(\frac{1}{R^2}+\frac{1}{RL_D}\right)e^{-\frac{R}{L_D}}\hat{R} \tag{2.73}$$

$$Q = \frac{|q_1 q_2|}{8\pi\epsilon k_B T x_c} \tag{2.74}$$

The negative is taken for opposite charges, and the positive for like charges.

Using elementary charges, reasonable values for temperature ($T \approx 310$K), and the relative permittivity of water ($\epsilon_r \approx 80$), this expression simplifies to $Q \approx \frac{3.367\text{Å}}{x_c}$.

The parameters of the ionic interaction in simulations are the normalized Debye length $L_D = \frac{l_D}{x_c}$ and the "electrostatic interaction strength" Q. The temperature, ionic strength, and choice of x_c will roughly set the values of these parameters. The Debye length at $T = 310$K is $L_D \approx \frac{3.09}{x_c\sqrt{I}}$Å where I is

the Molar ionic strength (equivalent to Molar concentration for monovalent ions) [75]. At 100mM NaCl, $L_D \approx \frac{9.8}{x_c}$Å. However, there is further indeterminacy in Q due to two details that are "coarse-grained" out in the model. First, the relative permittivity of water decreases near surfaces such as the interfaces of macromolecules [103, 104] and consequently may depend on the local conformation of the beads. Second, a bead generally represents multiple atoms or amino acids, and the actual location of the electric charge may not coincide with the bead center used to compute the forces in simulations [44, 105, 106].

2.3.7 Hydrodynamic Mobility Matrix

The remaining component of the simulation equation (Eq. 2.56) is the dimensionless mobility matrix $\bar{M} = \xi_0 \bar{\xi}^{-1}$, where $\xi_0 = 6\pi\eta a_0$ is the Stokes drag coefficient for a bead with hydrodynamic radius a_0. If we wish to obtain only ensemble averages from the simulations, it is sufficient to set $\bar{M}$ to the diagonal matrix with entries $\frac{a_0}{a_i}$ (the inverse relative hydrodynamic radius of each bead) and $\bar{H}$ to the square root of this matrix. This will model the viscous drag and the random forces on the beads due to the solvent. However, the moving beads will also disturb the solvent, leading to correlations between the motion of all the beads. These long-range correlations are captured by the off-diagonal terms of the matrices $\bar{M}$ and $\bar{H}$. The Oseen tensor [32] is the simplest representation of hydrodynamic interactions between and is valid at long distances. The Rotne-Prager-Yamakawa tensor [107] gives corrections for distances close to the radii of the beads and accounts for possible overlap of the beads due to the "soft" 8-6 Lennard-Jones potential used for repulsion. The matrix $\bar{H}$ must satisfy $\bar{H}\bar{H}^\intercal = \bar{M}$, hence we use the Cholesky decomposition to calculate a triangular matrix for $\bar{H}$ [98, 108].

$\bar{M}$ is a symmetric $3N \times 3N$ block matrix consisting of $N \times N$, 3×3 blocks. Each block represents the hydrodynamic interaction between bead i and bead j. For different sized beads with hydrodynamic radii a_i and a_j and separated by vector $\vec{r}_{ij}$, the dimensionless Rotne-Prager-Yamakawa tensor [109] defines the block $\bar{M}_{ij}$ of $\bar{M}$ (beginning at indices $(3i, 3j)$ of $\bar{M}$) as:

$$\bar{M}_{ij} = \begin{cases} \dfrac{A_0}{A_i}\bar{I} & \text{if } i = j \quad (2.75) \\ \dfrac{3A_0}{4R_{ij}}\left[\left(1 + \dfrac{A_i^2 + A_j^2}{3R_{ij}^2}\right)\bar{I} + \left(1 - \dfrac{A_i^2 + A_j^2}{R_{ij}^2}\right)\dfrac{\vec{R}_{ij}\vec{R}_{ij}}{R_{ij}^2}\right] & \text{if } i \neq j \text{ and } A_i + A_j < R_{ij} \quad (2.76) \\ \dfrac{A_0}{A_i A_j}\left[\dfrac{16R_{ij}^3(A_i + A_j) - ((A_i - A_j)^2 + 3R_{ij}^2)^2}{32R_{ij}^3}\bar{I}\right. & (2.77) \\ \left. + \dfrac{3((A_i - A_j)^2 - R_{ij}^2)^2}{32R_{ij}^3}\dfrac{\vec{R}_{ij}\vec{R}_{ij}}{R_{ij}^2}\right] & \text{if } i \neq j \text{ and } A_> - A_< < R_{ij} \leq A_i + A_j \\ \dfrac{A_0}{A_>}\bar{I} & \text{if } i \neq j \text{ and } R_{ij} \leq A_> - A_< \quad (2.78) \end{cases}$$

$$A_> = \max(A_i, A_j) \text{ and } A_< = \min(A_i, A_j)$$

Here, all the distances are normalized by x_c: $A_0 = \frac{a_0}{x_c}$, $A_i = \frac{a_i}{x_c}$, and $\vec{R}_{ij} = \frac{\vec{r}_{ij}}{x_c}$. The presence of A_0 is due to fact that $\bar{M}$ was normalized by the Stokes' drag coefficient of a bead of hydrodynamic radius a_0. Case 2.75 is the Stokes' drag on bead i, case 2.76 describes the hydrodynamic interaction between non-overlapping beads, case 2.77 describes partially overlapping beads, and case 2.78 describes when one bead completely encloses the other.

Unlike the short range Lennard-Jones potentials, long range interactions, including the $3N \times 3N$ RPY tensor, cannot be cut-off and require $O(N^2)$ operations to calculate. Even more computationally costly than $\bar{M}$ is the calculation of its Cholesky decomposition $\bar{H}$ which is $O(N^3)$ [98, 108]. Approximate methods, such as Fixman's Chebyshev polynomial expansion ($O(N^{2.25})$) and the Truncated Expansion Ansatz ($O(N^2)$) offer a speedup at the cost of reduced accuracy [110]. We utilize the exact method (Cholesky decomposition) in our simulations.If only equilibrium averages are required, hydrodynamic interactions may be ignored by keeping only the diagonal elements of $\bar{M}$ (Case 2.75), which reduces complexity from $O(N^3)$ to $O(N)$ [111].

2.3.8 Summary of Parameters

We set $A_0 = 1$ ($a_0 = x_c$). This choice is arbitrary and only redefines the time units. If we were instead to choose $a_0 = 2x_c$, at each step the deterministic forces would be multiplied by a factor of 2 and the random forces by $\sqrt{2}$, the time unit $t_c = \frac{\xi_0}{k}$ would also be double; we have sped up our simulation at the cost of accuracy. To remedy this we would need to halve the timestep ΔT. For example, a simulation with $A_0 = 1$ and $\Delta T = 0.002$ is identical to a simulation with $A_0 = 2$ and $\Delta T = 0.001$.

We set B_0, the dimensionless LJ repulsion distance between bonded monomers, to $\sqrt{\frac{3}{2}}$ and L_{max} of the bonds to $2B_0$. The former is motivated by the fact that for the Rouse model (no excluded volume interactions and Hookean bonds) the dimensionless root mean square distance between bonded monomers would be $\sqrt{\frac{3}{2}}$. We set E_{LJ}, the LJ repulsion strength between bonded monomers, to 1. With these parameters, the average bond length measured in simulations is slightly larger than B_0 (about 1.35).

The values of the soft LJ diameters B_i determine the sizes of the beads, relative to the bond length. For example, if we equate one bead in the model to one amino acid of an IDP, the bond length in simulations (≈ 1.35) corresponds to the distance between Cα atoms of adjacent peptides (≈ 3.8Å), and the LJ diameters can be estimated from the volumes of particular amino acids using this conversion. The spherical beads represent the more complicated atomic arrangements of amino acids and there is no guarantee that their hydrodynamic radii and LJ radii are the same, but as an estimate we set $A_i = \frac{B_i}{2}$; similar choices are used in other coarse grained models with implicit hydrodynamic interactions [30, 31, 112]. Finally, the parameters E_i, Q, and L_D control the cohesive and ionic interactions. These depend on the properties of the beads and the solvent.

The parameters for each of the interactions in the coarse-grained model are summarized in Table 2.1.

Interaction	Parameters	Description
Bond	L_{max}	Maximal bond extension (set to $2B_0$)
Excluded Volume	B_0	Separation between bonded beads (set to $\sqrt{\frac{3}{2}}$)
	B_i	LJ Diameter of bead i
	E_{LJ}	Strength of repulsion (set to 1)
Cohesive	E_i	Cohesive strength of bead i
Ionic	L_D	Debye length
	Q	Electrostatic strength
Hydrodynamic	A_i	Hydrodynamic radius of bead i (roughly $\frac{B_i}{2}$)

Table 2.1: Summary of coarse-grained bead model parameters.

Chapter 3

Surface Grafted FG Nucleoporin Layers

The results this chapter have been partially reported in Refs. [78] and [79].

3.1 Introduction

The Nuclear Pore Complex (NPC) is a bidirectional selective gate for cargo traveling across the nuclear envelope of eukaryotic cells. Understanding the mechanism of transport remains elusive due to the size and complexity of the NPC. A visualization of the key structural elements of the NPC is shown in Figure 3.1. The NPC consists of ∼500 proteins referred to as nucleoporins (nups) and is able to transport ∼1000 cargo molecules per second. There are only about 30 distinct types of nups and each is present in copies of 8-64 due to the eightfold rotational symmetry of the NPC's cylindrical structure. Sub-complexes referred to as the inner, outer, and membrane rings form the core structure of the NPC which is embedded into the nuclear envelope and surrounds the transport channel. The channel is about 35-50 nm in diameter and 50-80 nm in length, and is filled with ~ 200 intrinsically disordered proteins, referred to as FG nups due to the recurring Phenylalanine Glycine (FG) motif in their sequence. FG nups are anchored via their structured regions to the interior of the core structure and form the selectivity barrier by allowing passive diffusion of small particles but restricting larger particles that do not possess a translocation signal. Numerous theoretical, experimental, and computational approaches continue to attack this problem as often it is explicative to study a subsystem of the full *in vivo* NPC. In this chapter, I will present a model for surface grafted FG nups addressing a subset of *in vitro* experiments to elucidate the structure and function of the NPC. The model reconciles the apparent contradictions arising from several experiments and suggests how the NPC can maintain it's function despite significant structural perturbations [12, 113–118].

The FG domains lack secondary structure and show other features common to intrinsically disordered regions. The sequences of the FG nucleoporins share four letter hydrophobic short linear motifs, which are most commonly GLFG or FxFG [119, 120]. These motifs are separated by spacer sequences with an average distance of about 15 residues [121]. The spacer regions have low hydrophobicity and can be classified as having either a high or low fraction of charged residues. An FG domain's fraction of charged residues has been linked to its polymer dimensions [46]. The spacer sequences are not conserved

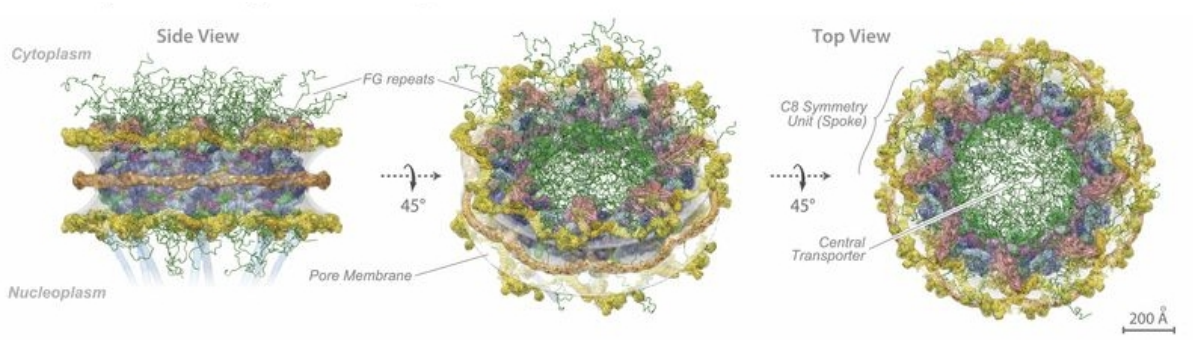


Figure 3.1: Core structure of the NPC (shown in yellow, orange, purple, and blue) and intrinsically disordered FG nups (shown as a single representative conformation in green). Figure from Kim et al. [118].

across all species while retaining similar functions, hinting that the underlying NPC mechanism is very general since it is partially independent of the primary amino acid sequences [114, 115, 122]. This generality is reinforced by the findings showing that transport proteins are able to function with the NPCs of a different organism [123] and yeast maintains viability even after the deletion of many of the FG nups [123–126].

Transport proteins in the nucleus and cytoplasm bind and shuttle large cargo that cannot passively diffuse through the NPC. Crystal structures and molecular dynamics simulations show that transport proteins have multiple, relatively weak, binding sites for the four amino acid hydrophobic patches of FG nups. The hydrophillic peptides surrounding the FG patches remain unstructured in the bound states. This binding is crucial for the *in vivo* transport mechanism, since mutations to the binding sites on the transport proteins reduces or eliminates transport efficiency and inhibits cell growth [122, 127–130].

Conversely, the specific interaction between FG patches and the binding sites of transport proteins is not completely essential for transport of all cargo. Several chemical modifications to the surfaces of molecules allow them to be shuttled through NPCs without transport proteins, indicating that interactions between cargo surfaces and FG nups such as hydrophobic, electrostatic, or cation-π are all sufficient for transport and the mechanism is non-specific [131–134].

Multiple models that hypothesize the mechanistic roles of FG nups and transport proteins in nucleocytoplasmic transport have been proposed [135]. In the "virtual gate" model, the FG nups contribute to an entropic barrier which hinders larger molecules. The barrier can be lowered by transport protein and FG nup binding [136–138]. In the related "brush" model, the FG nups undergo a recoverable collapse through interactions with the transport proteins allowing cargo to pass through [138–140]. In the alternative "selective phase" (or "gel") model, the barrier arises through the formation of a gel-like network of FG nups, linked by their hydrophobic FG motifs. The transport proteins bind to the FG domains and disentangle the network allowing cargo to pass through [125, 141, 142]. It is likely that the various effects described in these and other models all play a role in the NPC transport mechanism to some degree.

The size and complexity of the NPC prevents the experimental study of the transport process *in vivo* on the relevant time and length scales (several milliseconds and tens of nanometers) [143–146]. There are many types of intrinsically disordered FG nups, which differ across species and location in

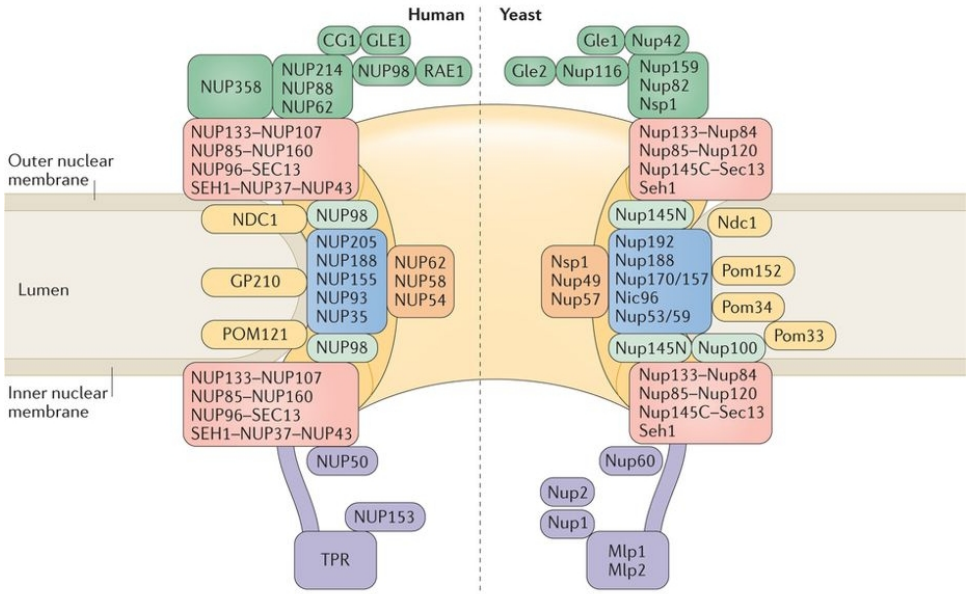


Figure 3.2: Locations of the Nucleoporins within the Human and Yeast NPC, organized into boxes corresponding to their subcomplexes. FG nups are present in the green, light blue, orange, and purple boxes. Figure from Beck and Hurt [117].

the NPC, which contribute to the selectivity barrier. The complexity of the FG nup system is further increased by their interactions with transport proteins and cargo during the transport process [137, 147–151]. *In vitro* studies provide a basis for understanding the structure of the FG nup assemblies and their interactions with the transport proteins. However, attempting to synthesize the insights from these studies to understand NPC transport is hindered by the fact that many *in vitro* studies appear to draw contradictory conclusions. Although it was shown that the mechanism of transport is highly resilient and non-specific, the contradictions suggest that the structure and behaviour of FG nups are highly dependent on the experimental conditions [122].

First, the relative strengths and importance of the possible interactions governing FG nup and transport protein binding remain unclear [122]. Molecular dynamics reveals up to ten possible binding spots for FG patches on the transport protein Importin-β, some of which were not present in crystal structures [130]. As stated earlier, these specific transport proteins and binding sites are not essential for passage through the NPC and transport can be induced with chemical modifications like the addition of hydrophobic moieties to cargo molecules [131, 133]. On FG nups, the non-FG spacer regions also have a role in binding transport proteins [152]. In addition to hydrophobicity, electrostatic interactions are important in the structure of FG nup assemblies and transport protein binding, as FG nups generally carry a net positive charge in their spacer regions, while transport proteins are negatively charged [132, 153]. Experimental results disagree about the affinities of FG nup and transport protein binding [122, 135, 152, 154]. Furthermore, *in vitro* measurements of binding affinities between FG nups and transport proteins suggest that the NPC would be saturated with transport proteins and lead to transport times that are too slow, conflicting with *in vivo* observations of transport on the timescale of milliseconds [122, 155, 156].

Different experimental evidence exists in support of the various proposed models of FG nup structure. Atomic Force Microscopy measurements indicate that surface grafted FG nups behave like an entropic polymer brush and the addition of transport proteins compacts the layer height, which recovers once the transport proteins are removed, supporting the "brush" model [138, 139]. Free FG domains in solution spontaneously form dense aggregates, which repulse inert cargo but permit transport proteins and bound cargo to enter. The cohesiveness is crucial to the formation of the dense phase and its selectivity properties, supporting the "gel" model [141, 142, 157].

The model used in this chapter addresses the surface grafted FG nup system and elucidates some of the conflicts emerging from *in vitro* experiments. In one typical experimental setup, rather than following the cylindrical geometry of the NPC, FG nups are grafted to a flat surface at one end, while the other end extends out into the solution, which allows measurement of the layer height using techniques such as Atomic Force Microscopy [139], Quartz Crystal Microbalance with Dissipation monitoring [158], or Surface Plasmon Resonance [150]. Transport proteins can penetrate this layer and alter its height. However, even such simplified systems led to initially conflicting observations: depending on the experimental conditions, the addition of transport proteins could cause both a compaction and expansion of the layer height and in some cases caused no change. A general behaviour did emerge once concentration was systematically increased. At low concentration of transport proteins, there is little change in the FG nup layer height. As the concentration increases, and transport proteins accumulate in the layer, the layer height decreases to varying degrees. Any further increase of the concentration of the transport protein solution leads to an increase in height which eventually surpasses the initial height and swells the layer. All combinations of FG nups and transport proteins studied exhibit this pattern of behaviour to some degree irrespective of the species of origin or location in the NPC [150, 158–161].

These observations support the idea that the NPC is a robust system governed by a few general principles which are not constrained by exact details such as amino acid sequences. A model which captures these principles allows for the systematic study of NPC transport. This idea has been put into practice by the design of synthetic channels inspired by the NPC which successfully mimic its transport properties [16, 162–164]. Furthermore, including more details into a model may be counterproductive as their effects can be obscured by experimental uncertainty and parametrization can lead to overfitting. Therefore, using a coarse-grained model is often sufficient and can be more insightful. The model for surface grafted FG nups and transport proteins utilized in this chapter aims to capture:

1. The polymer-like behaviour arising from the intrinsic disorder of the FG nups
2. The intra and inter-chain cohesiveness arising from the interactions between the FG nups
3. The attractive interactions between the FG nups and transport proteins

The model is based on the well known polymer brush theory, supported by coarse-grained Brownian dynamics simulations [86, 87], and is derived Section 2.2. Related coarse-grained models have been successfully applied to describe other biopolymer systems [165–168]. In Section 3.2, I explain how the general polymer brush model is used to describe the FG nup and transport protein system specifically. In Section 3.3, I will show that the model is able to semi-quantitatively explain the behaviour observed in *in vitro* experiments and analyze the implications of the model for the greater nucleocytoplasmic transport system.

The model reproduces the wide range of behaviour of the FG nup layers observed in experiments. The model also suggests a resolution of the apparent controversies arising from the measured binding

affinities, and can reconcile the debate among the proposed models over the importance of entropic and cohesive interactions in the formation of the FG nup selectivity barrier of the NPC. Using the model, we identify the regimes of physical variables that determine the behaviour of the FG nup assemply, which suggest ways in which this system can be controlled. In the future, the simple theoretical framework used in the model can be expanded with the addition of molecular and structural details to create a more accurate description of the system. Aside from providing insight into NPC transport, the model contributes to understanding the broad subjects of intrinsically disordered proteins and the design of artificial pores with nanotechnological applications such as sensing and sorting [169–171].

3.2 Mean Field Polymer Brush Model for Grafted FG Nucleoporin Layers

In this section, I will describe how to parametrize the polymer brush model constructed in Section 2.2 to represent a surface grafted FG nup layer in a solution of transport proteins. A schematic of the model is shown in Figure 2.2. Equation 2.50 gives the normalized grand potential of the system:

$$\begin{aligned}\tilde{\Phi}(\tilde{h},\phi) = &\frac{\tilde{h}^2}{2\bar{a}^2\bar{b}} \\ &+ \tilde{h}\Big(\frac{1}{\bar{v}}\phi\ln\phi + (\frac{1}{\bar{v}}-1)(1-\phi)\ln(1-\phi) + (1-\phi-\psi)\ln(1-\phi-\psi) + \frac{1}{\bar{v}}\chi\phi\psi + \frac{1}{2}\chi_{cr}\psi^2\Big) \\ &- \frac{\tilde{h}\phi}{\bar{v}}\ln\frac{c}{1-c} - \frac{\tilde{h}}{\bar{v}}\ln(1-c) \quad (3.1)\end{aligned}$$

A single polymer, composed of N monomers each of size l (and volume l^3) and connected by bonds of length b, represents an FG nup. χ_{cr} captures the average cohesiveness of a particular FG nup in a single parameter which represents monomer-monomer attractive strength. The polymers are tethered by one end to a planar surface at intervals of the grafting distance a. The nanoparticles in the model represent transport proteins. $\bar{v}$ describes their volume relative to the FG nup monomers and χ is the strength of their attractive interaction with the monomers. The variables $\tilde{h}$, ψ, ϕ describe the normalized height of the layer, the volume fraction of monomers, and the volume fraction of transport proteins in the layer, respectively. The volume fraction of monomers is inversely proportional to the layer height: $\psi = \frac{1}{\tilde{h}\bar{a}^2\bar{b}}$. The layer height $\tilde{h}$ has been normalized by the contour length Nb of a single FG nup. Aside from the layer height, all lengths are normalized by the monomer size l, indicated by a horizontal bar above the letter. Thus $\bar{a}$ is the normalized grafting distance (per monomer size), $\bar{b}$ is the normalized bond length (per monomer size), and $\bar{v}$ is the normalized nano-particle volume (per monomer volume) as already mentioned. c describes the volume fraction of the transport proteins in the solution outside of the layer. To obtain the equilibrium layer height and monomer and the number of transport proteins in the layer for a particular set of parameters, Equation 2.50 is minimized numerically over the normalized layer height $\tilde{h}$ (or equivalently the monomer volume fraction ψ) and the transport protein volume fraction ϕ.

The molecular mass of a particular transport protein combined with the average protein density ($\sim 1.2 - 1.5\text{g/cm}^3$ [172]) gives its volume. For Karyopherin-β1 (molecular mass $\approx 97 - 103$ kDa), the volume is $\sim 120 - 140$ nm^3. For NTF2 (molecular mass ≈33kDa), the volume is $\sim 35 - 45$ nm^3.

The bond length parameter b in our model is equivalent to the Kuhn length. Consecutive carbon-alpha atoms of amino acids in a polypeptide chain are separated by ~ 0.38 nm [100]. However, due

to the peptide backbone geometry, the outer two carbon-alpha atoms of three consecutive amino acids may be maximally separated by ~ 0.72 nm, and the effective contour length per amino acid is therefore ~ 0.36 nm [106, 173]. Polypeptides are not freely jointed chains and their Kuhn lengths can include several amino acids; different measurement techniques report values between 4-20 amino acids [44]. Lim et al. reported that the persistence length of the FG nup Nup153 was ~ 0.4 nm which translates to a Kuhn length of roughly 2 amino acids [139].

Amino acid volumes range on the order of $\sim 0.072 - 0.239$ nm^3, with sizes ranging between $\sim 0.52 - 0.77$ nm [100, 174]. The effective size of the amino acids may be further altered by their non-spherical geometries and the presence of bound ions within the Debye screening length.

Taking into account all the previous considerations, a single monomer in the model is equivalent to between one and four amino acids: $b \approx 0.4 - 1.6$ nm and $l \approx 0.5 - 1$ nm. The upper limit is motivated by the size of the hydrophobic GLFG or FxFG domains of FG nups and these particular motifs corresponding to specific binding sites on transport proteins [128, 129]. The sensitivity of the results to the specific choice of b and l are presented in Section 3.5, but these parameters do not affect the qualitative behaviour or the conclusions of this chapter.

It is straightforward to normalize experimentally measured layer heights by the contour length (number of amino acids multiplied by the amino acid separation $\sim 0.36 - 0.38$ nm [100]). Other experimental conditions include the grafting distance of the FG nups and the concentration of transport proteins in the external solution. The grafting distance is normalized by l and Section 2.2.3 describes the conversion between concentrations and volume fractions. Finally, the number of transport proteins in the layer per monomer of the FG nups n_{ads} is given by: $n_{ads} = \frac{\phi}{\bar{v}\psi} = \frac{\tilde{h}\phi\bar{a}^2\bar{b}}{\bar{v}}$ and can be easily related to experimental quantities such as the surface density of bound transport proteins.

The final parameters are the cohesiveness of the FG nup monomers χ_{cr} and attractive strength of monomer - transport protein interactions χ. Overall, the relative importance of hydrophobic, electrostatic, and other effects in the inter-FG and in FG nup - transport protein interactions are unknown. Although there are specific binding sites for FG patches on transport proteins, other chemical modifications to the surfaces of cargo are sufficient for transport [128, 129, 131, 133]. The parameters χ and χ_{cr} can depend on our exact definition of a monomer (since the number of amino acids included can vary) as well as the specific FG nup and transport proteins used and due to the simplifications in the model, it is difficult to relate them to biophysical quantities directly. They can be understood in terms of the attractive parts of the second virial coefficients of the interactions between them. The mean field parameter χ_{cr} is proportional to the strength of the cohesive interaction of the coarse grained model of Section 2.3 as explained in Section 2.3.5. We can infer χ_{cr} from measurements of pure brushes of FG nups without transport proteins as will be described in Section 3.3.1. χ is roughly proportional to the number of binding sites or surface area of the transport proteins and the interaction energy between the monomers of the FG nups and transport proteins. In Section 3.3.2, we find that the values of χ that best agree with experiments are physically reasonable and close to the values inferred from an analogous model for the phase separation of FG nup polymer solutions [78, 79].

3.3 Results and Discussion

3.3.1 Morphology of FG Nup Layers in the Absence of Transport Proteins

In order to validate the model, I first consider the case of pure FG nup layers in the absence of transport proteins. This reduces the number of variables in the model and enables parametrization of the size l and bond length b describing our FG nup monomers and cohesiveness χ_{cr} describing the interactions between them. A variant of the model describing the phase separation of FG nup polymer solutions has been shown to agree well with experimental data. The parameters l, b, and χ_{cr} in bulk solution are consistent with those found in this section for surface grafted FG nups [78]. In the absence of transport proteins, $\phi = 0$ and $c = 0$ in Equation 2.50 and after minimization over the layer height, the solutions of Equation 2.46 give the layer height at equilibrium. This is a simple description of a polymer brush: the polymers are tethered to a surface at regularly spaced intervals by one end while the other end extends into the solution. The model assumes that all polymers extend to the same height and the monomer density within the brush is uniform. The height of the brush depends on the balance of elastic and cohesive forces resisting swelling of the layer and steric repulsion preventing compaction.

The predictions of the model for the layer height as a function of the grafting distance are summarized in Fig. 3.3a. The height decreases with increasing grafting distance as the steric repulsion responsible for the stretching of the chains is reduced. The height also decreases with increasing monomer cohesiveness $|\chi_{cr}|$. This is expected since stronger monomer cohesiveness would favour more compact conformations which are ordinarily opposed by entropic elasticity and steric repulsion of the chains. Increasing the cohesiveness is equivalent to changing the solvent quality from good to poor in polymer physics terms [24, 88, 89, 161, 175–179]. The exponent g, which describes how layer height scales with grafting distance $h \sim a^{-g}$, quantifies this behaviour. For $\chi_{cr} = 0$, the layer behaves as a purely entropic brush with $g = \frac{2}{3}$. For high cohesiveness $|\chi_{cr}| > 2$, the exponent approaches $g = 2$, since in the fully collapsed case $ha^2 = Nl^3$. However, this behaviour of the exponent occurs long before the chain reaches the theoretical maximum compaction of $\psi = 1$ [24, 88, 89, 161, 175–179]. The two extremes of the exponent g are comparable to the coil-globule transition of a single polymer which occurs at $|\chi_{cr}| \approx 1$ and the scaling of the polymer size with the number of monomers (see Section 2.1.3).

