Determining the importance of hydrogen bond strength on the mechanical stability of proteins

Megan Lianne Hughes

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds
School of Physics and Astronomy

October 2015

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

For Tom Dixon and all of my family especially the big G-ma also known as my Gran. She is the strongest person I know and has taught me how to be a "tough old bird".

"If one tugs on something in nature, he finds it attached to the rest of the world" - J. Muir

Acknowledgements

I would like to thank my supervisors, Dr. Lorna Dougan and Dr David Brockwell. This Ph.D has been challenging at times but I have made it through with some stumbles along the way. Without their help and support I would never have made it to the end of the project. Their feedback and discussions have been invaluable and they have taught me so much during my Ph.D. Their enthusiasm for science has made our meetings so energetic and helped me to understand so many key topics. Thank you to Lorna who accepted a young over-enthusiastic undergraduate into her group. Her hard work and dedication to every project has been paramount to this work I have done. She has also helped me believe in myself, and the work I have done, even when I didn't believe in myself. I also admire her creative thinking, whether it is about a problem in science, or for one of our groups photos. Thank you to David who was so patient whilst teaching me everything I now know about protein expression and purification (among other scientific knowledge). He put up me causing chaos in the biology laboratories in my first year and will probably not miss the phrase "well the thing is" when something didn't quite go to plan. He has supported me so much and I would not have got to this point without him. I admire the way he considers problems in science and how much of a sponge for knowledge he is. They both have built up my confidence as a researcher so thank you. I would like to also thank the European Research Council for making this research possible.

One person in particular needs a special mention. I would not have been able to do this Ph.D without the support of Kasia. She is an amazing scientist and an even better friend. Every time I was stressed she was always there with a supportive hug and a brownie would mysteriously appear outside of the laboratory, or on my desk. I admire her ability to get everything done, in fact we coined the phrase "Super Kasia". There is literally nothing she cannot do and I really do look up to her. She has been a constant stream of loveliness during this Ph.D, she even let me live with her and Chris (thank you both)! I will miss the times she could not even run up a hill without dying, because this was the only time I was better than her at something. Thank you also for teaching me, watching experiments and helping me out with all of my work and AFM experiments along the way. Even though she is not in this country any more she still managed to check my work and I am so grateful to her. Thanks also needs to be given to the rest of the Dougan group. Especially Danielle Walsh who has been through all the painful moments with me. Due to our unique moments in biology, we are co-authoring the book: "Adventures of a Physicist in Biology". Thank you to her for also not killing me when I would laugh at something going wrong.

Thank you to everyone who has measured data for me and taught me how to use equipment. Thank you to Toni Hoffmann for also teaching me AFM and how to use the CD instrument, Michael Wilson for teach me fluorescence, Julia for teaching me how to use the stopped flow instrument and Dr James Ault for doing the mass spectrometry measurements. Also thank you to Dr Emanuele Paci, and Prof. Peter Olmsted for discussions. A thank you also needs to be given to Prof. Sheena Radford for the use of her laboratory and equipment.

Thank you to all of the physicists in MNP it has been so much fun working with you all. A special mention need to go to my exercise buddy Spanish (aka Queen Victoria). She has been so supportive with my stress, when I've cried. She is also doing a great job at teaching me Spanish! I also need to acknowledge everyone in Astbury. Especially Nasir, Rhys, Theo and Rachel who were always there to answer questions when I was stuck. Nasir was so welcoming when I first started using the laboratory as an alien physicist. And thank you to everyone else along the way.

I also need to thank some of my lovely friends, including Phili and Julia (who are definitely my Leeds support group!). Thank you Kerry for being amazing and always being there for me and Stevie for constantly checking I am alive.

I especially need to mention Mr Tom Dixon. He has been amazingly patient and understanding whilst I have been doing this Ph.D. He has provided me with late night dinners, unlimited lifts from university when I have worked late, and breaks from work when I've needed them the most. He is always there for me and has made me see I can get through the toughest of times. I am very grateful for all the amazing treats and trips he has bought me and for encouraging me to travel. This thesis would never have been finished without his support and I will always be grateful for his ongoing support even if he does still think I just research pieces of strings!

Last but not least, thank you to the support of my family. Thank you to the big G's who are always proud. Thank you to my mum especially: She is always there at the other end of the phone no matter what time I have needed her and is always there with a big cuddle when I visit. She's been through so much and she still manages to worry about me! My crazy sisters Ceri and Sian also need a mention. Despite years of sisterly abuse, they have always encouraged me along the way and I love them for it. You are all amazing and I would never have got to this point without you!

Abstract

Proteins are only marginally stable at room temperature and can easily be unfolded when subjected to denaturants such as chemicals or mechanical forces. Despite this, proteins are able to maintain a unique 3D structure held together by interactions such as hydrogen bonds and hydrophobic interactions in physiological conditions. We have used single molecule force spectroscopy (SMFS) to investigate the effect of hydrogen bond strength on the mechanical resilience of two proteins: I27, from the giant muscle protein titin, and protein L, an immunoglobulin binding domain from bacterial cell wall. By picking up and stretching a single protein, the unfolding force needed to unravel it can be determined; this infers information about the molecular interactions that confer its stability and flexibility.

To directly test the importance of hydrogen bond strength, we have exploited the use of deuterium (D or $_2$H) and deuterium oxide (D_2O). Deuterium has been shown to form strong hydrogen bonds than hydrogen (1.04 – 2.07 k_BT stronger). Protein engineering was used to create protonated and deuterated versions of the proteins and SMFS experiments have been completed on the proteins in both water (H_2O) and D_2O. Our single molecule studies indicate that the mechanical resilience of both proteins is sensitive to the hydrogen bond strength in both the protein and the solvent. Furthermore, the changes in the mechanical resilience of the protein is coupled with an changes in the spring constant of the protein chain. The changes observed in the properties of the proteins due isotopic substitution could aid differentiation between the interactions involved within the rate limiting step of mechanical unfolding. This study has illustrated the sensitivity of SMFS experiments to small changes in the strength of the interactions within proteins, and the importance of probing the net contribution of different interactions to protein stability.

Publications

Refereed Journals

K. M. Tych, T. Hoffmann, M. Batchelor, **M. L. Hughes**, K. E. Kendrick, D. L. Walsh, M. Wilson, D. J. Brockwell and L. Dougan. Life in extreme environments: Single molecule force spectroscopy as a tool to explore proteins from extremophilic organisms. *Biochemical Society Transactions*, 43:179-185 2015.

K. M. Tych, **M. L. Hughes**, J. Bourke, Y. Taniguchi, M. Kawakami, D. J. Brockwell and L. Dougan. Optimizing the calculation of energy landscape parameters from single-molecule protein unfolding experiments. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 91:012710 2015.

T. Hoffmann, K. M. Tych, **M. L. Hughes**, D. J. Brockwell and L. Dougan. Towards design principles for determining the mechanical stability of proteins. *Phy Chem Chem Phys*, 15:15767-15780 2013.

List of abbreviations

Experimental Technique	
AFM	Atomic force microscope
CD	Cicular dichroism
GF	Gel filtration/Size exclusion chromatography
IMAC	Immobilised metal-affinity chromatography
NMR	Nuclear magnetic resonance spectroscopy
MCS	Monte Carlo simulations
MDS	Molecular dynamic simulations
MS	Mass spectrometry
SEC	Size exclusion/Gel filtration chromatography
SMFS	Single molecule force spectrosccopy
Model/properties	
ΔL_c	Change in contour length
ΔH	Change in enthalpy
ΔS	Change in entropy
ΔU	Change in internal energy
ΔS_{conf}	Change in conformational entropy
ΔG	Change in Gibbs free energy
$\Delta G_{nonpolar-->water}$	Change in transfer free energy between a nonpolar environment to water
c_A^β	Chemical potential
c_A^α	Chemical potential
ΔE	Change in Helmholtz free energy
L_c	Contour length '
Δx_U	Distance between unfolded state and transition state
$\Delta G_{F\text{-}TS}$	Free energy difference between folded and transition state of a protein

$\Delta G_{F\text{-}U}$	Free energy difference between folded and unfolded state
$\Delta G_{u,chem}$	Chemical denaturation free energy difference between the folded and unfolded state
$\Delta G_{TS,force}$	Mechanical height of the free energy barrier to protein unfolding
KT	Kramer's theory
T_m	Midpoint temperature of unfolding
L_p	Persistence length
v	Pulling velocity
k_f	Rate of folding
C_p	Specific heat capacity
BM	The Bell Model
TS	Transition state
T	Temperature
T_m	Thermal melting temperature
P	Pressure
V	Volume
F	Force
x	distance
F_{UN}	Unfolding force
$k^0_{u,chem}$	Unfolding rate at 0M denaturant
$k^0_{u,force}$	Unfolding rate at zero force
WLC	Worm-like chain model
Ω	Number of accessible states
SASA	Solvent accessible surface area
ZPVE	Zero point vibrational energy
RCO	Relative contact order
w.t.	Wild-type

Quantities

g	gram
kg	kilogram
mg	milligram
μg	microgram
ml	millilitre

μl	microlitre
mM	millimolar
μM	micromolar
M	Molar
MWCO	Molecular weight
pN	piconewton
nN	nanonewton
nm	nanometer
cm	centimeter
s	seconds

Databases

BSDB	Biomolecular stretching data base
PDB	Protein data bank

Protein model system labelling

p(protein name)$_5$ in H_2O	Fully protiated protein in a H_2O solvent environment
p(protein name)$_5$ in D_2O	Fully protiated protein in a D_2O solvent environment
d(protein name)$_5$ in H_2O	Fully deuterated protein in a H_2O solvent environment
d(protein name)$_5$ in D_2O	Fully deuterated protein in a D_2O solvent environment

Biological molecules

DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
pL	Protein L

Interactions

HB	Hydrogen bond(ing) involving hydrogen (H) atoms

Important chemicals

H	Hydrogen
D	Deuterium
H_2O	Water
D_2O	Deuterium oxide

ATP	adenosine triphosphate
GdnHcl	Guanidine hydrochloride

Constants

k_B	Boltzmann's constant
h	Plank's constant
R	Gas constant

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter will begin by introducing the biomolecules under investigation in this project, namely proteins. The importance of hydrogen bonds, the solvent and hydrophobic interactions to the function, structure and stability of proteins will then be discussed, followed by the use of deuterium and deuterium oxide as methods to manipulate the strength of different interactions. Finally it will discuss the contribution single molecule force spectroscopy has made to understanding the importance of interactions governing protein folds, and the use of deuterium, an isotope of hydrogen, for investigating protein mechanical stability.

