Computational Studies of Macromolecular Crowding Effects on Proteins

by

Supervisor: Dr. Stefan Wallin

Committee members:

Dr. G. Todd Andrews

Dr. James LeBlanc

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Physics and Physical Oceanography

Faculty of Science

Memorial University of Newfoundland

October 2023

St. John's Newfoundland and Labrador

Abstract

The high concentration of macromolecules inside living cells leads to an effect called macromolecular crowding. One very basic consequence of crowding is a decrease in the volume physically available to the molecules in the solution. This consequence is called the excluded volume effect. It has been observed experimentally that the excluded volume effect impacts various processes in cells. In this thesis, we investigate the effect of macromolecular crowding on two different large-scale conformational transitions in proteins: folding, which is the process by which proteins become functional and achieve their native state; and fold switching, a process in which a protein exhibits two or more native states and reversibly interconverts between them. To address these issues, we develop and apply coarse-grained models at various levels of resolution. In particular, we use two different models. 1. A sequence-based model with 7 atoms per amino acid in which folding is driven by effective hydrophobic interactions and hydrogen bonding. 2. A model with one bead per amino acid, with a structure-based (Gō-like) potential, which provides an energetic bias towards two or more native states. Sampling of conformational space is performed using Monte Carlo techniques and Langevin dynamics. A long standing assumption in the crowding field is that the excluded volume effect always stabilizes the native states of proteins. However, by using our sequence-based model, we find this crowding effect can be destabilizing in some cases depending on the protein and crowding condition. To study crowding effects on fold switching, which has not been done before, we focus on the G_A/G_B fold switch system based on the 56 amino acid binding domain of protein G. This system is a designed fold switch system in which amino acid mutations drive the

switch between the two different folds adopted by G_A and G_B. We find that crowding impacts the population balance between the two different folds. Specifically, we find that crowding enhances the stability of G_B relative to G_A. Overall, this thesis provides insight into how crowding effects on proteins depend on factors such as protein fold, types of interactions between crowders and proteins, crowders concentration, and conformational landscape of the protein, further advancing our understanding of the intricate interplay between cellular environments and protein behavior.

General Summary

Protein folding is an essential process in biology where proteins, made up of chains of amino acids, acquire their unique three-dimensional shapes as their native state. The correct folding of proteins is essential to carry out their specific functions in our bodies. However, the crowded environment inside cells, filled with numerous other biomolecules, can impact protein folding. Macromolecular crowding refers to the high concentration of various macromolecules within cells. These molecules can take up space and create a crowded and inhomogeneous environment around proteins. This crowding effect can influence how proteins fold and adopt their native structures and also how they interconvert between native states reversibly in the case of metamorphic proteins. The presence of macromolecules in close proximity to folding proteins can either facilitate or hinder the folding process depending on temperature, crowders' volume fractions, and interactions. Studying the effects of macromolecular crowding on protein folding is essential to gain a deeper understanding of how proteins behave in their natural cellular environment. Scientists use experimental techniques such as nuclear magnetic resonance (NMR) and computational simulations to investigate how various factors, including macromolecular crowding, can affect proteins' structures, thermodynamics and kinetics. To gain a comprehensive understanding of the effects of macromolecular crowding on protein behavior, it is crucial to have a thorough knowledge of the techniques mentioned above. As technology continues to advance, we can anticipate further insights into the fascinating nature of proteins in the cellular environment. This progress will contribute to our overall comprehension of protein structure and function, enhancing our understanding of the crucial role of proteins in

living organisms.

Co-authorship Statement

Chapter 3 is based on the expanded and corrected version of the following paper:
Saman Bazmi and Stefan Wallin, "Crowding-induced protein destabilization in the absence of soft attractions", Biophysical Journal 121, (2022): 2503-2513.
The study and model are designed by Dr. Stefan Wallin. Dr. Stefan Wallin and I carried out simulations, and analyses, and wrote the manuscript.

Chapter 4 is based on the following paper:
Saman Bazmi and Stefan Wallin, "Exploring soft interactions in crowded systems: repulsive, nonspecific, hydrophobic and polypeptide crowders", Preprint.
Dr. Stefan Wallin and I designed the study and developed the model. I carried out simulations and analyses for spherical crowders. Dr. Stefan Wallin carried out simulations and analyses for peptide crowders. Dr. Stefan Wallin and I wrote the manuscript.

Chapter 5 is based on the following paper:
Saman Bazmi and Stefan Wallin, "Crowding effects on protein folding thermodynamics and kinetics: polymeric versus spherical crowders", Preprint.
Dr. Stefan Wallin and I designed the study and developed the model. I carried out simulations and analyses. Dr. Stefan Wallin and I wrote the manuscript.

Chapter 6 is based on the following paper:

Saman Bazmi, Bahman Seifi, and Stefan Wallin, "Simulations of a protein fold switch reveal crowding-induced population shifts driven by disordered regions." Communications Chemistry 6, 191 (2023).

In this project, I contributed with expertise on macromolecular crowding and Dr. Bahman Seifi contributed with expertise on fold switching. Overall, Dr. Bahman Seifi and I contributed equally. Dr. Stefan Wallin guided and supervised this project and contributed to developing the Cα model. Dr. Bahman Seifi and I analyzed all the simulations and prepared figures. Dr. Stefan Wallin, Dr. Bahman Seifi, and I wrote the manuscript.

During my PhD program, I am also contributed to the following paper:

Venketesh Thrithamara Ranganathan, Saman Bazmi, Stefan Wallin, Yun Liu, and Anand Yethiraj. "Is Ficoll a Colloid or Polymer? A Multitechnique Study of a Prototypical Excluded-Volume Macromolecular Crowder", Macromolecules (2022): 55, 9103–9112.

Acknowledgements

I am immensely grateful for the exceptional guidance and support provided by my supervisor, Dr. Stefan Wallin. Throughout the journey of completing this thesis, Dr. Wallin's unwavering belief in my abilities, expert knowledge, and invaluable insights have been instrumental in shaping the quality of my research. His dedication to excellence, patience, and willingness to go above and beyond have inspired and motivated me to push the boundaries of my capabilities. Working under his mentorship has been a privilege, and I am deeply appreciative of the transformative impact he has had on my academic growth.

I would also like to thank my committee members, Dr. James LeBlanc and Dr. Todd Andrews, for their support and helpful comments. Additionally, I extend my gratitude to Dr. Bahman Seifi, a member of the computational biophysics lab and my collaborator on the "Fold-Switching Proteins Under Crowded Conditions" project.

I am thankful to the Memorial University of Newfoundland for the financial support which enabled me to concentrate on my research and writing. I also appreciate everyone at the Department of Physics and Physical Oceanography for creating a pleasant working environment.

I would like to extend my heartfelt acknowledgment and deep appreciation to my parents, as well as my sister, and my friends Dr. Mohammad Mohammadi Baghmolaei and Ms. Mahin Hashemi for their unwavering support, unwavering belief in my abilities and invaluable guidance throughout my life.

Contents

Abstract **ii**

General Summary **iv**

Co-authorship Statement **vi**

Acknowledgements **viii**

List of Tables **xv**

List of Figures **xvi**

1 Introduction **1**

1.1 Macromolecular crowding effects 4

1.2 Proteins: folding, fold switching, and structural disorder 6

1.3 Molecular driving forces 9

1.3.1 Hydrogen bonding 9

1.3.2 Hydrophobic effect 10

1.4 Quantifying native state stability 11

1.5 Artificial polymers 12

1.6 Previous studies on the effect of crowding on protein folding and stability 15
1.7 Outline 16

2 Theory and Methods 28

2.1 Statistical mechanics 29
2.2 Molecular dynamics 31
2.3 Langevin dynamics 32
2.4 Metropolis-Hastings Monte Carlo 34
2.5 Simulated tempering 36
2.6 Coarse-graining in molecular simulations 38
2.6.1 C_β model 39
2.6.2 C_α model 41
2.7 Scaled particle theory 42

3 Crowding-induced protein destabilization in the absence of soft attractions 50

3.1 Introduction 53
3.2 Materials and Methods 56
3.2.1 Excluded volume crowders 56
3.2.2 Coarse-grained model for protein folding in the presence of crowders 57
3.2.3 Simulated tempering Monte Carlo 60
3.2.4 Simulation and analysis details 61
3.2.5 Observables 62
3.3 Results 63

3.3.1 Native structures are not changed by crowders 63
3.3.2 Nonnative interactions are promoted under crowded conditions 66
3.3.3 Excluded volume crowders can both increase and decrease native state stability 67
3.3.4 Crossover temperature 70
3.3.5 Compact β_{25} unfolded state allows for crowding induced destabilization at low T 72
3.3.6 How is β_{25} destabilized by crowders at T^+? 75
3.3.7 Apparent stabilization effect is observable dependent 79
3.4 Discussion 79
3.5 Conclusion 83
3.6 Acknowledgements 84

4 Exploring soft interactions in crowded systems: repulsive, nonspecific, hydrophobic and polypeptide crowders 95

4.1 Introduction 98
4.2 Methods 101
4.2.1 Coarse-grained model for protein folding 101
4.2.2 Crowders 104
4.2.3 Simulations and analysis details 106
4.2.4 Observables 107
4.3 Results 108
4.3.1 Spherical crowders 108

4.3.2 How can the stronger destabilizing effect of weak hydrophobic attraction be explained? 110
4.3.3 Balancing destabilizing soft interactions and stabilizing steric repulsions 115
4.3.4 Polypeptide crowders 117
4.4 Discussions and Conclusions 120

5 Crowding effects on protein folding thermodynamics and kinetics: polymeric versus spherical crowders 130

5.1 Introduction 132
5.2 Material and Methods 135
5.2.1 Native structures and contact maps 135
5.2.2 Coarse-grained model for protein folding 135
5.2.3 Excluded volume crowders 136
Spherical crowders 136
Polymeric crowders 137
5.2.4 Langevin Dynamics 138
5.2.5 Simulations and analysis details 139
5.2.6 Observable 140
5.3 Results 140
5.3.1 Polymeric crowders: size and volume exclusion 140
5.3.2 Native structures are not perturbed by polymeric or spherical crowders 143
5.3.3 Both polymeric and spherical crowders enhance stability . . . 144

5.3.4 Size dependence of stability for spherical and polymeric crowders 146
5.3.5 Folding rate enhanced by crowding agents 152
5.4 Discussion and Conclusions 155

6 Simulations of a protein fold switch reveal crowding-induced population shifts driven by disordered regions 167

6.1 Introduction 169
6.2 Model and Methods 170
6.2.1 Native structures and contact maps 170
6.2.2 Observables 171
6.2.3 Coarse-grained model for protein fold switching 171
6.2.4 Excluded volume crowders 172
6.2.5 Langevin dynamics 173
6.2.6 Simulation and analysis details 174
6.2.7 Theory 174
6.3 Results 176
6.3.1 Mimicking the mutational pathway between the G_A and G_B folds 176
6.3.2 Macromolecular crowding effects on the G_A/G_B fold switch . . 178
6.3.3 Disordered tails control the crowding effect on the fold switch 181
6.3.4 Comparing with crowding effects on single-fold proteins 183
6.3.5 The unfolded state changes character across the fold switch . . 184
6.4 Discussion 188
6.5 Acknowledgments 190
6.6 Authorship contributions 190

6.7 Code availability 191

7 Summary and Outlook **199**

A Supporting Information for Chapter 3 **205**

B Supporting Information for Chapter 4 **209**

C Supporting Information for Chapter 6 **211**

C.1 Development of computational model for the GA/GB fold switch . . 211

C.1.1 Single-basin structure-based model for protein folding 211

C.1.2 Dual-basin structure-based model for fold switching 215

C.2 Supplementary figures 217

List of Tables

3.1 Amino acid model sequences. 55

5.1 Radius of gyration, R_g, for isolated polymer crowder chains for different numbers of beads, L_{ch}. 142

List of Figures

1.1 The excluded volume effect 4

1.2 Amino acid structure 6

3.1 Folding curves and native structures. 65

3.2 Formation of nonnative interactions under crowded and non-crowded conditions. 67

3.3 Effect of crowding on native state stability and folding cooperativity. 69

3.4 Existence of a crossover temperature 71

3.5 Effect of crowders on the size of the native and unfolded states. . . . 73

3.6 Radius of gyration histograms 77

3.7 Compact nonnative state. 78

3.8 Impact of crowding on secondary structure content and end-to-end distance. 80

4.1 Schematic of crowders molecule interaction with the protein 102

4.2 Schematics of the pair potential for crowder-crowder and crowder-atom interactions. 105

4.3 The free energy of folding for α_{35} 108

4.4 Effect of crowding on the native state stability of α_{35} 111
4.5 Energetic stabilization of the unfolded state 112
4.6 Effect of hydrophobic crowders on the native state stability. 114
4.7 Native state stability and polypeptide crowders 116
4.8 Protein and peptide crowders interaction 118

5.1 Scaling of polymer chain size and crossover volume fraction 142
5.2 Snapshots of 1SHF and 1PRB with polymeric crowders 143
5.3 Folding curves and free energy profiles. 145
5.4 Midpoint temperature and folding free energy changes for 1PRB. . . 147
5.5 Midpoint temperature and folding free energy changes for 1SHF. . . . 148
5.6 Midpoint temperature changes for 1PRB and 1SHF. 149
5.7 Size dependency of stability for 1PRB and 1SHF. 150
5.8 Folding kinetic of 1PRB and 1SHF. 154
5.9 Folding kinetic size dependency of 1PRB and 1SHF. 156

6.1 The G_A/G_B fold-switch system 177
6.2 Crowding effects on absolute and relative fold populations in the G_A/G_B fold switch. 180
6.3 Disordered tail segments control crowding effects on the fold switch. . 182
6.4 Crowding effects on G_A and G_B single-fold proteins. 185
6.5 Changes to the unfolded state character across the fold switch. 187

A.1 Mittal and Best pair potential 206
A.2 Heat capacity 206

A.3 Two-state fits of folding curves. 207

A.4 Population of native state curves 208

B.1 Two-state fits of folding curves. 210

C.1 Non-terminal positions contacts 213

C.2 Dual basin pair potential 214

C.3 Merging potentials 217

C.4 G_A, G_B, and total G_A and G_B (P_{tot}) fold populations and κ^*. 217

C.5 Crowder concentration (ϕ_c) dependence of the free energy of fold switching 218

C.6 Free energy as function of the number of G_A and G_B contacts 219

C.7 Pair correlation function 220

Chapter 1

Introduction

Proteins are fundamental components of living organisms and play essential roles in cellular processes. They have many functions, including catalyzing chemical reactions [1], providing structural support to cells and tissues [2], and signaling between cells [3]. The amino acid sequence of a protein molecule determines its chemical and physical properties [4]. A variety of methods have been used to elucidate the structures and functions of these molecules, including nuclear magnetic resonance spectroscopy and X-ray crystallography.

Proteins, as linear polymer chains, exhibit a wide range of lengths. The typical length of a protein chain can vary from a few dozen amino acids to several thousand amino acids [4]. For example, insulin is a protein with 51 amino acids [5], while the protein titin, found in muscle cells, contains over 30,000 amino acids [6]. The amino acid sequence of a protein chain is intimately linked to its structure and function, as

different amino acid sequences fold into distinct three-dimensional shapes, enabling proteins to carry out specific tasks within the cell [4].

Proteins also display a remarkable diversity of structures, which can be used to classify them into different types. One such classification results in three main categories: globular, membrane, and fibrous proteins [7]. Globular proteins, as the name suggests, fold into compact, roughly spherical shapes and are typically found in the aqueous environment of cells, where they perform a wide range of functions, such as transporting oxygen [8], regulating blood sugar levels [9], and enzymatic activity [10]. Membrane proteins are embedded within cell membranes. Some have transmembrane structures, allowing them to interact with other biomolecules on either side of the membrane [11]. Lastly, fibrous proteins have the unique ability to spontaneously form large complex structures, such as amyloid fibrils, which are associated with certain diseases, including Alzheimer's and Parkinson's [12, 13]. Understanding the diverse types of proteins and their characteristic structures provides valuable insights into their biological roles and opens avenues for therapeutic interventions. Hemoglobin is an example of a globular proteins that has been extensively studied and has played a historic role in the development of the field of protein science [14–16].

In recent decades, computer simulations have emerged as a powerful tool for studying the structure and function of proteins and protein complexes. By using computational methods, we can gain insight into the detailed dynamics of these complex biomolecules in a way that is difficult to achieve experimentally. Computer simulations of biomolecules can be carried out with either coarse-grained models or detailed atomistic models. Coarse-grained models simplify the representation of the

protein structure, such as reducing atomic groups into a single bead or interaction site. Coarse-grained models have the advantage of being computationally more efficient and can access longer timescales, making them ideal for studying large systems or complex biological processes. On the other hand, detailed atomistic models provide more accurate representations of the interactions but are computationally more intensive and limit simulations to shorter timescales.

Most experiments in protein research have traditionally been carried out in dilute solutions in *vitro*. Similarly, computer simulations typically involve only a single protein or a small number of particles in the simulation box. However, the cellular environment where proteins function and evolve is inhomogenous and crowded, affecting components' physical and chemical properties, particularly proteins. While much has been learnt from studying isolated proteins, ultimately, proteins must be understood within the context of the crowded and complex cellular environment. Macromolecules such as proteins, carbohydrates, ribosomes, and nucleic acids are the molecules responsible for making this crowded intracellular environment [17]. While the concentration of individual macromolecules may be low, the total concentration in cells is remarkably high, making a fully complex cellular environment [18]. For example, the concentration of RNA and protein in the *Escherichia coli* cytosol is in the range 300-400 g/L and can reach volume fractions of around 30-40% of the cell interior volume [17]. This environment influences protein processes, such as protein folding and proteins' functions [19]. The presence of macromolecular crowders of different shapes and sizes increases the degree of complexity of these processes, which are already complex.

1.1 Macromolecular crowding effects

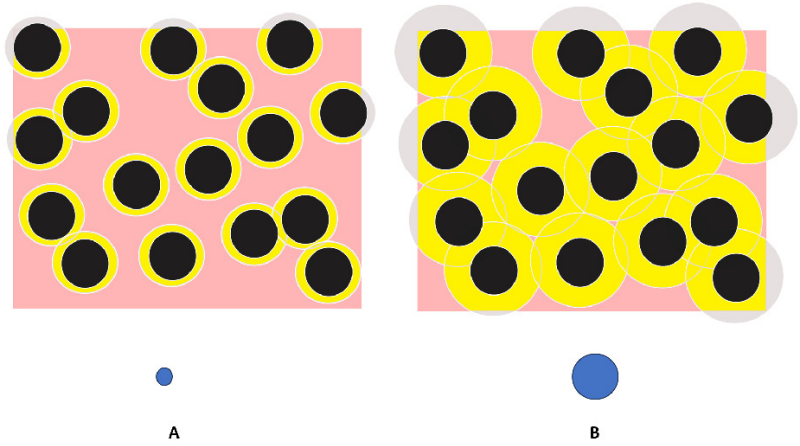

Figure 1.1: The excluded volume effect. Two proteins, A and B, are inserted into a volume (square) that contains crowding particles (black circles). Protein A is smaller than the crowders, while protein B is comparable in size to the crowders. The yellow area represents the volume not accessible to the center of mass of the protein that is being inserted. Note that the volume that is free (pink) depends on the size of the inserted protein. The figure is adapted from [20].

The effects of macromolecular crowding on the behavior of proteins can be divided into different classes based on the type of interaction between proteins and crowders. The most fundamental influence of crowders, which refer to all macromolecules except the protein under study, is to remove some of the volume that would otherwise be accessible to the protein. This excluded volume effect is the result of hard-core repulsion and is unavoidable under crowded conditions because it results from the impenetrability of atoms on the macromolecular crowding agents and the protein. It is always present independently of the magnitude of additional types of interactions.

The excluded volume effect is expected to stabilize the native state of a protein. Many experimental studies have sought to test the impact of the excluded volume effect on protein stability by using artificial polymers as crowding agents. The addition of such crowders indeed stabilizes proteins' native state [21, 22]. However, the effect is often weak, and a few exceptions have been found [23, 24].

It can easily be understood why hard-core repulsive interactions between crowders and the protein should be stabilizing [25]. The native state is typically more compact than the extended unfolded state and occupies a smaller volume. Therefore, the native state becomes entropically favored relative to the unfolded state when the protein is immersed in a solution that contains crowding agents that occupy space [21]. As illustrated in Figure 1.1, the folded state of a protein (A) is easier to immerse in a crowded environment rather than an unfolded state (B) because the volume available decreases with the size of the protein molecule.

Weak attractive interactions between proteins and other macromolecules can also be important [23, 24]. In the case of nonspecific interactions, crowders can interact with both the unfolded and the native states. It is expected that the unfolded state has a larger accessible surface area than the native state. Hence, the effect of this kind of interaction is destabilizing because it results in a net energetic stabilization of the unfolded state. If crowders interact favorably only with the native state, the effect will be stabilizing.

In this thesis, we have studied the effects of macromolecular crowding on two different conformational transformations in proteins: folding and fold switching. To this

end, we develop coarse-grained models for proteins that we will discuss in Chapter 2. Protein folding and fold switching will be discussed in the following.

1.2 Proteins: folding, fold switching, and structural disorder

Proteins are linear heteropolymer chains. The monomers that make them up are the 20 naturally occurring amino acids. All amino acids have the same basic chemical structure: a central carbon atom called C_α, which is covalently bonded to a carboxyl group (–COOH), a hydrogen atom, an amino group ($-NH_2$), and a side chain group, often denoted R, as illustrated in Figure 1.2. The covalent bonds that hold amino

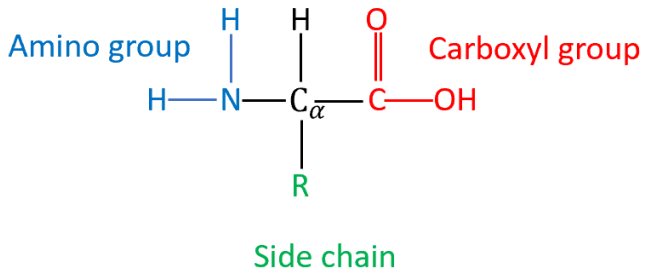

Figure 1.2: Amino acid structure. The basic structure of an amino acid includes amino group, carboxyl group, side chain, and central carbon.

acids together in a protein chain are called peptide bonds. For this reason, proteins are referred to as polypeptide chains. When an amino acid is incorporated into a polypeptide chain, it is often referred to as an amino acid residue.

The side chain group is responsible for the distinct physio-chemical properties of the amino acids. This group is the only difference in structure between the 20 or so amino acids that except proline, in which the side chain is connected covalently to both C_α and the nitrogen atom, forming a five-membered nitrogen-containing ring [26]. This unique cyclic structure can affect protein stability, folding, and function [27]. In order for proteins to become functional and biologically active, they need to adopt a unique three-dimensional structure called the native structure. The process in which a protein chain acquires its native structure is known as folding. Even under ideal conditions, small structural fluctuations always occur around the native structure, and this ensemble of native structures is often called the native state. According to Anfinsen's dogma [28], the native state is the global free energy minimum of the system consisting of the protein and the surrounding solvent molecules. Hence, at least for small globular proteins, the native structure is determined only by the protein's amino acid sequence [29]. This makes a "sequence-structure" relationship for proteins.

Proteins have structures at different levels. Primary structure refers to the sequence of amino acids that constitute the protein. Secondary structure refers to commonly occurring structural motifs, which include α-helices and β-sheets. α-helices are local motifs, i.e., they involve amino acids close in sequence, while β-sheets can be either local or involve non-local chain segments. For example, α-helices are stabilized by hydrogen bonds between the CO group of amino acid i, and NH group of amino acid i+4. For β-sheets, hydrogen bonds can occur between amino acids with larger separation sequences. The tertiary structure designates the organization of the sec-

ondary structure elements. The arrangement of multiple protein chains or subunits into a yet higher-level structure is known as a protein's quaternary structure. Each subunit of a protein at this level contains its unique primary, secondary, and tertiary structures. There is also a higher level of structures up to the quinary structure, which refers to the organization of proteins in highly dense solutions, such as in the cell interior [30].

In addition to proteins with only a single fold, there are metamorphic (fold-switching) proteins. Metamorphic proteins are a class of proteins that have the unique ability to adopt multiple stable and functionally distinct native conformations [31]. These conformational changes often occur reversibly in response to environmental cues such as changes in temperature, pH, or the presence of ligands or other proteins. This ability enables them to perform various biological functions [32, 33]. For example, a protein fold switch operates in the circadian oscillator clock in cyanobacteria [34]. Other examples of metamorphic proteins are Lymphotactin, Mad2 spindle checkpoint protein (Mad2) [35], Chloride intracellular channel 1 (CLIC1) protein [32], protein RfaH [36], and Protein G [37].

In this thesis, we focus on protein G and study macromolecular effects on the thermodynamics of their different folds. Protein G has two binding domains called G_A and G_B. These two domains have similar lengths, around 56 amino acids, but entirely different structures (folds). It was shown that by applying sequence mutations, an abrupt switch between these two folds can be obtained [38]. We study the fold-switching and thermodynamics of protein G in different temperatures in the presence of crowders with various volume fractions.

Furthermore, there are some biologically active and functional proteins without any stable native structure, so-called intrinsically disordered proteins (IDPs) [39]. They can be totally or partly unstructured. Intrinsically disordered regions (IDRs) in protein sometimes adopt a well-defined structure when they bind to other macromolecules [40]. Interestingly, metamorphic proteins often include IDRs[37].

1.3 Molecular driving forces

Proteins are only marginally stable, and it is estimated that the free energy difference between the native state of a protein and its unfolded state is usually in the range of 5-15 $k_{\mathrm{B}}T$ [41, 42]. Proteins achieve their stability by various forces and effects. Below, we will discuss two important driving forces in protein folding processes: backbone-backbone hydrogen bonds and hydrophobic interactions. Other forces include electrostatic interactions between partial charges and van der Waals interactions, which are weak interactions that arise due to fluctuations in electron densities around atoms [43].

1.3.1 Hydrogen bonding

A hydrogen bond can be understood as an electrostatic interaction in which a hydrogen atom is shared by two electronegative atoms; the hydrogen bond donor and the hydrogen bond acceptor. The hydrogen bond donor is an atom that is covalently bonded to the hydrogen. There is partial positive charge on the hydrogen atom due

to the electronegativity difference between the hydrogen and the donor atom. The hydrogen bond acceptor likewise has a partially negative charge. Aligning two electronegative atoms and hydrogen along a line is electrostatically favorable, which is the hydrogen bond. An important example for protein folding is the hydrogen bond formed between the NH and CO groups of the protein backbone (see Fig 1.2), where N is the donor, and O is the acceptor. The backbone-backbone hydrogen bonds are responsible for stabilizing α-helices and β-sheets. The energy of a fully formed such hydrogen bond is about 2-10 kcal/mol [44, 45]. The accepted (and most frequently observed) geometry for a hydrogen bond is a distance of less than 2.5 Å between the hydrogen and the acceptor and a donor-hydrogen-acceptor angle of between 90° and 180° [43, 46].

1.3.2 Hydrophobic effect

Another interaction that is important in protein folding is the hydrophobic effect. The hydrophobic effect is the tendency for nonpolar molecules to minimize contact with water. Nonpolar molecules cannot participate in the network of hydrogen bonds that is always present in bulk water, and, as a result, they dissolve poorly. Hence, nonpolar molecules tend to come together, and it leads to an effective attraction between them in water. The hydrophobic interaction is entropy-driven because water molecules close to nonpolar surfaces are more ordered than water molecules in bulk. As a result, when two nonpolar molecules come together, the total surface area exposed to the water molecules will be reduced, and the total entropy increases.

Some amino acids' side chains are hydrophobic, and these amino acids play an essential role in protein folding. Because of their non-polar character, hydrophobic side chains are typically found in the interior of a protein structure, where they form a hydrophobic core. Once the protein structure has been formed, the nonpolar core of the protein can be further stabilized through attractive van der Waals forces. Because the hydrophobic effect is entropically driven, the strength of this interaction decreases with the temperature [47], and this is probably the major cause of cold-denaturation in proteins [48]. In addition to the temperature, the strength of hydrophobic interactions depends on several other factors, such as the shape and size of the non-polar molecules [43, 49].

1.4 Quantifying native state stability

Proteins have different thermodynamic properties, which can be impacted by macromolecular crowders. One of the most important of these properties is the stability of the native state. Maintaining protein stability is important in vivo because proteins need to cope with changes in environmental cellular conditions such as salt concentration, pH, and temperature [50]. An understanding of protein stability is also essential for practical purposes, e.g. optimizing protein expression, storage, and structural studies [51]. One way to quantify the native state stability is the difference in the (Helmholtz) free energy of the folded state, F_{f}, and the free energy of the unfolded state, F_{u}, $\Delta F_{\mathrm{f}} = F_{\mathrm{f}} - F_{\mathrm{u}}$. The quantity ΔF_{f} is called the folding free energy, which indicates whether the protein is stable ($\Delta F_{\mathrm{f}} < 0$) or unstable ($\Delta F_{\mathrm{f}} > 0$). In compari-

son with covalent bonds (50-150 $k_{\mathrm{B}}T$), ΔF_{f} of globular proteins are relatively small (5-15 $k_{\mathrm{B}}T$) [52, 53]. The folding free energy can be divided into an energy component (ΔE_{f}) and an entropy component (ΔS_{f}) via the equation:

$$\Delta F_{\mathrm{f}} = \Delta E_{\mathrm{f}} - T\Delta S_{\mathrm{f}}. \tag{1.1}$$

Depending on whether the crowders are purely repulsive or include nonspecific attractive interactions, the change in the stability can originate from the entropic or energetic components, respectively. ΔF_{f} is the natural quantity in simulations with a fixed volume. Experiments are typically carried out under fixed pressure conditions. For that reason, experimental papers typically report Gibbs folding free energies, ΔG_{f}, in which the enthalpy (ΔH_{f}) takes the role of ΔE_{f}. However, because volume changes in protein solutions are often negligible, the difference between ΔG_{f} and ΔF_{f} can be assumed to be small.

Another way to measure the stability of the native state is the melting (midpoint) temperature T_{m}, which is defined as the temperature in which the native and unfolded states are equally populated. It should be mentioned that increasing the melting point indicates a higher amount of stability.

1.5 Artificial polymers

In general view, polymers are molecules composed of basic units called monomers, which are covalently bonded together in a repeated fashion. In the realm of biology, many biomolecules found in our cells are polymers. Proteins, polysaccharides, and

nucleic acids are prominent examples of these natural polymers. Proteins, in particular, are naturally occurring polymers that are composed of long chains of amino acids, with each amino acid serving as a monomer unit. Beyond the realm of biology, artificial polymers have been created, including those used in the production of plastics [52, 54]. Polymer-based materials are a major focus of soft-condensed matter physics [52].

Polymers can be divided into different categories. Homopolymers are the simplest form of polymers because they are made up of a single type of monomer. If polymers are composed of more than one type of monomer, they heteropolymers. Furthermore, based on the properties of the monomers and the way that they are synthesized, polymers may be linear or branched. In the former, monomers are attached to each other, forming a linear structure; in the latter, the structure can include branched points. For example, proteins are linear heteropolymers.

In many experimental studies of macromolecular crowding, Ficoll has been used to mimic an artificial crowding agent [55–57]. Ficoll is a neutral and highly branched homopolymer that dissolves easily in aqueous solutions and is made up of repeating units, sucrose. Ficoll is often assumed to be an inert, hard, and spherical crowder. As such it would be an ideal molecule to study the excluded volume effect. However, recent studies by Fissel *et al.* show that Ficoll is a soft and non-spherical particle rather than a hard-sphere crowder [58, 59].

The size of a polymer can be quantified theoretically in different ways. One way to characterize the size of a linear polymer chain is to consider the polymer as a random

walk. A random walk polymer is a theoretical model that describes the conformation of the chain as a series of randomly oriented links $\vec{a}_{\mathrm{i}}$ that connects monomer i to the following monomer i+1. Each link $\vec{a}_{\mathrm{i}}$ can point in any direction. The end-to-end vector $\vec{R}_{\text{e-e}}$ of the polymer chain is the displacement from the first to the last monomer. Accordingly,

$$\vec{R}_{\text{e-e}} = \sum_{\mathrm{i}=1}^{N} \vec{a}_{\mathrm{i}}, \tag{1.2}$$

where N is the number of links. When the directions of different links are uncorrelated, we obtain the average squared end-to-end distance:

$$\langle R_{\text{e-e}}^2 \rangle = Na^2, \tag{1.3}$$

where a is the length of the links $\vec{a}_{\mathrm{i}}$. It means the overall size of a random walk polymer with N monomers is proportional to $\sqrt{N}$. For branched polymers, a unique end-to-end distance cannot be defined. A quantity that can be used to characterize the size of both branched and linear chains is the radius of gyration R_{g}, defined by:

$$R_{\mathrm{g}}^2 = \frac{1}{N} \sum_{\mathrm{i}=1}^{N} \left(\vec{R}_{\mathrm{i}} - \vec{R}_{\mathrm{cm}} \right)^2, \tag{1.4}$$

where $\vec{R}_{\mathrm{i}}$ is the position of monomer i, m_{i} is the mass of monomer i, and

$$\vec{R}_{\mathrm{cm}} = \frac{\Sigma_{\mathrm{i}} m_{\mathrm{i}} \vec{R}_{\mathrm{i}}}{M_{\mathrm{polymer}}} \tag{1.5}$$

is the polymer's center of mass, where

$$M_{\mathrm{Polymer}} = \sum_{\mathrm{i}=1}^{\mathrm{N}} m_{\mathrm{i}}, \tag{1.6}$$

is the total polymer's mass. An interpretation of R_{g}^2 is the average squared distance between the monomers and the center of mass [52].

For polymer solutions, there are other relevant properties that can be quantified, e.g., the volume fraction ϕ, which is the ratio of the volume V_{Polymer} occupied by the polymer in the solution and the total volume V_{Solution} of the solution. Then, we have:

$$\phi = \frac{V_{\text{polymer}}}{V_{\text{solution}}} = N\frac{V_{\text{monomer}}}{V_{\text{solution}}}, \tag{1.7}$$

where V_{monomer} is the volume of a monomer.

1.6 Previous studies on the effect of crowding on protein folding and stability

Many experimental studies have investigated the effects of macromolecular crowding on protein folding. Minton was the first to point out that volume exclusion of crowder molecules should have stabilizing effects on proteins [60]. By using artificial crowders such as Ficoll, which are often assumed not to interact strongly with proteins, Minton's prediction was confirmed experimentally [61–63]. However, protein molecules used as crowding agents are destabilizing [64]. This is explained by weak nonspecific attractive interaction between crowders and protein, which counteracts the stabilizing excluded volume effect [65–67]. In general, the net effect of macromolecular crowding on the stability of proteins will be determined by a competition between the stabilizing effect of excluded volume and the destabilizing nonspecific attractions.

Because attractive interactions have an energetic (or enthalpic) component, the impact of crowders should be temperature-dependent. Indeed, this was observed by

Zhou *et al.*, who studied the stability of the protein CI2 in the presence of protein crowders [22]. CI2 was found to be destabilized at low temperatures and stabilized at high temperatures. This allowed the authors to define a so-called crossover temperature at which there was no net effect on the stability of CI2 [68].

The effects of crowding on protein stability have also been studied by computer simulations. Mittal *et al.* have shown that crowding increased the free energy of the unfolded state relative to that of the folded state [63]. In one study, Cheung *et al.* showed that purely repulsive macromolecular crowding enhances the native state stability for the all-β WW domain in the presence of large spherical particles [61]. Computer simulations were also used to analyze the effect of crowding agents on the folding cooperativity. Tsao *et al.* found that macromolecular crowding induces polypeptide compaction and decreases folding cooperativity [69]. They found the folding temperature increases with increasing volume fraction of crowders, but the height of the heat capacity peak decreases. This height is a measure of folding cooperativity [70].

1.7 Outline

This thesis focuses on the effects of macromolecular crowding on protein processes, mainly protein folding and protein fold switching. Chapter 1 and Chapter 2 are introductions to the literature and the methods we are using. Chapter 1 includes macromolecular crowding effects concepts and protein structures, and Chapter 2 includes statistical and computational physics concepts, which underlie Monte Carlo

and molecular dynamics simulation techniques.

Chapter 3 describes a research project that investigates the thermodynamic features of proteins in the presence of crowding agents. We investigate the effect of crowding on the folding of three model proteins with different native structures. The main question is how purely repulsive crowding agents can impact these proteins in terms of thermodynamic properties such as native state stability, folding cooperativity, nonnative contacts, and size of proteins. The basic approach is a coarse-grained simulation model for proteins [40] and crowders. Despite long-standing expectations that excluded volume crowders stabilize the native state, we found that this effect can be destabilizing, for some protein folds at high enough crowder concentrations.

Having studied the excluded volume effect in Chapter 3, we turned to the impact of different types of weak (soft) attractive interactions between the protein and crowders on protein folding and protein stability in Chapter 4. For this follow-up study, we utilized the same model in Chapter 3 but modified to include hydrophobic and nonspecific attractive crowding agents in the form of short polypeptide chains and spherical crowders. We found both hydrophobic and nonspecific crowders can lead to a net destabilization and counteract the stabilizing effect from hard-core repulsions. Also, we find that, for the same interaction strength and crowder concentration, spherical crowders with a hydrophobic character are more destabilizing than crowders with only of nonspecific interactions.

In the research project of Chapter 5 of this thesis, we investigate how polymeric crowders affect the protein's native state stability and their kinetics behavior. For

this goal, we apply a structure-based model and the Langevin equation to study two proteins with Protein Data Bank (PDB) ID 1PRB and 1SHF, under crowded conditions and examine their response based on their topologies. 1SHF is a SH3 domain from the human fyn protein[71], and 1PRB is an albumin binding domain on protein G[72].

Chapter 6 focuses on the effects of macromolecular crowding on protein fold switching. We test the effects of purely repulsive excluded volume crowders on the protein G with two folds as G_A and G_B. The main question in this research project is how crowding can change the populations of each fold at different temperatures and crowder volume fractions. Specifically, we examine the role of intrinsically disordered regions within the protein for the change in fold-switching behavior. The same model as in Chapter 5 has been used for both projects to develop a potential energy function to examine fold-switching thermodynamics.

Finally, the thesis ends in chapter 7 with a summary and outlook.

Bibliography

[1] H. Neurath and K. A. Walsh. Role of proteolytic enzymes in biological regulation (a review). *Proc Natl Acad Sci of the USA*, 73:3825–3832, 1976.

[2] D. M. Chudakov, M. V. Matz, S. Lukyanov, and K. A. Lukyanov. Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol Rev*, 90:1103–1163, 2010.

[3] T. Pawson. Protein modules and signalling networks. *Nature*, 373:573–580, 1995.

[4] D. Voet and J. G. Voet. *Biochemistry*. John Wiley & Sons, fourth edition, 2021.

[5] J. Nielsen, J. Brandt, T. Boesen, Tina T. Hummelshøj, R. Slaaby, G. Schluckebier, and P. Nissen. Structural investigations of full-length insulin receptor dynamics and signalling. *J Mol Biol*, 434:167458, 2022.

[6] S. Labeit, B. Kolmerer, and W. A. Linke. The giant protein titin: emerging roles in physiology and pathophysiology. *Circ Res*, 80:290–294, 1997.

[7] A. V. Finkelstein and O. Ptitsyn. *Protein physics: a course of lectures*. Elsevier, second edition, 2016.

[8] M. F. Perutz, M. G. Rossmann, A. F. Cullis, H. Muirhead, G. Will, and A. North. Structure of hæmoglobin: a three-dimensional fourier synthesis at 5.5-å. resolution, obtained by x-ray analysis. *Nature*, 185:416–422, 1960.

[9] G. Dodson and D. Steiner. The role of assembly in insulin's biosynthesis. *Curr Opin Struct Biol*, 8:189–194, 1998.

[10] D. M. Chipman and N. Sharon. Mechanism of lysozyme action: Lysozyme is the first enzyme for which the relation between structure and function has become clear. *Science*, 165:454–465, 1969.

[11] G. Guidotti. Membrane proteins. *Annu Rev Biochem*, 41:731–752, 1972.

[12] M. L. Huggins. The structure of fibrous proteins. *Chem Rev*, 32:195–218, 1943.

[13] T. Liu and G. Bitan. Modulating self-assembly of amyloidogenic proteins as a therapeutic approach for neurodegenerative diseases: strategies and mechanisms. *Chem Med Chem*, 7:359–374, 2012.

[14] A. Dong and W. S. Caughey. Infrared methods for study of hemoglobin reactions and structures. *Meth Enzymol*, 232:139–175, 1994.

[15] J. M. Rifkind and E. Nagababu. Hemoglobin redox reactions and red blood cell aging. *Antioxid Redox Signal*, 18:2274–2283, 2013.

[16] G. L. River, A. B. Robbins, S. O. Schwartz, et al. Sc hemoglobin: a clinical study. *Blood J Hematol*, 18:385–416, 1961.

[17] S. B. Zimmerman and S. G. Trach. Estimation of macromolecule concentrations

and excluded volume effects for the cytoplasm of escherichia coli. *J Mol Biol*, 222:599–620, 1991.

[18] A. P. Minton. Implications of macromolecular crowding for protein assembly. *Curr Opin Struct Biol*, 10:34–39, 2000.

[19] A. Christiansen, Q. Wang, M. S. Cheung, and P. Wittung-Stafshede. Effects of macromolecular crowding agents on protein folding in vitro and in silico. *Biophys Rev*, 5:137–145, 2013.

[20] G. Rivas and A. P. Minton. Macromolecular crowding in vitro, in vivo and in between. *Trends Biochem Sci*, 41:970–981, 2016.

[21] A. P. Minton. The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences. *Mol Cell Biochem*, 55:119–140, 1983.

[22] H. X. Zhou. Protein folding and binding in confined spaces and in crowded solutions. *J Mol Recognit*, 17:368–375, 2004.

[23] A. C. Miklos, M. Sarkar, Y. Wang, and G. J. Pielak. Protein crowding tunes protein stability. *J Am Chem Soc*, 133:7116–7120, 2011.

[24] A. P. Minton. Influence of macromolecular crowding upon the stability and state of association of proteins: predictions and observations. *J Pharm Sci*, 94:1668–1675, 2005.

[25] M. Sarkar, C. Li, and G. J. Pielak. Soft interactions and crowding. *Biophys Rev*, 5:187–194, 2013.

[26] S. Lehmann, D. Funck, L. Szabados, and D. Rentsch. Proline metabolism and transport in plant development. *J Amino Acids*, 39:949–962, 2010.

[27] M. T. Fisher. Proline to the rescue. *Proc Natl Acad Sci of USA*, 103:13265–13266, 2006.

[28] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181:223–230, 1973.

[29] C. B. Anfinsen and H. A. Scheraga. Experimental and theoretical aspects of protein folding. *Adv Protein Chem Struct Biol*, 29:205–300, 1975.

[30] R. D. Cohen and G. J. Pielak. A cell is more than the sum of its (dilute) parts: A brief history of quinary structure. *Protein Sci*, 26:403–413, 2017.

[31] M. Lella and R. Mahalakshmi. Metamorphic proteins: emergence of dual protein folds from one primary sequence. *Biochem*, 56:2971–2984, 2017.

[32] L. L. Porter and L. L. Looger. Extant fold-switching proteins are widespread. *Proc Natl Acad Sci of USA*, 115:5968–5973, 2018.

[33] Y. He, Y. Chen, P. A. Alexander, P. N. Bryan, and J. Orban. Mutational tipping points for switching protein folds and functions. *Struct*, 20:283–291, 2012.

[34] Y. G. Chang, S. E. Cohen, C. Phong, W. K. Myers, Y. I. Kim, Roger J. Lin R. Tseng, L. Zhang, J. S. Boyd, and Y. Lee. A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. *Science*, 349:324–328, 2015.

[35] X. Luo, Z. Tang, G. Xia, K. Wassmann, T. Matsumoto, J. Rizo, and H. Yu.

The mad2 spindle checkpoint protein has two distinct natively folded states. *Nat Struct Mol Biol*, 11:338–345, 2004.

[36] P. K. Zuber, K. Schweimer, P. Rösch, I. Artsimovitch, and S. H. Knauer. Reversible fold-switching controls the functional cycle of the antitermination factor rfah. *Nat Commun*, 10:702, 2019.

[37] P. N. Bryan and J. Orban. Implications of protein fold switching. *Curr Opin Struct Biol*, 23:314–316, 2013.

[38] P. A. Alexander, Y. He, Y. Chen, J. Orban, and P. N. Bryan. A minimal sequence code for switching protein structure and function. *Proceedings of the National Academy of Sciences*, 106:21149–21154, 2009.

[39] V. N. Uversky. Unusual biophysics of intrinsically disordered proteins. *Biochim Biophys Acta - Proteins Proteom*, 1834:932–951, 2013.

[40] A. Bhattacherjee and S. Wallin. Coupled folding-binding in a hydrophobic/polar protein model: impact of synergistic folding and disordered flanks. *Biophys J*, 102:569–578, 2012.

[41] R. Nikam, A. Kulandaisamy, K. Harini, D. Sharma, and M. M. Gromiha. Prothermdb: thermodynamic database for proteins and mutants revisited after 15 years. *Nucleic acids research*, 49:D420–D424, 2021.

[42] E. E. Lattman and G. D. Rose. Protein folding–what's the question? *Proc Natl Acad Sci USA*, 90:439–441, 1993.

[43] K. A. Dill. Dominant forces in protein folding. *Biochem*, 29:7133–7155, 1990.

[44] L. Pauling, R. B. Corey, and H. R. Branson. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci USA*, 37:205–211, 1951.

[45] M. M. Deshmukh and S. R. Gadre. Estimation of n- h··· o c intramolecular hydrogen bond energy in polypeptides. *J Phys Chem A*, 113:7927–7932, 2009.

[46] S. M. Habermann and K. P. Murphy. Energetics of hydrogen bonding in proteins: A model compound study. *Protein Sci*, 5:1229–1239, 1996.

[47] E. Van Dijk, A. Hoogeveen, and S. Abeln. The hydrophobic temperature dependence of amino acids directly calculated from protein structures. *PLoS Comput Biol*, 11:e1004277, 2015.

[48] G. Graziano. On the mechanism of cold denaturation. *Phys Chem Chem Phys*, 16:21755–21767, 2014.

[49] P. Atkins and J. De Paula. *Physical chemistry for the life sciences*. Oxford University Press, USA, 2011.

[50] P. L. Privalov. Stability of proteins small globular proteins. *Adv Protein Chem Struct Biol*, 33:167–241, 1979.

[51] M. C. Deller and Leopold B. Rupp L. Kong. Protein stability: a crystallographer's perspective. *Acta Crystallogr*, 72:72–95, 2016.

[52] R. A. L. Jones. *Soft Condensed Matter*. Oxford University Press, 2002.

[53] J. A. Rupley, E. Gratton, and G. Careri. Water and globular proteins. *Trends Biochem Sci*, 8:18–22, 1983.

[54] M. M. Coleman. *Fundamentals of polymer science: An introductory text*. Routledge, second edition, 2019.

[55] A. J. Guseman, G. M. Perez Goncalves, S. L. Speer, G. B. Young, and G. J. Pielak. Protein shape modulates crowding effects. *Proc Natl Acad Sci USA*, 115:10965–10970, 2018.

[56] K. Nasreen, Z. A. Parray, S. Ahamad, F. Ahmad, A. Ahmed, S. Freeh Alamery, T. Hussain, M. Hassan, and A. Islam. Interactions under crowding milieu: chemical-induced denaturation of myoglobin is determined by the extent of heme dissociation on interaction with crowders. *Biomolecules*, 10:490, 2020.

[57] J. Kundu, U. Kar, S. Gautam, S. Karmakar, and P. K. Chowdhury. Unusual effects of crowders on heme retention in myoglobin. *FEBS letters*, 589:3807–3815, 2015.

[58] W. H. Fissell, S. Manley, A. Dubnisheva, J. Glass, J. Magistrelli, A. N. Eldridge, A. J. Fleischman, A. L. Zydney, and S. Roy. Ficoll is not a rigid sphere. *Am J Physiol Ren*, 293:F1209–F1213, 2007.

[59] W. H. Fissell, C. L. Hofmann, E. Smith, and M. H. Chen. Size and conformation of ficoll as determined by size-exclusion chromatography followed by multiangle light scattering. *Am J Physiol Ren*, 298:F205–F208, 2010.

[60] A. P. Minton. Excluded volume as a determinant of protein structure and stability. *Biophys J*, 32:77, 1980.

[61] M. S. Cheung, D. Klimov, and D. Thirumalai. Molecular crowding enhances

native state stability and refolding rates of globular proteins. *Proc Natl Acad Sci of the USA*, 102:4753–4758, 2005.