Figure 3.3b shows the results of the model along with the experimental results of Lim and coauthors [150, 159, 160]. In experiments, domains of FG nups were covalently tethered to a gold surface, and the grafting distance, the layer height, and the surface density of transport proteins in the layer were obtained using Surface Plasmon Resonance. The original locations of the FG nups in the NPC are shown in Figure 3.2. In Figure 3.3b, the layers heights of five FG nup segments are plotted against their grafting distance along with the model predictions for brushes of varying cohesiveness, but constant monomer bond length and size (b=1.52 nm, l=1 nm). Short and long segments of Nsp1 behave similarly, reinforcing the polymer brush model of FG nups. The heights of the FG nup layers decay with increasing grafting distance faster than $h \sim a^{-\frac{2}{3}}$ but slower than $h \sim a^{-2}$. This comparison indicates that, all FG nups have significant cohesiveness but due to experimental uncertainty, it is impossible to determine the exponent g and the strength of their cohesive interactions exactly. Based on the relative positions of the experimental data in the figure Nup62 and Nup98 likely have the highest cohesion, followed by Nup153, and Nsp1 has the weakest with χ_{cr} falling between -0.8 and -1.4.

It is impossible to learn the values of the monomer size and bond length parameters due to uncertainty in the experimental data, but the choice of $l = 1$ nm and $b = 1.52$ nm, corresponding to the size of one FG

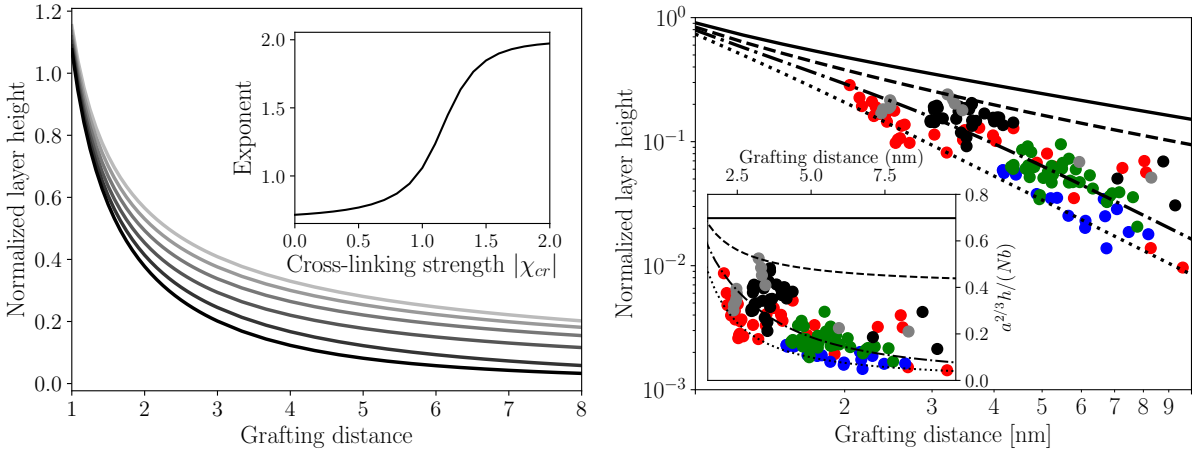


Figure 3.3: (a) Cohesion makes FG nup layers more compact: theoretical predictions. Layer height h/L normalized by the chain length as a function of the normalized grafting distance a/l for increasing cohesiveness. χ_{cr} varies from 0 (grey) to -1.5 (black). For any value of χ_{cr}, the curve is well approximated by the dependence $h \sim a^{-g}$. The inset shows that the exponent g increases from $\frac{2}{3}$ to 2 as the absolute value of the cohesion strength $|\chi_{cr}|$.
(b) FG nup layer height depends on the grafting distance: theory vs. experiment. The dots are the experimentally measured layer heights from [159] and [160] normalized by the FG nup length. The colours represent: Nup62, Nup98, Nup153, long **Nsp1**, and short Nsp1 segments. Solid line: $h \sim a^{-\frac{2}{3}}$ is the ideal brush ($\chi_{cr} = 0$) behavior obtained from the model. Dotted line: $h \sim a^{-2}$ is the behavior of a strongly collapsed brush with $\chi_{cr} = -2.5$. All the FG Nups lie between these two regimes, indicating a significant amount of cohesion. The dashed line is for $\chi_{cr} = -0.8$. The dashed-dotted line is for $\chi_{cr} = -1.4$. To enhance the contrast, inset shows the same data with the height h normalized by the ideal brush height ($h \sim a^{-\frac{2}{3}}$). b=1.52 nm, l=1 nm.

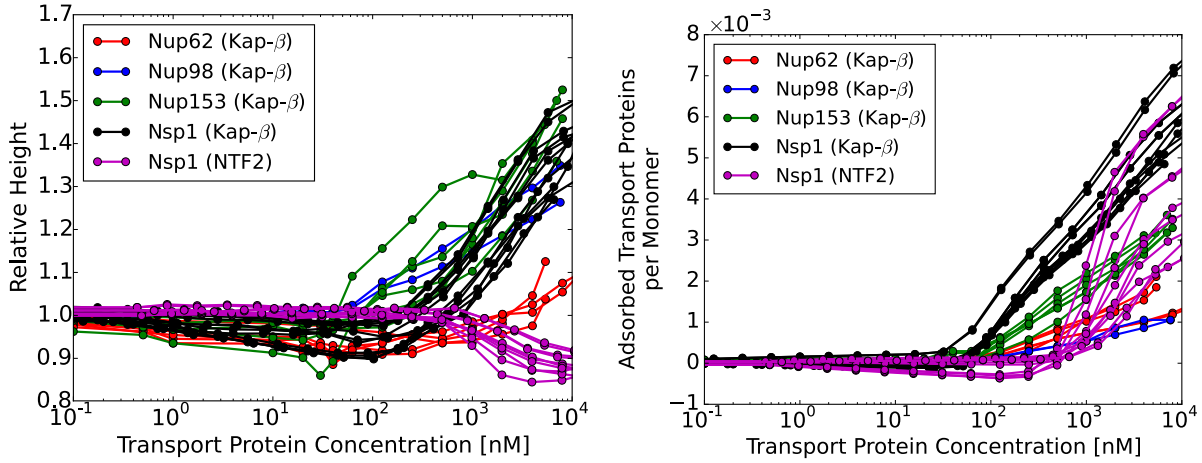
patch or four amino acids, is reasonable since the experimental values all fall between the good and bad solvent regime. The main results of this chapter are insensitive to the exact choice of these parameters as shown in Section 3.5. The goal of the model is not to determine the exact parameters corresponding to the individual FG nups but to explain the contrasting behaviour observed in experiments. The agreement with experiments for the pure FG nup brush identifies reasonable values for monomer size, bond length, and cohesiveness and supports the application of the model to the more complicated system of FG nups in the presence of transport proteins.

3.3.2 Collapse and Expansion of the FG Nup Layer in the Presence of Transport Proteins

I next describe the predictions of the model when transport proteins are added into the solution above the FG nup layer and compare with experimental observations. Although initially it seems that the addition of transport proteins to an FG nup layer can lead to highly variable behaviour, such as an increase or a decrease in the layer height depending on the exact conditions, the experiments all follow the general trend shown in Figure 3.4, which reproduces data of Kapinos et al. [159] and Wagner et al. [160]. At first, the layer height is unperturbed by an increasing concentration of transport proteins. A further increase in concentration results in some degree of compaction, eventually followed by extension of the layer at

high transport protein concentrations. The collapse of the layer height can be negligible compared to the pure brush, as in the case of Nup98. The collapse of the layer is concurrent with the accumulation of transport proteins inside the layer. The width and the magnitude of the collapse, transport protein concentration at the layer collapse, as well as the curve describing accumulation of transport proteins in the layer depends on the grafting distance, the transport proteins, as well as the specific FG nups forming the layer. In particular, there is a clear difference between the two transport proteins Karyopherin-β1 and NTF2. NTF2 is roughly 3 times smaller in volume and the collapse of the layer height as well as the accumulation of transport proteins in the layer occur at a transport protein concentration in solution which is at least an order of magnitude greater than for Karyopherin-β1. Based on these observations, we can define three qualitatively different regimes of behaviour of the layer height as the concentration of transport proteins in the external solution is increased:

1. Initial collapse followed by recovery and extension of the layer height
2. Only extension of the layer height (no/negligible collapse)
3. Only collapse and no extension of the layer height at physiological concentrations of transport proteins

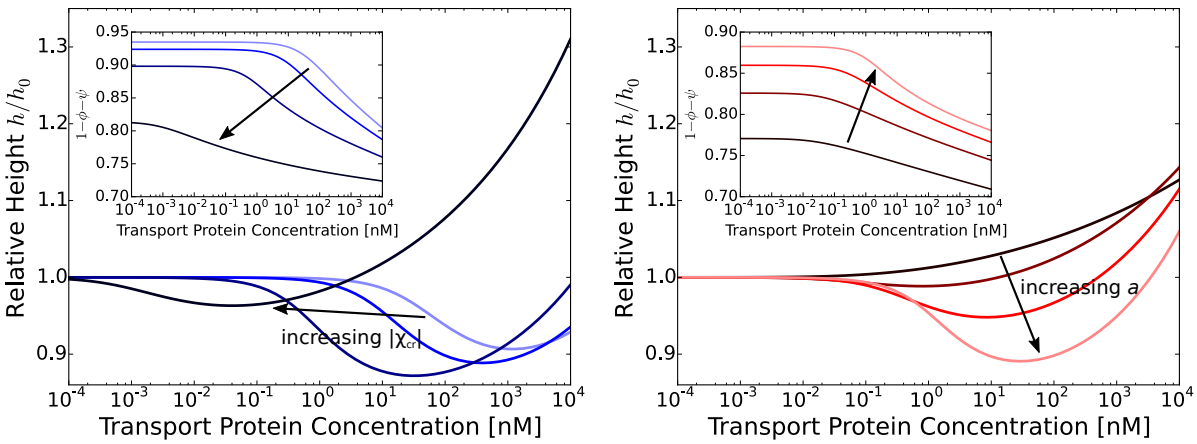


(a) (b)

Figure 3.4: **Characteristic responses of FG nup layers to the transport proteins: experimental results.**
(a) Change in the layer height relative to the unperturbed layer as a function of the transport protein concentration in the outside solution.
(b) Number of the transport proteins in the layer per unit length of the FG nup chain.
Each line corresponds to a different run with a different initial layer height and grafting distance. Different colors correspond to different FG nups, which all exhibit qualitatively similar behavior. The colours indicate different combinations of transport proteins and FG nups: Karyopherin-β1 on Nup62, Nup98, Nup153, and **Nsp1**, and NTF2 on Nsp1. The corresponding average grafting distances are $\sim$ 2.5 nm, $\sim$ 4.2 nm, $\sim$ 4.5 nm, $\sim$ 3.75 nm, $\sim$ 3.75 nm. The data are from Refs. [159, 160]

Figure 3.5 demonstrates that the model reproduces the same qualitative behaviour as the experiments. As the concentration of transport proteins in solution increases we see the same collapse and subsequent extention of the FG nup layer height corresponding to the accumulation of transport proteins inside the layer. Increasing the strength of monomer-monomer interactions χ_{cr} in the model, which could vary depending on the FG nups used as well as the solvent properties in the experiments, shifts

the location of the minimum of the layer height to lower transport protein concentrations, and has a non-monotonic effect on the relative magnitude of the collapse, at first increasing, then reducing, and almost eliminating it entirely. The role of monomer cohesiveness is discussed in detail in Section 3.3.3. Increasing the grafting distance, reduces the monomer density, which has the effect of enhancing the relative magnitude of the collapse and shifting the location of the minimum layer height in the opposite direction - to higher transport proteins concentrations. A higher grafting distance reduces the monomer density allowing for a greater degree of collapse. The other model properties that modulate this behaviour will depend on the transport proteins used - their size and interaction strength with the FG nups.



(a) (b)

Figure 3.5: **Collapse and recovery: effect of cohesion and of the grafting distance.**
(a) Theoretical curves show that FG nup cohesion can convert layer collapse to swelling. The cross-linking strengths are $\chi_{cr} = 0, -0.4, -0.8, -1.1$ for $a = 5$ nm and $\chi = -550$.
(b) Increasing grafting distance increases the magnitude of the layer compaction. The lines correspond to model predictions for $a = 3, 4, 5, 6$ nm for $\chi = -530$ and $\chi_{cr} = -1$.
The insets show that the fraction of free space in the layer, calculated as $1 - \phi - \psi$, decreases with the addition of the transport proteins. $b = 1$ nm, $l = 0.67$ nm in both panels.

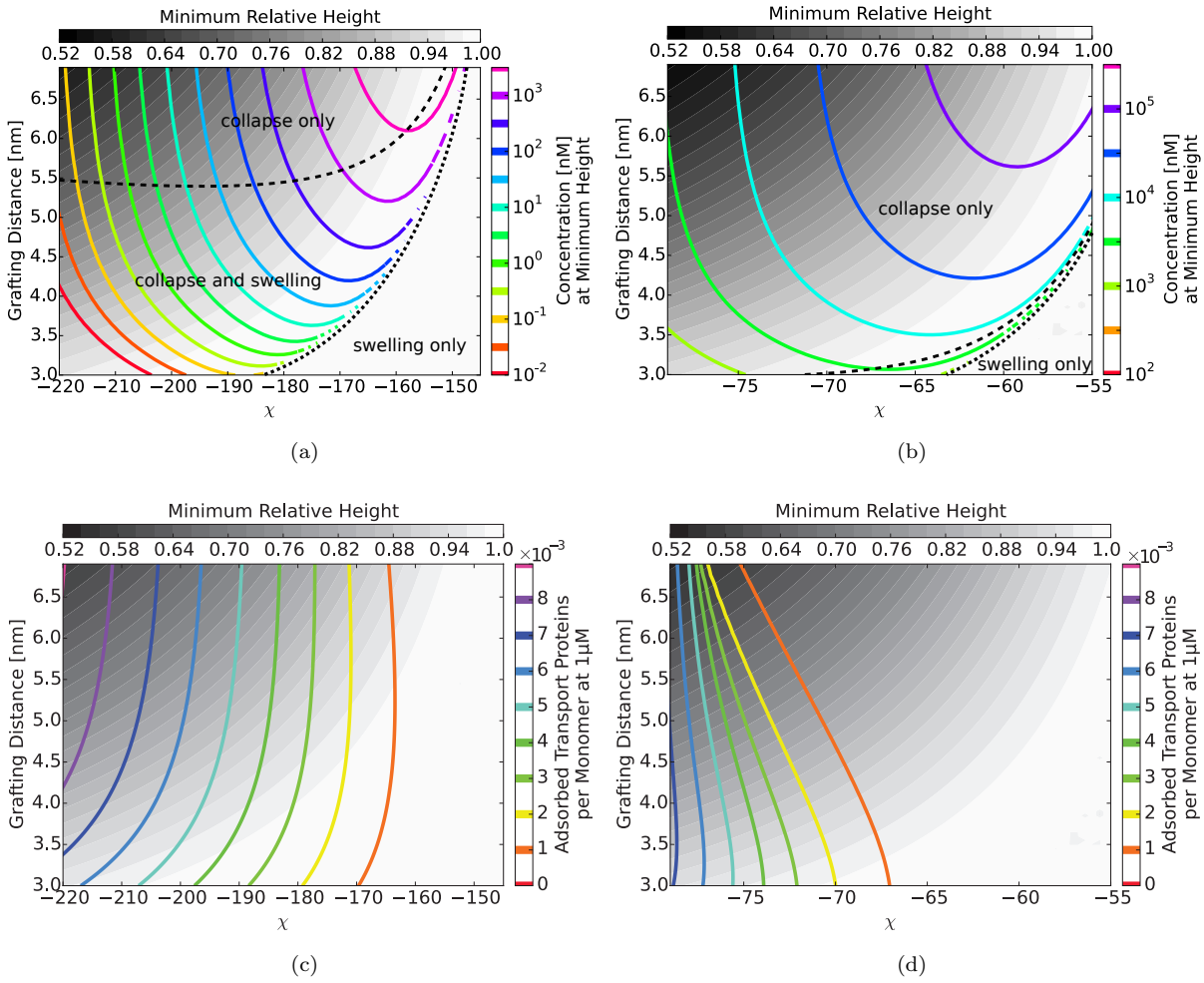


Figure 3.6: **The "phase diagrams" of predicted behaviors for the conformational transitions and amount of transport protein in the layer**
In all panels, the grayscale color denotes the degree of layer compaction, h_{min}/h_0, relative to the unperturbed layer (color legend is on top). (a) and (c): $\bar{v} = 125$, roughly corresponding to Karyopherin-β1. (b) and (d): $\bar{v} = 40$, roughly corresponding to NTF2.
(a) and (b): The colored contour lines indicate the corresponding bulk concentration c_{min} at which the minimal layer height is achieved (legend on the right side). Above the dashed line, which separates the "collapse only" and "collapse and swelling" regions, there is no swelling of the layer height for up to 1 μM transport protein concentration. The "swelling only" region is where the response of layer height to bulk transport protein concentration (e.g. Fig. 3.5) does not have a minimum, and the boundary between the "collapse and swelling" and "swelling only" regions is the common curve approached by the "Concentration at Minimum Height" contour lines. A dotted line has been added manually to indicate this boundary. The overall phase diagram topology is similar in both cases, but for smaller protein, the collapse is more pronounced and occurs at lower χ.
(c) and (d): The colored contour lines show the amount of adsorbed proteins in the layer per chain monomer. Higher degree of collapse is correlated with higher accumulation of the proteins in the layer.
In all panels, $b = 1.52, l = 1$ nm, corresponding to the "monomer" size of roughly four amino acids.

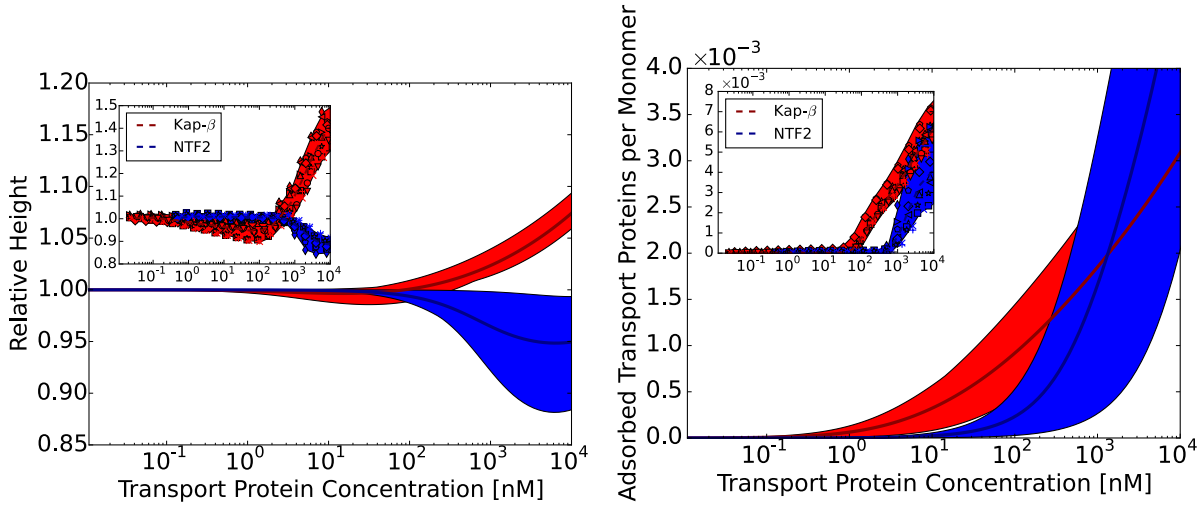
In order to get a more comprehensive understanding of the model, its predictions are synthesized in a "phase diagram" of behaviours shown in Figure 3.6. The diagrams show the effect of varying monomer-transport protein interaction strength and FG nup grafting distance for two transport proteins

of different sizes: Karyopherin-β1 and NTF2. Both Figures show the relative magnitude of maximal collapse using grayscale shading in the background. Using coloured contour lines, Figure 3.6ab shows the transport protein concentration at the minimum of the layer height. On the second set of diagrams (Figure 3.6cd), the coloured contour lines show the number of transport proteins in the layer for a fixed transport protein solution concentration of 1 μM. The upper phase diagrams (Figure 3.6ab) are partitioned into three regions corresponding to different regimes of behaviour of the layer height. In the lower left (weak monomer-transport protein interaction strength and low grafting distance), there is no noticeable collapse of the layer height and only extension. In the region above the black dashed line, there is no swelling of the layer at concentrations of transport proteins in the external solution below 1 μM. In the intermediate region between the two, there is initially a collapse followed by a recovery and extension of the layer height. This region is effectively eliminated for NTF2.

Different experimental conditions correspond to different positions on the phase diagrams, which can resolve some of the contradictions from experiments such as the lack of significant change in the layer height reported by other groups [158, 161]. The collapse of the layer height is not an intrinsic feature of the FG nups and can be reduced and eliminated entirely for some experimental conditions. The response of the layer height to increasing transport protein concentration depends on the grafting distance, monomer cohesion (see Section 3.3.3 for an in depth discussion), and transport protein interaction strength, the latter two of which can be affected by solvent properties such as pH, salt, and denaturant. The phase diagrams also demonstrate the diverse effects of transport protein size on the layers. For smaller transport proteins, the concomitant collapse and extension for concentrations below 1μM is eliminated. Another interesting difference is in the effect of grafting distance on large and small transport proteins. For the large Karyopherin-β1, increasing the grafting distance increases the number of transport proteins that adsorb into the layer, because the monomer volume fraction decreases, reducing the repulsive barrier. For small transport proteins such as NTF2, the repulsion is less important and an increase in grafting distance hinders their entry into the layer as there is a lower density of FG nup monomers for them to bind to.

Figure 3.7 shows a semi-quantitative comparison of the model with experimental measurements of layers of Nsp1 in solution with Karyopherin-β1 or NTF2 from [160]. Due to the relatively large experimental uncertainty, a range of grafting distances and interaction strengths are used to make predictions from the model. The easily discernible effects of the two transport proteins on layers of the FG nup Nsp1 are: greater magnitude of the collapse but at a higher transport protein concentration for NTF2, as well as a delayed but steeper increase of the number of adsorbed transport proteins in the layer again for the smaller NTF2. The model succeeds in reproducing these qualitative differences, but a more precise comparison would require both less uncertainty in the experimental measurements as well as a more sophisticated model. In general, as transport protein volume varies and χ per transport protein surface area remains fixed, the relative magnitude of the collapse is more pronounced for smaller transport proteins, and occurs at higher concentrations in solution.

The physics behind the penetration of a single transport protein into the FG nup layer relies on the balance between the entropic cost and energetic gain of entering the layer. Although there is significant free space within the layer, this entropic repulsion is sufficient for creating the permeability barrier for non-binding molecules and no cohesive interactions or formation of a network between FG nups are necessary, although monomer cohesiveness will further strengthen the barrier to inert molecules. If the FG nup - transport protein interaction is sufficiently strong, the transport protein will enter the layer.



(a) (b)

Figure 3.7: **Comparison of the theoretical predictions with the experimental data.** Theoretical predictions for the range of the parameter values approximately corresponding to the experimental ones for Nsp1 layers infiltrated by Karyopherin-β1 and NTF2.
(a) Layer height vs bulk concentration of the transport factor.
(b) Amount of adsorbed transport protein in the layer as a function of the concentration in the solution. The shaded regions correspond to $3.5 < a < 4$ nm and $-185 < \chi < -175$ for Karyopherin-β1 and $-73 < \chi < -63$ for NTF2. For all lines, $b = 1.52, l = 1$ nm and $\chi_{cr} = -1$. The insets show the corresponding experimental data from [160].

At low concentrations, the transport proteins do not cause significant conformational changes as there is ample empty space in the layer for them to occupy. Once a sufficiently high number of transport proteins has accumulated in the layer or for high interaction strengths, further addition of the transport proteins causes a cooperative transition of the FG nups leading to either collapse or swelling of the layer. The collapse is due to the energetic gain of more contacts between the FG nups and transport proteins outweighing the entropic cost of a denser volume fraction in the layer. The swelling is due to the already dense layer needing to increase it's height in order to make room for and continue to adsorb more transport proteins.

3.3.3 The Role of Monomer Cohesiveness

Section 3.3.1 showed that in the absence of transport proteins, an increase in the cohesive strength between monomers ($|\chi_{cr}|$) always causes a decrease in the layer height. The presence of transport proteins in the solution results their penetration into the FG nup layer. As described in Section 3.3.2, increasing the concentration of transport proteins in solution causes an initial collapse, which can be negligible, followed by a recovery and swelling of the FG nup layer height. This behaviour is determined by the balance of the energetic gain of contacts formed between the FG nups and transport proteins and the entropic cost of entering the brush. Intuitively, an increase in monomer cohesiveness should enhance the FG nup barrier of the NPC. In this section we will see that without altering the transport protein properties or monomer-transport protein interaction strength, monomer cohesiveness alone can also enhance the collapse of the FG nup brush and penetration of transport proteins into the layer.

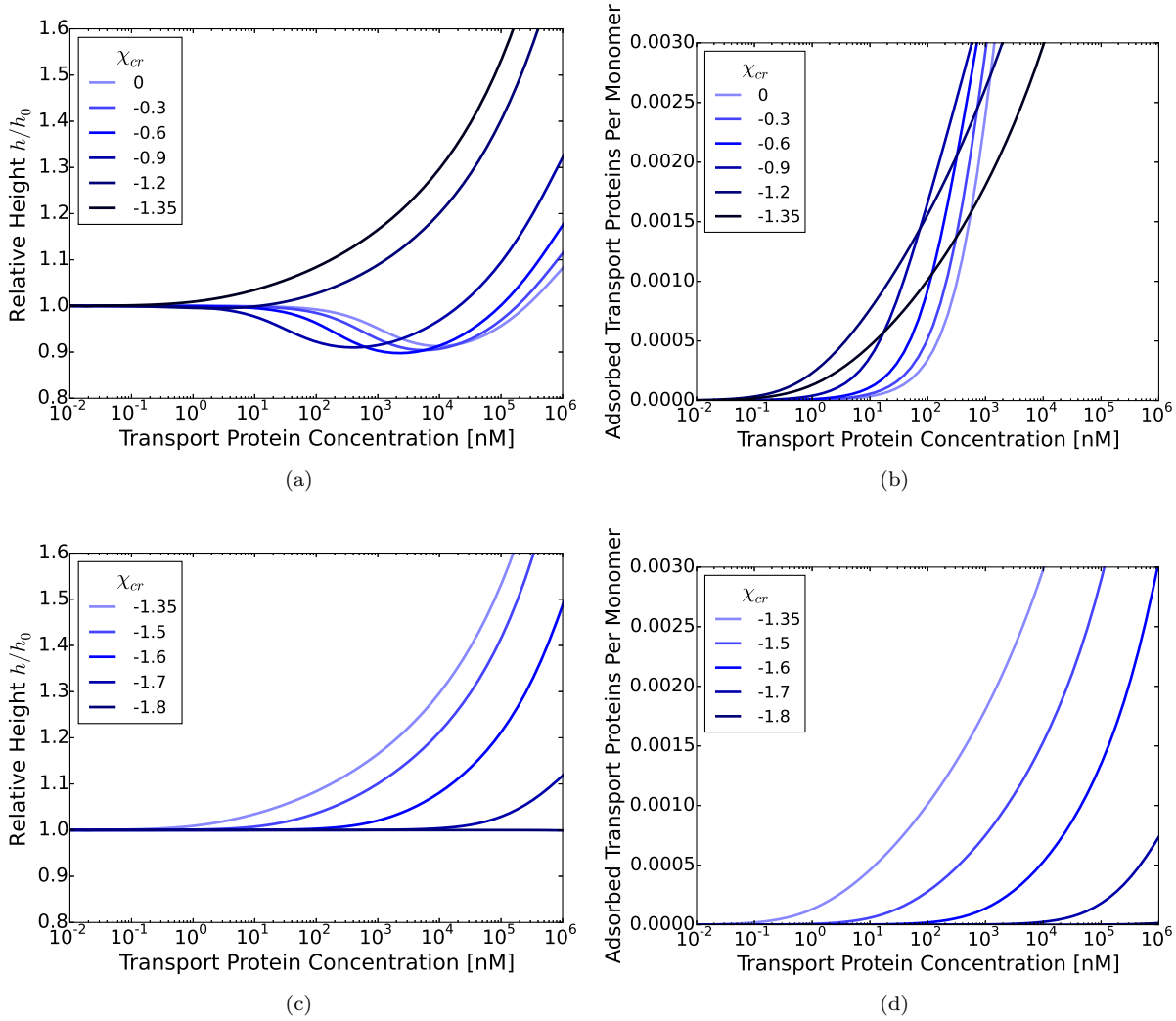
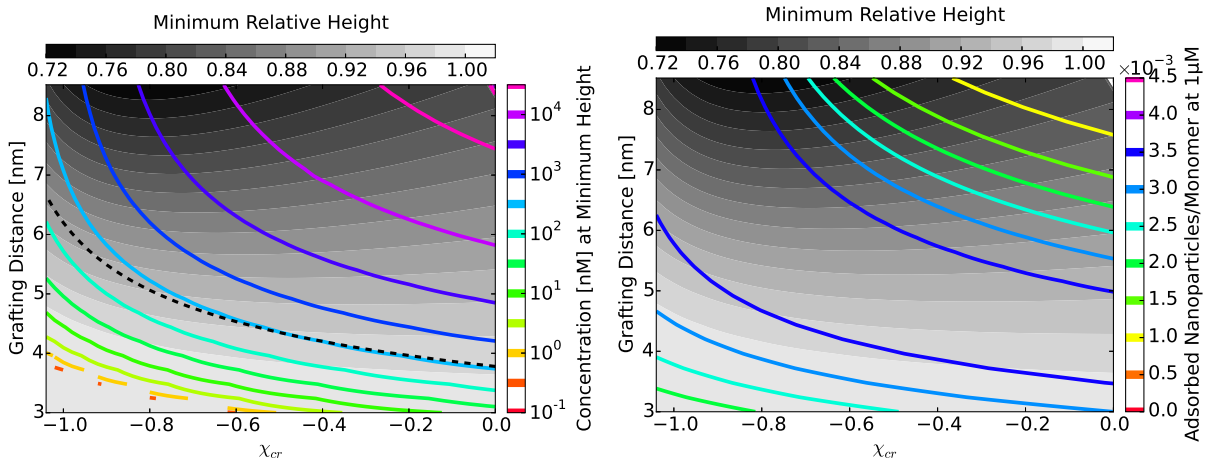


Figure 3.8: **Effect of cohesiveness on the layer height.**
(a) and (b): χ_{cr} is varied between 0 and -1.35.
(c) and (d): χ_{cr} is varied between -1.35 and -1.8.
(a) and (c): Collapse and recovery of normalized layer height
(b) and (d): Number of adsorbed transport proteins in the layer.
The transport protein properties are: $\chi = -187.5$ and $\bar{v} = 125$. The monomers have size $l = 1$ nm and bond length $b = 1.52$ nm.