1.1 Proteins

The word protein originates from the Greek for "first rank importance" [1]. These biological molecules are important for all biological processes such as: catalysis of reactions, mechanical support, transportation of molecules and transmitters of signals [1]. Understanding the biology, chemistry and physics of these molecules is of fundamental importance. Proteins are long chain polymers (also known as polypeptides) made up of small subunits called amino acids[1] covalently linked together via peptide bonds [2].

1.1.0.1 The primary structure of proteins

The typical chemical structure of an amino acid is shown in Figure 1.1.0.1 where R denotes the side chain. There are 20 naturally occurring amino acids each with a unique side chain conferring different physical and chemical properties. The 20 naturally occurring amino acids are shown in Figure 1.1.0.2 and are grouped according to the

[1] An amino acid is an organic molecule consisting of a central carbon atom (C_α) covalently bound to a hydrogen, a carboxylic acid COOH group, an amino NH_2 and a side chain. This side chain is unique to each amino acid.

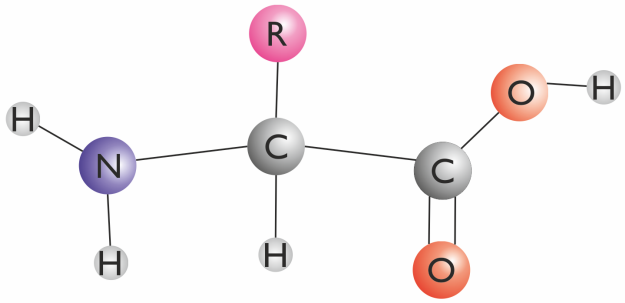


Figure 1.1.0.1: Proteins are made up of subunits called amino acids. This schematic depicts the general structure of an amino acid. An amino acid is made up of a central carbon atom, known as the α-carbon, covalently bound to a hydrogen atom, a carboxyl group (COOH), an amino group (NH_2) and an R group. The R denotes the characteristic side-chain of atoms which is unique for every type of amino acid. This side chain gives an amino acid different chemical properties.

chemical properties of their side chains. The R group is responsible for an increase in hydrophobicity[2] or an increase in the overall charge of the molecule. The order of amino acids in the polypeptide chain is known as the sequence of the protein and forms the primary structure of the protein as shown in Figure 1.1.0.3. Amino acids in the primary sequence determine how the chain folds, amino acids are able to interact and fold into different shapes according to their sequence (Figure 1.1.0.3).

Figure 1.1.0.2: Figure showing the 20 naturally occurring amino acids and their side chains. Amino acids are the building blocks of proteins. The amino acids are grouped according to the properties of their side chains [3]. Nonpolar, aliphatic amino acids (yellow) are amino acids with long chain hydrocarbons. The sidechains of these amino acids are typically hydrophobic which means it does not interact favourably with water molecules. Aromatic amino acids (pink) are amino acids that contain an aromatic ring on their side chain. These amino acids can be nonpolar or polar. The aliphatic polar uncharged amino acids (green) contain an amino or hydroxyl group on their sidechain. These amino acids can interact and form hydrogen bonds with other polar amino acids, water molecules and the backbone. The sulfur containing amino acids (orange) include cysteine and methionine. Cysteine amino acids contain a sulfur-hydryl (SH) group. This amino acid is able to form covalent disulphide bonds with other cysteine residues. The methionine has a hydrophobic side chain. The charged amino acids carry either a negative (acidic, red) or positive (basic, blue) charge at neutral pH.

[2]Tendency for a molecule to not interact with water.

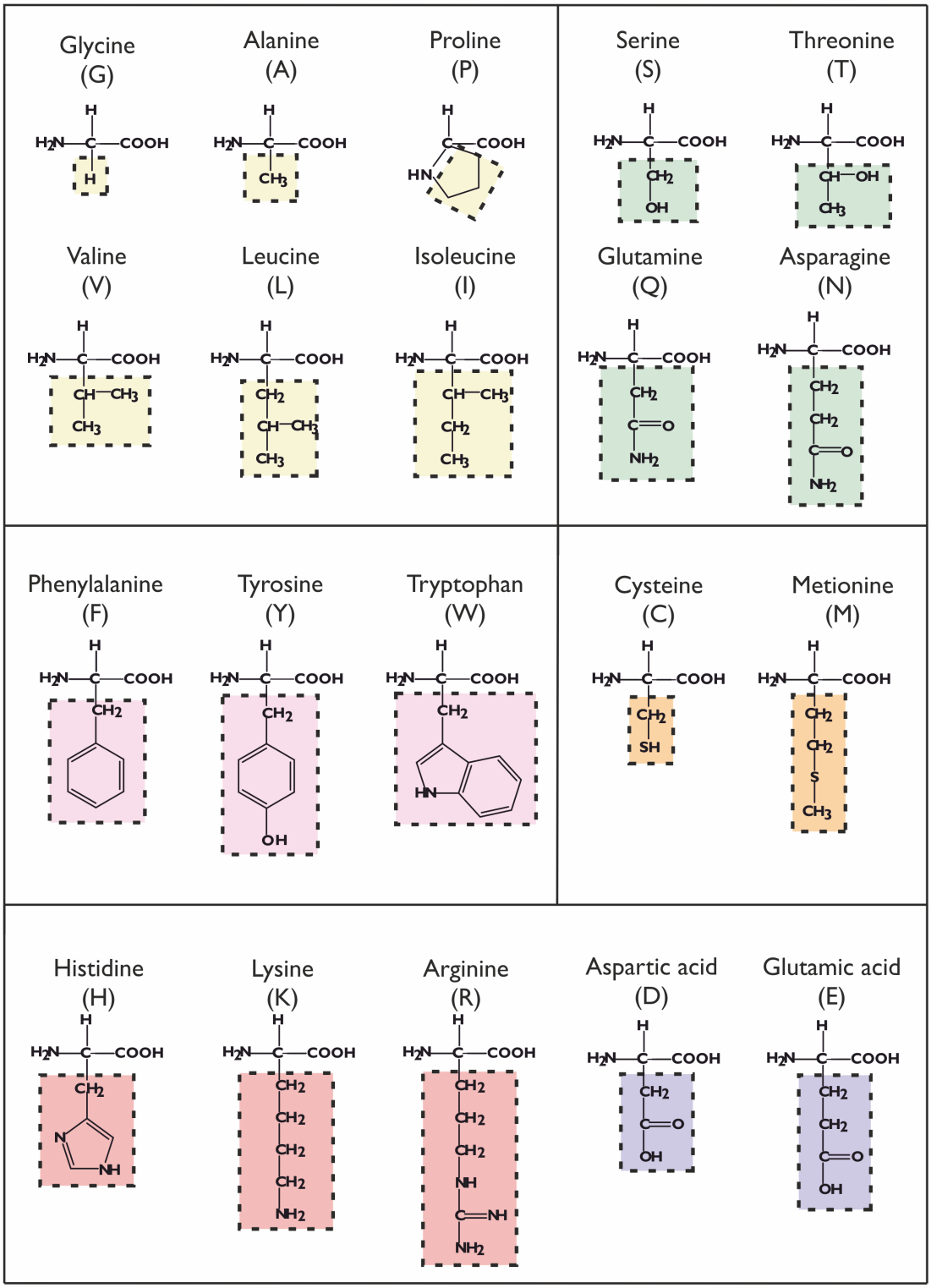


Figure 1.1.0.2: (Caption on the previous page)

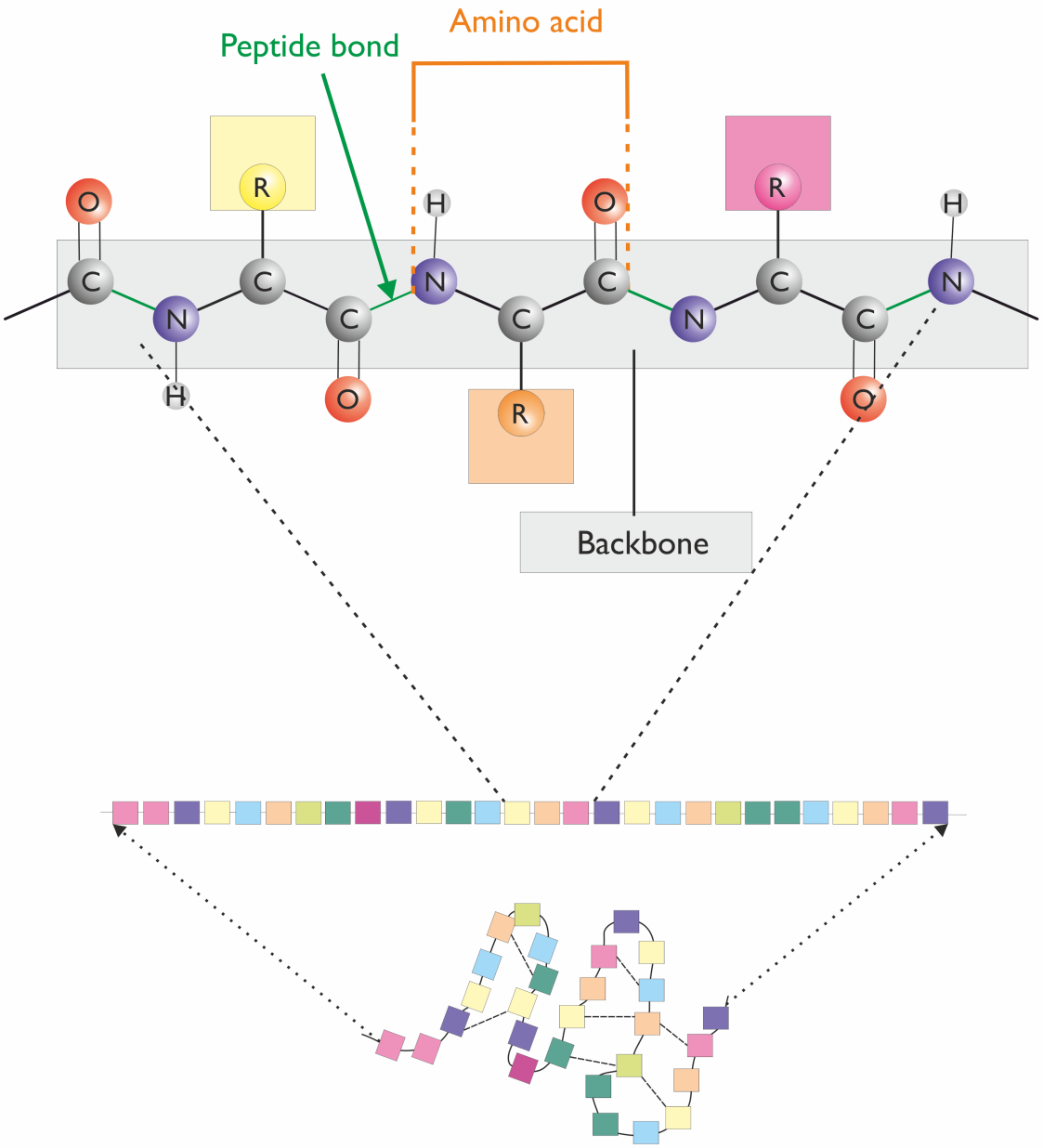


Figure 1.1.0.3: Schematic showing the basis for the primary and secondary structures of proteins. The primary structure of the protein refers to the order of amino acids in the polypeptide chain. Each amino acid has a side chain (shown as the coloured square). This polypeptide chain can fold into different shapes based on the order of the amino acids. These folded peptide chains form the secondary structure of a protein.