[62] J. Batra, K. Xu, S. Qin, and H. X. Zhou. Effect of macromolecular crowding on protein binding stability: modest stabilization and significant biological consequences. *Biophys J*, 97:906–911, 2009.

[63] J. Mittal and R. B. Best. Dependence of protein folding stability and dynamics on the density and composition of macromolecular crowders. *Biophys J*, 98:315–320, 2010.

[64] M. Sarkar and G. J. Pielak. An osmolyte mitigates the destabilizing effect of protein crowding. *Protein Sci*, 23:1161–1164, 2014.

[65] M. Jiao, H. Li, J. Chen, A. P. Minton, and Y. Liang. Attractive protein-polymer interactions markedly alter the effect of macromolecular crowding on protein association equilibria. *Biophys J*, 99:914–923, 2010.

[66] J. F. Douglas, J. Dudowicz, and K. F. Freed. Crowding induced self-assembly and enthalpy-entropy compensation. *Phys Rev Lett*, 103:135701, 2009.

[67] Y. Phillip, V. Kiss, and G. Schreiber. Protein-binding dynamics imaged in a living cell. *Proc Natl Acad Sci of the USA*, 109:1461–1466, 2012.

[68] S. Timr and F. Sterpone. Stabilizing or destabilizing: simulations of chymotrypsin inhibitor 2 under crowding reveal existence of a crossover temperature. *J Phys Chem*, 12:1741–1746, 2021.

[69] Douglas D. Tsao and N. V. Dokholyan. Macromolecular crowding induces polypeptide compaction and decreases folding cooperativity. *Phys Chem Chem Phys*, 12:3491–3500, 2010.

[70] T. Bereau, M. Deserno, and M. Bachmann. Structural basis of folding cooperativity in model proteins: Insights from a microcanonical perspective. *Biophys J*, 100:2764–2772, 2011.

[71] ME. Noble, A. Musacchio, M. Saraste, S. A. Courtneidge, and R. K. Wierenga. Crystal structure of the sh3 domain in human fyn; comparison of the three-dimensional structures of sh3 domains in tyrosine kinases and spectrin. *The EMBO journal*, 12:2617–2624, 1993.

[72] M. U. Johansson, M. de Château, M. WikstroÈm, S. Forsén, T. Drakenberg, and L. Björck. Solution structure of the albumin-binding ga module: a versatile bacterial protein domain. *Journal of molecular biology*, 266:859–865, 1997.

Chapter 2

Theory and Methods

Computer simulations are an increasingly common method to study physical phenomena in fields such as physics, chemistry, and biology. Simulation of molecular systems is based on concepts in statistical mechanics and computational physics. In order to overcome different kinds of limitations, such as accessing long time- and length-scales with currently available computational resources, various techniques have been developed, such as enhanced conformational sampling techniques and coarse-graining of models. In this chapter, we describe some of the sampling techniques and models used in this thesis. We start from the point of view of statistical mechanics.

2.1 Statistical mechanics

In a classical system with N particles in a volume V, any microstate at time t can be defined by specifying the positions $\vec{q}_1$,...,$\vec{q}_N$ and momenta $\vec{p}_1, ..., \vec{p}_N$ of the particles. Hence, to specify a microstate requires $3N$ position and $3N$ momentum coordinates. We will denote these sets as $q = \{\vec{q}_\mathrm{i}\}_{\mathrm{i}=1}^{N}$ and $p = \{\vec{p}\}_{\mathrm{i}=1}^{N}$, respectively, such that each point (q, p) represents a distinct microstate. The space of all (q, p) points available to the system is called the phase space. In this thesis, the vectors $\vec{q}_\mathrm{i}$ and $\vec{p}_\mathrm{i}$ might represent positions and momenta of the atoms in a protein chain and surrounding solvent molecules. The set of positions (q) is often referred to the conformation of the system. Then, the Hamiltonian of the system will be given by:

$$H(q, p) = E(q) + \sum_{\mathrm{i}=1}^{N} \frac{\vec{p}_\mathrm{i}^{\,2}}{2m_\mathrm{i}}, \tag{2.1}$$

where m_i is the mass of particle i, and $E(q)$ is the potential energy of the system.

For a classical thermodynamic system at temperature T, the total partition function can be written $Z = Z_\mathrm{ideal} Z_\mathrm{conf}$ [1]. The two factors Z_ideal and Z_conf are obtained by integrating over the momenta and position coordinates, respectively. The probability of finding the system in a conformation q is given by the Boltzmann distribution,

$$P_\mathrm{B}(q) = \frac{\exp\left[-\beta E(q)\right]}{Z_\mathrm{conf}}, \tag{2.2}$$

where $\beta = 1/k_\mathrm{B}T$, and

$$Z_\mathrm{conf} = \int d\vec{q}_1 ... \int d\vec{q}_N \exp\left[-\beta E(q)\right]. \tag{2.3}$$

The average of an observable $A(q)$ takes the form of an integral over the conforma-

tional space of the system. Specifically, the thermodynamic average is given by

$$\langle A \rangle = \frac{\int d\vec{q}_1 ... \int d\vec{q}_N A(q) \exp\left[-\beta E(q)\right]}{\int d\vec{q}_1 ... \int d\vec{q}_N \exp\left[-\beta E(q)\right]}. \tag{2.4}$$

For example, in a protein chain, the observable could be the end-to-end distance given by equation 1.3.

Furthermore, in the canonical ensemble, we have $\ln Z = -\beta F$, and F is the Helmholtz free energy of the system. This relation makes a connection with thermodynamics. For example, from F we can determine the internal energy of the system,

$$\langle E \rangle = \frac{\partial(\beta F)}{\partial \beta}. \tag{2.5}$$

Some quantities reflect how the system responds to changes in intensive variables such as T, and are called response functions. Two examples are the heat capacity at constant volume C_{v} and the isothermal compressibility κ_{v}. The heat capacity is defined as

$$C_{\mathrm{v}} = \frac{d\langle E \rangle}{dT} = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_{\mathrm{B}} T^2}, \tag{2.6}$$

where the second equality can be derived from equation 2.4. The isothermal compressibility is given by

$$\kappa_{\mathrm{v}} = -\frac{1}{V} \left(\frac{\partial V}{\partial p} \right)_T, \tag{2.7}$$

where p is pressure. These two quantities are accessible experimentally for biomolecular systems. For example, heat capacity can be measured with isothermal titration calorimetry, and κ_{v} can be found from scattering experiments.

2.2 Molecular dynamics

The basic idea of molecular dynamics simulations is to numerically integrate the equations of motion for the system under study. In biomolecular simulations, it is often classical equations of motions that are integrated; for example, for particle i,

$$\dot{\vec{q}_{\mathrm{i}}} = \frac{\partial H}{\partial \vec{p_{\mathrm{i}}}}, \tag{2.8}$$

and,

$$\dot{\vec{p}_{\mathrm{i}}} = -\frac{\partial H}{\partial \vec{q_{\mathrm{i}}}}, \tag{2.9}$$

where, H is the Hamiltonian of the system [2]. Generally, for long simulation times, a molecular dynamic trajectory should generate microscopic states consistent with the equilibrium distribution. For example, if the simulation is carried out at constant temperature and volume, the Boltzmann distribution should be generated. Thermodynamic averages (equation 2.4) can thus be determined for suitable observables and compared with experimental measurements. In addition to equilibrium properties, molecular dynamics simulations should also provide any dynamic properties: transport coefficients, time-dependent responses to perturbations, and rheological properties. Molecular dynamics simulations have many applications in biophysics, chemical physics, and material science. There are many variants of molecular dynamics simulations [3]. Molecular dynamics simulations are not always carried out using the equations of motion in 2.8 and 2.9. An alternative is the Langevin equation, which will be discussed in the next section.

Because macromolecular systems are complex and consist of a large number of particles, it can be hard in practice to determine the properties of such complex systems

precisely from molecular dynamics simulations. Issues that can limit the precision of molecular dynamics simulations include configurational sampling problems [4] and errors resulting from the discretization of time in the numerical integration [5]. A common choice in calculating trajectories of particles in molecular dynamics simulations is the Verlet algorithm [6]. Because of fast vibrational motions, the time step of numerical integration in detailed molecular dynamics simulations of protein systems must be of the order of $dt = 1 - 2$ fs [7]. This limits the size and time scales that can be probed via this method.

2.3 Langevin dynamics

In a typical detailed protein simulation, a large part of the computational resources is devoted to the surrounding solvent molecules. In Langevin dynamics [3], solvent molecules are no longer explicitly represented in the simulation. Instead, they are replaced with a friction term in the equations of motion. For example, Langevin dynamics have been used to study protein folding kinetics [7, 8]. Langevin dynamics is based on the Langevin equation, which is a stochastic differential equation describing the time evolution of a system subject to both deterministic and random forces. For a given particle i in the system, the Langevin equation can be written as:

$$m_{\mathrm{i}}\ddot{\vec{q}}_{\mathrm{i}} = -\zeta\dot{\vec{q}}_{\mathrm{i}} + \vec{F}_{\mathrm{i}}^{c} + \vec{\Gamma}_{\mathrm{i}} = \vec{F}_{\mathrm{i}}^{\mathrm{T}}, \tag{2.10}$$

where m_{i} is the mass of the particle, ζ is the friction coefficient, $\vec{q}_{\mathrm{i}}$ is the position of particle i. The total force of the particle i, $\vec{F}_{\mathrm{i}}^{\mathrm{T}}$ includes $\vec{F}_{\mathrm{i}}^{\mathrm{c}}$, the conformational force equal to the negative gradient of the potential energy of the system ($\vec{F}_{\mathrm{i}}^{\mathrm{c}} = \frac{-\partial E}{\partial \vec{q}_{\mathrm{i}}}$), and

$\vec{\Gamma}_{\rm i}$, the random force representing the thermal motion of solvent particles.

The equation 2.10 can be numerically integrated by using the velocity form of the Verlet algorithm [9], which works in the following way. The Verlet algorithm updates the positions and velocities of particles in a simulation by approximating the equations of motion. If the integration step is dt, the position of a particle at the time $t+dt$ is described to the second-order in dt as

$$\vec{q}_{\rm i}(t+dt) = \vec{q}_{\rm i}(t) + dt\dot{\vec{q}}_{\rm i}(t) + \frac{dt^2}{2m_{\rm i}}\vec{F}_{\rm i}^{T}(t). \tag{2.11}$$

The velocity at the time $t+dt$ is similarly approximated to the second order in dt , i.e.,

$$\dot{\vec{q}}_{\rm i}(t+dt) = \dot{\vec{q}}_{\rm i}(t) + \frac{dt}{m_{\rm i}}\vec{F}_{\rm i}^{T}(t) + \frac{dt^2}{2m_{\rm i}}\dddot{\vec{q}}_{\rm i}. \tag{2.12}$$

Equation 2.12 can be written

$$\dot{\vec{q}}_{\rm i}(t+dt) = (1-A)\left(\left(1-A+A^2\right)\right)\dot{\vec{q}}_{\rm i}(t) + \frac{dt}{2m_{\rm i}}\left(\left(1-A+A^2\right)\right)\times \\ \left(\vec{F}_{\rm i}^{c}(t) + \vec{\Gamma}_{\rm i}(t) + \vec{F}_{\rm i}^{c}(t+dt) + \vec{\Gamma}_{\rm i}(t+dt)\right) + \mathcal{O}(dt^3), \tag{2.13}$$

in which, $A = 1 - \frac{\zeta dt}{2m_{\rm i}}$; The random force is uncorrelated at different times, i.e.,

$$\langle \Gamma_{{\rm i}\alpha}(t)\Gamma_{{\rm i}\alpha}(t')\rangle = 0, \tag{2.14}$$

where $\Gamma_{{\rm i}\alpha}$ is the component of the random force on particle i in the $\alpha = x, y, z$ direction, and $t \neq t'$. The magnitude of the random force sets the temperature of the system. Specifically,

$$\langle \Gamma_{{\rm i}\alpha}^2(t)\rangle = 2\zeta k_{\rm B}T. \tag{2.15}$$

Equations 2.14 and 2.15 can be combined into a single equation,

$$\langle \Gamma_{{\rm i}\alpha}(t)\Gamma_{{\rm i}\alpha}(t'))\rangle = 2\zeta k_{\rm B}T\delta(t-t'), \tag{2.16}$$

where $\delta(t)$ is the Dirac delta function. To solve the Langevin equation 2.10 numerically, time is discretized. Equation 2.16 then becomes

$$\langle \Gamma_{\mathrm{i}\alpha}(t)\Gamma_{\mathrm{i}\alpha}(t+ndt))\rangle = \frac{2\zeta_{\mathrm{i}}k_{\mathrm{B}}T}{dt}\delta_{0,\mathrm{n}}, \tag{2.17}$$

where $\delta_{0,n}$ is the Kroncker delta and n= $0, 1, 2, ...$ [10].

2.4 Metropolis-Hastings Monte Carlo

Monte Carlo is a simulation approach that is entirely different from molecular dynamics, using random numbers to evolve the system rather than the equation of motion. The Metropolis-Hastings Monte Carlo algorithm is a Markov chain Monte Carlo (MCMC) method for sampling states from the Boltzmann distribution $P_{\mathrm{B}}(s)$. If we consider a sequence generated of states $s_1, s_2, ..., s_{\mathrm{i}}, ..., s_{\mathrm{M}}$, the key characteristic of a Markov chain is that the state s_{i} only depends on the previous state $s_{\mathrm{i}-1}$. To see why this is useful, note that because the states are biased according to the Boltzmann distribution, the thermodynamic average of an observable A can be estimated by:

$$\langle A\rangle \approx \frac{1}{M}\sum_{\mathrm{i}}^{M} A(s_{\mathrm{i}}). \tag{2.18}$$

The equation 2.18 is accurate in the limit of large M and when the s_{i}'s are not all correlated.

Practically, the Metropolis-Hastings algorithm works in the following way [11, 12]:

1. Prepare the system in an initial state s.
2. Pick a new trial state $s^{'}$ with probability $\alpha(s \rightarrow s')$.

3. Accept the new state s' with the probability,

$$P_{\text{acc}}(s \to s') = \min[1, \frac{\alpha(s' \to s)}{\alpha(s \to s')} \exp(-\beta \Delta E)], \tag{2.19}$$

where $\Delta E = E(s') - E(s)$ is the change in energy. If the trial move is not accepted, restore the system to state s.

4. Return to step 2.

The Metropolis-Hastings algorithm can be understood from the so-called detailed balance condition, which states that

$$P_{\text{B}}(s)\pi(s \to s') = P_{\text{B}}(s')\pi(s' \to s), \tag{2.20}$$

where $\pi(s \to s')$ is the probability of transitioning to state s' in one step, given that the current state is s, and $\pi(s' \to s)$ is the probability of transitioning to state s in one step, given that the current state is s'.

According to the detailed balance condition, transitions between any two states, s and s', occur equally frequently in both directions. Many possible transition probabilities π can satisfy the equation 2.20. Because generating a new state s' is performed in two steps, a trial step and an accept/reject step, the transition probability can be written

$$\pi(s \to s') = \alpha(s \to s')P_{\text{acc}}(s \to s'), \tag{2.21}$$

where $P_{\text{acc}}(s \to s')$ is the probability of accepting the transition $s \to s'$. Then, the equation 2.20 becomes

$$P_{\text{B}}(s)\alpha(s \to s')P_{\text{acc}}(s \to s') = P_{\text{B}}(s')\alpha(s' \to s)P_{\text{acc}}(s' \to s), \tag{2.22}$$

and we obtain

$$\frac{P_{\text{acc}}(s \to s')}{P_{\text{acc}}(s' \to s)} = \frac{\alpha(s' \to s)P_{\text{B}}(s')}{\alpha(s \to s')P_{\text{B}}(s)} = \frac{\alpha(s' \to s)}{\alpha(s \to s')} \exp\left[-\beta \Delta E\right]. \tag{2.23}$$

In order to maximize the acceptance probability $P_{\text{acc}}(s \to s')$ we make the choice

$$P_{\text{acc}}(s \to s') = \min[1, \frac{\alpha(s' \to s)}{\alpha(s \to s')} \exp\left[-\beta \Delta E\right]], \tag{2.24}$$

which is the acceptance probability of the Metropolis-Hastings algorithm.

It should be noticed that generally, α is not symmetric i.e. $\alpha(s \to s') \neq \alpha(s' \to s)$. However, in the original Metropolis algorithm [11], α is symmetric ($\alpha(s \to s') = \alpha(s' \to s)$), and the equation 2.24 simplifies to

$$P_{\text{acc}}(s \to s') = \min[1, \exp\left(-\beta \Delta E\right)]. \tag{2.25}$$

where $\Delta E = E(s') - E(s)$. This choice will satisfy the equation 2.20 for both cases as $E(s') > E(s)$ or $E(s) > E(s')$ [3].

2.5 Simulated tempering

Many physical systems have local minima in their energy landscape that are separated by high energy barriers. As a result, in a standard Monte Carlo or molecular dynamics simulation, the system can become trapped in one of these local minima, which means that we cannot explore all configurations. Such kinetic traps will occur, especially at low temperatures, where the probability of moving to states with higher energies becomes small.

One class of algorithms that have been proposed to solve this problem is generalized ensemble (GE) methods. These methods aim to enable the system to overcome high energy barriers and allow the system to sample rare states. Examples include $1/k$ sampling [13], multicanonical algorithm [14] and simulated tempering [15, 16]. In this thesis, we use simulated tempering.

Simulated tempering works by using Metropolis-Hastings Monte Carlo to sample from the following non-canonical distribution [17]:

$$P(s, j) \propto \exp\left[-E(s)/k_{\mathrm{B}}T_{\mathrm{j}} + g_{\mathrm{j}}\right], \tag{2.26}$$

where $\mathrm{j} = 1, ..., K$ is a new dynamic variable and $T_1 ..., T_{\mathrm{K}}$ is a set of pre-determined temperature values. By introducing the control variable j, the simulation can explore both high and low temperatures. Visits to high temperatures will help escape from local energy minima and enhance sampling at low T. $g_1, ..., g_{\mathrm{K}}$ is a set of simulation parameters which control the marginal distribution $P(j)$. An appropriate set of g_{j} provides a flat $P(j)$, thereby achieving good conformational sampling at all temperatures.

In simulated tempering, there are thus two types of MC updates: a conventional conformational update $s \to s'$ and an update in the control variable $j \to j'$. The acceptance probability for $s \to s'$ updates becomes the same as for the Metropolis algorithm, such as equation 2.24. For the temperature update, the acceptance probability becomes:

$$P_{\mathrm{acc}}(j \to j') = \min[1, \exp\left[-E(s)\Delta\beta + \Delta g\right]], \tag{2.27}$$

where $\Delta\beta = \frac{1}{k_{\mathrm{B}}T_{\mathrm{j}'}} - \frac{1}{k_{\mathrm{B}}T_{\mathrm{j}}}$ and $\Delta g = g_{\mathrm{j}'} - g_{\mathrm{j}}$. At a given value of the control parameter

j, the probability distribution function sampled will be:

$$P(s|j) \propto \exp\left[-E(s)/k_{\mathrm{B}}T_{\mathrm{j}}\right]. \tag{2.28}$$

Therefore, conformational states generated at a given j will be distributed according to the Boltzmann distribution at the temperature T_j. Note, in particular, that the distribution 2.28 is independent of the simulation parameter g_{j}. Therefore, while the g_{j} parameters should be set so that all temperatures are reasonably well sampled, they do not need to be fine-tuned.

2.6 Coarse-graining in molecular simulations

Molecular simulations can be carried out at different levels of resolution. All-atom simulations have the potential to provide the most accurate results. However, they are often limited to small systems or short simulation times. One way to tackle this problem is by using coarse-grained (CG) models. The primary idea behind CG models is to reduce the representation of the system such that it can be simulated over a longer time while retaining the essential physics of the system. CG models are using a degree of simplification that varies depending on the system under study. They also typically result in faster dynamics, which further speed up conformational sampling [18]. A challenge with CG models is to formulate an energy function for the reduced system. Hence, it must be tested against more detailed models or experimental data [19, 20].

A type of coarse-graining that has been used in folding simulations of biomolecules is so-called Go-type or structure-based models [21]. The interactions in a structure-

based model are not based on physio-chemical principles. Rather, they are constructed so that interactions that are present in a reference structure, typically the native structure, are made attractive. Other (non-native) interactions are left neutral or are even made repulsive. Surprisingly, structure-based models have successfully reproduced detailed experimental data on protein folding [22–28].

In contrast to structure-based models, a sequence-based model has a potential energy function that depends only on the primary amino acid sequence and is independent of any particular reference structure. In the following, we describe two coarse-grained models that are used in this thesis, one sequence-based (C_β) model and one structure-based (C_α) model.

2.6.1 C_β model

Our C_β model is a coarse-grained sequence-based model with 7 atoms per amino acid. It has only three amino acid types, which are a simplification of the 20 types found in nature. One polar amino acid, one hydrophobic amino acid, and glycine. We denote them h,p, and t, respectively [29]. Different sequences lead to making different structures. In this method, amino acid sequences can be constructed in the 3-letter code. It should be noticed that not all sequences behave protein-like, as is the case also for real sequences.

Basic protein design principle can be used to construct sequences with stable native conformations. For example, a sequence pattern with a hydrophobic (h) amino acid at every 3 or 4 positions is consistent with amphipathic α-helix. By alternating

p and h amino acids, it is possible to make amphipathic β-strands. Also, due to the lack of C_β atom, turn residues (t) are flexible and can be used to create pliable turns in a sequence. The C_β model thus provides a simplified model for the sequence-structure relationship in proteins. Although the 3-letter sequences are not real, the native folds that the model produces are protein-like. For example, in Chapter 3 we study a sequence folding to a β-sheets structure that is similar to the so-called WW domain [30].

As we mentioned earlier, we use the C_β model as a framework to investigate the impacts of crowding on protein thermodynamics. The potential energy function in this model can be written as:

$$E_{C_\beta} = E_{\text{exvol}} + E_{\text{local}} + E_{\text{hbond}} + E_{\text{hp}} \tag{2.29}$$

These four terms represent excluded volume, local partial charge interaction, hydrogen bonding, and hydrophobic interaction [29]. We will discuss the details of the model in Chapter 3 and Chapter 4.

2.6.2 C_α model

In the C_α model, each amino acid is represented by a single interaction site located at the C_α atom. The structure-based energy function has the following form [22]:

$$E_{C_\alpha} = \sum_{\text{bonds}} K_\text{b} \left(b_\text{i} - b_\text{i}^\text{n}\right)^2 + \sum_{\text{angles}} K_\theta \left(\theta_\text{i} - \theta_\text{i}^\text{n}\right)^2 + \sum_{\text{dihedrals}} \left[K_\phi^{(1)} \left[1 - \cos\left(\phi_\text{i} - \phi_\text{i}^\text{n}\right)\right] + K_\phi^{(3)} \left[1 - \cos 3(\phi_\text{i} - \phi_\text{i}^\text{n})\right]\right] + \sum_{\text{i<j}-3}^{\text{native}} \epsilon \left[5\left(\frac{r_\text{ij}^\text{n}}{r_\text{ij}}\right)^{12} - 6\left(\frac{r_\text{ij}^\text{n}}{r_\text{ij}}\right)^{10}\right] + \sum_{\text{i<j}-3}^{\text{non-native}} \epsilon \left(\frac{r_\text{rep}}{r_\text{ij}}\right)^{12}. \quad (2.30)$$

The parameter ϵ sets the energy scale of the model, and b_i, θ_i, ϕ_i, and r_ij are virtual bond lengths, bond angles, torsion angles, and C_α-C_α distance between residues i and j, respectively. The interaction strengths K_b, K_θ, $K_\phi^{(1)}$, $K_\phi^{(3)}$ are identical to those in Ref. [22]. b_i^n, θ_i^n, ϕ_i^n, and r_ij^n are their values at the native state for a specific experimentally determined protein structure. Structures can e.g. be provided by the Protein Data Bank (PDB). To obtain this model's kinetic and thermodynamic properties, we use Langevin dynamics, as described in section 2.3.

We use our C_α model in Chapter 5 to study the crowding effect on protein thermodynamics and kinetics. In chapter 6, we apply it to study fold-switching proteins. However, in that case, the model needs to be modified to take into account the two different native states of metamorphic proteins. We will discuss the details of this modification in Chapter 6.

2.7 Scaled particle theory

The scaled particle theory (SPT) is a theoretical framework for understanding the thermodynamic properties of simple liquids based on the assumption that a hard sphere potential can approximate the interparticle potential [31]. Lebowitz *et al.* first developed the theory in 1960 [32], and it has since been extended by a number of researchers [33, 34].

In the context of macromolecular crowding, SPT can be used to consider non-ideality arising from the presence of cosolutes. This theory assumes that the interaction energy between molecules is purely repulsive. The usefulness of SPT comes from providing a point of comparison to experimental [35, 36] or computational data [37].

The SPT provides a way to calculate the thermodynamic properties in the context of the hard-sphere model. One important application of SPT is to calculate the free energy of inserting a hard sphere of radius R in a hard sphere fluid of particles with radius R_c, as illustrated in Fig. 1.1. The free energy is given by [32]:

$$\beta F = (3y + 3y^2 + y^3)\psi + (\frac{9y^2}{2} + 3y^3)\psi^2 + 3y^3\psi^3 - \ln(1 - \phi_c)\,, \tag{2.31}$$

where $y = \frac{R}{R_c}$, $\psi = \frac{\phi_c}{1-\phi_c}$, and ϕ_c is fluid volume fraction. We discuss equation 2.31 in Chapter 6 and show how to calculate changes to fold switching due to crowding by SPT.

SPT has been applied to a range of problems in soft matter physics, including the behavior of colloids [38], polymers [39–41], and liquid crystals [42]. While the theory is based on a number of simplifying assumptions, it has proven to be a useful tool for

understanding the behavior of simple liquids. It has provided insight into a number of important phenomena, such as the nature of the phase transition [43].

Bibliography

[1] D. Chandler. *Introduction to modern statistical mechanics*. Oxford University Press, USA, first edition, 1987.

[2] R. Swendsen. *An introduction to statistical mechanics and thermodynamics*. Oxford University Press, USA, first edition, 2020.

[3] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications*. Elsevier, 2001.

[4] M. Pájaro, A. Alonso, I. Otero-Muras, and C. Vázquez. Stochastic modeling and numerical simulation of gene regulatory networks with protein bursting. *J Theor Biol*, 421:51–70, 2017.

[5] D. Bucher, L. CT. Pierce, J. A. McCammon, and P. RL. Markwick. On the use of accelerated molecular dynamics to enhance configurational sampling in ab initio simulations. *J Chem Theory Comput*, 7:890–897, 2011.

[6] H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol Simul*, 6:121–142, 1991.

[7] Z. Guo and D. Thirumalai. Kinetics of protein folding: nucleation mechanism, time scales and pathways. *Biopolymers*, 36:83–102, 1995.

[8] J. D. Honeycutt and D. Thirumalai. The nature of folded states of globular proteins. *Biopolymers*, 32:695–709, 1992.

[9] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J Chem Phys*, 76:637–649, 1982.

[10] T. Veitshans, D. Klimov, and D. Thirumalai. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold Des*, 2:1–22, 1997.

[11] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Introduction of the Metropolis algorithm for molecular-dynamics simulation. *J Chem Phys*, 21:1987, 1953.

[12] W. K. Hastings. Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, 57:97–109, 1970.

[13] B. Hesselbo and R. B. Stinchcombe. Monte Carlo simulation and global optimization without parameters. *Phys Rev Lett*, 74:2151, 1995.

[14] W. Janke. Multicanonical Monte Carlo simulations. *Phys A: Stat Mech Appl*, 254:164–178, 1998.

[15] U. H. E. Hansmann and Y. Okamoto. Generalized-ensemble Monte Carlo method for systems with rough energy landscape. *Phys Rev E*, 56:2228, 1997.

[16] W. Nadler and U. H. E. Hansmann. Generalized ensemble and tempering simulations: A unified view. *Phys Rev E*, 75:026109, 2007.

[17] E. Marinari and G. Parisi. Simulated Tempering: a new Monte Carlo scheme. *Europhys Lett*, 19:451, 1992.

[18] N. Singh and W. Li. Recent advances in coarse-grained models for biomolecules and their applications. *Int J Mol Sci*, 20:3774, 2019.

[19] S. Izvekov and G. A. Voth. A multiscale coarse-graining method for biomolecular systems. *J Phys Chem B*, 109:2469–2473, 2005.

[20] E. Brini, E. A. Algaer, P. Ganguly, C. Li, F. Rodríguez-Ropero, and N. F. A. van der Vegt. Systematic coarse-graining methods for soft matter simulations–a review. *Soft Matter*, 9:2108–2119, 2013.

[21] C. Clementi, H. Nymeyer, and J. N. Onuchic. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? an investigation for small globular proteins. *J Mol Biol*, 298:937–953, 2000.

[22] S. Wallin and H. S. Chan. Conformational entropic barriers in topology-dependent protein folding: perspectives from a simple native-centric polymer model. *J Phys Condens Matter*, 18:307–328, 2006.

[23] N. M. Mascarenhas and S. Gosavi. Understanding protein domain-swapping using structure-based models of protein folding. *Prog Biophys*, 128:113–120, 2017.

[24] H. Lammert, A. Schug, and J. N. Onuchic. Robustness and generalization of structure-based models for protein folding and function. *Proteins: Struct Func*, 77:881–891, 2009.

[25] A. M. Rubio and A. Rey. Design of a structure-based model for protein folding from flexible conformations. *Phys Chem Chem Phys*, 21:6544–6552, 2019.

[26] J. Especial, A. Nunes, A. Rey, and P. F. N Faísca. Hydrophobic confinement modulates thermal stability and assists knotting in the folding of tangled proteins. *Phys Chem Chem Phys*, 21:11764–11775, 2019.

[27] C. Clementi. Coarse-grained models of protein folding: toy models or predictive tools? *Curr Opin Struct Biol*, 18:10–15, 2008.

[28] M. S. Cheung, D. Klimov, and D. Thirumalai. Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc Natl Acad Sci of USA*, 102:4753–4758, 2005.

[29] A. Bhattacherjee and S. Wallin. Coupled folding-binding in a hydrophobic/polar protein model: impact of synergistic folding and disordered flanks. *Biophys J*, 102:569–578, 2012.

[30] P. Freddolino, F. Liu, M. Gruebele, and Klaus K. Schulten. Ten-microsecond

molecular dynamics simulation of a fast-folding ww domain. *Biophys J*, 94:L75–L77, 2008.

[31] J. Hansen and I. R. McDonald. *Theory of simple liquids: with applications to Soft Matter*. Academic press, third edition, 2013.

[32] E. Helfand, H. Reiss, H. L. Frisch, and J. L. Lebowitz. Scaled particle theory of fluids. *J Chem Phys*, 33:1379–1385, 1960.

[33] J. L. Lebowitz, E. Helfand, and E. Praestgaard. Scaled particle theory of fluid mixtures. *J Chem Phys*, 43:774–779, 1965.

[34] R. M. Gibbons. The scaled particle theory for particles of arbitrary shape. *Mol Phys*, 17:81–86, 1969.

[35] A. J. Guseman, G. M. Perez Goncalves, S. L. Speer, G. B. Young, and G. J. Pielak. Protein shape modulates crowding effects. *Proc Natl Acad Sci USA*, 115:10965–10970, 2018.

[36] A. P. Minton. Influence of macromolecular crowding upon the stability and state of association of proteins: predictions and observations. *J Pharm Sci*, 94:1668–1675, 2005.

[37] H. Choi, H. Kang, and H. Park. Computational prediction of molecular hydration entropy with hybrid scaled particle theory and free-energy perturbation method. *J Chem Theory Comput*, 11:4933–4942, 2015.

[38] P. G. Bolhuis, A. Stroobants, D. Frenkel, and H. N. W. Lekkerkerker. Numerical

study of the phase behaviour of rodlike colloids with attractive interactions. *J Chem Phys*, 107:1551–1564, 1997.

[39] M. A. Cotter. Hard-rod fluid: scaled particle theory revisited. *Phys Rev A*, 10:625, 1974.

[40] S. D. Zhang, P. A. Reynolds, and J. S. Van Duijneveldt. Phase behavior of mixtures of colloidal platelets and nonadsorbing polymers. *J Chem Phys*, 117:9947–9958, 2002.

[41] D. S. Simmons and I. C. Sanchez. Scaled particle theory for the coil–globule transition of an isolated polymer chain. *Macromolecules*, 46:4691–4697, 2013.

[42] Martha A Cotter. Hard spherocylinders in an anisotropic mean field: a simple model for a nematic liquid crystal. *J Chem Phys*, 66:1098–1106, 1977.

[43] Howard H. Reiss and A. D. Hammerich. Hard spheres: scaled particle theory and exact relations on the existence and structure of the fluid/solid phase transition. *J Phys Chem*, 90:6252–6260, 1986.

Chapter 3

Crowding-induced protein destabilization in the absence of soft attractions

Abstract

It is generally expected that volume exclusion by macromolecular crowders universally stabilizes the native states of proteins, and destabilization suggests soft attractions between crowders and proteins. This expectation is based on the assumption that the unfolded state, U, is a conformational ensemble that is extended under all conditions and thereby becomes entropically disfavored relative to the native state, N, under crowded conditions. Here we show that proteins can be left neutral or even be destabilized by crowders that are purely repulsive. With a coarse-grained sequence-based model, we study the folding thermodynamics of three sequences with different native folds, a helical hairpin, a 5-stranded β-barrel, and a 3-stranded β-sheet, in a range of crowder volume fractions, ϕ_{c}. We find that, for all three sequences, N remains structurally unchanged under all studied crowded conditions, while the size of the unfolded state, U, decreases with ϕ_{c}. A simple entropy-centric view holds for the helical hairpin protein, which is stabilized under all crowded conditions as quantified by changes in either the folding midpoint temperature, T_{m}, or the free energy of folding. Although the size of U is temperature dependent, it is always large enough to be entropically disfavored relative to N, leading to a general stabilization of the protein. We find, however, that the stability of the β-barrel protein is not affected strongly under any conditions, and the 3-stranded β-sheet protein is destabilized by crowders at low Ts. This destabilization of the β-sheet protein can be understood from a highly compact U at low T, which is even more compact than the rather open native structure for this protein, such that U is entropically favored over N. At higher Ts, U is more expanded than N at $\phi_{\mathrm{c}} = 0$. Destabilization by repulsive crowders

is nonetheless possible at these higher Ts due to the crowding-induced compaction of U, which increases the fraction of U conformations that occupy highly compact non-native conformations.

3.1 Introduction

Many valuable advances in protein biophysics have come from experiments on proteins in dilute solutions. However, the native environment of proteins, i.e., the cell interior, is spatially inhomogeneous and highly crowded [1]. A substantial fraction of this space is occupied by macromolecules. They can occupy up to 40% of a cell's volume and reach concentrations of 100-400 g/L. [2]. Macromolecular crowding effects have been shown to impact a range of protein processes, such as folding [3], aggregation [4], and liquid-liquid phase separation [5].

An unavoidable consequence of macromolecular crowders is that they reduce the volume available to other molecules in the solution [6]. Minton was first to quantify the impact of this excluded volume effect on protein stability, predicting it to be universally stabilizing [7]. Experiments using artificial polymer crowders, such as Ficoll, dextran or polyethelene glycol, as excluded volume agents, indeed very often stabilize proteins as indicated by, e.g., an increase in the free energy of unfolding or the folding midpoint temperature [8–15]. Some exceptions have also been found [11, 16–18]. For example, the protein apoazurin exhibits a decreased unfolding free energy at low concentrations of Ficoll-70 [18]. Computational studies [19–26] have also examined the effect of volume exclusion on protein stability, showing results that range from negligible to robustly stabilizing depending on crowding conditions and protein studied. Most of these computational studies, although not all [20], relied on so-called structure-based or Gō-type [27] models for folding in which attractive nonnative interactions are typically ignored. To the best of our knowledge, no computational

crowding study has reported a decrease in protein stability from purely repulsive crowders.

More recent experiments [28–32] and theory [32, 33] have shown that the stabilizing effect of hard-core steric repulsions can be opposed by nonspecific soft attractions or "chemical" interactions between crowders and protein, by energetically favoring the unfolded state. These soft attractions can even dominate over hard-core effects, leading to a net destabilization [31], which appears to be common when the crowder molecules are proteins [34–36]. As showed by Zhou [33], soft attractions may lead to a cross-over temperature, $T_{\rm cross}$, at which crowding effects switch from destabilizing ($T < T_{\rm cross}$) to stabilizing ($T > T_{\rm cross}$). Examining the temperature dependence has provided additional insights by allowing the enthalpic and entropic components of the unfolding free energy to be determined. In the case of artificial polymer crowders, stabilization was in some cases found to be driven by enthalpy rather than entropy, contrary to the expectation of the excluded volume effect [10, 29, 32]. These results have been interpreted in terms of a preferential hydration effect [29, 37], akin to the protective mechanism of osmolytes [38], in which the crowders are preferentially excluded from the protein-water interface.

Here, we revisit the issue of the excluded volume effect on protein folding and stability, focusing on the role of nonnative interactions in folding. To this end, we use a coarse-grained model [39] for folding with 3 amino acid types, and a potential energy function based on hydrogen bonding and effective hydrophobic attractions. In this model, sequences can be designed using basic design principles [40] to fold into thermally stable states with protein-like native structures. As a focus of our

study, we take two 35- and one 25-amino acid sequences [41, 42] (see Table 1) that adopt a helical hairpin, a 5-stranded β-barrel, and a 3-stranded β-sheet structure, as shown in Fig 3.1A-C. We study their folding in the presence of spherical crowders with purely repulsive interactions. Under excluded volume crowders, α_{35} is always stabilized, β_{35} is stabilized at high Ts but neither stabilized nor destabilized at low temperatures. Unexpectedly, we find that the 3-stranded β-sheet structure protein is destabilized at low temperatures and high volume fraction of crowders, contrary to the expectation of the excluded volume effect. Our results can, however, be fully rationalized by a crowder-driven population shift towards more compact states, both native and nonnative.

α_{35}	$\mathrm{p(phpphhp)_2ptttp(phpphhp)_2p}$
β_{35}	$\mathrm{p(hp)_3tt(ph)_2(hp)_3tt(ph)_2tt(hp)_3}$
β_{25}	$\mathrm{(ph)_3tt(hp)_4tt(ph)_3p}$

Table 3.1: Amino acid model sequences.

The three amino acid model sequences studied in this work. The three types of amino acids, p (polar), h (hydrophobic) and t (turn), are described in Methods. Subscripts are used to indicate repeats. For instance, $(\mathrm{ph})_2$ means phph.

3.2 Materials and Methods

3.2.1 Excluded volume crowders

We model crowder-protein and crowder-crowder interactions using the pair potential suggested by Mittal and Best [24],

$$V(r) = k_{\mathrm{cr}} \left(\frac{\sigma}{r - \rho + \sigma} \right)^{12}, \tag{3.1}$$

where $r \geq \rho - \sigma$ is the center-to-center distance either between two crowders or between a crowder and a protein atom. For $r < \rho - \sigma$, V is taken to be infinite. The two parameters ρ and σ control, respectively the range and softness of the interaction, as illustrated in Fig. A1. For crowder-crowder pairs, we set $\rho = 2R_{\mathrm{c}}$ and $\sigma = 2\sigma_{\mathrm{cr}}$, where R_{c} is the crowder radius and $\sigma_{\mathrm{cr}} = 3$ Å is a parameter setting the softness of the crowders. The choice of ρ and σ_{cr} are the same as Mittal and Best[24]. For crowder-atom pairs, we set $\rho = R_{\mathrm{c}} + \sigma_{\mathrm{a}}$ and $\sigma = \sigma_{\mathrm{cr}} + \sigma_{\mathrm{a}}$, where σ_{a} is the atom radius. The interaction strength is set to $k_{\mathrm{cr}} = 1.0$. As typical, we describe the concentration of crowders with the fraction of the total simulation volume V occupied by the crowders, $\phi_{\mathrm{c}} = 4\pi R_{\mathrm{c}}^3 N_{\mathrm{cr}}/3V$. The number of crowding particles in our protein-crowder simulations range from $N_{\mathrm{cr}} = 6$ for $\phi_{\mathrm{c}} = 0.10$ and the largest crowder ($R_{\mathrm{c}} =$ 16 Å) to $N_{\mathrm{cr}} = 207$ for $\phi_{\mathrm{c}} = 0.44$ and the smallest crowder ($R_{\mathrm{c}} = 8$ Å).

3.2.2 Coarse-grained model for protein folding in the presence of crowders

As a model for protein folding, we use the coarse-grained "C_β-model" developed in Ref. [39], which includes three types of amino acids: hydrophobic (h), polar (p), and turn (t). While this three-letter alphabet is limited, it suffices to construct sequences with amphipathic secondary structure character. For example, repeats of the α-segment phpphhp can produce amphipathic α-helices and repeats of the β-segment ph can produce amphipathic β-strands. Such segments can be further organized into higher-order, tertiary structures through hydrogen bonding and effective hydrophobic attraction. We study three sequences: α_{35}, constructed from two identical 16-amino acid α-segments linked by ttt; β_{35}, five β-strand type segments, e.g., phphph, linked by four tt-segments; and β_{25}, constructed similarly to β_{35} but with only three β-segments, linked by two tt-segments (see table 3.1). In our model, α_{35} folds spontaneously into a stable helical hairpin [42], β_{35} fold into a stable 5-stranded β-barrel and β_{25} folds into a 3-stranded β-sheet structure.

Geometrically, the C_β-model includes seven atoms per amino acid. All backbone atoms are explicitly included (N, C_α, C', H, $H_{\alpha 1}$, and O) while the sidechain is represented by a single large C_β-atom. In contrast to p and h, type t lacks a C_β atom, which is replaced by an $H_{\alpha 2}$ atom. Hence, t strongly resembles a glycine residue. All bond lengths and angles and some dihedral angles (e.g. the peptide plane angle $\omega = 180°$) are held fixed at standard values. As a result of these constraints, the internal conformation of a chain with N amino acids is completely specified by the

$2N$ Ramachandran angles $\{\phi_{\rm i}, \psi_{\rm i}\}_{\rm i=1}^{N}$.

To incorporate crowders we extend here the C_β-model energy function $E_{\rm pr}$ [39] to include also terms for crowder-crowder ($E_{\rm cr}$) and crowder-protein ($E_{\rm cr\text{-}pr}$) interactions. The total energy function is thus $E = E_{\rm pr} + E_{\rm cr} + E_{\rm cr\text{-}pr}$. Crowder interactions are assumed to be pairwise additive and modeled with the pair potential $V(r)$, as described in section 3.2.1. Hence, for a system with $N_{\rm cr}$ crowders and $N_{\rm a}$ protein atoms,

$$E_{\rm cr} = \sum_{\rm i<j}^{N_{\rm cr}} V(r_{\rm ij}) \tag{3.2}$$

and

$$E_{\rm cr\text{-}pr} = \sum_{\rm i}^{N_{\rm cr}} \sum_{\rm j}^{N_{\rm a}} V(r_{\rm ij})\,, \tag{3.3}$$

where $r_{\rm ij}$ is the center-to-center distance between the crowder i and crowder j (Eq. 3.2) or between crowder i and protein atom j (Eq. 3.3).

The C_β-model energy function, $E_{\rm pr} = E_{\rm hp} + E_{\rm hbond} + E_{\rm exvol} + E_{\rm loc}$, is described in the following (see also Ref. [39]). The first term,

$$E_{\rm hp} = -k_{\rm hp} \sum_{\rm ij} e^{-(r_{\rm ij}-\sigma_{\rm hp})^2/2}\,, \tag{3.4}$$

describes effective hydrophobic interactions. The sum goes over all pairs of hydrophobic C_β atoms, excluding nearest and next-nearest amino acid neighbors along the chain. The strength of hydrophobicity is $k_{\rm hp} = 0.805$, and the energetically optimal C_β-C_β distance is $\sigma_{\rm hp} = 5$ Å. The second term represents hydrogen bonding between NH and C′O groups and can be written

$$E_{\rm hbond} = k_{\rm hbond} \sum_{\rm ij} \gamma_{\rm ij} \left[5 \left(\frac{\sigma_{\rm hb}}{r_{\rm ij}} \right)^{12} - 6 \left(\frac{\sigma_{\rm hb}}{r_{\rm ij}} \right)^{10} \right] \times (\cos\alpha_{\rm ij} \cos\beta_{\rm ij})^{\frac{1}{2}}\,, \tag{3.5}$$

where the sum is over all CO and NH pairs, but with the restriction that i and j are separated by at least two CO groups or two NH groups, and the strength is controlled by $k_{\mathrm{hbond}} = 3.1$. The strength is modified by a sequence-dependent scale factor taken to be $\gamma_{\mathrm{ij}} = 1$ for hh, hp, and pp pairs, and 0.75 for tt, th, and tp pairs. The reduced hydrogen bonding capacity of t amino acids is meant to mimic the tendency for glycine residues to break secondary structure through weaker hydrogen bonds [43]. For a hydrogen bond the optimal HO distance is $\sigma_{\mathrm{hb}} = 2.0$ Å. A directional dependence is implemented via the factor $(\cos\alpha_{\mathrm{ij}} \cos\beta_{\mathrm{ij}})^{\frac{1}{2}}$, where α_{ij} and β_{ij} are the NHO and HOC′ angles, respectively. Additionally, for any bond with either $\alpha_{\mathrm{ij}} < 90°$ or $\beta_{\mathrm{ij}} < 90°$, the bond contribution is set to zero. The third term is the excluded volume energy, which can be expressed as

$$E_{\mathrm{exvol}} = k_{\mathrm{exvol}} \sum_{\mathrm{i<j}} \left(\frac{\lambda_{\mathrm{ij}} \sigma_{\mathrm{ij}}}{r_{\mathrm{ij}}} \right)^{12}, \tag{3.6}$$

where the $k_{\mathrm{exvol}} = 0.10$. The sum is taken over all pairs of atoms ij connected by more than 2 covalent bonds. The scale factor $\lambda_{\mathrm{ij}} = 1.00$ for atom pairs connected by three covalent bonds. With two exceptions, all other global atom pairs have $\lambda_{\mathrm{ij}} = 0.75$. The two exceptions are carboxyl OO pairs and amide HH pairs for which $\lambda_{\mathrm{ij}} = 1.00$ and 1.25, respectively. The reduction factor $\lambda_{\mathrm{ij}} = 0.75$ for most global pairs is meant to accommodate the reduced flexibility of a chain with fixed bond lengths and angles. The parameter $\sigma_{\mathrm{ij}} = \sigma_{\mathrm{i}} + \sigma_{\mathrm{j}}$ is the sum of i and j atom radii, taken to be 1.75, 1.42, 1.55 and 1.00 Å for carbon, oxygen, nitrogen, and hydrogen atoms, respectively. An exception is C_β-C_β pairs for which $\sigma_{\mathrm{ij}} = 5.0$ Å, thereby accounting for some of the bulkiness of sidechains. The last term captures local interactions between partial

charges in adjacent peptide planes,

$$E_{\text{loc}} = k_{\text{local}} \sum_{\text{I}} \sum_{\text{i,j}} \frac{q_{\text{i}} q_{\text{j}}}{r_{\text{ij}}/\text{Å}}\,, \tag{3.7}$$

where the strength is $k_{\text{local}} = 50$ and the partial charges q_{i} are -0.2, $+0.2$, $+0.42$, -0.42 for N, H, C$'$, and O, respectively. The outer sum is over all amino acids I and the inner sum over the NC$'$, NO, HC$'$, and HO atom pairs of amino acid I.

3.2.3 Simulated tempering Monte Carlo

To find the thermodynamic behavior of various protein-crowder systems, as determined by the amino acid sequence, number of crowders, and the energy function $E(r)$, we use simulated tempering Monte Carlo (MC) [44–46]. In addition to a random walk in conformational space, as in basic Monte Carlo, simulated tempering also carries out a random walk in temperature while keeping the simulation at equilibrium. This is achieved by defining a set of temperatures, $\{T_{\text{j}}\}_{\text{j}=1}^{\text{M}}$, and simulating the joint probability distribution

$$P(r, j) \propto e^{-\beta_j E(r) + g_j}, \tag{3.8}$$

where $\beta_j = 1/k_{\text{B}}T_j$, k_{B} is Boltzmann's constant, and j has been made a dynamic parameter. The g_{j}'s are M simulation parameters that control the marginal distribution $P(j)$. Jumps between temperatures, $j \to j'$, are accomplished as MC updates, with acceptance probability

$$P_{\text{acc}}(r, j \to j') = \min\left[1, e^{-E(r)\left(\beta_{j'} - \beta_j\right) + g_{j'} - g_j}\right]\,. \tag{3.9}$$

A common choice for the g_j parameters, which we follow here, is to select them such that $P(j)$ is roughly flat, ensuring that sampling of conformations takes place at each temperature T_j.