Figure 3.8 shows the response of the relative height of the FG nup layer and the number of transport proteins adsorbed into the layer as the concentration of transport proteins in solution is increased, for different strengths of monomer cohesiveness. As $|\chi_{cr}|$ is increased from 0 to 1.35, the relative swelling of the layer increases and the collapse region is shifted to lower concentrations and eventually disappears. The relative magnitude of the collapse also increases initially with increasing $|\chi_{cr}|$. The number of adsorbed transport proteins similarly increases for relevant concentrations in solution, but reaches a maximum at a lower value of $|\chi_{cr}|$. When the cohesiveness is further strengthened beyond this point, the number of adsorbed transport proteins, relative magnitude of the collapse, and the relative swelling of the all layer decrease.



(a) (b)

Figure 3.9: **The "phase diagrams" of predicted behaviors: effect of grafting distance and monomer cohesiveness.** The grayscale color denotes the degree of layer compaction, h_{min}/h_0, relative to the unperturbed layer (color legend is on top).
(a): The colored contour lines indicate the corresponding bulk concentration c_{min} at which the minimal layer height is achieved (legend on the right side). There is no layer swelling above the dashed line (up to 1 μM transport protein concentration).
(b): The colored contour lines show the amount of adsorbed proteins in the layer per chain monomer. In the lower panel we see a non-monotonic variation in the number of transport proteins in the layer at 1 μM for cohesiveness and grafting distance.
The transport protein properties are: $\chi = -187.5$ and $\bar{v} = 125$. The monomers have size $l = 1$ nm and bond length $b = 1.52$ nm.

Figure 3.8 demonstrated that increasing monomer cohesiveness results in an increased number of transport proteins in the layer for certain concentrations of transport proteins in solution. However, the non-monotonic behaviour is not universal to all choices of parameters as more clearly shown in Figure 3.9 for varying grafting distance and monomer cohesiveness. Above a grafting distance of about 4 nm, increasing the strength of monomer cohesiveness $|\chi_{cr}|$, causes an initial increase in the number of adsorbed transport proteins per monomer for a concentration of 1μM of transport proteins in the external solution. The number of adsorbed transport proteins per monomer peaks between 3.5×10^{-3} and 4.0×10^{-3} at a monomer cohesiveness which increases with grafting distance. The region of grafting distances and monomer-transport protein interaction strengths where this behaviour persists is summarized in the contour plot in Figure 3.10. The critical χ_{cr} is the value at which the number of transport proteins in the layer reaches its maximum for a particular concentration in solution. In the region where the critical χ_{cr} is 0, any increase of monomer cohesiveness only lowers the number of transport proteins in the layer.

It is surprising that for certain combinations of parameters, increasing the cohesiveness between monomers in the layer would result in greater adsorption of transport proteins into the layer. Intuitively, it would seem that cohesiveness would compact the layer, increasing entropic repulsion and only forcing transport proteins out. However, a higher cohesiveness results in a higher density of monomers and more binding events between monomers and transport proteins, increasing the energetic gain for a transport protein to enter the layer. The balance between these two contributions can be tuned using the grafting

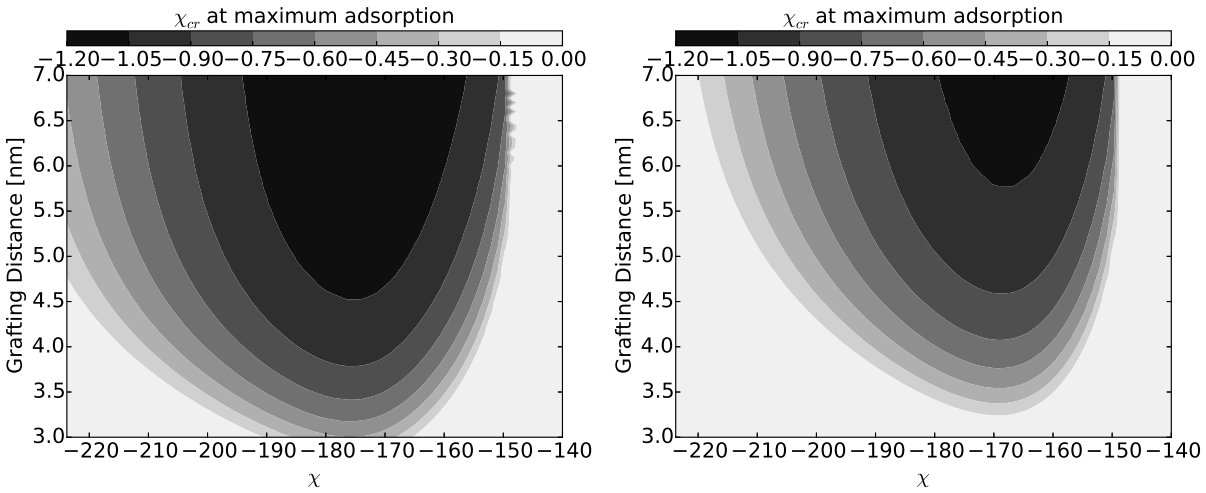


Figure 3.10: A contour map showing the monomer cohesiveness at which the number of transport proteins in the layer is maximized for 100 nM (a) and 1000 nM (b) of transport proteins in solution. The transport protein volume is $\bar{v} = 125$. The monomers have size $l = 1$ nm and bond length $b = 1.52$ nm.

distance. At low grafting distances, the monomer density is already high at $\chi_{cr} = 0$ and for transport proteins entering the layer, the entropic penalty outweighs any energetic gain. At high grafting distances, the monomer density is fairly low and the permeability can be improved by an increase in cohesiveness as the energetic gain would outweigh the negligible entropic cost. As the cohesiveness is increased further ($|\chi_{cr}| >\sim 1.2$), the entropic cost dominates in all cases, reducing the permeability of the FG nup layer.

3.4 Conclusions

The model introduced in Section 2.2 provides a rigorous physical framework for investigating the structure of surface grafted FG nups and their interactions with transport proteins. The complexity of the FG nups and transport proteins is encapsulated by a small number of key parameters. The model explains the behaviour of collapse and extension and accumulation of transport proteins in the layer and the regimes of parameters which control this behaviour have been identified. The model is able to qualitatively and semi-quantitatively reproduce the range of *in vitro* experimental observations and explains how initial contradictions could arise between them.

One of the discrepancies in the experimental measurements of binding affinities of transport proteins to FG nups is the range of measured dissociation constants from nanomolars to micromolars. Some of these affinities are also inconsistent with transport times in the millisecond range [119, 122, 143, 147, 155, 156, 180]. These discrepancies might stem from the fundmental statistical physics of the transport protein-FG nup interaction. The adsorption of transport proteins into the layer is a cooperative process and is not well described by a single Langmuir isotherm typically used to interpret binding assays. Additionally, the entry of transport proteins into the layer is not only due to energetic but also entropic effects and therefore the measured affinity does not reflect the binding energies directly. Even the energetic component of the interaction can vary with experimental conditions, since the average number

of monomers available to bind to a transport protein depends on the monomer concentration which itself depends on the layer height and grafting distance [150, 159, 180, 181]. These points reinforce the idea that the classical characterization of transport protein-FG nup binding by a single affinity value is not informative for the complex interactions of spatially extended objects such as FG nup layers.

The model can reconcile conflicting theories of the roles of FG nups and transport proteins in NPC transport. The analysis of the model suggests that both entropic ("brush"-like) and cohesive ("gel"-like) effects naturally cooperate in determining the structure of the FG nup and transport protein layers. Quantitative comparison with experiments shows that all FG nups likely possess some degree of cohesiveness, which modifies the structural transition and enhances the selectivity barrier for some transport proteins and reduces it for others. The different classes of "extended" and "collapsed" FG nups [46] can be mapped to different values of the cohesiveness parameter χ_{cr}.

Partially due to uncertainty of experiments, but also due to the simplicity of the model, a full quantitative comparison with experiments is not possible, and a more sophisticated model would need to be developed. The limitations of the model include the way amino acid and transport protein geometries are represented, the uniform density assumption of the FG nup layer, and the mean field approximation of the energy of the monomer-monomer and monomer-transport protein interactions. Nevertheless, the results of the model are in good agreement with more detailed simulations [86, 87]. Furthermore, one needs to be careful when using *in vitro* experiments to make inferences about the NPC transport mechanism and the same can be said for our model. The model aims to address surface grafted FG nups, which differ from the cylindrical geometry of the NPC. Some recent work of polymers in channel geometries allows us to evaluate which predictions of the model are applicable to the NPC [171]. Polymers in relatively wide channels, such that the center is unoccupied, will collapse towards the walls and extend towards the center with increasing transport protein concentrations analogous to brushes on a flat surface collapsing and extending their height as in the model. However, if the polymers are long enough relative to the channel width, there will be accumulation of chains at the center, this case is not reducible to the planar brush model. More work is needed to understand where between these two extremes lies the structure of FG nups in the NPC. Because the model assumes that the transport proteins are isotropic, it cannot address the effect that the shapes of transport proteins and cargos could have on the morphology of the FG nup layers. Even if the shape does not significantly alter the equilibrium properties of the layers, Moussavi-Baygi and Mofrad [182] showed, using Brownian dynamics simulations, that elongated cargo require fewer binding sites and a smaller nudging force in order to penetrate into an FG nup assembly in a channel geometry, when compared to spherical cargo, indicating that elongated shapes may be transported more efficiently through the NPC.

The success of the model relies on the robust physical mechanisms underlying the function of the NPC. The collapse and expansion of the layer height as the concentration of transport proteins increases is not unique to FG nups but general for all surface grafted polymer systems with adsorbent molecules. Thus, the results are applicable not only to understanding NPC structure and function and but also the design of biomimetic transporters. Recently, several NPC mimics have been developed, featuring a DNA scaffold for the central channel and tunable parameters such as the number, type, and position of FG nups [183, 184]. The analysis of the model identified and classified the important parameters and behaviours and shows how an analogous system can be controlled. Another interesting observation which can be understood within the framework of the model, is the ability of the NPC to maintain viability even when significant numbers of the FG nups are deleted. The deletion of FG nups roughly parallels

lowering the grafting distance in the model, while maintaining the qualitative regime of behaviour. When this deletion is performed in the NPC, the neighbouring chains may simply swell without significantly altering the selectivity properties of the layer [124, 126].

Finally, although the qualitative behaviour described by the model and observed experimentally is very general, the specific quantitative features such as the exact layer height (which is normalized out in the analysis of the model) and the degree of compaction are rather sensitive to the exact parameter values, such as the grafting distance, transport protein concentration, and the interaction strength. This raises the question of how the NPC can maintain its function with large variations in size, spatial organization, and divergence of FG nup sequences across species, and despite large perturbations such as deletion of many of the FG nups? One possible solution is that the NPC is finely tuned and only works correctly when all of the parameters are just right: FG nup types, amounts, and localization, local transport protein and other molecule concentration, pH, and ionic strength, are all exactly set by cellular homeostatis. The other possibility is that the NPC is robust: any structure of FG nups with approximately the right physical properties will function efficiently enough to maintain selective transport, which would provide an example of function conservation in the absence of sequence conservation.

3.5 Parameter Sensitivity Analysis

The model used in this chapter required assumptions about the monomer size and bond length. We made the choice: $l = 1$ nm, $b = 1.52$ nm, corresponding to about 4 amino acids per monomer. In this section, we demonstrate that the model is not sensitive to changes in these parameters. We present results for the parameters b and l corresponding to around 2 or 3 amino acids in Figure 3.11, although we tested other combinations with similar predictions for reasonable values. The phase diagrams recapitulate the qualitative behaviour from before. The same behaviour of collapse and expansion is observed and the model is able to differentiate between the behaviour of large Kap-β1 transport proteins, and small NTF2.

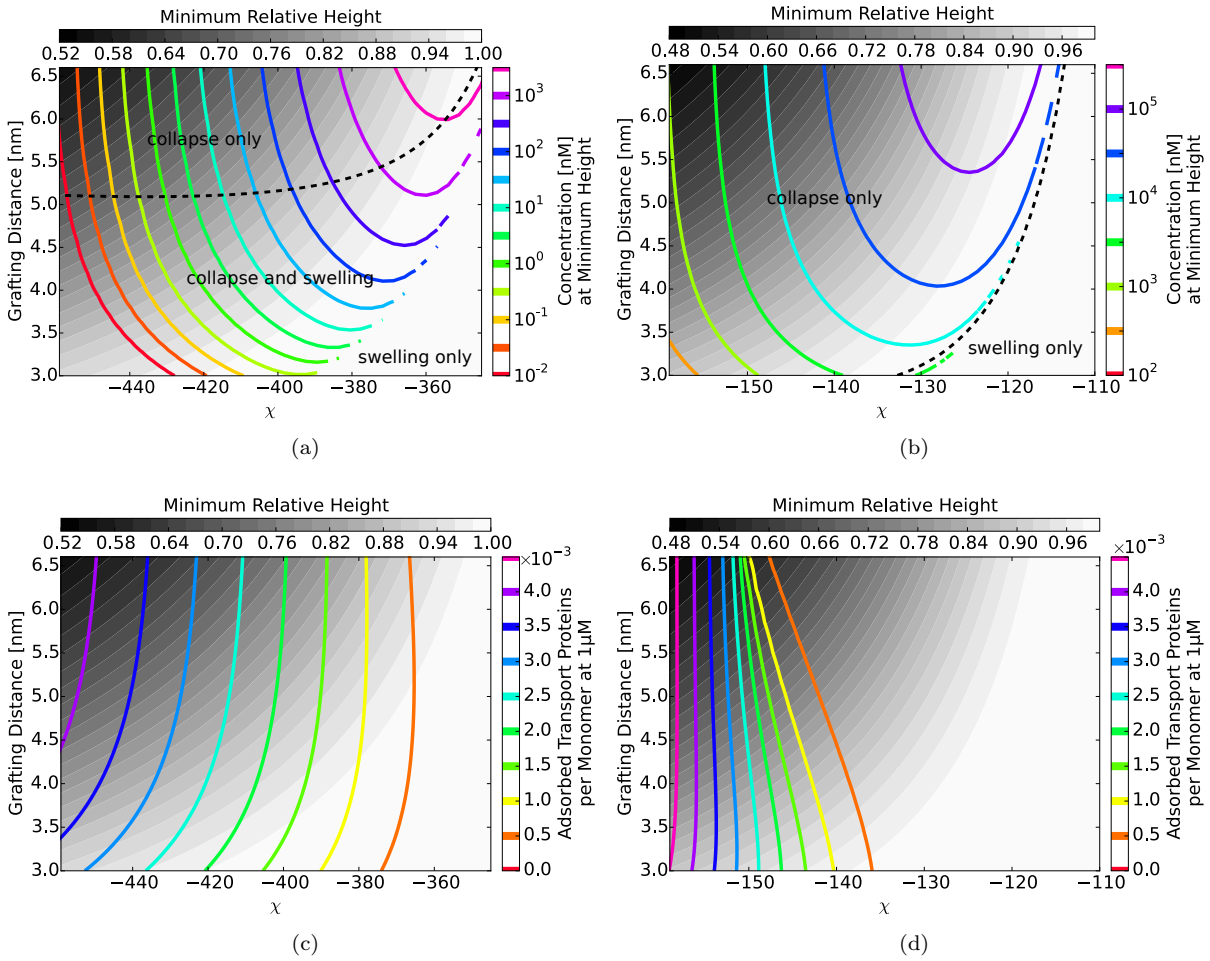


Figure 3.11: **The "phase diagrams" of predicted behaviors.** The grayscale color denotes the degree of layer compaction, h_{min}/h_0, relative to the unperturbed layer (color legend is on top). *Upper panels (a) and (b):* conformational transitions of the layer. The colored contour lines indicate the corresponding bulk concentration c_{min} at which the minimal layer height is achieved (legend on the right side). There is no layer swelling above the dashed line (up to 1 μM transport protein concentration). *Lower panels (c) and (d):* amount of transport protein in the layer. The colored contour lines show the amount of adsorbed proteins in the layer per chain monomer (legend on the right side). Higher degree of collapse is correlated with higher accumulation of the proteins in the layer. *Left panels (a) and (c):* $\bar{v} = 295$, roughly corresponding to Karyopherin-β1; *Right panels (b) and (d):* $\bar{v} = 95$, roughly corresponding to NTF2. The overall phase diagram topology is similar in both cases, but for smaller protein, the collapse is more pronounced and occurs at lower interaction strengths χ. In all panels $b = 1.4, l = 0.75$ nm, corresponding to the "monomer" size of roughly between two and three amino acids.

Chapter 4

Relationships Between Amino Acid Composition and IDP Dimensions

The results of this chapter are being prepared for publication.

4.1 Introduction

In Chapter 1, we reviewed how the structures of IDPs may be viewed through the framework of polymer physics. An IDP's ensemble of conformations lies somewhere on the disorder-to-order continuum, and depends on its amino acid composition, sequence, and the properties of the solvent. At the disordered extreme, all polypeptides may be described by good solvent polymer models, and sequence heterogeneity is unimportant. At the ordered extreme, natively folded proteins adopt unique conformations encoded by their sequences, requiring atomistic detail for a proper description. IDPs are located at intermediate locations on this continuum [3, 23, 77]. Simple polymer theories can lead to discrepancies when interpreting experimental results, most notably between the polymer dimensions obtained using FRET and SAXS [60, 61]. On the other hand, all-atom descriptions are often unfeasible due the vast numbers of possible conformations of IPDs and are prone to overfitting [70]. Some properties of the sequences of IDPs suggest that the atomistic details may be coarse-grained out [4, 9, 18, 21, 42]. In this Chapter, I use coarse-grained models to investigate the effects of monomer cohesiveness and sequence patterning on the polymer dimensions of IDPs.

The mean field model used in Chapter 3 and analogous model successfully described both surface grafted layers and polymer solutions of FG nucleoporins [78, 79]. The inter-monomer interactions and solvent properties were encapsulated by a single parameter χ_{cr} representing the overall cohesiveness of the FG nups. Despite their successes, mean-field models fail to differentiate between different polymer dimensions (end-to-end distance, radius of gyration, and hydrodynamic radius) and cannot capture the effects of sequence heterogeneity. FG nup sequences contain patches of cohesive amino acids, interspersed by spacer regions which can contain charged amino acids [46, 120, 121]. Although specific cases have been investigated, the effect and importance of the patterning of the different types of amino acids on the dimensions and functions of the FG nups and IDPs in general are unknown [23, 27, 46, 50, 51].

In Section 4.2, I present the results of simulations of a homopolymer model, which includes a minimal level of detail but is capable of discerning between the ensembles of IDPs with different locations on the

disorder-to-order continuum. Similar to the mean-field model used in Chapter 3, all information about an IDP's sequence and the properties of the solvent are encapsulated in a single cohesiveness parameter ϵ. Unlike the mean-field model, the simulations are able to differentiate between the various polymer dimensions: end-to-end distance, radius of gyration and hydrodynamic radius, and how the relative polymer dimensions change with ϵ. These results can potentially address the discrepancies arising from experimental results due to the use of simple monomer theories to infer the polymer dimensions of IDPs. For example, if an IDP at a particular denaturant concentration, is associated with a unique cohesiveness value, it is not surprising that the radius of gyration measured with SAXS would disagree with the radius of gyration inferred from the end-to-end distance information from a FRET measurement under the assumption that the ratio between end-to-end distance and radius of gyration is constant [36, 53].

In Section 4.3, to investigate the effects of sequence heterogeneity on the ensemble of conformations, the homopolymer model is expanded to a four letter model: every amino acid is represented by a monomer of one of four types (cohesive, neutral, positively charged, or negatively charged). This model can distinguish between disorder and order promoting amino acids within a sequence as well as the patterning of charged amino acids. Instead of focusing on specific IDPs, I infer general relationships between sequence patterning and the polymer dimensions, and show how the relative polymer dimensions can deviate from the predictions of the homopolymer model. Specifically, I investigate three sequence features, which can qualitatively be described as: (i) the segregation of charged residues, (ii) the size of cohesive patches, and (iii) the segregation of charged and cohesive residues. In each case, the composition (fraction of each type of monomer) remains fixed as the sequence varies. Sequences of 50 and 60 monomers are studied to match the amino acid lengths of sequences previously studied using the all-atom ABSINTH model [50] and the typical length of IDP fragments used in FRET characterization [35, 36].

4.2 Ratios of Polymer Dimensions as a Ruler for IDP Conformations

4.2.1 Homopolymer Model

In the homopolymer model the parameters for the interactions, described in Section 2.3, are identical for all monomers. All properties of the solvent, and the composition and sequence of an IDP are captured by a single parameter ϵ (or $E = \frac{\epsilon}{kT}$), which is the cohesiveness between the monomers in the coarse-grained model. A low ϵ represents a protein in high denaturant conditions or an IDP with many disorder promoting amino acids in its sequence. Increasing ϵ represents decreasing the denaturant concentration or substituting order promoting amino acids into an IDP sequence.

In simulations, the range (diameters) of excluded volume and cohesive interactions was $B = B_0 = \sqrt{\frac{3}{2}}$. The maximal extension of the FENE bonds between monomers was $L_{max} = 2B_0$. The strength of excluded volume interactions was $\frac{\epsilon_{LJ}}{kT} = 1$. The hydrodynamic radii of the monomers were $A = \frac{B}{2}$. Cohesive forces were cutoff at a distance of $4B$. There were no electrostatic interactions between the monomers.

Simulations were performed for $N = 100$ monomers and cohesive interactions strengths ranging from $\frac{\epsilon}{kT} = 0$ to $\frac{\epsilon}{kT} = 1.9$ inclusive, in intervals of 0.1. For each ϵ, 40 runs were performed, each lasting

10^8 steps, with a time step of $\Delta T = 0.001$. Each run began from a self-avoiding random walk initial condition. The first 10^6 steps were excluded from the analysis in order to avoid biasing the results by the initial conditions, and averages were taken over steps over runs and steps.

4.2.2 Ensemble of Conformations and Dimensions Determined by ϵ

In this section, we explore the effects of the cohesiveness parameter ϵ on a polymer's ensemble of conformations. Chemically denatured proteins follow the scaling laws of the good solvent or SAW model [34, 35]; this corresponds to $\epsilon = 0$. Although the homopolymer model cannot differentiate between the unique structures of folded proteins, the ordered extreme can be approximated by high cohesiveness between the monomers (roughly $\epsilon > 1.5$), which is equivalent to a poor solvent [41]. Homopolymers adopt collapsed globule conformations at high cohesiveness [77, 80]. Based on measurements of the polymer dimensions and scaling behaviour of IDPs [35, 36, 38], on average, IDP cohesiveness is equivalent to the θ point of the homopolymer. For our model, the θ point occurs at $\frac{\epsilon}{kT} \approx 0.7 - 0.75$, and is investigated in Appendix B. However, individual IDPs deviate from this average and as a first approximation we can associate each one with a unique ϵ.

In general, all of the polymer dimensions compact with increasing ϵ and are shown in Figure 4.1a. The figure shows the square roots of the ensemble averages of square end-to-end distance and radius of gyration, and the hydrodynamic radius. The end-to-end distance has been scaled down by a factor of $\sqrt{6}$ to be comparable to the other dimensions. Overall, all of the polymer dimensions undergo a similar transition at the θ point when $\frac{\epsilon}{kT} \approx 0.7 - 0.75$. The dimensions decrease monotonically as the chain compacts from a coil to a globule. The end-to-end distance undergoes the greatest relative compaction, while the hydrodynamic radius experiences the least change. The inset of Figure 4.1a shows the predictions of the mean-field model of Section 2.1.3 for the equilibrium polymer size R_*, which does not correspond to any of the dimensions directly [37]. The mean-field model predicts that the θ point occurs at a corresponding $\epsilon \approx 0.64$ kT (see Section 2.3.5).

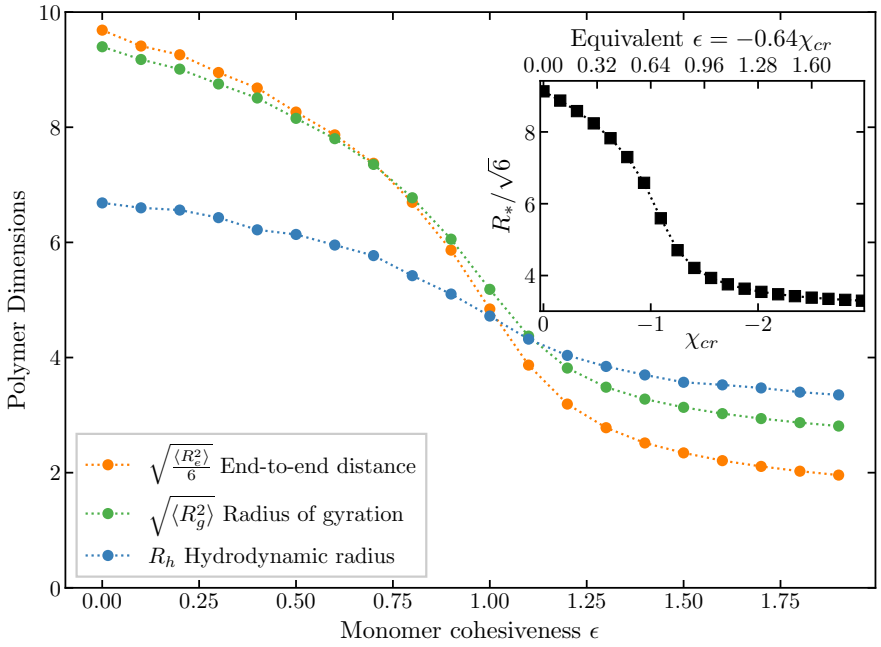
Another common measure of the shapes of polymer conformations is the asphericity δ (sometimes referred to as the shape anisotropy) [36, 58]. Compared to the end-to-end distance, the asphericity is less sensitive to the positions of the chain ends, and describes the global shape of the polymer, however it is not readily accessible by experiment. For a rod $\delta = 1$ and for a sphere $\delta = 0$. The ensemble average of the asphericity is:

$$\langle \delta \rangle = 1 - \left\langle \frac{3(\lambda_x\lambda_y + \lambda_y\lambda_z + \lambda_z\lambda_x)}{(\lambda_x + \lambda_y + \lambda_z)^2} \right\rangle \tag{4.1}$$

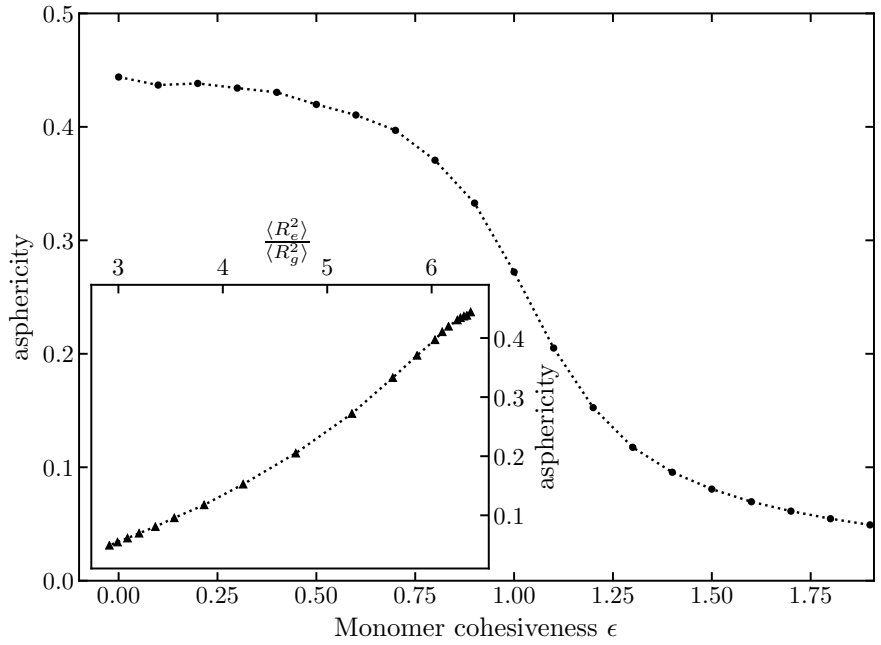
where λ_x, λ_y, and λ_z are the eigenvalues of the 3×3 gyration tensor for a single conformation, whose entries are:

$$S_{xy} = \frac{1}{N}\sum_{i=1}^{N}(R_{i,x} - R_{c,x})(R_{i,y} - R_{c,y}) = \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(R_{i,x} - R_{j,x})(R_{i,y} - R_{j,y}) \tag{4.2}$$

$R_{i,x}$ and $R_{c,x}$ are the x-components of the position of monomer i and the center of mass respectively. The radius of gyration for that conformation is: $R_g^2 = \lambda_x + \lambda_y + \lambda_z$. Although, the ensemble average monomer density will be isotropic for any polymer, the individual conformations may not be, giving a non-zero average asphericity. Figure 4.1b shows the asphericity for the different values of monomer cohesiveness, which decreases from ~ 0.45 for the coil to close to 0 for compact globule conformations.