1.1.0.2 Dihedral angles and Ramachandran plots

The peptide bond is rigid due to a "partial" double bond forming due to interactions between the electron clouds of the N-, C'- atoms [2, 4]. The rotation about the C'-N bond is therefore restricted. When the protein folds this means the number of conformations available to the chain decreases which has significant consequences to the stability of a protein (Section 1.2). Understanding why a protein adopts a particular conformation is important to understanding how they remain folded. The relative positions of side chains about the carbon backbone bonds can be described by dihedral angles. These angle are formed between two planes as depicted in Figure 1.1.0.4. In this figure one plane is formed between the (1) and (a) atoms and the other is formed between the (2) and (b) atoms. The angle between these planes is known as the dihedral angle of the (a) and (b) groups about the (1)–(2) bond. The dihedral angle increases if one plane rotates anticlockwise but decreases in the clockwise direction.

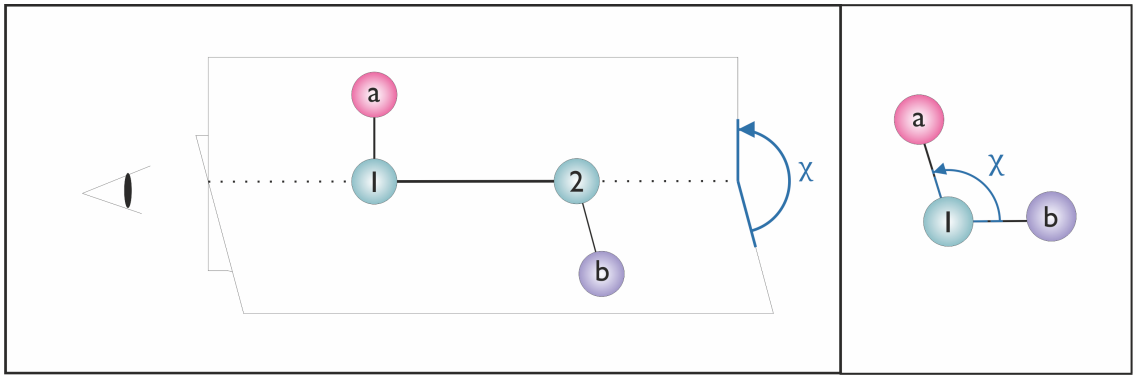


Figure 1.1.0.4: Schematic diagram showing how dihedral angles are measured. In this molecule containing atom (a) covalently bound to atom (1) this atom is also bound to atom (2) which, in turn, is bound to (b)(a–1–2–b). The angle of rotation of the side chains a and b is taken about the 1–2 bond. The a–1 and 2–b form the two intersecting planes to measure the angle χ between as shown in the left-hand figure. The angle is also illustrated from the viewpoint of the eye (right).

The protein backbone (Figure 1.1.0.3) can be described in terms of the angles ϕ, ψ and ω as shown in Figure 1.1.0.5. The ϕ angle is the rotation of the two adjoining C' groups around the N–C_α bond. The ψ angle describes the rotation of the N groups around the C_α–C' bond. The ω angle of rotations describes the position of the two C_α groups around the N–C' bond. The magnitude of the angles depends on the positions of the electron clouds. If electron clouds, already involved in covalent bonds, become too close there is a repulsion, known as a steric clash. This is based on the Pauli exclusion principle: Only two electrons (with opposite spins[3]) are allowed in a single orbital [2]. If

[3]Spins refer to the moments of rotation of an electron. Electrons have two spin state: spin up and spin down.

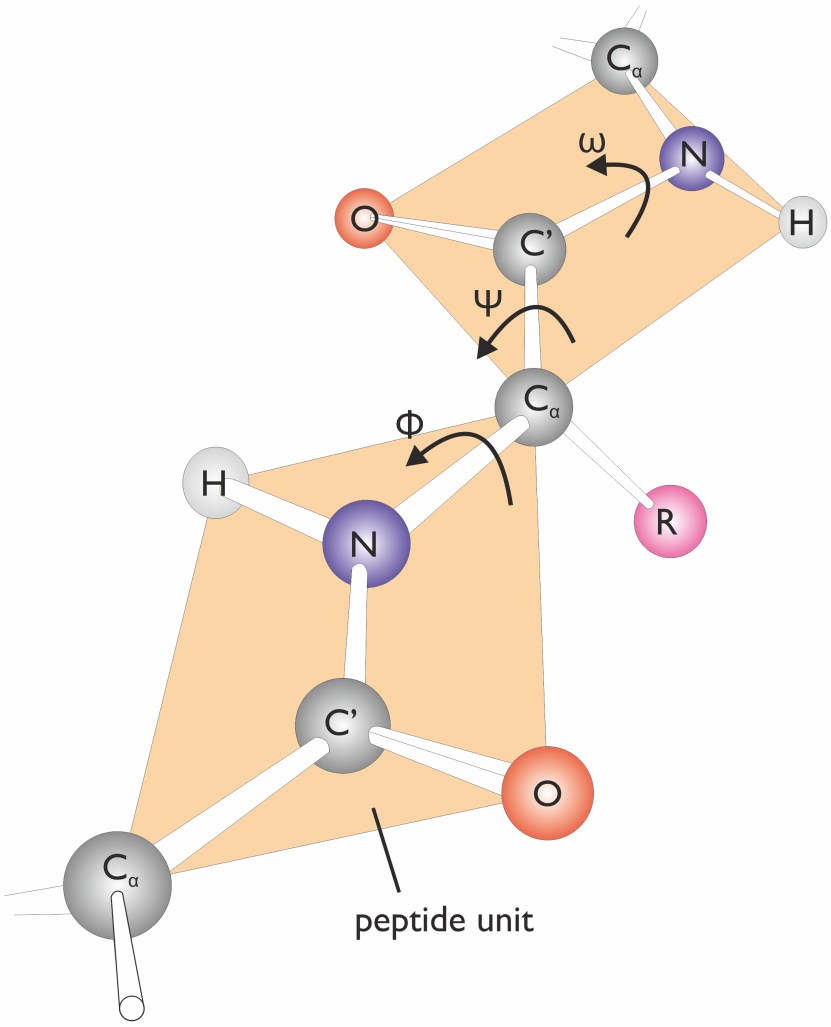


Figure 1.1.0.5: Figure illustrating the angles of rotation in the backbone of a polypeptide bond. Nitrogen atoms, N, are blue spheres, carbon atoms, C, are grey, hydrogen atoms, H, oxygen atoms, O, are red and side chains, R, are pink. The peptide unit is highlighted by the orange rectangles, this is formed by the covalent peptide bond between the amide N and carboxyl C' of amino acids. The C_α is the central carbon atom in an amino acid bound to the characteristic side chain labelled R. Three angles are illustrated: The ω angle is the angle of rotation around the C'–N peptide bond. The ψ angle is the rotation around the C_α–C' bond. The ϕ is the angle of rotation around the N–C_α bond. The angles of rotations are taken between the planes corresponding to the largest groups covalently bound to the two atoms.

an electron cloud contains two electrons it is said to be saturated and therefore repels other saturated electron clouds if they overlap [2]. This leads to potential barriers on the rotations on each of the bonds. At larger distances, however, the atoms will attract one another due to coordinated electron vibrations in the orbital. The distance (r) dependence on the interaction potential, U, between two atoms can be described by a Lennard-Jones (LJ) potential [2]:

$$U(r) = E_0 \left[\left(\frac{r_o}{r} \right)^{12} - 2 \left(\frac{r_0}{r} \right)^{6} \right] \tag{1.1}$$

Where r_o is the optimum distance where U is at a minimum. This value of U corresponds to E_0. This potential describes the attraction and repulsion observed between atoms.

The ω angle has two minima and two maxima in the potential barriers within a 360 ° rotation. These potential minima occur at 0 ° and 180 ° and corresponds to the angles at which the electron clouds of the N and C' atoms are at their closest. The potential barriers are high due to the partial double bond nature of the peptide bond and therefore tend to only fluctuate $\sim$ 5–10 ° due to thermal fluctuations [2]. If ω=0 °then the molecule is said to be in a cis configuration. This is shown in Figure 1.1.0.6. The two C_α are in the same plane. This is energetically unfavourable due to the large size of these atoms, the distance between them in this configuration is $\sim$ 0.28 nm [2]. The minimum distance for two carbon atoms in the crystalline state (0.30 nm), r_{min}, is the approximate distance at which the LJ potential goes from being attractive to repulsive at short distances [2]. The cis configuration is therefore less energetically favourable than the trans configuration, where ω=180 °and the two C_α atoms are on opposites sides of the C'–N bond. The peptide bonds are therefore found in the trans configuration with the exception of proline, where two carbons are bound to the peptide nitrogen. This results in a 20 % chance of the proline amino acid being found in the cis conformation [2].

Unlike the peptide bond, the N–C_α (described by the ϕ angle) and the C_α–C' (described by the χ angle) are purely single bonds. These bonds are free to rotate and determine the path of the polypeptide chain [5]. However, the angles of these bonds are also limited by steric clashes of the main chain and bulky side chains. A Ramachandran plot is often used to illustrate the regions of "allowed" and "disallowed" ϕ and χ angles. The area of the number of allowed angles is much smaller than the region of disallowed.