3.2.4 Simulation and analysis details

Protein and crowders are placed in a $V = 100 \times 100 \times 100$ Å^3 box with periodic boundary conditions. Crowder positions are updated using single-particle translational MC moves, with random direction and maximum distance 8.7 Å. Two different types of MC moves are used for the protein chain: pivot moves and Biased Gaussian Steps (BGS) [47]. In pivots, a single $\psi_{\rm i}$ or $\phi_{\rm i}$ angle is chosen and assigned a new random value. This rotates the chain around a NC_α bond or a $C_\alpha C'$ bond, respectively. In BGS, eight consecutive $\psi_{\rm i}, \phi_{\rm i}$ angles are changed in a coordinated way to provide a roughly local chain deformation. The frequency of different updates are chosen as follows. Updates are divided equally between crowder particles and protein. The relative frequency between pivots and BGS is chosen to be temperature dependent, such that pivot dominates at high T and BGS dominates at low T. Temperature updates at attempted every 100 MC steps.

Simulated tempering runs for protein-crowder systems are carried out using eight and ten different temperatures in the range $k_{\rm B}T = 0.48$–0.68 for α_{35} and $k_{\rm B}T = 0.40$–0.70 for β_{35} and β_{25}. For each system with crowders, 8-10 independent runs are carried. Each simulation has at least 5×10^8 elementary MC steps for α_{35} and 4×10^9 elementary MC steps for both β-proteins. Initial conformations are created

by picking a random protein conformation and random crowder positions, followed by a relaxation step in which hard core steric clashes involving the crowders are removed.

Statistical analysis is carried out using the multistate Bennett acceptance ratio (MBAR) technique [48], which optimally combines statistics from different thermodynamic conditions. Respectively, we apply MBAR to combine the statistics from the 8 and 10 different T_j-values for α_{35} and both β proteins in our simulated tempering runs, and use it to determine the thermodynamic averages of various observables at narrowly spaced temperatures in the range $k_{\rm B}T = 0.40$–0.70 for all sequences. Since for α_{35} the range of simulated values is $k_{\rm B}T$ =0.48–0.68, the MBAR analysis involves an extrapolation to the range $k_{\rm B}T$ =0.40–0.70. Separate test simulations of β_{25}, covering a wider temperature range, confirm that the MBAR extrapolation is valid. Statistical errors are estimated from the 8-10 independent runs.

3.2.5 Observables

The number of native contacts is defined by $Q_{\rm nat} = \sum_{\rm i<j-3} \Delta_{\rm ij} C_{\rm ij}$, where the sum goes over pairs of residues ij, and $\Delta_{\rm ij} = 1$ if ij has formed a contact and 0 otherwise. A contact between amino acids i and j is considered formed if the distance between their C_β atoms is < 7 Å (positions with a t amino acid type, which lacks a C_β atom, do not contribute towards contact counts). The native contact set, C, is defined so that $C_{\rm ij} = 1$ if ij is a native contact, and otherwise 0. We use the native contact sets taken from our previous study [41]. We define the native state as $Q_{\rm nat} \geq Q_{\rm cut}$, where $Q_{\rm cut} = 50$, 68 and 33 for α_{35}, β_{35}, and β_{25} respectively. For β_{35} and β_{25}, $Q_{\rm cut}$ was

selected based on the peak in the free energy barrier along the order parameter $Q_{\rm nat}$, as shown in Ref. [41]. For α_{35}, which does not exhibit a clear folding barrier, $Q_{\rm cut}$ is selected such that the folding midpoint temperature, $T_{\rm m}^0$, roughly coincides with the temperature $T_{C_{\rm v,max}}$ at the maximum in the heat capacity curve (see Fig. A2). For $Q_{\rm cut} = 50$, we obtain $T_{\rm m}^0/T_{C_{\rm v,max}} = 0.988$). The number of nonnative contacts is determined from $Q_{\rm nonnat} = Q - Q_{\rm nat}$, where $Q = \sum_{\rm i<j-3} \Delta_{\rm ij}$ is the total number of contacts. The root-mean-square deviation, RMSD, is determined over all C_α-atoms. As a reference (native) structures for α_{35} and β_{35}, we pick the lowest energy conformations found from simulations with no crowders. These reference structures are similar to those shown in Fig 3.1.

3.3 Results

3.3.1 Native structures are not changed by crowders

Using the model for protein folding and the Monte Carlo sampling techniques described in Methods, we determine the thermodynamic behavior of the three model sequences, α_{35}, β_{35}, and β_{25} given in Table 1, in the presence of crowders with radii $R_{\rm c} = 12$ or 8 Å and volume fractions in the range $0 \leq \phi_{\rm c} \leq 0.44$. The crowders are comparable in size to the native structures of all three sequences, which have radii of gyration ranging from ≈ 8-9 Å. The radii of gyration of the unfolded state depends on T, as will be shown, but are for α_{35} and β_{35} in the range ≈14-15 Å and for β_{25}≈10-12 Å, at high temperatures. The excluded-volume effect is expected to affect

proteins when the concentration of crowders is high and the crowder size is similar, or smaller, that of the protein [49]. Because β_{25} is a shorter chain than α_{35} and β_{35}, our main analysis is carried out with crowders of size $R_c = 12$ Å for α_{35} and β_{35} and $R_c = 8$ Å for β_{25}.

As a first step in our analysis, we examine the temperature dependence of the native state population, P_{nat}, in the absence of crowders ($\phi_c = 0$), as shown in Fig. 3.1(D-F). For all the model proteins, the equilibrium folding curve is well described by a two-state equation, with only two free fit parameters (see Fig. A3). One of the free parameters is the folding midpoint temperature, T_m^0. We find (in model units) $k_B T_m^0 = 0.535$, 0.517, and 0.495 for α_{35}, β_{35}, and β_{25} respectively. In calculating P_{nat}, we define the native state, N, as $Q \geq Q_{cut}$, where Q is the number of native contacts, and Q_{cut} is chosen as described in Methods.

For simulations at $\phi_c > 0$, the protein chain must avoid overlapping with the hard cores of the crowding particles. Conversely, of course, the crowders must similarly avoid the protein chain. Because the volume available to the crowders is reduced when the protein is expanded, compact conformations will be entropically favored under crowded conditions. In our model, atoms on the protein chain can penetrate the soft shell of the crowder particles at an energetic cost. The thickness of this soft shell is controlled by a softness parameter, which we hold fixed at $\sigma_{cr} = 3.0$ Å (see Methods). At this σ_{cr}, the total protein-crowder and crowder-crowder repulsive energy, $E_{cr\text{-}pr} + E_{cr\text{-}cr}$, turns out to be small. For example, for $R_c = 12$ Å and $\phi_c = 0.20$, this repulsive energy per crowding particle is $\approx 0.1 k_B T$.

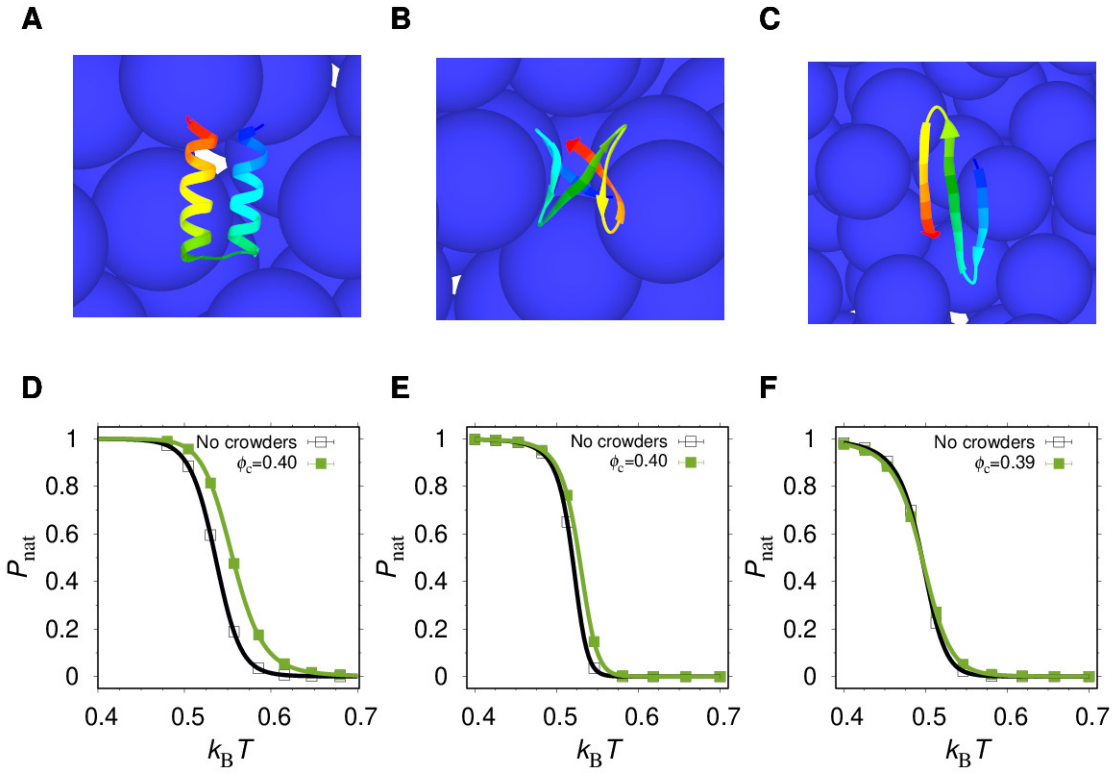

Figure 3.1: Folding curves and native structures. Representative low-energy conformations from Monte Carlo simulations of (A) α_{35}, (B) β_{35}, and (C) β_{25} in the presence of excluded volume crowders (blue spheres). The protein structures are shown in ribbon representation and rainbow color scheme (from N-terminus in blue to C-terminus in red). The native state population, $P_{\rm nat}$, as a function of temperature, for (D) α_{35}, (E) β_{35}, and (F) β_{25} in the absence of crowders (open black squares) and in the presence of crowders with radius $R_{\rm c} = 12$ Å (α_{35} and β_{35}) and radius $R_{\rm c} = 8$ Å (β_{25}) and volume fraction $\phi_{\rm c} \approx 0.40$ (solid green squares). Solid curves are obtained by using the multistate Bennett acceptance ratio (MBAR) reweighting technique [48]. Statistical errors are smaller than the plot symbols.

We find that the folding curves for the $\phi_c = 0$ and $\phi_c > 0$ cases are generally different, as illustrated in Fig. 3.1D-F (the folding curves for all ϕ_c values are given in Fig. A4). For all studied ϕ_c, $P_{nat} \approx 1$ at very low T. We conclude that for $\phi_c \leq 0.40$ the α_{35}, β_{35}, and β_{25} native structures are not substantially perturbed by the crowders.

3.3.2 Nonnative interactions are promoted under crowded conditions

Figure 3.2 shows the number of nonnative contacts, Q_{nonnat}, as a function of temperature across different ϕ_c. It is clear that Q_{nonnat} increases monotonically with ϕ_c except at very low T. Interestingly, for all model proteins, the Q_{nonnat} curve exhibits a peak at intermediate T, which can be understood in the non-crowding case in the following way. At high T, the chain is in an entropy dominated state with contacts formed and unformed basically at random, leading to a mix of native and nonnative contacts. The total number of contacts, $Q_{nat} + Q_{nonnat}$, is not maximal at this T, however, because the chain is expanded (see section 3.3.5). As T decreases, Q_{nonnat} initially increases because the chain becomes more compact but then abruptly decreases when T approaches T_m^0 due to folding (and Q_{nat} abruptly increases). The net result is a peak in Q_{nonnat} at $T \approx 1.05T_m$ for both α_{35} and β_{35}, and $T \approx 1.08T_m$ for β_{25}. As it turns out, the Q_{nonnat} peak remains under crowding conditions. Overall, we find that the excluded volume crowders generally promote the formation of nonnative interactions, except at very low T where the native state is thermodynamically

dominant. We note that the increase in Q_{nonnat} with ϕ_{c} is more than linear, as shown in Fig 3.2(insets).

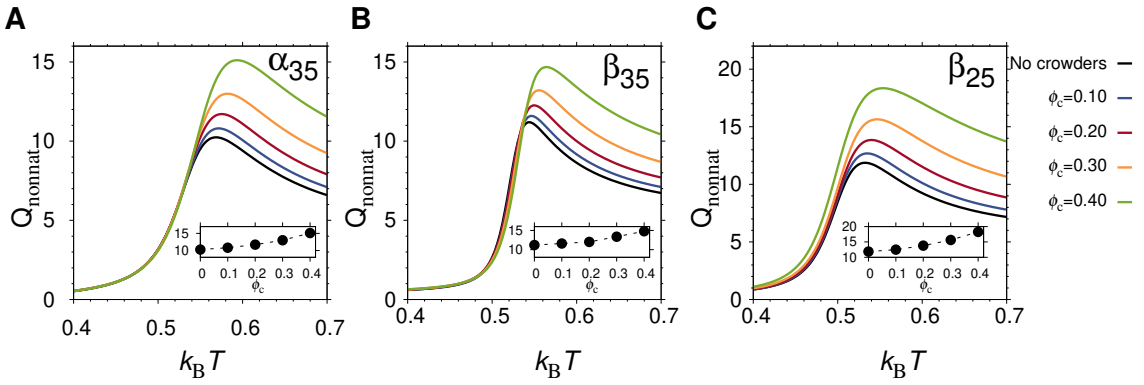

Figure 3.2: Formation of nonnative interactions under crowded and non-crowded conditions. Number of nonnative contacts, Q_{nonnat}, as a function of temperature for (A) α_{35} and (B) β_{35}, and β_{25} at different crowder volume fractions ϕ_{c}. Insets: Q_{nonnat} as a function of ϕ_{c} (solid circles) and fits to $Q_{\mathrm{nonnat}} \propto {\phi_{\mathrm{c}}}^{\gamma}$, giving $\gamma = 1.8$ for α_{35}, $\gamma = 1.9$ for β_{35}, and $\gamma = 1.6$ for β_{25}, (dashed curves) taken at $k_{\mathrm{B}}T = 0.588$ for α_{35}, $k_{\mathrm{B}}T = 0.569$ for β_{35}, and $k_{\mathrm{B}}T = 0.545$ for β_{25}, in α_{35} and β_{35} corresponding to a temperature $T \approx 1.10T_{\mathrm{m}}^{0}$ and to a temperature $T \approx 1.13T_{\mathrm{m}}^{0}$ for β_{25}. Results are for $R_{\mathrm{c}} = 12$ Å in α_{35} and β_{35} and for $R_{\mathrm{c}} = 8$ Å in β_{25}.

3.3.3 Excluded volume crowders can both increase and decrease native state stability

We turn now to the effect of crowders on the stability of α_{35}, β_{35}, and β_{25}. Because crowding effects can be strongly dependent on solution conditions [11, 33], we examine stability changes at two different temperatures, above ($T^{+} = 1.05T_{\mathrm{m}}^{0}$) and below

($T^- = 0.95T_{\rm m}^0$) the folding midpoint temperature in the absence of crowders ($T_{\rm m}^0$). As a direct measure of stability, we use the free energy of folding,

$$\Delta F = F_{\rm N} - F_{\rm U} = -k_{\rm B}T \ln \frac{P_{\rm nat}}{1 - P_{\rm nat}}\,, \tag{3.10}$$

where $F_{\rm N}$ and $F_{\rm U}$ are the free energies of the native and unfolded states, respectively, and we have assumed that the unfolded state population is $1 - P_{\rm nat}$. Figures 3.3A and B show the crowding-induced change in the free energy, $\Delta\Delta F(\phi_{\rm c}) = \Delta F(\phi_{\rm c}) - \Delta F^0$, where $\Delta F^0 = \Delta F(\phi_{\rm c} = 0)$, as a function of $\phi_{\rm c}$. At T^+, $\Delta\Delta F$ is negative for all sequences in all $\phi_{\rm c}$, indicating stabilization, except for β_{25} at the highest packing fraction $\phi_{\rm c} = 0.44$ (See Fig. 3.3B). At T^-, the situation is more complex. The stability of α_{35} increases with $\phi_{\rm c}$ as indicated by a monotonically decreasing $\Delta\Delta F$ (see Fig 3.3A). The β_{35} sequence is neither stabilized nor destabilized at T^- since $\Delta\Delta F$ is almost zero except at $\phi_{\rm c} = 0.30 - 0.40$ for which there is a subtle stabilization. The β_{25} sequence is destabilized at packing fractions $\phi_{\rm c} \geq 0.30$ as shown by positive $\Delta\Delta F$ (see Fig 3.3A). In terms of the midpoint temperature $T_{\rm m}$, over the entire range of volume fractions $\phi_{\rm c} = 0.0 - 0.40$, α_{35} and β_{35} are stabilized as indicated by an increase in $T_{\rm m}$. The β_{25} protein, as assessed by $T_{\rm m}$, exhibits a weak stabilization at low $\phi_{\rm c}$ but a sharp turn towards destabilization at $\phi_{\rm c} \approx 0.30$, in line with the picture obtained from considerations of $\Delta\Delta F$ at T^-. In particular, we note that, at the highest studied packing fraction, $\phi_{\rm c} = 0.44$, β_{25} is destabilized at both T^- and T^+ and in terms of both $\Delta\Delta F$ and $T_{\rm m}$.

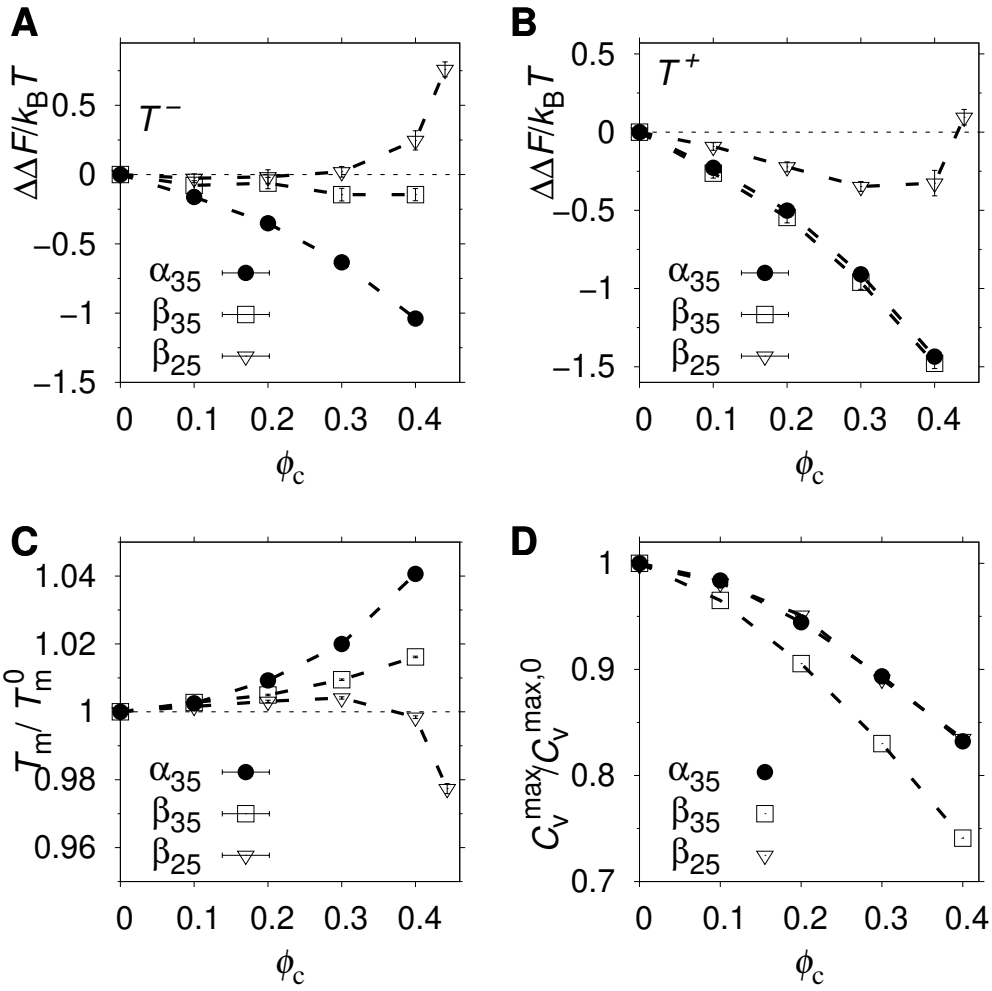

Figure 3.3: Effect of crowding on native state stability and folding cooperativity. Change in the free energy of folding, $\Delta\Delta F = \Delta F - \Delta F^0$, as a function of ϕ_c, at temperatures (A) $T^- = 0.95T_m^0$ and (B) $T^+ = 1.05T_m^0$. $\Delta\Delta F < 0$ indicates stabilization. (C) Midpoint folding temperature, T_m, and (D) maximum heat capacity, C_v^{max}, as functions of ϕ_c, where the heat capacity is determined from $C_v = \left(\langle E^2\rangle - \langle E\rangle^2\right)/k_BT^2$. Results are shown for α_{35} (filled circles) and β_{35} (open squares). ΔF^0, T_m^0 and $C_v^{max,0}$ are the $\phi_c = 0$ values of ΔF, T_m and C_v^{max}, respectively. The crowder radius is $R_c = 12$ Å for α_{35} and β_{35} and $R_c = 8$ Å for β_{25}. Dashed lines between points are drawn to guide the eye.

3.3.4 Crossover temperature

The temperature-dependent crowding effect for β_{25} (cf. Fig 3.3A-C) suggests there is a temperature, $T_{\rm cross}$, such that crowders enhance stability at $T > T_{\rm cross}$ but reduce stability at $T < T_{\rm cross}$. The β_{25} sequence does indeed exhibit such a crossover temperature, as seen in Fig 3.4C. The crossover temperature $k_{\rm B}T_{\rm cross} = 0.50$ turns out to be similar to $T_{\rm m}$ at $\phi_{\rm c} = 0.40$. The α_{35} sequence does not have a crossover temperature because it is always stabilized, even though the magnitude of the stabilization is still temperature dependent (see Fig 3.4A). Also, β_{35} appears to cross over into the destabilizing regime $\Delta\Delta F > 0$ at low T (see Fig. 3.4B), but it may not be statistically significant. The situation for β_{25} can be compared to that of ubiquitin, which was studied in the presence of synthetic (PVP or Ficoll) and protein (BSA or lysozyme) crowders [11]. Crossover temperatures were later determined for these systems [33]. For example, $T_{\rm cross}$ for ubiquitin in the presence of Ficoll at concentration 100 g/L was estimated to be 301 K, much lower than the folding midpoint temperature of this protein (370 K). The existence of $T_{\rm cross}$ was proposed to originate from "soft" attractive interactions between protein and crowders [33]. That such soft attractions can occur in the case of protein crowders is by now well established [1, 20, 36]. It is much less clear, however, if they occur for all synthetic crowders. Our results demonstrate that destabilization at low T can occur even in the absence of protein-crowder attractions.

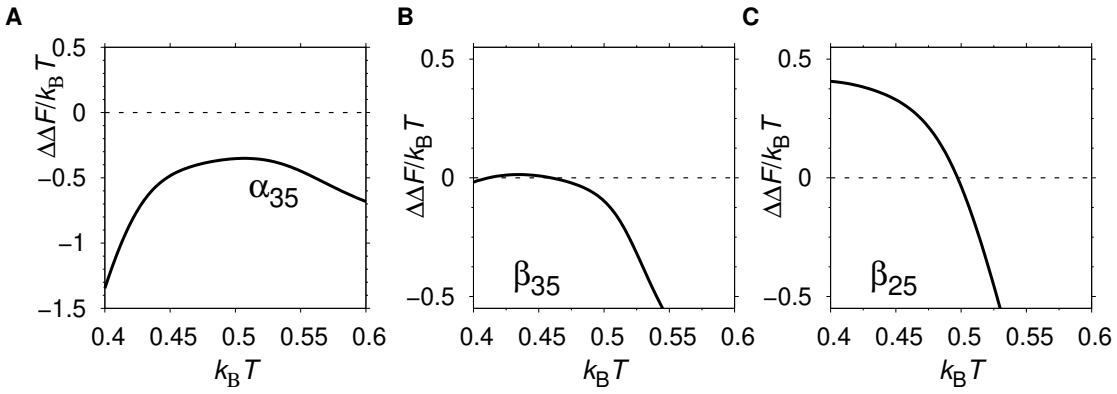

Figure 3.4: Crossover temperature. Existence of a crossover temperature, T_{cross}, in a system with only excluded volume crowders. $\Delta\Delta F$ as a function of temperature for (A) α_{35}, (B) β_{35}, and (C) β_{25} under fixed crowding conditions ($\phi_c = 0.20$ and $R_c = 12$ Å for α_{35} and β_{35}; $\phi_c = 0.40$ and $R_c = 8$ Å for β_{25}). A crossover temperature, defined by $\Delta\Delta F = 0$, exists for β_{25} and β_{35} but not for α_{35}.

3.3.5 Compact β_{25} unfolded state allows for crowding induced destabilization at low T

In order to understand the distinct responses of α_{35}, β_{35}, and β_{25} to crowded conditions, we examine more closely the character of the folding transition (see Fig 3.5.). All proteins undergo a chain collapse at low T, as seen Fig 3.5A, C and E (solid curves). At the highest studied T, the (average) radius of gyration of α_{35} and β_{35} is $R_{\rm g} \approx 14$–15 Å, which can be compared with the value ≈ 15.6 Å obtained from the scaling law $R_{\rm g} = R_0 N^{\nu}$ where R_0 is a constant and $\nu = 0.588$, which holds for fully chemically denatured proteins [50]. For β_{25}, $R_{\rm g} \approx 12$ Å at the highest T, compared to 12.8 Å from the scaling law. Hence, all three sequences transition from a random coil at high T to a much more compact folded state at low T.

It is instructive to consider also the size of the folded ($R_{\rm g}^{\rm N}$) and unfolded ($R_{\rm g}^{\rm U}$) state ensembles. In particular, the size (and shape) of the unfolded state is important for how volume exclusion affects protein stability [51]. The unfolded state is very sensitive to temperature changes (see Fig 3.5A, C, E). For all three sequences above $T_{\rm m}$, the unfolded state becomes increasingly compact as conditions become more stabilizing. Below $T_{\rm m}$, β_{35} and β_{25} unfolded states remain compact following the chain collapse, while a re-expansion occurs for α_{35}, i.e., the unfolded state becomes more expanded with decreasing T. This explains why α_{35} exhibits robust stabilization at both high and low Ts (see Fig. 3.4A), and an increase in $T_{\rm m}$ (see Fig. 3.3C). It also highlights the need for measuring crowding-induced stability changes over a range of temperatures [11].

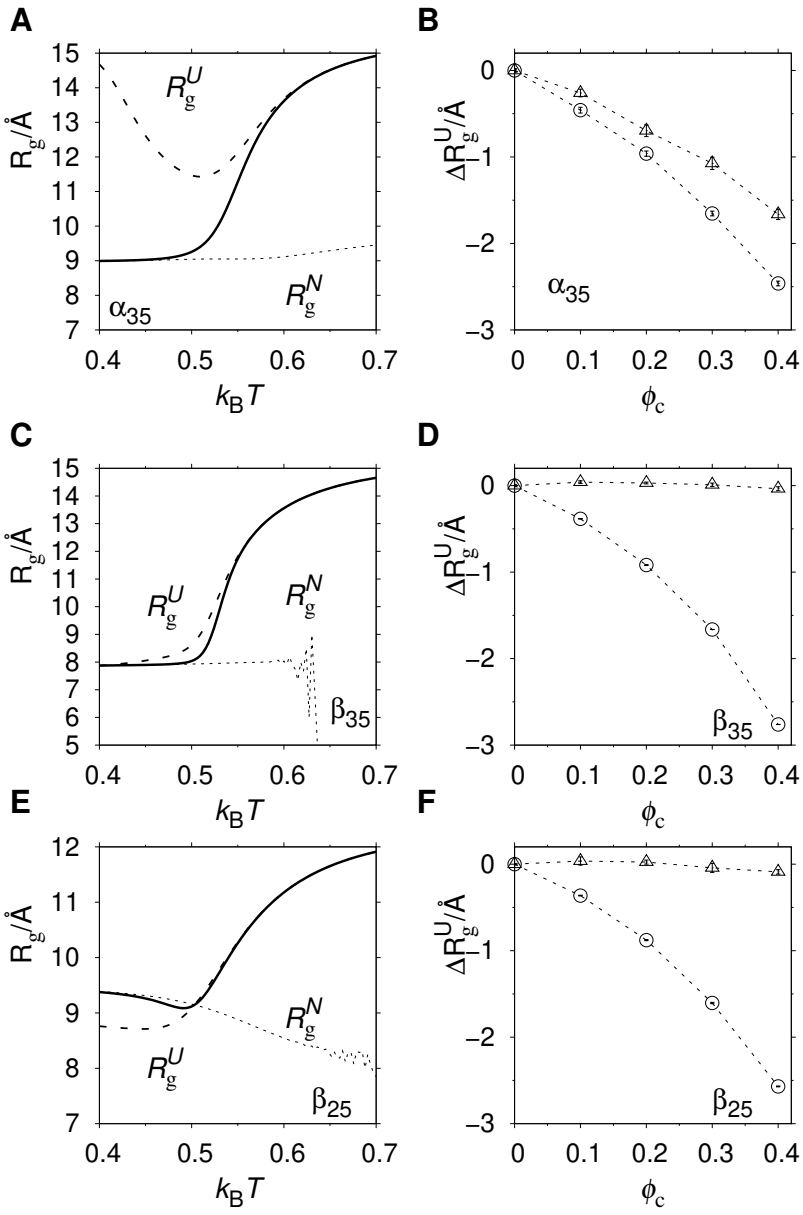

Figure 3.5: Effect of crowders on the size of the native and unfolded states. The average radius of gyration, $R_{\rm g}$, of (A) α_{35}, (C) β_{35}, and (E) β_{25} as a function of temperature (solid curves). Shown are also $R_{\rm g}$ determined over the unfolded ($R_{\rm g}^{\rm U}$, dashed) and native ($R_{\rm g}^{\rm N}$, dotted) ensembles. For β_{25}, $R_{\rm g}^{\rm U} < R_{\rm g}^{\rm N}$ for temperatures below $T = 0.503/k_{\rm B}$. Estimates of $R_{\rm g}^{\rm N}$ become unreliable at $k_{\rm B}T \gtrsim 0.6$ due to the small native state population at these Ts, especially for β_{25} and β_{35}. The change in unfolded state radius of gyration, $\Delta R_{\rm g}^{\rm U}(\phi_{\rm c}) = R_{\rm g}^{\rm U}(\phi_{\rm c}) - R_{\rm g}^{\rm U}(\phi_{\rm c} = 0)$, as function of $\phi_{\rm c}$, for (B) α_{35}, (D) β_{35} and (F) β_{25}, taken at the highest ($k_{\rm B}T = 0.70$, circles) and lowest ($k_{\rm B}T = 0.40$, triangles) studied temperatures.

Visual inspection of structures reveals that unfolded α_{35} is characterized by a partial or complete opening of the helical hairpin. As T decreases, the two α-helices become increasingly stable on their own and thus stiffer. At very low T, these stiff helices are unable to accommodate a hydrophobically collapsed unfolded state, but must instead dissociate while remaining well formed, which causes $R_{\mathrm{g}}^{\mathrm{U}}$ to increase. For β_{35} at low T, the unfolded state is almost as compact as the native state (see Fig. 3.5C), as assessed by R_{g}, which is in line with the close-to-neutral effect on stability from the crowders. Low-T unfolded β_{35} structures are characterized by a partial loss of the β-barrel organization, with one or two strands detached. These strands remain close to the remaining part of the barrel through hydrophobic attractions, which are strong at low T in our model, thereby leaving $R_{\mathrm{g}}^{\mathrm{U}}$ small.

The average R_{g} of the β_{25} unfolded state at the lowest studied temperature is $\approx$ 8.8 Å (see Fig. 3.1E). Hence, at this temperature, the R_{g} for the β_{25} unfolded state is slightly smaller than for the native state, as assessed using either the average value taken over the native state ensemble ($R_{\mathrm{g}} = 9.4$ Å; see Fig. 3.5E) or the value of the native conformation ($R_{\mathrm{g}} = 9.5$ Å; See Fig. 3.1C). For β_{25}, this unusual situation, in which the unfolded state is more compact than the native state, persists until the temperature reaches approximately T_{m} (see Fig. 3.5E). It is made possible by the combination of a relatively open native structure for this sequence (see Fig. 3.1C) and a compact unfolded state. Visual inspection of low-T unfolded β_{25} structures reveal a β-sheet rich state with two common features: (1) nonnative arrangements of the three β-strands, e.g., β-sheet structures with the N-terminal strand (residues 1-6) connected to the C-terminal strand (residues 19-25); and (2) the presence of β-hairpins

with hydrogen bonds between strands that are out-of-register, i.e., with β-hairpins with turns occurring at position different from that of the native structure. Overall, in the presence of excluded volume crowders, an expanded state will typically become more compact. The β_{25} unfolded state is already compact at low Ts (see Fig. 3.5E), and it will be basically unaffected by crowders (see Fig. 3.5F).

3.3.6 How is β_{25} destabilized by crowders at T^+?

As mentioned above, we showed that the β_{25} unfolded state is slightly more compact than its native state at low Ts. Accordingly, the crowder-induced destabilization of β_{25} at low Ts is easily understood. However, β_{25} is also destabilized at T^+, where the radius of gyration of the unfolded state is larger than that of the native state. Specifically, at this temperature, $R_{\rm g}^{\rm U} = 9.4$ Å while $R_{\rm g}^{\rm N} = 9.1$ Å. Then, how can β_{25} be destabilized under these conditions?

Crowding-induced protein destabilization must arise from a shift in the protein's conformational ensemble such that the population of the native state decreases relative to other states. For crowders with soft attractions, such a population shift away from the native state can be driven by favorable crowder-protein interactions that energetically stabilize the unfolded state. In order to grasp the reason why β_{25} is destabilized in high temperatures and $\phi_{\rm c} = 0.44$, (see Fig. 3.3B), it is necessary to consider the compactness of the unfolded state and its population relative to the native state.

As shown in Fig 3.6, at $\phi_{\rm c} = 0$, there is a small but nonnegligible population of

unfolded conformations that are more compact than the native structure. The fact that $R_g^U > R_g^N$ at this temperature is due to a rather long tail in the unfolded state distribution at high values of R_g. Naturally, such a high-R_g-tail is absent from the native state, which requires conformations to be highly compact. Upon the addition of crowders, there is a shift in the unfolded state ensemble towards more compact conformations. In particular, this shift almost entirely removes the high-R_g tail of the distribution (see Figure 3.6B). At $\phi_c = 0.44$, R_g^U is less than R_g^N. At this volume fraction the average radius of gyration for the unfolded state is $R_g^U = 7.9$ Å compared to $R_g^N = 8.6$ Å for the native state, which explains the destabilization. The shift towards more compact conformations can also be seen by considering the fraction of unfolded (nonnative) conformations that are more compact than the native structure (i.e., with $R_g < R_g^N$, where R_g^N is determined at $\phi_c = 0$), which is ≈ 40 % at $\phi_c = 0$, and increases by more than a factor of 2 to ≈ 90 % at $\phi_c = 0.44$. Note that this fraction slightly exaggerates the compactedness of nonnative conformation relative to the native state, because R_g^N decreases slightly with increasing ϕ_c.

Additional insight about the conformational ensembles can be found by considering the free energy surfaces $F(\mathrm{RMSD}, R_g)$, where the root-mean-square deviation RMSD is taken with respect to the native structure, as shown in figure 3.7 (A-F) for β_{25} and α_{35}. α_{35}, which is a sequence that is stabilized by the crowders at all Ts, exhibits large-RMSD (hence unfolded) conformations that are extended (high R_g). By contrast, such conformations are absent for β_{25} at very low Ts. At $T \approx T_m$ for β_{25}, extended large-R_g conformations do occur but these can be suppressed by the crowders (see Figure 3.6), leading to destabilization at high enough ϕ_c.

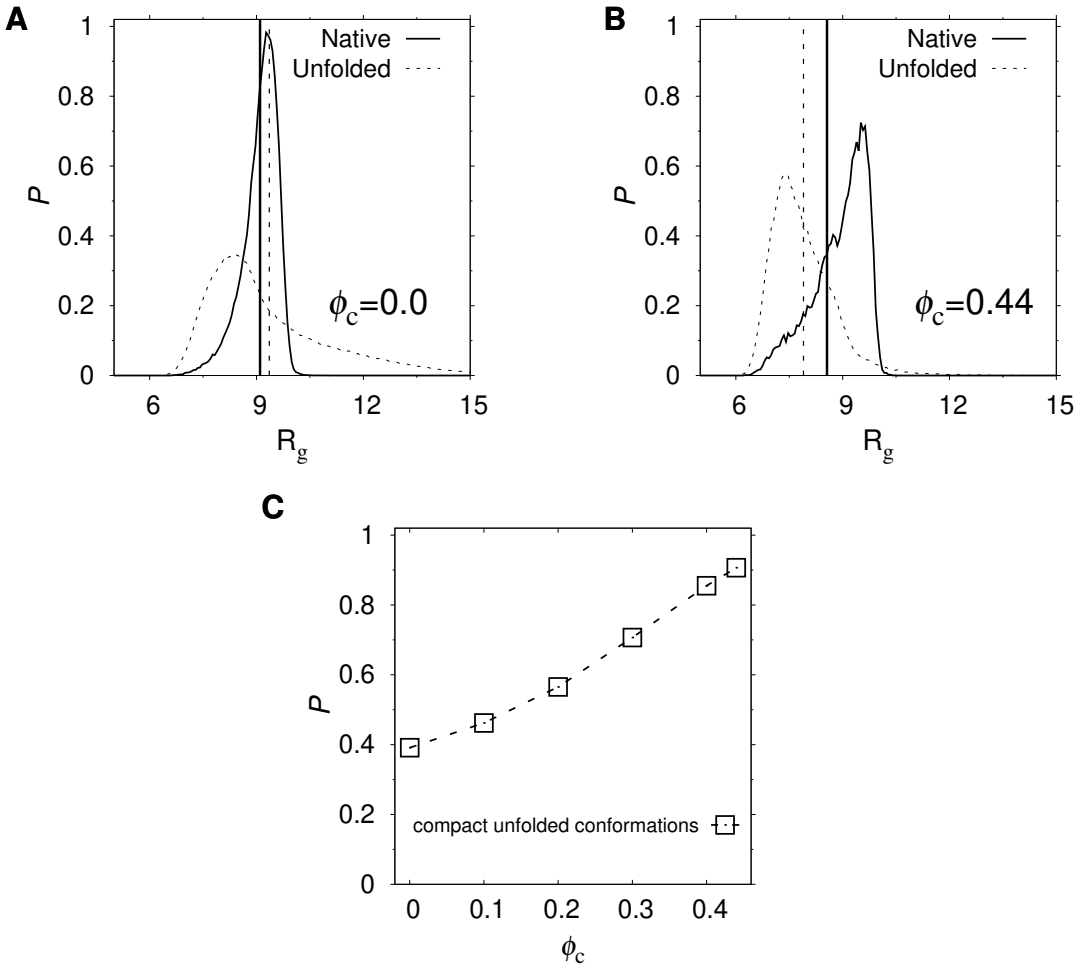

Figure 3.6: Radius of gyration histograms. The probability distribution function of the radius of gyration $R_{\rm g}$ for native (solid line) and unfolded (dots) states of β_{25} for (A) the no crowder case and (B) the highest studied volume fraction, $\phi_{\rm c} = 0.44$. Vertical lines show the average values, $R_{\rm g}^{\rm U}$ (dashed) and $R_{\rm g}^{\rm N}$ (solid). (C) The fraction of β_{25} unfolded state conformations with $R_{\rm g} < R_{\rm g}^{\rm N}$ as a function of the volume fraction of crowders $\phi_{\rm c}$. All values are taken at the temperature T^{+}. The radius of crowders is 8 Å.

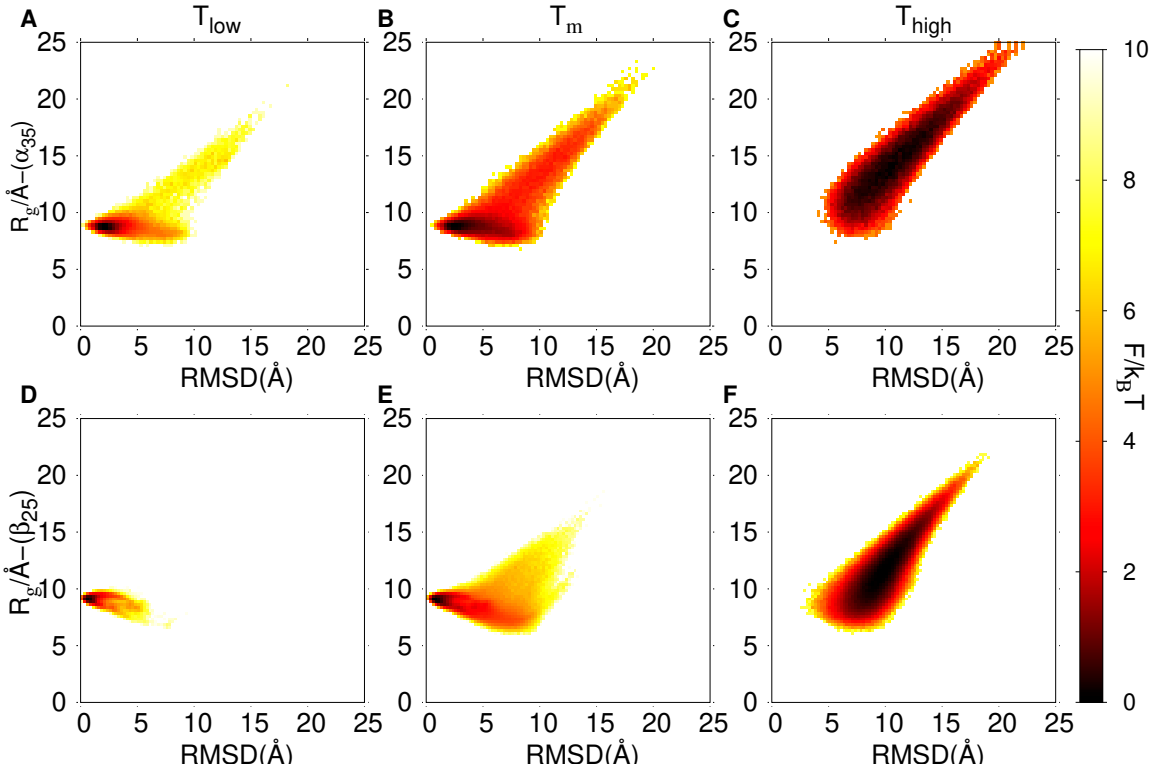

Figure 3.7: Compact nonnative state in the folding of β_{25}. Free-energy surface $F(\mathrm{RMSD}, R_\mathrm{g}) = -k_\mathrm{B}T \ln P(\mathrm{RMSD}, R_\mathrm{g})$, where the joint probability distribution $P(\mathrm{RMSD}, R_\mathrm{g})$ is taken at T_low, T_m, and T_high, which are the lowest studied temperature ($k_\mathrm{B}T = 0.40$), the midpoint temperature (T_m^0), and the highest studied temperature ($k_\mathrm{B}T = 0.70$), respectively, for (A-C) α_{35} and (D-F) β_{25}. RMSD is taken with respect to the representative native structure of α_{35} and β_{25}(see Fig. 1A and C).

3.3.7 Apparent stabilization effect is observable dependent

Finally, we examine the folding progress in variables other than $P_{\rm nat}$. Figure 3.8 shows the $\phi_{\rm c}$-dependence of the end-to-end distance, $R_{\rm ee}$, and the secondary structure content. We note especially that the destabilization of β_{25} is not apparent in $R_{\rm ee}$, which follows a trend closely related to that of $R_{\rm g}$. Examining the secondary structure for β_{25} shows the β-structure content is rather insensitive to an increase in the crowder concentration for $\phi_{\rm c} \leq 0.40$, but there is a detectable change at $\phi_{\rm c} = 0.44$ (see Fig. 3.8 D). This change thus coincides with the decrease in $P_{\rm nat}$ at T^+, which occurs at $\phi_{\rm c} = 0.44$ (see Fig. A4).

3.4 Discussion

It was first realized on theoretical grounds that the native state of proteins should be stabilized by the presence of surrounding macromolecules if these macromolecules are inert and simply occupy space [7]. The reason is that the unfolded state, on account of its conformationally expanded character, will leave a smaller volume for the crowder molecules to occupy than the volume left by the more compact native conformation. As a result, the native state will be entropically favored relatively to the unfolded state, which should stabilize the protein. Indeed, there is wide support for at least a moderately stabilizing effect from experiments [8–16], and from theory [19–25, 34].

However, as noted by Minton [52], the above theoretical argument can be applied not just to the native state but to any compact nonnative state, which will also be sta-

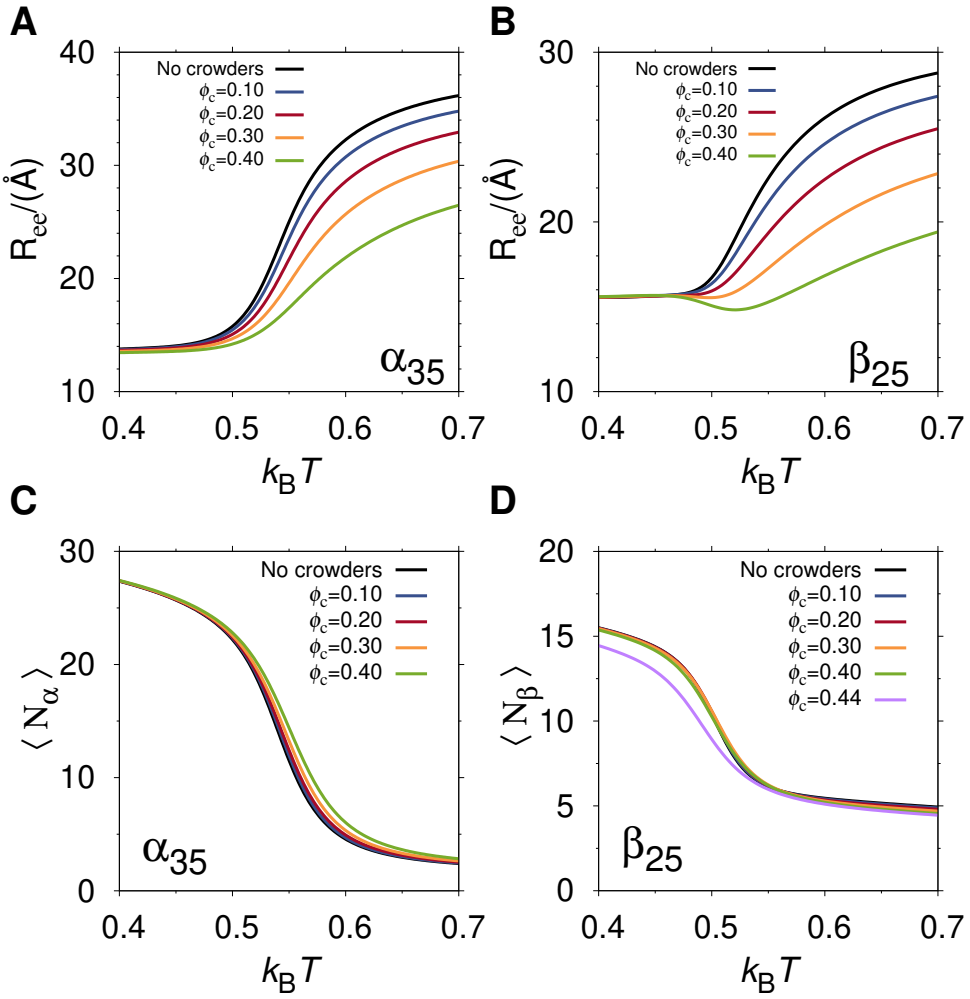

Figure 3.8: Impact of crowding on secondary structure content and end-to-end distance. Averages of (A, B) end-to-end distance, R_{ee}, (C) number of helical amino acids, N_α, and (D) number of β-sheet amino acids, N_β, as functions of temperature, shown for α_{35} and/or β_{25} and different ϕ_c as indicated. In determining N_α and N_β, a residue position i is classified in the following way: a helical state if $-90° < \phi_i < -30°$ and $-77° < \psi_i < -17°$, and a β-sheet state if $-160° < \phi_i < -50°$ and $100° < \psi_i < 160°$, where the ϕ_i and ψ_i are the Ramachandran angles of residue i. R_{ee} is the C_α-C_α distance between the terminal amino acids.

bilized relative to more expanded conformations. Indeed, compaction of the unfolded state ensemble under crowding has been observed for several proteins [13, 53, 54], although an exception was recently found [55]. Intrinsically disordered proteins also tend to become more compact under crowded conditions [56]. Even for a strictly two-state protein, the excluded volume effect is not expected to be strongly stabilizing if nonnative, attractive interactions result in a collapsed unfolded state. Computer simulations of structure-based models for protein folding, which do not permit energetically favorable nonnative interactions within the protein, consistently lead to enhanced stability along with compaction of the unfolded state [23–25].