(a)



(b)

Figure 4.1: (a) Polymer dimensions of a homopolymer for varying monomer cohesiveness. *Inset:* equilibrium polymer size predicted by the mean-field model of Section 2.1.3 with $b = 1.35$, $v_0 \approx 4.52$, and $N = 100$. b was based on the average bond length in the simulations. The estimate of v_0 and the conversion between χ_{cr} and ϵ shown on the top x-axis is explained in Section 2.3.5.
(b) Asphericity of a homopolymer for varying monomer cohesiveness. *Inset:* the ratio of end-to-end distance to radius of gyration for the corresponding asphericity.
The good solvent corresponds to $\epsilon = 0$, the θ solvent occurs at $\epsilon \approx 0.7 - 0.75$ kT, and poor solvents are $\epsilon > 1.5$ kT. The number of monomers is $N = 100$. Polymer dimensions are in units $\sqrt{\frac{2}{3}}b_0$ where b_0 is the 8-6 LJ repulsion distance between bonded monomers described in Section 2.3.4. ϵ is in units of kT.

Coarse-grained simulations can complement experimental measurements by providing the ensemble of conformations of a polymer directly. Inference of the polymer dimensions from FRET experiments requires an appropriate polymer model (see Section 1.4.1). For example, in order to extract the radius of gyration, a model based on the Sanchez theory [56] is sometimes used. Rather than the end-to-end distance, the model only provides a probability distribution for radius of gyration $p(R_g)$ directly. As proposed by Ziv and Haran [57], the end-to-end distance distribution of conformations with a particular radius of gyration $p(R_e|R_g)$ is the probability distribution of distances between two random points in a sphere (see Equation 2.9) of radius $\sqrt{5}R_g$. In Figure 4.2, this assumption is compared to the conditional distributions of $p(R_e|R_g)$ obtained from simulations. The distributions are not noticeably affected by the strength of the cohesive interactions ϵ. The distribution of the distance between two random points in a sphere matches the simulations for compact conformations, which usually have high ϵ, but underestimates the end-to-end distance for large conformations, which have low ϵ. In order to match the raw experimental data, assuming the distribution of random points in a sphere for $p(R_e|R_g)$ would tend to overestimate the radius of gyration for polymers with low cohesiveness or in good solvents.

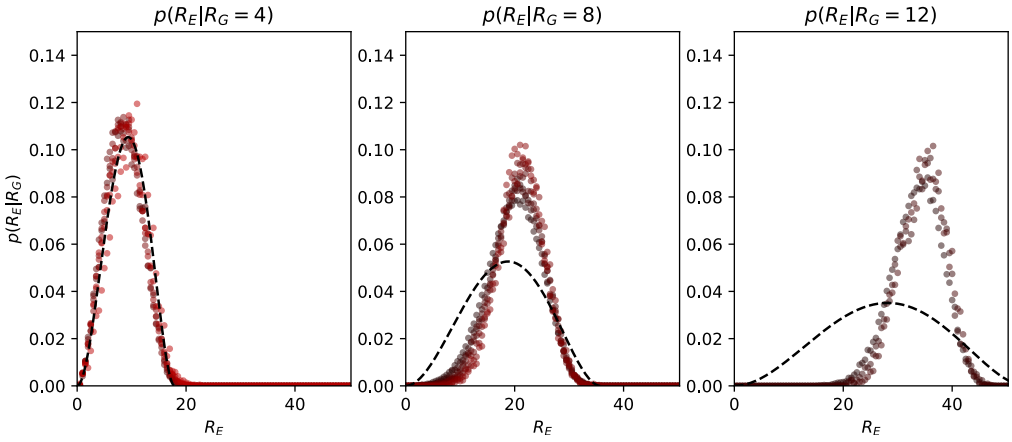


Figure 4.2: Probability distributions of the end-to-end distance, conditioned on the radii of gyration of the individual conformations (not ensemble average in both cases). The circles indicate simulation results with the amount of grey and red representing low to high ϵ. The bin size for simulation data was 0.5, so the probability at $R_E = 20$ for $P(R_E|R_G = 8)$ in the figure means the probability that $20 \leq R_e < 20.5$ for conformations with $8.0 \leq R_g < 8.5$. The $p(R_E|R_G)$ for each ϵ and R_G was plotted only if the number of conformations used to calculate it exceeded 3000. The black dashed line shows the distribution of the end-to-end distance if the ends were at two points chosen randomly within a sphere of radius $\sqrt{5}R_g$ (see Equation 2.9). The number of monomers is $N = 100$. Polymer dimensions are in units $\sqrt{\frac{2}{3}}b_0$ where b_0 is the 8-6 LJ repulsion distance between bonded monomers described in Section 2.3.4.

4.2.3 Relative Dimensions of a Homopolymer

We can more readily identify a polymer's position on the order-disorder continuum by studying the ratios between the various polymer dimensions rather than the quantities themselves. From simple models we know the universal values of these ratios at the extremes (good, θ, and poor solvents) and as the number of monomers $N \to \infty$. These values are calculated in Chapter 2. R_e^2/R_g^2 is 6 for an ideal chain or at the θ point, when excluded volume interactions are balanced by attractive forces; this ratio is slightly

higher than 6 for a good solvent and decreases for more compact conformations. In the compact regime (high cohesiveness), the polymer can be approximated as a uniformly dense sphere. If the positions of the ends are independent and uniformly distributed inside the sphere, $R_e^2/R_g^2 = 2$. R_g/R_h approaches ~ 1.5 in a good solvent for $N \to \infty$ and decreases to ~ 0.774 when the polymer can be approximated as a uniform sphere. The Kirkwood approximation predicts a different lower limit of $R_g/R_k \sim 0.93$.

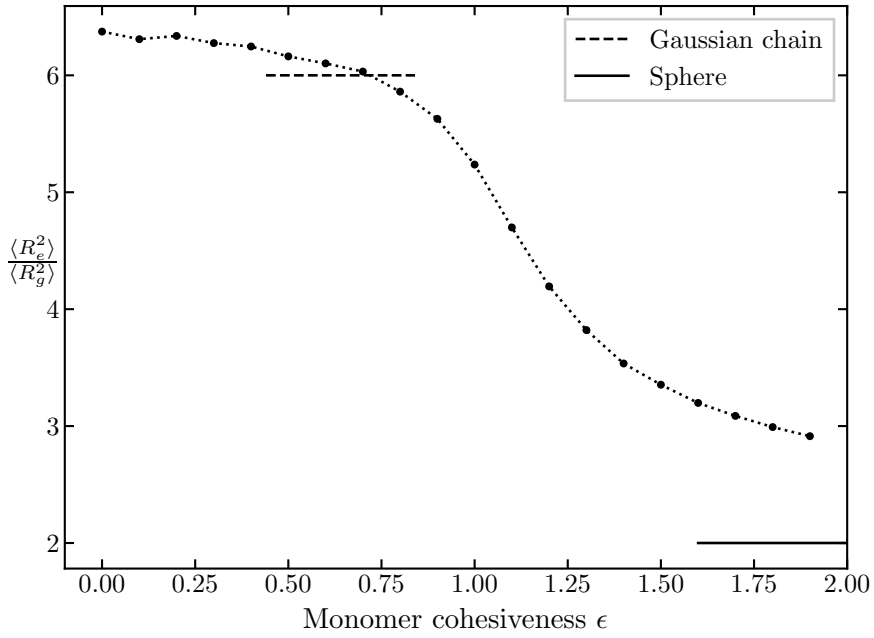
Figure 4.3 shows the ratios of the square end-to-end distance to the square radius of gyration and radius of gyration to hydrodynamic radius for varying values of polymer cohesiveness ϵ. The ratios obtained from simulations approach the theoretical limits, and the two are expected to coincide as $N \to \infty$ [185] and for higher ϵ in the case of the poor solvent. Specifically, the ratio of end-to-end distance to radius of gyration in simulations agrees with the Gaussian chain prediction at the θ point. For the homopolymer model, the asphericity is well correlated with the ratio of end-to-end distance to radius of gyration (*Inset* of Fig. 4.1b).

There has been a discrepancy between the polymer dimensions of IDPs and chemically denatured proteins measured using different experimental techniques, most prominently FRET and SAXS [60, 61]. In FRET, assumptions about the distribution of the end-to-end distance are needed to infer the ensemble average value. Further assumptions about the polymer model are made to infer the radius of gyration, which has led to the discrepancy between the radii of gyration obtained from FRET and SAXS experiments. Song et al. [58] have emphasized that conventional models used in FRET inference overestimate the radius of gyration and proposed a subensemble method which finds the most probable radius of gyration which is consistent with the experimental FRET efficiency. Borgia et al. [63] and Fuertes et al. [36] have shown that end-to-end distance and radius of gyration values are decoupled: one can reweight the conformations to create many ensembles with similar radii of gyration but drastically different end-to-end distances. The initial discrepancies between FRET and SAXS radii of gyration have been attributed to the assumptions of the hompolymer models (usually Gaussian chain or Sanchez theory) used in FRET inference. The true ensembles of IDPs are heterogeneous and aspherical. In this section, we found that a homopolymer model can predict aspherical conformations and differing ratios for the polymer dimensions although both are still correlated with ϵ and chain compaction. In the next section (4.3), we investigate to what degree the variation of the polymer dimensions is due to the heterogeneity of the sequence versus overall chain cohesiveness and compaction.

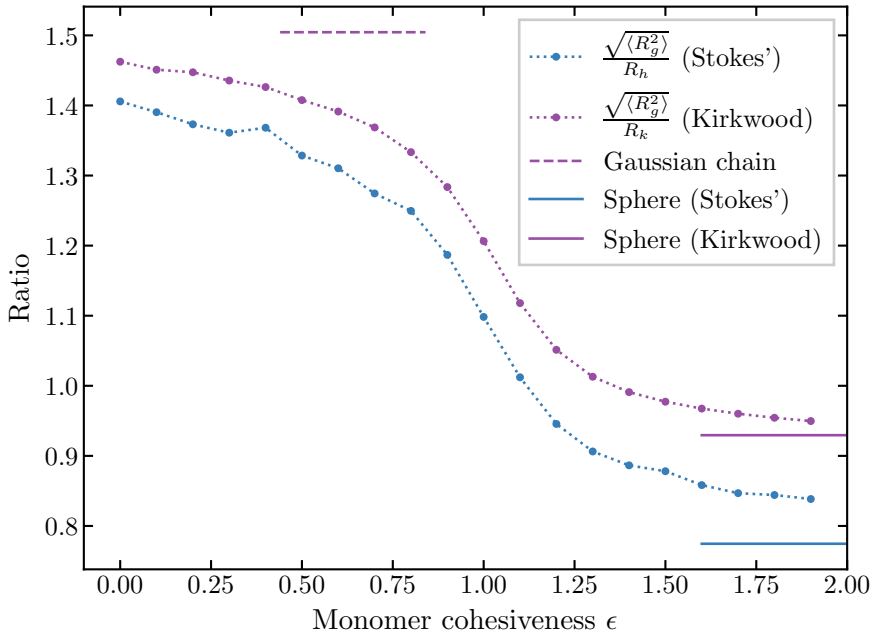
4.2.4 Hydrodynamic Radius and the Kirkwood Approximation

Accurate simulations of dynamics are required to measure the hydrodynamic radius or long time limit of the diffusion coefficient. The Kirkwood approximation gives the short time limit of the hydrodynamic radius and diffusion coefficient, and can be obtained from the equilibrium ensemble of conformations without requiring dynamics. The ensembles can be generated from Monte Carlo or Brownian dynamics simulations. By performing Brownian dynamics simulations with the addition of implicit hydrodynamic interactions, Liu et al. [30] and Schmidt et al. [110] have previously found that the Kirkwood approximation overestimates the hydrodynamic radius by $< 4\%$ for a SAW and a worm-like chain model. In this section, we extend the comparison between the Kirkwood approximation and the hydrodynamic radius to all values of cohesiveness.

From the simulations, we calculate a time dependent diffusion coefficient by varying the time interval over which we measure the centre of mass displacement of the polymer. In the limit of short times this time dependent diffusion coefficient approaches the Kirkwood approximation (Eq. 1.6). As the interval



(a)



(b)

Figure 4.3: Ratios between polymer dimensions of a homopolymer for varying monomer cohesiveness. *(a)* Square end-to-end distance to square radius of gyration. *(b)* Radius of gyration to hydrodynamic radius. The Kirkwood approximation is shown in purple and the long time limit is shown in blue. The dashed lines correspond to the Gaussian chain predictions, the solid lines correspond to a uniform sphere. The ratio of square end-to-end distance to square radius of gyration agrees with the Gaussian chain prediction ($R_e^2/R_g^2 = 6$) at the θ point ($\epsilon \approx 0.7 - 0.75$ kT). The number of monomers is $N = 100$. ϵ is in units of kT.

is increased and the polymer samples more conformations as it diffuses, the diffusion coefficient relaxes to its long time limit (Eq. 1.5 and Fig. 4.4a).

As the polymer cohesiveness increases and the polymer adopts more compact configurations, the diffusion coefficient increases as expected. However, the relative difference between the short and long time limits of the diffusion coefficients also increases. If we convert to the hydrodynamic radius we can see that for low cohesiveness, ranging from good to θ solvents, the Kirkwood approximation overestimates the true diffusion coefficient by 3-5% in agreement with other studies [30, 110]. In the poor solvent regime the relative difference increases to beyond 10% and is larger for longer polymers (Fig. 4.4b).

4.3 Effects of Amino Acid Sequence Heterogeneity on IDP Dimensions

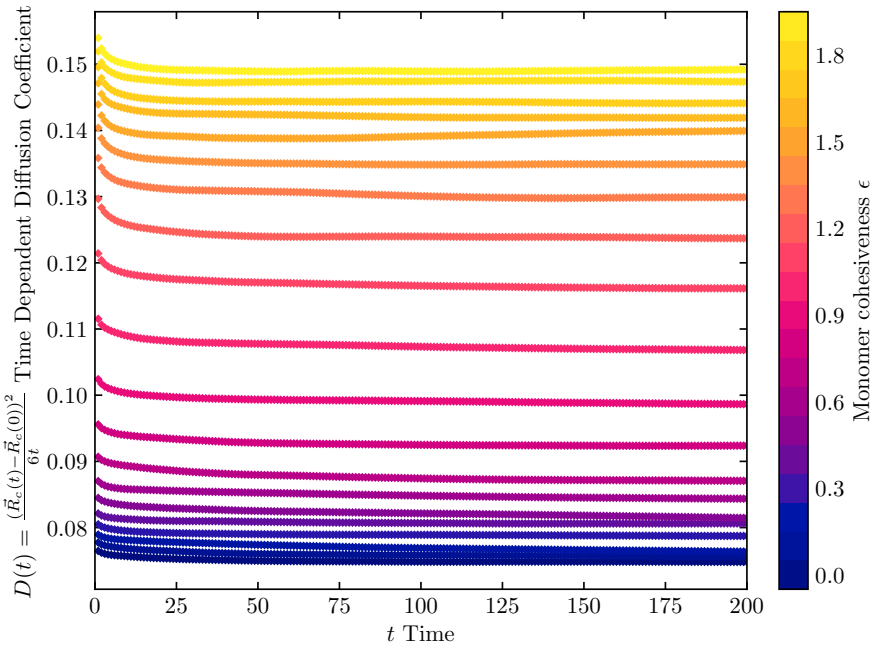
4.3.1 Sequence Heterogeneity Model

To capture the importance of sequence effects on IDP structures, the homopolymer model can be augmented by ascribing different properties to the monomers resulting in heterogeneous sequences. These can include: monomer size (LJ diameters representing the range of the excluded volume and cohesive interactions), variable strengths of the cohesive interactions, and variable charges and electrostatic interactions between the monomers. The parameters were originally introduced in Section 2.3.

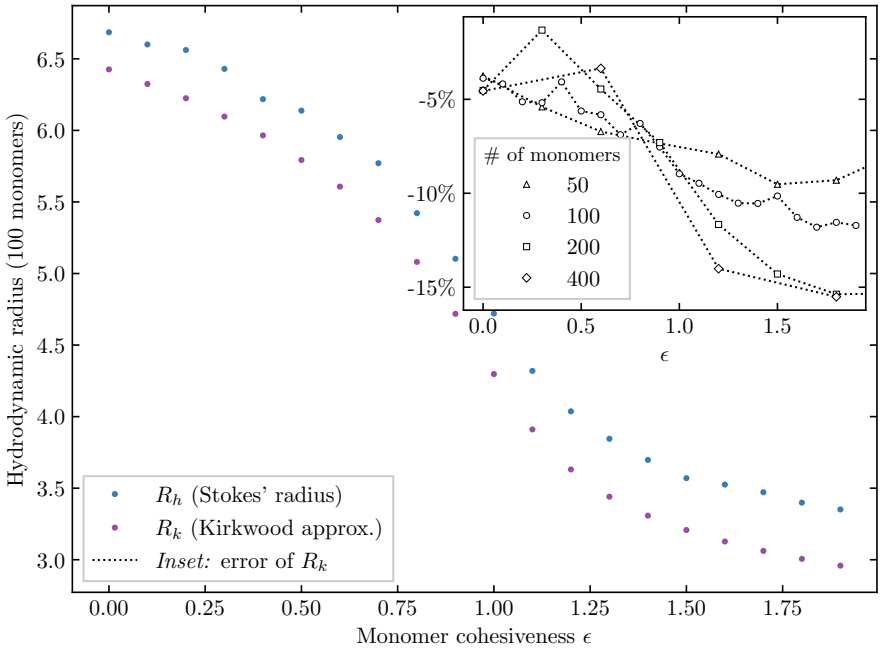
Rather than attempt to map the amino acids to specific coarse-grained model parameters, we focus on discovering universal relationships between sequence and polymer dimensions. We introduce a four letter model ("HP+-"), where monomers can be either: neutral/disorder promoting ("P"), cohesive/order promoting ("H"), positively charged ("+"), or negatively charged ("-"). The first two types of monomers are inspired by the Hydrophobic-Polar model of proteins [186]. Overall, this model takes into account the basic features of IDP sequences: that polymer dimensions are correlated with order and disorder promoting amino acids [38], that net charge and hydrophobicity can distinguish IDPs from natively folded proteins [1], and the importance of the sequence of charged residues on polymer dimensions [50]. The charged monomers directly represent charged amino acids, while the cohesive monomers represent order promoting or hydrophobic amino acids, and the neutral monomers are polar or disorder promoting amino acids. Neutral monomers do not interact via cohesive or electrostatic interactions ($\epsilon_i = 0$ and $q_i = 0$). Cohesive monomers interact identically only with other cohesive monomers via the cohesive interaction (overall strength $E = \frac{\epsilon}{kT}$). Charged monomers interact only with other charged monomers via the electrostatic interactions.

In simulations, the range (diameters) of excluded volume interactions between bonded monomers was $B_0 = \sqrt{\frac{3}{2}}$. The maximal extension of the FENE bonds between monomers was $L_{max} = 2B_0$. The strength of excluded volume interactions was $E_{LJ} = \frac{\epsilon_{LJ}}{kT} = 1$. For excluded volume and cohesive interactions between non-bonded monomers, the range (diameters) B_i varied for each specific simulation. The diameters were originally inspired by the volumes of real amino acids [100, 174], however the results can be applied more generally and we expect the small variation in radii to not have had a great effect. The hydrodynamic radii of the monomers were $A_i = \frac{B_i}{2}$. Cohesive forces were cutoff at a distance of $4B$. The strength of electrostatic interactions $Q = 2$ and the Debye length was $L_D = 4$ in simulation units (see Section 2.3.6 for an explanation of these parameters).

The simulations can be directly compared to real polypeptides in the following way: the distance

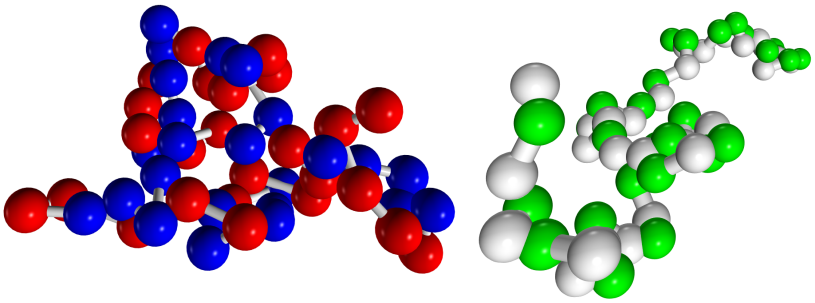


(a)



(b)

Figure 4.4: (a) Time Dependence of Diffusion coefficient. Homopolymer model with $N = 100$.
(b) Error of the Kirkwood Approximation. Homopolymer model with $N = 100$. *Inset:* relative difference between the Kirkwood approximation and the long limit of the hydrodynamic radius for polymers with 50, 100, 200 and 400 monomers.
ϵ is in units of kT. Distance and time are in simulation units as defined in Section 2.3.2.



(a) (b)

Figure 4.5: Visualizations of example conformations of heterogeneous sequences.
(a) Sequence "sv19" (shown in Fig. 4.6).
(b) The $(HP)_{30}$ sequence from Section 4.3.3.
Each monomer is represented by a sphere of diameter $0.7B_i$. White, green, red and blue colours represent neutral, cohesive, positively and negatively charged monomers respectively. Bonded monomers are connected using a cylinder.

between two Cα atoms of adjacent amino acids is roughly 0.38 nm [100], this corresponds to the distance between bonded monomers in simulations which was ~ 1.35. The Debye length corresponds to ~ 1.1 nm, which is equivalent to ~ 75 mM of NaCl. However, the goal of this section is not to create a mapping between amino acids and the coarse-grained model parameters, but instead investigate what degree of detail is important when modeling IDPs and investigate the general relationships between sequence and IDP dimensions.

For each sequence and interaction parameters ϵ, Q, and L_D, 8 runs were performed, each lasting 10^8 steps, with a time step of $\Delta T = 0.001$. Each run began with a self-avoiding random walk initial condition. The first 10^6 steps were excluded from the analysis, and averages were taken over steps over runs and steps.

4.3.2 Sequences of Charged Residues

Many IDPs contain higher fractions of charged amino acids in their sequences, when compared with natively folded proteins [23, 27]. Das and Pappu [50] introduced a patterning parameter κ to quantify the segregation of positive and negative charges in a sequence. This value is low for well mixed sequences and high for completely segregated sequences, and is further normalized by a hypothetical maximally segregated sequence where all of the positive charges lie at one end and the negative charges at the opposite end. From simulations using the ABSINTH model, they found that the radius of gyration decreased for increasing κ for sequences of equal numbers of positively and negatively charged amino acids. In the ABSINTH model, ensembles are generated using the Monte Carlo method. All atoms of the amino acids and ions in solution are modeled explicitly, while the solvent molecules are included implicitly [187].

The definition of the sequencing parameter is inspired by the charge asymmetry σ of a polyelectrolytic sequence:

$$\sigma = \frac{(f_+ - f_-)^2}{f_+ + f_-} \tag{4.3}$$

f_+ and f_- are the fractions of positively and negatively charged amino acids respectively. For the se-

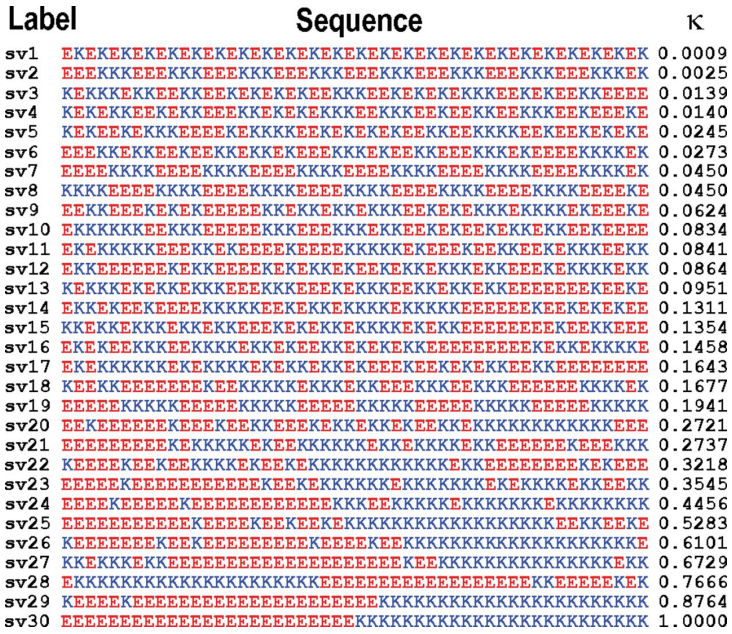


Figure 4.6: IDPs with identical compositions but differing sequences. Each is composed of identical numbers of positively ("K") and negatively ("E") negatively charged amino acids, and the sequence is quantified by the charge segregation parameter κ. Figure from Das and Pappu [50].

quences of equal numbers of positively and negatively charged amino acids studied by Das and Pappu [50]: $\sigma = 0$. To differentiate between sequences with identical compositions, the local charge asymmetry σ_i of a consecutive group i of amino acids is introduced:

$$\sigma_i = \frac{(f_+ - f_-)_i^2}{(f_+ + f_-)_i} \tag{4.4}$$

σ_i is typically calculated for every group of $g = 5$ or 6 consecutive amino acids in the sequence, including overlaps. The average square deviation of the local charge asymmetry from the global value quantifies the segregation of positive and negative charges:

$$\delta = \frac{\sum_{i=1}^{N_g} (\sigma_i - \sigma)^2}{N_g} \tag{4.5}$$

Where $N_g = N + 1 - g$ is the number of groups for which σ_i is calculated. The sequence parameter $\kappa = \frac{\delta}{\delta_{max}}$ is the average square deviation of local charge asymmetry (δ), normalized by the value δ_{max} for a maximally charge segregated sequence with the same composition (all the positive charges placed at the C-terminus and all the negative charges at the N-terminus). κ can also be thought of as roughly quantifying the average net charge of groups of 5 or 6 consecutive of amino acids or the average size of contiguous like-charged patches.

In order to evaluate the coarse-grained model relative to atomistic simulations, we performed simulations of 30 sequences with κ values ranging from 0 to 1 of an IDP composed of 25 negatively charged and 25 positively charged monomers. The sequences are shown in Figure 4.6. Ignoring atomistic details, we only consider the screened Coulomb interactions between the charges of the monomers, the bonds

between consecutive monomers along the chain, and excluded volume interactions through the LJ diameters which are set to $B_- = 2.29$ and $B_+ = 2.44$ reflecting the relative volumes of the corresponding amino acids ("E" and "K"). The charged interaction strength $Q = 2$ and the Debye length is $L_D = 4$.

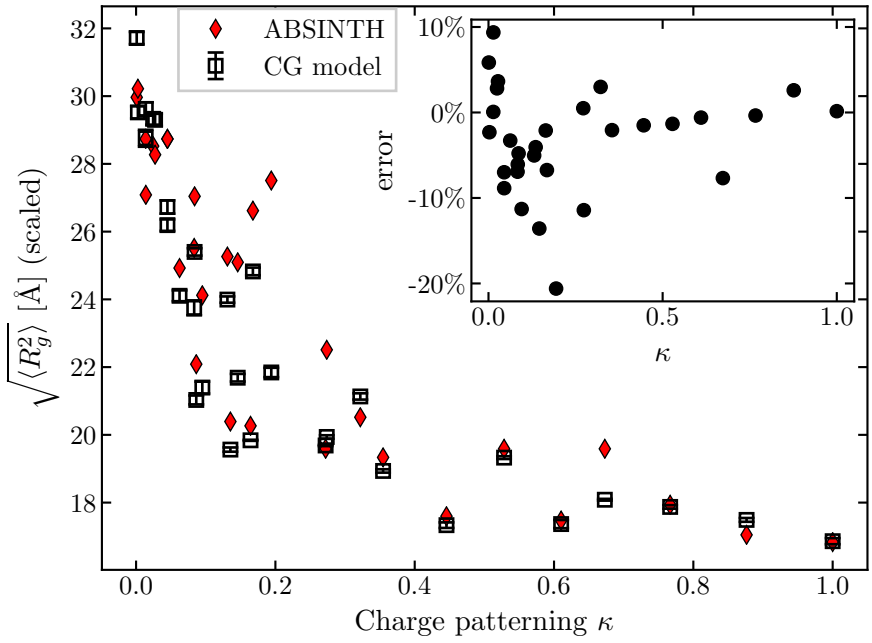
We reproduce the general trend of decreasing radius of gyration with increasing κ (Fig. 4.7a), as well as the partially reproducing the undulations within this dependence. In order to minimize the percent difference (Inset of Fig. 4.7) between our results and those of Das and Pappu [50], our radii of gyration are rescaled by a factor of ~ 1.4 (the ratio of the average radii of gyration over all sequences between our results and those of Das and Pappu). This difference could be due to several assumptions of the coarse grained model: the bond angle restrictions between subsequent amino acids are neglected, amino acids are treated as spherically symmetric and side-chain geometries are ignored, and the amino acid size in the 8-6 LJ repulsive potential is based on the volumes of amino acids from folded proteins, which could differ from the excluded volume of amino acids in IDPs [100, 174]. The agreement between the two models is better for sequences with higher κ, and for low κ there are a few sequences for which the coarse-grained model predicts greater compaction. The sequence with the highest disagreement is "sv19" (shown in Fig. 4.6) with $\kappa = 0.1941$. This sequence is made up of repeating motifs of 5 negative followed by 5 positive amino acids. It is likely that the simple bead geometries and lack of bond angle restrictions in the CG model, results in more compact conformations specifically for this sequence.