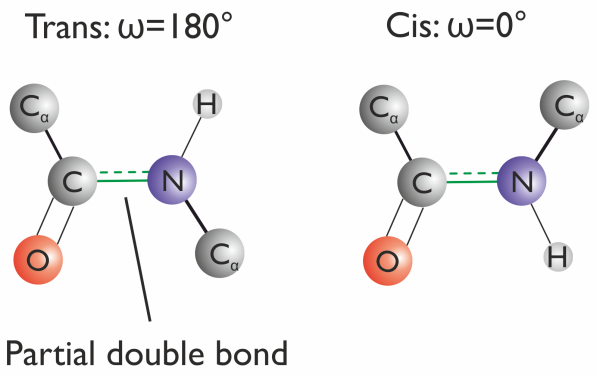


Figure 1.1.0.6: Schematic of the cis and trans bonds around a rigid N–C' bond. The cis configuration has both C_α atoms on the plane (i.e, ω=0 from the dihedral angles). This is sterically unfavourable due to the close proximity of the electron clouds of the atoms. The trans configuration has the C_α atoms on the opposite sides. This is more favourable due to the distance between the electron clouds of the two carbon atoms.

The regions of allowed angles are important for protein folding. The 3D structures that proteins adopt have well defined regions associated with them on a Ramachandran plot [6]. Figure 1.1.0.7 shows a schematic of the regions associated with different types of 3D structures. Therefore the order of amino acids can determine whether a protein has a propensity for a certain structure.

1.1.0.3 Secondary, tertiary and quaternary structure of proteins

The common shapes (or structural motifs) a protein adopts are known as β-sheets, formed by parallel or anti-parallel β-strands. The β-structures are shown by arrows in Figure 1.1.0.8, and α-helices shown as the helical ribbon. These structures, among other less common structural motifs, are known as the secondary structure of the protein. The existence of these structures were predicted as early as 1951 [4, 7]. In 1951 Linus Pauling won a Nobel prize for his contribution to predicting a number of secondary structures [4]. Typically one polypeptide chain forms multiple secondary structures. These secondary structures can interact to form a 3D shape. This is known as the tertiary structure of the protein as shown in Figure 1.1.0.8. Protein tertiary structure can be classed in three main categories: all β structure; all α structure and α/β structure. A number of different protein structures are shown in Section 1.6.3. Quaternary structure of a protein refers to interactions between multiple polypeptide chains or subunits. When a protein folds, it will "test" multiple different 3D structures before adopting the most favourable conformation. This unique structure is held together by many chemical forces between the parts of the polypeptide and the surrounding environment. These chemical forces

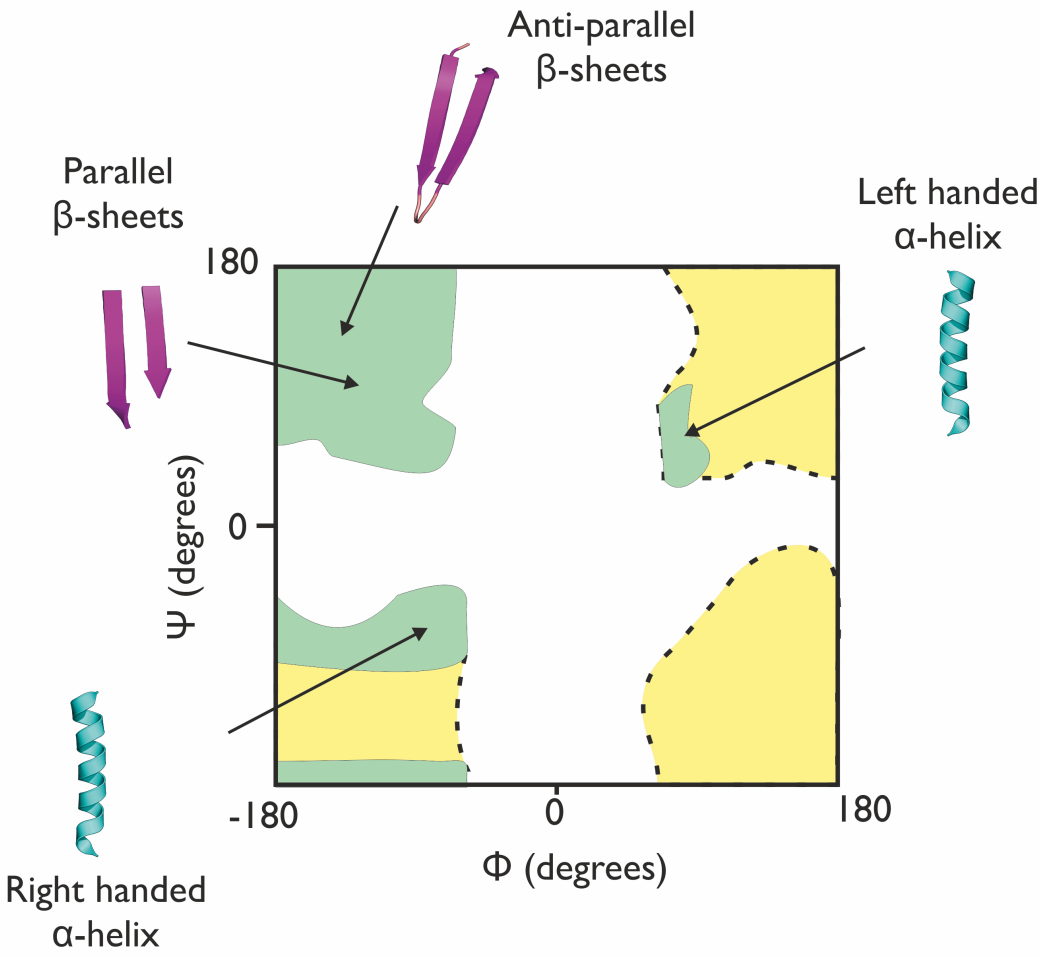


Figure 1.1.0.7: Schematic of a Ramachandran plot of the ϕ angle against the χ angle. These are the angles of rotation around the N–C_α and the C_α–C' respectively. The approximate regions of the most energetically favourable angles for all amino acids are shown as green, less favourable as yellow and disallowed regions as white. The regions correspond to different types of secondary structure and these are illustrated. The disallowed region occurs due to steric clashes of the backbone and side chains.

are known as non-covalent interactions and typically depend on the primary structure of the polypeptide. For example, the environment surrounding a protein is aqueous, and consequently, hydrophobic amino acids (Figure 1.1.0.2) that are "water fearing" and are typically found to be shielded from water in the centre of the protein fold. Hydrophilic ("water loving") amino acids, on the other hand, are usually found on the outside of proteins. It is these interactions that determine how stable the folded structure is.

1.2 Protein stability

The first folded structure of a protein was solved by Perutz and Kendrew in 1958 [2]. However the idea that proteins folded into specific 3D structures was first suggested in the 1860s by Hoppe–Zeiler from studying hemoglobulin crystals [2]. For proteins to be able to carry out important functions in the cell they typically need to be folded [11]. Proteins can be unfolded by small changes to its environmental conditions such as an increase in temperature or a decrease in pH. Understanding how the protein maintains its folded state at physiological conditions is important. One way to quantify the stability of the folded protein is by determining the change in free energy associated with the reaction[11]:

$$\text{Unfolded}(U) \longleftrightarrow \text{Folded}(F) \tag{1.2}$$

1.2.1 Gibbs free energy and the protein stability problem

When a polypeptide folds many interactions form and the structural conformation of the polymer changes. This causes energetic changes in the system. The free energy of a system is a measure of how favourable a state is; the most favourable state of a system will have the lowest energy. For a reaction, if a change in free energy between two states is negative, a reaction will occur spontaneously. On the other hand if the difference in free energy is positive the reaction will require energy to occur. In the context of protein unfolding the conformational stability of the protein is defined as the change in free energy of the unfolded protein and the free energy of the folded protein. This change in free energy is depicted in Figure 1.2.0.1 and is the difference between the two minima in the energy profile of the protein.

The change in free energy is defined as the work required to transfer a body from one state to another whilst the body exchanges heat with the environment [2]. For a system at constant pressure (isobaric), and constant temperature (isothermal), measuring the

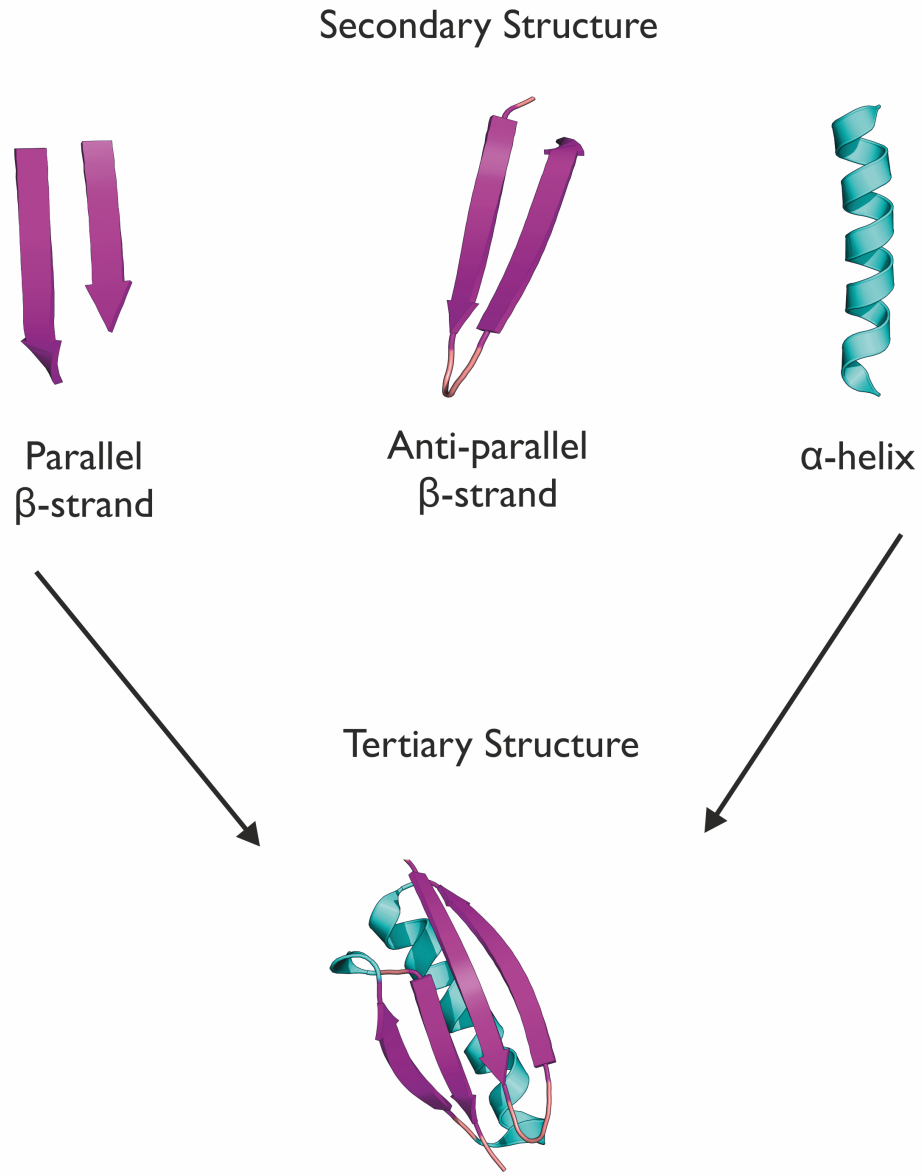


Figure 1.1.0.8: Schematic showing different types of secondary structure in proteins: the parallel β-strands shown by two arrows in the same direction, the anti-parallel β-strands shown as arrows in opposite directions and the α-helix shown as a twisting ribbon. These secondary structures can form in different parts of the polypeptide chain. The arrangement of secondary structures defines the tertiary structure of a protein. The secondary and tertiary structures of the protein (PDB ID 1HZ6 [8, 9] was made using the softwater PyMOL [10]