The perspective provided by our model is partly different in that it takes into account attractive nonnative (intra-protein) interactions. We have found that such nonnative interactions during folding become increasingly prevalent with increasing crowder concentration, even when the native structure is left structurally unchanged. Moreover, these nonnative interactions can turn the excluded volume effect from being stabilizing to destabilizing under some conditions (see Fig 3.3). The destabilization we observe for β_{25} occurs at low temperatures with volume fraction $\phi_c \geq 0.30$ and high temperatures with volume fraction $\phi_c > 0.40$, through two main factors. Firstly, this protein has a relatively compact unfolded state at low temperatures. As a result, it is weakly favored entropically by the crowders. At higher T, even though the unfolded state is less compact than the native state, the protein can still be destabilized by the excluded volume effect because of the crowder-induced compaction of the unfolded state.

Sparsely populated nonnative states have been detected for several small globular

proteins [30, 57–59], e.g., using nuclear magnetic resonance (NMR) spectroscopy [58], hydrogen-deuterium exchange [30] and other methods [59], and may be more common than previously thought [60]. For example, using relaxation dispersion NMR, Neudecker *et al.* [57] showed that a β-barrel native fold exhibits a compact near-native intermediate state with an $\approx$2% population stabilized by both native and nonnative interactions.

Although artificial polymer crowders, such as Ficoll or dextran, are typically stabilizing to proteins (summarized in recent reviews [61, 62]), exceptions have been seen [11, 16–18]. For example, weak destabilization of ubiquitin [11] and apoazurin [18] were observed with Ficoll-70 as a crowder. Crowding induced destabilization is often interpreted as evidence for soft attractive (or chemical) interactions between the unfolded protein and crowders. Because of the presence of an energetic component, soft attractions will lead to a crossover temperature below which crowding becomes destabilizing [33]. Here, we have shown that nonnative states during folding can lead to a destabilization of the native state, as well as a crossover temperature, without soft attractions. This idea could be tested through crowding experiments on proteins for which sparsely populated compact nonnative (or intermediate) states have already been characterized.

All of our model proteins exhibit a compaction of the unfolded state with decreasing temperature, i.e., as conditions increasingly promote folding, which is in line with data from single-molecule FRET (smFRET) and small-angle X-ray scattering (SAXS) experiments [63]. Interestingly, Radford *et al.*, using smFRET, additionally detected an expansion of the unfolded state of the α-helical protein Im9 at low very

denaturant concentrations [64]. This behavior mirrors the re-expansion at low T we observe for α_{35}, which does not occur for β_{35} and β_{25} (see Fig 3.5). The stabilization of α_{35} at both high and low Ts results in a clear increase in T_{m} with ϕ_{c}. In contrast, β_{25} is destabilized at T^{-} and T^{+} as changes in T_{m}. Indeed, capturing small changes in stability may require stability measurements across a range of temperatures, as was pointed out previously [11, 65].

3.5 Conclusion

In summary, we have used a coarse-grained sequence-based model to study the folding and stability of three different sequences in the presence of excluded volume crowders over a range of sizes and concentrations. We find that during folding, nonnative interactions are generally promoted by the crowders. Moreover, under low and high Ts, in sufficiently high-ϕ_{c} conditions for each temperature, the excluded volume effect of crowders can lead to a destabilization of the protein even when its native structure remains unchanged and the population of nonnative conformations is relatively small, as observed in our model. Such destabilization may, however, not be apparent in observables reporting on the total content of secondary structure or overall chain size. The results suggested by our model may be tested experimentally, for example, using an artificial crowder molecule, such as Ficoll, on proteins for which a low population of compact nonnative conformations has been detected [57, 66, 67].

3.6 Acknowledgements

SW acknowledges support from the Natural Sciences and Engineering Research Council of Canada (grant RGPIN-2016-05104). This research was enabled in part by the computational resources provided by the Digital Research Alliance of Canada.

Bibliography

[1] F. X. Theillet, A. Binolfi, T. Frembgen-Kesner, K. Hingorani, M. Sarkar, C. Kyne, C. Li, P. B. Crowley, L. Gierasch, G. J. Pielak, A. H. Elcock, A. Gershenson, and P. Selenko. Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs). *Chem Rev*, 114:6661–6714, 2014.

[2] K. Luby-Phelps. Cytoarchitecture and physical properties of cytoplasm: volume, viscosity, diffusion, intracellular surface area. *Int rev cyt*, 192:189–221, 1999.

[3] R. J. Ellis. Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci*, 26:597–604, 2001.

[4] T. Niwa, R. Sugimoto, L. Watanabe, S. Nakamura, T. Ueda, and H. Taguchi. Large-scale analysis of macromolecular crowding effects on protein aggregation using a reconstituted cell-free translation system. *Front Microbiol*, 6:1113, 2015.

[5] A. A. M. André and E. Spruijt. Liquid-liquid phase separation in crowded environments. *Int J Mol Sci*, 21:5908, 2020.

[6] A. P. Minton. The influence of macromolecular crowding and macromolecu-

lar confinement on biochemical reactions in physiological media. *J Biol Chem*, 276:10577–10580, 2001.

[7] A. P. Minton. Excluded volume as a determinant of protein structure and stability. *Biophys J*, 32:77–79, 1980.

[8] K. Sasahara, P. McPhie, and A. P. Minton. Effect of dextran on protein stability and conformation attributed to macromolecular crowding. *J Mol Biol*, 326:1227–1237, 2003.

[9] D. S. Spencer, K. Xu, T. M. Logan, and H. X. Zhou. Effects of pH, salt, and macromolecular crowding on the stability of FK506-binding protein: an integrated experimental and theoretical study. *J Mol Biol*, 351:219–232, 2005.

[10] L. A. Benton, A. E. Smith, G. B. Young, and G. J. Pielak. Unexpected effects of macromolecular crowding on protein stability. *Biochemistry*, 51:9773–9775, 2012.

[11] Y. Wang, M. Sarkar, A. E. Smith, A. S. Krois, and G. J. Pielak. Macromolecular crowding and protein stability. *J Am Chem Soc*, 134:16614–16618, 2012.

[12] B. Köhn and M. Kovermann. Macromolecular crowding tunes protein stability by manipulating solvent accessibility. *Chem Bio Chem*, 20:759–763, 2019.

[13] J. Hong and L. M. Gierasch. Macromolecular crowding remodels the energy landscape of a protein by favoring a more compact unfolded state. *J Am Chem Soc*, 132:10445–10452, 2010.

[14] S. Mittal and L. R. Singh. Denatured state structural property determines protein stabilization by macromolecular crowding: a thermodynamic and structural approach. *PLOS One*, 8:e78936, 2013.

[15] A. Christiansen and P. Wittung-Stafshede. Synthetic crowding agent dextran causes excluded volume interactions exclusively to tracer protein apoazurin. *FEBS Lett*, 588:811–814, 2014.

[16] A. Malik, J. Kundu, S. K. Mukherjee, and P. K. Chowdhury. Myoglobin unfolding in crowding and confinement. *J Phys Chem B*, 116:12895–12904, 2012.

[17] K. Nasreen, S. Ahamad, F. Ahmad, M. I. Hassan, and A. Islam. Macromolecular crowding induces molten globule state in the native myoglobin at physiological pH. *Int J Biol Macromol*, 106:130–139, 2018.

[18] A. Christiansen and P. Wittung-Stafshede. Quantification of excluded volume effects on the folding landscape of Pseudomonas aeruginosa apoazurin in vitro. *Biophys J*, 105:1689–1699, 2013.

[19] S. R. McGuffee and A. H. Elcock. Protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLOS Comput Biol*, 6:e1000694, 2010.

[20] A. Bille, S. Mohanty, and A. Irbäck. Peptide folding in the presence of interacting protein crowders. *J Chem Phys*, 144:175105, 2016.

[21] I. Yu, T. Mori, T. Ando, R. Harada, J. Jung, Y. Sugita, and M. Feig. Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *Elife*, 5, 2016.

[22] M. Candotti and M. Orozco. The Differential Response of Proteins to Macromolecular Crowding. *PLOS Comput Biol*, 12:e1005040, 2016.

[23] M. S. Cheung, D. Klimov, and D. Thirumalai. Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc Natl Acad Sci USA*, 102:4753–4758, 2005.

[24] J. Mittal and R. B. Best. Dependence of protein folding stability and dynamics on the density and composition of macromolecular crowders. *Biophys J*, 98:315–320, 2010.

[25] D. Tsao and N. V. Dokholyan. Macromolecular crowding induces polypeptide compaction and decreases folding cooperativity. *Phys Chem Chem Phys*, 12:3491–3500, 2010.

[26] B. Macdonald, S. McCarley, S. Noeen, and A. E. van Giessen. β-hairpin crowding agents affect α-helix stability in crowded environments. *J Phys Chem B*, 120:650–659, 2016.

[27] N. Go and H. Taketomi. Respective roles of short- and long-range interactions in protein folding. *Proc Natl Acad Sci USA*, 75:559–563, 1978.

[28] A. C. Miklos, C. Li, N. G. Sharaf, and G. J. Pielak. Volume exclusion and soft interaction effects on protein stability under crowded conditions. *Biochemistry*, 49:6984–6991, 2010.

[29] A. H. Gorensek-Benitez, A. E. Smith, S. S. Stadmiller, G. M. Perez Goncalves,

and G. J. Pielak. Cosolutes, crowding, and protein folding kinetics. *J Phys Chem B*, 121:6527–6537, 2017.

[30] R. D. Cohen and G. J. Pielak. Quinary interactions with an unfolded state ensemble. *Protein Sci*, 26:1698–1703, 2017.

[31] M. Sarkar, C. Li, and G. J. Pielak. Soft interactions and crowding. *Biophys Rev*, 5:187–194, 2013.

[32] M. Senske, L. Törk, B. Born, M. Havenith, C. Herrmann, and S. Ebbinghaus. Protein stabilization by macromolecular crowding through enthalpy rather than entropy. *J Am Chem Soc*, 136:9036–9041, 2014.

[33] H-X Zhou. Polymer crowders and protein crowders act similarly on protein folding stability. *FEBS Lett*, 587:394–397, 2013.

[34] R. Harada, N. Tochio, T. Kigawa, Y. Sugita, and M. Feig. Reduced native state stability in crowded cellular environment due to protein-protein interactions. *J Am Chem Soc*, 135:3696–3701, 2013.

[35] A. C. Miklos, M. Sarkar, Y. Wang, and G. J. Pielak. Protein crowding tunes protein stability. *J Am Chem Soc*, 133:7116–7120, 2011.

[36] S. Timr and F. Sterpone. Stabilizing or destabilizing: simulations of chymotrypsin inhibitor 2 under crowding reveal existence of a crossover temperature. *J Phys Chem Lett*, 12:1741–1746, 2021.

[37] L. Sapir and D. Harries. Origin of enthalpic depletion forces. *J Phys Chem Lett*, 5:1061–1065, 2014.

[38] R. Politi and D. Harries. Enthalpically driven peptide stabilization by protective osmolytes. *ChemComm*, 46:6449–6451, 2010.

[39] A. Bhattacherjee and S. Wallin. Coupled folding-binding in a hydrophobic/polar protein model: impact of synergistic folding and disordered flanks. *Biophys J*, 102:569–578, 2012.

[40] M. H. Cordes, A. R. Davidson, and R. T. Sauer. Sequence space, folding and protein design. *Curr Opin Struct Biol*, 6:3–10, 1996.

[41] D. Trotter and S. Wallin. Effects of topology and sequence in protein folding linked via conformational fluctuations. *Biophys J*, 118:1370–1380, 2020.

[42] C. Holzgräfe and S. Wallin. Local versus global fold switching in protein evolution: insight from a three-letter continuous model. *Phys Biol*, 12:026002, 2015.

[43] J. S Merkel and L. Regan. Aromatic rescue of glycine in β sheets. *Fold Des*, 3:449–456, 1998.

[44] E. Marinari and G. Parisi. Simulated Tempering: a new Monte Carlo scheme. *Europhys Lett*, 19, 1992.

[45] A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov. New approach to Monte Carlo calculation of the free energy: method of expanded ensembles. *J Chem Phys*, 96:1776–1783, 1992.

[46] A. Irbäck and F. Potthast. Studies of an off-lattice model for protein folding: Sequence dependence and improved sampling at finite temperature. *J Chem Phys*, 103:10298, 1995.

[47] G. Favrin, A. Irbäck, and F. Sjunnesson. Monte Carlo update for chain molecules: biased Gaussian steps in torsional space. *J Chem Phys*, 114:8154–8158, 2001.

[48] M. R. Shirts and J. D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys*, 129:124105, 2008.

[49] H-X Zhou, G Rivas, and A P Minton. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu Rev Biophys*, 37:375–397, 2008.

[50] 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. *Proc Natl Acad Sci USA*, 101:12491–12496, 2004.

[51] A. Politou and P. A. Temussi. Revisiting a dogma: the effect of volume exclusion in molecular crowding. *Curr Opin Struct Biol*, 30:1–6, 2015.

[52] A P Minton. Implications of macromolecular crowding for protein assembly. *Curr Opin Struct Biol*, 10:34–39, 2000.

[53] R. Engel, A. H. Westphal, D. H. E. W. Huberts, S. M. Nabuurs, S. Lindhoud, A. J. W. G. Visser, and C. P. M. van Mierlo. Macromolecular crowding compacts unfolded apoflavodoxin and causes severe aggregation of the off-pathway intermediate during apoflavodoxin folding. *J Biol Chem*, 283:27383–27394, 2008.

[54] T. Mikaelsson, J. Adén, L. B. Johansson, and P. Wittung-Stafshede. Direct ob-

servation of protein unfolded state compaction in the presence of macromolecular crowding. *Biophys J*, 104:694–704, 2013.

[55] F. C. Zegarra, D. Homouz, A. G. Gasic, L. Babel, M. Kovermann, P. Wittung-Stafshede, and M. S. Cheung. Crowding-induced elongated conformation of urea-unfolded apoazurin: investigating the role of crowder shape in silico. *J Phys Chem B*, 123:3607–3617, 2019.

[56] D. Johansen, C. M. Jeffries, B. Hammouda, J. Trewhella, and D. P. Goldenberg. Effects of macromolecular crowding on an intrinsically disordered protein characterized by small-angle neutron scattering with contrast matching. *Biophys J*, 100:1120–1128, 2011.

[57] P. Neudecker, P. Robustelli, A. Cavalli, P. Walsh, P. Lundström, A. Zarrine-Afsar, S. Sharpe, M. Vendruscolo, and L. E. Kay. Structure of an intermediate state in protein folding and aggregation. *Science*, 336:362–366, 2012.

[58] A. Sekhar and L. E. Kay. NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc Natl Acad Sci USA*, 110:12867–12874, 2013.

[59] S. A. Raab, T. J. El-Baba, D. W. Woodall, W. Liu, Y. Liu, Z. Baird, D. A. Hales, A. Laganowsky, D. H. Russell, and D. E. Clemmer. Evidence for many unique solution structures for chymotrypsin inhibitor 2: A thermodynamic perspective derived from vT-ESI-IMS-MS Measurements. *J Am Chem Soc*, 142:17372–17383, 2020.

[60] D. J. Brockwell and S. E. Radford. Intermediates: ubiquitous species on folding energy landscapes? *Curr Opin Struct Biol*, 17:30–37, 2007.

[61] A. Christiansen, Q. Wang, M. S. Cheung, and P. Wittung-Stafshede. Effects of macromolecular crowding agents on protein folding in vitro and in silico. *Biophys Rev*, 5:137–145, 2013.

[62] S. Shahid, M. I. Hassan, A. Islam, and F. Ahmad. Size-dependent studies of macromolecular crowding on the thermodynamic stability, structure and functional activity of proteins: in vitro and in silico approaches. *Biochim Biophys Acta Gen Subj*, 1861:178–197, 2017.

[63] R. B. Best. Emerging consensus on the collapse of unfolded and intrinsically disordered proteins in water. *Curr Opin Struct Biol*, 60:27–38, 2020.

[64] T. Tezuka-Kawakami, C. Gell, D. J. Brockwell, S. E. Radford, and D. A. Smith. Urea-induced unfolding of the immunity protein Im9 monitored by spFRET. *Biophys J*, 91:L42–44, 2006.

[65] C. Alfano, D. Sanfelice, S. R. Martin, A. Pastore, and P. A. Temussi. An optimized strategy to measure protein stability highlights differences between cold and hot unfolded states. *Nat Commun*, 8:15428, 2017.

[66] A. Dasgupta and J. B. Udgaonkar. Transient non-native burial of a Trp residue occurs initially during the unfolding of a SH3 domain. *Biochemistry*, 51:8226–8234, 2012.

[67] A. Zarrine-Afsar, S. Wallin, A. M. Neculai, P. Neudecker, P. L. Howell, A. R.

Davidson, and H. S. Chan. Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. *Proc Natl Acad Sci USA*, 105:9999–10004, 2008.

Chapter 4

Exploring soft interactions in crowded systems: repulsive, nonspecific, hydrophobic and polypeptide crowders

Abstract

The equilibrium stability of a protein is determined by its amino acid sequence and the solution conditions, such as temperature, pH, and chemical denaturant. The stability of a single protein in two identical solutions can nonetheless differ if other macromolecules, termed cosolutes or crowders, are present in one of the solutions at concentrations high enough to occupy a substantial fraction of the solution volume. This effect, due solely to the presence of the crowders, decreases or increases the stability depending on the interactions between the protein and crowders. Hard-core repulsions, which are always present, typically increase the stability by entropically favoring the compact native conformation over extended unfolded conformations. Soft attractive interactions between protein and crowders can counteract the stabilizing effect from hard-core repulsions, even leading to a net destabilization. Soft interactions are typically assumed to be nonspecific. Here we use a coarse-grained model for protein folding to assess the impact of different types of soft interactions, by considering crowders that are spherical particles or short polypeptide chains. In particular, we study the stability of a 35-amino acid model sequence folding into a helical bundle fold. We find that, for the same crowder concentration and interaction strength, spherical crowders with a hydrophobic character are more destabilizing than crowders capable only of nonspecific interactions. The hydrophobic crowders are more destabilizing even though the magnitude of the crowder-protein interaction energy is much smaller for this crowder type. Short polypeptide crowders, which are capable of hydrogen bonding with the protein, have a destabilizing effect even at relatively low crowder concentrations, especially if the sequence of the peptide crowders includes

hydrophobic amino acids. Such destabilization is in part driven by misfolding of the protein through the formation of inter-molecular β-sheets. These findings emphasize the importance of understanding the interplay between energetic and entropic effects in determining the structure and stability of proteins under crowded conditions.

4.1 Introduction

Most biophysical experiments on biomolecules or biomolecular systems are performed under dilute solution conditions in which the macromolecular concentration rarely exceeds 10 g/L [1]. However, the cellular environment is often anything but dilute [2]. For example, the concentration of macromolecules in *Escherichia coli* can reach up to 300-400 g/L, corresponding to a volume occupancy of around 30-40% [3]. This crowded milieu has been shown to impact a wide range of biophysical processes, including protein folding [4–6], assembly [7] and aggregation [8–10], DNA replication [11], and liquid-liquid phase separation [12–14].

One of the major issues in macromolecular crowding, in fact since the inception of the field [15], is the impact of macromolecular crowding on the equilibrium stability of proteins. In the simplest case, the protein populates mainly two states: the structurally coherent native state (N) and the unfolded state (U). Different types of interactions between crowder and protein are expected to differently impact the stability. Hard-core repulsions, arising simply from the fact that two macromolecules cannot simultaneously occupy the same region in space, is expected to be a stabilizing factor for protein. The reason is that N is typically more compact than the extended conformations that make up the U ensemble, which thus become entropically disfavored relative to N under crowded conditions (see Fig. 4.1A). Experimental studies employing artificial polymer macromolecules as crowders, typically [16–21], but not always [22, 23], stabilize N as measured by, e.g., the midpoint temperature of the folding transition or the free energy of folding. Recent computer simulations of small

model proteins have shown that volume exclusion by purely repulsive crowders can lead to destabilization under some conditions [24].

In addition to hard-core steric repulsions, so-called soft interactions (also "chemical" interactions) between the protein and surrounding crowder molecules also impact protein stability. Soft interactions can be either attractive or repulsive [25], and they are generally assumed to be nonspecific and originate from a range of different effects, including charge-charge, van der Waals, hydrogen bonding, and hydrophobic interactions [26–28]. Repulsive soft interactions are expected to enhance the effect of volume exclusion, i.e., make the protein even more stable. Soft interactions that are attractive counteract the excluded volume effect and act to decrease the stability [25]. Recent studies have been performed to study the effects of attractive crowder-protein interactions [28–32]. In general, the net effect of macromolecular crowding on the stability of proteins will be determined by a competition between stabilizing effect of excluded volume and destabilizing chemical attractions. Because soft interactions have an energetic (or enthalpic) component, the impact of crowders should be temperature-dependent. Indeed, this was observed by Zhou *et al.*, who studied the stability of the protein CI2 in the presence of protein crowders [33]. CI2 was found to be destabilized at low temperatures and stabilized at high temperatures. This change allowed the authors to define a so-called crossover temperature at which there was no net effect on the stability of CI2 [34].

Here we aim to delineate the effects of different types of soft attractive interactions. In particular, we compare the effect from "nonspecific crowders", which are capable of energetically favorable interactions with any part of the protein, and hydrophobic

crowders, for which these favorable interactions are limited to nonpolar (hydrophobic) amino acids. Hydrophobic crowders are expected to be destabilizing because U but not N will be favored by effective attractive interactions with the crowders, assuming a native structure with a completely buried hydrophobic core. However, it is unclear a priori if hydrophobic crowders should be more or less destabilizing than crowders with nonspecific interactions, because the overall destabilization is determined by the net effect of crowder interactions with U and N (see Fig. 4.1). For comparison, we also consider the results from excluded volume crowders which are stabilizing. In addition to spherical crowders, we also consider the crowding effect from short polypeptide chains, which can favorably interact with the protein through hydrogen bonding via backbone NH and CO groups.

To explore these issues, we use a coarse-grained model for folding which combines an all-atom backbone geometry with a one-bead sidechain representation, an enlarged C_β atom. This model relies on a simplified amino acid alphabet with 3 types: polar (p), hydrophobic (h), and turn (t). Folding is driven by backbone-backbone hydrogen bonding and effective hydrophobic interactions (pairwise hh-attractions). Different sequences can be designed using simple principles [35] leading the chain to adopt various protein-like folds [36, 37]. We focus here on a 35-amino acid sequence that folds into a stable α-helical hairpin fold at low temperatures [38], and is stabilized in the presence of excluded volume crowders [24]. We implement nonspecific interactions by making contacts between crowders and any C_β atom on the protein (p and h amino acids) and limit the interactions to h amino acids in the case of hydrophobic crowders. For a given strength of the attraction, we find that the hydrophobic

crowders are more destabilizing than the crowders interacting nonspecifically with the protein. Moreover, because the hydrophobic crowders have fewer interaction sites on the protein compared to the nonspecific crowders, the overall crowder-protein association is much weaker for the hydrophobic crowders despite their stronger destabilizing effect. Crowders that drive protein destabilization through nonspecific interactions rely on a difference in accessible interaction sites in U and N, and therefore lead to a rather strong crowder-protein association at the point where they can overcome the stabilizing effect of volume exclusion.

4.2 Methods

4.2.1 Coarse-grained model for protein folding

To model protein folding, we use the coarse-grained "C_β-model" described in Ref. [36]. It is a model with three different types of amino acids: polar (p), hydrophobic (h), and turn (t) amino acids. Geometrically, the protein chain is described using an atomistic backbone (C_α, C', N, H, $H_{\alpha 1}$, and O) and simplified sidechains using an enlarged C_β atom. The t amino acids differ from p and h in that it does not contain a C_β atom, which is instead replaced by an $H_{\alpha 2}$ atom. Hence, t is strongly related to glycine and more flexible than p and h. Chain conformations are completely specified by the $2N$ Ramachandran angles $\{\phi_{\rm i}, \psi_{\rm i}\}_{\rm i=1}^{N}$. Hence, bond lengths, bond angles, and dihedral angles (e.g. the peptide plane angle $\omega = 180°$) are held fixed at standard values. The potential energy function of the model, $E_{\rm p}$, includes four

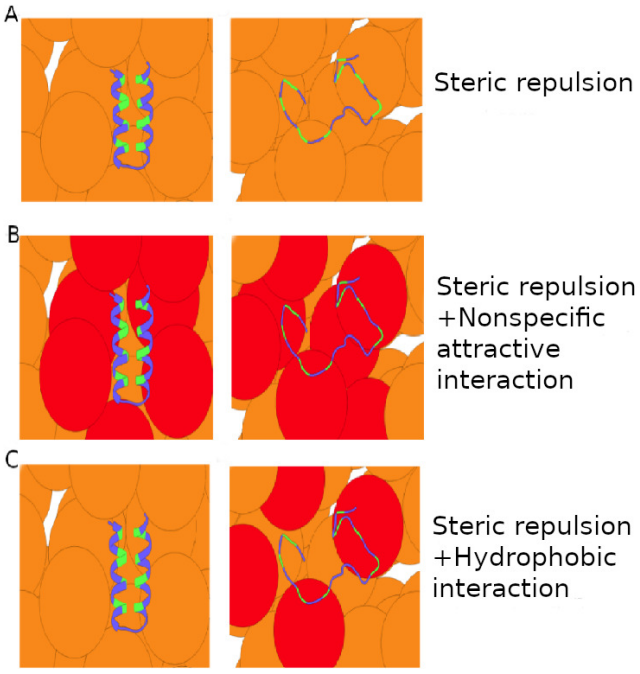

Figure 4.1: Schematic of crowders molecule interaction with the protein. Crowder molecules (orange and red spheres) interacting with a two-state protein (ribbon) through (A) hard-core steric repulsions, (B) nonspecific attractive interactions, and (C) hydrophobic interactions. Crowders (red) make energetically favorable contacts with unfolded and native protein conformations in the case of nonspecific interactions, and with only unfolded conformations in the case of hydrophobic interactions if the native conformation has a core of hydrophobic amino acids (green) inaccessible to other molecules.

terms $E_{\rm p} = E_{\rm loc} + E_{\rm exvol} + E_{\rm hbond} + E_{\rm hp}$, described in detail in Ref. [36]. Briefly, the first term, $E_{\rm exvol}$, provides atom-atom repulsions with a range $\sigma_{\rm a}$ determined by the sum of the atoms' van der Waals radii, $\sigma_{\rm a} = \sigma_{\rm i}^{\rm vdW} + \sigma_{\rm j}^{\rm vdW}$. The second term, $E_{\rm loc}$, represents electrostatic interactions between partial charges of adjacent peptide planes. This term is included to ensure sampled $\phi_{\rm i}$, $\psi_{\rm j}$ angles agree with statistics from real protein structures (Ramachandran plots).

The two last terms, $E_{\rm hbond}$ and $E_{\rm hp}$, represent hydrogen bonding and effective hydrophobic interactions, respectively. These terms are essential for stabilizing and driving the formation of the native state. Hydrogen bonds are modeled using

$$E_{\rm hbond} = k_{\rm hb} \sum_{\rm ij} \gamma_{\rm ij} \left[5 \left(\frac{\sigma_{\rm hb}}{r_{\rm ij}} \right)^{12} - 6 \left(\frac{\sigma_{\rm hb}}{r_{\rm ij}} \right)^{10} \right] \times (\cos\alpha_{\rm ij} \cos\beta_{\rm ij})^{\frac{1}{2}} \,, \tag{4.1}$$

where the sum is over pairs of CO, NH groups, $r_{\rm ij}$ is the OH distance, $\sigma_{\rm hb} = 2.0$ Å, and $k_{\rm hbond} = 3.1$. The interaction strength is modified by a sequence-dependent scale factor $\gamma_{\rm ij}$ taken to be 1 for hh, hp, and pp pairs, and 0.75 for tt, th, and tp pairs. The reduced hydrogen bonding capacity of t amino acids mimics the secondary structure breaking ability of glycine. The directional dependence is implemented via the factor $(\cos\alpha_{\rm ij} \cos\beta_{\rm ij})^{\frac{1}{2}}$, where $\alpha_{\rm ij}$ and $\beta_{\rm ij}$ are the NHO and HOC$'$ angles, respectively. Additionally, if either $\alpha_{\rm ij} < 90°$ or $\beta_{\rm ij} < 90°$, the ij contribution is set to zero. The hydrophobic effect is modeled using pairwise additive effective attractive interactions between hydrophobic amino acids. Specifically,

$$E_{\rm hp} = -k_{\rm hp} \sum_{\rm ij} g(r_{\rm ij}; \sigma_{\rm hp}) \,, \tag{4.2}$$

where the sum is over pairs of C_β atoms of h amino acids, excluding nearest and

next-nearest neighbors along the chain, and

$$g(r; r^0) = \exp\left(-\frac{(r - r^0)^2}{2}\right) \tag{4.3}$$

is a function with a maximum at $r = r^0$. We set $\sigma_{\mathrm{hp}} = 5$ Å and the interaction strength $k_{\mathrm{hp}} = 0.805$. In this model, tuning the relative strength of hydrogen bonding and hydrophobic forces is essential to obtain folding into stable and protein-like structures [39, 40].

4.2.2 Crowders

Repulsive interactions are modeled using the pair potential [41]

$$V(r) = \left(\frac{\sigma}{r - \rho + \sigma}\right)^{12}, \tag{4.4}$$

where r is a crowder-crowder or crowder-atom distance. Because $V(\rho) = \epsilon$ and $V \to \infty$ as $r \to \rho - \sigma$ (for $r < \rho - \sigma$, we set $V = \infty$) the two parameters ρ and σ determine the range and "sharpness" of the repulsion, respectively. We set $\rho = 2R_{\mathrm{c}}$ for crowder-crowder interactions and $\rho = \rho_{\mathrm{cp}} = R_{\mathrm{c}} + \sigma_{\mathrm{a}}$ for crowder-atom interactions, where R_{c} and σ_{a} are the radii of the crowder and atom, respectively. We set $\sigma = 6$ Å for crowder-crowder interactions and 3 Å + σ_{a} for crowder-atom interactions. A crowder-atom attractive interaction is modeled with the potential $V(r) - \epsilon_{\mathrm{att}} g(r; \rho_{\mathrm{cp}})$, where the function g is given by Eq. 4.3. The form of this potential for different attraction strengths ϵ_{att} is shown in Fig. 4.2. For a system consisting of a protein with N_{a} atoms and N_{cr} crowder particles, the total potential energy then becomes

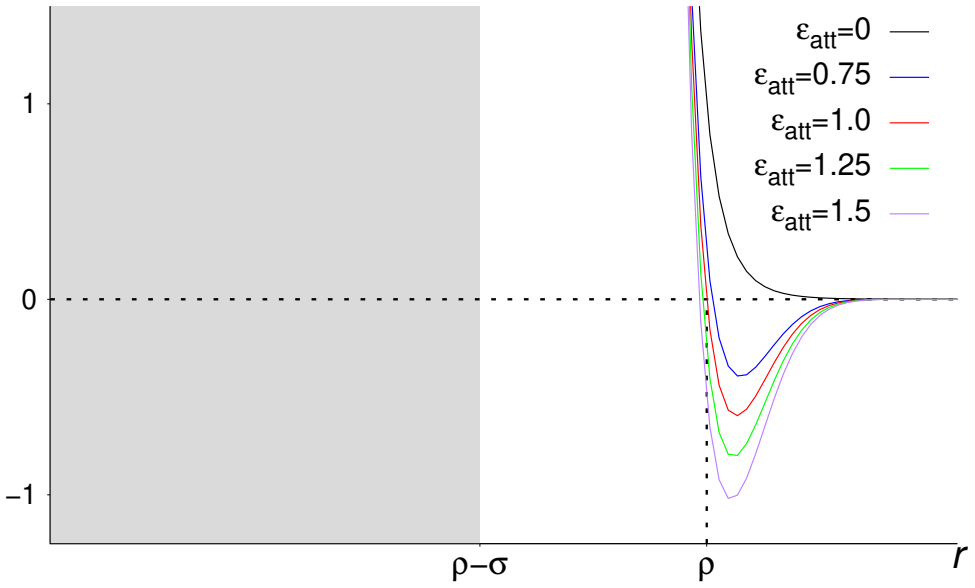

Figure 4.2: Schematics of the pair potential for crowder-crowder and crowder-atom interactions. Shown is $V(r) + g(r)$ for different attraction strengths, ϵ_{att} (see Eqs. 4.4 and 4.3). Note that the $\epsilon_{\text{att}} = 0$ curve is the repulsive potential $V(r)$ given by Eq. 4.4

$E = E_{\rm p} + E_{\rm cc} + E_{\rm cp}$, where

$$E_{\rm cc} = \sum_{\rm i}^{N_{\rm cr}} \sum_{\rm j=i+1}^{N_{\rm cr}} V(r_{\rm ij}) \tag{4.5}$$

and

$$E_{\rm cp} = \sum_{\rm i}^{N_{\rm cr}} \left[\sum_{\rm j}^{N_{\rm a}} V(r_{\rm ij}) - \epsilon_{\rm att} \sum_{\rm j} g(r_{\rm ij}; \rho_{\rm cp}) \right] \tag{4.6}$$

are the crowder-crowder and crowder-protein energies, respectively. In Eq. 4.6, the second sum within square brackets controls which atoms are subject to crowder attraction. For crowders with nonspecific attraction to the protein, the sum is over all C_β atoms (p and h amino acids). For hydrophobic crowder, the sum is over the C_β atoms of h amino acids. For excluded volume crowders, $\epsilon_{\rm att} = 0$.

4.2.3 Simulations and analysis details

Equilibrium behaviors of crowder-peptide systems are determined using simulated tempering Monte Carlo simulations [42]. Runs are carried out with either 8 temperatures in the range $k_{\rm B}T = 0.48 - 0.68$ or 10 temperatures in the range 0.40-0.70. For each system, 10 independent runs with 5×10^9 elementary MC updates are performed. Initial configurations are obtained by picking random chain conformations and random spherical crowder positions, followed by a relaxation step to remove any hard-core steric clashes. Monte Carlo chain updates are divided equally between the protein chain and the crowder particles. Two different types of chain moves are used: Biased Gaussian Steps (BGS) [43], which produce approximately local chain deformations, and pivot moves, which produce global changes. In the latter type, a single $\phi_{\rm i}$ or $\psi_{\rm i}$ angle is set to a new random value such that the chain rotates around the

NC_α or $C_\alpha C'$ bond. The frequency of chain updates are set so that BGS is most frequent at low Ts and pivot is most frequent at high Ts. Temperature updates are attempted every 100 MC step. Spherical crowder simulations are carried out in a cubic box of side length $L = 100$ Å while for simulations of polypeptide crowders $L = 65$ Å. The number of spherical crowding agents are 14, 28, 42, and 56, corresponding to packing fractions $\phi_c = 0.10$, 0.20, 0.30, and 0,40, respectively. The crowder radius is $R_c = 12$ Å. For simulation carried out with 8 temperatures in the range $k_B T = 0.48 - 0.68$, the multistate Bennett acceptance ratio (MBAR) technique was applied to determine thermodynamic averages in the range $k_B T = 0.40$–0.70.

4.2.4 Observables

We quantify native state stability using two quantities, ΔF and T_m. The free energy of folding is determined using

$$\Delta F = F_N - F_U = -k_B T \ln \frac{P_{nat}}{1 - P_{nat}}, \tag{4.7}$$

where F_U and F_N are the free energies of the unfolded and native states, respectively, and P_{nat} is the population of the native state. The native state population P_{nat} is determined as in Ref. [24]. Conformations are considered part of the native state if $Q_{nat} \geq Q_{cut}$ where Q_{nat} is the number of formed native contacts and $Q_{cut} = 50$. The folding midpoint temperature, T_m, is determined by fitting the temperature dependence of P_{nat} to a two-state model with two fit parameters. The fit for α_{35} in the absence of crowders is given in Supporting Information and gives $k_B T_m^0 = 0.535$.

4.3 Results

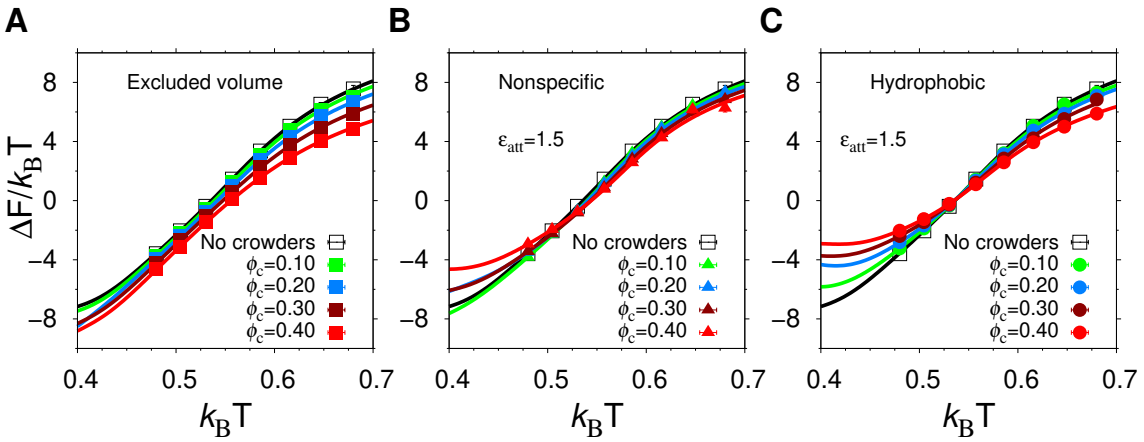

Figure 4.3: The free energy of folding for α_{35}. Temperature dependence of the free energy of folding of α_{35} in the absence and presence of (A) excluded volume crowders, (B) nonspecific crowders, and (C) hydrophobic crowders, at different volume fractions, ϕ_c, and (B, C) for one strength of the crowder-protein attractions, $\epsilon_{att} = 1.5$.

4.3.1 Spherical crowders

Using simulated tempering MC simulations, we determined the thermodynamics behavior of our model protein, α_{35}, both as an isolated chain with no crowders and in the presence of the three different types of spherical crowders in Fig 4.3. All crowders have the same radius, $R_c = 12$ Å. We varied the crowder volume fraction, ϕ_c, and the strength of the attractive interactions between crowders and protein, ϵ_{att}. In general,

we find that the equilibrium stability of the α_{35} native state, as quantified by the free energy of folding, $\Delta F = -k_{\rm B}T \ln[P_{\rm nat}/(1-P_{\rm nat})]$, where $P_{\rm nat}$ is the population of the native state, depends on both $\phi_{\rm c}$ and $\epsilon_{\rm att}$. As an illustration of our results, Fig 4.3 shows the temperature dependence of ΔF at $\epsilon_{\rm att} = 1.5$, for different $\phi_{\rm c}$ and crowder types.

In the absence of crowders (solid black curve in Fig 4.3), the helical hairpin fold of α_{35} is highly stable at low T. For example, at the lowest studied temperature T, which corresponds to $T \approx 0.75T_{\rm m}^0$, where $T_{\rm m}^0$ is the midpoint temperature of the α_{35} folding transition for $\phi_{\rm c} = 0$, $\Delta F \approx -8k_{\rm B}T$. This is within the range of typical stabilities of real single-domain proteins [44]. The midpoint temperature $T_{\rm m}^0$, i.e., the temperature for which $\Delta F = 0$, is also commonly used as a measure of protein stability. For α_{35}, $k_{\rm B}T_{\rm m}^0 = 0.535$, as obtained previously [24] by fitting the of α_{35} folding curve i.e., $P_{\rm nat}$ as function of T, to a two-state model with two free parameters (see Fig B1). For the values of $\epsilon_{\rm att}$ considered here, α_{35} remains stable in the low T region even when the crowder packing fraction reaches $\phi_{\rm c} = 0.40$. However, clear variations in results between crowders can be seen (see Fig 4.3).

At a temperature just below the folding midpoint, $T^- = 0.95T_{\rm m}^0$, we observe the following trends. Upon the addition of excluded volume crowders, ΔF decreases monotonically with $\phi_{\rm c}$, as shown in Fig 4.3A, meaning a strict stabilization of the protein. The soft interactions of the nonspecific and hydrophobic crowders are expected to provide a destabilizing effect. However, because these two crowder types occupy the same amount of space as the excluded volume crowders, they also provide a stabilizing effect. The net crowding effect will be determined by a competition be-

tween stabilizing volume exclusion and destabilizing soft interactions. Indeed, at low attraction strength ($\epsilon_{\rm att} = 1.0$), the addition of nonspecific or hydrophobic crowders still provide an overall stabilization (see Fig 4.4A). At $\epsilon_{\rm att} = 1.5$, the hydrophobic crowders overcome the stabilizing excluded volume effect, leading to a net destabilization, while the nonspecific crowders are net stabilizing. A similar view is obtained by considering the changes in the midpoint temperature. As shown in Fig 4.4B, there is a relative monotonic increase in $T_{\rm m}$ for both the excluded volume crowder and for the hydrophobic and nonspecific crowder with $\epsilon_{\rm att} = 1.0$. For $\epsilon_{\rm att} = 1.5$, the nonspecific and hydrophobic crowders exhibit monotonically increasing and decreasing midpoint temperatures with increases $\phi_{\rm c}$, respectively. Overall, our results show that excluded volume crowders provide a stabilizing effect on α_{35}, which can be counteracted by the presence of soft attractive interactions. The hydrophobic crowders provide a stronger destabilizing effect than the nonspecific crowders.

4.3.2 How can the stronger destabilizing effect of weak hydrophobic attraction be explained?

To address the question of why hydrophobic attractive interactions are more destabilizing than nonspecific interactions, we consider the interaction energy between crowders and protein, $E_{\rm cp}$ (see Methods). $E_{\rm cp}$ is a mix of repulsive and attractive energy contributions. We consider, in particular, $E_{\rm cp}^{\rm U}$ and $E_{\rm cp}^{\rm N}$, i.e., the crowder-protein energy determined separately for the U and N states. When the difference $\Delta E_{\rm cp} = E_{\rm cp}^{\rm U} - E_{\rm cp}^{\rm N} < 0$ there is a net energetic stabilization of U, which should have

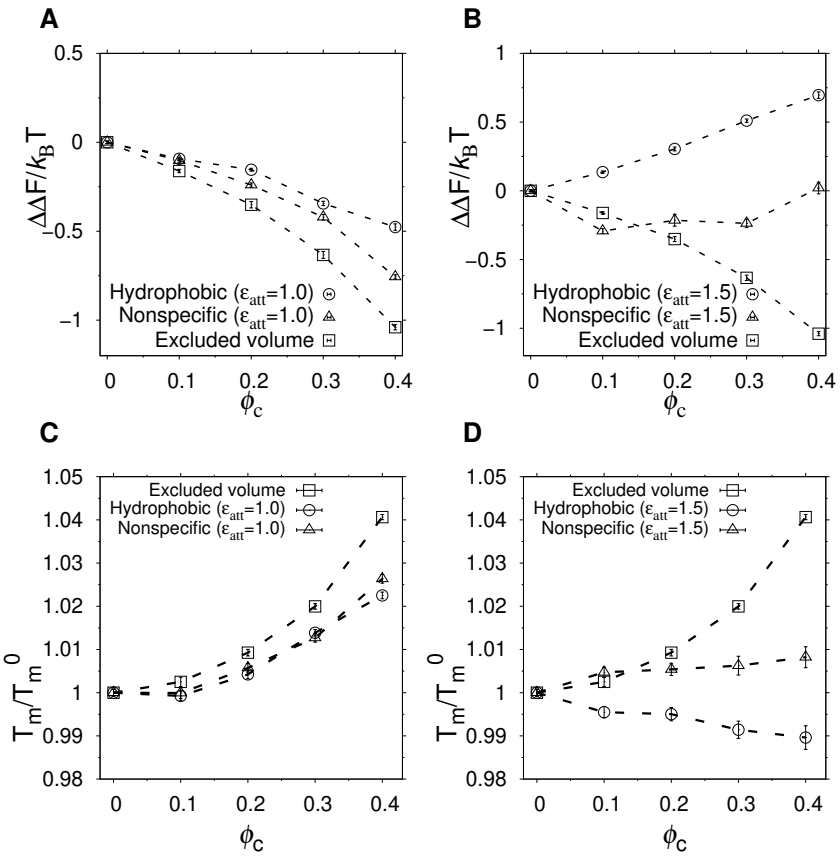

Figure 4.4: Effect of crowding on the native state stability. The change in free energy of folding, $\Delta\Delta F(\phi_c) = \Delta F(\phi_c) - \Delta F_0$, where ΔF_0 is the folding free energy in the absence of crowders, as a function of ϕ_c, for excluded volume (squares), nonspecific (triangles), and hydrophobic (circles) crowders with attraction strengths (A) $\epsilon_{att} = 1.0$ and (B) 1.5, obtained at the temperature $T^- = 0.95T_m$ ($k_B T^- = 0.508$). Relative change in midpoint folding temperature, T_m/T_m^0, where T_m^0 is the midpoint temperature at $\phi_c = 0$, as a function of ϕ_c, for (C) $\epsilon_{att} = 1.0$ and (D) 1.5. Dashed lines between points are drawn to guide the eye.

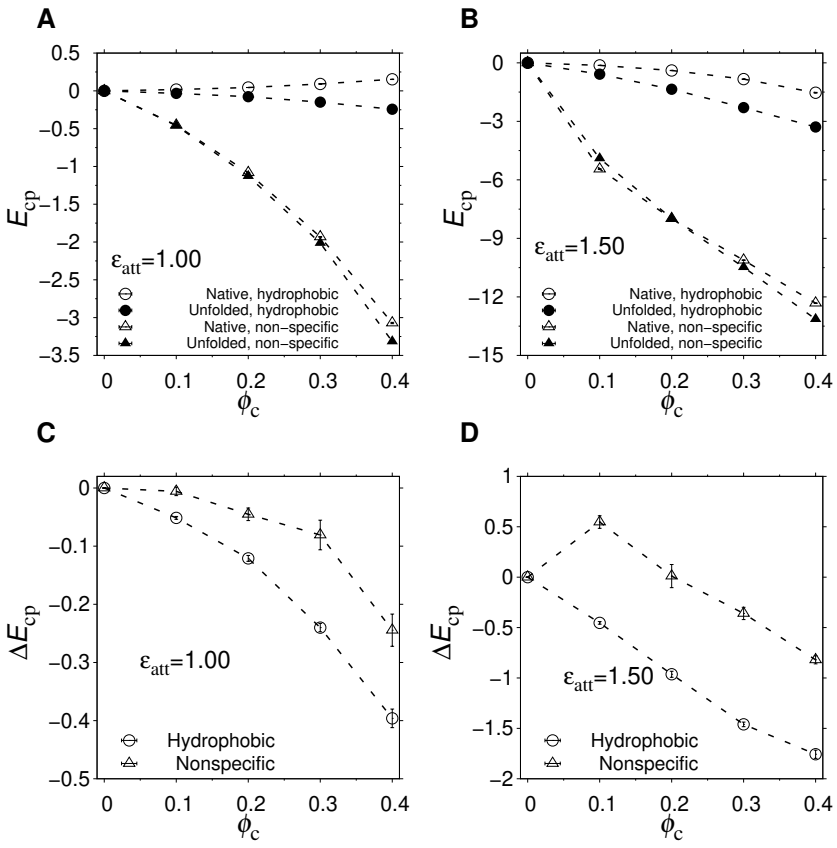

Figure 4.5: Energetic stabilization of the unfolded state. The crowder-protein attractive interaction $E_{\rm cp}$ with the native (open) and the unfolded (solid) states, as a function of $\phi_{\rm c}$, for non-specific (triangles) and hydrophobic crowders (circles) with (A) $\epsilon_{\rm att} = 1.0$ and (B) $\epsilon_{\rm att} = 1.5$. Change in the energetic stabilization of the unfolded state $\Delta E_{\rm cp}$ as a function $\phi_{\rm c}$ for hydrophobic (circle) and non-specific (triangle) interactions with (C) $\epsilon_{\rm att} = 1.0$ and (D) $\epsilon_{\rm att} = 1.5$ at low temperature. Dashed lines between points are drawn to guide the eye.

a destabilizing effect on the protein.