In Figure 4.7b, the results are reordered according to the Sequence Charge Decoration (SCD) parameter introduced by Sawle and Ghosh [188]. Lin and Chan [189] found that the radius of gyration simulated by Das and Pappu had a smoother dependence on SCD than on κ. For the coarse-grained model results, we find a similar improvement in the correlation of radius of gyration with SCD over κ.

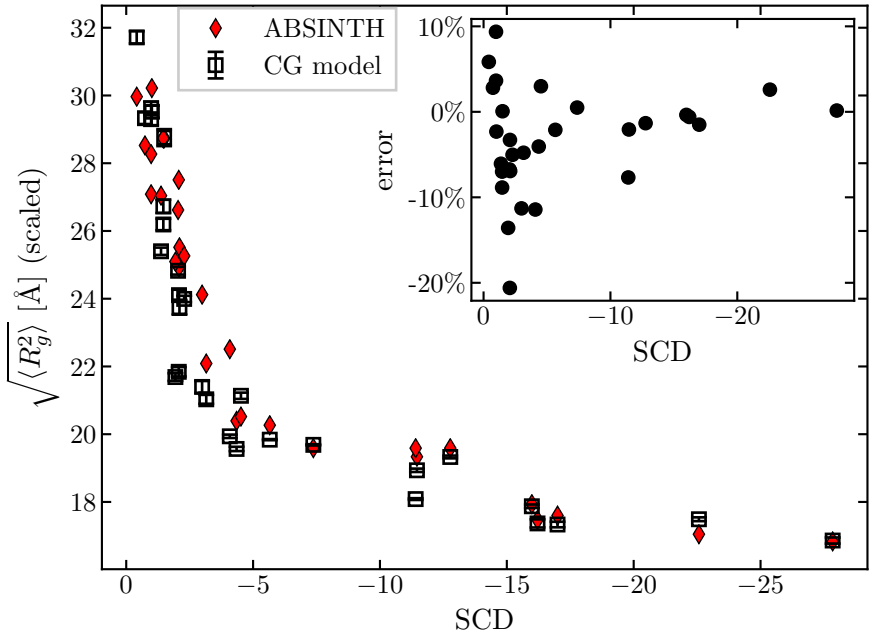
For the homopolymer, the dimensions were controlled by the cohesiveness parameter ϵ. Without varying any of the interactions, and only the sequence, we observed a similar compaction of the radius of gyration, now correlated with the parameters κ and SCD. Using a random-phase-approximation polymer theory, Lin and Chan [189] found that SCD and R_g were both correlated with the critical temperature describing the tendency of an IDP to phase separate, which provides a connection to the χ_{cr} parameter of mean-field models. Thus, the effect of modifying the sequence of charged monomers quantified by the parameters κ or SCD on the radius of gyration and phase separation behaviour of IDPs is analogous to adjusting the global cohesiveness parameter of a homopolymer. Figure 4.8 shows the ratios of the polymer dimensions for the different sequences. For the homopolymer, the dimensions were decoupled, but the ratios still varied with overall compaction controlled by the cohesiveness parameter ϵ. For the polyampholytic sequences, the ratio of end-to-end distance to radius of gyration is very sensitive to the specific sequence. A FRET measurement will be more indicative of local structure of the polymer ends, and cannot be used to infer the other polymer dimensions. On the other hand, the ratio of radius of gyration to the hydrodynamic radius is correlated with κ or SCD and overall compaction, and determines the position of the ensemble on the disorder-to-order continuum.

4.3.3 Effect of Length of Cohesive Patches

In this section, I investigate the patterning of neutral (disorder promoting) and cohesive (order promoting) monomers, specifically the length of continuous cohesive patches. Martin et al. [51] augmented the charge patterning parameter κ of Das and Pappu [50] to incorporate all amino acids. Since, in general, Proline and charged amino acids are disorder promoting, their positions in a sequence are expected to influence the dimensions of an IDP. They introduced the parameter Ω, defined analogously to κ, except



(a)



(b)

Figure 4.7: Radius of gyration of sequences composed of 25 positively and 25 negatively charged amino acids. *Inset:* percent error of the coarse-grained model (relative to ABSINTH model [50]). (a) The κ parameter introduced by Das and Pappu [50] is on the x-axis. (b) Sequence Charge Decoration (SCD) introduced by Sawle and Ghosh [188] is on the x-axis.

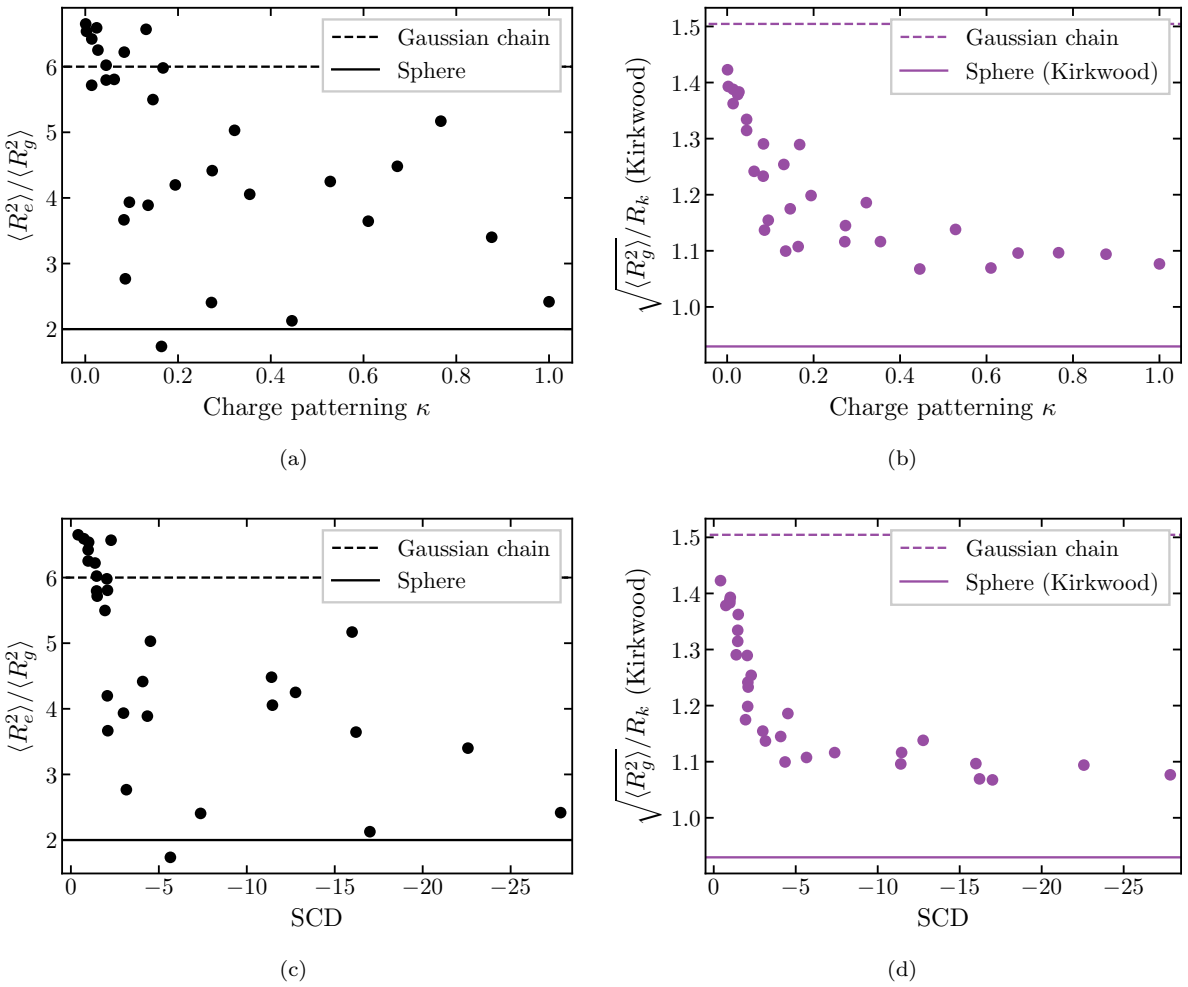


Figure 4.8: Ratios between polymer dimensions of polyampholyte sequences. *(a)* Square end-to-end distance to square radius of gyration. *(b)* Radius of gyration to hydrodynamic radius (Kirkwood approximation). The dashed lines correspond to the Gaussian chain predictions, the solid lines correspond to a uniform sphere. *Top:* The κ parameter introduced by Das and Pappu [50] is on the x-axis. *Bottom:* Sequence Charge Decoration (SCD) introduced by Sawle and Ghosh [188] is on the x-axis.

$f_{P,+,-}$, the combined fraction of Proline and charged residues, takes the place of f_+ and f_{others}, the fraction of all other amino acids, takes the place of f_i. Using ABSINTH simulations, they found that Ω was correlated with compaction, meaning sequences where disorder promoting amino acids were well spread out, were more expanded.

We simulated 5 different sequences of 30 cohesive ("H") and 30 neutral ("P") monomers using the coarse-grained model. The LJ diameters of the beads were set to $\sqrt{1.5}$ for all monomers. The sequences, shown in Table 4.1, vary in the size of the cohesive and neutral clusters increasing from 1 to 5. The values of Ω increase as the consecutive cluster size increases and the order and disorder promoting amino acids become more and more segregated.

Ω	Sequence
0.00086	$(HP)_{30}$
0.00414	$(PHHP)_{15}$
0.00086	$(PHHHPP)_{10}$
0.04056	$PP(HHHHPPPP)_7HH$
0.17943	$(PPHHHHHPPP)_6$

Table 4.1: Sequences of cohesive ("H") and neutral ("P") monomers with different sizes (1, 2, 3, 4, or 5) of cohesive (and neutral) clusters.

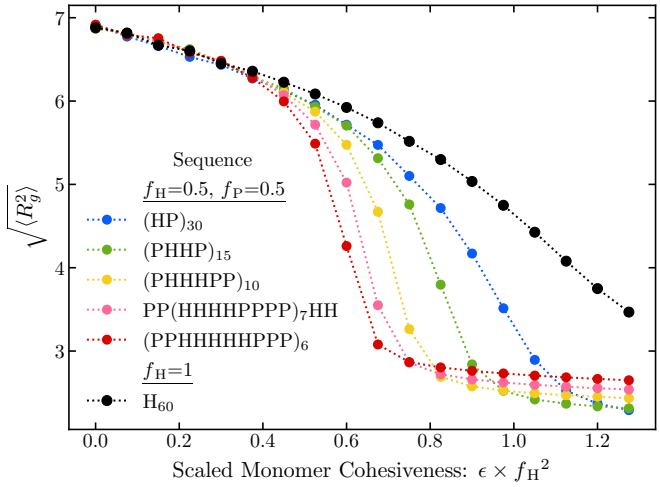


Figure 4.9: Radius of gyration of sequences composed of 30 cohesive and 30 neutral monomers for varying monomer cohesiveness. The size of the hydrophobic patches varies from 1 to 5 and the exact sequence is shown in the legend. The subscript indicates the number of times that motif is repeated. For comparison, a homopolymer sequence of 60 cohesive monomers is shown in black. On the x-axis, the monomer cohesiveness has been rescaled according to the mean-field assumption. f_H refers to the fraction of cohesive monomers in the sequence. ϵ is in units of kT. Radius of gyration is in units $\sqrt{\frac{2}{3}}b_0$ where b_0 is the 8-6 LJ repulsion distance between bonded monomers described in Section 2.3.4.

We varied the cohesiveness ϵ of the "H" monomers and measured the effects of the overall cohesiveness and the size of cohesive clusters on polymer dimensions. At low cohesiveness, the radii of gyration of the different sequences converged to the same value (Fig. 4.9). The radii of gyration for a homopolymer of 60 cohesive monomers is shown for comparison. On the x-axis, the monomer cohesiveness parameter

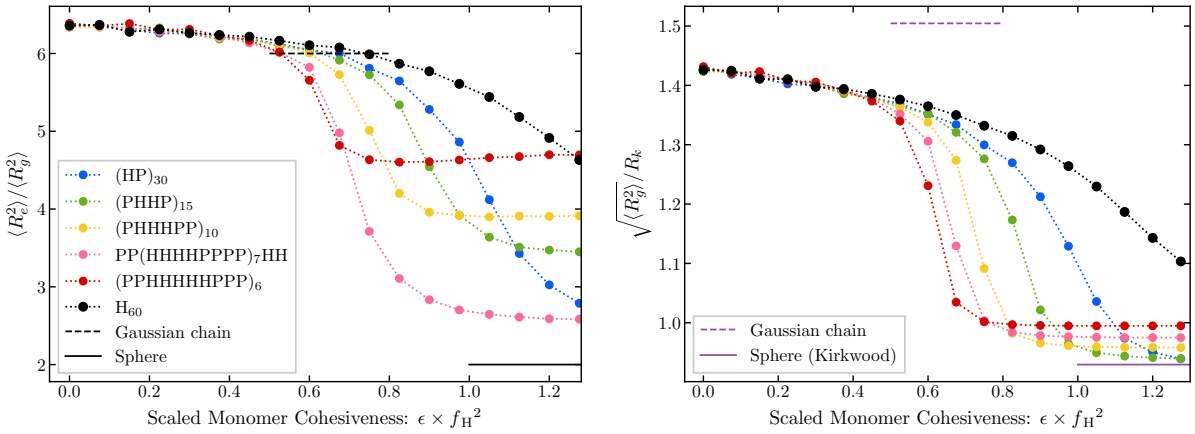


Figure 4.10: Ratios between polymer dimensions of sequences composed of 30 cohesive and 30 neutral monomers for varying monomer cohesiveness. (a) Square end-to-end distance to square radius of gyration. (b) Radius of gyration to hydrodynamic radius (Kirkwood approximation). The dashed lines correspond to the Gaussian chain predictions, the solid lines correspond to a uniform sphere. For comparison, a homopolymer sequence of 60 cohesive monomers is shown in black. On the x-axis, the monomer cohesiveness has been rescaled according to the mean-field assumption. f_H refers to the fraction of cohesive monomers in the sequence. ϵ is in units of kT.

ϵ is rescaled by the square fraction of cohesive monomers, according to the mean-field assumption under which the polymer dimensions would be equivalent. The correspondence with the homopolymer begins to break down around the θ-point ($\epsilon \approx 0.7 - 0.75$ kT). For intermediate cohesiveness, the sequences with larger cluster sizes exhibited an earlier and steeper coil to globule transition. At the highest values of cohesiveness, there was a non-monotonic relationship between the size of the clusters and the polymer dimensions. This is likely due to the sequences close to the ends having an exaggerated effect due to the short sequence lengths. This behaviour is more pronounced when we look at the ratio of the end-to-end distance to the radius of gyration compared with the ratio of radius of gyration to hydrodynamic radius (Fig. 4.10). For the former, the ratios vary greatly even for collapsed sequences, and the length of the neutral tails at the end of each sequence likely have a greater effect on the end-to-end distance than the other dimensions. On the other hand, the ratio of radius of gyration to hydrodynamic ratio follows the overall compaction of the polymer. These results emphasize again how end-to-end distance how discrepancies may arise when inferring the other polymer dimensions from the end-to-end distance.

4.3.4 Patterning of Charged and Cohesive Monomers

Mao et al. [48] proposed a classification of the conformational ensembles of IDPs by composition based on their mean hydrophobicity, and fractions of positively and negatively charged amino acids. Polyelectrolytic sequences are abundant in either positively or negatively charged amino acids, resulting in a high fraction of charged residues and a high net charge per residue, and are expected to adopt swollen coil structures. In this section, we investigate the effects of sequence patterning of cohesive and charged amino acids on the dimensions of polyelectrolytic IDPs [190]. Beginning with the sequences composed of positively and negatively charged amino acids used by Das and Pappu [50] that looked at the effects

of charge segregation (shown in Figure 4.6), the negatively charged amino acids are replaced with cohesive amino acids. The results apply identically to sequences of negative and cohesive monomers. The patterning parameter Ω proposed by Martin et al. [51] describes the segregation of disorder promoting amino acids (Proline and charged amino acids) and order promoting amino acids (all others), and in this case the values of Ω of the charged/cohesive sequences are equivalent to the values of κ for the corresponding original polyampholytic sequences. Low Ω for these sequences corresponds to good mixing of the positively charged and cohesive residues with the lowest value given to the sequence composed of the "H+" motif repeated. The maximum value of $\Omega = 1$ is assigned to the sequence with maximum segregation: 25 positively charged amino acids followed by 25 cohesive. Martin et al. [51] considered permutations of the sequence of Ash1, which had a combined total of 35% Proline and charged amino acids ($\approx$ 19% positively charged and $\approx$ 1.2% negatively charged). They observed a monotonic decrease of radius of gyration with increasing Ω. For comparison, in our sequences 50% of the residues are all positively charged and the rest are order promoting (cohesive). We varied the monomer cohesiveness, ϵ from 0 to 5.1, in intervals of 0.3. This is similar to starting from all disorder promoting amino acids in the uncharged regions, and increasing the fraction of order promoting or cohesive amino acids in those regions.

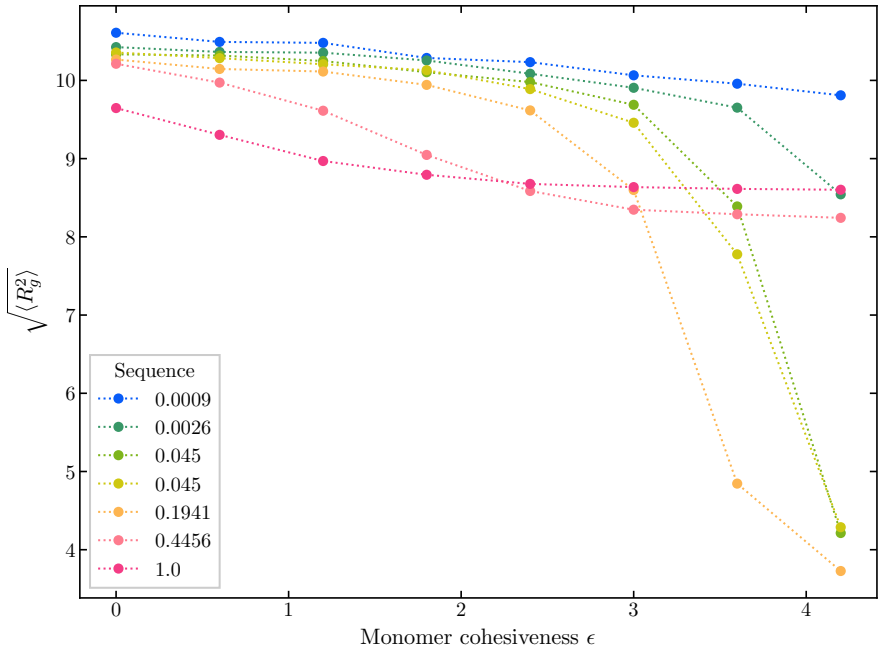
For well mixed sequences, repulsive interactions dominate, swelling their conformations. As ϵ, the strength of attraction between cohesive monomers, increases we observe an initial decrease in the radius of gyration that is more pronounced for sequences with higher Ω. This is expected as higher segregation means larger size of cohesive patches which resulted in an earlier coil to globule transition as shown in the previous section. However, as ϵ increases further, the sequences with the highest values of Ω fail to collapse, and the smallest radii of gyration are observed for intermediate sequences (Fig. 4.11a) This can be explained by the long positively charged regions within sequences with high Ω resisting the global compaction and overpowering the cohesive monomers found within them.

Figure 4.11b shows the non-monotonic relationship between radius of gyration and the sequence parameter Ω more clearly. At sufficiently high ϵ, as we increase the segregation between charged and hydrophobic residues we initially observe a decrease in radius of gyration, as we increase the segregation parameter further, we observe a recovery in the polymer size. As before, the ratio of end-to-end distance to radius of gyration is highly sequence specific, and there is no trend between it and Ω or the radius of gyration.

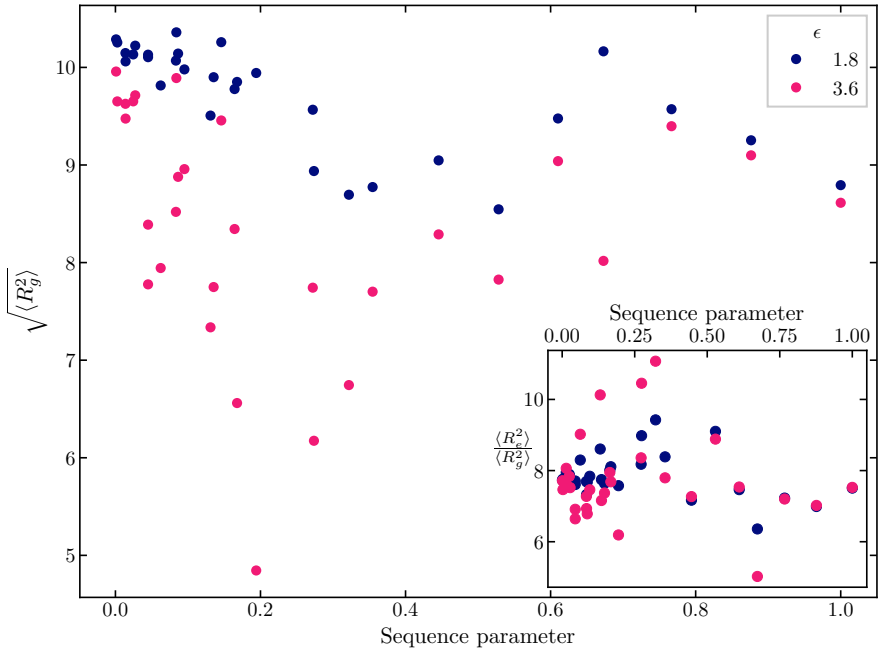
4.4 Conclusions

In Section 4.2, we investigated the dimensions of a homopolymer for varying cohesiveness using coarse-grained simulations. An increase in cohesiveness resulted in a compaction of all the polymer dimensions (end-to-end distance, radius of gyration, and hydrodynamic radius). The compaction factor differs for each of the polymer dimensions and their relative values (the ratio of the square end-to-end distance to the square radius of gyration and the ratio of the radius of gyration to the hydrodynamic radius) also vary with cohesiveness. Additionally, we found that the conformations of the homopolymer are aspherical for low values of ϵ and the ratio of end-to-end distance to radius of gyration was correlated with asphericity. For the homopolymer model, the dimensions are decoupled in the sense that they vary with ϵ, however they are still correlated with the cohesiveness and compaction of the polymer.

In Section 4.3, we investigated the role of sequence patterning by introducing four types of monomers



(a)



(b)

Figure 4.11: Radius of gyration of sequences composed of 25 cohesive and 25 positively charged monomers. The sequence parameter Ω describes the segregation of the cohesive and charged monomers. (a) Monomer cohesiveness ϵ is on the x-axis and the results for 7 sequences with varying Ω are shown. (b) The sequence parameter Ω is on the x-axis and the results for two values of the strength of cohesive interactions ϵ are shown. *Inset:* the ratio of square end-to-end distance to radius of gyration. ϵ is in units of kT. Radius of gyration is in units $\sqrt{\frac{2}{3}}b_0$ where b_0 is the 8-6 LJ repulsion distance between bonded monomers described in Section 2.3.4.

(neutral, cohesive, positively charged, and negatively charged). In general, we found that sequence heterogeneity can modulate the polymer dimensions independently of composition or the overall cohesiveness. The sequences studied included polyampholytes composed of positively and negatively monomers described by the κ parameter [50], sequences composed of cohesive and neutral monomers with differing cluster size, and polyelectrolytes composed of positively charged and cohesive monomers described by the Ω parameter [51]. The dimensions of the polyampholytes were similar to those predicted by an all-atom model (within 5% error for over half of the sequences). For polymers composed of cohesive and neutral monomers, an increase in the size of cohesive patches resulted in compaction of the polymer dimensions at a weaker overall cohesiveness value. The dimensions of chains composed of charged and cohesive monomers had a non-monotonic relationship with the sequence parameter describing the segregation of the two types of monomers: well mixed or mostly segregated sequences adopted expanded conformations but those in between the two extremes tended to be more compacted. Overall, the ratio of radius of gyration and hydrodynamic radius was correlated with sequence patterning parameters and is a good measure of global conformational features of an IDP. On the other hand, end-to-end distance was highly sequence specific and the ratio of end-to-end distance to radius of gyration can be used to describe local conformational features of the IDP's ends. Future investigations could include sequences composed of three or more of the monomer types of the four letter model (e.g. cohesive, positive, and negative) requiring both sequence parameters. Also, the relationships between polymer dimensions and sequence parameters almost certainly depend on the composition (fraction of cohesive, neutral, positive, and negative monomers), and future investigations should look at what these relationships are for different compositional classes of IDPs.

After scaling the radius of gyration by a factor of ~ 1.4, the coarse-grained model was in good agreement with the all atom ABSINTH model for sequences of positively and negatively charged amino acids [50]. Based on the sequence whose radius of gyration differed the most and the overall conversion factor, neglecting backbone geometry in the coarse grained model is the reason that a direct quantitative comparison with IDP dimensions impossible. Further development of the model could include the addition of the backbone geometry information in the form of bond angle potentials since this would not perceptibly increase the runtime of the simulations. Validation could include more comprehensive matching to all atom simulations or experimental results.

The results of this Chapter might be able to clarify the discrepancies arising between the FRET and SAXS inferred radii of gyration. The models commonly used to infer the radius of gyration from FRET measurements, such as the Gaussian chain and Sanchez models [22, 53, 57], assume a constant ratio between the end-to-end distance and the radius of gyration. The homopolymer model showed that this ratio varied with monomer cohesiveness. Furthermore, the four letter model showed that the end-to-end distance and subsequently the the ratio of end-to-end distance to radius of gyration is highly sequence dependent and not correlated with overall compaction, and therefore cannot reliably be used to infer the other dimensions. Our coarse grained model offers an improvement over the simple models commonly used in FRET analysis without resorting to all atom simulations as it gives the probability distributions of the individual polymer dimensions. Rather than being the source of a discrepancy, the combined measurements of several polymer dimensions can be used to describe the conformational ensemble, and guide theoretical and computational models of IDPs. This is analagous to the use of the heights of the pure FG nup layers at various grafting distances to infer the cohesiveness of the polymers in the brush model of Chapter 3. For example, the ratio between radius of gyration and hydrodynamic radius

can reveal the location of a particular IDP on the disorder-to-order continuum and the cohesiveness parameter of the IDP, while the ratio of end-to-end distance to radius of gyration may reveal the local conformation of the IDP's ends and whether or not a homopolymer model is appropriate.

Chapter 5

Polymer Dynamics of IDPs

5.1 Introduction

In the previous two chapters, we have studied the equilibrium ensembles of conformations of IDPs: specifically the heights of surface grafted FG nup layers and transport proteins in Chapter 3 and the polymer dimensions of single IDPs and and their relationship to the sequence and solution properties in Chapter 4. In this chapter, we continue to use the coarse-grained models of Chapter 4 to study the dynamics of IDP ensembles, specifically the dynamics of the end-to-end distance which can be characterized experimentally using Förster resonance energy transfer (FRET) in combination with Fluorescence Correlation Spectroscopy (FCS) [22, 52, 191]. Understanding of the link between the amino acid composition and sequence of an IDP, and their dynamics will lead to understanding of their diverse functions such as: how fast they find their binding targets, coupled folding upon binding transitions, and the dynamics of flexible linkers and entropic bristles [2, 4].

FRET, which gives structural information about the ensemble of conformations, can be combined with FCS, to obtain the dynamics of inter-conversion between the conformations. The fluorescence intensities of FRET dyes, typically attached at the two ends of an IDP or natively folded protein, depend on the distance between them. As an IDP transitions between different conformations, fluctuations of the distance between the dyes will cause fluctuations in their fluorescence intensities. The decay times of the correlation functions of the fluorescence intensities will capture the dynamics of the inter-dye distance. This can be related to the decay of the reconfiguration time of the chain defined as the relaxation time of the end-to-end distance auto-correlation function [22, 52–54, 191].

One characteristic experiment quantifying the dynamics of disordered proteins was performed by Soranno et al. [66]. A natively folded protein and two IDPs were labeled with FRET dyes close to their termini. The native protein was a 68 residue mutant of cold shock protein (Csp). The IDPs were a segment of HIV-1 integrase (IN) with low net charge and low hydrophobicity, and a highly net negatively charged segment of human prothymosin α (ProTα), both 60 residues long. FRET measurements of end-to-end distance and FCS measurements of the reconfiguration time were performed at various denaturant concentrations (0-7 M [GdmCl]). Overall, decreasing the denaturant concentrations led to more compact chains. However, the reconfiguration times first decreased as denaturant concentration was lowered and then increased as the solvent approached native conditions. Despite having similar reconfiguration times at high denaturant, all three proteins exhibited different behaviours as the GdmCl concentration

decreased. Csp had the greatest slowing of the end-to-end distance dynamics, while those of ProTα were almost unchanged, and the behaviour of IN was in between the other two. Some of these changes are due to the changing polymer dimensions, in this case radius of gyration inferred from FRET, of the proteins: for Csp and IN the radii of gyration were comparable until the lowest denaturant concentrations (<2 M [GdmCl]) where Csp became more compact, and ProTα had a larger radius of gyration than the other two and compacted similarly to IN. Naively, one would expect the dynamics of the end-to-end distance to speed up with chain compaction simply due to the reduction of the polymer dimensions as is the case for the Gaussian chain model. However, as the chain compacts, the amount of inter-monomer interactions increases and it is not unexpected that the reconfiguration time is a function of the protein sequence and the denaturant concentration.