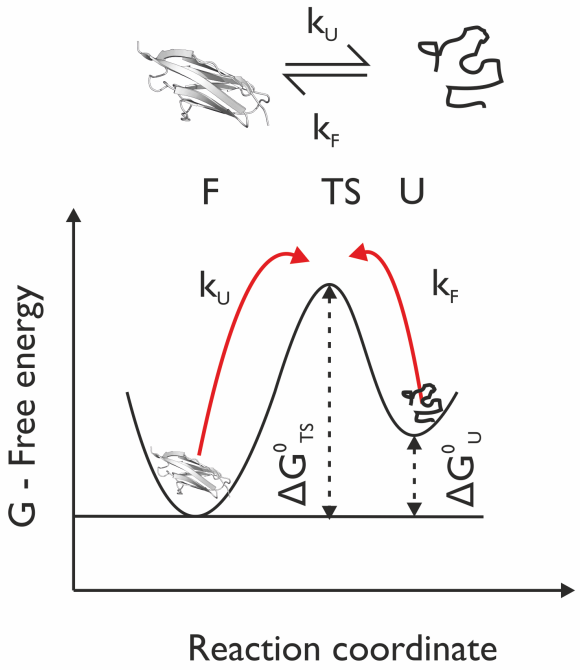


Figure 1.2.0.1: Schematic of a two-state energy landscape of a protein. Top: The forward arrow indicate the reaction of the folded (F) protein being unfolded (U). The probability of this reaction is governed by the rate of unfolding (k_U). The backwards arrow is the folding reaction of the protein. The probability of this reaction is governed by the rate of folding (k_F). Bottom: Schematic of a plot of the free energy of the system against the reaction coordinate being followed. The reaction coordinate is the parameter being probed when the system is perturbed. For example, in single molecule force spectroscopy the reaction coordinate is the distance between the N- and C- termini. There are energy minima in the landscape where protein can be in the folded (F) or unfolded (U) state. The difference in the free energy of these two states, $\Delta G_{U\text{-}F}$ determines the stability of F. Proteins can sample the energy landscape around the energy minima. A barrier to unfolding separates the two states of the protein. This barrier is known as the transition state (TS) of the protein and the height of this barrier relative to the F state is $\Delta G^0_{TS\text{-}F}$.

change in Gibbs free energy, Δ G, can determine whether a reaction is favourable or unfavourable and is defined as[2]:

$$\Delta G = \Delta H - T\Delta S \tag{1.3}$$

where ΔH is the change in enthalpy of the system and ΔS is the change in entropy of the system. The change in enthalpy is defined as [2]:

$$\Delta H = \Delta E + P\Delta V \tag{1.4}$$

where ΔE is the change internal energy of the system and PΔV is the work done on the system at constant pressure. The entropy of a system is dependent of the number of available states in the system. Each state has an associated energy with is E_{state}. The greater the number of available states, the higher the entropy because the energy dispersion is greater. The entropy is defined as [2]:

$$S = k \times \ln \Omega \tag{1.5}$$

Where Ω is the number of accessible states of a single particle. A similar expression exists for a system at constant volume (isochoric) and constant temperature the change in Helmholtz free energy, ΔU, is defined as:

$$\Delta U = \Delta E - T\Delta S \tag{1.6}$$

The Gibbs free energy and Helmholtz free energy only differ by PV. In the case of liquids PV is negligible because the volume is small and the pressure is low, therefore the value is less than the thermal energy, kT, of the system [2]. There is therefore a negligible difference between the Helmholtz and Gibbs free energy. In this thesis, the change in free energy or simply free energy refers to the change in the Gibbs free energy of the system. There are a number of points to note from equation 1.3:

- A large decrease in entropy (and therefore the number of states available in a system) is unfavourable.
- A large increase in enthalpy (and therefore a large amount of energy required for the reaction to take place) is unfavourable.

Many proteins can be repeatedly unfolding and refolded. If a single barrier exists between the folded and unfolded state the reaction is known as a two-state process. Figure 1.2.0.1 shows an example of two-state energy landscape. Most proteins have more complex landscapes [12], where proteins fold via stable intermediates (Figure 1.2.1.1) due to multiple barriers along the reaction coordinate probed. In fact protein energy landscapes are actually hypothesised to be funnel shaped (Figure 1.2.1.2) [12, 13]. The proteins are biased to the folded state and therefore many intermediates and folding pathways exist causing the energy landscape to be rough.

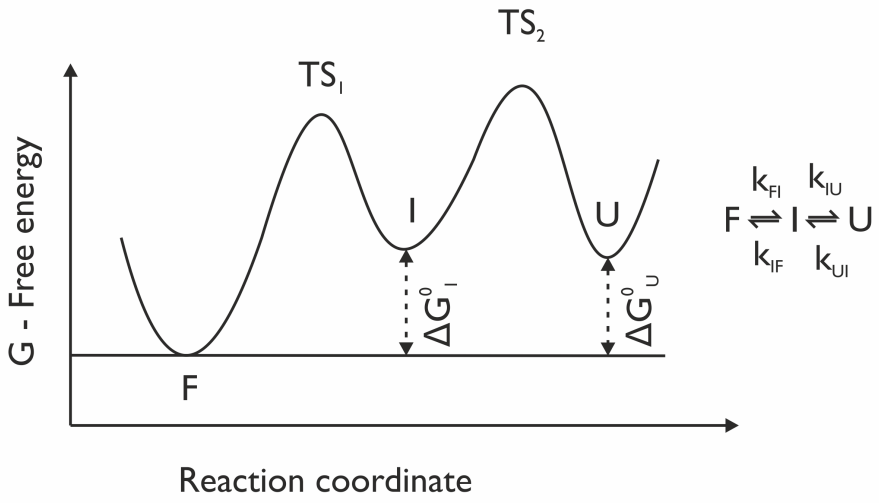


Figure 1.2.1.1: Schematic of the free energy, G, of a protein against the reaction coordinate for a protein that populates three-states during folding. In a three-state energy landscape there are three energy minima in the profile corresponding to a folded (F) protein, a stable intermediate (I) state and the unfolded (U) polypeptide chain. There are two energy barriers to unfolding. The first, TS_1, is the barrier between the F and I states. The height of this barrier is $\Delta G^0_{TS\text{-}F}$ with respect to the F state. The second barrier, TS_2, separates the I and U protein states. There are energy minima in the landscape where the protein can be in the folded (F) or unfolded (U) state. The probability of the protein going from one state to the other is determined by the rate of the reaction (right). For example, the probability of the protein going from the F to the I state is determined from the rate of F to I transition k_{FI}. Therefore, there are two rates associated with the protein going from the F to U state (k_{FI} and k_{IU}).

For protein folding to be spontaneous, the free energy between the unfolded and folded state, $\Delta G_{U\text{-}F}$, must be negative [2]. This energy difference is the difference between: the total energy of the interactions between the unfolded protein with the solvent and the energy of the interactions when the protein is folded. When a protein is unfolded there are multiple conformations available to the polypeptide chain; when a protein folds it has a fewer conformations available to it [14]. This results in a decrease in the entropy of this

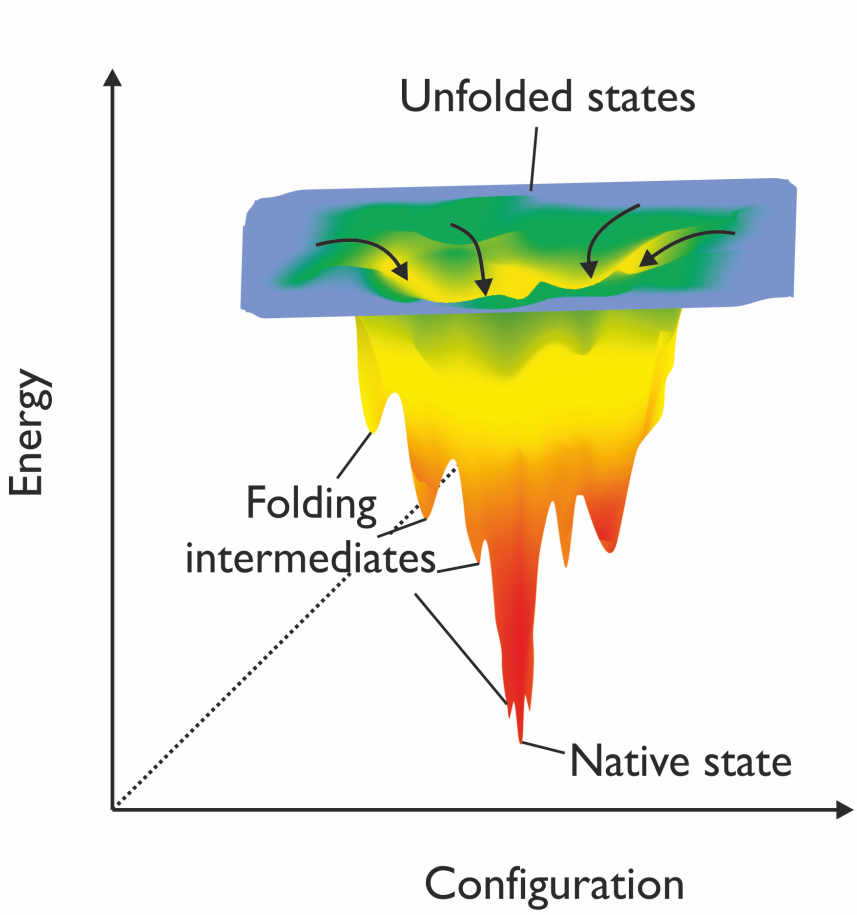


Figure 1.2.1.2: Schematic of the 3-D energy landscape profile of a protein, where the y axis is the potential energy and the x-axis is the configuration of protein. When a protein is in the unfolded state, the protein has a larger amount of conformational freedom in 3-D conformational space. As the protein folds this freedom decreases which reduces the conformational freedom of the chain. This decrease is shown by the transition of colours from blue to red. Where blue indicates multiple states (or conformations) available to the protein and red indicates a single conformation. Proteins can sample many conformations between the unfolded and folded states. The protein folding landscape represents a funnel where the protein folding reaction, indicated by black arrows, can occur via many pathways. The protein therefore can form many unfolding intermediates between the unfolded and folded states, these are indicated by small local minima along the folding pathway. As the free energy of the systems decreases, the protein tends towards the native folded protein state of the protein. This native state corresponds to the global minimum in the schematic. If no folding intermediates were present, the energy landscape of the protein would illustrate a smooth funnel leading to the native state.