Fig 4.5A,B shows the ϕ_c-dependence of E_{cp}^{U} and E_{cp}^{N} for two different values of ϵ_{att}. For hydrophobic crowders, E_{cp}^{N} (open circles) exhibits a slight positive slope for $\epsilon_{att} = 1.0$ and a slight negative slope for $\epsilon_{att} = 1.5$, but is overall rather flat. The slight positive slope for $\epsilon_{att} = 1.0$ means that repulsive interactions dominate the interaction when the protein populates the N state. The negative slope for $\epsilon_{att} = 1.5$, indicates that, at this strength of attractive interactions, the native state becomes slightly distorted to accommodate favorable contacts between hydrophobic sidechains and crowder particles. However, the interactions are rather limited. At the highest packing fractions, $\phi_c = 0.40$, $E_{cp}^{N} \approx -1.5$, thus corresponding one fully formed interaction between an h amino acid and a crowder. For the nonspecific case, interactions between crowders and protein are much more prevalent, for both U and N. E_{cp}^{U} is a sharply decreasing function of ϕ_c for both $\epsilon_{att} = 1.0$ and 1.5, which will help to stabilize the U state at high ϕ_c. However, the decrease in E_{cp}^{U} is nearly compensated for by the decrease in E_{cp}^{N}. Comparing the two types of crowders, we find that the net energetic effect, ΔE_{cp}, is more negative for the hydrophobic crowder as compared to the nonspecific crowders, as shown in Fig 4.5C and D.

An interesting difference between the two crowder types is that, while their net effect on stability is similar, they differ substantially in the degree of association with the protein. While hydrophobic crowders associate weakly with the protein, even in U, the protein-crowder association in the case of nonspecific interactions is greater by approximately an order of magnitude, as quantified by the magnitude of the interaction energy E_{cp} (see Fig 4.5D).

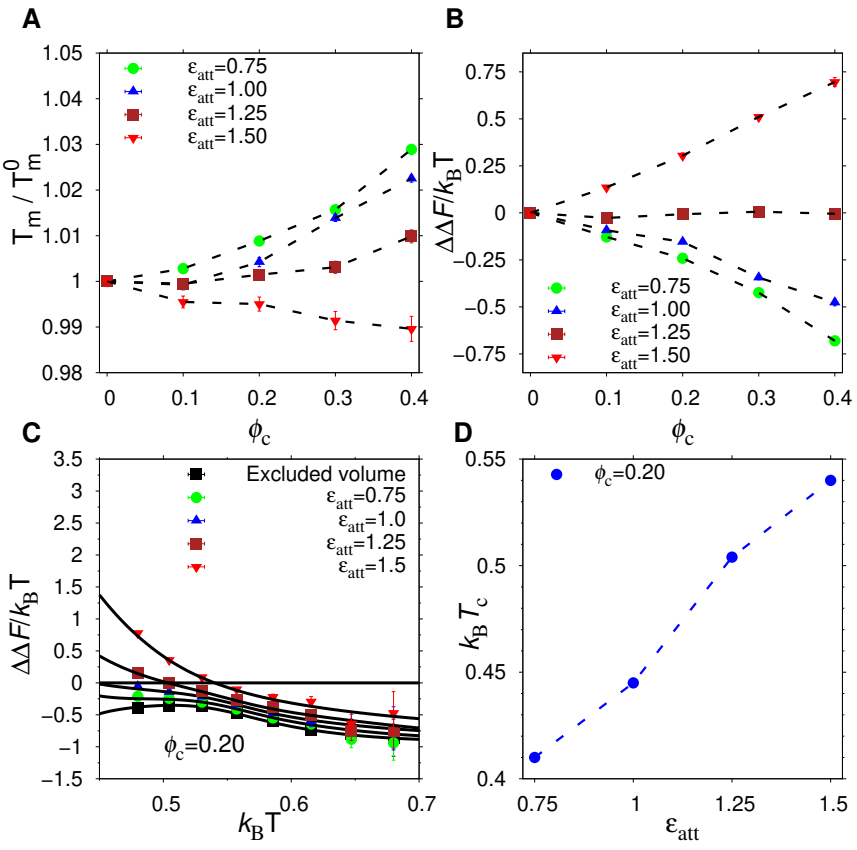

Figure 4.6: Effect of hydrophobic crowders on the native state stability. (A) Change in the folding midpoint temperature, and (B) change in the folding free energy at T^-, as functions of ϕ_c for different attraction strengths ϵ_{att}. (C) $\Delta\Delta F$ as a function of temperature for difference values of ϵ_{att}. A crossover temperature, T_{cross} is defined by $\Delta\Delta F = 0$. (D) T_{cross} as a function of attraction strengths for fixed volume fraction $\phi_c = 0.20$. Dashed lines between points are drawn to guide the eye.

4.3.3 Balancing destabilizing soft interactions and stabilizing steric repulsions

The α_{35} protein is net stabilized by the presence of hydrophobic crowders at $\epsilon_{\mathrm{att}} = 1.0$, but net destabilized at $\epsilon_{\mathrm{att}} = 1.5$. This suggests a critical attraction strength at which there is no net change in stability. Indeed, for $\epsilon_{\mathrm{att}} = 1.25$, the change in folding free energy, $\Delta\Delta F$, is roughly zero, as shown in Fig 4.6B. Interestingly, this balance between stabilizing and destabilizing effects is largely independent of ϕ_{c}. The midpoint temperature T_{m} also remains roughly constant, except at the very high packing fractions ($\phi_{\mathrm{c}} = 0.40$), where there is a slight increase. Crowding induced stabilization at a given ϵ_{att} is, however, dependent on temperature. Fig 4.6C shows the T-dependence of $\Delta\Delta F$. For excluded volume crowders, $\Delta\Delta F < 0$, reflecting the strictly stabilizing effect. There is a T dependence even for this crowder type, which originates from the fact that the size of U is not constant [24]. For hydrophobic crowders, due to the energetic stabilization of U relative to N, at low enough temperatures, the net effect crosses over from stabilizing to destabilization, i.e., $\Delta\Delta F$ changes from negative to positive. This crossover temperature, T_{c}, [45] increases with the contact strength ϵ_{att}. The findings for α_{35} are similar to previous studies on other proteins, such as ubiquitin, which showed a crossover temperature in the presence of either synthetic polymers or protein crowders [19, 45].

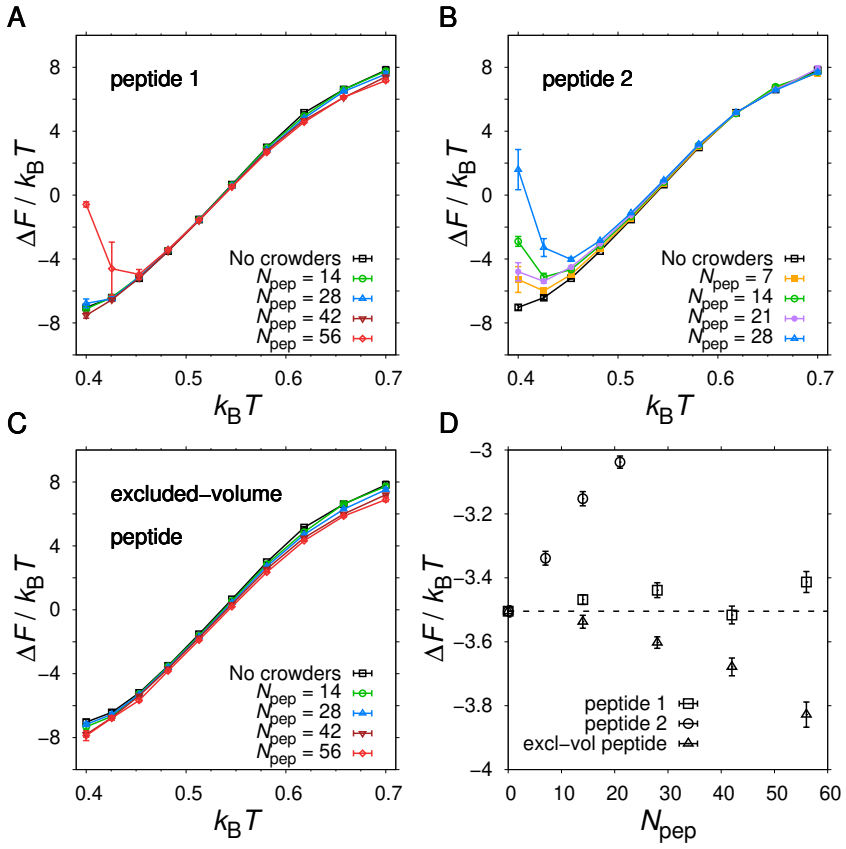

Figure 4.7: Native state stability of α_{35} in the presence of different types of polypeptide crowders. Folding free energy, ΔF, as function of temperature T for three peptide crowder types: (A) peptide 1 with sequence ppppp, (B) peptide 2 with sequence pphpp, and (C) a peptide geometrically identical to peptides 1 and 2 but with inter-chain hydrophobic and hydrogen bond interactions turned off. (D) ΔF as function of the number of peptides, $N_{\rm pep}$, at $T \approx T^-$ ($k_{\rm B}T = 0.482$).

4.3.4 Polypeptide crowders

We turn now to the folding of α_{35} in the presence of short polypeptide chains. We use the same model for the peptides as for α_{35} (see section 4.2.1), with the restriction that energetically favorable interactions (i.e., effective hydrophobic and hydrogen bond interactions) between crowding peptides are turned off. This avoids the formation of oligomeric peptide structures, which would complicate the analysis. We consider two different 5-amino acid sequences: ppppp (peptide 1) and pphpp (peptide 2). Both peptides can thus interact with the α_{35} chain through backbone-backbone hydrogen bonding and peptide 2, due to its central h amino acid, can additionally interact with α_{35} through hydrophobic attractions. Fig 4.7(A) and (B) show the T-dependence of the free energy of folding, ΔF, for different numbers $N_{\rm pep}$ of peptide 1 or peptide 2 chains added to the system. For $N_{\rm pep} \leq 42$, peptide 1 induces a weak stabilization at high T and weak destabilization at low T. This behavior is consistent with a competition between entropy-driven stabilization due to volume exclusion by the peptides and energy-driven destabilization due to inter-chain hydrogen bonding. Interestingly, there is a rather broad temperature range around $T_{\rm m}^0$ ($k_{\rm B}T \approx 0.45 - 0.60$) with no detectable change in ΔF (see Fig 4.7(A) and (D)). The apparent lack of crowding effects in this temperature range can arise either because (i) the crowding effects are overall weak at the studied peptide concentrations or (ii) the two opposing crowding effects, peptide excluded volume and peptide-protein attractions, are equal in magnitude and therefore cancel each other out. To determine which scenario holds, we performed additional simulations with peptides that only allowed to interact with other chains via repulsive interactions. These "excluded volume peptides"

significantly stabilize α_{35} across all Ts (see Fig. 4.7(C) and (D)), which means that scenario (ii) above holds. Hence, in a relatively broad range around the midpoint temperature of the protein, the excluded volume effect due to the peptides is almost perfectly counteracted by soft interactions in the form of hydrogen bonds. Peptide 2

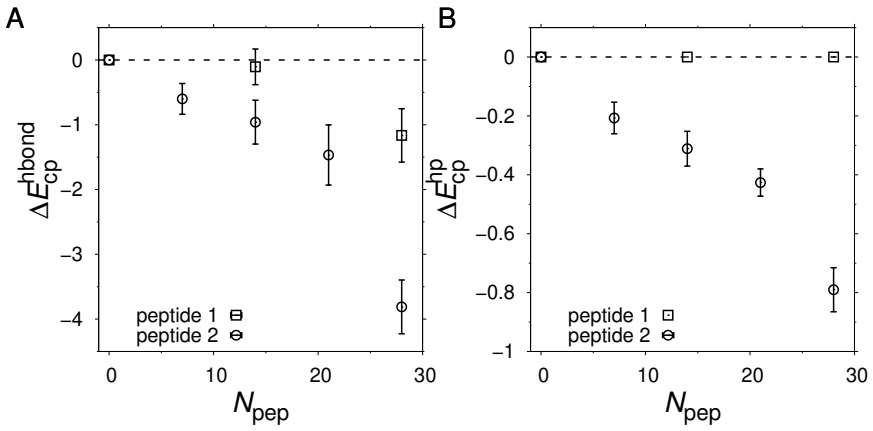

Figure 4.8: Energetically favorable interactions between crowder peptides and the protein. Hydrophobic and hydrogen-bond peptide-protein interaction energies are measured for the native ($E_{cp}^{N,hp}$ and $E_{cp}^{N,hbond}$) and unfolded ($E_{cp}^{U,hp}$ and $E_{cp}^{U,hbond}$) states. Shown are (A) $\Delta E_{cp}^{hbond} = E_{cp}^{U,hbond} - E_{cp}^{N,hbond}$ and (B) $\Delta E_{cp}^{hp} = E_{cp}^{U,hp} - E_{cp}^{N,hp}$ as function of N_{pep}. Negative ΔE_{cp} values means that the peptides interact more favorably with U relative to N. The temperature is the same as in Fig 4.7(D).

provides a stronger destabilizing effect on α_{35} than peptide 1 at a given concentration, as seen in Fig 4.7(B) and (D). For example, at $T \approx T^-$, ΔF monotonically increases with the number of added peptide 2 chains in contrast to the flat ΔF exhibited by peptide 1. At very low T and $N_{pep} = 28$, peptide 2 even leads α_{35} to become net unstable ($\Delta F > 0$), although the error bars are larger at lowest studied Ts. A similar behavior is seen for peptide 1 and $N_{pep} = 56$. It is instructive to get a rough idea

of the volume fraction ϕ_c occupied by the peptides in our systems. Assuming each atom in our C_β-model occupies volume according to its van der Waals radius, we obtain $\phi_c \approx 0.07$ for $N_{pep} = 28$. Alternatively, if we assume amino acids are spheres with radius 3.8 Å (typical C_α-C_α distances of peptide bonds) we obtain a slightly higher value, $\phi_c \approx 0.12$. What makes peptide 2 more destabilizing than peptide 1? Because the two peptides are geometrically identical, any difference must derive from soft interactions. Fig 4.8(A) and (B) show the peptide concentration dependence of the quantity $\Delta E_{cp} = E_{cp}^{U} - E_{cp}^{N}$ (Fig 4.5), determined separately for peptide (crowder)-protein hydrogen bond (ΔE_{cp}^{hbond}) and hydrophobic (ΔE_{cp}^{hp}) energies. For peptide 1, $\Delta E_{cp}^{hp} = 0$, since this peptide lacks a hydrophobic amino acid. For peptide 2, $\Delta E_{cp}^{hp} < 0$ for all peptide-protein systems. This means that, vis-à-vis hydrophobic interactions, peptide 2 interacts more favorably with U than with N, causing α_{35} destabilization. The formation of hydrogen bonds between peptides and the protein are also destabilizing because generally $\Delta E_{cp}^{hbond} < 0$. Interestingly, ΔE_{cp}^{hbond} is more negative for peptide 2 than for peptide 1, such that the presence of a hydrophobic amino acid on peptide 2 also enhances the formation of hydrogen bonds with the protein chain. In other words, there is a cooperative effect between hydrophobic and hydrogen bond interactions in peptide 2, which further enhances the destabilizing effect of this peptide relative to peptide 1 which has only polar amino acids.

4.4 Discussions and Conclusions

We have used a sequence-based coarse-grained protein model to study crowding-induced changes to the stability of a model protein with a sequence that folds to α-helical hairpin at low temperatures. In particular, we investigated crowders making different types of soft interactions with the protein, i.e., interactions different from hard-core steric repulsions. We found that crowders with a hydrophobic character, i.e., crowders interacting favorably with only nonpolar amino acids, provide a stronger destabilizing effect compared to crowders that interact nonspecifically, i.e., favorably with both polar and nonpolar amino acids, with the protein. Both types of soft interactions are counteracting the generally stabilizing effect of excluded volume. Therefore, for weak attraction strengths, these crowders still increase the stability of the protein relative to the dilute limit where there are no crowding effects. At a critical strength of the attraction, the stabilizing and destabilizing effects cancel leading to a net zero change in protein stability over a wide range of crowder concentrations. Similar results were obtained in a structure-based one-bead-per-amino acid model studied by Mittal *et al.* [46].

The destabilization effect of either hydrophobic or nonspecific crowder interactions arise because interactions with U is energetically more favorable than the interactions with N. This is manifest by the net energetic effect of the crowder-protein interactions, $\Delta E_{\rm cp}$, which is negative at high $\phi_{\rm c}$ for both crowder types. However, perhaps counterintuitively, the overall interactions of the hydrophobic crowders is much less pronounced than for the nonspecific crowders at a given degree of destabilization.

Quantitatively, we find that the magnitude of the interaction energy between crowders and protein, E_{cp}, is larger for nonspecific crowders than for the hydrophobic crowders by roughly an order of magnitude. The reason for this difference is the way the two crowder types achieve destabilization. Nonspecific crowders rely on a difference between two rather large favorable interaction energies for U and N. Hydrophobic crowders, for which favorable interactions with N are almost absent due to hydrophobic amino acids being buried in the native structure, obtain destabilization even for relatively weak favorable interactions with U. This difference between hydrophobic and nonspecific soft interactions might be tested experimentally using protein crowders, and varying the chemical nature of surface-exposed amino acids through mutations.

Our results also confirm that the addition of crowders attracted to a protein can lead to temperature-dependent crowding effects, as demonstrated experimentally [19]. Specifically, we found that the our model protein exhibits a crossover temperature, T_{cross}, below which the crowders are destabilizing and above which the crowders are stabilizing. The presence of such a crossover temperature is well established in the literature and has been observed in the presence of both synthetic polymer crowders and protein crowders [19, 34, 45]. The origin of this crossover temperature is believed to be due to the fact that energetic effects are becoming more important at low T, while at high T entropic effects dominate [34, 45]. It should be noted that the hydrophobic effect is itself partly entropically driven [47], while in our model it is treated as an effective energetically driven interaction. Destabilization driven entirely by soft hydrophobic interactions are not guaranteed to increase with decreasing T.

We have also investigated how the folding and stability of our model protein are impacted by short polypeptide chains, which differ from our spherical crowder in that they are able to interact with the protein through hydrogen bonding. Results for this system indicate an interplay between hydrogen bonds and hydrophobic interactions. Peptide chains with partial hydrophobic character exhibit much stronger destabilizing effects than peptides with only polar amino acids. Part of this destabilization is due to hydrophobic interactions directly stabilizing U, in the same way as found for the hydrophobic spherical crowders. However, we found that the presence of a hydrophobic amino acid in the sequence of our peptide crowder also tends to promote the formation of additional hydrogen bonds with the protein (see Fig 4.8), thus further strengthening the destabilization. It should be pointed out, however, that our polypeptide crowders are relatively short. It would be interesting to test the crowding effect of much longer popypeptide chains in our model, which could mimic macromolecular crowders capable of both substantial excluded volume effects and soft interactions in the form of hydrophobic interactions and hydrogen bonding.

In summary, our sequenced-based model provides insights into the effect of attractive interactions between crowders and protein on the stability of the protein's native state. Our results highlight the importance of considering both the type and strength of soft crowder-protein interactions when evaluating the impact of crowding on protein stability.

Bibliography

[1] F. X. Theillet, A. Binolfi, T. Frembgen-Kesner, K. Hingorani, M. Sarkar, C. Kyne, C. Li, P. B. Crowley, L. Gierasch, G. J. Pielak, A. H. Elcock, A. Gershenson, and P. Selenko. Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs). *Chem Rev*, 114:6661–6714, 2014.

[2] A. P. Minton. The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences. *Mol Cell Biochem*, 55:119–140, 1983.

[3] S. B. Zimmerman and S. O. Trach. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J Mol Biol*, 222:599–620, 1991.

[4] R. J. Ellis. Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci*, 26:597–604, 2001.

[5] H. Dong, S. Qin, and H. Zhou. Effects of macromolecular crowding on protein conformational changes. *PLoS Comput Biol*, 6:1000833, 2010.

[6] B. van den Berg, R. J. Ellis, and C. M. Dobson. Effects of macromolecular crowding on protein folding and aggregation. *EMBO J*, 18:6927–6933, 1999.

[7] A P Minton. Implications of macromolecular crowding for protein assembly. *Curr Opin Struct Biol*, 10:34–39, 2000.

[8] L. A. Munishkina, E. M. Cooper, V. N. Uversky, and A. L. Fink. The effect of macromolecular crowding on protein aggregation and amyloid fibril formation. *J Mol Recognit*, 17:456–464, 2004.

[9] J. S. Schreck, J. Bridstrup, and J. Yuan. Investigating the effects of molecular crowding on the kinetics of protein aggregation. *J Phys Chem B*, 124:9829–9839, 2020.

[10] G. A. Siddiqui and A. Naeem. Aggregation of globular protein as a consequences of macromolecular crowding: A time and concentration dependent study. *Int J Biol Macromol*, 108:360–366, 2018.

[11] B. Akabayov, S. R. Akabayov, S. J. Lee, G. Wagner, and C. C. Richardson. Impact of macromolecular crowding on dna replication. *Nat Commun*, 4:1–10, 2013.

[12] A. A. André and E. Spruijt. Liquid–liquid phase separation in crowded environments. *Int J Mol Sci*, 21:5908, 2020.

[13] A. AM. André, N. A. Yewdall, and E. Spruijt. Crowding-induced phase separation and solidification by co-condensation of peg in npm1-rrna condensates. *Biophys J*, 122:397–407, 2023.

[14] K. Julius, J. Weine, M. Gao, J. Latarius, M. Elbers, M. Paulus, M. Tolan, and R. Winter. Impact of macromolecular crowding and compression on protein–protein interactions and liquid–liquid phase separation phenomena. *Macromolecules*, 52:1772–1784, 2019.

[15] A P Minton. Excluded volume as a determinant of protein structure and stability. *Biophys J*, 32:77–79, 1980.

[16] K. Sasahara, P. McPhie, and A. P. Minton. Effect of dextran on protein stability and conformation attributed to macromolecular crowding. *J Mol Biol*, 326:1227–1237, 2003.

[17] D. S. Spencer, K. Xu, T. M. Logan, and H. X. Zhou. Effects of pH, salt, and macromolecular crowding on the stability of FK506-binding protein: an integrated experimental and theoretical study. *J Mol Biol*, 351:219–232, 2005.

[18] L. A. Benton, A. E. Smith, G. B. Young, and G. J. Pielak. Unexpected effects of macromolecular crowding on protein stability. *Biochemistry*, 51:9773–9775, 2012.

[19] Y. Wang, M. Sarkar, A. E. Smith, A. S. Krois, and G. J. Pielak. Macromolecular crowding and protein stability. *J Am Chem Soc*, 134:16614–16618, 2012.

[20] J. Hong and L. M. Gierasch. Macromolecular crowding remodels the energy landscape of a protein by favoring a more compact unfolded state. *J Am Chem Soc*, 132:10445–10452, 2010.

[21] A. Christiansen and P. Wittung-Stafshede. Synthetic crowding agent dextran

causes excluded volume interactions exclusively to tracer protein apoazurin. *FEBS Lett*, 588:811–814, 2014.

[22] F. C. Zegarra, D. Homouz, A. G. Gasic, L. Babel, M. Kovermann, P. Wittung-Stafshede, and M. S. Cheung. Crowding-induced elongated conformation of urea-unfolded apoazurin: investigating the role of crowder shape in silico. *J Phys Chem B*, 123:3607–3617, 2019.

[23] A. Christiansen, Q. Wang, M. S. Cheung, and P. Wittung-Stafshede. Effects of macromolecular crowding agents on protein folding in vitro and in silico. *Biophys Rev*, 5:137–145, 2013.

[24] S. Bazmi and S. Wallin. Crowding-induced protein destabilization in the absence of soft attractions. *Biophys J*, 121:2503–2513, 2022.

[25] M. Sarkar, A. E. Smith, and G. J. Pielak. Impact of reconstituted cytosol on protein stability. *Proc Natl Acad Sci USA*, 110:19342–19347, 2013.

[26] M. Sarkar, C. Li, and G. J. Pielak. Soft interactions and crowding. *Biophys Rev*, 5:187–194, 2013.

[27] M. Sarkar, J. Lu, and G. J. Pielak. Protein crowder charge and protein stability. *Biochemistry*, 53:1601–1606, 2014.

[28] J. Rosen, Y. C. Kim, and J. Mittal. Modest protein- crowder attractive interactions can counteract enhancement of protein association by intermolecular excluded volume interactions. *J Phys Chem B*, 115:2683–2689, 2011.

[29] Y. C. Kim and J. Mittal. Crowding induced entropy-enthalpy compensation in protein association equilibria. *Phys Rev Lett*, 110:208102, 2013.

[30] Y. Phillip, V. Kiss, and G. Schreiber. Protein-binding dynamics imaged in a living cell. *Proc Natl Acad Sci USA*, 109:1461–1466, 2012.

[31] M. Jiao, H. Li, J. Chen, A. P. Minton, and Y. Liang. Attractive protein-polymer interactions markedly alter the effect of macromolecular crowding on protein association equilibria. *Biophys J*, 99:914–923, 2010.

[32] J. F. Douglas, J. Dudowicz, and K. F. Freed. Crowding induced self-assembly and enthalpy-entropy compensation. *Phys Rev Lett*, 103:135701, 2009.

[33] H. X. Zhou. Protein folding and binding in confined spaces and in crowded solutions. *J Mol Recognit*, 17:368–375, 2004.

[34] S. Timr and F. Sterpone. Stabilizing or destabilizing: simulations of chymotrypsin inhibitor 2 under crowding reveal existence of a crossover temperature. *J Phys Chem Lett*, 12:1741–1746, 2021.

[35] M. H. Cordes, A. R. Davidson, and R. T. Sauer. Sequence space, folding and protein design. *Curr Opin Struct Biol*, 6:3–10, 1996.

[36] A. Bhattacherjee and S. Wallin. Coupled folding-binding in a hydrophobic/polar protein model: impact of synergistic folding and disordered flanks. *Biophys J*, 102:569–578, 2012.

[37] D. Trotter and S. Wallin. Effects of topology and sequence in protein folding linked via conformational fluctuations. *Biophys J*, 118:1370–1380, 2020.

[38] C. Holzgräfe and S. Wallin. Local versus global fold switching in protein evolution: insight from a three-letter continuous model. *Phys Biol*, 12:026002, 2015.

[39] A. Irbäck, F. Sjunnesson, and S. Wallin. Three-helix-bundle protein in a Ramachandran model. *Proc Natl Acad Sci USA*, 97:13614–13618, 2000.

[40] A. Irbäck, F. Sjunnesson, and S. Wallin. Hydrogen bonds, hydrophobicity forces and the character of the collapse transition. *J Biol Phys*, 27:169–179, 2001.

[41] J. Mittal and R. B. Best. Dependence of protein folding stability and dynamics on the density and composition of macromolecular crowders. *Biophys J*, 98:315–320, 2010.

[42] E. Marinari and G. Parisi. Simulated Tempering: a new Monte Carlo scheme. *Europhys Lett*, 19(6), 1992.

[43] G. Favrin, A. Irbäck, and F. Sjunnesson. Monte Carlo update for chain molecules: biased Gaussian steps in torsional space. *J Chem Phys*, 114:8154–8158, 2001.

[44] J. A. Rupley, E. Gratton, and G. Careri. Water and globular proteins. *Trends Biochem Sci*, 8:18–22, 1983.

[45] H-X Zhou. Polymer crowders and protein crowders act similarly on protein folding stability. *FEBS Lett*, 587:394–397, 2013.

[46] Young C Kim, Apratim Bhattacharya, and Jeetain Mittal. Macromolecular crowding effects on coupled folding and binding. *J Phys Chem B*, 118:12621–12629, 2014.

[47] D. Chandler. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437:640–647, 2005.

Chapter 5

Crowding effects on protein folding thermodynamics and kinetics: polymeric versus spherical crowders

Abstract

Computational studies of macromolecular crowding effects on protein folding and stability often represent crowding agents as spheres, while experiments often use artificial polymers, such as polyethylene glycol and Ficoll. Here, we study and compare the effects of two different types of crowding agents, spherical and polymeric crowders, on a protein with an all-α structure (PDB id: 1PRB) and a protein with an all-β structure (PDB id: 1SHF), in a range of crowder volume fractions and temperatures. To this end, we use a structure-based (Gō-like) model and Langevin dynamics sampling, which allow us to study effects on both the folding thermodynamics and kinetics of the two proteins. We show first that both proteins are stabilized by adding either type of crowding agent, but the stabilization is slightly more pronounced for β-structure protein than for α-structure protein. Additionally, we analyze the size dependency of the crowding effects by varying the radius of the spheres and the length of the polymers. For spherical crowders, we find that, for a fixed crowder volume fraction, stabilization increases with decreasing size of the spheres. By contrast, for polymer crowders, we find only a relatively weak effect on chain length. We also examined the effect of crowding on the folding kinetics for both proteins and we find that the folding rate is enhanced by adding either type of crowder.

5.1 Introduction

The cellular environment is crowded, heterogeneous, and complex [1]. It has been known for a long time that this environment plays essential roles in numerous cell biological and biochemical processes, including protein folding [2], aggregation [3], liquid-liquid phase separation [4], oligomerization [5], protein-protein interactions [6] and enzymatic activity [7]. The high concentration of macromolecules, such as nucleic acids, proteins, and other biomolecules, in the cellular environment creates a crowded environment that impacts the behavior and properties of these molecules. This impact is known as macromolecular crowding. For instance, the Escherichia coli cytoplasm typically contains 300–400 g/L of macromolecules, which corresponds to volume percentages of 30–40% [8]. Hence, understanding the effect of crowding is essential for elucidating the biophysical and biochemical processes occurring in living cells.

Crowding effects can be classified into two main categories based on the type of interaction made between crowders and protein: hard core repulsion and weak nonspecific interaction. The hard core repulsion emerges because macromolecules occupy space and impede the free movement of other macromolecules [9], so they become spatially constrained. This leads to an entropic effect, which favors the compact native state of proteins, thereby enhancing their stabilities [10]. Weak nonspecific crowder-protein interactions other than steric repulsions are often called soft or chemical interactions [11]. Soft interactions can be either repulsive or attractive, leading to enhancement or reduction of the effects from hard core repulsions [12]. Hence,

depending on the type of soft interaction in various conditions, they can lead to net stabilization or destabilization for proteins [13].

Experimental studies have been carried out to investigate the effect of crowding on the kinetics and thermodynamics of proteins by using synthetic polymers, such as polyethylene glycol [14, 15], dextran [16], and Ficoll [17], as excluded volume crowding agents. Typically, the result of adding these crowders to the solution is a net stabilizing effect on the protein as quantified by an increase in the folding mid-point temperature or unfolding free energy of proteins [18–25]. A net stabilization of the excluded volume effect is generally consistent with theoretical prediction by Minton [26]. In terms of kinetics, impacts on both folding and unfolding rates of proteins from macromolecular crowding have been demonstrated [27–31]. These effects can be nontrivial. For example, Mukherjee *et al.* [29] found that the addition of either Ficoll or dextran resulted in a decrease in the folding rate of a 16-amino acid β-hairpin, but found no effect on two other small α-helical proteins. Dhar *et al.* [17] observed a nonmonotonic dependence of the folding rate on the crowder concentration for the protein phosphoglycerate kinase [32].

Computational studies have also shown that the excluded volume effect stabilizes the native state of protein models [32–42]. However, we found, within a coarse-grained protein model, that small proteins can be left neutral by the excluded volume effect, or even be destabilized, for some protein folds and under conditions where the unfolded state is highly compact [43]. In order to probe the effects of crowding, it is common in computational studies to represent crowders as spherical particles. However, it is clear that synthetic crowders are not hard-core spheres [44]. This is rather obvious

for linear polymers, such as PEG. Ficoll, however, which is a cross-linked polymer of sucrose monomers, have been described as both hard [45] and spherical [46]. However, even Ficoll has been shown to be a soft [47] and highly hydrated macromolecule with strong polymer character [44].

Here, we study the effects of crowding on native state stability and folding kinetics by explicitly treating crowders as linear polymers of different lengths. For comparison, we also include spherical crowders. We focus on two proteins 1PRB and 1SHF that fold into α-helical bundle and β-barrel folds, respectively. In particular, we address the issue of whether spherical or polymeric crowders are most effective at stabilizing proteins, at a given concentration of crowders. We also study how protein stability and kinetics are dependent on the size of crowders, i.e., the radius of spheres and the length of the polymers. To address these issues, we use a structure-based (Gō-like) model to simulate the folding and kinetics of our two proteins. Polymer crowders are treated using a simple bead-spring model with some chain stiffness. Both spherical and polymer crowders exert only repulsive interactions. We find that both crowder types are generally stabilizing. While the magnitude of the stabilizing effect changes with the size of the spherical crowders, the polymeric crowders exhibit a size dependence that is very weak, or even absent. Comparing the two proteins, we find that the stabilization effect is slightly more pronounced for the 1SHF protein relative to 1PRB.

5.2 Material and Methods

5.2.1 Native structures and contact maps

The structures of 1PRB (PDB id: 1PRB) and SH3 domain of FYN (PDB id: 1SHF) were downloaded from the Protein Data Bank (PDB). The former folds to 3α-helices and the latter folds to 5-stranded β-barrel, including 53 and 59 amino acids, respectively. The two generated contact maps contain 82 and 129 contacts, respectively. Contact maps are calculated by defining a cutoff distance of 4.5Å between amino acid pairs using the atomic coordinates extracted from the PDB files.

5.2.2 Coarse-grained model for protein folding

In our model [48], each amino acid is represented by a single interaction site located at the C_α atom. The model is a so-called Gō-type or structure-based model. The conformation of an N-amino-acid chain can therefore be presented by the bead positions $\mathbf{r}_i$, where $i = 1, ..., N$. The interactions in a structure-based model are not based on physio-chemical principles. Rather, they are constructed so that interactions that are present in a reference structure, typically the native structure, are made attractive. Other (non-native) interactions are left neutral or even repulsive. In this particular

model, the structure based energy function has the following form [48]:

$$\begin{aligned} V &= \sum_{\mathrm{i}}^{\mathrm{bonds}} K_{\mathrm{b}}(b_{\mathrm{i}} - b_{\mathrm{i}}^{\mathrm{n}})^2 + \sum_{\mathrm{i}}^{\mathrm{angles}} K_{\theta}(\theta_{\mathrm{i}} - \theta_{\mathrm{i}}^{\mathrm{n}})^2 \\ &+ \sum_{\mathrm{i}}^{\mathrm{dihedrals}} K_{\phi}^{(1)}[1 - \cos(\phi_{\mathrm{i}} - \phi_{\mathrm{i}}^{\mathrm{n}})] + K_{\phi}^{(3)}\left[1 - \cos 3(\phi_{\mathrm{i}} - \phi_{\mathrm{i}}^{\mathrm{n}})\right] \\ &+ \sum_{\mathrm{i<j-3}}^{\mathrm{nonnative}} \epsilon \left(\frac{\sigma}{r_{\mathrm{ij}}}\right)^{12} + \sum_{\mathrm{i<j-3}}^{\mathrm{native}} \epsilon \left[5\left(\frac{r_{\mathrm{ij}}^{\mathrm{n}}}{r_{\mathrm{ij}}}\right)^{12} - 6\left(\frac{r_{\mathrm{ij}}^{\mathrm{n}}}{r_{\mathrm{ij}}}\right)^{10}\right] \end{aligned} \tag{5.1}$$

where b_{i}, θ_{i}, ϕ_{i}, and r_{ij} are virtual bond lengths, bond angles, torsion angles, and $C_\alpha - C_\alpha$ distance between residues i and j, respectively. $b_{\mathrm{i}}^{\mathrm{n}}$, $\theta_{\mathrm{i}}^{\mathrm{n}}$, $\phi_{\mathrm{i}}^{\mathrm{n}}$, and $r_{\mathrm{ij}}^{\mathrm{n}}$ are their values at the native state, which have found in experiment [49].

In the equation 5.1, ϵ sets the energy scale of the model and $r_{\mathrm{ij}} = |\mathbf{r}_{\mathrm{j}} - \mathbf{r}_{\mathrm{i}}|$. The first three terms represent bonded interactions with strengths set to $K_{\mathrm{b}} = 100\epsilon$, $K_\theta = 20\epsilon$, $K_\phi^{(1)} = \epsilon$ and $K_\phi^{(3)} = 0.5\epsilon$. The fourth term represents steric repulsions between bead pairs that do not form contacts in the native structure. The repulsion range is set to $\sigma = 4$ Å. The summation in the last term is over contacts that belong to the native contact set of each protein (see 5.2.1.).

5.2.3 Excluded volume crowders

Spherical crowders

In the case of mimicking crowding agents as spheres, crowder-crowder and crowder-bead interactions are expressed using the potential function suggested by Mittal and Best[37],

$$V(r) = \epsilon \left(\frac{\sigma}{r - \rho + \sigma}\right)^{12}, \tag{5.2}$$

for distances $r > \rho - \sigma$, and $V(r) = \infty$ otherwise. Hence, crowders have a soft repulsive shell over of a hard-core steric repulsion. The parameters σ and ρ, respectively, regulate the width of the soft repulsive shell and the range of the interaction. For crowder-crowder interactions, we set $\rho = 2R_{\mathrm{c}}$ and $\sigma = 2\sigma_{\mathrm{cr}}$, where $\sigma_{\mathrm{cr}} = 3$Å regulates the width of the crowders' soft shells. We set $\rho = R_{\mathrm{c}} + \sigma_{\mathrm{b}}$ and $\sigma = \sigma_{\mathrm{cr}} + \sigma_{\mathrm{b}}$ for crowder-bead interactions, where $\sigma_{\mathrm{b}} = 4$Å is the bead radius. Spherical crowder concentration is a fraction ϕ_{c} of the overall simulation volume V occupied by spherical crowders, i.e. $\phi_{\mathrm{c}} = 4\pi {R_{\mathrm{c}}}^3 N_{\mathrm{cr}}/3V$. The number of crowding particles N_{cr} in our simulations ranges from 3 for $\phi_{\mathrm{c}} = 0.05$ to 92 for $\phi_{\mathrm{c}} = 0.20$.

Polymeric crowders

Polymeric crowders can be described as chains of beads without any attractive native contact energy terms, which is the last term in Eq. 5.1. Parameters for bond length, bond angle and dihedral angles are set the same for all amino acids i. The bonding length for polymer chains is set to $b_{\mathrm{i}}^{\mathrm{n}} = b_{\mathrm{l}} = 3.8$ Å, and both dihedral and bond angle parameters are set to 120°, i.e., $\theta_{\mathrm{i}}^{\mathrm{n}} = \phi_{\mathrm{i}}^{\mathrm{n}} = 120°$. The purely repulsive interactions between these polymer and protein beads are the fourth term in Eq. 5.1, with the sum taken over all bead pairs in the system. All interaction strength parameters are the same as for the protein chain.

The polymer crowders are thus represented as chains of beads with diameter $\sigma = 4$ Å meaning that consecutive bead spheres along the chain overlap slightly. Therefore, we estimate the volume V_{chain} of a polymer chain with a tube of diameter

$d_{\mathrm{p}} = \sigma$ and length $L_{\mathrm{ch}} b_{\mathrm{l}}$, i.e.,

$$V_{\mathrm{chain}} = L_{\mathrm{ch}} b_{\mathrm{l}} \pi (d_{\mathrm{p}}/2)^2 \,. \tag{5.3}$$

The parameter L_{ch} is the number of beads per chain as chain length. Accordingly, the polymeric crowder concentration ϕ_{p} is defined as the fraction of the simulation volume V occupied by the polymer chains, i.e.,

$$\phi_{\mathrm{p}} = N_{\mathrm{ch}} \frac{V_{\mathrm{chain}}}{V}, \tag{5.4}$$

where N_{ch} is the number of polymer crowder chains in the system. In our simulations, the polymer chain concentration ranges from $\phi_{\mathrm{p}} = 0.0058$ for 6 chains with 20 beads to $\phi_{\mathrm{p}} = 0.068$ for 24 chains with 60 beads.

5.2.4 Langevin Dynamics

The conformational sampling technique uses Langevin dynamics, as described in ref [48] (see section 2.3). The equation $m\dot{v}(t) = F_{\mathrm{conf}} - m\gamma v(t) + \eta(t)$ describes the time evolution of the system, where m, v, $\dot{v}$, γ, F_{conf} and $\eta(t)$ are mass, velocity, acceleration, friction coefficient, conformation force, and random force, respectively. The random force $\eta(t)$ is taken from a Gaussian distribution, the variance of which determines the system's temperature. The velocity form of the Verlet algorithm[50] is used for numerical integration of the equation of motion, with an integration time step of $\delta t = 0.005\tau$. Simulations are carried out in the low-friction limit, where $-m\gamma v(t)$ is small relative to the inertial term $m\dot{v}(t)$. A natural unit of time for the dynamics is $\tau = \sqrt{ml^2/\epsilon}$ [51], where ϵ is the magnitude of typical interactions and l is a length

scale, which we set to 4 Å, in this limit. The friction coefficient for beads is chosen to be $\gamma_{\mathrm{b}} = 0.05\tau^{-1}$, and all units are set so that the mass of a bead is $m_{\mathrm{b}} = 1.0$. For crowders, the mass and friction coefficient are set to $m_{\mathrm{c}} = 9.0$ and $\gamma_{\mathrm{c}} = 0.017\tau^{-1}$.

5.2.5 Simulations and analysis details

Spherical and polymeric crowders were simulated by placing them in a cubic box with length 100Å, together with protein, plus applying the periodic boundary conditions. The equilibrium behavior of systems including spherical crowders- and polymeric crowders-proteins are characterized by crowder concentrations ϕ_{c} and ϕ_{p}, respectively, was determined using Langevin dynamics in a range of different temperatures. Simulations were performed at either a fixed temperature or with simulated tempering, as described by [52], in which temperature alters dynamically between a set of temperatures. Temperatures were updated every 100 time steps during the simulated tempering runs. The temperature ranges for 1PRB and 1SHF are varied and are $k_{\mathrm{B}}T = 0.80 - 1.00$ and $1.00 - 1.20$, respectively. Those temperature ranges are chosen to cover the entire transition from the native state to the unfolded state for both proteins. 5 separate runs of 2×10^9 time steps were performed for each system and utilized to estimate averages and statistical uncertainties. The simulations started with a random protein structure (random torsional angles ϕ_{i}) and random crowder positions, then a Monte Carlo-based relaxation step plus removing all hard-core steric clashes.

For simulations with spherical crowders in the range $\phi_{\mathrm{c}} = 0.0 - 0.20$, the radii

of spheres are determined as $R_c = 8, 12, 16$Å. Simulations for polymeric crowders in the range $\phi_p = 0.0058 - 0.068$ are performed in different chain lengths $L_{ch} = 5, 10, 20, 30, 40, 50,$ and 60 with assigned number of chains N_{ch}. For simulations at ϕ_c and $\phi_p > 0$, the protein chains must avoid overlapping with the crowding agents. Accordingly, the crowders must similarly avoid the protein chains.

5.2.6 Observable

The contact criterion used to determine the fractions of native contacts formed, Q_{nat}, is as follows: Two amino acids i and j are said to be in contact if their distances are $r_{ij} = 1.2r^0_{ij}$, where r_{ij} is the distance between the C_α atoms and r^0_{ij} is the distance in the native structure. In calculating P_{nat}, we define the native state, N, as $Q \geq Q_{cut}$, where Q is the number of native contacts, and Q_{cut} is 45 and 61 for 1PRB and 1SHF, respectively.

5.3 Results

5.3.1 Polymeric crowders: size and volume exclusion

We start by characterizing the behavior of our polymeric crowders. For simplicity, we use the same model for protein and polymers in this work, with the main difference that the polymers are not subject to any attractive forces between beads (see Methods for details). Figure 5.1 and Table 5.1 show the radius of gyration, R_g, of isolated

polymer chains, as function of the number of beads, $L_{\rm ch}$. For $L_{\rm ch} \geq 20$, the length dependence of $R_{\rm g}$ fits well to a scaling law, $R_{\rm g} \sim L_{\rm ch}^{\nu}$, where exponent $\nu = 0.61 \pm 0.01$, (see Figure 5.1 A), which is in line with the original Flory value of the exponent (3/5) [53]. It should be pointed out, however, that the range of chain lengths considered here might not permit a very accurate estimation of ν. We assume that our polymer crowders behave roughly as self-avoiding random walks for chain lengths of 20 beads or more.

In the case of many polymer chains in our system, we must distinguish between two different concentration regimes, dilute and semidilute [54]. At low concentrations, in the dilute regime, polymers are well separated and form objects roughly of size $R_{\rm g}$. As the concentration increases, the polymers will start to fill the entire system and entangle, thereby entering the semidilute regime. The transition point between the two regimes occurs at the overlap concentration $\phi_{\rm p}^*$, given by

$$\phi_{\rm p}^* = \frac{V_{\rm chain}}{(4/3)\pi R_{\rm g}^3}, \tag{5.5}$$

where $V_{\rm chain}$ is the volume occupied by the polymer chain. The crossover volume fraction $\phi_{\rm p}^*$ thus corresponds to the volume fraction occupied by the chain within its pervaded volume, and is strongly dependent on $L_{\rm ch}$. Because $V_{\rm chain} \sim L_{\rm ch}$ and $R_{\rm g} \sim L_{\rm ch}^{0.61}$, we expect a scaling $\phi_{\rm p}^* \sim L_{\rm ch}^{-0.83}$, which indeed fits well with calculated values of $\phi_{\rm p}^*$ for $L_{\rm ch} \geq 20$, as shown in Figure 5.1B. Our crowded systems are mostly in the dilute regime, except for the longest chains ($L_{\rm ch} = 60$) with 24 chains for which the volume fraction $\phi_{\rm p} = 0.068$ which is very similar to the crossover volume fraction ${\phi_{\rm p}}^* = 0.064$ at this length.

L_{ch}	R_{g} (Å)
5	4.3
10	7.0
20	11.2
30	14.5
40	17.2
50	19.7
60	22.0

Table 5.1: Radius of gyration, R_{g}, for isolated polymer crowder chains for different numbers of beads, L_{ch}.

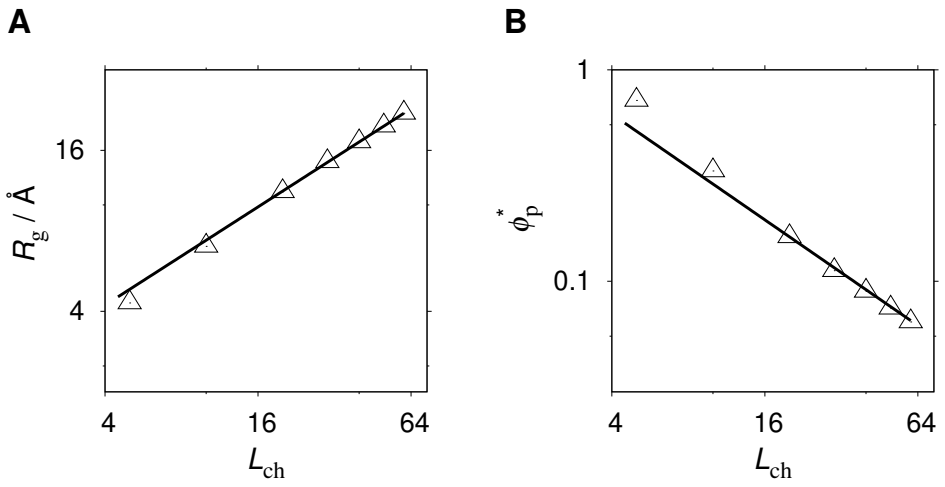

Figure 5.1: Scaling of polymer chain size and crossover volume fraction. Dependence of (A) the radius of gyration, R_{g}, and (B) the crossover volume fraction, ϕ_{p}^*, on the polymer length, L_{ch}, as given in table 5.1. For $L_{\mathrm{ch}} \geq 20$, calculated values are well described by the relations $R_{\mathrm{g}} \sim L_{\mathrm{ch}}^{0.61}$ and $\phi_{\mathrm{p}}^* \sim L_{\mathrm{ch}}^{-0.83}$ (see text), shown as black solid lines.