Although there is still a debate over the molecular details, denaturants, such as Guanidinium Chloride (GdmCl), weaken the inter-residue interactions of a protein, either directly or indirectly by disrupting the hydrophobic effect, leading to unfolding or a swelling of its dimensions [77, 101]. The effect of denaturant on reconfiguration times is not as straightforward. At high denaturant, larger dimensions would lead to longer global reconfiguration times (slower dynamics), however this is accompanied by weaker and less numerous inter-residue interactions which both lead to faster dynamics. Another complication arises as increasing denaturant concentration increases the solvent viscosity leading to a slowing of dynamics [192].

The overdamped Langevin dynamics equation (2.54) states that any time scale is proportional to the solvent viscosity. The solvent viscosity can be varied independently of denaturant concentration in experiments, by varying glycerol concentration, and in molecular dynamics simulation with explicit solvent, by changing the mass of the water molecules [66, 68, 193]. Both experiments and molecular dynamics simulations show a deviation from inverse proportionality with solvent viscosity for protein folding rates [194]. Similarly, Soranno et al. [66] found that the deviation of end-to-end distance reconfiguration time from proportionality with the solvent viscosity increased as GdmCl concentration was lowered. These observations are attributed to the "internal friction" of the polypeptides [54, 66, 195]. The presence of internal friction indicates that some of the assumptions of the overdamped Langevin dynamics equation break down at the level of molecular interactions between the solvent and amino acids. These assumptions include: that monomers are spherical beads experiencing Stokes drag forces, that the random forces due to collisions with the solvent are uncorrelated in time, and that the timescale of monomer motion is significantly longer than the relaxation time due to solvent drag [73, 91, 98].

Multiple explicit solvent molecular dynamics studies of short polypeptides have been performed in order to understand the molecular origins of internal friction [67, 68, 193, 196]. The mechanisms that contribute to internal friction are likely a combination of dihedral angle transitions, hydrogen bonding and other inter-residue interactions. Although several heuristic relationships exist, it is difficult to isolate the contributions to dynamics from solvent and internal friction [194].

Highly coarse-grained models such as the Rouse (or Zimm) model with internal friction can phenomenologically capture the experimental effect of solvent viscosity on reconfiguration time [66, 197, 198]. The standard Rouse model describes a polymer as a series of beads connected by Hookean springs undergoing Brownian dynamics. Aside from the linear forces between neighbouring beads along the chain, there are no other interactions (excluded volume or other interactions between non-neighbouring beads) [25, 37]. The Rouse model can be extended to the "Rouse model with internal friction" by adding a bond friction term that acts to dampen changes in the relative velocities of the neighbouring beads. This term is inherently independent of the solvent viscosity and leads to expressions for the relaxation

times of the normal modes of the form:

$$\tau_r = \tau_i + \tau_s = \tau_i + \frac{\eta}{\eta_0}\tau_s(\eta_0) \tag{5.1}$$

where τ_i and τ_s are the relaxation times associated with internal friction and solvent respectively and η is the solvent viscosity [198]. When experimental reconfiguration times are interpreted using this model, any deviation from proportionality with solvent viscosity can be explained by the bond friction term. Although the Rouse model with internal friction can capture the experimental effects, only bond interactions are included, and other interactions which may contribute to the increase in reconfiguration time for decreasing denaturant are ignored. Additionally, more complicated models and molecular dynamics simulations show that the contribution to reconfiguration time from internal friction is also dependent on the solvent viscosity and may not be so easily separated using the linear assumption [198].

It must be emphasized that there are two separate effects observed in the experimentally inferred reconfiguration times: the deviation from solvent viscosity proportionality as denaturant concentrations are lowered and the increase in the reconfiguration time in the same regime. Both are concomitant with the compaction of the polymer dimensions and the associated increase in strength and number of inter-residue interactions [23, 77]. However, to address the former requires an investigation of the molecular origins of internal friction, which are not accessible using the coarse-grained Langevin dynamics model.

Instead I will focus on the latter: investigating how the strength of inter-residue interactions which lead to IDP compaction affect the end-to-end distance dynamics. Additionally, I will isolate the effects of cohesive inter-residue interactions from sequence heterogeneity by comparing homopolymer and heteropolymer simulations. The coarse-grained model introduced in Chapter 2 can include several effects not captured in simpler Rouse and Zimm-like models: excluded volume repulsion, cohesive interactions between non-bonded monomers, and sequence heterogeneity. Simulations of the coarse-grained model can be thought of as occurring at a sufficiently high but constant solvent viscosity, so that the contributions from internal friction are neglected.

In the rest of the chapter, I will quantify the end-to-end dynamics of polymer chains, using the coarse-grained model introduced in Chapter 2. Section 5.2.1 includes a description of the end-to-end vector and end-to-end distance auto-correlation functions, emphasizing the distinction between the two. Section 5.2.2 reviews the homogeneous and heterogeneous models used to represent compaction of an IDP for various compositions, sequences, and denaturant concentrations. Section 5.3 presents the auto-correlation times of the end-to-end vector and end-to-end distance for the homopolymer and two heterogeneous sequences. The results suggest that sequence heterogeneity, and not solely inter-monomer interactions, is responsible for the experimental observations of increase in reconfiguration time. The analysis of the dynamics of the end-to-end distance offers advantages over the model commonly used in experimental inference and has various implications for IDP functions: the increase in reconfiguration time is due to the IDP exploring conformation states that are more separated in space.

5.2 Definitions and Methods

5.2.1 Correlation Times of End-to-end Dynamics

The end-to-end distance reconfiguration time, which is inferred from FRET and FCS experiments [22, 52, 191], is distinct from the end-to-end vector relaxation time (approximated by the Rouse or Zimm

time) [25, 197]. The latter includes global rotations of the chain in addition to changes in only the distance. Here I define how these quantities are calculated in simulations.

The normalized autocorrelation function of the end-to-end vector is defined as:

$$c_{\vec{R}_e}(t) = \frac{\langle\langle \vec{R}_e(t) \cdot \vec{R}_e(0)\rangle\rangle}{\langle R_e^2\rangle} \tag{5.2}$$

The decay time of this function is referred to as the relaxation time of the end-to-end vector or the rotation time [25, 199]. The double angle brackets represent averaging over instantiations and realizations (or initial conditions and sample paths).

The normalized autocorrelation function of the end-to-end distance is defined as:

$$c_{R_e}(t) = \frac{\langle\langle |\vec{R}_e(t)||\vec{R}_e(0)|\rangle\rangle - \langle|\vec{R}_e|\rangle^2}{\langle \vec{R}_e^2\rangle - \langle|\vec{R}_e|\rangle^2} \tag{5.3}$$

The decay time of this function is referred to as the relaxation time of the end-to-end distance or the reconfiguration time [22, 199].

For the Rouse model, the auto-correlation of the end-to-end vector can be expressed as a sum over the odd normal modes, each of which follows a simple exponential decay. The Rouse time is defined as the longest relaxation time, corresponding to the decay of the first mode and follows $\tau_R \propto \frac{\eta R_g^4}{k_B T} \propto N^2$ for a θ solvent. The Zimm model provides an improvement by including a pre-averaging approximation of the hydrodynamic interactions. The longest relaxation time (Zimm time) in good and θ solvent regimes follows $\tau_Z \propto \frac{\eta R_g^3}{k_B T} \propto N^{3\nu}$ (with a differing prefactor for the two regimes) [25].

5.2.2 Computational Model to Investigate End-to-end Dynamics

In FRET and FCS experiments, the reconfiguration times are calculated for natively folded proteins and IDPs for varying concentrations of chemical denaturant [66, 195, 200]. In addition to increasing the solvent viscosity, the denaturant inhibits the inter-residue cohesive interactions and swells the proteins, measured as an increase in the FRET derived end-to-end distance.

In our model, we aim to capture two effects: the effect of denaturant on weakening the inter-residue interactions, and heterogeneity of the protein sequences and inter-residue interactions. We do not consider the change in solvent viscosity with denaturant concentration. As in Chapter 4, we make use of the coarse-grained overdamped Langevin dynamics with hydrodynamic interactions model described in Section 2.3 and first implemented in Section 4.2.1. In this model, the strength of the cohesive interactions between monomers is controlled by the parameter ϵ; small $\epsilon \approx 0$ corresponds to high denaturant concentrations (e.g. 6-8 M [GdmCl]), while an increase in ϵ is corresponds to a decrease in GdmCl concentration.

We investigate three sequences of $N = 100$ monomers. The first is the homopolymer model introduced in Section 4.2.1. The other two are heterogeneous sequences composed of cohesive ("H") and neutral ("P") monomers first described in Section 4.3.1. These sequences consist of: the repeated "HP" motif and the repeated "HPP" motif. The difference between the homopolymer and heterogeneous sequences in the simulations is that only the cohesive monomers interact via the cohesive/attractive potential, and the neutral do not. All other parameters (LJ diameters, hydrodynamic radii, etc.) are the same.

Because the sequences have different fractions of cohesive monomers, different values of global monomer cohesiveness are required to obtain the same change in polymer dimensions. For the ho-

mopolymer the cohesive interactions strengths ranged from $E = \frac{\epsilon}{kT} = 0$ to $E = \frac{\epsilon}{kT} = 1.9$ inclusive, in intervals of 0.1. For the "$(HP)_{50}$" sequence, the cohesive interaction strengths were: 0.5, 1, 1.5, 2.0, 2.2, 2.4, 2.6, 2.8, 3, 3.2, 3.4, 3.6, 3.8, 4, 4.2, 4.4, 4.6, and 4.8. For the "$(HPP)_{33}H$" sequence, the cohesive interaction strengths were: 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, and 6.

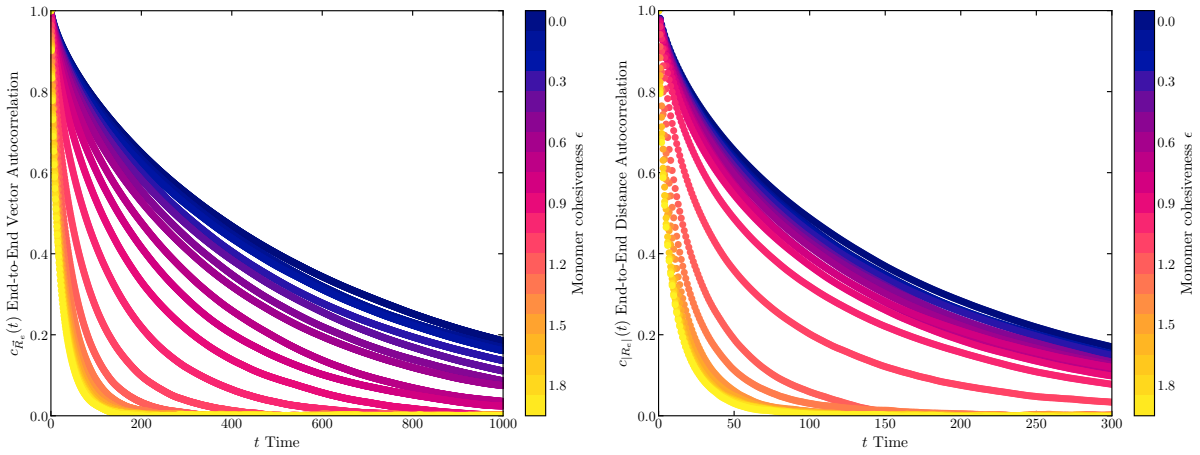
For each $E = \frac{\epsilon}{kT}$, 240 runs were performed, each lasting $\sim 1.8 \times 10^7$ steps, with a time step of $\Delta T = 0.001$. Each run began with a SAW initial condition. The first 2×10^6 steps were excluded from the analysis, and averages were taken over steps over runs and steps. For each run, the autocorrelation function was calculated using the Fast Correlation Algorithm, which uses the Fast Fourier Transform, described in Section 4.1 of Ref. [201] and then the functions were averaged over different runs for each ϵ.

When analyzing simulation results, we define the relaxation times τ of the end-to-end vector and distance as the integral of their normalized autocorrelation functions: $\tau = \int_0^{3\tau_e} c(t)dt$ where $c(t)$ is $c_{\vec{R}_e}(t)$ or $c_{R_e}(t)$, and τ_e is a preliminary estimate of the decay time, which satisfies $c(\tau_e) = e^{-1}$. Other methods, such as fitting a single exponentially decaying function, produce substantially the same results, but more investigations are required to understand the shapes of the autocorrelation functions. Alternatively, the auto-correlation time of the end-to-end vector can be obtained from simulations using the expression $\tau_Z \approx \frac{\langle R_g^2 \rangle}{6D_e}$ [202], since the longest relaxation time should be proportional to the time it takes the chain to move a distance comparable to its size [203]. Here D_e is the long time limit of the diffusion coefficient of the center of mass of the polymer.

5.3 Effects of Cohesiveness and Sequence Heterogeneity on End-to-end Dynamics of IDPs

The Zimm model predicts that in good and poor solvents, the autocorrelation function of the end-to-end vector is a sum of exponentially decaying terms. The slowest of these has a relaxation time of $\tau_Z \sim \frac{\eta N^{3\nu}}{k_B T} \sim R_g^3$ [25]. From simulations with hydrodynamic interactions we calculate the autocorrelation functions of the end-to-end vector and end-to-end distance (shown for the homopolymer in Fig. 5.1) directly. For all three sequences (homopolymer and two heterogeneous sequences), the end-to-end vector relaxation time decreases monotonically with ϵ and the simulation results for the homopolymer are compared to the Zimm scaling prediction in Figure 5.2. For all sequences, both the vector and distance relaxation times (defined as the integral of the normalized autocorrelation function) are shown in Figure 5.3a. This figure compares the sequences by mean square end-to-end distance (on the x-axis), controlled by the global cohesiveness parameter ϵ in the simulations, which differs for the three sequences at each $\sqrt{\langle R_e^2 \rangle}$. For the homopolymer and $(HP)_{50}$ sequence, the end-to-end distance reconfiguration times decrease monotonically with increasing ϵ. For the more heterogeneous $(HPP)_{33}H$ sequence, the end-to-end distance reconfiguration time has a non-monotonic dependence on ϵ, similar to experimental measurements for varying denaturant concentrations [66].

In Figure 5.2 the relaxation time of the end-to-end vector of a homopolymer in all solvent regimes is compared with the Zimm scaling prediction of $\tau \sim R_g^3$. The deviation of the simulation results from the Zimm scaling prediction is not unexpected as the simulations give the relaxation time of a polymer with $N = 100$ in different solvent regimes where the prefactor to the Zimm scaling relationship could change, rather than the relaxation time for different N and unchanging solvent properties. For weak cohesiveness, the relaxation time of the end-to-end vector agrees with $\frac{\langle R_g^2 \rangle}{D_e}$ (where D_e is the long time



(a) (b)

Figure 5.1: Normalized autocorrelation functions of the end-to-end vector (a) and end-to-end distance (b). Homopolymer model with $N = 100$.

limit of the diffusion coefficient of the centre of mass of the polymer) which is the time needed for the polymer to diffuse it's own size [204]. They diverge for stronger cohesiveness and more compact conformations. This may be due to end-to-end distance fluctuations having a stronger contribution to the vector auto-correlation but not affecting the overall diffusion coefficient of the polymer.

The end-to-end distance relaxation time, shown in Figure 5.3a, excludes contributions from rotational dynamics of the entire polymer, and is closer to the reconfiguration time captured by FRET and FCS experiments [22, 197, 199]. In the good and θ solvent regimes, the reconfiguration time (relaxation of the end-to-end distance) follows a different scaling relationship with the polymer dimensions than the rotation time. Beyond the θ point and for poor solvents, the reconfiguration time of the homopolymer and the $(\mathrm{HP})_{50}$ sequence shows a decrease similar to the rotation time (relaxation of the end-to-end vector). However for the $(\mathrm{HPP})_{33}H$ sequence there is a larger deviation from the Zimm scaling relationship of the rotation time and an increase in reconfiguration time when the polymer compacts beyong the θ point, the latter being similar to the behaviour observed for chemically denatured IDPs and natively folded proteins in FRET and FCS experiments. Excluded volume effects and cohesive inter-monomer interactions which lead to chain compaction are not enough to cause an increase in reconfiguration time, instead sufficient heterogeneity in the sequence is necessary.

In order to infer reconfiguration times from experiments [22], 1D normal diffusion in a potential defined by the end-to-end distance of a Gaussian chain is commonly assumed (see Appendix C). Under these assumptions, for a particular τ and $\langle R_e^2 \rangle$, the equivalent normal diffusion coefficient will be proportional to $\frac{\langle R_e^2 \rangle}{\tau}$, and therefore an increase in τ as $\langle R_e^2 \rangle$ decreases will predict a slower end-to-end distance diffusion coefficient. Using simulations, we can look at the distributions of end-to-end distance (Fig. 5.3b) to gain an understanding of the configurational landscape of the different sequences. As cohesiveness increases and the polymers compact, there is a shift to lower end-to-end distances in all distributions. For the homompolymer and the $(\mathrm{HP})_{50}$ sequence, the distributions have a single peak and the Gaussian chain assumption is reasonable. However, for the $(\mathrm{HPP})_{33}\mathrm{H}$ sequence, multiple peaks emerge immediately after the polymer compacts beyond the θ solvent condition. This is illustrated in

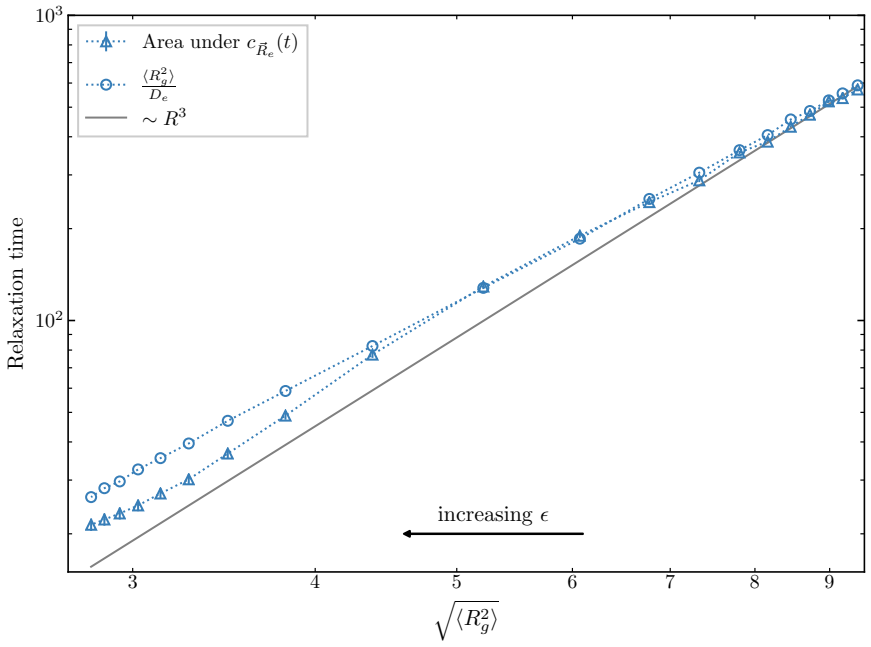


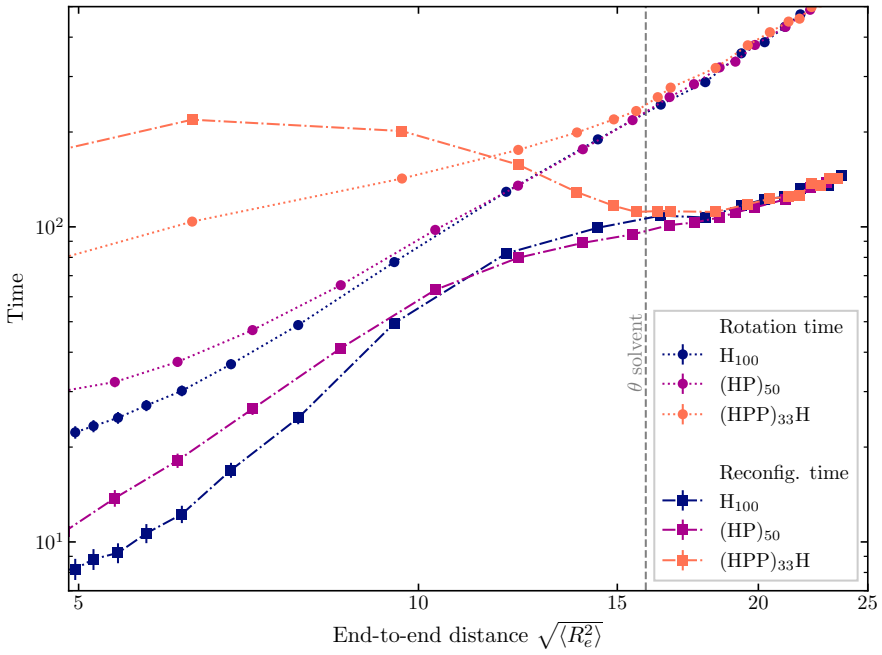
Figure 5.2: The relaxation time of the end-to-end vector autocorrelation function and $\frac{\langle R_g^2 \rangle}{D_e}$ for a homopolymer model with $N = 100$. The x-axis shows the mean squared radius of gyration set by monomer cohesiveness. The grey solid line is the expected scaling with polymer size from the Zimm model, for constant cohesiveness/solvent conditions (by varying N) [25]. Distance and time are in simulation units.

the inset of the figure by the difference between the variances of the end-to-end distances of the three sequences. For the more homogeneous sequences, the variance decreases with compaction, while for the heterogeneous sequence there is a regime where the variance dependence on end-to-end distance is relatively flat. In conclusion, the increase in reconfiguration time is due to heterogeneity of the inter-monomer interactions. The energy landscape in conformation space is wider and more rugged, and the polymer is sampling a few highly probable conformations rather than smoothly transitioning between conformations of a Gaussian chain.

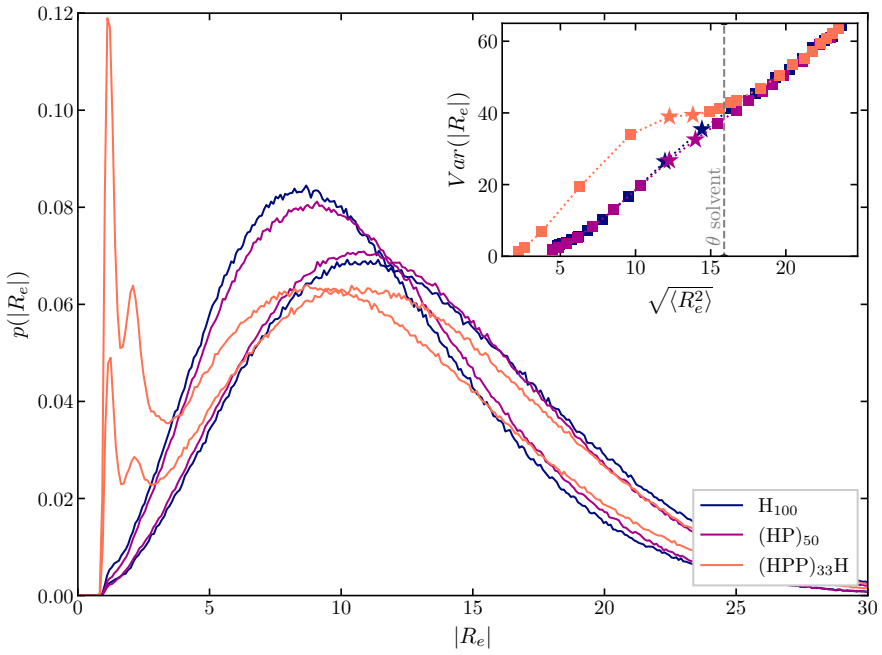
5.4 Conclusions

Both IDPs and chemically denatured proteins can exhibit a non-monotonic dependence of end-to-end distance reconfiguration times on denaturant concentrations and the associated chain compaction [66, 195]. Molecular dynamics studies have proposed various inter-monomer interactions as sources for this behaviour, but its microscopic origins still remain unclear, and the role of sequence has not been investigated [67, 68]. Theoretical approaches based on Rouse (and Zimm) like models can capture the experimentally observed effects but often assume the end-to-end distance dynamics resemble those of the end-to-end vector [197–199]. In this chapter, we have quantified both end-to-end vector and end-to-end distance dynamics, and investigated the role of general inter-monomer interactions and sequence heterogeneity on the dynamics of the polymer ends.

In summary, we have investigated how general inter-monomer interactions affect the dynamics of



(a)



(b)

Figure 5.3: (a) Relaxation time of the end-to-end vector (rotation time) and the end-to-end distance (reconfiguration time) for the different sequences indicated in the legend. The x-axis shows the mean square end-to-end distance controlled by monomer cohesiveness in the simulations. The θ solvent line is the linear interpolation of $\sqrt{\langle R_e^2 \rangle}$ at $\frac{\epsilon}{kT} = 0.725$ for the homopolymer (based on Appendix B). (b) Two end-to-end distance probability distribution functions (with different monomer cohesiveness ϵ) for each of the three sequences described in the legend. *Inset:* variance of the end-to-end distance VS the root mean square end-to-end distance (controlled by monomer cohesiveness in the simulations). The values of root mean square end-to-end distance and variance of the end-to-end distance of the distributions plotted in the main figure are indicated using stars in the inset.

the ends of a polymer using coarse-grained simulations with hydrodynamic interactions. These results are valid in the overdamped regime where solvent viscosity dominates all timescales. Our model includes monomer connectivity, excluded volume interactions, and intra-chain cohesion is modulated by an attractive force between monomers. As cohesiveness varies, the scaling of the end-to-end vector relaxation time with radius of gyration is similar to the prediction of the Zimm model for good and θ solvents, and the deviations can be attributed to the differing prefactor between the poor and good solvent regimes. For the homopolymer, the end-to-end distance reconfiguration time decreases with increasing cohesiveness and decreasing polymer dimensions but does not follow a power law akin to the end-to-end vector relaxation time. Therefore, only increasing the strength of intra-chain interactions is not sufficient to cause an increase in the reconfiguration time. Heterogeneity in the sequence was included via a simple hydrophobic-polar model, and we found that the "$(HPP)_{33}H$" sequence exhibits an increase in reconfiguration time with the compaction of the polymer dimensions, qualitatively following the experimental behaviour [66]. In contrast to the 1D normal diffusion models commonly used in experimental analysis [22], this increase was simultaneous with the emergence of multiple peaks in the end-to-end distance distribution, and was primarily due to the increase of the variance of end-to-end distance (a wider conformational landscape), when compared with the homopolymer.

Although further investigations are required in order to make a direct comparison to experiments possible, our model allows a potential qualitative interpretation of the experimental measurements of end-to-end distance reconfiguration times of IDPs and unfolded states of natively folded proteins by Soranno et al. [66], which were introduced in Section 5.1. At high denaturant, all polypeptides follow the behaviour of the self-avoiding walk model, equivalent to the low cohesiveness regime in the simulations. As the denaturant concentration is lowered, the different responses of reconfiguration time can be understood in terms of the overall amino acid composition as well as the sequence heterogeneity. The unfolded state of Csp likely has a more rugged conformational landscape than the IDPs at lower denaturant concentrations leading to longer reconfiguration times. The two IDPs, ProTα and IN, likely have a smoother conformational landscape leading to shorter reconfiguration times [3]. ProTα also carries a high net charge, which maps to an overall weaker cohesiveness parameter accounting for its lack of increase in reconfiguration time at lower denaturant concentrations. Thus the results of this chapter are in qualitative agreement with experimental measurements on real IDPs and natively folded proteins [66].

For the most heterogeneous sequence ($(HPP)_{33}H$), the increase in reconfiguration time and widening of the end-to-end distance distribution occurred after the cohesive strength increased beyond the conditions of the θ solvent. Experimental measurements of IDP dimensions at native conditions point towards equivalent solvent qualities close to the θ point [35, 36, 38]. These results suggest that there are multiple structural and dynamical behaviours of IDPs modulated by their sequence heterogeneity and which can be probed experimentally. Those with more homogeneous sequences explore Gaussian chain-like conformational landscapes and have faster end-to-end distance reconfiguration times, while those with more heterogeneous sequences explore conformational states that are more distant in space and therefore have slower reconfiguration times. This difference between the conformational ensembles would not appear in a static measurement of polymer dimensions.

Chapter 6

Conclusions

Our understanding of IDPs is hindered by their complexity: IDP structures dynamically explore vast conformational ensembles. Viewing IDPs through the framework of polymer physics and analysis of the sequences of IDPs suggest that, in some cases, the atomistic details of the amino acids may be coarse-grained out. In this thesis, I employed several polymer models and simulations in investigations of the structures and dynamics of IDPs. The models encapsulated the complexity of the IDPs using a small number of key parameters.

The mean field polymer brush model used in Chapter 3 provides a rigorous physical framework for investigating the structure of surface grafted FG nups and their interactions with transport proteins. The model explains the behaviour of collapse and extension and accumulation of transport proteins in the layer and the regimes of parameters which control this behaviour have been identified. The results can qualitatively and semi-quantitatively reproduce the range of *in vitro* experimental observations and explain how initial contradictions could arise between them. The mean-field model reconciles several theories of transport and shows how surface grafted polymer and nanoparticle systems may be controlled.