system. This entropy is known as the conformational entropy of the system, ΔS_{conf} and contributes unfavourably to the ΔS in Equation 1.3 [11, 14]. When a protein is unfolded, amino acids are free to rotate about the peptide bond leading to multiple conformations. However when a protein folds the number of dihedral angles available to the backbone and side chains reduces (see Section 1.1.0.2). This reduces the number of states available to each amino acid. It has been estimated that each amino acid contributes $\sim$ 7 kJ/mol to the ΔG due to the increase in ΔS_{conf} [11, 14]. In fact experimental evidence has suggested that the entropic penalty at 25 °C is 5.7 $\pm$ 0.6 $kJmol^{-1}residue^{-1}$, which is consistent with theoretical predictions [15, 16]. As the size of the protein increases this entropy penalty will be more significant. Proteins with enhanced stability have also been engineered by reducing this conformational entropy [17]. Therefore for protein folding to be spontaneous there must be favourable contributions to ΔH and $T\Delta S$ that compensate the unfavourable ΔS_{conf}.

The favourable contribution to ΔH arises from non-covalent interactions between parts of the polypeptide chain. These interactions include: hydrogen bonds, Van der Waal's interactions, salt bridges and ion pairs [2, 11, 18, 19]. The hydrophobic effect, discussed in Section 1.4.3 contributes favourably to the $T\Delta S$ [14, 18]. Hydrophobic interactions and hydrogen bonds are thought to be two of the most important contributions to protein stability [11]. For example, Figure 1.2.1.3 shows the predicted contributions of different interactions to the protein Ribonuclease Sa. This protein is only 96 amino acids long. The destabilising contribution of ΔS_{conf} towards protein folding is predicted to be $\sim$ 700 kJ/mol. This is counteracted by the combination of hydrogen bonds ($\sim$ -350 kJ/mol) and hydrophobic interactions ($\sim$ -400 kJ/mol). These competing effects result in a small, predicted, negative $\Delta G \sim$ -50 kJ/mol which is two times greater than the measured ΔG of the protein ($\sim$ -25 kJ/mol). This small ΔG is typical for proteins 21 – 42 $kJmol^{-2}$ [18]. This is a small difference compared to the observed value of their entropy and enthalpy [17, 20]. It is therefore challenging to understand why proteins remain so unstable despite the ability to evolve to form more stable constructs [11, 21]. This suggest the small stability of the protein could play an important role in the proteins function [22], for example to provide flexibility in enzymatic catalysis[23]. Understanding the contribution of different interactions to protein stability can help design and engineer proteins with increased stabilities or optimised functions. It can also aid with the understanding of how proteins misfold and aggregate [24]. In this thesis, two interactions are going to be discussed in detail: Hydrogen bonds and hydrophobic interactions. Hydrogen bonds play an important role in the stability, flexibility and dynamics of proteins. Furthermore

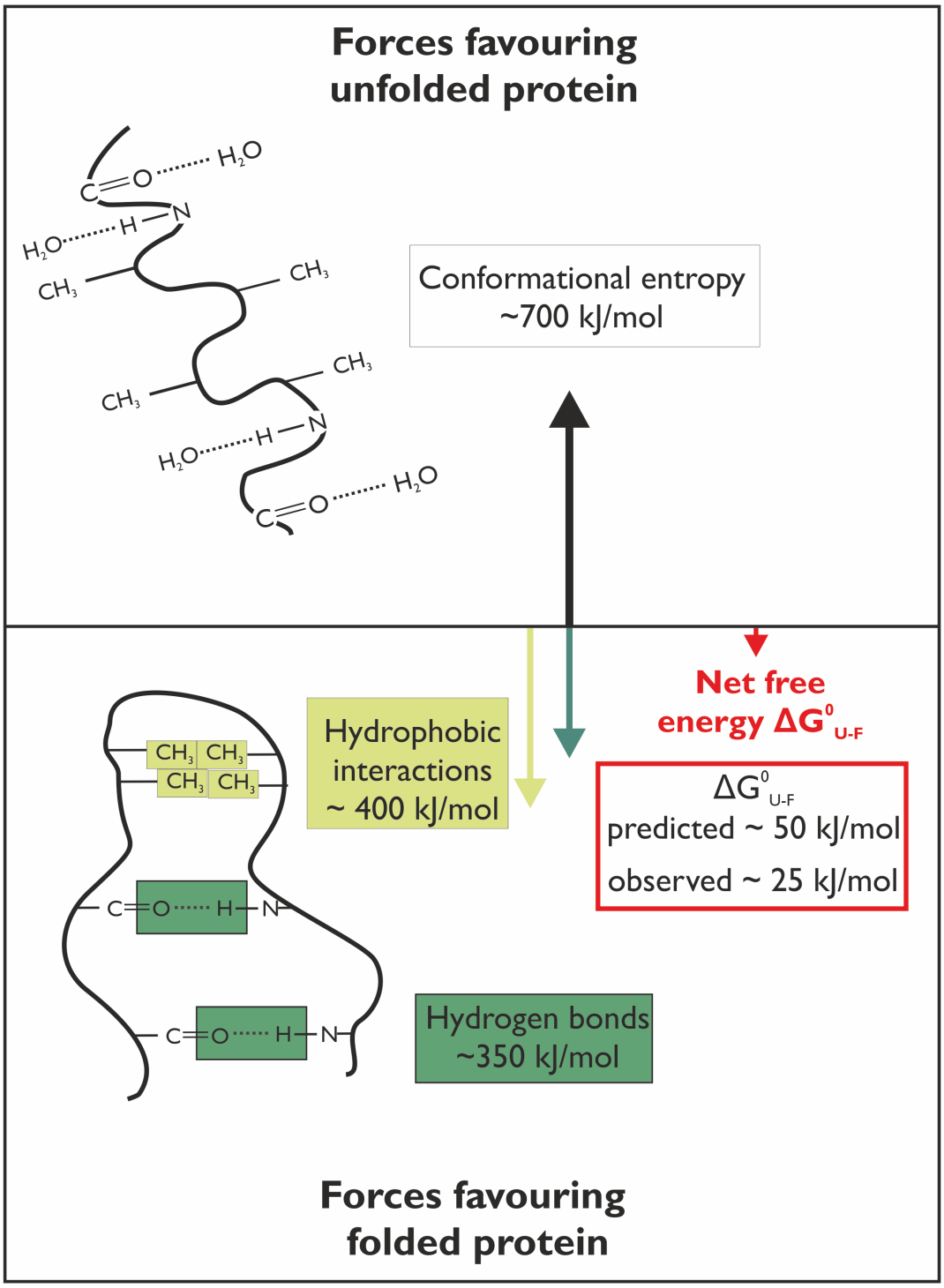


Figure 1.2.1.3: (a) Schematic of the protein folding reaction showing the formation of hydrogen bonds and hydrophobic interactions. (b) Figure showing the negative and positive contributions to free energy of the folded protein Ribonuclease Sa (25 °C, pH 7). Conformational entropy destabilises the folded protein whereas hydrogen bonds and hydrophobic interactions contribute favourably to protein stability. The values are taken from experimental studies [14]. Figure and values based on Figure 2 (a and b) Pace, C. N., Grimsley, G. R., Scholtz, J. M. and Shaw, K. L. 2014. Protein Stability. eLS.

they underpin the the structures of proteins. Hydrogen bonding within the solvent is also the reason the interior of a protein is mostly made of hydrophobic residues. These hydrophobic interactions are therefore important for protein folding and determining the tertiary structure of the protein.

1.3 Hydrogen bonds

1.3.1 Definition and history of hydrogen bonds

Interaction	Strength (k_BT)	Ref.
Covalent bonds	60–380	[25]
Ionic bonds	>400	[25]
Hydrogen bonds	1.6–5.2	[26]
Van der Waals interactions	0.8–1.6	[26]
CH - π bonds	0.9–1.5	[27]

Table 1.3.1.1: Table listing the different interactions found between atoms and molecules and their approximate strengths at 298 K in k_BT which is the Boltzmann constant times temperature.

Hydrogen bonds (HBs) are non-covalent interactions that are weaker than covalent and ionic interactions but stronger than Van der Waals interactions. Table 1.3.1.1 illustrates the difference in strength of a number of different interactions. Some of the earliest suggestions about the existence of these non-covalent interactions date back to 1902 [28]. The term "hydrogen bond" was first coined in 1912 by Moore and Winmill to describe the weak dissociation of a basic compound [29–31]. In 1920 interactions between hydrogen atoms and "free" pairs of electrons were used to describe the interaction between water molecules [31, 32]. Their importance in the structure of proteins was first made apparent by Pauling *et al.*, who predicted the existence of α-helical and β-sheets in the secondary structures of proteins based solely on the knowledge of the peptide bond and the existence of HBs [7, 33, 34].