5.3.2 Native structures are not perturbed by polymeric or spherical crowders

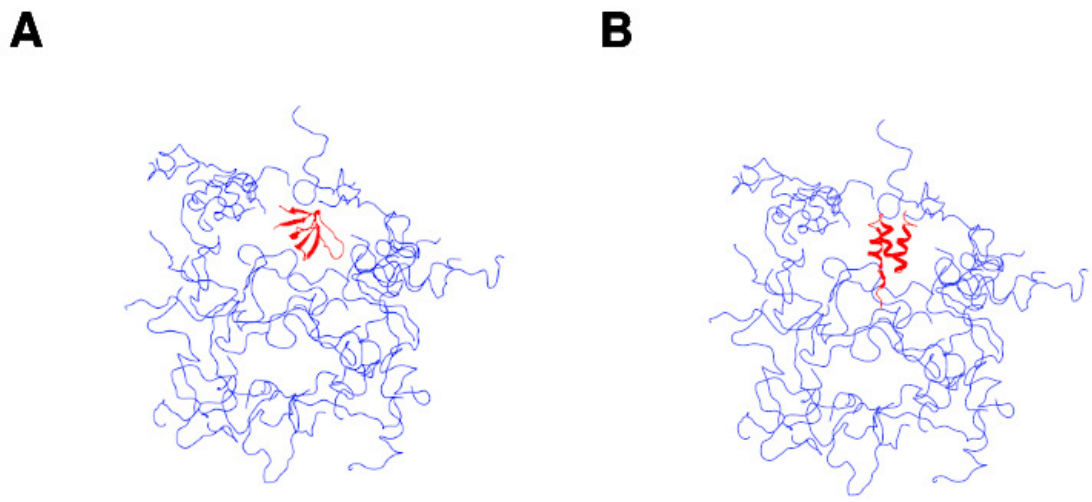

Figure 5.2: Snapshots of 1SHF and 1PRB with polymeric crowders. Snapshots of folding simulations of (A) 1SHF and (B) 1PRB (red), taken with the proteins in their respective native states, in the presence of polymeric crowders with 40 beads per chain (blue). In both cases there are 24 polymeric crowder chains corresponding to an occupied volume fraction $\phi_{\rm p} = 0.046$.

Next, we study the folding of our two proteins, 1PRB and 1SHF, in the presence of polymer crowders of different lengths, $L_{\rm ch}$, and in the presence of spherical crowders of different radii, $R_{\rm c}$. An illustration of these simulations is shown in Fig. 5.2. For comparison, we also study folding in the absence of crowders. To vary the crowder concentrations, we include different numbers of crowder chains or spheres in our simulation box, which we keep fixed.

As a progress variable for folding, we use the fraction of formed native contacts, Q. We picked this variable because it has been shown that Q is a good reaction coordinate for folding in several different models [55, 56], i.e., folding rates are well captured by the height of the folding barrier. Figure 5.3 C and D show representative free energy profiles, $F(Q)$, taken close to the midpoint temperature $T_{\rm m}^0$ of the respective proteins in the absence of crowders. In general, we find that the overall two-state character of the folding of our proteins remains even under crowded conditions. In particular, this means that we can use a single definition of the native state, $Q \geq Q_{\rm cut}$, where the cutoff values $Q_{\rm cut}$ are chosen based on the location of the free energy barrier at $\phi = 0$. For both dilute and crowded cases, the population of the native state, $P_{\rm nat}$, approaches unity at temperatures well below $T_{\rm m}^0$, and $P_{\rm nat}$ as a function of T is well described by a two-state equation, as shown in Fig 5.3A and B.

5.3.3 Both polymeric and spherical crowders enhance stability

Next we turn to the change in equilibrium stability of 1PRB and 1SHF upon the addition of crowders. As shown in Fig. 5.3C and D the destabilization of unfolded state due to the presence of crowders is visible from an upward shift in the free energy profile near the unfolded state basin. As done previously [43], we quantify the native state stability using two different quantities: the midpoint folding temperatures, $T_{\rm m}$, and the free energy of folding $\Delta F = -k_{\rm B}T \ln[P_{\rm nat}/(1 - P_{\rm nat})]$. Stabilization of the native state is indicated by an increase in $T_{\rm m}$ and a decrease in ΔF.

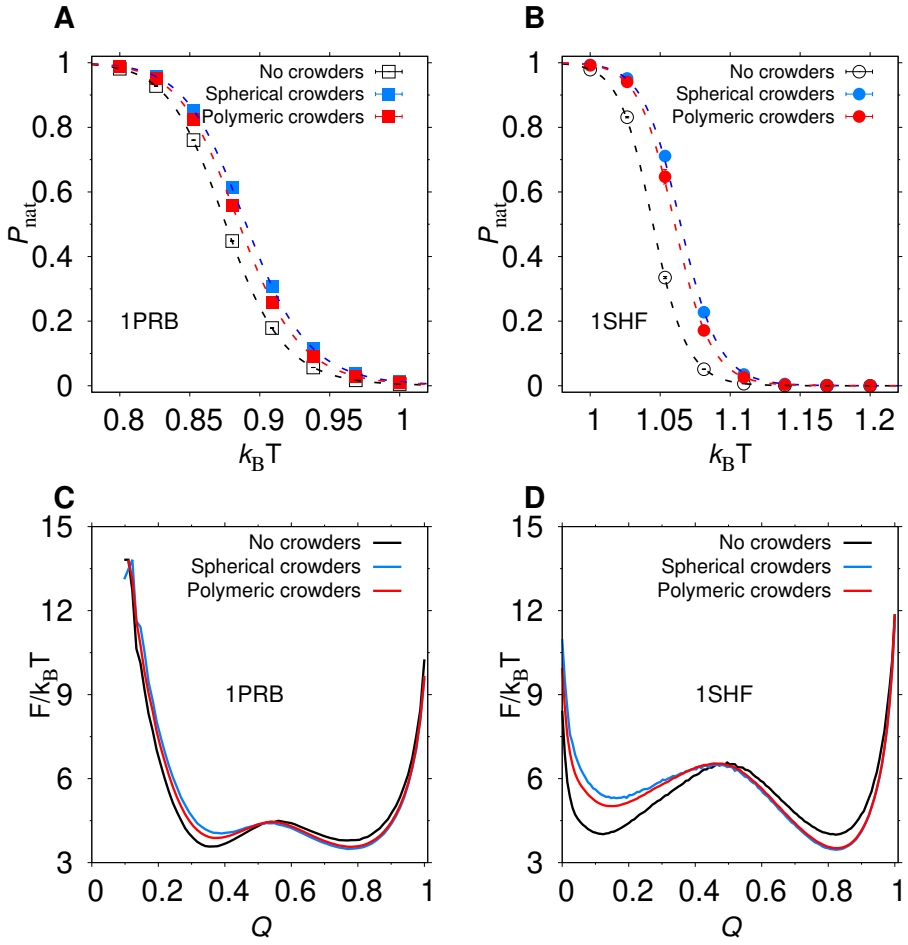

Figure 5.3: Folding curves and free energy profiles. The native state population, $P_{\rm nat}$, as a function of temperature T, for (A) 1PRB and (B) 1SHF in the absence of crowders (open squares or open circles) and in the presence of spherical crowder with $R_{\rm c} = 12$ Å and $\phi_{\rm c} = 0.20$ (solid blue squares or solid blue circles) and in the presence of polymeric crowder with $L_{\rm ch} = 5$ and $\phi_{\rm p} = 0.05$ (solid red squares or solid red circles) crowders. These curves are fitted to the equation $P_{\rm nat} = K/(1+K)$ (two-state model), where $K = \exp[-\Delta E(1/k_{\rm B}T - 1/k_{\rm B}T_{\rm m})]$, and ΔE and $T_{\rm m}$ are free parameters (dashed curves). Free energy profiles $F(Q) = -k_{\rm B}T \ln P(Q)$, where Q is the number of native contacts and $P(Q)$ is a probability distribution, for (C) 1PRB and (D) 1SHF, for each of the systems shown in (A) and (B), respectively. The temperature in (C) and (D) is the respective midpoint temperatures ($T = T_{\rm m}^0$) of the two proteins in the absence of crowders.

Our results for both polymer and spherical crowders across a range of crowder sizes and concentrations, are summarized in Figures 5.4 (1PRB) and 5.5 (1SHF). The general trends are the following. 1PRB and 1SHF are stabilized by the addition of either polymeric or spherical crowders, and the stabilization increases with the number of added crowders. Moreover, stabilization increases with both the polymer length L_{ch} and the radius of the spherical crowders, R_{c}, at a given number of crowders. Hence, the stability of 1PRB and 1SHF is a monotonically increasing function of the crowder concentration (ϕ_{c} or ϕ_{p}) for both types of crowder agents.

The two proteins thus follow similar trends in terms of stabilization, as quantified either by folding midpoint temperature or folding free energy (see Fig 5.4 and Fig 5.5). However, the stabilization is slightly stronger for 1SHF than 1PRB under the same crowding conditions. For example, in the presence of polymeric crowders of length $L_{\mathrm{ch}} = 60$ with $N_{\mathrm{ch}} = 24$ chains, the relative increase in the midpoint temperature is larger for 1SHF than for 1PRB (see Figs 5.4 A and 5.5 A). A similar trend emerges with spherical crowders, e.g., the increase in the midpoint temperature for $R_{\mathrm{c}} = 8$ Å with $N_{\mathrm{cr}} = 92$ crowders, is larger for 1SHF than for 1PRB (see Figs 5.4 C and 5.5 C.).

5.3.4 Size dependence of stability for spherical and polymeric crowders

The addition of small spherical crowders leads to a larger increase in protein stability compared to the addition of large spherical crowders, at a given fixed volume fraction

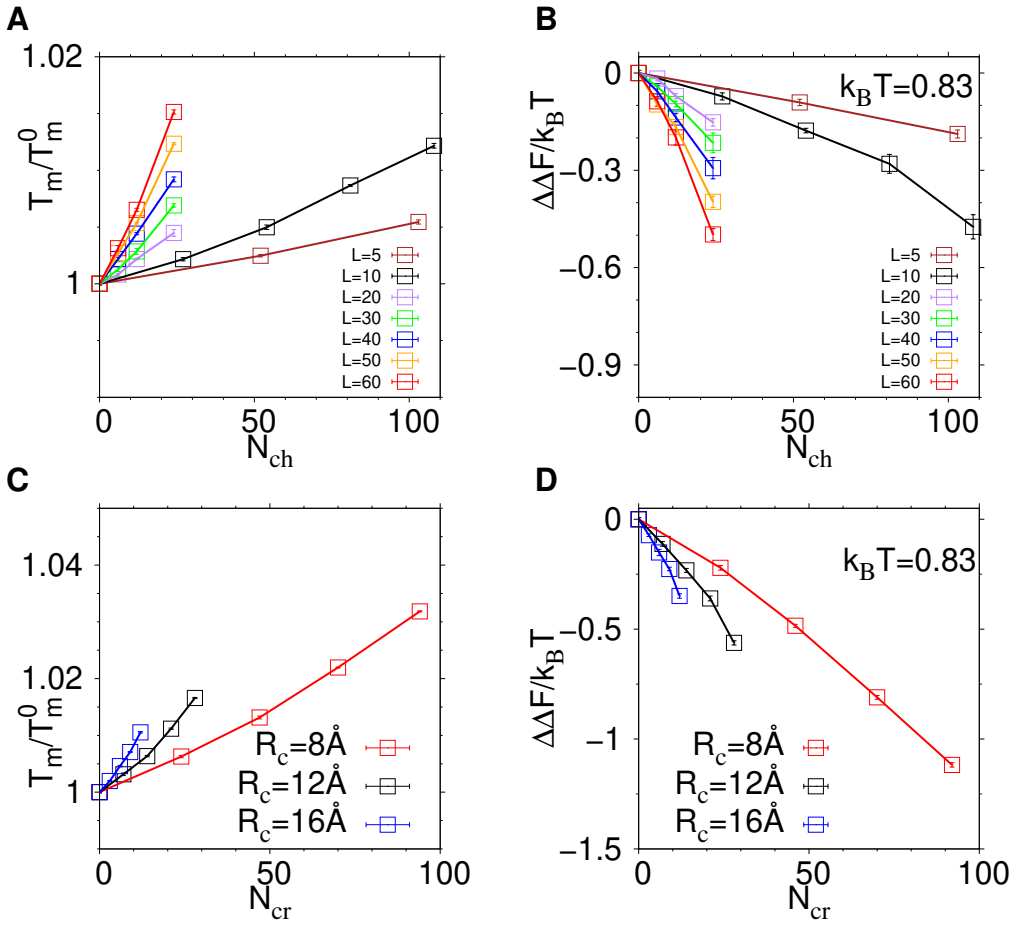

Figure 5.4: Midpoint temperature and folding free energy changes for 1PRB. Change in the folding midpoint temperature, T_{m}, versus (A) the number of polymer chains N_{ch} and (C) the number of spherical crowders, N_{cr}. Change in the free energy of folding $\Delta\Delta F$ as a function of (B) number of polymer chains N_{ch} and (D) number of spherical crowders N_{cr}, taken at the temperature $k_{\mathrm{B}}T = 0.83$ in the model unit.

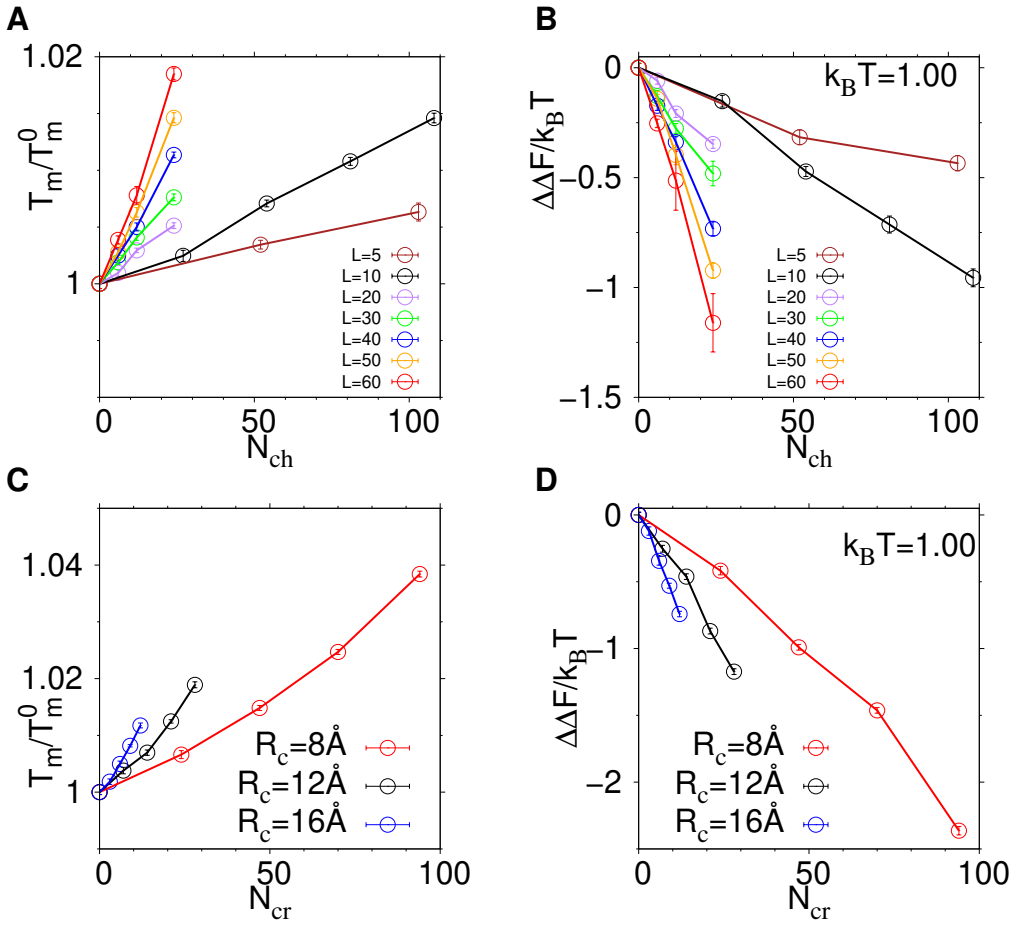

Figure 5.5: Midpoint temperature and folding free energy changes for 1PRB. Change in the folding midpoint temperature, T_{m}, versus (A) the number of polymer chains N_{ch} and (C) the number of spherical crowders, N_{cr}. Change in the free energy of folding $\Delta\Delta F$ as a function of (B) number of polymer chains N_{ch} and (D) number of spherical crowders N_{cr}, taken at the temperature $k_{\mathrm{B}}T = 1.00$ in the model unit.

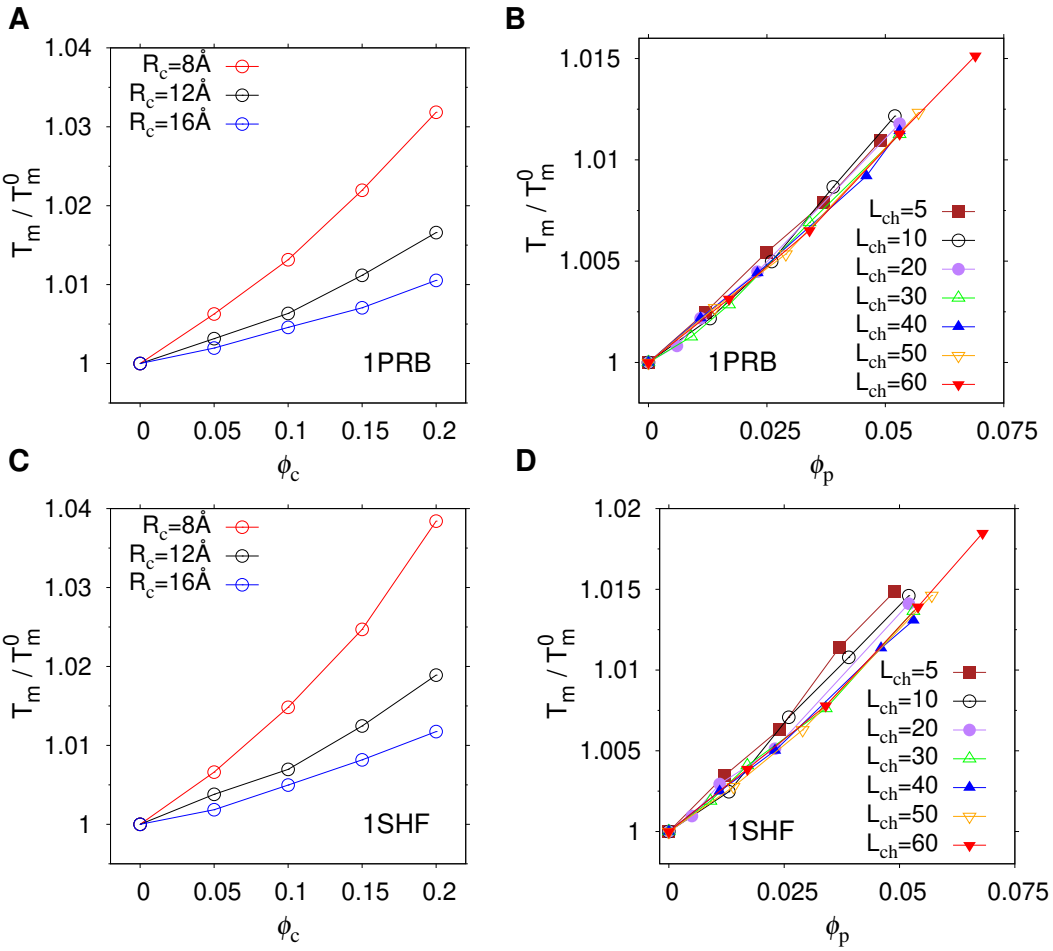

Figure 5.6: Midpoint temperature changes for 1PRB. Midpoint temperature of (A and B) 1PRB and (C and D) 1SHF as a function of volume fraction of (A and C) spherical and (B and D) polymeric crowders.

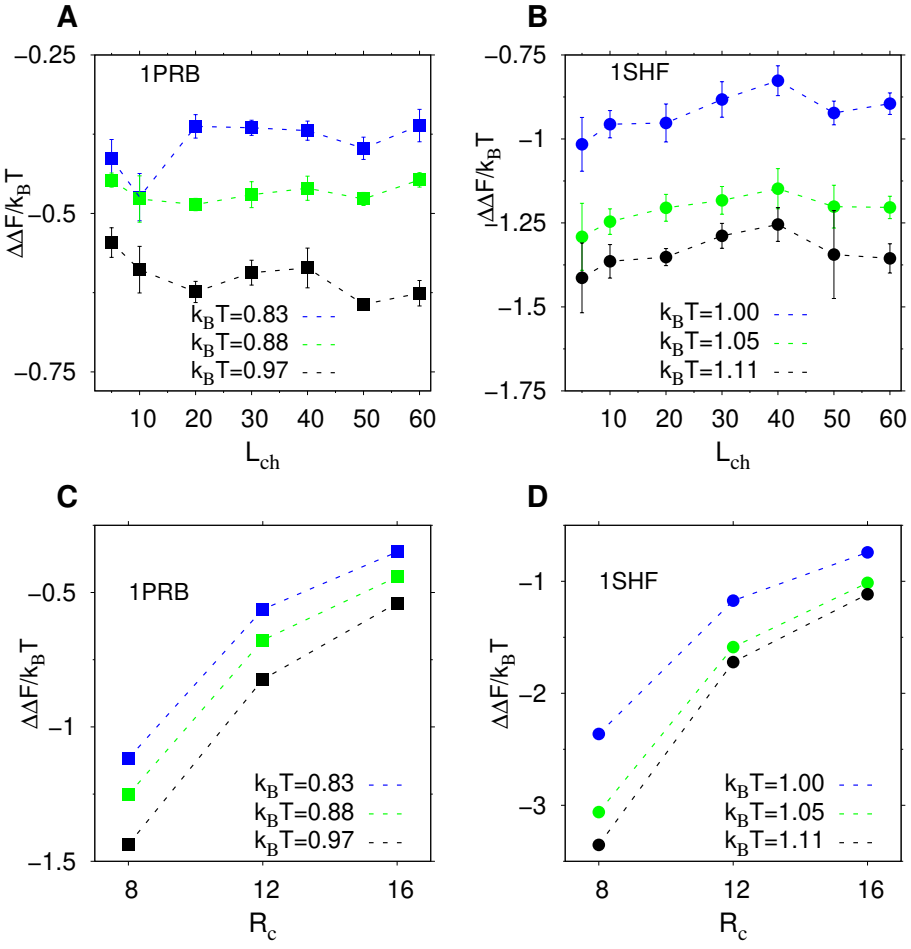

Figure 5.7: Size dependency of stability for 1PRB and 1SHF. Change in the free energy of folding $\Delta\Delta F$ (A (1PRB) and B (1SHF)) as a function of polymers chain length $L_{\rm ch}$ and as a function of (C (1PRB) and D (1SHF)) crowders radii $R_{\rm c}$, taken at different temperatures for $\phi_{\rm c} = 0.20$ and $\phi_{\rm p} = 0.053$.

ϕ_c, as can be seen in Fig 5.5A and C. Such a size dependence for spherical crowders have been found in previous computational studies [37, 38, 43] and is expected from theory [10]. In contrast, we find that the effect of polymeric crowders on stability does not show a very strong size dependence. Spanning chain lengths from 5–60 beads, the increase in T_m is mostly determined by the volume fraction, ϕ_p, and not L_{ch}, as can be seen in Fig 5.6B and D. However, there is a weak length dependence for 1SHF to be discussed below. The lack of a strong dependence on L_{ch} may in part be due to the fact that we are only able to explore relatively small volume fractions, ϕ_p, for the polymer crowders.

In order to more closely examine the length dependence for the polymer crowders, we consider $\Delta\Delta F$ as a function of L_{ch} at three different temperatures: below, above, and in the proximity of the folding midpoint temperature. Numerically, these temperatures correspond to $k_B T = 0.83$, 0.88 and 0.97, for 1PRB, and $k_B T = 1.00$, 1.05, and 1.11, for 1SHF. In general, the stabilization is stronger (lower $\Delta\Delta F$) at higher temperatures for both polymer and spherical crowders. This likely results from the fact that, at higher T, the unfolded state is more expanded, which leads a stronger entropic disfavoring of the unfolded state relative to the compact native state. Regarding the length dependence, for 1PRB, no clear trend in chain length L_{ch} can be discerned, as mentioned above. For 1SHF, the weakest impact on stability, as indicated by the least negative $\Delta\Delta F$, occurs for $L_{ch} = 40$ for all three temperatures. Interestingly, in a recent study, Pielak *et al.* [57] studied the impact of stabilizing PEG crowders on an SH3 domain, which has the same native state topology as our protein 1SHF, across different polymer lengths, and observed that the impact on stability was

smallest for medium sized PEGs with molecular mass in the interval 400-1000 g/mol. Using 44 g/mol for the molecular mass of the PEG monomer (ethylene glycol), these medium sized PEG chains correspond to $\approx$ 9-23 monomer units.

5.3.5 Folding rate enhanced by crowding agents

In this study, we aim to investigate the folding rate k_{f} of the protein using Langevin dynamics simulations. The folding rate, a fundamental kinetic parameter characterizing the folding process, quantifies the rate at which the protein transits from its initial unfolded configuration to its stable native state. This parameter sheds light on the energetics and mechanisms underlying the folding process. To determine k_{f}, we employ the following equation:

$$k_{\mathrm{f}} = \frac{1}{t_{\mathrm{f}}}, \tag{5.6}$$

where t_{f} is the average number of Langevin dynamics steps it takes to transition from unfolded states U to the native state, N. The number of native contacts Q has been used to determine the unfolded and native states. Accordingly, for 1PRB, we classify a conformation to be in the unfolded state if $Q < Q^{\mathrm{U}}_{\mathrm{cut}} = 33$, and in the native state if $Q > Q^{\mathrm{N}}_{\mathrm{cut}} = 60$. Similarly, an unfolded state defined as $Q < Q^{\mathrm{U}}_{\mathrm{cut}} = 34$, and the native state as $Q > Q^{\mathrm{N}}_{\mathrm{cut}} = 90$, for 1SHF. The Q_{cut} cutoff values for the unfolded and native states were determined based on the free energy profile of 1PRB and 1SHF (see Fig. 5.3 C and D). These cutoffs in Q implement more restrictive definitions of U and N than those for the calculation of the equilibrium quantity P_{nat}, previously (see 5.2.1). The more restrictive definitions are necessary here so that

transient fluctuations within U, which reach the peak of the free energy barrier and return to the U basin, are not counted as folding transitions.

We investigate the effect of excluded volume crowding on the kinetics of two proteins, 1PRB and 1SHF. To examine this effect, we conducted fixed-T simulations in the presence of polymeric or spherical crowding agents at different volume fractions at $T = T_{\mathrm{m}}^{0}$. For each protein-crowder system, 5 separate runs of 2×10^{9} steps were performed to estimate averages of the quantity t_{f}. The parameter t_{f} is calculated as the average number of simulation time steps it takes during a trajectory to reach an N state conformations, defined as $Q > Q_{\mathrm{cut}}^{\mathrm{N}}$, from the time the trajectory first entered into the U state, defined as $Q < Q_{\mathrm{cut}}^{\mathrm{U}}$.

We study the relative increase in the folding rate $k_{\mathrm{f}}/k_{\mathrm{f}}^{0}$ upon adding any type of crowders where k_{f}^{0} is the rate at $\phi_{\mathrm{c}} = \phi_{\mathrm{p}} = 0$. The values for the folding rate in the absence of crowders are $k_{\mathrm{f}}^{0} = 1.7 \times 10^{-5}$ and $k_{\mathrm{f}}^{0} = 1.6 \times 10^{-6}$ for 1PRB and 1SHF, respectively in the model time unit. We find that the folding rate k_{f} increases monotonically with the crowder volume fraction for both proteins and for both crowder types over the studied range of crowder sizes and concentrations. The excluded volume crowding effect on the 1SHF folding rate is more strongly enhanced by both polymer crowders (see Fig 5.8A and C) and spherical crowders (see Fig 5.8B and D) than 1PRB.

We also investigate the effect of crowding agent size on the folding rate of both proteins at a constant volume fraction $\phi_{\mathrm{p}} = 0.053$ for polymeric crowders and at a constant volume fraction $\phi_{\mathrm{c}} = 0.20$ for spherical crowders. Interestingly, for polymer

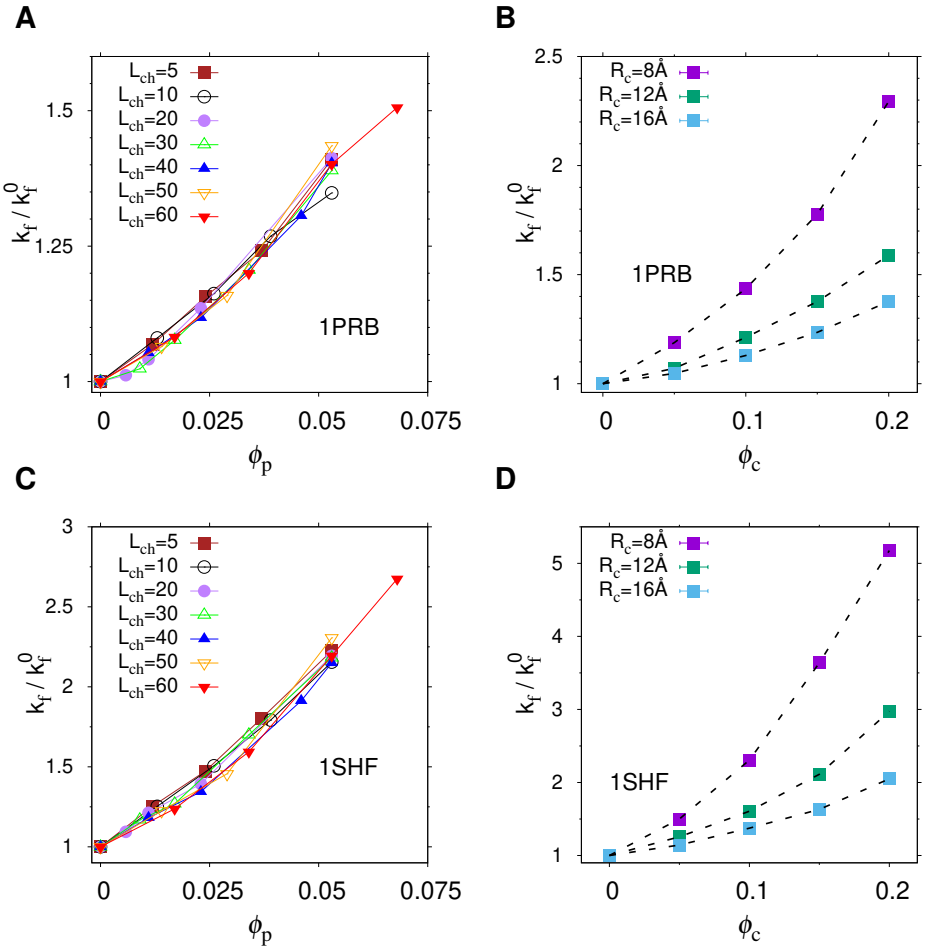

Figure 5.8: Folding kinetic of 1PRB and 1SHF. The relative increase in folding rate k_f/k_f^0 for the (A and B) 1PRB and (C and D) 1SHF proteins as a function of crowders volume fraction. (A and C) are for polymeric crowders and (B and D) are for spherical crowders. The volume fraction for polymeric and spherical crowders are ϕ_p and ϕ_c, respectively. All values are taken at the folding midpoint temperatures.

crowders, we find that the folding rate does not depend strongly on the chain length, L_{ch}, as shown in Fig. 5.9 A and C. This result holds for both proteins. The limited impact of L_{ch} on the folding rate might be attributed, at least in part, to the relatively small polymer crowder volume fractions studied here. Spherical crowders show a different trend. At a constant volume fraction, increasing the size of spherical crowders R_{c} leads to a smaller effect on the folding rate k_{f}.

Overall, our main finding is that the effect of excluded volume crowding is to monotonically increase the rate of folding. However, the extent of the speedup of folding depends on the protein and the crowder type. Furthermore, the sharp increase in the folding rate observed for 1SHF at higher volume fractions ($\phi_{\mathrm{c}} > 0.20$) of spherical crowders suggests that the effect of crowding on protein folding kinetic can be non-linear and complex in our model.

5.4 Discussion and Conclusions

In this study, we investigated the effect of purely repulsive excluded volume crowding on the stability and folding kinetics of two different proteins, 1PRB and 1SHF, using both polymeric and spherical crowding agents by molecular dynamic simulations and a coarse-grained model. Our results demonstrate that while both proteins are generally stabilized by both types of crowders, the nature and degree of the stabilizing effect depend on the crowder agent used and the protein being studied.

1SHF exhibits a greater crowder-induced stabilization than 1PRB in response to

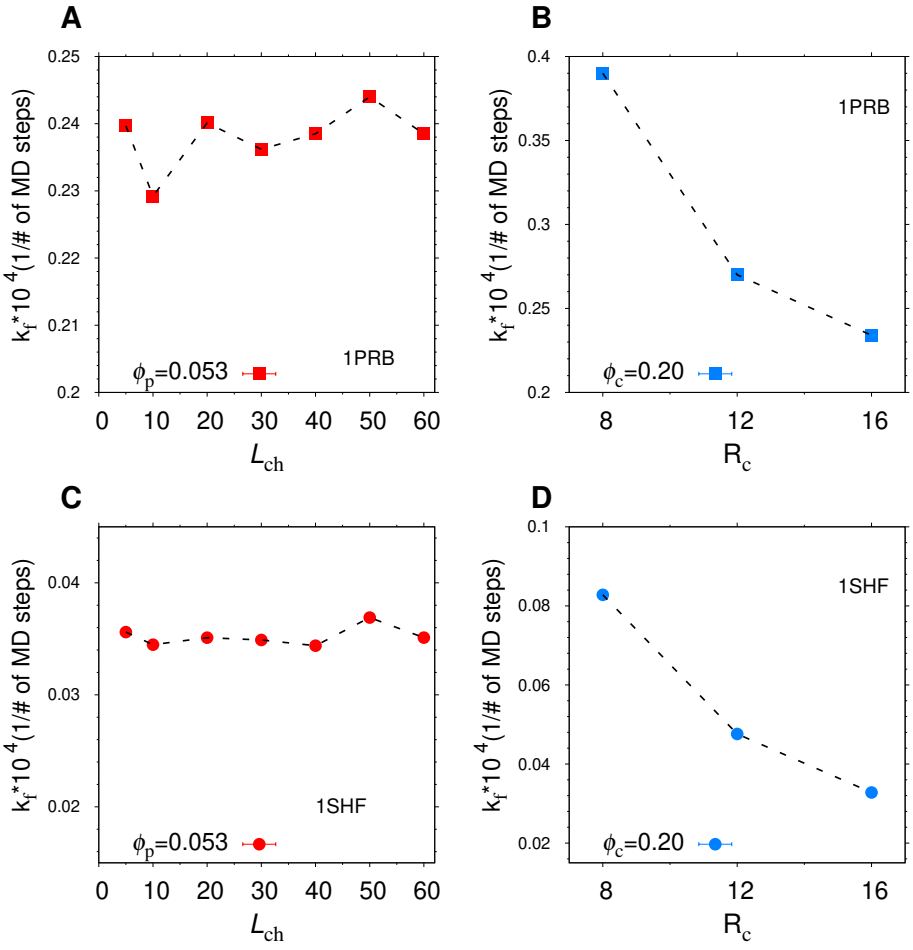

Figure 5.9: Folding kinetic size dependency of 1PRB and 1SHF. The folding rate k_f for the (A) 1PRB and (C) 1SHF protein as a function of polymeric crowder size L_{ch} at constant volume fraction $\phi_p = 0.053$. The folding rate k_f for (B) 1PRB and (D) 1SHF protein as a function of spherical crowders size R_c at constant volume fraction $\phi_c = 0.20$. All values are taken at the folding midpoint temperatures.

both spherical and polymeric crowding agents. This difference may be related to the distinct topologies of the two proteins, because 1SHF is a β-sheet protein and 1PRB is an α-helical protein. However, the difference between the two proteins in terms of the stabilization observed is rather small.

For polymeric crowders, we find that the stabilization of 1PRB by polymer crowders exhibit basically no dependence on the chain length, L_{ch}, for a given crowder volume fraction. For 1SHF, the greatest stabilization is provided by the smallest chains ($L_{\mathrm{ch}} = 5$ and 10) while the smallest stabilization is provided by $L_{\mathrm{ch}} = 40$. Although the length dependence we observe is weak also for 1SHF, the results mirror those of Stewart *et al.* [57] who studied native state stability of an SH3 domain in the presence of ethylene glycol (EG) and polyethylene glycol (PEG) of various sizes. These authors divided their crowders into small (< 400 g/ml), medium (400-1000 g/mol), and large (> 1000 g/mol) sized PEGs. The greatest stabilization occurred for small PEGs, the smallest stabilization occurred for the medium sized PEGs, and the large PEGs had a stabilization in between the small and medium types. In this background, it would be interesting to carry out additional simulations within our model and modify our polymer crowders to mimic PEG chains specifically. In contrast, the difference in impact of spherical crowders of different sizes is more pronounced. Smaller spherical crowders are observed to enhance the stability of the native state more for both proteins at a fixed volume fraction compared to large spherical crowders. This is illustrated in Fig. 5.7 (C and D).

Notably, we observed that the folding rate of both proteins increased upon adding both spherical and polymeric crowding agents. Cheung *et al.* [32], using a model

similar to ours and spherical crowders, found a non-monotonic dependence of the folding rate k_{f} on the crowder packing fraction. Specifically, they found an increase in k_{f} with crowder concentration until $\phi_{\mathrm{c}} \approx 0.10$, followed by a sharp decrease at higher ϕ_{c}. We did not observe such non-monotonicity in our model, either for spherical or polymer crowders. A difference between our simulations and those of Cheung *et al.* [32] is that these authors carried out their simulations using Brownian dynamics while we have relied on Langevin dynamics. Specifically, our findings suggest that the size and shape of crowding agents and the specific protein being studied must be considered when designing experimental or computational studies to investigate the effects of excluded volume crowding on protein behavior.

In conclusion, our simulations demonstrate that excluded volume crowders can significantly impact the native state stabilities and folding rates of proteins. We found that adding crowders generally stabilizes the native state of both 1PRB and 1SHF proteins, as indicated, e.g., by a decrease in the folding free energy. The magnitude of this effect varied depending on the type and size of the crowders. Moreover, we observed that the effect of both spherical polymeric crowders on the folding rate is more pronounced for 1SHF compared to 1PRB (see Fig. 5.8). It would be interesting to explore, in future studies, if this difference between 1SHF and 1PRB is related to the character of the respective unfolded states of the two proteins, and how this state interacts with the different crowders.

Bibliography

[1] F. X. Theillet, A. Binolfi, T. Frembgen-Kesner, K. Hingorani, M. Sarkar, C. Kyne, C. Li, P. B. Crowley, L. Gierasch, G. J. Pielak, A. H. Elcock, A. Gershenson, and P. Selenko. Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs). *Chem Rev*, 114:6661–6714, 2014.

[2] R. J. Ellis. Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci*, 26:597–604, 2001.

[3] T. Niwa, R. Sugimoto, L. Watanabe, S. Nakamura, T. Ueda, and H. Taguchi. Large-scale analysis of macromolecular crowding effects on protein aggregation using a reconstituted cell-free translation system. *Front Microbiol*, 6:1113, 2015.

[4] A. A. M. André and E. Spruijt. Liquid-liquid phase separation in crowded environments. *Int J Mol Sci*, 21:5908, 2020.

[5] R. J. Ellis. Protein misassembly: macromolecular crowding and molecular chaperones: Molecular aspects of the stress response: Chaperones, membranes and networks. *Adv Exp Med Biol*, 594:1–13, 2007.

[6] T. C. Jarvis, D. M. Ring, S. S. Daube, and PH. Von Hippel. "macromolecular

crowding": thermodynamic consequences for protein-protein interactions within the t4 dna replication complex. *J Biol Chem*, 265:15160–15167, 1990.

[7] I. Pozdnyakova and P. Wittung-Stafshede. Non-linear effects of macromolecular crowding on enzymatic activity of multi-copper oxidase. *Biochim Biophys Acta*, 1804:740–744, 2010.

[8] S. B. Zimmerman and S. O. Trach. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J Mol Biol*, 222:599–620, 1991.

[9] A. P. Minton. Excluded volume as a determinant of macromolecular structure and reactivity. *Biopolymers*, 20:2093–2120, 1981.

[10] A. P. Minton. The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences. *Mol Cell Biochem*, 55:119–140, 1983.

[11] A. C. Miklos, M. Sumpter, and H. X. Zhou. Competitive interactions of ligands and macromolecular crowders with maltose binding protein. *PLoS One*, 8:e74969, 2013.

[12] M. Sarkar, C. Li, and G. J. Pielak. Soft interactions and crowding. *Biophys Rev*, 5:187–194, 2013.

[13] J. Rosen, Y. C. Kim, and J. Mittal. Modest protein- crowder attractive interactions can counteract enhancement of protein association by intermolecular excluded volume interactions. *J Phys Chem B*, 115:2683–2689, 2011.

[14] A. AM. André, N. A. Yewdall, and E. Spruijt. Crowding-induced phase separation and gelling by co-condensation of peg in npm1-rrna condensates. *Biophys J*, 122:397–407, 2023.

[15] Md. S. Rahman, M. A. Gulshan, S. Matsumura, and Y. Ikawa. Polyethylene glycol molecular crowders enhance the catalytic ability of bimolecular bacterial rnase p ribozymes. *Nucleosides Nucleotides Nucleic Acids*, 39:715–729, 2020.

[16] S. Biswas and P. K. Chowdhury. Unusual domain movement in a multidomain protein in the presence of macromolecular crowders. *Phys Chem Chem Phys*, 17:19820–19833, 2015.

[17] A. Dhar, A. Samiotakis, S. Ebbinghaus, L. Nienhaus, D. Homouz, M. Gruebele, and M. S. Cheung. Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding. *Proc Natl Acad Sci USA*, 107:17586–17591, 2010.

[18] K. Sasahara, P. McPhie, and A. P. Minton. Effect of dextran on protein stability and conformation attributed to macromolecular crowding. *J Mol Biol*, 326:1227–1237, 2003.

[19] D. S. Spencer, K. Xu, T. M. Logan, and H. X. Zhou. Effects of pH, salt, and macromolecular crowding on the stability of FK506-binding protein: an integrated experimental and theoretical study. *J Mol Biol*, 351:219–232, 2005.

[20] L. A. Benton, A. E. Smith, G. B. Young, and G. J. Pielak. Unexpected effects of macromolecular crowding on protein stability. *Biochemistry*, 51:9773–9775, 2012.

[21] Y. Wang, L. A. Benton, V. Singh, and G. J. Pielak. Disordered protein diffusion under crowded conditions. *J Phys Chem Lett*, 3:2703–2706, 2012.

[22] B. Köhn and M. Kovermann. Macromolecular crowding tunes protein stability by manipulating solvent accessibility. *Chembiochem*, 20:759–763, 2019.

[23] J. Hong and L. M. Gierasch. Macromolecular crowding remodels the energy landscape of a protein by favoring a more compact unfolded state. *J Am Chem Soc*, 132:10445–10452, 2010.

[24] S. Mittal and L. R. Singh. Denatured state structural property determines protein stabilization by macromolecular crowding: a thermodynamic and structural approach. *PLOS One*, 8:e78936, 2013.

[25] A. Christiansen and P. Wittung-Stafshede. Synthetic crowding agent dextran causes excluded volume interactions exclusively to tracer protein apoazurin. *FEBS Lett*, 588:811–814, 2014.

[26] A P Minton. Excluded volume as a determinant of protein structure and stability. *Biophys J*, 32:77–79, 1980.

[27] B. van den Berg, R. Wain, C. M. Dobson, and R. J. Ellis. Macromolecular crowding perturbs protein refolding kinetics: implications for folding inside the cell. *The EMBO journal*, 19:3870–3875, 2000.

[28] A. H. Gorensek-Benitez, A. E. Smith, S. S. Stadmiller, G. M. Perez Goncalves, and G. J. Pielak. Cosolutes, crowding, and protein folding kinetics. *J Phys Chem B*, 121:6527–6537, 2017.

[29] S. Mukherjee, M. M. Waegele, P. Chowdhury, L. Guo, and F. Gai. Effect of macromolecular crowding on protein folding dynamics at the secondary structure level. *J Mol Biol*, 393:227–236, 2009.

[30] G. B. Ralston. Effects of" crowding" in protein solutions. *J Chem Educ*, 67:857, 1990.

[31] D. De Sancho, J. Mittal, and R. B. Best. Folding kinetics and unfolded state dynamics of the gb1 hairpin from molecular simulation. *J Chem Theory Comput*, 9:1743–1753, 2013.

[32] M. S. Cheung, D. Klimov, and D. Thirumalai. Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc Natl Acad Sci USA*, 102:4753–4758, 2005.

[33] S. R. McGuffee and A. H. Elcock. Protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLOS Comput Biol*, 6:e1000694, 2010.

[34] A. Bille, S. Mohanty, and A. Irbäck. Peptide folding in the presence of interacting protein crowders. *J Chem Phys*, 144:175105, 2016.

[35] I. Yu, T. Mori, T. Ando, R. Harada, J. Jung, Y. Sugita, and M. Feig. Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *Elife*, 5, 2016.

[36] M. Candotti and M. Orozco. The Differential Response of Proteins to Macromolecular Crowding. *PLOS Comput Biol*, 12:e1005040, 2016.

[37] J. Mittal and R. B. Best. Dependence of protein folding stability and dynamics on the density and composition of macromolecular crowders. *Biophys J*, 98:315–320, 2010.

[38] D. Tsao and N. V. Dokholyan. Macromolecular crowding induces polypeptide compaction and decreases folding cooperativity. *Phys Chem Chem Phys*, 12:3491–3500, 2010.

[39] B. P. Thuy, H. T. T. Huong, and T. X. Hoang. Effects of macromolecular crowding on protein folding. *J Phys Conf Ser*, 627:012027, 2015.

[40] B. Macdonald, S. McCarley, S. Noeen, and A. E. van Giessen. β-hairpin crowding agents affect α-helix stability in crowded environments. *J Phys Chem B*, 120:650–659, 2016.

[41] D. Gomez, K. Huber, and S. Klumpp. On protein folding in crowded conditions. *J Phys Chem Lett*, 10:7650–7656, 2019.

[42] Y. C. Kim, A. Bhattacharya, and J. Mittal. Macromolecular crowding effects on coupled folding and binding. *J Phys Chem B*, 118:12621–12629, 2014.

[43] S. Bazmi and S. Wallin. Crowding-induced protein destabilization in the absence of soft attractions. *Biophys J*, 121:2503–2513, 2022.

[44] V. T. Ranganathan, S. Bazmi, S. Wallin, Y. Liu, and A. Yethiraj. Is ficoll a colloid or polymer? a multitechnique study of a prototypical excluded-volume macromolecular crowder. *Macromolecules*, 55:9103–9112, 2022.

[45] Loren Stagg, Shao-Qing Zhang, Margaret S Cheung, and Pernilla Wittung-Stafshede. Molecular crowding enhances native structure and stability of α/β protein flavodoxin. *Proc Natl Acad Sci USA*, 104:18976–18981, 2007.

[46] A. Christiansen, Q. Wang, A. Samiotakis, M. S. Cheung, and P. Wittung-Stafshede. Factors defining effects of macromolecular crowding on protein stability: an in vitro/in silico case study using cytochrome c. *Biochemistry*, 49(31):6519–6530, 2010.

[47] William H Fissell, Christina L Hofmann, Ross Smith, and Michelle H Chen. Size and conformation of ficoll as determined by size-exclusion chromatography followed by multiangle light scattering. *Am J Physiol - Ren Physiol*, 298:F205–F208, 2010.

[48] S. Wallin and H. S. Chan. Conformational entropic barriers in topology-dependent protein folding: perspectives from a simple native-centric polymer model. *J Condens Matter Phys*, 18:S307, 2006.

[49] P. K. Zuber, I. Artsimovitch, M. NandyMazumdar, Z. Liu, Y. Nedialkov, K. Schweimer, P. Rösch, and S. H. Knauer. The universally-conserved transcription factor rfah is recruited to a hairpin structure of the non-template dna strand. *ELife*, 7:e36349, 2018.

[50] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J Chem Phys*, 76:637–649, 1982.

[51] T. Veitshans, D. Klimov, and D. Thirumalai. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold Des*, 2:1–22, 1997.

[52] E. Marinari and G. Parisi. Simulated Tempering: a new Monte Carlo scheme. *Europhys Lett*, 19:451, 1992.

[53] P. G. De Gennes. *Scaling concepts in polymer physics*. Cornell university press, first edition, 1979.

[54] Michael Rubinsten. *Polymer physics*. Oxford University Press: Oxford, 2003.

[55] S. Yang, J. N. Onuchic, and H. Levine. Effective stochastic dynamics on a protein folding energy landscape. *J Chem Phys*, 125, 2006.

[56] R. B. Best and G. Hummer. Coordinate-dependent diffusion in protein folding. *Proc Nat Acad Sci of USA*, 107:1088–1093, 2010.

[57] C. J. Stewart, G. I. Olgenblum, A. Propst, D. Harries, and G. J. Pielak. Resolving the enthalpy of protein stabilization by macromolecular crowding. *Protein Sci*, 32:e4573, 2023.