In order to better understand the limitations of the mean field model, the effects of monomer cohesiveness and sequence patterning on the dimensions of homopolymers and heteropolymers were investigated using coarse-grained simulations in Chapter 4. For the homopolymer model, all of the dimensions were correlated with the cohesiveness and compaction of the polymer, although their ratios varied. We found that sequence heterogeneity can modulate the polymer dimensions independently of composition or the overall cohesiveness. Overall, the ratio of radius of gyration and hydrodynamic radius was correlated with sequence patterning parameters and is a good measure of global conformational features of an IDP. On the other hand, the end-to-end distance was highly sequence specific and the ratio of end-to-end distance to radius of gyration can be used to describe local conformational features of the IDP's ends. The results of Chapter 4 can clarify some of the discrepancies arising between the FRET and SAXS inferred polymer dimensions. The models commonly used to infer the radius of gyration from FRET measurements assume a constant ratio between the end-to-end distance and the radius of gyration. The homopolymer model showed that this ratio varied with monomer cohesiveness. Furthermore, the heteropolymer models showed that the end-to-end distance and consequently the the ratio of end-to-end distance to radius of gyration is highly sequence dependent and is not correlated with overall compaction, and therefore cannot reliably be used to infer the other dimensions. Our coarse grained model offers an improvement over the simple models commonly used in FRET analysis without resorting to all atom

simulations as it gives the probability distributions of the individual polymer dimensions. Rather than being the source of a discrepancy, the combined measurements of several polymer dimensions can be used to describe the conformational ensemble, and guide theoretical and computational models of IDPs. For example, the ratio between radius of gyration and hydrodynamic radius can reveal the location of a particular IDP on the disorder-to-order continuum and the cohesiveness parameter of the IDP, while the ratio of end-to-end distance to radius of gyration may reveal the local conformation of the IDP's ends and whether or not a homopolymer model is appropriate.

In Chapter 5, coarse-grained simulations were used to study the effects of monomer cohesiveness and sequence heterogeneity on the end-to-end dynamics of polymers. For homopolymers, the end-to-end distance reconfiguration time decreases with increasing cohesiveness and decreasing polymer dimensions. Therefore, only increasing the strength of intra-chain interactions is not sufficient to cause an increase in the reconfiguration time. Sufficient heterogeneity in the sequence does lead to an increase in reconfiguration time as cohesiveness increases, qualitatively following the experimental behaviour of proteins in varying concentrations of chemical denaturant. The increase in reconfiguration time is due to a wider conformational landscape, in contrast to the assumptions of the 1D normal diffusion model commonly used in experimental analysis. The results of Chapter 5 suggest how the multiple structural and dynamical behaviours of IDPs, modulated by their sequence heterogeneity, can be probed experimentally. For IDPs of similar lengths and polymer dimensions, those with shorter end-to-end distance reconfiguration times likely explore Gaussian chain-like conformational landscapes and can be modeled using homopolymer models. Longer end-to-end distance reconfiguration times indicate that those IDPs explore more rugged landscapes with conformational states that are more distant in space.

One possible direction for future investigations is to refine the amino acid representation of the coarse-grained model to allow for a more quantitative experimental comparison for specific IDP sequences. This would involve using experimental data of the dimensions for specific sequences to determine parameters for the individual amino acids and create a straightforward one bead per amino acid mapping. After the mapping is obtained, the model could be applied to study specific systems involving IDPs or in FRET and FCS analysis. Another direction is to continue the investigation of the general relationships between amino acid sequence and an IDP's ensemble of conformations and dynamics. Possible investigations could include sequences composed of three or more of the monomer types of the four letter model (e.g. cohesive, positive, and negative) requiring at least two sequence parameters (e.g. κ and Ω) to capture the sequence heterogeneity. Also, the relationships between polymer dimensions and sequence parameters almost certainly depend on the composition (fraction of cohesive, neutral, positive, and negative monomers), and future investigations should look at what these relationships are for different compositional classes of IDPs. Lastly, a more thorough investigation of the effect of sequence on end-to-end distance dynamics is needed since it was so far limited to only three sequences and only two monomer types (neutral and cohesive).

Appendix A

Derivation of the Kirkwood Approximation or Short Time Diffusion Coefficient

The following derivation is adapted from Ref. [30]. In order to find the short time limit of the translational diffusion coefficient, we calculate the center of mass displacement of a polymer whose monomer motions can be simulated using the Ermak-McCammon algorithm [98].

The hydrodynamic radius R_H of a polymer is the radius of a spherical bead with the same translational diffusion coefficient.

$$D_e = \frac{k_B T}{6\pi\eta R_H} \tag{A.1}$$

A single spherical particle in a fluid with low Reynolds number moves with velocity v under the influence of a force F:

$$\vec{v} = \frac{1}{\xi}\vec{F} \tag{A.2}$$

$$\xi = 6\pi\eta a = \frac{k_B T}{D_e} \tag{A.3}$$

where ξ is called the friction constant and a is the radius of the particle.
If a force $\vec{F}$ is exerted on a fluid, the velocity perturbation $\vec{v_h}$ at a point $\vec{r}$ away will be:

$$\vec{v}_h(r) = \bar{M}(\vec{r})\vec{F} \tag{A.4}$$

$$\bar{M}(\vec{r}) = \frac{1}{8\pi\eta r}\left(\boldsymbol{I} + \frac{\vec{r}\vec{r}}{r^2}\right) \tag{A.5}$$

where the Oseen tensor is used for the hydrodynamic interactions $\bar{M}$. Consider a polymer of N spherical monomers each under the influence of a force $\vec{F}_n$:

$$\vec{v}_n = \frac{1}{\xi_n}\vec{F}_n + \sum_{m\neq n}^{N} \bar{M}(\vec{r}_n - \vec{r}_m)\vec{F}_m \tag{A.6}$$

or

$$\vec{v}_n = \sum_{m=1}^{N} \bar{M}_{nm}(\vec{r})\vec{F}_m \tag{A.7}$$

where $\bar{M}_{nm}$ for two monomers n and m is defined as:

$$\bar{M}_{nm}(\vec{r}) = \bar{M}(\vec{r}_n - \vec{r}_m) \text{ for } n \neq m \tag{A.8}$$

$$\bar{M}_{nn} = \frac{1}{\xi_n}\boldsymbol{I} \tag{A.9}$$

For the model of Section 2.3, the displacement of a monomer along 1 dimension over a single time step of the simulation is:

$$\Delta x_n = \Delta t \sum_m M_{nm}F_m + \sum_m H_{nm}\Delta w_m$$

The matrix $\bar{H}$ satisfies $\bar{H}\bar{H}^T = 2\boldsymbol{D}$, and $\boldsymbol{D}_{nm} = k_BT\bar{M}_{nm}$. The elements of $\vec{\Delta w}$ are uncorrelated with $\langle \Delta w_n \rangle = 0$ and $\langle \Delta w_n \Delta w_m \rangle = \Delta t \delta_{nm}$. The displacement of the polymer's center of mass along 1 dimension is:

$$\begin{aligned}
\Delta x_{CM} =& \frac{1}{N}\sum_n \Delta x_n \\
=& \frac{1}{N}(\Delta t \sum_n \sum_m M_{nm}F_m + \sum_n \sum_m H_{nm}\Delta w_m) \\
\langle (\Delta x_{CM})^2 \rangle =& \frac{\Delta t^2}{N^2}\langle (\sum_n \sum_m M_{nm}F_m)^2 \rangle + \\
&+ \frac{1}{N^2}\langle \sum_n \sum_m \sum_i \sum_j H_{nm}\Delta w_m H_{ij}\Delta w_j \rangle + \\
&+ \frac{\Delta t}{N^2}\langle \sum_n \sum_m \sum_i \sum_j M_{nm}F_m H_{ij}\Delta w_j \rangle \\
=& \frac{\Delta t^2}{N^2}\langle (\sum_n \sum_m M_{nm}F_m)^2 \rangle + \frac{\Delta t}{N^2}\langle \sum_n \sum_m \sum_i H_{nm}H^T_{mi} \rangle + \\
&+ \frac{\Delta t}{N^2}\langle \sum_n \sum_m \sum_i \sum_j M_{nm}F_m H_{ij} \rangle \langle \Delta w_j \rangle \\
=& \frac{\Delta t^2}{N^2}\langle (\sum_n \sum_m M_{nm}F_m)^2 \rangle + \frac{2\Delta t}{N^2}\langle \sum_n \sum_i D_{ni} \rangle
\end{aligned}$$

The diffusion coefficient is:

$$\begin{aligned}
\frac{\langle (\Delta x_{CM})^2 \rangle}{2\Delta t} =& \frac{\Delta t}{2N^2}\langle (\sum_n \sum_m M_{nm}F_m)^2 \rangle + \frac{1}{N^2}\langle \sum_n \sum_i D_{ni} \rangle \\
D_K = \lim_{t\to 0} \frac{\langle (\Delta x_{CM})^2 \rangle}{2\Delta t} =& \frac{1}{N^2}\langle \sum_n \sum_i D_{ni} \rangle \\
D_K =& \frac{1}{N^2}\sum_n \sum_m \langle D_{nm} \rangle
\end{aligned}$$

In the 3D case:

$$D_K = \lim_{t\to 0} \frac{\langle (\Delta \vec{r}_{CM})^2 \rangle}{6\Delta t} = \frac{1}{3N^2} \sum_n \sum_m \mathrm{Tr} \langle \boldsymbol{D}_{nm} \rangle$$

Substituting in the Oseen tensor:

$$\begin{aligned} D_K =& \frac{k_B T}{3N^2} \left(\sum_n \frac{3}{6\pi\eta b_n} + \sum_n \sum_{m\neq n} \langle \frac{1}{8\pi\eta r_{nm}} (3+1) \rangle \right) \\ =& \frac{k_B T}{6\pi\eta N^2} \left(\sum_n \frac{1}{b_n} + \sum_n \sum_{m\neq n} \langle \frac{1}{r_{nm}} \rangle \right) \end{aligned}$$

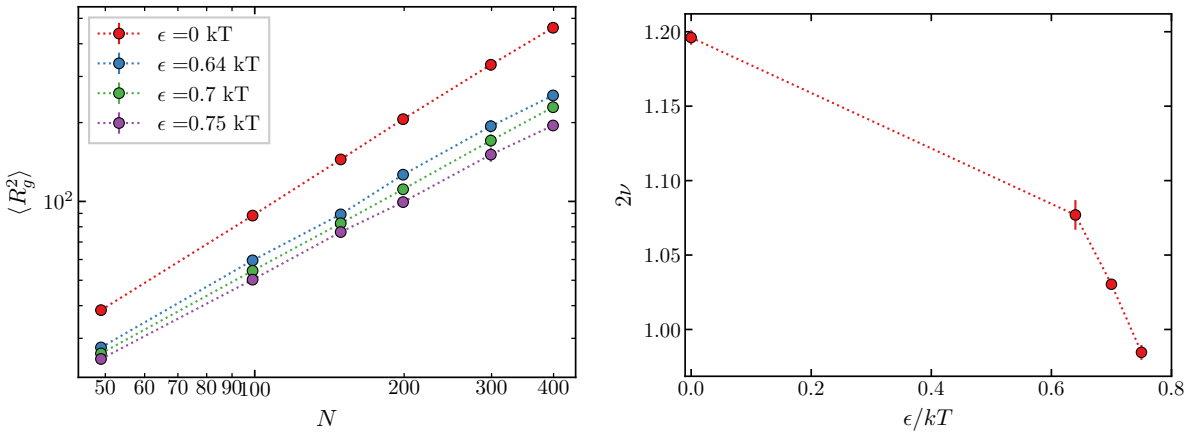
For identical particles:

$$D_K = \frac{k_B T}{6\pi\eta} \left(\frac{1}{Nb} + \frac{1}{N^2} \sum_n \sum_{m\neq n} \langle \frac{1}{r_{nm}} \rangle \right)$$

Appendix B

Scaling Exponent of Radius of Gyration

Using the coarse-grained model of Section 2.3, we investigate the effects of monomer cohesiveness ϵ on the scaling exponent ν of the radius of gyration of a homopolymer with the number of bonds: $\langle R_g^2 \rangle \propto N^{2\nu}$. In Section 2.3.5, we found that the second virial coefficient for the excluded volume and cohesive interactions is 0 when $\epsilon \approx 0.64$ kT.



(a) (b)

Figure B.1: (a) Dependence of radius of gyration on the number of bonds in the chain for different monomer cohesiveness ϵ. (b) The variation of the scaling exponent of $\langle R_g^2 \rangle \propto N^{2\nu}$ with monomer cohesiveness ϵ.

Figure B.1 shows the dependence of radius of gyration on the number of bonds N in the homopolymers for different monomer cohesiveness ϵ and the fitted scaling exponent ν. At $\epsilon = 0$, we find good agreement with a more exact prediction of 0.588 ± 0.001 [25] for the self-avoiding walk. We also find that ϵ at the θ point (where $\nu = 0.5$) is between 0.7 kT and 0.75 kT, which is larger than the predicted value of 0.64 kT. The exact deviation is dependent on the details of the repulsive and attractive potentials used in the model [205].

We performed simulations of homopolymers with $N+1 = 50, 100, 150, 200, 300, 400$ monomers at $\epsilon = 0, 0.64, 0.7, 0.75$ kT. All other simulation parameters were the same as the homopolymer model of Section 4.2.1. The total runtime (number of steps) and number of independent runs (from different initial conditions) varied. Simulations where initialized from a self-avoiding walk ($N+1 = 50, 100, 150, 200$) or a random walk initial condition ($N = 200, 300, 400$). Correlation functions of $R_g^2(t)$ were calculated for $N+1 = 50, 100, 150$ and fit with exponential decays in order to estimate the correlation times τ. The correlation times for $N+1 = 50, 100, 150$ were then fit with a power law in order to obtain a relationship of the form $\tau_N \propto N^\alpha$. The initial $2\tau_N$ or more of each simulation were excluded from the analysis. The error was estimated as $\mathrm{Var}(\langle R_g^2 \rangle) = \frac{\mathrm{Var}(R_g^2)}{n_s n_t}$ where n_s is the number of independent runs and $n_t = \frac{t_r}{2\tau_N}$, where t_r is the minimum simulation time included in the analysis for that N [111].

Appendix C

1D Model for Diffusion of End-to-end Distance

In FCS experiments, the observables are the correlation functions of the fluoresence intensities of the dyes, often attached at the two ends of a polypeptide chain. Combined with the inferred end-to-end distance from FRET ($\langle R_e^2 \rangle$), the relaxation times of the end-to-end distance is extracted by modeling fluctuations of the distance between the two dyes as normal diffusion in a 1D potential of mean force given by $U(r) = -k_B T \ln p_{eq}(r)$, where $p_{eq}(r)$ is the probability distribution of the end-to-end distance of a Gaussian chain (Equation 2.1). Along with a model of the FRET photo-physics, the diffusion coefficient D_G that best reproduces the measured fluoresence intensity correlation functions is extracted. The relaxation time of the end-to-end distance or "reconfiguration time" is calculated from $\frac{\langle R_e^2 \rangle}{D_G}$ [22].

In this model, the Langevin equation for the end-to-end distance is:

$$\frac{dr}{dt} = -\frac{D_G}{k_B T}\frac{dU}{dr} + \xi(t) = D_G\left(\frac{2}{x} - \frac{3x}{\langle r^2 \rangle}\right) + \xi(t)$$

where $\langle \xi(t) \rangle = 0$ and $\langle \xi(t)\xi(t') \rangle = 2D_G\delta(t - t')$. We can non-dimensionalize the above by setting $R = \frac{r}{x_c}$, $T = \frac{t}{t_c}$, $D_G = \frac{x_c^2}{t_c}$, and $\langle r^2 \rangle = x_c^2 \langle R^2 \rangle$, where x_c and t_c are the units of distance and time, the Ito stochastic differential equation becomes:

$$dR = \left(\frac{2}{R} - \frac{3R}{\langle R^2 \rangle}\right) dT + \sqrt{2} dW$$

The dimensionless mean square end-to-end distance is the only parameter of the Langevin equation. But it is straightforward to verify that for any observed end-to-end distance $\langle r_e^2 \rangle$ and end-to-end distance reconfiguration time τ from experiments or simulations, the diffusion coefficient of the end-to-end distance according to the 1D model will follow: $D_G = \frac{C\langle r_e^2 \rangle}{\tau}$, where C is a numerical constant. The 1D model, which is commonly used in experimental analysis [22], assumes normal diffusion in a potential of width comparable to the end-to-end distance.

Bibliography

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

[2]J. Habchi, P. Tompa, S. Longhi, and V. N. Uversky, "Introducing protein intrinsic disorder", Chemical Reviews **114**, 6561–6588 (2014).

[3]V. N. Uversky, "Unusual biophysics of intrinsically disordered proteins", Biochimica et Biophysica Acta - Proteins and Proteomics **1834**, 932–951 (2013).

[4]R. Van Der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright, and M. M. Babu, "Classification of intrinsically disordered regions and proteins", Chemical Reviews **114**, 6589–6631 (2014).

[5]B. Xue, A. K. Dunker, and V. N. Uversky, "Orderly order in protein intrinsic disorder distribution: Disorder in 3500 proteomes from viruses and the three domains of life", Journal of Biomolecular Structure and Dynamics **30**, 137–149 (2012).

[6]A. K. Dunker, C. J. Oldfield, J. Meng, P. Romero, J. Y. Yang, J. W. Chen, V. Vacic, Z. Obradovic, and V. N. Uversky, "The unfoldomics decade: An update on intrinsically disordered proteins", in Bmc genomics, Vol. 9, SUPPL. 2 (Sept. 2008), S1.

[7]P. Tompa, "Intrinsically disordered proteins: A 10-year recap", Trends in Biochemical Sciences **37**, 509–516 (2012).

[8]V. N. Uversky, V. Davé, L. M. Iakoucheva, P. Malaney, S. J. Metallo, R. R. Pathak, and A. C. Joerger, "Pathological unfoldomics of uncontrolled chaos: Intrinsically disordered proteins and human diseases", Chemical Reviews **114**, 6844–6879 (2014).

[9]V. N. Uversky, "A decade and a half of protein intrinsic disorder: Biology still waits for physics", Protein Science **22**, 693–724 (2013).

[10]Z. Liu and Y. Huang, "Advantages of proteins being disordered", Protein Science **23**, 539–550 (2014).

[11]V. N. Uversky, "The alphabet of intrinsic disorder: II. Various roles of glutamic acid in ordered and intrinsically disordered proteins", Intrinsically Disordered Proteins **1**, e24684–1 (2013).

[12]S. R. Wente and M. P. Rout, "The nuclear pore complex and nuclear transport.", Cold Spring Harbor perspectives in biology **2**, a000562 (2010).

[13]T. Jamali, Y. Jamali, M. Mehrbod, and M. R. K. Mofrad, "Nuclear pore complex: biochemistry and biophysics of nucleocytoplasmic transport in health and disease.", in *International review of cell and molecular biology*, Vol. 287 (Academic Press, Jan. 2011) Chap. 6, pp. 233–86.

[14]A. Dickmanns, R. H. Kehlenbach, and B. Fahrenkrog, "Nuclear Pore Complexes and Nucleocytoplasmic Transport: From Structure to Function to Disease", in *International review of cell and molecular biology*, Vol. 320 (Academic Press, Jan. 2015) Chap. 5, pp. 171–233.

[15]C. W. Pouton, K. M. Wagstaff, D. M. Roth, G. W. Moseley, and D. A. Jans, "Targeted delivery to the nucleus", Advanced Drug Delivery Reviews **59**, 698–717 (2007).

[16]T. Jovanovic-Talisman, J. Tetenbaum-Novatt, A. S. McKenney, A. Zilman, R. Peters, M. P. Rout, and B. T. Chait, "Artificial nanopores that mimic the transport selectivity of the nuclear pore complex.", Nature **457**, 1023–7 (2009).

[17]B. He, K. Wang, Y. Liu, B. Xue, V. N. Uversky, and A. K. Dunker, "Predicting intrinsic disorder in proteins: An overview", Cell Research **19**, 929–949 (2009).

[18]C. J. Oldfield and A. K. Dunker, "Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions", Annual Review of Biochemistry **83**, 553–584 (2014).

[19]S. Rauscher and R. Pomes, "Molecular simulations of protein disorder", Biochemistry and Cell Biology **88**, 269–290 (2010).

[20]V. M. Burger, T. Gurry, and C. M. Stultz, "Intrinsically disordered proteins: Where computation meets experiment", Polymers **6**, 2684–2719 (2014).

[21]V. N. Uversky, "Paradoxes and wonders of intrinsic disorder: Complexity of simplicity", Intrinsically Disordered Proteins **4**, e1135015 (2016).

[22]B. Schuler, A. Soranno, H. Hofmann, and D. Nettels, "Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins", **45**, 207–31 (2016).

[23]A. H. Mao, N. Lyle, and R. V. Pappu, "Describing sequence-ensemble relationships for intrinsically disordered proteins.", The Biochemical journal **449**, 307–18 (2013).

[24]P. G. De Gennes, *Scaling concepts in polymer physics* (Cornell university press, Ithaca N.Y., 1979).

[25]M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon Press, Oxford, 1986).

[26]C. Wu and X. Wang, "Globule-to-Coil Transition of a Single Homopolymer Chain in Solution", PHYSICAL REVIEW LETTERS **80**, 4092–4094 (1998).

[27]R. K. Das, K. M. Ruff, and R. V. Pappu, "Relating sequence encoded information to form and function of intrinsically disordered proteins", Current Opinion in Structural Biology **32**, 102–112 (2015).

[28]V. Tozzini, "Coarse-grained models for proteins.", Current opinion in structural biology **15**, 144–50 (2005).

[29]S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski, "Coarse-Grained Protein Models and Their Applications", Chemical Reviews **116**, 7898–7936 (2016).

[30]B. Liu and B. Dünweg, "Translational diffusion of polymer chains with excluded volume and hydrodynamic interactions by Brownian dynamics simulation", The Journal of Chemical Physics **118**, 5057 (2003).

[31]R. Rodríguez Schmidt, J. G. Hernández Cifre, and J. García de la Torre, "Translational diffusion coefficients of macromolecules.", The European physical journal. E, Soft matter **35**, 9806 (2012).

[32]J. G. Kirkwood and J. Riseman, "The intrinsic viscosities and diffusion constants of flexible macromolecules in solution", The Journal of Chemical Physics **16**, 565–573 (1948).

[33]M. Benhamou and G. Mahoux, "Long polymers in good solvent : -expansion of the ratio of the radius of gyration to the end to end distance To cite this version :" Journal de Physique Lettres **46**, 689–693 (1985).

[34]J. E. Kohn, I. S. Millett, J. Jacob, B. Zagrovic, T. M. Dillon, N. Cingel, R. S. Dothager, S. Seifert, P. Thiyagarajan, T. R. Sosnick, M. Z. Hasan, V. S. Pande, I. Ruczinski, S. Doniach, and K. W. Plaxco, "Random-coil behavior and the dimensions of chemically unfolded proteins", Proceedings of the National Academy of Sciences **101**, 12491–12496 (2004).

[35]H. Hofmann, A. Soranno, A. Borgia, K. Gast, D. Nettels, and B. Schuler, "Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy", Proceedings of the National Academy of Sciences **109**, 16155–16160 (2012).

[36]G. Fuertes, N. Banterle, K. M. Ruff, A. Chowdhury, D. Mercadante, C. Koehler, M. Kachala, G. Estrada Girona, S. Milles, A. Mishra, P. R. Onck, F. Gräter, S. Esteban-Martín, R. V. Pappu, D. I. Svergun, and E. A. Lemke, "Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements", Proceedings of the National Academy of Sciences **114**, 201704692 (2017).

[37]M. Doi, *Introduction to polymer physics* (Clarendon Press, Oxford, 1996).

[38]J. A. Marsh and J. D. Forman-Kay, "Sequence determinants of compaction in intrinsically disordered proteins.", Biophysical journal **98**, 2383–90 (2010).

[39]J. Dayantis, "On the physics of collapse of isolated polymer chains", Makromol. Chem **187**, 1035–1075 (1986).

[40]M. Parry and E. Fischbach, "Probability distribution of distance in a uniform ellipsoid: Theory and applications to physics", Journal of Mathematical Physics **41**, 2417–2433 (2000).

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

[42]F.-X. Theillet, L. Kalmar, P. Tompa, K.-H. Han, P. Selenko, A. K. Dunker, G. W. Daughdrill, and V. N. Uversky, "The alphabet of intrinsic disorder I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins", Intrinsically Disordered Proteins **1**, e24360–1 (2013).

[43]E. A. Weathers, M. E. Paulaitis, T. B. Woolf, and J. H. Hoh, "Reduced amino acid alphabet is sufficient to accurately recognize intrinsically disordered protein", FEBS Letters **576**, 348–352 (2004).

[44]J. N. Bright, T. B. Woolf, and J. H. Hoh, "Predicting properties of intrinsically unstructured proteins", Progress in Biophysics and Molecular Biology **76**, 131–173 (2001).

[45]H. S. Ashbaugh and H. W. Hatch, "Natively unfolded protein stability as a coil-to-globule transition in charge/hydropathy space", Journal of the American Chemical Society **130**, 9536–9542 (2008).

[46]J. Yamada, J. L. Phillips, S. Patel, G. Goldfien, A. Calestagne-Morelli, H. Huang, R. Reza, J. Acheson, V. V. Krishnan, S. Newsam, A. Gopinathan, E. Y. Lau, M. E. Colvin, V. N. Uversky, and M. F. Rexach, "A Bimodal Distribution of Two Distinct Categories of Intrinsically Disordered Structures with Separate Functions in FG Nucleoporins", Molecular & Cellular Proteomics **9**, 2205–2224 (2010).

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

[48]A. H. Mao, S. L. Crick, A. Vitalis, C. L. Chicoine, and R. V. Pappu, "Net charge per residue modulates conformational ensembles of intrinsically disordered proteins", Proceedings of the National Academy of Sciences **107**, 8183–8188 (2010).

[49]S. Milles and E. A. Lemke, "Single molecule study of the intrinsically disordered FG-repeat nucleoporin 153.", Biophysical journal **101**, 1710–9 (2011).

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

[51]E. W. Martin, A. S. Holehouse, C. R. Grace, A. Hughes, R. V. Pappu, and T. Mittag, "Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation", Journal of the American Chemical Society **138**, 15323–15335 (2016).

[52]B. Schuler, "Single-molecule FRET of protein structure and dynamics - a primer.", Journal of nanobiotechnology **11** (2013).

[53]G.-N. Gomes and C. C. Gradinaru, "Insights into the conformations and dynamics of intrinsically disordered proteins using single-molecule fluorescence", Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1865**, 1696–1706 (2017).

[54]B. Schuler, "Perspective: Chain dynamics of unfolded and intrinsically disordered proteins from nanosecond fluorescence correlation spectroscopy combined with single-molecule FRET", Journal of Chemical Physics **149**, 20901 (2018).

[55]E. P. O'Brien, G. Morrison, B. R. Brooks, and D. Thirumalai, "How accurate are polymer models in the analysis of Förster resonance energy transfer experiments on proteins?", The Journal of Chemical Physics **130** (2009).

[56]I. C. Sanchez, "Phase Transition Behavior of the Isolated Polymer Chain", Macromolecules **12**, 980–988 (1979).

[57]G. Ziv and G. Haran, "Protein folding, protein collapse, and Tanford's transfer model: Lessons from single-molecule FRET", Journal of the American Chemical Society **131**, 2942–2947 (2009).

[58]J. Song, G. N. Gomes, C. C. Gradinaru, and H. S. Chan, "An Adequate Account of Excluded Volume Is Necessary to Infer Compactness and Asphericity of Disordered Proteins by Forster Resonance Energy Transfer", Journal of Physical Chemistry B **119**, 15191–15202 (2015).

[59]J. Song, G. N. Gomes, T. Shi, C. C. Gradinaru, and H. S. Chan, "Conformational Heterogeneity and FRET Data Interpretation for Dimensions of Unfolded Proteins", Biophysical Journal **113**, 1012–1024 (2017).

[60]T. Y. Yoo, S. P. Meisburger, J. Hinshaw, L. Pollack, G. Haran, T. R. Sosnick, and K. Plaxco, "Small-angle X-ray scattering and single-molecule FRET spectroscopy produce highly divergent views of the low-denaturant unfolded state", Journal of Molecular Biology **418**, 226–236 (2012).

[61]H. M. Watkins, A. J. Simon, T. R. Sosnick, E. A. Lipman, R. P. Hjelm, and K. W. Plaxco, "Random coil negative control reproduces the discrepancy between scattering and FRET measurements of denatured protein dimensions", Proceedings of the National Academy of Sciences **112**, 6631–6636 (2015).

[62]G. H. Zerze, R. B. Best, and J. Mittal, "Modest influence of FRET chromophores on the properties of unfolded proteins", Biophysical Journal **107**, 1654–1660 (2014).

[63]A. Borgia, W. Zheng, K. Buholzer, M. B. Borgia, A. Schüler, H. Hofmann, A. Soranno, D. Nettels, K. Gast, A. Grishaev, R. B. Best, and B. Schuler, "Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods", Journal of the American Chemical Society **138**, 11714–11726 (2016).

[64]A. Ortega, D. Amorós, and J. García De La Torre, "Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models", Biophysical Journal **101**, 892–898 (2011).

[65]W. Zheng, A. Borgia, K. Buholzer, A. Grishaev, B. Schuler, and R. B. Best, "Probing the Action of Chemical Denaturant on an Intrinsically Disordered Protein by Simulation and Experiment", Journal of the American Chemical Society **138**, 11702–11713 (2016).

[66]A. Soranno, B. Buchli, D. Nettels, R. R. Cheng, S. Müller-Späth, S. H. Pfeil, A. Hoffmann, E. A. Lipman, D. E. Makarov, and B. Schuler, "Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy", Proceedings of the National Academy of Sciences **109**, 17800–17806 (2012).

[67]I. Echeverria, D. E. Makarov, and G. A. Papoian, "Concerted dihedral rotations give rise to internal friction in unfolded proteins", Journal of the American Chemical Society **136**, 8708–8713 (2014).