The classical definition of the HB[30] states that they are interactions between a hydrogen atom bound to an electronegative atom, known as the hydrogen bond donor (D), and another electronegative atom, called the hydrogen bond acceptor (A) [37]. This is often written as D-H$\cdots$A. The electronegative atoms include oxygen (O) and nitrogen (N). The electronegative donor atom is able to attract the electron density from the hydrogen atom, leaving it with a slightly positive charge. The partially charged hydrogen atom is free to interact with the lone pair of electrons on another electronegative atom. It has been argued by Pimetal *et al.*, however, that hydrogen bonds exist between any

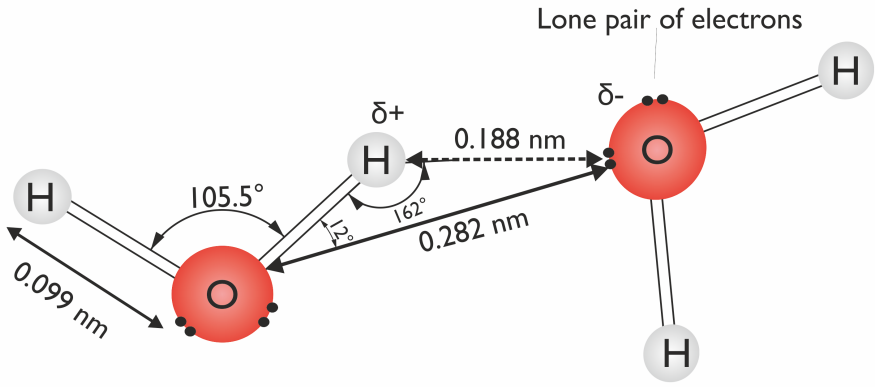


Figure 1.3.1.1: Figure illustrating a hydrogen bond between two water molecules. The oxygen is electronegative atom, therefore it is able to attract the electron density away from hydrogen. This leaves the hydrogen atom with a partial positive charge ($\delta+$). The oxygen has two pairs of free electrons (black dots) which are able to interact with the hydrogen atom. This interaction is known as a hydrogen bond. The parameters are taken from [35, 36].

D-H$\cdots$A , as long as the bond involves a hydrogen already bonded [38]. Figure 1.3.1.1 illustrates a hydrogen bond between two water molecules with dimensions calculated from *ab initio* studies [35, 36]. The strength of a HB depends on the angle and distance between atoms [2, 35]. The donor angle is measured between O-H$\cdots$O. If the angle is 180°, the strength depends linearly on the distance between the atoms, with the shortest bonds being the strongest [35]. The HB strength is less sensitive to the acceptor angle [2]. The favourable ΔG associated with hydrogen bond formation between water molecules is estimated to be $\sim$ -5.7 kJ/mol at 25 °C [35]. Therefore breaking of these bonds is energetically unfavourable. Temperature and pressure have been found to affect the hydrogen bond strength [39, 40].

Proteins contain many atoms that are capable of forming HBs. The amide and carbonyl groups on the backbone and polar groups on the amino acid side chains can participate in HBs. The amino acids that can be involved in hydrogen bonding are the polar amino acids shown in Figure 1.1.0.2. Proteins have been found to form 1.1–1.2 HBs per residue when a protein folds [41, 42]. The majority (68.1 %) of these occur between the amide and carbonyl groups of the backbone [19, 41, 43]. However, 31.9 % of the hydrogen bonds involve the polar side chains of the amino acids [41, 43]. HBs are therefore important components of proteins.

1.3.2 The role of hydrogen bonds in protein structure

In Section 1.1.0.3 different types of protein secondary structure were introduced. When a protein is unfolded HBs can occur between the polar water molecules and the polar

atoms in the protein chain. When a protein folds, the atoms in the peptide backbone can no longer interact with water molecules. To overcome the loss of these interactions, intra-molecular HBs are formed between parts of the protein backbone. These HBs underpin the secondary structure of the protein: The protein chain maximises the number of hydrogen bonded pairs by forming the secondary structure elements. Figure 1.3.2.1 illustrates the HBs formed within the different types of secondary structure. The pitch of a helix is determined by the HB pattern in the protein chain. The α-helix is a 4_{13}–helix which means that the the n and the n+4 residue is involved in the hydrogen bond and 13 atoms are involved in the loop [2]. In β-structures, HBs form between opposite amide and carbonyl groups in the backbone.

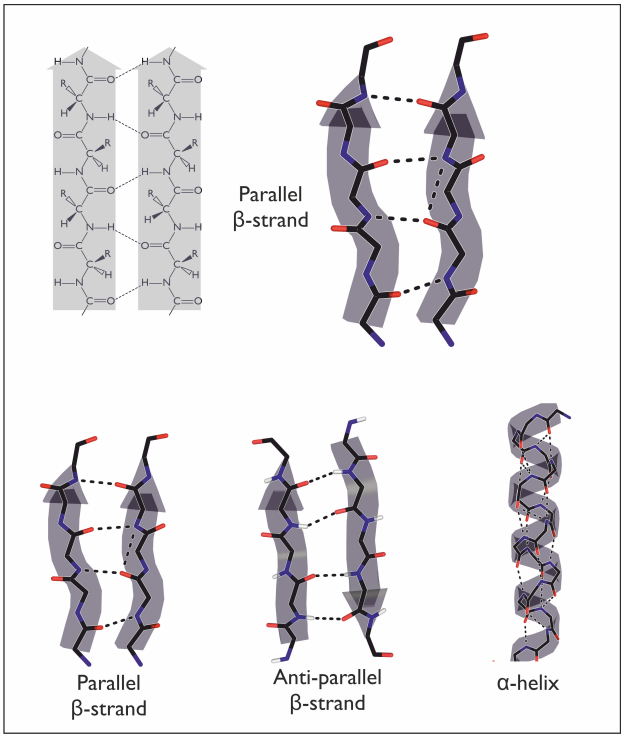


Figure 1.3.2.1: Schematic indicating the hydrogen bonding between the carbonyl and the hydrogen from the amide groups in the protein backbone. The hydrogen bonds are illustrated by dashed lines. The oxygen of the carbonyl is coloured red and the nitrogen of the amide group is coloured blue. The carbon backbone is coloured black. These hydrogen bonds underpin the structural stability of α-helices and β strands. β-strands can be arranged in anti-parallel or parallel geometry based on the chain direction.

Figure 1.3.2.2 shows examples of the classifications of HBs that can occur in protein systems. HBs in a protein mostly occur between the backbone of the protein (68.1 %), but also form between the backbone and sidechains (20.3 %) and between side chains (10.6 %) [41, 43]. Side chain HBs help stabilise the tertiary structures of proteins. The strength of the HBs within a protein will be variable due to the large range of bond lengths and bond angles that can occur. Baker *et al.* have suggested a maximum distance

between the hydrogen - acceptor of 2.5$\AA$ and a donor angle between 90 and 180 °C [44].

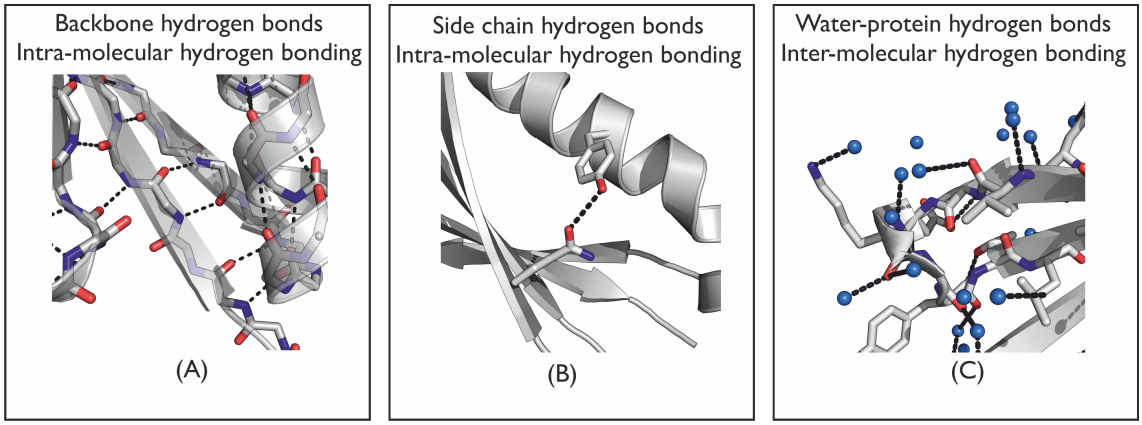


Figure 1.3.2.2: Figure illustrating hydrogen bond networks within protein L (PDB ID:1HZ6). Hydrogen bonds, shown as black dashes, can form between the amide (blue) and carbonyl (red) backbone groups (a), between polar side chains (b) and between the protein and water molecules (blue spheres) (c).

1.3.3 Role of hydrogen bonds in protein stability

Polar side chains (up to 65 %), peptide groups (up to 70%) and charged groups (up to 54 %) in proteins that are able to HB are buried[4] during protein folding [45]. Many studies have been carried out to try and elucidate the contribution of HBs to protein stability [19, 41, 46–48]. Despite this, it is still not fully understood how HBs contribute to protein stability [19]. To determine whether HBs are more stable within the folded protein, or in the unfolded protein it is useful to study the change in free energy associated with the reaction [19]:

$$(\text{–NH}\cdots\text{O=C–})_{\text{protein}} + 2\text{H}_2\text{O} \longleftrightarrow \text{H}_2\text{O}\cdots\text{HN–} + \text{–C=O}\cdots\text{H}_2\text{O} \tag{1.7}$$

Whilst some experimental[19, 41, 47–52] studies have suggested this reaction contributes favourably to the thermodynamic stability of proteins, other theoretical [53–56] and experimental [57] studies have suggested that this reaction is unfavourable or contributes little to protein stability. This is attributed to the loss of water-hydrogen bonds upon protein folding. However, studies show that the likelihood of an unsatisfied HB is unlikely [58].

In general, HBs have been shown experimentally to contribute 4.6 ± 3.4 kJ/mol to the thermodynamic stability of proteins [46]. This is almost 25 % of the protein's folded

[4]This means that these amino acids, or parts of the amino acid, can be removed from contact with water

stability. The large error associated with this value is attributed to the context of the HB environment [14, 19, 46]. When polar groups are buried upon folding, they are often surrounded by both polar and nonpolar groups. It has been suggested that the burial of the polar groups can contribute more to protein stability than nonpolar groups [59, 60]. In the nonpolar environment of the core of the protein, the dielectric constant is lower, therefore HBs have been found to be up to 5.0 kJ/mol stronger than if they are solvent exposed [59, 60]. Furthermore, HBs between the protein and the solvent have also been found to contribute favourably to protein stability (4.8 $\pm$ 0.7 kJ/mol)[52]. The location of hydrogen bonds are also thought to play a key role to the flexibility of proteins [61]. Livesay *et al.* (2008) determined a correlation with the rigidity of a hinge region within a protein and the number of HBs. The hinge had fewer HBs in an open conformation, leading to additional flexibility in this region.