Chapter 6

Simulations of a protein fold switch reveal crowding-induced population shifts driven by disordered regions

Abstract

Macromolecular crowding effects on globular proteins, which usually adopt a single stable fold, have been widely studied. However, little is known about crowding effects on fold-switching proteins, which reversibly switch between distinct folds. Here, we study the mutationally driven switch between the folds of G_A and G_B, the two 56-amino acid binding domains of protein G, using a structure-based dual-basin model. We show that, in the absence of crowders, the fold populations P_A and P_B can be controlled by the strengths of contacts in the two folds, κ_A and κ_B. A population balance, $P_A \approx P_B$, is obtained for $\kappa_B/\kappa_A = 0.92$. The resulting model protein is subject to crowded conditions with different packing fractions, ϕ_c. We find that crowding increases the G_B population and reduces the G_A population, reaching $P_B/P_A \approx 4$ at $\phi_c = 0.44$. We analyze the ϕ_c-dependence of the crowding-induced G_A-to-G_B fold switch using scaled particle theory, which provides a qualitative, but not quantitative, fit of our data, suggesting effects beyond a spherical description of the folds. We show that the terminal regions of the protein chain, which are intrinsically disordered only in G_A, play a dominant role in determining the response of the fold switch to crowding effects.

6.1 Introduction

Most globular proteins rely on a single fold to carry out their function. However, recently, proteins have been discovered with an ability to switch between different folds [1–4], a phenomenon called fold switching. By adopting an alternative structure, these fold-switching proteins (also termed metamorphic [5] or transformer [6] proteins) gain the ability to carry out an additional unrelated function. For example, a switch from a helical hairpin to a β-barrel transforms the *Escherichia coli* protein RfaH from a transcription factor to a translational activator [7]. Consistent with this view, fold switching is often regulated [8]. A range of cellular signals has been associated with fold switching, such as changes in salt concentration [9], redox conditions [10], and oligomerization [11]. Fold switching also underpins evolutionary changes in protein structure [12–14], in which case fold switching is driven by mutations.

In this work, we investigate the effects of macromolecular crowding on fold switching. To this end, we focus on the binding domains of Protein G, G_A and G_B, which form one of the most well-characterized fold switch systems [15] (see Figure 6.1a). It was demonstrated that a set of substitution mutations can be found which drastically increases the sequence identity of G_A and G_B, while still retaining their respective native structures and binding partners [15]. For example, the variants G_A95 and G_B95 differ in only 3 amino acid positions. Hence, three additional substitutions (L20A, I30F and L45Y) applied to G_A95 cause an abrupt switch from the 3α fold of G_A to the $4\beta + \alpha$ fold of G_B. Later it was shown that there are multiple ways in which a single substitution can tip the balance from one fold to the other, e.g., L20A applied

to the variant G_B98-T25I [16]. These experiments on G_A and G_B were, however, carried out in dilute protein solutions and therefore in the absence of any crowding effects.

We carry out our simulations with a coarse-grained structure-based model, which we develop and test on the G_A/G_B fold switch in the absence of crowders (see Methods). The structure-based approach involves constructing a potential energy landscape with a single basin of attraction by making native contacts attractive and non-native contacts repulsive. This type of modeling has provided important insights into several aspects of protein folding [17–20]. The natural extension to fold switching is a potential with dual basins of attractions corresponding to the two alternative folds [21–26]. Our dual-basin model for G_A/G_B fold switching permits us to mimic the progression of mutations along a pathway from one fold to the other by tuning the relative interaction strengths of residue-residue contacts in the G_A and G_B folds (see Figure 6.1b). To understand the effect of crowding, we focus on the point along the mutational pathway where the G_A and G_B folds exhibit roughly equal fold propensities, which we reasoned should be especially susceptible to crowding effects.

6.2 Model and Methods

6.2.1 Native structures and contact maps

The experimentally determined structures of G_A95 (PDB id 2KDL) and G_B95 (2KDM) [15] were downloaded from the Protein Data Bank (PDB). Both structures were submit-

ted to the smog webserver (https://smog-server.org/) to obtain contact maps as prescribed by the shadow map method [27]. The two contact maps contain 106 and 145 contacts, respectively.

6.2.2 Observables

The fractions of native contacts were determined using $Q_{\mathrm{A}} = N_{\mathrm{A}}/106$ and $Q_{\mathrm{B}} = N_{\mathrm{B}}/145$, where N_{A} (N_{B}) is the number of $\mathrm{G_A}$ ($\mathrm{G_B}$) contacts formed. A contact between two amino acids i and j was considered formed if $r_{\mathrm{ij}} < 1.2 r_{\mathrm{ij}}^0$, where r_{ij} is the distance between the C_α atoms and r_{ij}^0 is the distance in the native structure (2KDL or 2KDM). In determining the fold populations, P_{A} and P_{B}, we classified a conformation to be in the $\mathrm{G_A}$ fold if $N_{\mathrm{A}} > N_{\mathrm{A}}^{\mathrm{cut}} = 58$ and in the $\mathrm{G_B}$ fold if $N_{\mathrm{B}} > N_{\mathrm{B}}^{\mathrm{cut}} = 76$, where $N_{\mathrm{A}}^{\mathrm{cut}}$ and $N_{\mathrm{B}}^{\mathrm{cut}}$ were determined based on the free energy profiles $F(N_{\mathrm{A}})$ and $F(N_{\mathrm{B}})$ for $\mathrm{G}_{\mathrm{AB}}^*$ (see Figure C.6). The root-mean-square deviation, RMSD, was calculated over all C_α positions of the chain.

6.2.3 Coarse-grained model for protein fold switching

Simulations were carried out using a dual-basin structure-based model in which each amino acid is represented by a single bead located on the C_α-atom position. The starting point for developing this model was a modified version of the single-basin structure-based model in Ref. [18] with a potential energy function with 5 terms, $E = E_{\mathrm{bond}} + E_{\mathrm{bend}} + E_{\mathrm{torsion}} + E_{\mathrm{rep}} + E_{\mathrm{cont}}$, representing bond stretching, bond flex-

ing, torsional rotations, repulsions between bead pairs, and attractive native contact interactions. We applied this model separately to the native structures of G_A95 and G_B95 resulting in two structure-based energy functions, $E^{(A)}$ and $E^{(B)}$, with single basins of attraction (either the G_A fold or the G_B folds). Using the exponentially-weighted mixing approach of Best et al. [28], we then merged $E^{(A)}$ and $E^{(B)}$ into a single (dual basin) energy function, $E^{(db)}$. The strength of G_A and G_B contacts, κ_A and κ_B, were left as free parameters in $E^{(db)}$, allowing the relative depth of the G_A and G_B basins of attraction to be controlled. Full details of the model are given in Supporting Information.

6.2.4 Excluded volume crowders

Crowder-crowder and crowder-protein interactions are described using the potential function [29]

$$V(r) = \epsilon \left(\frac{\sigma}{r - \rho + \sigma} \right)^{12} \tag{6.1}$$

for distances $r > \rho - \sigma$, and $V(r) = \infty$ otherwise. Hence, our crowders have a soft repulsive shell over a hard core. The parameters ρ and σ control the range of the interaction and the width of the soft repulsive shell, respectively. We determined these parameters using $\sigma = \sigma_i + \sigma_j$ and $\rho = R_i + R_j$, where i and j are two interacting elements. When i, j are crowders we set $\sigma_i = \sigma_j = 3$ Å and $R_i = R_j = 12$ Å, and when one of i, j is a crowder and the other is a chain bead we set for the bead (assuming j) $\sigma_j = R_j = \sigma_b$, where $\sigma_b = 4$ Å is the bead radius. With this choice of ρ and σ, an approximate value for the crowder radius R_c is ≈ 12 Å. A more precise

value was obtained from the radial distribution function $g(r)$ for a large crowder-only system, which indicated 12.5 Å (see Figure C.7). We therefore use $R_c = 12.5$ Å for the crowder radius throughout this work. The crowder concentration, as quantified by the fraction of the total simulation volume V occupied by the crowders, is then given by $\phi_c = 4\pi R_c^3 N_{cr}/3V$. In our simulations, the number of crowder particles N_{cr} ranges from 9 for $\phi_c = 0.073$ to 54 for $\phi_c = 0.442$.

6.2.5 Langevin dynamics

Conformational sampling was carried out using Langevin dynamics following the approach of Ref. [18] (see 2.3.). The time evolution of the system is then governed by the equation, $m\dot{v}(t) = F_{conf} - m\gamma v(t) + \eta(t)$, where m, v, $\dot{v}$, γ, F_{conf} and $\eta(t)$ are the mass, velocity, acceleration, friction coefficient, conformational force and random force, respectively. The random force $\eta(t)$ is drawn from a Gaussian distribution, the variance of which sets the temperature of the system. For computational reasons, simulations were carried out in the low-friction (underdamped) limit, where $-m\gamma v(t)$ is small relative to the inertial term $m\dot{v}(t)$. In this limit, a natural unit of time for the dynamics is $\tau = \sqrt{ml^2/\epsilon}$ [30], where ϵ is the magnitude of typical interactions and l is a length scale, which we set to 4 Å. The friction coefficient for beads was taken to be $\gamma_b = 0.05\tau^{-1}$. Units were set so that the mass of a bead is $m_b = 1.0$. Numerical integration of the equation of motion was carried out using the velocity form of the Verlet algorithm [31] with an integration time step $\delta t = 0.005\tau$. For crowders, the mass and friction coefficient were set to $m_c = 9.0$ and $\gamma_c = 0.017\tau^{-1}$.

6.2.6 Simulation and analysis details

Simulations were carried out by placing the protein and crowders in a cubic box with side 100 Å. Periodic boundary conditions were applied. Langevin dynamics simulations were used to determine the equilibrium behavior of various systems characterized by different G_B contact strengths κ_B and crowder concentrations ϕ_c. Simulations were performed at either fixed temperature or using simulated tempering [32], in which temperature changes dynamically between a predetermined set of values. In the simulated tempering runs, temperatures were updated every 100 time steps. For each system, 5-10 independent runs of $(4-5)\times 10^9$ time steps each were carried out and used to estimate averages and statistical uncertainties. All simulations were initiated from a random protein conformation (random torsional angles ϕ_i) and random crowder positions, followed by a Monte Carlo-based relaxation step in which all hard core steric clashes were removed.

6.2.7 Theory

Simulation results were analyzed using scaled particle theory (SPT) [33] and generalized fundamental measure theory (GFMT) [34, 35]. According to SPT, the free energy cost of inserting a hard sphere of radius R in a hard sphere fluid of particles with radius R_c is [33]

$$\beta F = (3x + 3x^2 + x^3)\psi + (\frac{9x^2}{2} + 3x^3)\psi^2 + 3x^3\psi^3 - \ln(1-\phi_c)\,, \qquad (6.2)$$

where $\beta = 1/k_{\rm B}T$, T is the temperature, $k_{\rm B}$ is the Boltzmann constant, $x = \frac{R}{R_{\rm c}}$, $\psi = \frac{\phi_{\rm c}}{1-\phi_{\rm c}}$, and $\phi_{\rm c}$ is fluid volume fraction. Minton showed that SPT predicts a strong stabilizing effect on the native states of single-fold proteins if the unfolded state size is modeled as an ideal Gaussian chain [36]. Here SPT was used to model the G_A/G_B folds as spheres of different radii.

GFMT accounts for geometric features of the protein structure using three quantities, linear size $l_{\rm p}$, surface area $s_{\rm p}$, and volume $\nu_{\rm p}$. These quantities are obtained by sampling the crowder-excluded surface, which is influenced by the crowder radius, chain bead radius, and the protein conformation. The free energy cost of inserting a protein conformation into a crowder fluid is then estimated as [34, 35]

$$\beta F = \beta\Pi_{\rm c}\nu_{\rm p} + \beta\gamma_{\rm c}s_{\rm p} + \beta\kappa_{\rm c}l_{\rm p} - \ln(1-\phi_{\rm c})\,, \tag{6.3}$$

where $\Pi_{\rm c}$, $\gamma_{\rm c}$ and $\kappa_{\rm c}$ are osmotic pressure, surface tension, and bending rigidity (curvature), respectively, of the crowder fluid. To calculate the change in fold switching free energy at different $\phi_{\rm c}$ using GFMT, we used a freely available code (`http://pipe.rcc.fsu.edu/gfmt/`) and set the radii of crowder and beads to 12.5 and 4.0 Å, respectively. GFMT calculations were performed on 200 random chain conformations for each of the G_A and G_B folds, taken from our simulations of the model protein $G^*_{\rm AB}$ at temperature T_0.

6.3 Results

6.3.1 Mimicking the mutational pathway between the G_A and G_B folds

We first simulate the G_A/G_B system in the absence of crowders at a fixed temperature, T_0, sufficiently low for low-energy folded conformations to dominate over those in the unfolded state (U). By varying the strength κ_B of G_B contacts, keeping the strength of G_A contacts fixed ($\kappa_A = 1$), we can control the relative population of the two folds in our model, as shown in Figure 6.1c. While G_B is the dominant state at high κ_B ($\gtrsim 0.97$) G_A dominates at low κ_B ($\lesssim 0.85$), where there is also a non-zero population of U. At an intermediate value, $\kappa_B = \kappa^* = 0.92$, the populations of G_A (P_A) and G_B (P_B) are almost equal, $P_A \approx P_B \approx 0.39 - 0.42$. The drastic population shifts between states G_A, G_B, and U, can be seen from the free energy surfaces $F(Q_A, Q_B)$, where Q_A and Q_B are the fractions of formed G_A and G_B contacts, respectively, taken at different κ_B values (see Figure 6.1d-f).

The sharp structural transition around $\kappa_B \approx \kappa^*$ is reminiscent of experiments showing that very few mutational steps (or even a single step) is sufficient to tip the balance from G_A to G_B, or vice versa, for carefully selected mutational pathways [15]. Moreover, the minimum in the total folded population $P_{tot} = P_A + P_B$ at $\kappa_B \approx \kappa^*$ (see Figure 6.1a) is in line with the partial loss of stability seen for G_A and G_B sequences close to the transition point, e.g., G_A98 and G_B98, [15] as well as for other fold switching proteins [1, 37]. These results allow us to interpret κ_B as a continuous parameter

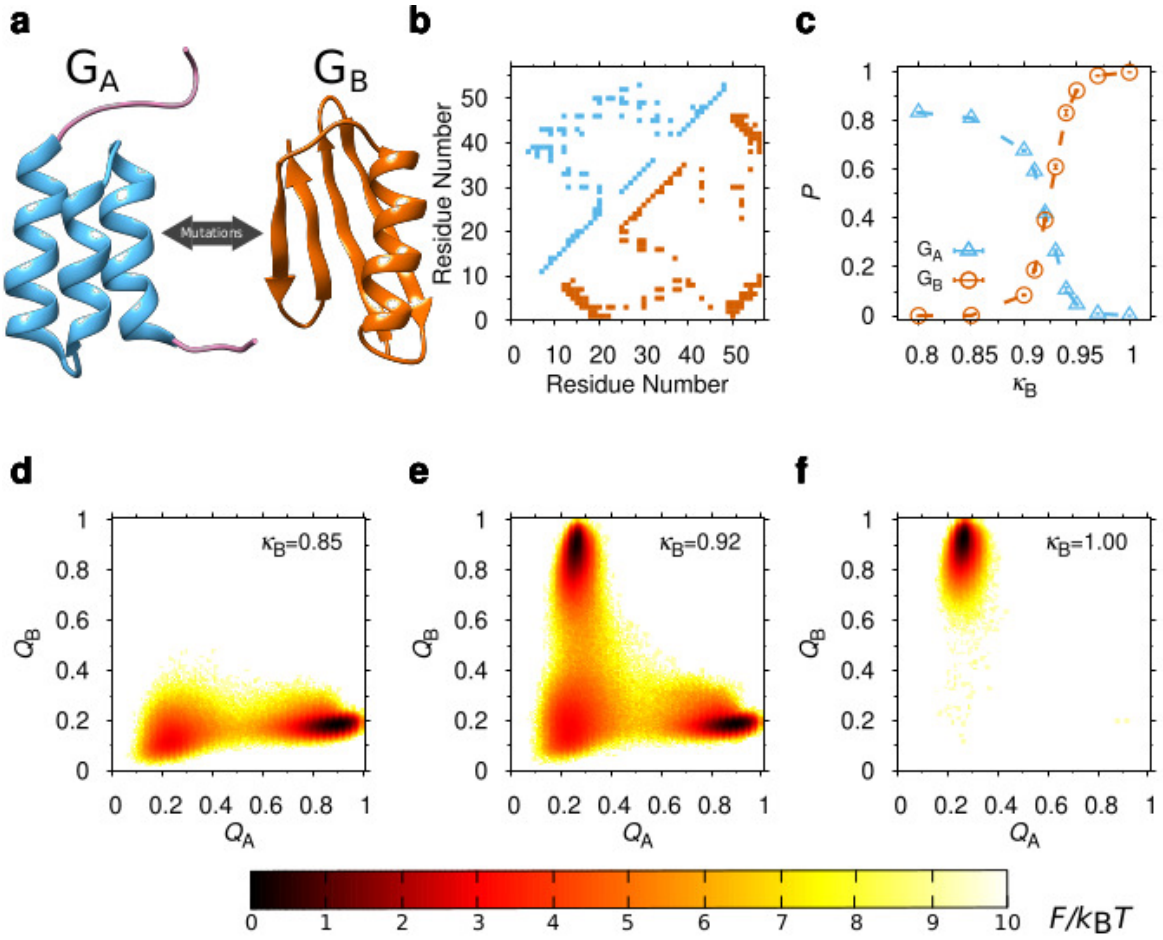

Figure 6.1: (a) Representative experimental structures of the G_A and G_B folds shown in ribbon: G_A95 (PDB id 2KDL; blue) and G_B95 (PDB id 2KDM; orange). In G_A95, residue positions 1-7 and 53-56 are intrinsically disordered (purple). (b) Contact maps of the G_A95 (above diagonal) and G_B95 (below diagonal) structures. (c) Populations of the G_A (triangles) and G_B (circles) folds as functions of the G_B contacts strengths, κ_B. (d-e) Free energy surface $F(Q_A, Q_B) = -k_B T \ln P(Q_A, Q_B)$, where Q_A and Q_B are the fractions of G_A and G_B contacts, respectively, T is the temperature, k_B is the Boltzmann constant, and $P(Q_A, Q_B)$ is a probability distribution, obtained at three different values of κ_B. Results in panels (c)-(f) are taken at temperature T_0 (in model units, $k_B T_0 = 0.88$, where k_B is the Boltzmann constant). Error bars in (c) and other figures unless otherwise stated, represent 1σ standard error of the mean estimated from independent simulations.

mimicking the number of steps taken along a mutational pathway connecting the G_A and G_B folds. The point $\kappa_B = \kappa^*$ thus represents a sequence located on the border between G_A and G_B. Although a sequence with a perfect G_A and G_B population balance was not reported, it has been found for other fold switching systems, e.g., the E48S variant of RfaH [7] and the N11L mutant of the Switch Arc protein [38]. At $\kappa_B = \kappa^*$, $P_{tot} \approx 0.82$ meaning there is a minor population of U under these conditions. It is possible to achieve a higher P_{tot} while maintaining the G_A and G_B population balance by lowering the temperature below T_0 and adjusting κ^* (see Figure C.4). In the following, we focus our analysis on T_0 and refer to our $\kappa_B = \kappa^*$ model protein as G^*_{AB}.

6.3.2 Macromolecular crowding effects on the G_A/G_B fold switch

Next we introduce spherical crowder particles with an effective radius $R_c = 12.5$ Å (see Methods) into our simulation box, thereby probing the effect of volume exclusion on the G_A/G_B switch from objects of roughly the size of the protein chain in either folded state. Because of steric repulsions, the protein chain must at all times avoid the space occupied by the crowders. Such loss of free volume typically stabilizes the native state of single-fold proteins because any extended conformation in U becomes entropically disfavored relative to compact, folded conformations [29]. The same argument can be applied to each fold of a metamorphic protein. Hence, the overall stability of all folded states should increase. Indeed, as shown in Figure 6.2a, the

addition of crowders increases the total population $P_{\rm tot} = P_{\rm A} + P_{\rm B}$ across all values of $\kappa_{\rm B}$. Interestingly, poor stability is a common feature of fold-switching proteins [1]. For example, sequences on either side of the G_A/G_B switch point exhibit reduced stabilities relative to wild-type G_A or G_B [15]. Crowding effects, if indeed providing an overall stabilization, might therefore alleviate the partial loss of stability suffered by bridge sequences in evolutionary fold-switch transitions [39].

To investigate how the relative population of the G_A and G_B folds is affected by crowders we focus on G^*_{AB}. Figure 6.2b shows that, as $\phi_{\rm c}$ increases, the population balance exhibited by G^*_{AB} at $\phi_c = 0$ swings towards G_B at the expense of G_A, i.e., $P_{\rm B}$ increases while $P_{\rm A}$ decreases. The effect on G^*_{AB} is not small. For example, $P_{\rm A}/P_{\rm B} \approx 4$ at $\phi_{\rm c} = 0.44$ as compared to ≈ 1 at $\phi_{\rm c} = 0$. Hence, the effect of steric repulsions between crowders and protein is to favor to G_B over G_A.

To quantitatively analyze this population shift we apply scaled particle theory (SPT) [33]. In this theory, the free energy cost of inserting a hard sphere of radius R into a fluid of hard spheres of radii $R_{\rm c}$ with packing fraction $\phi_{\rm c}$ can be analytically expressed (see Methods). SPT has been used to model crowding-induced changes to the unfolding free energy $\Delta F_{\rm unf} = F_{\rm U} - F_{\rm N}$, where $F_{\rm U}$ and $F_{\rm N}$ are the free energies of U and N, respectively [29, 36, 40]. Here we adapt SPT to fold switching by treating the G_A and G_B folds as spheres of radii $R_{\rm A}$ and $R_{\rm B}$. With the parametrization $R_{\rm A} = R_0 - \delta$ and $R_{\rm B} = R_0 + \delta$, where R_0 and δ are two parameters, the free energy difference can be written

$$\beta\Delta F_{\rm SPT} = 6\left[(a + 2ab + ab^2 + \frac{a^3}{3})\psi + (3ab + 3ab^2 + a^3)\psi^2 + (3ab^2 + a^3)\psi^3\right], \quad (6.4)$$

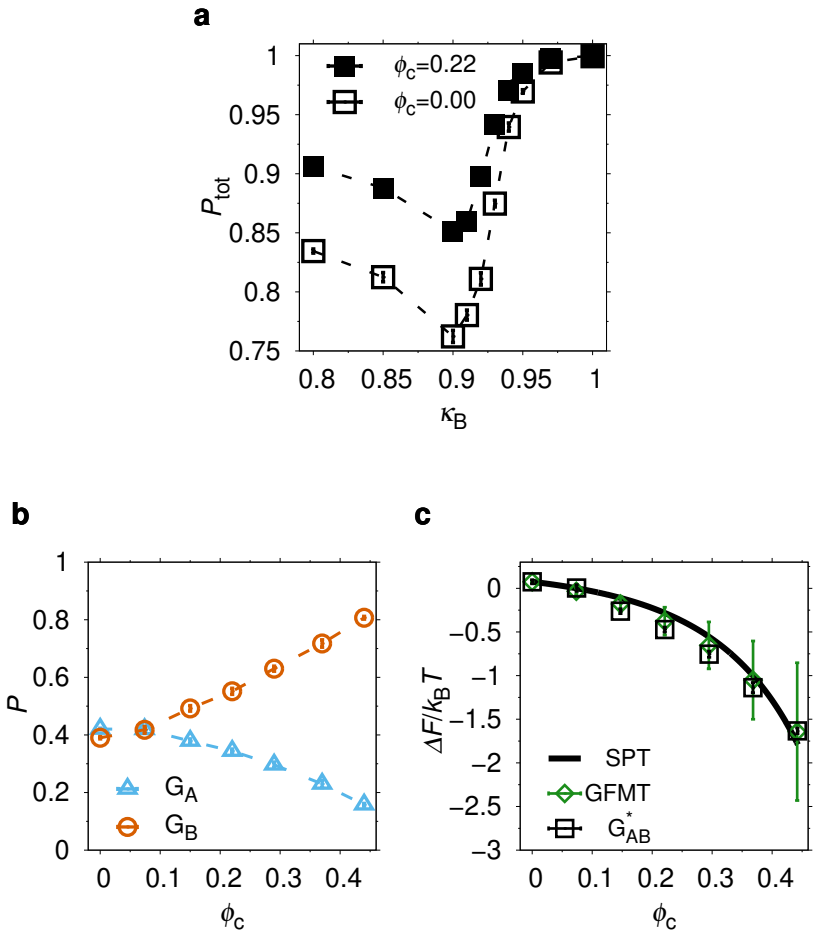

Figure 6.2: (a) The total native population $P_{\rm tot} = P_{\rm A} + P_{\rm B}$ as function of the contact strength $\kappa_{\rm B}$ in the absence (open squares) and presence (filled squares) of crowders at packing fraction $\phi_{\rm c} = 0.22$. (b) G_A (P_A; triangles) and G_B (P_B; circles) fold populations as functions of ϕ_c. (c) Free energy of fold switching $\Delta F_{\rm switch} = -k_{\rm B}T \ln P_{\rm B}/P_{\rm A}$ (squares) as function of ϕ_c, fitted to Eq. 6.2 with δ as a single free parameter (solid curve). Green rhombuses are average $\Delta F_{\rm switch}$ values calculated using GFMT (Eq. 6.3) for a representative set of G_A and G_B conformations taken from simulations, and error bars indicate standard deviations over this set. The temperature is the same as in Figure 1. Dashed lines between points are drawn to guide the eye.

where $a = \delta/R_c$, $b = R_0/R_c$, $\psi = \phi_c/(1-\phi_c)$ and $\beta = 1/k_B T$. We fit the measured crowding induced changes in free energy of fold switching, $\Delta F_{switch} = F_B - F_A = -k_B T \ln[P_B/P_A]$ to Eq. 6.4 with δ as a single free parameter, fixing $R_0 = R_g^{av} + \sigma_b$, where $R_g^{av} = 10.9$ Å is the average radius of gyration of the G_A95 and G_B95 native structures (see Figure 6.1a), and $\sigma_b = 4.0$ Å is the radius of the beads in our protein chain. The fit is shown in Figure 6.2c and gives $\delta = -0.35 \pm 0.03$ Å. The size difference 2δ is in rough agreement with that calculated for the radii of gyration of the G_A95 and G_B95 native structures, $R_g^A = 11.4$ Å and $R_g^B = 10.5$ Å. The quality of the fit ($\chi^2/(n-1) = 22$, sample size $n = 7$) indicates, however, that SPT poorly describes the observed crowding effects on ΔF_{switch}. A better agreement can be obtained by applying the generalized fundamental measure theory (GFMT) of Qin and Zhou [34, 35], which takes into account both the shape of protein conformations and their fluctuations (see Figure 2c). The generally good agreement obtained for GFMT indicates, in particular, that accounting for fluctuations in chain size, which are absent in SPT, is necessary to describe the observed ϕ_c-dependence of ΔF_{switch}.

6.3.3 Disordered tails control the crowding effect on the fold switch

The two terminal segments of the G_A95 structure, residues 1-7 and 53-56, are intrinsically disordered (see Figure 6.1a). Hence, the G_A-to-G_B fold switch involves a disorder-order transition of these tail regions. Given their flexible nature, it is likely that the tails contribute substantially to the volume excluded by the protein when

occupying the G_A fold. Indeed, if the terminal segments are ignored, the radius of gyration of G_A95 is reduced by $\approx$ 22 %, $R_g^{A,8\text{-}52} = 8.9$ Å. By contrast, the radius of gyration of G_B95 determined over the same segment is $R_g^{B,8\text{-}52} = 10.8$ Å, which is a slight increase compared to value for the full chain. Together with the poor fit with SPT (see Figure 6.2c), these results suggest a potential role for the tail segments in how the G_A/G_B fold switch is impacted by crowding.

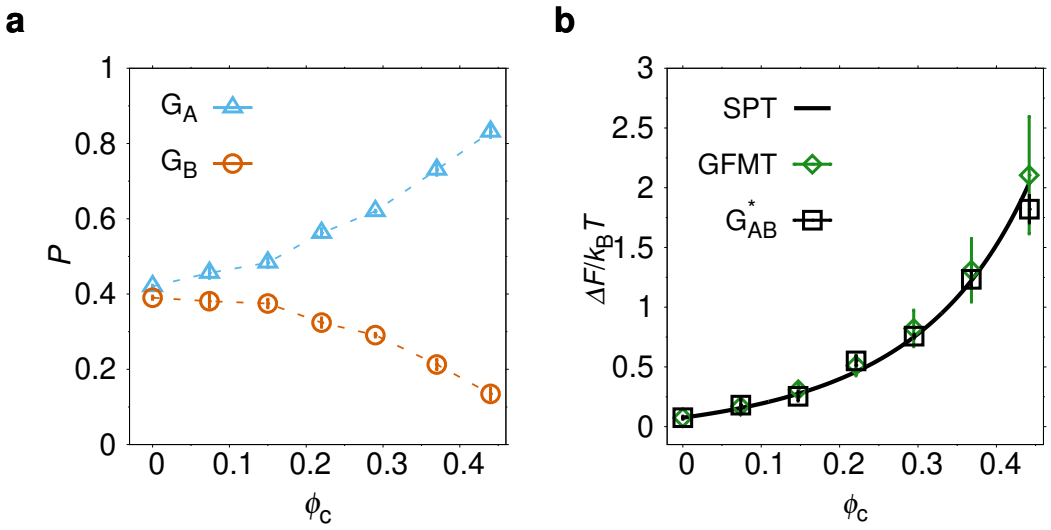

Figure 6.3: Results are shown for simulations with a modified potential energy function that ignores hard-core steric repulsions between any crowder and beads in chain segments 1-7 and 53-56 (see text). (a) G_A (triangles) and G_B (circles) fold populations as function of ϕ_c. (b) Fit of ΔF_{switch} (squares) to scaled particle theory (solid curve). Green rhombuses are ΔF_{switch} values calculated as in Figure 6.2 but with GFMT applied only to the chain segment 8-52. The temperature is the same as in Figure 1.

To show that this is indeed the case, we carry out crowding simulations with a modified potential energy function, $E_{mod}^{(db)}$, in which all crowder-protein interactions

have been turned off for residues in the 1-7 and 53-56 regions. Hence, in these simulations, the N- and C-terminal segments become invisible to the crowders, which thus freely overlap with the residues. Although unphysical, this computational experiment logically tests the role of the tail regions in our model under crowded conditions. Note that crowders can overlap with the tails regardless of which state is populated by the protein. Moreover, at $\phi_c = 0$, the model remains the same because intra-chain interactions are unaffected. The results are shown in Figure 6.3. Strikingly, with the modified potential $E_{\rm mod}^{\rm (db)}$, the impact of crowding reverses such that the G_A fold becomes increasingly favored over G_B with increasing ϕ_c. Moreover, the fit to the SPT (see Fig 6.3b), obtained using $R_0 = 13.9$Å and giving $\delta = 0.42 \pm 0.01$Å, is now much better ($\chi^2/(n-1) \approx 1.5$). We also carried out simulations with the modifications in $E_{\rm mod}^{\rm (db)}$ applied separately to the 1-7 and 53-56 regions. As it turns out, the effects on $\Delta F_{\rm switch}$ is roughly additive (see Figure C.5), suggesting that the two tail regions independently reduce the volume available to the crowders. Taken together, our computational experiment shows that the volume excluded by the disordered tails in the G_A fold is the dominant factor affecting the balance between the folds in the presence of crowders.

6.3.4 Comparing with crowding effects on single-fold proteins

Above we have shown that the crowders induce a population shift in G_{AB}^*, which is due to the presence of disordered tails. For single-fold (monomorphic) proteins, purely

repulsive crowders typically enhance the stability of the native state [41]. Naively, one may therefore expect that the native state of monomorphic G_B ($\kappa_B > \kappa^*$) would be more strongly stabilized by the crowders than monomorphic G_A ($\kappa_B < \kappa^*$). To test this idea, we determine the folding midpoint temperature, T_m, for the model proteins with $\kappa_B = 0.85$, which adopts the single fold G_A, and $\kappa_B = 1.00$, which adopts the single fold G_B (see Figure 6.1), over a range of ϕ_c. As seen in Figure 6.4a-c, both proteins exhibit a monotonic increase in T_m with increasing ϕ_c, indicating stabilization. The relative increase in T_m for monomorphic G_B is indeed somewhat larger than for monomorphic G_A. The difference is relatively small, however. We also perform similar simulations using the single-basin energy functions $E^{(A)}$ and $E^{(B)}$, (see Methods) which lack entirely a bias towards the alternative fold. For these models, the crowding-induced increases in T_m are almost identical (see Figure 6.4d). Taken together, these results suggest that determining the crowding response of a fold switcher with two "co-existing" folds may not be easily obtained from experiments on single-fold proteins representative of the two different folds.

6.3.5 The unfolded state changes character across the fold switch

The results in Figure 6.4 are at first surprising because ΔF_{switch} for a fold switching protein can be obtained from the relation

$$\Delta F_{switch} = \Delta F^{A}_{unf} - \Delta F^{B}_{unf}\,, \tag{6.5}$$

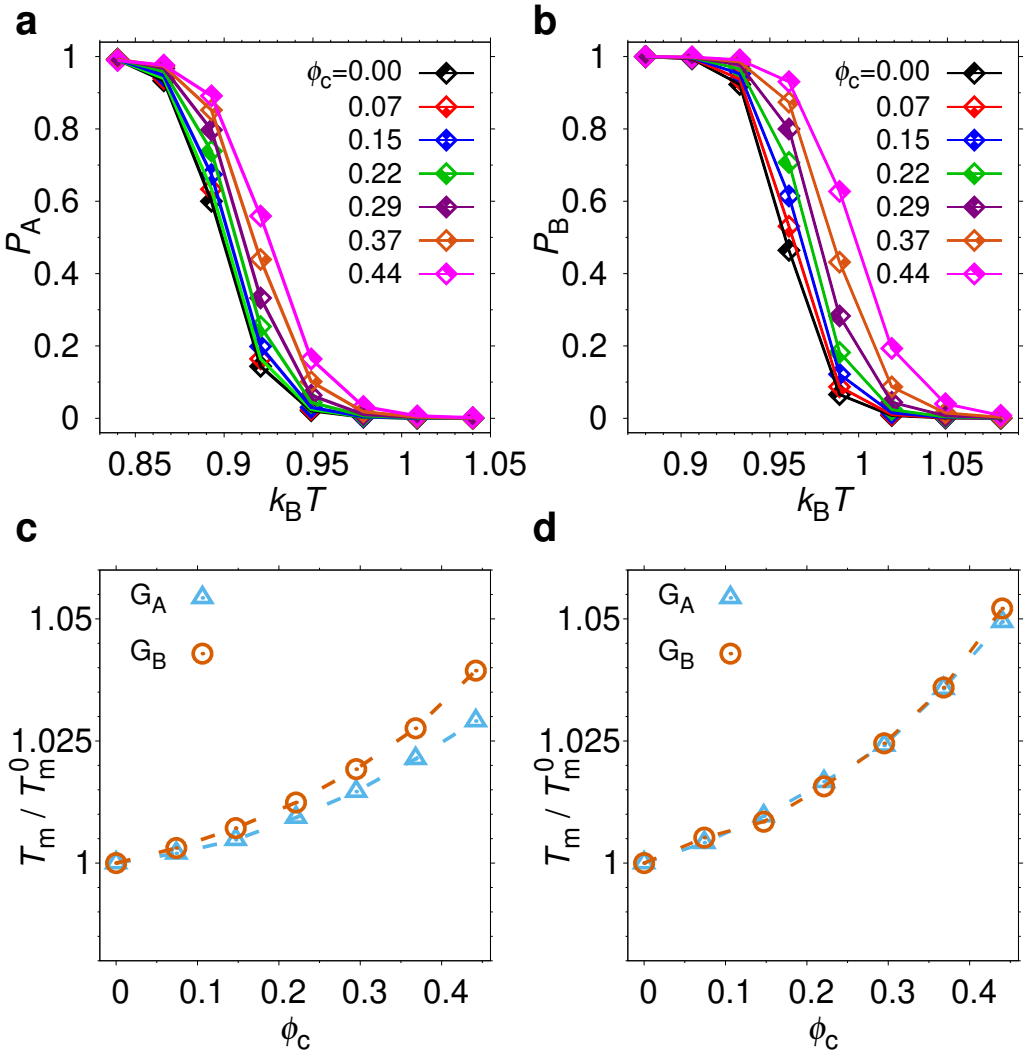

Figure 6.4: (a) G_A fold population obtained with our dual-basin structure-based model with weak G_B contacts ($\kappa_B = 0.85$), as function of temperature. (b) G_B fold population obtained with the same model but with strong G_B contacts ($\kappa_B = 1.00$). (c) Midpoint temperature, T_m, as function of ϕ_c. T_m is obtained by fitting the folding curves in (a) and (b) to a two-state model. (d) T_m as function of ϕ_c, obtained with single-basin structure-based models for G_A and G_B. In both (c) and (d), T_m^0 is the value of T_m at $\phi_c = 0$.

where $\Delta F^{\mathrm{A}}_{\mathrm{unf}} = F_{\mathrm{U}} - F_{\mathrm{A}}$ and $\Delta F^{\mathrm{B}}_{\mathrm{unf}} = F_{\mathrm{U}} - F_{\mathrm{B}}$ are defined in direct analogy with the unfolding free energy of a single fold protein. Equation 6.5 expresses that a decrease in $\Delta F_{\mathrm{switch}}$ results when the crowding-induced stabilization of fold $\mathrm{G_B}$ relative to U is stronger than the stabilization of fold $\mathrm{G_A}$. However, Eq. 6.5 is only guaranteed to hold when $\Delta F_{\mathrm{switch}}$, $\Delta F^{\mathrm{A}}_{\mathrm{unf}}$ and $\Delta F^{\mathrm{B}}_{\mathrm{unf}}$ are determined for the same protein for which U provides a common reference. We therefore examine if the drastic structural shift for low energy (folded) conformations in the $\mathrm{G_A}/\mathrm{G_B}$ fold switch is accompanied by changes in U.

We first characterize U across the fold switch in the absence of crowders, i.e., upon changing the contact strength κ_{B}, as shown in Figure 6.5a and b. With increasing κ_{B}, and therefore increasing $\mathrm{G_B}$ population, the unfolded state radius of gyration $R^{\mathrm{U}}_{\mathrm{g}}$ decreases. Additionally, U becomes more "$\mathrm{G_B}$-like" as shown by the increase in $Q^{(\mathrm{U})}_{\mathrm{B}}$, i.e., the fraction of formed $\mathrm{G_B}$ contacts in U. These results are in line with simulations of single-fold proteins showing that native contacts in β-proteins tend to promote chain collapse during folding more efficiently than α-proteins [42]. In the crowding-induced $\mathrm{G_A}$-to-$\mathrm{G_B}$ fold switch we similarly find a compaction of U (see Figure 6.5c and d). For $\phi_{\mathrm{c}} > 0.20$, $R^{\mathrm{U}}_{\mathrm{g}}$ becomes smaller than for any value of κ_{B} in the case of no crowders. Moreover, $Q^{(\mathrm{U})}_{\mathrm{A}}$ and $Q^{(\mathrm{U})}_{\mathrm{B}}$ both increase with ϕ_{c}. In summary, fold switching driven either by mutation or crowding substantially impacts the structural characteristics of U. Both chain compaction and the formation of residual structure due to crowding have been observed for various single-fold proteins [43, 44].

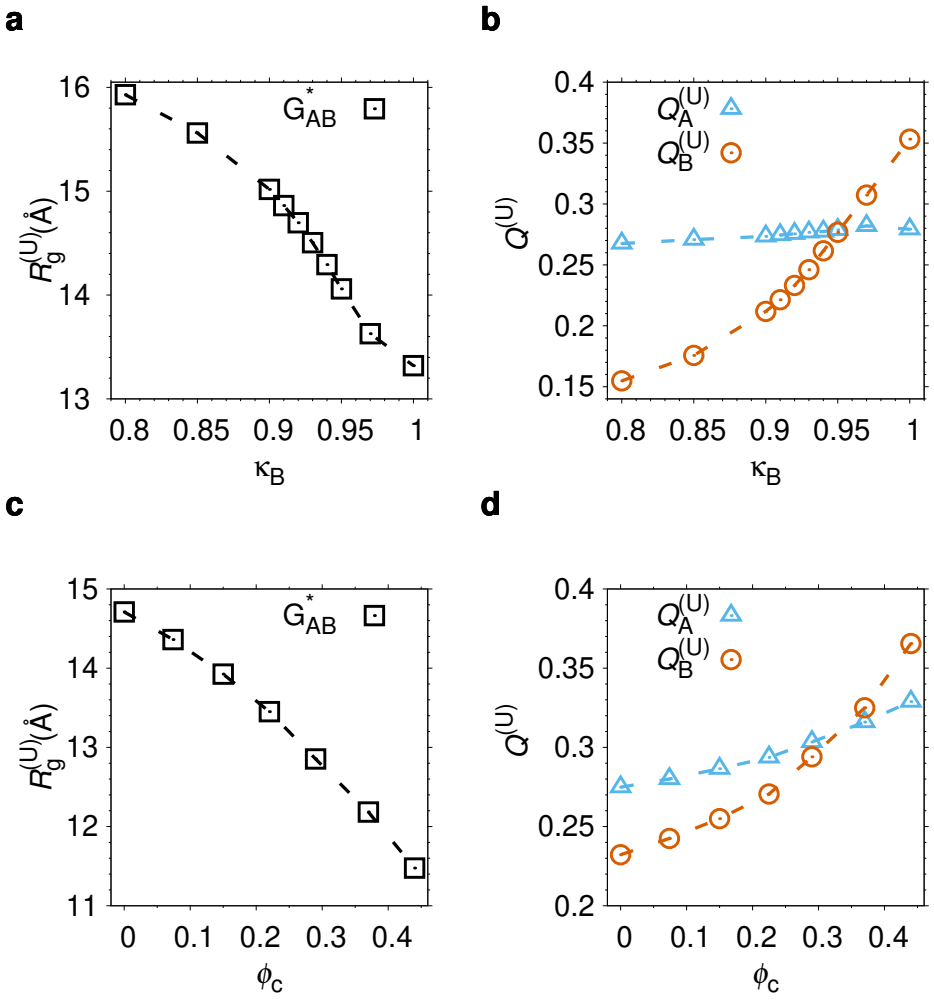

Figure 6.5: Changes to the unfolded state character across the fold switch. (a) $R_{\mathrm{g}}^{\mathrm{U}}$, (b) $Q_{\mathrm{A}}^{(\mathrm{U})}$ (triangles) and $Q_{\mathrm{B}}^{(\mathrm{U})}$ (circles) as functions of the contact strength κ_{B} at $\phi_{\mathrm{c}} = 0$, where $R_{\mathrm{g}}^{\mathrm{U}}$, $Q_{\mathrm{A}}^{(\mathrm{U})}$, and $Q_{\mathrm{B}}^{(\mathrm{U})}$ are the radius of gyration, fraction of $\mathrm{G_A}$ contacts, and fraction of $\mathrm{G_B}$ contacts, respectively, determined for the unfolded state, U. (c) $R_{\mathrm{g}}^{\mathrm{U}}$, (d) $Q_{\mathrm{A}}^{(\mathrm{U})}$ (triangles) and $Q_{\mathrm{B}}^{(\mathrm{U})}$ (circles) as functions of ϕ_{c}, obtained for $\mathrm{G}_{\mathrm{AB}}^{*}$ ($\kappa_{\mathrm{B}} = 0.92$). The temperature is the same as in Figure 1.

6.4 Discussion

Fold switching in proteins involves major structural changes, including in shape and amino acid composition of surface regions. As a result, fold switching should be inherently susceptible to crowding effects. Here we tested this idea by applying a dual-basin structure-based protein model and purely repulsive crowders to the G_A/G_B fold switch. We found that the addition of crowders indeed alters the free energy balance between the two folds. The effect increases monotonically with ϕ_c. At $\phi_c = 0.44$, the change in ΔF_{switch} is $\approx 2\ k_B T$ in magnitude. While no experiment probing crowding effects on the G_A/G_B fold switch is available for comparison, a key role for molecular shape in crowding has been demonstrated in a study that exploited alternative dimer forms of two almost identical sequences [45]. Very recently, it was shown using nuclear magnetic resonance spectroscopy that the addition of 90 g/L Ficoll400, polyethylene glycol (PEG10) or BSA to the solution impacted the relative fold population of the two metamorphic proteins KaiB and XCL1 [46].

Our results show that crowding effects on the G_A/G_B system may be determined by chain segments at the N- and C-terminal ends, which are intrinsically disordered only in the G_A fold. The volume excluded by these disordered segments leads to an entropic stabilization of G_B relative to G_A. Interestingly, order-disorder transitions occur frequently in protein fold switching [1]. One example besides G_A/G_B is human chemokine XCL1, which switches folds upon dimerization. In its monomeric (chemokine) fold, XCL1 adopts an α-helix in its C-terminal region, which becomes disordered when the protein transforms to its dimeric fold-switched state [12]. It

should be pointed out that crowder interactions other than hard-core steric repulsions can modify the crowding effects. For single-fold proteins, non-specific attractive (soft) interactions between protein and crowders generally counteract the stabilizing effect of volume exclusion [47], and can even lead to a net destabilization [48].

Most studies on fold switching have quite naturally focused on the structure and dynamics of the different folded states, and their interconversions. However, our simulations of the G_A/G_B switch reveal that fold switching may be accompanied by substantial changes in U (see Figure 6.5). Under conditions favoring G_A, we find that U is rather expanded and dominated by local contacts while becoming more compact and forming more non-local contacts as the conditions shift to favor G_B. In previous simulations of the metamorphic RfaH [25], we showed that the isolated C-terminal domain (CTD), which adopts a stable β-barrel in isolation, exhibits a propensity for α-helical structure in U. This helical propensity was demonstrated experimentally by Zuber et al. [37], who suggested further that the presence of residual helical structure may help initiate the reverse fold switching of RfaH, i.e., the transformation from the β-barrel to its alternative all-α fold. Taken together, the above considerations suggest that an improved understanding of U may give further insights into fold switching mechanisms, as well as effects from crowding.

In addition to changes to the relative population of the two folds, we have found that the presence of crowders increases the total population of the G_A and G_B folds relative to U. An overall stabilization of ordered states might be especially beneficial to fold-switching proteins, which often exhibit reduced stabilities [1]. Poor stabilities of bridge sequences at the border between folds may hamper evolutionary transitions [16,

49, 50]. A recent study suggests fold switching within the context of multidomain proteins, in which non-switching domains can act as stabilizing scaffolds, may help stabilize such bridge sequences and facilitate fold transitions [13]. Our results suggest that additional stabilization may be provided by crowding effects.

Our study opens up for additional experimental and theoretical investigations into the effects of crowding on fold switching. Recent advances in the fold switching field are improving our understanding of this phenomenon within functional [37, 51] and evolutionary [3, 12, 13] contexts. These efforts should also include a characterization of the impact of macromolecular crowding on equilibrium and kinetic properties of fold switching proteins.

6.5 Acknowledgments

SW acknowledges support from the Natural Sciences and Engineering Research Council of Canada (grant RGPIN-2016-05104). This research was enabled in part by the computational resources provided by the Digital Research Alliance of Canada.

6.6 Authorship contributions

SW designed the study. SB, BS and SW carried out the simulations, performed the analyses, and wrote the paper.

6.7 Code availability

Simulations were carried out using in-house software written in C. The software will be made available for academic use upon request to the corresponding author.

Bibliography

[1] P. N. Bryan and J. Orban. Proteins that switch folds. *Curr Opin Struct Biol*, 20:482–488, 2010.

[2] A. F. Dishman and B. F. Volkman. Unfolding the mysteries of protein metamorphosis. *ACS Chem Biol*, 13:1438–1446, 2018.

[3] A. K. Kim and L. L. Porter. Functional and regulatory roles of fold-switching proteins. *Structure*, 29:6–14, 2021.

[4] I. Artsimovitch and C. A. Ramírez-Sarmiento. Metamorphic proteins under a computational microscope: Lessons from a fold-switching RfaH protein. *Comput Struct Biotechnol J*, 20:5824–5837, 2022.

[5] A. G. Murzin. Biochemistry. Metamorphic proteins. *Science*, 320(5884):1725–1726, Jun 2008.

[6] S. H. Knauer, I. Artsimovitch, and P. Rösch. Transformer proteins. *Cell Cycle*, 11:4289–4290, 2012.