[68]D. De Sancho, A. Sirur, and R. B. Best, "Molecular origins of internal friction effects on protein-folding rates", Nature Communications **5** (2014).

[69]R. B. Best, "Computational and theoretical advances in studies of intrinsically disordered proteins", Current Opinion in Structural Biology **42**, 147–154 (2017).

[70]S. Rauscher, V. Gapsys, M. J. Gajda, M. Zweckstetter, B. L. De Groot, and H. Grubmüller, "Structural ensembles of intrinsically disordered proteins depend strongly on force field: A comparison to experiment", Journal of Chemical Theory and Computation **11**, 5513–5524 (2015).

[71]D. Mercadante, J. A. Wagner, I. V. Aramburu, E. A. Lemke, and F. Gräter, "Sampling Long- versus Short-Range Interactions Defines the Ability of Force Fields to Reproduce the Dynamics of Intrinsically Disordered Proteins", Journal of Chemical Theory and Computation **13**, 3964–3974 (2017).

[72]S.-H. Chong, P. Chatterjee, and S. Ham, "Computer Simulations of Intrinsically Disordered Proteins", Annual Review of Physical Chemistry **68**, 117–34 (2017).

[73]G. W. Slater, C. Holm, M. V. Chubynsky, H. W. de Haan, A. Dubé, K. Grass, O. A. Hickey, C. Kingsburry, D. Sean, T. N. Shendruk, and L. Zhan, "Modeling the separation of macromolecules: A review of current computer simulation methods", Electrophoresis **30**, 792–818 (2009).

[74]K. M. Ruff, R. V. Pappu, and A. S. Holehouse, "Conformational preferences and phase behavior of intrinsically disordered low complexity sequences: insights from multiscale simulations", Current Opinion in Structural Biology **56**, 1–10 (2019).

[75]J. N. Israelachvili, *Intermolecular and Surface Forces* (Academic Press, 1992).

[76]A. Karshikoff, *Non-Covalent Interactions in Proteins* (Imperial College Press, London, 2006).

[77]A. S. Holehouse and R. V. Pappu, "Collapse Transitions of Proteins and the Interplay Among Backbone, Sidechain, and Solvent Interactions", Annual Review of Biophysics **47** (2018).

[78]A. Vovk, C. Gu, M. G. Opferman, L. E. Kapinos, R. Y. Lim, R. D. Coalson, D. Jasnow, and A. Zilman, "Simple biophysics underpins collective conformations of the intrinsically disordered proteins of the nuclear pore complex", eLife **5** (2016).

[79]C. Gu, A. Vovk, T. Zheng, R. D. Coalson, and A. Zilman, "The Role of Cohesiveness in the Permeability of the Spatial Assemblies of FG Nucleoporins", Biophysical Journal **116**, 1204–1215 (2019).

[80]A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (AIP Press, New York, 1994).

[81]W. Van Saarloos, "On the hydrodynamic radius of fractal aggregates", Physica A: Statistical Mechanics and its Applications **147**, 280–296 (1987).

[82]P. J. Flory, *Principles of polymer chemistry* (Cornell University Press, 1953).

[83]K. Dill, *Molecular Driving Forces* (Garland Science, Taylor & Francis Group, LLC, 2018).

[84]S. Alexander, "Adsorption of chain molecules with a polar head a scaling description", Journal de Physique **38**, 983–987 (1977).

[85]P. G. de Gennes, "Conformations of Polymers Attached to an Interface", Macromolecules **13**, 1069–1075 (1980).

[86]M. G. Opferman, R. D. Coalson, D. Jasnow, and A. Zilman, "Morphological control of grafted polymer films via attraction to small nanoparticle inclusions", Physical Review E **86**, 031806 (2012).

[87]M. G. Opferman, R. D. Coalson, D. Jasnow, and A. Zilman, "Morphology of polymer brushes infiltrated by attractive nanoinclusions of various sizes.", Langmuir : the ACS journal of surfaces and colloids **29**, 8584–91 (2013).

[88]S. T. Milner, T. A. Witten, and M. E. Cates, "Theory of the Grafted Polymer Brush", Macromolecules **21**, 2610–2619 (1988).

[89]P. Y. Lai and A. Halperin, "Polymer Brushes in Mixed Solvents: Chromatography and Collapse", Macromolecules **25**, 6693–6695 (1992).

[90]G. T. Nolan and P. E. Kavanagh, "Computer simulation of random packing of hard spheres", Powder Technology **72**, 149–155 (1992).

[91]P. Szymczak and M. Cieplak, "Hydrodynamic effects in proteins", Journal of Physics Condensed Matter **23**, 33102–14 (2011).

[92]C. Loken, D. Gruner, L. Groer, R. Peltier, N. Bunn, M. Craig, T. Henriques, J. Dempsey, C. H. Yu, J. Chen, L. J. Dursi, J. Chong, S. Northrup, J. Pinto, N. Knecht, and R. V. Zon, "SciNet: Lessons learned from building a power-efficient top-20 system and data centre", in Journal of physics: conference series, Vol. 256, 1 (Nov. 2010), p. 012026.

[93]J. G. De La Torre and V. A. Bloomfield, "Hydrodynamic properties of macromolecular complexes. I. Translation", Biopolymers **16**, 1747–1763 (1977).

[94]B. Noetinger, "Fluctuating hydrodynamics and Brownian motion", Physica A: Statistical Mechanics and its Applications **163**, 545–558 (1990).

[95]Y. Harpaz, M. Gerstein, and C. Chothia, "Volume changes on protein folding", Structure **2**, 641–649 (1994).

[96]K. Jacobs, *Stochastic processes for physicists : understanding noisy systems* (Cambridge University Press, 2010), p. 188.

[97]L. C. Evans, *An introduction to stochastic differential equations* (American Mathematical Society, 2013), p. 151.

[98]D. L. Ermak and J. A. McCammon, "Brownian dynamics with hydrodynamic interactions", The Journal of Chemical Physics **69**, 1352–1360 (1978).

[99]G. Stirnemann, D. Giganti, J. M. Fernandez, and B. J. Berne, "Elasticity, structure, and relaxation of extended proteins under force", Proceedings of the National Academy of Sciences **110**, 3847–3852 (2013).

[100]M. Levitt, "A simplified representation of protein conformations for rapid simulation of protein folding", Journal of Molecular Biology **104**, 59–107 (1976).

[101]J. L. England and G. Haran, "Role of Solvation Effects in Protein Denaturation: From Thermodynamics to Single Molecules and Back", Annual Review of Physical Chemistry **62**, 257–277 (2011).

[102]D. V. Schroeder, *An Introduction to Thermal Physics* (Addison Wesley Longman, 2000), p. 334.

[103]O. Teschke, G. Ceotto, and E. F. de Souza, "Interfacial water dielectric-permittivity-profile measurements using atomic force microscopy", Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics **64**, 10 (2001).

[104]M. Ahmad, W. Gu, T. Geyer, and V. Helms, "Adhesive water networks facilitate binding of protein interfaces", Nature Communications **2** (2011).

[105]M. A. Blanco, E. Sahin, A. S. Robinson, and C. J. Roberts, "Coarse-grained model for colloidal protein interactions, B22, and protein cluster formation", Journal of Physical Chemistry B **117**, 16013–16028 (2013).

[106]A. Ghavami, E. van der Giessen, and P. R. Onck, "Coarse-Grained Potentials for Local Interactions in Unfolded Proteins", Journal of Chemical Theory and Computation **9**, 432–440 (2013).

[107]H. Yamakawa, "Transport Properties of Polymer Chains in Dilute Solution: Hydrodynamic Interaction", The Journal of Chemical Physics **53**, 436 (1970).

[108]M. Fixman, "Inclusion of Hydrodynamic Interaction in Polymer Dynamical Simulations", Macromolecules **14**, 1710–1717 (1981).

[109]P. J. Zuk, E. Wajnryb, K. A. Mizerski, and P. Szymczak, "Rotne-Prager-Yamakawa approximation for different-sized particles in application to macromolecular bead models", Journal of Fluid Mechanics **741**, 5 (2014).

[110]R. R. Schmidt, J. G. Cifre, and J. G. De La Torre, "Comparison of Brownian dynamics algorithms with hydrodynamic interaction", Journal of Chemical Physics **135**, 84116 (2011).

[111]D. Frenkel and B. Smit, *Understanding molecular simulation : from algorithms to applications* (Academic Pres, 2002), p. 638.

[112]T. T. Pham, M. Bajaj, and J. R. Prakash, "Brownian dynamics simulation of polymer collapse in a poor solvent: Influence of implicit hydrodynamic interactions", Soft Matter **4**, 1196–1207 (2008).

[113]C. M. Feldherr and D. Akin, "The location of the transport gate in the nuclear pore complex.", Journal of cell science **110**, 3065–70 (1997).

[114]L. J. Terry and S. R. Wente, "Flexible gates: Dynamic topologies and functions for FG nucleoporins in nucleocytoplasmic transport", Eukaryotic Cell **8**, 1814–1827 (2009).

[115]J. D. Aitchison and M. P. Rout, "The yeast nuclear pore complex and transport through it.", Genetics **190**, 855–83 (2012).

[116]J. Fernandez-Martinez and M. P. Rout, "A jumbo problem: mapping the structure and functions of the nuclear pore complex", Current Opinion in Cell Biology **24**, 92–99 (2012).

[117]M. Beck and E. Hurt, "The nuclear pore complex: Understanding its function through structural insight", Nature Reviews Molecular Cell Biology **18**, 73–89 (2017).

[118]S. J. Kim, J. Fernandez-Martinez, I. Nudelman, Y. Shi, W. Zhang, B. Raveh, T. Herricks, B. D. Slaughter, J. A. Hogan, P. Upla, i. E. Chemmama, R. Pellarin, I. Echeverria, M. Shivaraju, a. S. Chaudhury, J. Wang, R. Williams, J. R. Unruh, C. H. Greenberg, E. Y. Jacobs, Z. Yu, M. J. De La Cruz, R. Mironska, D. L. Stokes, J. D. Aitchison, m. F. Jarrold, J. L. Gerton, S. J. Ludtke, c. W. Akey, B. T. Chait, A. Sali, and m. P. Rout, "Integrative structure and functional anatomy of a nuclear pore complex", Nature **555**, 475–482 (2018).

[119]D. P. Denning, S. S. Patel, V. Uversky, A. L. Fink, and M. Rexach, "Disorder in the nuclear pore complex: The FG repeat regions of nucleoporins are natively unfolded", Proceedings of the National Academy of Sciences **100**, 2450–2455 (2003).

[120]E. A. Lemke, "The Multiple Faces of Disordered Nucleoporins", Journal of molecular biology **428**, 2011–2024 (2016).

[121]N. Matsushima, T. Tanaka, and R. H. Kretsinger, "Non-globular structures of tandem repeats in proteins", Protein Pept Lett **16**, 1297–1322 (2009).

[122]J. Tetenbaum-Novatt and M. P. Rout, "The mechanism of nucleocytoplasmic transport through the nuclear pore complex", Cold Spring Harbor Symposia on Quantitative Biology **75**, 567–584 (2010).

[123]C. Feldherr, D. Akin, T. Littlewood, and M. Stewart, "The molecular mechanism of translocation through the nuclear pore complex is highly conserved.", Journal of cell science **115**, 2997–3005 (2002).

[124]L. A. Strawn, T. Shen, N. Shulga, D. S. Goldfarb, and S. R. Wente, "Minimal nuclear pore complexes define FG repeat domains essential for transport", Nature Cell Biology **6**, 197–206 (2004).

[125]B. B. Hülsmann, A. A. Labokha, and D. Görlich, "The Permeability of Reconstituted Nuclear Pores Provides Direct Evidence for the Selective Phase Model", Cell **150**, 738–751 (2012).

[126]P. Popken, A. Ghavami, P. R. Onck, B. Poolman, and L. M. Veenhoff, "Size-dependent leak of soluble and membrane proteins through the yeast nuclear pore complex.", Molecular biology of the cell **26**, 1386–94 (2015).

[127]R. Bayliss, K. Ribbeck, D. Akin, H. M. Kent, C. M. Feldherr, D. Görlich, and M. Stewart, "Interaction between NTF2 and xFxFG-containing nucleoporins is required to mediate nuclear import of RanGDP", Journal of Molecular Biology **293**, 579–593 (1999).

[128]R. Bayliss, T. Littlewood, and M. Stewart, "Structural Basis for the Interaction between FxFG Nucleoporin Repeats and Importin-β in Nuclear Trafficking", Cell **102**, 99–108 (2000).

[129]R. Bayliss, S. W. Leung, R. P. Baker, B. B. Quimby, A. H. Corbett, and M. Stewart, "Structural basis for the interaction between NTF2 and nucleoporin FxFG repeats.", The EMBO journal **21**, 2843–53 (2002).

[130]T. A. Isgro and K. Schulten, "Binding Dynamics of Isolated Nucleoporin Repeat Regions to Importin-β", Structure **13**, 1869–1879 (2005).

[131]B. Naim, D. Zbaida, S. Dagan, R. Kapon, and Z. Reich, "Cargo surface hydrophobicity is sufficient to overcome the nuclear pore complex selectivity barrier", EMBO Journal **28**, 2697–2705 (2009).

[132]L. J. Colwell, M. P. Brenner, and K. Ribbeck, "Charge as a selection criterion for translocation through the nuclear pore complex", PLoS Computational Biology **6**, edited by M. Gilson, e1000747 (2010).

[133]M. Kumeta, H. Yamaguchi, S. H. Yoshimura, and K. Takeyasu, "Karyopherin-independent spontaneous transport of amphiphilic proteins through the nuclear pore", Journal of Cell Science **125**, 4979–4984 (2012).

[134]S. Frey, R. Rees, J. Schünemann, S. C. Ng, K. Fünfgeld, T. Huyton, and D. Görlich, "Surface Properties Determining Passage Rates of Proteins through Nuclear Pores", Cell **174**, 202–217 (2018).

[135]R. Peters, "Translocation through the nuclear pore: Kaps pave the way", BioEssays **31**, 466–477 (2009).

[136]M. P. Rout, J. D. Aitchison, M. O. Magnasco, and B. T. Chait, "Virtual gating and nuclear transport: The hole picture", Trends in Cell Biology **13**, 622–628 (2003).

[137]A. Zilman, S. Di Talia, B. T. Chait, M. P. Rout, and M. O. Magnasco, "Efficiency, selectivity, and robustness of nucleocytoplasmic transport", PLoS Computational Biology **3**, 1281–1290 (2007).

[138]R. Y. H. Lim, B. Fahrenkrog, J. Köser, K. Schwarz-Herion, J. Deng, and U. Aebi, "Nanomechanical basis of selective gating by the nuclear pore complex.", Science (New York, N.Y.) **318**, 640–3 (2007).

[139]R. Y. H. Lim, N.-P. Huang, J. Köser, J. Deng, K. H. A. Lau, K. Schwarz-Herion, B. Fahrenkrog, and U. Aebi, "Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport.", Proceedings of the National Academy of Sciences of the United States of America **103**, 9512–7 (2006).

[140]R. Y. H. Lim, U. Aebi, and B. Fahrenkrog, "Towards reconciling structure and function in the nuclear pore complex", Histochemistry and Cell Biology **129**, 105–116 (2008).

[141]S. Frey, R. P. Richter, and D. Görlich, "FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties.", Science (New York, N.Y.) **314**, 815–7 (2006).

[142]S. Frey and D. Görlich, "A Saturated FG-Repeat Hydrogel Can Reproduce the Permeability Properties of Nuclear Pore Complexes", Cell **130**, 512–523 (2007).

[143]W. Yang, J. Gelles, and S. M. Musser, "Imaging of single-molecule translocation through nuclear pore complexes", Proceedings of the National Academy of Sciences **101**, 12887–12892 (2004).

[144]T. Dange, D. Grünwald, A. Grünwald, R. Peters, and U. Kubitscheck, "Autonomy and robustness of translocation through the nuclear pore complex: A single-molecule study", Journal of Cell Biology **183**, 77–86 (2008).

[145]D. Grünwald, R. H. Singer, and M. Rout, "Nuclear export dynamics of RNA-protein complexes", Nature **475**, 333–341 (2011).

[146]F. Cardarelli, L. Lanzano, and E. Gratton, "Capturing directed molecular motion in the nuclear pore complex of live cells.", Proceedings of the National Academy of Sciences of the United States of America **109**, 9863–8 (2012).

[147]W. Yang and S. M. Musser, "Nuclear import time and transport efficiency depend on importin β concentration", Journal of Cell Biology **174**, 951–961 (2006).

[148]A. Zilman, "Effects of multiple occupancy and interparticle interactions on selective transport through narrow channels: theory versus experiment.", Biophysical journal **96**, 1235–48 (2009).

[149]A. Zilman, S. Di Talia, T. Jovanovic-Talisman, B. T. Chait, M. P. Rout, and M. O. Magnasco, "Enhancement of transport selectivity through nano-channels by non-specific competition.", PLoS computational biology **6**, e1000804 (2010).

[150]R. L. Schoch, L. E. Kapinos, and R. Y. H. Lim, "Nuclear transport receptor binding avidity triggers a self-healing collapse transition in FG-nucleoporin molecular brushes.", Proceedings of the National Academy of Sciences of the United States of America **109**, 16911–6 (2012).

[151]A. R. Lowe, J. H. Tang, J. Yassif, M. Graf, W. Y. Huang, J. T. Groves, K. Weis, and J. T. Liphardt, "Importin-β modulates the permeability of the nuclear pore complex in a Ran-dependent manner", eLife **2015** (2015).

[152]B. Pyhtila and M. Rexach, "A Gradient of Affinity for the Karyopherin Kap95p along the Yeast Nuclear Pore Complex", Journal of Biological Chemistry **278**, 42699–42709 (2003).

[153]M. Borg, T. Mittag, T. Pawson, M. Tyers, J. D. Forman-Kay, and H. S. Chan, "Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity", Proceedings of the National Academy of Sciences **104**, 9650–9655 (2007).

[154]S. Otsuka, S. Iwasaka, Y. Yoneda, K. Takeyasu, and S. H. Yoshimura, "Individual binding pockets of importin-beta for FG-nucleoporins have different binding properties and different sensitivities to RanGTP.", Proceedings of the National Academy of Sciences **105**, 16101–16106 (2008).

[155]J. Ma and W. Yang, "Three-dimensional distribution of transient interactions in the nuclear pore complex obtained from single-molecule snapshots", Proceedings of the National Academy of Sciences **107**, 7305–7310 (2010).

[156]J. Tetenbaum-Novatt, L. E. Hough, R. Mironska, A. S. McKenney, and M. P. Rout, "Nucleocytoplasmic Transport: A Role for Nonspecific Competition in Karyopherin-Nucleoporin Interactions", Molecular & Cellular Proteomics **11**, 31–46 (2012).

[157]H. B. Schmidt and D. Görlich, "Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity", eLife **2015**, 4251 (2015).

[158]N. B. Eisele, S. Frey, J. Piehler, D. Görlich, and R. P. Richter, "Ultrathin nucleoporin phenylalanine-glycine repeat films and their interaction with nuclear transport receptors", EMBO Reports **11**, 366–372 (2010).

[159]L. E. Kapinos, R. L. Schoch, R. S. Wagner, K. D. Schleicher, and R. Y. H. Lim, "Karyopherin-centric control of nuclear pores based on molecular occupancy and kinetic analysis of multivalent binding with FG nucleoporins.", Biophysical journal **106**, 1751–62 (2014).

[160]R. S. Wagner, L. E. Kapinos, N. J. Marshall, M. Stewart, and R. Y. H. Lim, "Promiscuous Binding of Karyopherinβ1 Modulates FG Nucleoporin Barrier Function and Expedites NTF2 Transport Kinetics.", Biophysical journal **108**, 918–27 (2015).

[161]N. B. Eisele, A. A. Labokha, S. Frey, D. Görlich, and R. P. Richter, "Cohesiveness tunes assembly and morphology of FG nucleoporin domain meshworks - Implications for nuclear pore permeability", Biophysical Journal **105**, 1860–1870 (2013).

[162]Y. Caspi, D. Zbaida, H. Cohen, and M. Elbaum, "Synthetic mimic of selective transport through the nuclear pore complex", Nano Letters **8**, 3728–3734 (2008).

[163]S. W. Kowalczyk, L. Kapinos, T. R. Blosser, T. Magalhães, P. Van Nies, R. Y. Lim, and C. Dekker, "Single-molecule transport across an individual biomimetic nuclear pore complex", Nature Nanotechnology **6**, 433–438 (2011).

[164]T. Jovanovic-Talisman, B. T. Chait, and M. P. Rout, "NPC mimics: Probing the mechanism of nucleocytoplasmic transport", Methods in Cell Biology **122**, 379–393 (2014).

[165]F. A. M. Leermakers, Y. S. Jho, and E. B. Zhulina, "Modeling of the 3RS tau protein with self-consistent field method and Monte Carlo simulation", Soft Matter **6**, 5533–5540 (2010).

[166]S. Attili, O. V. Borisov, and R. P. Richter, "Films of end-grafted hyaluronan are a prototype of a brush of a strongly charged, semiflexible polyelectrolyte with intrinsic excluded volume", Biomacromolecules **13**, 1466–1477 (2012).

[167]A. Akinshina, E. Jambon-Puillet, P. B. Warren, and M. G. Noro, "Self-consistent field theory for the interactions between keratin intermediate filaments", BMC Biophysics **6** (2013).

[168]Y. Jung, C. Jeon, J. Kim, H. Jeong, S. Jun, and B. Y. Ha, "Ring polymers as model bacterial chromosomes: Confinement, chain topology, single chain statistics, and how they interact", Soft Matter **8**, 2095–2102 (2012).

[169]M. A. Stuart, W. T. Huck, J. Genzer, M. Müller, C. Ober, M. Stamm, G. B. Sukhorukov, I. Szleifer, V. V. Tsukruk, M. Urban, F. Winnik, S. Zauscher, I. Luzinov, and S. Minko, "Emerging applications of stimuli-responsive polymer materials", Nature Materials **9**, 101–113 (2010).

[170]M. Tagliazucchi and I. Szleifer, "Transport mechanisms in nanopores and nanochannels: can we mimic nature?", Materials Today **18**, 131–142 (2015).

[171]R. D. Coalson, A. Eskandari Nasrabad, D. Jasnow, and A. Zilman, "A Polymer-Brush-Based Nanovalve Controlled by Nanoparticle Additives: Design Principles", Journal of Physical Chemistry B **119**, 11858–11866 (2015).

[172]M. L. Quillin and B. W. Matthews, "Accurate calculation of the density of proteins", Acta Crystallographica Section D Biological Crystallography **56**, 791–794 (2000).

[173]T. Bereau and M. Deserno, "Generic coarse-grained model for protein folding and aggregation", Journal of Chemical Physics **130**, 235106 (2009).

[174]A. Zamyatnin, "Protein volume in solution", Progress in Biophysics and Molecular Biology **24**, 107–123 (1972).

[175]E. B. Zhulina, O. V. Borisov, V. A. Pryamitsyn, and T. M. Birshtein, "Coil-Globule Type Transitions in Polymers. 1. Collapse of Layers of Grafted Polymer Chains", Macromolecules **24**, 140–149 (1991).

[176]A. Halperin, M. Kröger, and E. B. Zhulina, "Colloid-brush interactions: The effect of solvent quality", Macromolecules **44**, 3622–3638 (2011).

[177]L. C. H. Moh, M. D. Losego, and P. V. Braun, "Solvent quality effects on scaling behavior of poly(methyl methacrylate) brushes in the moderate- and high-density regimes", Langmuir **27**, 3698–3702 (2011).

[178]O. Peleg, M. Tagliazucchi, M. Kröger, Y. Rabin, and I. Szleifer, "Morphology control of hairy nanopores", ACS Nano **5**, 4737–4747 (2011).

[179]D. Osmanović, A. Fassati, I. J. Ford, and B. W. Hoogenboom, "Physical modelling of the nuclear pore complex", Soft Matter **9**, 10442 (2013).

[180]L. C. Tu, G. Fu, A. Zilman, and S. M. Musser, "Large cargo transport by nuclear pores: Implications for the spatial organization of FG-nucleoporins", EMBO Journal **32**, 3220–3230 (2013).

[181]A. Sethi, B. Goldstein, and S. Gnanakaran, "Quantifying intramolecular binding in multivalent interactions: A Structure-Based synergistic study on Grb2-Sos1 complex", PLoS Computational Biology **7**, edited by G. Hummer, e1002192 (2011).

[182]R. Moussavi-Baygi and M. R. Mofrad, "Rapid Brownian Motion Primes Ultrafast Reconstruction of Intrinsically Disordered Phe-Gly Repeats Inside the Nuclear Pore Complex", Scientific Reports **6** (2016).

[183]P. Ketterer, A. N. Ananth, D. S. Laman Trip, A. Mishra, E. Bertosin, M. Ganji, J. Van Der Torre, P. Onck, H. Dietz, and C. Dekker, "DNA origami scaffold for studying intrinsically disordered proteins of the nuclear pore complex", Nature Communications **9**, 902 (2018).

[184]P. D. Fisher, Q. Shen, B. Akpinar, L. K. Davis, K. K. H. Chung, D. Baddeley, A. Šarić, T. J. Melia, B. W. Hoogenboom, C. Lin, and C. P. Lusk, "A Programmable DNA Origami Platform for Organizing Intrinsically Disordered Nucleoporins within Nanopore Confinement", ACS Nano **12**, 1508–1518 (2018).

[185]B. Dünweg, D. Reith, M. Steinhauser, and K. Kremer, "Corrections to scaling in the hydrodynamic properties of dilute polymer solutions", Journal of Chemical Physics **117**, 914–924 (2002).

[186]K. A. Dill, "Theory for the folding and stability of globular proteins", Biochemistry **24**, 1501–1509 (1985).

[187]A. Vitalis and R. V. Pappu, "ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions", Journal of Computational Chemistry **30**, 673–699 (2009).

[188]L. Sawle and K. Ghosh, "A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins", Journal of Chemical Physics **143**, 085101 (2015).

[189]Y. H. Lin and H. S. Chan, "Phase Separation and Single-Chain Compactness of Charged Disordered Proteins Are Strongly Correlated", Biophysical Journal **112**, 2043–2046 (2017).

[190]S. Ulrich, A. Laguecir, and S. Stoll, "Titration of hydrophobic polyelectrolytes using Monte Carlo simulations", Journal of Chemical Physics **122**, 134905 (2005).

[191]B. Schuler and H. Hofmann, "Single-molecule spectroscopy of protein folding dynamics-expanding scope and timescales", Current Opinion in Structural Biology **23**, 36–47 (2013).

[192]A. Möglich, F. Krieger, and T. Kiefhaber, "Molecular basis for the effect of urea and guanidinium chloride on the dynamics of unfolded polypeptide chains", Journal of Molecular Biology **345**, 153–162 (2005).

[193]J. C. Schulz, L. Schmidt, R. B. Best, J. Dzubiella, and R. R. Netz, "Peptide chain dynamics in light and heavy water: Zooming in on internal friction", Journal of the American Chemical Society **134**, 6273–6279 (2012).

[194]S. J. Hagen, "Solvent Viscosity and Friction in Protein Folding Dynamics", Current Protein & Peptide Science **999**, 1–11 (2010).

[195]D. Nettels, I. V. Gopich, A. Hoffmann, and B. Schuler, "Ultrafast dynamics of protein collapse from single-molecule photon statistics", Proceedings of the National Academy of Sciences **104**, 2655–2660 (2007).

[196]A. Das and D. E. Makarov, "Dynamics of Disordered Proteins under Confinement: Memory Effects and Internal Friction", Journal of Physical Chemistry B **122**, 9049–9060 (2018).

[197]R. R. Cheng, A. T. Hawk, and D. E. Makarov, "Exploring the role of internal friction in the dynamics of unfolded proteins using simple polymer models", Journal of Chemical Physics **138**, 74112 (2013).

[198]A. Soranno, F. Zosel, and H. Hofmann, "Internal friction in an intrinsically disordered protein - Comparing Rouse-like models with experiments", Journal of Chemical Physics **148**, 123326 (2018).

[199]J. J. Portman, "Non-Gaussian dynamics from a simulation of a short peptide: Loop closure rates and effective diffusion coefficients", The Journal of Chemical Physics **118**, 2381–2391 (2003).

[200]A. Holla, F. Dingfelder, A. Soranno, D. E. Makarov, B. Schuler, and D. Nettels, "Integrated view of internal friction in unfolded proteins from single-molecule FRET, contact quenching, theory, and simulations", Proceedings of the National Academy of Sciences **114**, E1833–E1839 (2017).

[201]V. Calandrini, E. Pellegrini, P. Calligari, K. Hinsen, and G. Kneller, "nMoldyn - Interfacing spectroscopic experiments, molecular dynamics simulations and models for time correlation functions", École thématique de la Société Franaise de la Neutronique **12**, 201–232 (2011).

[202]B. Dunweg and K. Kremer, "Microscopic verification of dynamic scaling in dilute polymer solutions: A molecular-dynamics simulation", Physical Review Letters **66**, 2996–2999 (1991).

[203]T. T. Pham, U. D. Schiller, J. R. Prakash, and B. Dünweg, "Implicit and explicit solvent models for the simulation of a single polymer chain in solution: Lattice boltzmann versus brownian dynamics", Journal of Chemical Physics **131**, 164114 (2009).

[204]B. Dünweg and K. Kremer, "Molecular dynamics simulation of a polymer chain in solution", The Journal of Chemical Physics **99**, 6983–6997 (1993).

[205]M. O. Steinhauser, "A molecular dynamics study on universal properties of polymer chains in different solvent qualities. Part I. A review of linear chain properties", Journal of Chemical Physics **122**, 94901 (2005).