[7] B. M. Burmann, S. H. Knauer, A. Sevostyanova, K. Schweimer, R. A. Mooney, R. Landick, I. Artsimovitch, and P. Rösch. An α helix to β barrel domain

switch transforms the transcription factor RfaH into a translation factor. *Cell*, 150:291–303, 2012.

[8] L. L. Porter and L. L. Looger. Extant fold-switching proteins are widespread. *Proc Natl Acad Sci USA*, 115:5968–5973, 2018.

[9] R. L. Tuinstra, F. C. Peterson, S. Kutlesa, E. S. Elgin, M. A. Kron, and B. F. Volkman. Interconversion between two unrelated protein folds in the lymphotactin native state. *Proc Natl Acad Sci USA*, 105:5057–5062, 2008.

[10] D. R. Littler, S. J. Harrop, W. D. Fairlie, L. J. Brown, G. J. Pankhurst, S. Pankhurst, M. Z. DeMaere, T. J. Campbell, A. R. Bauskin, R. Tonini, M. Mazzanti, S. N. Breit, and P. M. Curmi. The intracellular chloride ion channel protein CLIC1 undergoes a redox-controlled structural transition. *J Biol Chem*, 279:9298–9305, 2004.

[11] Y. G. Chang, S. E. Cohen, C. Phong, W. K. Myers, Y. I. Kim, R. Tseng, J. Lin, L. Zhang, J. S. Boyd, Y. Lee, S. Kang, D. Lee, S. Li, R. D. Britt, M. J. Rust, S. S. Golden, and A. LiWang. A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. *Science*, 349:324–328, 2015.

[12] A. F. Dishman, R. C. Tyler, J. C. Fox, A. B. Kleist, K. E. Prehoda, M. M. Babu, F. C. Peterson, and B. F. Volkman. Evolution of fold switching in a metamorphic protein. *Science*, 371:86–90, 2021.

[13] B. Ruan, Y. He, Y. Chen, E. J. Choi, Y. Chen, D. Motabar, T. Solomon, R. Simmerman, T. Kauffman, D. T. Gallagher, J. Orban, and P. N. Bryan. Design and characterization of a protein fold switching network. *Nat Commun*, 14:431, 2023.

[14] I. Yadid, N. Kirshenbaum, M. Sharon, O. Dym, and D. S. Tawfik. Metamorphic proteins mediate evolutionary transitions of structure. *Proc Natl Acad Sci USA*, 107:7287–7292, 2010.

[15] P. A. Alexander, Y. He, Y. Chen, J. Orban, and P. N. Bryan. A minimal sequence code for switching protein structure and function. *Proc Natl Acad Sci USA*, 106:21149–21154, 2009.

[16] Y. He, Y. Chen, P. A. Alexander, P. N. Bryan, and J. Orban. Mutational tipping points for switching protein folds and functions. *Structure*, 20:283–291, 2012.

[17] L. L. Chavez, J. N. Onuchic, and C. Clementi. Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J Am Chem Soc*, 126:8426–8432, Jul 2004.

[18] S. Wallin and H. S. Chan. Conformational entropic barriers in topology-dependent protein folding: perspectives from a simple native-centric polymer model. *J Phys Condens Matter*, 18:S307, 2006.

[19] S. G. Estáciocio, C. S. Fernandes, H. Krobath, P. F. Faísca, and E. I. Shakhnovich. Robustness of atomistic Gō models in predicting native-like folding intermediates. *J Chem Phys*, 137:085102, 2012.

[20] A. Kluber, T. A. Burt, and C. Clementi. Size and topology modulate the effects of frustration in protein folding. *Proc Natl Acad Sci USA*, 115:9234–9239, 09 2018.

[21] M. Kouza and U. H. Hansmann. Folding simulations of the A and B domains of protein G. *J Phys Chem B*, 116:6645–6653, 2012.

[22] L. Sutto and C. Camilloni. From A to B: a ride in the free energy surfaces of protein G domains suggests how new folds arise. *J Chem Phys*, 136:185101, 2012.

[23] C. A. Ramirez-Sarmiento, J. K. Noel, S. L. Valenzuela, and I. Artsimovitch. Interdomain contacts control native state switching of RfaH on a dual-funneled landscape. *PLOS Comput Biol*, 11:e1004379, 2015.

[24] L. Xiong and Z. Liu. Molecular dynamics study on folding and allostery in RfaH. *Proteins*, 83:1582–1592, 2015.

[25] B. Seifi and S. Wallin. The C-terminal domain of transcription factor RfaH: Folding, fold switching and energy landscape. *Biopolymers*, 112:e23420, 2021.

[26] P. Galaz-Davison, E. A. n, and C. A. rez Sarmiento. The N-terminal domain of RfaH plays an active role in protein fold-switching. *PLOS Comput Biol*, 17:e1008882, 2021.

[27] J. K. Noel, M. Levi, M. Raghunathan, H. Lammert, R. L. Hayes, J. N. Onuchic, and P. C. Whitford. SMOG 2: a versatile software package for generating structure-based models. *PLOS Comput Biol*, 12:e1004794, 2016.

[28] R. B. Best, Y. G. Chen, and G. Hummer. Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor. *Structure*, 13:1755–1763, 2005.

[29] J. Mittal and R. B. Best. Dependence of protein folding stability and dynamics on the density and composition of macromolecular crowders. *Biophys J*, 98:315–320, 2010.

[30] T. Veitshans, D. Klimov, and D. Thirumalai. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold Des*, 2:1–22, 1997.

[31] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J Chem Phys*, 76:637–649, 1982.

[32] E. Marinari and G. Parisi. Simulated Tempering: a new Monte Carlo scheme. *Europhys Lett*, 19:451, 1992.

[33] J. L. Lebowitz and J. S. Rowlinson. Thermodynamic properties of mixtures of hard spheres. *J Chem Phys*, 41:133–138, 1964.

[34] S. Qin and H. X. Zhou. Generalized fundamental measure theory for atomistic modeling of macromolecular crowding. *Phys Rev E*, 81:031919, 2010.

[35] S. Qin, J. Mittal, and H. X. Zhou. Folding free energy surfaces of three small proteins under crowding: validation of the postprocessing method by direct simulation. *Phys Biol*, 10:045001, 2013.

[36] A. P. Minton. Models for excluded volume interaction between an unfolded pro-

tein and rigid macromolecular cosolutes: macromolecular crowding and protein stability revisited. *Biophys J*, 88:971–985, 2005.

[37] P. K. Zuber, T. Daviter, R. mann, U. Persau, K. Schweimer, and S. H. Knauer. Structural and thermodynamic analyses of the β-to-α transformation in RfaH reveal principles of fold-switching proteins. *Elife*, 11:e76630, 2022.

[38] M. H. Cordes, R. E. Burton, N. P. Walsh, C. J. McKnight, and R. T. Sauer. An evolutionary bridge to a new protein fold. *Nat Struct Biol*, 7:1129–1132, 2000.

[39] T. Sikosek, E. Bornberg-Bauer, and H. S. Chan. Evolutionary dynamics on protein bi-stability landscapes can potentially resolve adaptive conflicts. *PLOS Comput Biol*, 8:e1002659, 2012.

[40] H. X. Zhou. Protein folding in confined and crowded environments. *Arch Biochem Biophys*, 469:76–82, 2008.

[41] A. P. Minton. The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences. *Mol Cell Biochem*, 55:119–140, 1983.

[42] H. S. Samanta, P. I. Zhuravlev, M. Hinczewski, N. Hori, S. Chakrabarti, and D. Thirumalai. Protein collapse is encoded in the folded state architecture. *Soft Matt*, 13:3622–3638, 2017.

[43] J. Hong and L. M. Gierasch. Macromolecular crowding remodels the energy landscape of a protein by favoring a more compact unfolded state. *J Am Chem Soc*, 132:10445–10452, 2010.

[44] T. Mikaelsson, J. Adén, L. B. Johansson, and P. Wittung-Stafshede. Direct observation of protein unfolded state compaction in the presence of macromolecular crowding. *Biophys J*, 104:694–704, 2013.

[45] A. J. Guseman, G. M. P. Goncalves, S. L. Speer, G. B. Young, and G. J. Pielak. Protein shape modulates crowding effects. *Proc Natl Acad Sci USA*, 115:10965–10970, 2018.

[46] N. Zhang, W. Guan, S. Cui, and N. Ai. Crowded environments tune the fold-switching in metamorphic proteins. *Commun Chem*, 6:117, 2023.

[47] M. Sarkar, C. Li, and G. J. Pielak. Soft interactions and crowding. *Biophys Rev*, 5:187–194, 2013.

[48] H. X. Zhou. Polymer crowders and protein crowders act similarly on protein folding stability. *FEBS Lett*, 587:394–397, 2013.

[49] C. Holzgräfe and S. Wallin. Smooth functional transition along a mutational pathway with an abrupt protein fold switch. *Biophys J*, 107:1217–1225, 2014.

[50] C. Holzgräfe and S. Wallin. Local versus global fold switching in protein evolution: insight from a three-letter continuous model. *Phys Biol*, 12:026002, 2015.

[51] Y. G. Chang, S. E. Cohen, C. Phong, W. K. Myers, Y. I. Kim, R. Tseng, J. Lin, L. Zhang, J. S. Boyd, Y. Lee, S. Kang, D. Lee, S. Li, R. D. Britt, M. J. Rust, S. S. Golden, and A. LiWang. Circadian rhythms. A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. *Science*, 349:324–328, 2015.

Chapter 7

Summary and Outlook

In this thesis, our goal has been to investigate through theory and computational simulations of coarse-grained models, the impact of macromolecular crowding effects on protein folding, stability, and fold switching. Perhaps surprisingly, the central issue in the crowding field of how the stability of proteins is impacted by various types of crowder interactions, is not yet fully understood. Our focus has mainly been to study the balance between the excluded volume effect and weak chemical attractions on equilibrium and kinetic properties of proteins, such as native state stability and folding rates. To study macromolecular crowding effects on fold-switching proteins, we focused on the G_A/G_B system, which is one of the best characterized systems in fold switching.

To study crowding effects on protein native state stability, we developed a sequence-based model for proteins in which folding is driven by effective hydrophobic and hy-

drogen bond interactions. Crowders were mimicked as spheres. One of the most widespread conceptions in the crowding field is that excluded volume crowders generally enhance the native state stability of proteins. However, depending on the volume fraction of crowders and temperature, we found that volume exclusion can have a neutral effect on the native state stability, or even lead to a net destabilization. Specifically, we found this behavior for two model proteins with only β-structure, which both exhibit highly compact nonnative conformations, especially at low T.

We also used the same model to study the impact of different kinds of soft interactions, especially hydrophobic interactions and hydrogen bonding, for a helical protein that is stabilized by the excluded volume effect. As expected, both types of soft interactions counteract the stabilizing excluded volume effect. We also found that crowders with hydrophobic character provide a stronger destabilization relative to crowders that interact nonspecifically with the protein. Additionally, we have shown that hydrophobic peptide chains with partial hydrophobic character demonstrate much stronger destabilizing effects than peptides with only polar amino acids.

Using a structure-based C_α model, we showed that polymeric crowders stabilize the native state of proteins. Unlike the C_β model, there are no attractive nonnative interactions to stabilize compact nonnative conformations in the C_α model. Hence, proteins are likely universally stabilized by excluded volume effects in these types of structure-based models, as observed by us and others. Two proteins 1PRB and 1SHF were investigated under both polymeric and spherical crowding conditions across different temperatures and crowder concentrations. We have shown that polymeric crowders with varying chain lengths, but with the total volume fraction of crowders

kept constant, have approximately the same impact on stability and folding kinetics. However, the enhancement of both the stability and the folding rate found upon the addition of spherical crowders, decreased with increasing size of spherical crowders for a given fixed total crowder volume fraction.

Using a dual-basin structure-based potential applied to the G_A/G_B fold switch system, we study macromolecular crowding effects on protein G. We found that increasing the volume fraction of crowders enhances the total stability of the native states and also shifts the folded population towards the G_B native state. Our analysis showed that it is the presence of intrinsically disordered tails, which only appear in the G_A state, that drives the population shift. It would be extremely interesting to compare our results with an experimental study of G_A/G_B sequences close to the fold switch point in the presence of various types of artificial and protein crowders.

The effect of macromolecular crowding on protein stability has been studied by many groups [1–3]. Generally, it was expected under excluded volume effects of crowders, the native state of protein is entropically favored relative to the unfolded state [4–6]. However, as we have shown here, it might not always be the case. This idea might be tested by crowding experiments on proteins with populated compact nonnative conformations [7–9]. One challenge with such experiments is that at low Ts, which is where we observe destabilization, the population of the unfolded state is very small. The unfolded state signal may therefore be hard to distinguish from the native state signal in bulk experiments. However, single-molecule experiments, in which populations in individual states in a single protein can be observed, might be able to address this issue. Single molecule experiments with the artificial crowder PEG were

recently carried out by Schuler et al. on intrinsically disordered proteins [10].

To further our research in protein fold-switching, we continue to explore the effects of macromolecular crowding by applying our model to the fold-switching protein RfaH. We also plan to investigate the fold-switching kinetics and effect of attractive interactions between protein and crowders. A remaining question to explore is how the presence of attractive interactions between the disordered tails of the G_A fold, and the crowder, will impact the fold switch. For these investigations, we will be able to utilize our coarse-grained structure-based C_α model.

Particularly, our C_α simulation method is general in the sense that it can be applied to other proteins, including other fold-switching proteins, as long as structures of the alternative folds are available. Furthermore, we can conduct fold-switching tests in a more realistic crowded environment using different crowder particles, such as crowders of other proteins, which would be particularly interesting.

Bibliography

[1] B. van den Berg, R. J. Ellis, and C. M. Dobson. Effects of macromolecular crowding on protein folding and aggregation. *The EMBO J*, 18:6927–6933, 1999.

[2] A. P. Minton. Implications of macromolecular crowding for protein assembly. *Curr Opin Struct Biol*, 10:34–39, 2000.

[3] N. Tokuriki, M. Kinjo, S. Negi, M. Hoshino, Y. Goto, I. Urabe, and T. Yomo. Protein folding by the effects of macromolecular crowding. *Protein Sci*, 13:125–133, 2004.

[4] L. Stagg, S. Zhang, M. S. Cheung, and P. Wittung-Stafshede. Molecular crowding enhances native structure and stability of α/β protein flavodoxin. *Proc Natl Acad Sci USA*, 104(48):18976–18981, 2007.

[5] M. Perham, L. Stagg, and P. Wittung-Stafshede. Macromolecular crowding increases structural content of folded proteins. *FEBS letters*, 581:5065–5069, 2007.

[6] A. P. Minton. Influence of macromolecular crowding upon the stability and state of association of proteins: predictions and observations. *J Pharm Sci*, 94:1668–1675, 2005.

[7] A. Dasgupta and J. B. Udgaonkar. Transient non-native burial of a Trp residue occurs initially during the unfolding of a SH3 domain. *Biochem*, 51:8226–8234, 2012.

[8] P. Neudecker, P. Robustelli, A. Cavalli, P. Walsh, P. Lundström, A. Zarrine-Afsar, S. Sharpe, M. Vendruscolo, and L. E. Kay. Structure of an intermediate state in protein folding and aggregation. *Science*, 336:362–366, 2012.

[9] A. Zarrine-Afsar, S. Wallin, A. M. Neculai, P. Neudecker, P. L. Howell, A. R. Davidson, and H. S. Chan. Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. *Proc Natl Acad Sci USA*, 105:9999–10004, 2008.

[10] F. Zosel, A. Soranno, K. J. Buholzer, D. Nettels, and B. Schuler. Depletion interactions modulate the binding between disordered proteins in crowded environments. *Proc Natl Acad Sci USA*, 117:13480–13489, 2020.

Appendix A

Supporting Information for Chapter 3

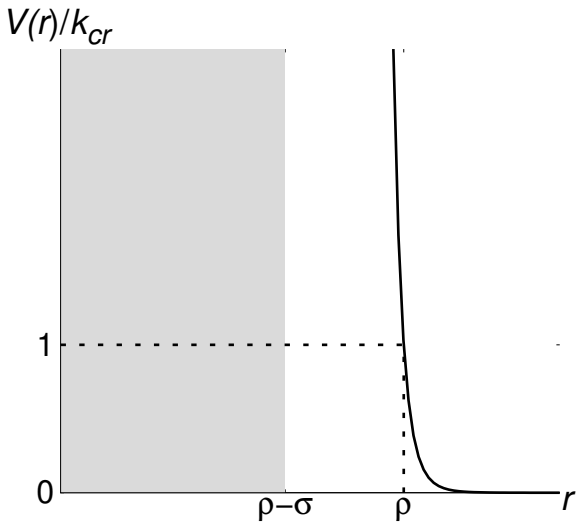

Figure A.1: The potential used in this work, $V(r)$, plotted as function of inter-particle distance r, (black curve) is controlled by three parameters: ρ (range), σ (softness) and $k_{\rm cr}$ (strength). For $r \leq \rho - \sigma$, $V(r)$ is infinite (a gray area).

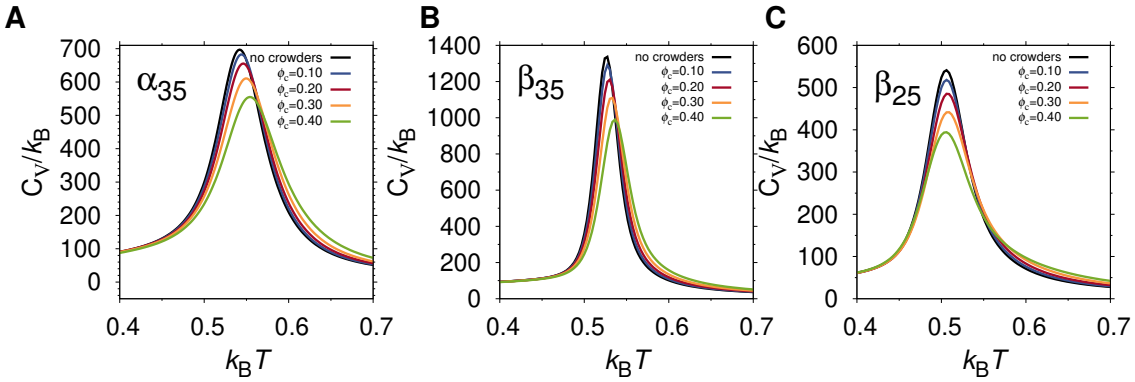

Figure A.2: Heat capacity $C_{\rm v}$ as a function of temperature for (A) α_{35}, (B) β_{35}, and (C) β_{25} in the absence and presence of excluded volume crowders with radius $R_{\rm c} = 12$ Å for α_{35} and β_{35} and $R_{\rm c} = 8$ Åfor β_{25} at different packing fractions, $\phi_{\rm c}$.

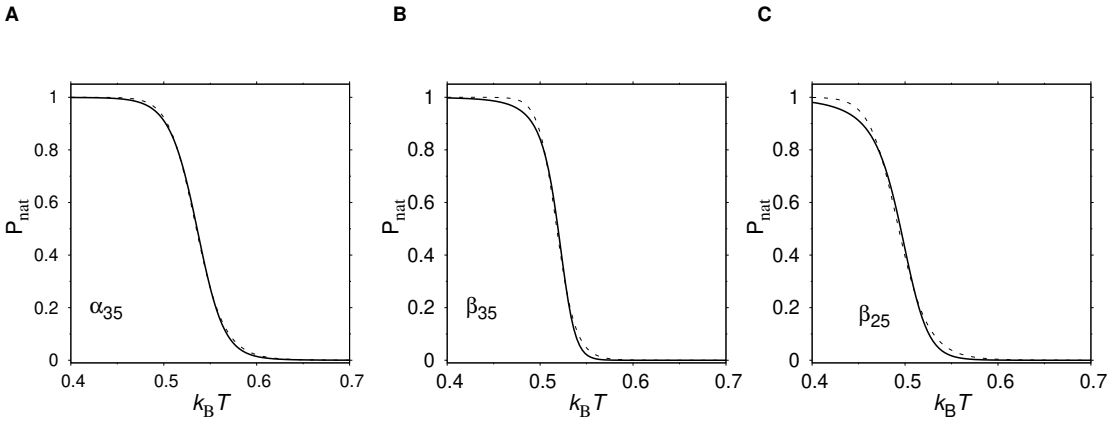

Figure A.3: Population of the native state, $P_{\rm nat}$, as a function of temperature for (A) α_{35} and (B) β_{35} (solid black curves). These curves are fitted to the equation $P_{\rm nat} = K/(1+K)$, where $K = \exp[-\Delta E(1/k_{\rm B}T - 1/k_{\rm B}T^0_{\rm m})]$, and ΔE and $T^0_{\rm m}$ are free parameters (dashed curves). Optimal fits give $\Delta E = -19.0$ and $k_{\rm B}T^0_{\rm m} = 0.535$ for α_{35}, $\Delta E = -26.8$ and $k_{\rm B}T^0_{\rm m} = 0.517$ for β_{35}, and $\Delta E = -15.4$ and $k_{\rm B}T^0_{\rm m} = 0.495$ for β_{25}. The errors on all 6 fit parameters are small, $< 0.5\%$.

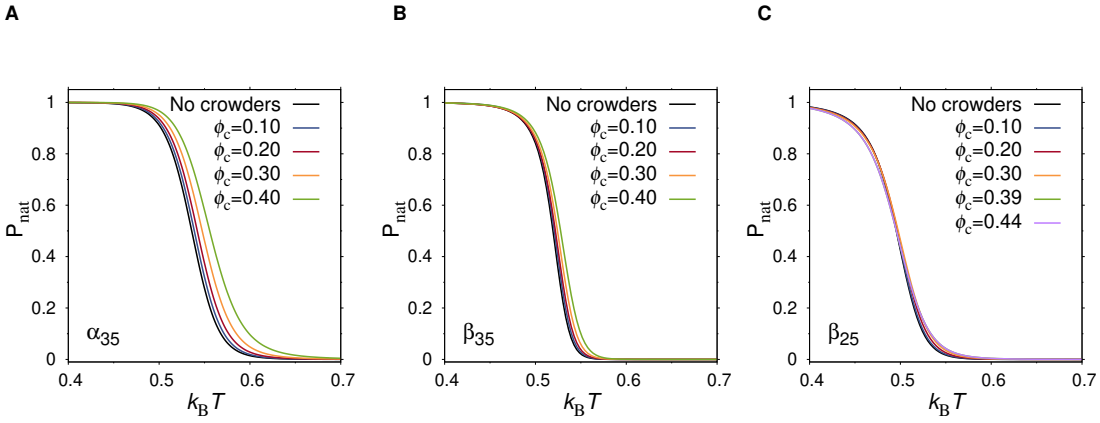

Figure A.4: Population of native state curve. P_{nat} as a function of temperature for (A) α_{35}, (B) β_{35} and (C) β_{25} in the absence and presence of crowders with radius $R_{\mathrm{c}} = 12$ Å for α_{35} and β_{35} and $R_{\mathrm{c}} = 8$ Åfor β_{25} at different packing fractions, ϕ_{c}.

Appendix B

Supporting Information for Chapter 4

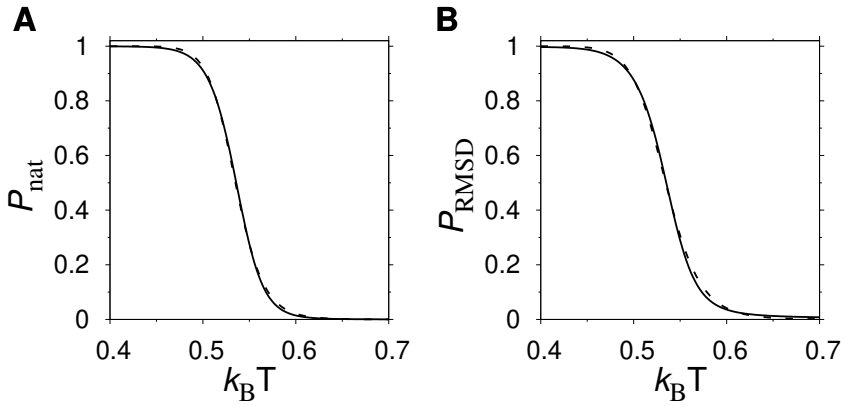

Figure B.1: Two-state fits of folding curves. Population of the native state, $P_{\rm nat}$ (A), and $P_{\rm RMSD}$ (B) as a function of temperature for α_{35} (solid black curves), where native state is defined as in the main manuscript or using the root-mean square deviation, RMSD. The native state is defined as RMSD $\leq$ RMSD$_{\rm cut}$, where RMSD$_{\rm cut}$ = 6.0Å for α_{35}. These curves are fitted to the equation $P_{\rm nat} = K/(1+K)$, where $K = \exp[-\Delta E(1/k_{\rm B}T - 1/k_{\rm B}T^0_{\rm m})]$, and ΔE and $T^0_{\rm m}$ are free parameters (dashed curves). Optimal fits give $\Delta E = -19.1$ and $k_{\rm B}T^0_{\rm m} = 0.535$ The errors on all 2 fit parameters are small, $< 0.5\%$.

Appendix C

Supporting Information for Chapter 6

C.1 Development of computational model for the GA/GB fold switch

C.1.1 Single-basin structure-based model for protein folding

As a starting point for the development our dual-basin structure-based model, which we apply in this work to the G_A/G_B switch, we take the single-basin model for protein folding developed in Ref. [1]. We start by describing this single-basin model along with a modification introduced here to enhance the conformational specificity of native contact interactions. We find that the enhanced contact specificity is necessary to

make the two folds structurally well defined in the dual-basin model. Geometrically, the protein is represented by beads located at the C_α atom positions. The conformation of an N-amino-acid chain can therefore be described by the bead positions $\mathbf{r}_\mathrm{i}$, where $\mathrm{i} = 1, ..., N$. Alternatively, a conformation can be described by the bond lengths, b_i, bond angles, θ_i, and dihedral angles, ϕ_i, defined by the $N-1$ (pseudo) C_α-C_α bonds of the chain. We denote by b_i^0, θ_i^0, and ϕ_i^0 the values of b_i, θ_i and ϕ_i in the native conformation. The potential energy E can be written as a sum of five terms:

$$\begin{aligned} E &= \sum_{\mathrm{i}}^{\mathrm{bonds}} K_\mathrm{b}(b_\mathrm{i} - b_\mathrm{i}^0)^2 + \sum_{\mathrm{i}}^{\mathrm{angles}} K_\theta(\theta_\mathrm{i} - \theta_\mathrm{i}^0)^2 \\ &+ \sum_{\mathrm{i}}^{\mathrm{dihedrals}} K_\phi^{(1)}[1 - \cos(\phi_\mathrm{i} - \phi_\mathrm{i}^0)] + K_\phi^{(3)}[1 - \cos 3(\phi_\mathrm{i} - \phi_\mathrm{i}^0)]) \\ &+ \sum_{\mathrm{i<j-3}}^{\mathrm{nonnative}} \epsilon \left(\frac{\sigma}{r_\mathrm{ij}}\right)^{12} + \sum_{\mathrm{i<j-3}}^{\mathrm{native}} \epsilon(h_\mathrm{ij} - f_\mathrm{ij}) \,, \end{aligned} \tag{C.1}$$

where ϵ sets the energy scale of the model and $r_\mathrm{ij} = |\mathbf{r}_\mathrm{j} - \mathbf{r}_\mathrm{i}|$. The first three terms represent bonded interactions with strengths set to $K_\mathrm{b} = 100\epsilon$, $K_\theta = 20\epsilon$, $K_\phi^{(1)} = \epsilon$ and $K_\phi^{(3)} = 0.5\epsilon$. The fourth term represents steric repulsions between bead pairs that do not form a contact in the native structure. The repulsion range is set to $\sigma = 4$ Å. These first four terms in Eq. C.1 are identical to Ref. [1].

The final term in Eq. C.1 represents native contact interactions, which in the previous model [1] were described by the Lennard-Jones potential $f_\mathrm{LJ}(r_\mathrm{ij}) = (r_\mathrm{ij}^0/r_\mathrm{ij})^{12} - 2(r_\mathrm{ij}^0/r_\mathrm{ij})^6$. Here we separate the interaction into a repulsive part (h_ij) and an attractive part (f_ij), such that they can be independently controlled. The repulsive part is

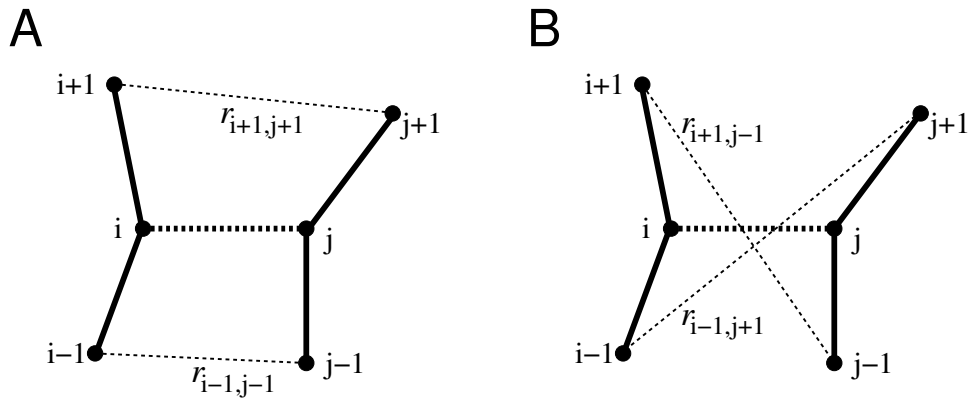

Figure C.1: A contact between two non-terminal positions, i and j, (thick dashed line) has four different nearest neighbor-nearest neighbor distances (thin dashed lines): (A) $r_{\mathrm{i-1,j-1}}$ and $r_{\mathrm{i+1,j+1}}$ and (B) $r_{\mathrm{i-1,j+1}}$ and $r_{\mathrm{i+1,j-1}}$. In evaluating the factor $g_{\xi_2}(r'_{\mathrm{ij}})g_{\xi_2}(r''_{\mathrm{ij}})$ in Equation C.1, r'_{ij} and r''_{ij} are the distances shown in (A), if $\Sigma_{\mathrm{A}} < \Sigma_{\mathrm{B}}$, or in (B), if $\Sigma_{\mathrm{B}} < \Sigma_{\mathrm{A}}$, where $\Sigma_{\mathrm{A}} = r^0_{\mathrm{i-1,j-1}} + r^0_{\mathrm{i+1,j+1}}$ and $\Sigma_{\mathrm{B}} = r^0_{\mathrm{i-1,j+1}} + r^0_{\mathrm{i-1,j+1}}$.

described by a Weeks-Chandler-Anderson type function,

$$h_{\mathrm{ij}} = \begin{cases} \left(\frac{r^0_{\mathrm{ij}}}{r_{\mathrm{ij}}}\right)^{12} - 2\left(\frac{r^0_{\mathrm{ij}}}{r_{\mathrm{ij}}}\right)^{6} + 1, & \text{if } r_{\mathrm{ij}} < r^0_{\mathrm{ij}}\,, \\ 0, & \text{if } r_{\mathrm{ij}} \geq r^0_{\mathrm{ij}}\,, \end{cases} \tag{C.2}$$

where r^0_{ij} is the distance between beads i and j in the native structure. The attractive part takes the form

$$f_{\mathrm{ij}} = g_{\xi_1}(r_{\mathrm{ij}})g_{\xi_2}(r'_{\mathrm{ij}})g_{\xi_2}(r''_{\mathrm{ij}})\,, \tag{C.3}$$

where $g_\xi(r) = \exp[-(r - r^0)^2/2\xi^2]$. With the construct in Eq. C.3, the distance r_{ij} as well as the two nearest neighbor distances, r'_{ij} and r''_{ij}, (see Figure C1) must assume their respective native values r^0_{ij}, r'^0_{ij} and r''^0_{ij} for ij to become a fully formed native contact, which then contributes $-\epsilon$ towards the total potential energy E. The parameter ξ_1 sets the width of the attractive well $-\epsilon g_{\xi_1}(r_{\mathrm{ij}})$. The combination of this attractive well and the repulsive part of the interaction results in a function, $h_{\mathrm{ij}} - g_{\xi_1}$,

with gross features similar to a Lennard-Jones potential (see Figure C2).

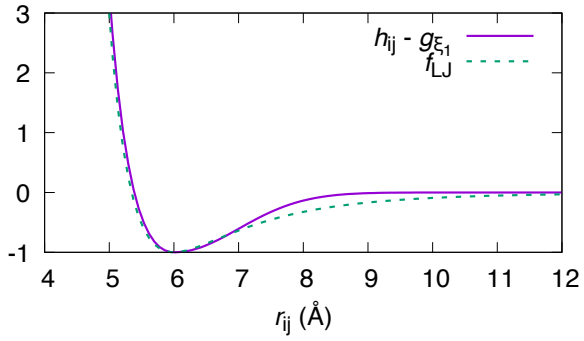

Figure C.2: The potentials $h_{ij} - g_{\xi_1}$ and f_{LJ} (see text) as functions of r_{ij} using $r_{ij}^0 = 6$ Å.

The factor $g_{\xi_2}(r'_{ij})g_{\xi_2}(r''_{ij})$ is included in f_{ij} in order to increase the conformational specificity of the native interactions. For a contact between residues i and j, this factor promotes the local chain segments $(i-1, i, i+1)$ and $(j-1, j, j+1)$ to adopt relative orientations close to that found in the native structure. The strength of this effect is controlled by the parameter ξ_2. It is weak when $\xi_2 \gg \xi_1$ and becomes strong when $\xi_2 \approx \xi_1$. Test simulations on a few small single domain proteins show that decreasing ξ_2 leads to increased co-operativity in the folding transition (data not shown). We picked $\xi_1 = 1.0$ Å and $\xi_2 = 5.0$ Å. We note also that there are terms in Eq. C.3 for which r'_{ij} or r''_{ij} is undefined because i or j is a terminal bead. In those cases, we set the corresponding factor g equal to unity.

The effect from the factor $g_{\xi_2}(r'_{ij})g_{\xi_2}(r''_{ij})$ in Eq. C.3 is similar to so-called local-nonlocal coupling [2], which also leads to increased folding co-operativity. Our effect is not exactly the same, however, because it does not provide a direct constraint on

the local internal conformation around beads i and j. Such a constraint does exist in local-nonlocal coupling.

C.1.2 Dual-basin structure-based model for fold switching

Next we extend the model of the previous section to a dual-basin (db) model, which provides bias towards two different reference structures "(a)" and "(b)". Such a bias can be achieved by first obtaining the two single-basin energy potentials $E^{(\mathrm{a})}$ and $E^{(\mathrm{b})}$ using Eq. C.1, and thereafter merging them into a single energy surface, $E^{(\mathrm{db})}$. Naively, one may attempt to put $E^{(\mathrm{db})} = E^{(\mathrm{a})} + E^{(\mathrm{b})}$. However, this strategy is problematic for some types of interactions, as pointed out by Ramirez-Sarmiento et al. [3]. For example, the sum of two quadratic bond terms $K_{\mathrm{b}}[(b_{\mathrm{i}} - b_{\mathrm{i}}^{(\mathrm{a})})^2 + (b_{\mathrm{i}} - b_{\mathrm{i}}^{(\mathrm{b})})^2]$ is another quadratic function with minimum at $(b_{\mathrm{i}}^{(\mathrm{a})} + b_{\mathrm{i}}^{(\mathrm{b})})/2$. Hence, this would abolish both minima. We combine the two single-basin potentials $E^{(\mathrm{a})}$ and $E^{(\mathrm{b})}$ using the procedure described below, which avoids these problems. This procedure is then applied to the $\mathrm{G_A}$ and $\mathrm{G_B}$ folds to produce the dual-basin potential used in this work.

Bonded terms. The bonded interactions are represented by the first three terms in Eq. C.1. Consider two individual energy terms, $e^{(\mathrm{a})}(x)$ and $e^{(\mathrm{b})}(x)$, with global minimum at $x = x^{\mathrm{a}}$ and $x = x^{\mathrm{b}}$, respectively. The functions $e^{(\mathrm{a})}(x)$ and $e^{(\mathrm{b})}(x)$ could be, e.g., the bond angle terms corresponding to a particular bond, in which case $x = \theta_{\mathrm{i}}$. To "mix" $e^{(\mathrm{a})}(x)$ and $e^{(\mathrm{b})}(x)$ into a single function $e(x)$, we use [4]

$$e(x) = \beta_{\mathrm{mix}}^{-1} \ln \left[e^{-\beta_{\mathrm{mix}} e^{(\mathrm{a})}(x)} + e^{-\beta_{\mathrm{mix}} e^{(\mathrm{b})}(x)} \right] , \tag{C.4}$$

where β_{mix} is a parameter controlling the smoothness of the mixing. We pick $\beta_{\text{mix}} = 10$ for the bond term, and $\beta_{\text{mix}} = 5$ for the angle and torsion terms. Examples of three different terms for the G_A and G_B folds are given in Figure C3.

Non-bonded terms. For the native contact term, we include all contact interactions present in either $E^{(a)}$ or $E^{(b)}$. Although this is straightforward in principle, care must be taken to avoid double counting interactions for common contacts, i.e., contacts that occur in both (a) and (b). Moreover, we want to insert parameters κ_A and κ_B such that strengths of the attractive wells $-\epsilon f_{ij}^{(a)}$ and $-\epsilon f_{ij}^{(a)}$ can be controlled. Hence, our dual-basin contact term becomes

$$\sum_{ij}^{(a)} \epsilon(h_{ij}^{(a)} - \kappa_A f_{ij}^{(a)}) + \sum_{ij}^{(b)} \epsilon(h_{ij}^{(b)} - \kappa_B f_{ij}^{(b)}) + \sum_{ij}^{\text{common}} \epsilon \left\{ \tilde{h}_{ij} - \max \left[\kappa_A f_{ij}^{(a)}, \kappa_B f_{ij}^{(b)} \right] \right\} ,$$

where the first two sums are taken over native contacts in (a) and native contacts in (b), respectively, *excluding* all common contacts, and the final sum is taken over these common contacts. Note that, for each common contact, only the energetically most favorable attraction is retained. The repulsive part, $\tilde{h}_{ij}$, is evaluated as h_{ij} using the smallest of the two reference distances, i.e., $r_{ij}^0 = \min \left[r_{ij}^{(a)}, r_{ij}^{(b)} \right]$. Picking r_{ij}^0 this way is necessary to guarantee that both conformations (a) and (b) can be formed without suffering a strong steric repulsion in one of the conformations, which would otherwise happen when $r_{ij}^{(a)}$ and $r_{ij}^{(b)}$ are very different. Note also that r_{ij}^0 for common contacts can be calculated before a simulation and that $\tilde{h}_{ij}$ does not change form during the simulation. The nonnative repulsive energy term, i.e., the fourth term in Eq. C.1, is evaluated over all pairs ij that are not contacts in either (a) or (b).

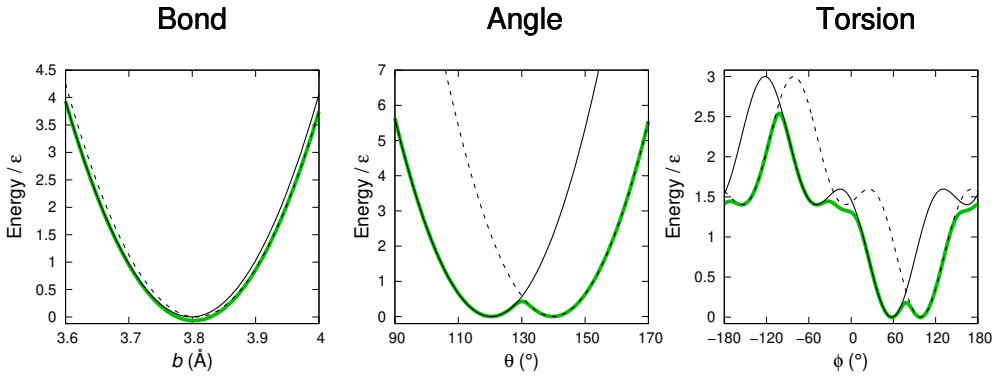

Figure C.3: Examples of the merging of different bonded potentials for G_A and G_B (thin black solid/dashed curves) into a single potential (thick solid green curves) using the "mixing" equation C.4

C.2 Supplementary figures

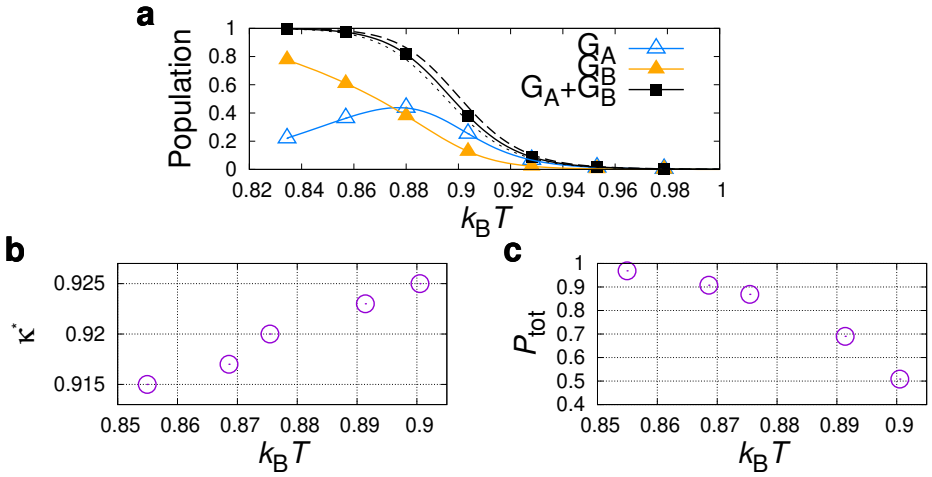

Figure C.4: (a) G_A, G_B, and total G_A and G_B (P_{tot}) fold populations for the protein G^*_{AB}, i.e., $\kappa_B = 0.92$, as function of temperature T. Smooth curves are obtained from MBAR analysis [5]. (b) κ^* as function of T, obtained from simulations with $\kappa_B = 0.915, 0.917, 0.920, 0.923$ and 0.925. (c) P_{tot} as function of T, for the five κ_B values in (b). Shown in (a) is also P_{tot} for $\kappa_B = 0.915$ (short dashed curve) and $\kappa_B = 0.925$ (long dashed curve).

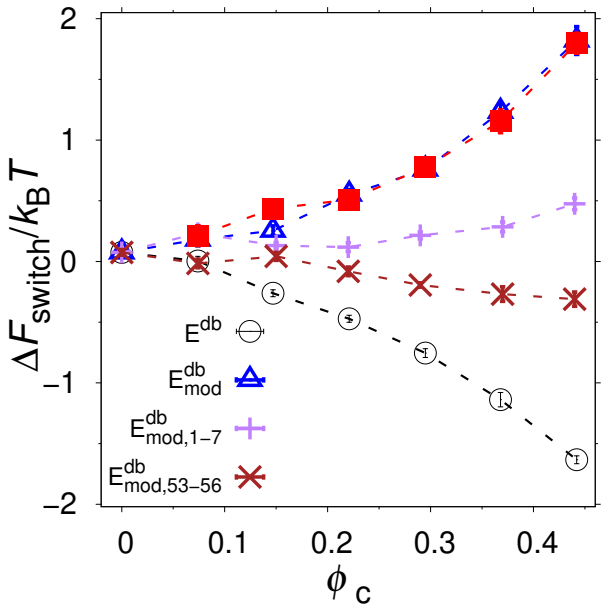

Figure C.5: Crowder concentration (ϕ_c) dependence of the free energy of fold switching (ΔF_{switch}) for the protein G^*_{AB} using four different potential energy functions: $E^{(db)}$ (original dual-basin model; see section 2.2 of this document), $E^{(db)}_{mod,1\text{-}7}$ (crowder-bead interactions turned off for the segment 1-7), $E^{(db)}_{mod,53\text{-}56}$ (crowder-bead interactions turned off for the segment 53-56), and $E^{(db)}_{mod}$ (crowder-bead interactions turned off for both 1-7 and 53-56). Also shown, for $\phi_c > 0$, are the results obtained by adding the changes in ΔF_{switch} for $E^{(db)}_{mod,1\text{-}7}$ and $E^{(db)}_{mod,53\text{-}56}$ relative to the original model $E^{(db)}$ (filled squares). The good agreement with the results for $E^{(db)}_{mod}$ indicates that the effects of the modifications introduced in $E^{(db)}_{mod,1\text{-}7}$ and $E^{(db)}_{mod,53\text{-}56}$ are roughly additive in the quantity ΔF_{switch}.

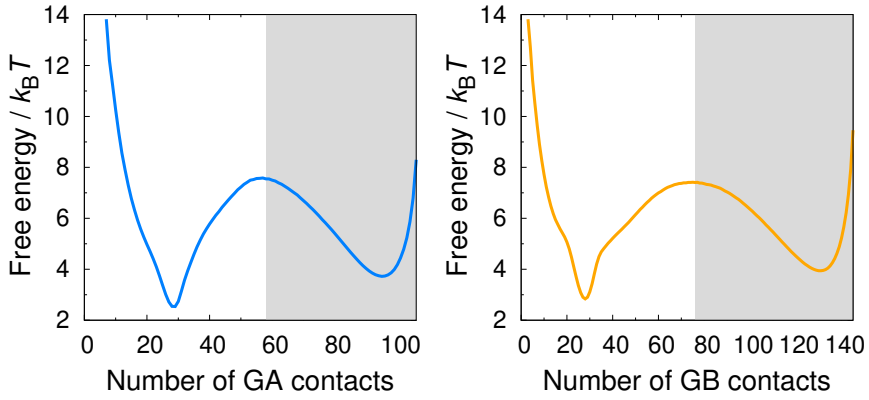

Figure C.6: Free energy as function of the number of G_A contacts and the number of G_B contacts, obtained for G^*_{AB} at $k_B T_0 = 0.88$. In determining fold populations, we consider the G_A fold formed if the number of G_A contacts is > 58 and we consider the G_B fold formed if the number of G_B contacts is > 76, respectively (shaded areas).

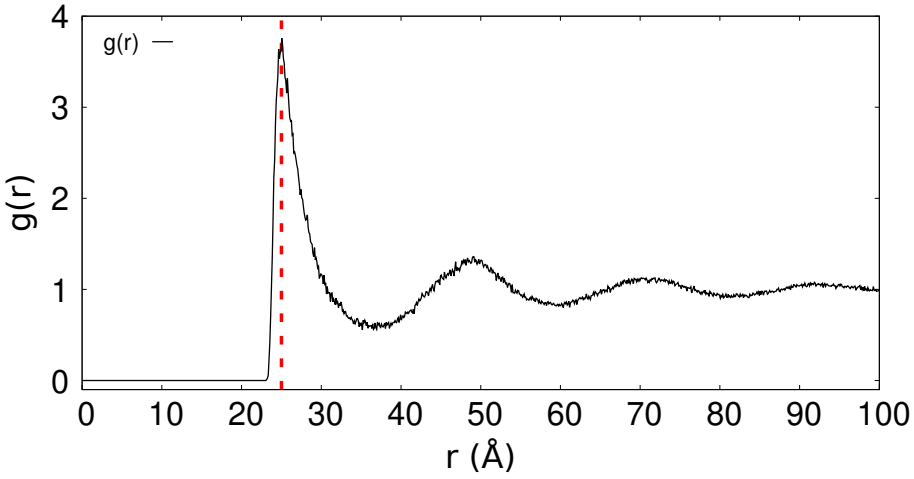

Figure C.7: Pair correlation function, $g(r)$, obtained from a simulations of 1755 crowder particles in a cubic box with side 300 Å and periodic boundary conditions. The temperature is held fixed at T_0 ($k_{\mathrm{B}}T_0 = 0.88$). An effective radius of the crowders ($R_{\mathrm{c}} = 12.5$ Å) was determined based on the location of the first peak in $g(r)$, which occurs at the interparticle distance $r = 25$ Å (vertical dashed line).

Bibliography

[1] S. Wallin and H. S. Chan. Conformational entropic barriers in topology-dependent protein folding: perspectives from a simple native-centric polymer model. *J Condens Matter Phys*, 18:S307, 2006.

[2] H. S. Chan, Z. Zhang, S. Wallin, and Z. Liu. Cooperativity, local-nonlocal coupling, and nonnative interactions: principles of protein folding from coarse-grained models. *Annu Rev Phys Chem*, 62:301–326, 2011.

[3] C. A. Ramirez-Sarmiento, J. K. Noel, S. L. Valenzuela, and I. Artsimovitch. Interdomain contacts control native state switching of RfaH on a dual-funneled landscape. *PLOS Comp Biol*, 11:e1004379, 2015.

[4] R. B. Best, Y. G. Chen, and G. Hummer. Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor. *Structure*, 13:1755–1763, 2005.

[5] M. R. Shirts and J. D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys*, 129:124105, 2008.