The HB network within water has also been found to play an important role in the flexibility, structure and dynamics[62–66] and, therefore, the function [67–69] of proteins. Indeed, in the absence of water, proteins have been shown to lose their activity [70, 71]. The structural changes in the polypeptide chain associated with protein folded has been hypothesised to be facilitated by hydrogen bonding between the protein and the water [65, 72, 73]. Simulations of protein folding have shown that the majority of protein structure is formed before water is expelled [74]. Water is "squeezed" out from the structure by peptide hydrogen bonding at the final stages of protein folding. Neutron diffraction experiments have also suggested that water acts a nucleation side between sections of a polypeptide during the formation of a β-turn [72]. The water molecules in this study were found to mediate conformational changes during folding and to stabilise interactions within the peptide [72].

Water forms a hydration layer around proteins. In the first hydration layer the molecules can interact with the protein via Van der Waals interactions or HBs [66, 70, 75–78]. This hydration layer is dynamic and exchanges with bulk water due to thermal fluctuations on the surface of the protein [70]. The structure of the protein can be maintained by this hydration layer. For example the unfolding of a protein has been coupled with a loss of the network of water hydrogen bonds surrounding the protein [79]. Furthermore, conformational changes in the protein have been shown to be coupled with the dissolution of large clusters of hydrogen bonded waters into smaller clusters [80]. The hydrogen bonding network within the water must therefore by dynamic and flexible to facilitate structural changes associated with the function of the protein [81]. Studies have

suggested that other solvents would not be able to facilitate biological processes [82].

Understanding the role of protein-protein and protein-water HBs is therefore crucial to understanding how proteins remain stable. It is also important to determine experimentally how hydrogen bonds in the protein and or solvent affect protein flexibility and dynamics. Few studies have been performed to determine systematically how the HB network of the solvent and protein affects the transition barrier to folding. The tertiary structure of proteins is also believed to be "glued" together by hydrophobic interactions. In fact, some studies have suggested that hydrophobic interactions provide the major energetic contribution to protein stability [83]. However, the importance of hydrophobic interactions are difficult to probe due to the inhomogeneous core of the protein.

1.4 Hydrophobic interactions

Hydrophobic interactions refer to the aggregation of nonpolar solutes to minimise their interaction with water. A nonpolar solute is defined as hydrophobic if it has a poor solubility in water and a good solubility in nonpolar solvents [84]. Ideas about the hydrophobic effect started as early as 1937 [85], in this study it was postulated that the driving force behind the hydrophobic interaction would be entropic due to the temperature dependence of the interaction. The importance of hydrophobic interactions in the structure of proteins was then introduced in 1939. Hydrophobic residues were hypothesised to "hold" the protein together, not due to attractive interactions, but due to repulsion of water molecules [86]. However, the importance of hydrophobic interactions in biological systems did not become widely accepted until 1959 after a review by Kauzmann [87]. In this review, Kauzmann predicted hydrophobic interactions to be a major driving force in protein folding. This review was controversial due to the lack of protein structures existing at that time and the widely accepted idea that bonds of a certain length and angle would be responsible for protein folding [88]. The term hydrophobic bonds was coined in this review, however the hydrophobic effect was introduced by Tanford in 1980 to describe this phenomenom. This term is more widely accepted due to issues with the term "bond" referring to interactions that did not have any characteristics of chemical bonds [84, 89]. These interactions are not driven by the interactions between the atoms involved, but the medium surrounding them [2]. Furthermore, there are favourable Van der Waals interactions between "hydrophobic" solutes and water [90].

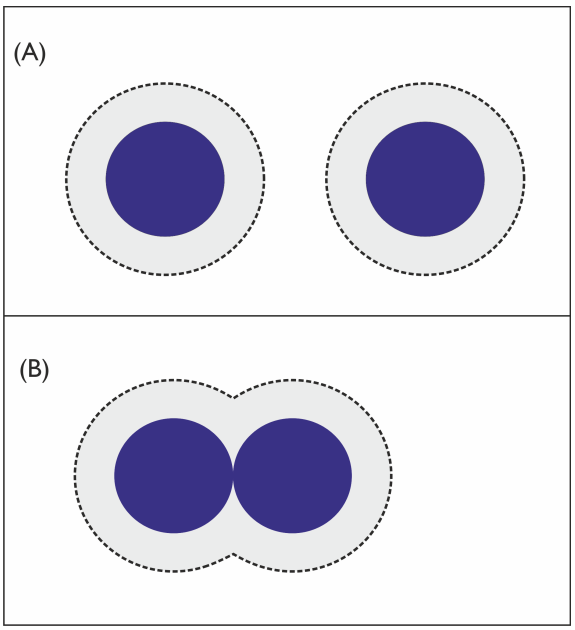


Figure 1.4.1.1: A schematic of the Van der Waals radius of a nonpolar solute (blue circle). The solvent accessible area is the area surrounded by the dashed line encompassing the nonpolar solute. Two individual non polar solutes have a larger accessible area to a water molecule (a) compared to the surface area if they cluster together (b).

1.4.1 What drives the hyrophobic interaction?

When a nonpolar molecule (solute) is introduced to water (solvent) a larger volume is accessible to the water molecules. Typically nonpolar molecules are considered to be hydrophobic because they are unable to form hydrogen bonds. Water molecules are unable to interact with this molecule. Therefore, to reduce the solvent accessible surface are (SASA)[5], non polar molecules are clustered together. As Figure 1.4.1.1 shows, this significantly reduces the SASA available to the water molecules. As the nonpolar solute increases in size more volume will be solvent accessible. This is the reason smaller nonpolar solutes are easier to dissolve in water than long chain molecules [84]. To determine the driving force behind the hydrophobic interaction it is useful to look at how the free energy of the system changes when nonpolar solutes are added to water.

To determine the ΔG from introducing a nonpolar solute to water, the nonpolar molecule is typically considered to move from a nonpolar liquid to water. The free energy of transfer ($\Delta G_{\text{nonpolar-water}}$) of particle A from phase the α (nonpolar environment), to the phase β (water) is given by Eqaution 1.8 [91].

$$\Delta G_{nonpolar->water} = -kT \ln \frac{c_A^\beta}{c_A^\alpha} \tag{1.8}$$

where c_A^β is the equilibrium concentration of the molecule A in phase β and c_A^α is the concentration in phase α for a given volume. The $\Delta G_{\text{nonpolar-water}}$ is deduced from the

[5]The SASA is the area described by the centre of a 1.4 $\mathring{A}$ sphere that rolls over the Van der Waals envelope of a molecule [2]

interactions of the molecule with the surrounding solvent and includes the change entropy and enthalpic interactions that occur with the reactions (see Section 1.2.1.

This reaction for a nonpolar solute, such as a hydrocarbon, is associated with a positive ΔG [2, 84, 92, 93]. This is typically several times k_BT, therefore the process is energetically unfavourable [2, 91]. The hydrophobic interaction is also more pronounced for larger nonpolar solutes [94]. For nonpolar solutes the free energy changes, per the SASA of the nonpolar side chain, by $\sim$ 0.02 kJmol^{-1}Å^{-1} [94]. This accounts for the low solubility of hydrocarbons at room temperature. When a non polar solute is transferred to water, it is expected that hydrogen bonds would be broken between the water molecules to encompass the solute. This would result in a large increase in the enthalpy of the system and would be energetically unfavourable. However, the ΔH, at room temperature and pressure, of the system has been shown to be small and positive in some experiments, and negative in others [95, 96]. The ΔS, on the other hand, is always large and negative [95, 96]. For example, for methane, the $\Delta G_{\text{nonpolar-water}}$ is $\sim$ 25.5 kJ/mol, the ΔH is $\sim$ -19.7 kJ/mol but the $T\Delta S$ was $\sim$ -45.2 kJ/mol [95]. The large positive $\Delta G_{\text{nonpolar-water}}$ is therefore driven by unfavourable entropic forces. Based on the definition of ΔS, this suggests a decrease in the number of states in the system. This change in entropy is also coupled with an increase in heat capacity in the system. Heat capacity is defined as[89]

$$\Delta C_p = \frac{d\Delta H}{dT} = T\frac{\Delta S}{dT} \tag{1.9}$$

A large heat capacity suggest, ordering of water molecules within the solvent and strengthening of intermolecular interactions [91]. The heat capacity is determined from the change in free energy with temperature. The driving force behind hydrophobic interaction is therefore temperature dependent. As the temperature increases, the ΔS increases but the ΔH also increases. At high temperatures the hydrophobic interaction is driven by enthalpic interactions[89]. Models have been developed to try and explain the driving force of the hydrophobic interaction and its temperature dependence. These will be discussed in the next section.

1.4.2 Models to explain the hydrophobic interactions

The driving force behind the hydrophobic interaction is under debate. The classical model for hydrophobic interactions is known as the "ice-berg" model (Figure 1.4.2.1). The model was first suggested by Frank and Evans in 1945 [97]. In this description of the hydrophobic interaction, the large negative $T\Delta S$ from hydration of nonpolar solutes

and is explained by the ordering of water molecules around the nonpolar solute. The decrease in ΔS occurs due to the formation of a "clathrate-like" shell of waters at the surface of the nonpolar solute. The waters rearrange themselves to form hydrogen bonds with other water molecules and minimise their interaction with the non-polar solute. This constrains the number of possible orientations of the water molecules and partially "freezes" the motion of water molecules at the surface of the solute. The enthalpy gain by loss of hydrogen bonding is minimised but the penalty is an increase in entropy. It is estimated that the loss of a hydrogen bond energetically costs $\sim$ 21 kJ/mol whereas the loss of entropy is $\sim$ 1 kJ/mol [2]. As the temperature increases initially the "shell" of waters is maintained but broadens, increasing the entropy. This broadening of the shell causes hydrogen bonds to break thus increasing the enthalpy. This model explains the positive heat capacity attributed to the hydration of nonpolar solutes [89]. These "clathrate" cages have been observed in crystalline hydrates [97, 98]. Furthermore some experimental evidence suggests that there is a propensity for the OH of a water molecule to be frozen at the surface of a nonpolar solute [99]. However, the water molecules are not "frozen" in this model; freezing of the water molecules would result in a decrease in molar volume, whereas introduction of nonpolar solutes to water have been shown to decrease it[89]. Theoretical studies have supported the idea of more ordered waters in the water shell around a nonpolar solute [100]. However, it has been suggested that this ordering would occur over a very narrow temperature range [89, 101]. There is strong experimental [102, 103] and theoretical[104] evidence to challenge this model. Some experimental studies have suggested that the water molecules at the surface of a nonpolar solute are less ordered than free water molecules [102].

Experiments suggest that, despite no evidence for the ordering of water, the dynamics of water molecules at the surface of the nonpolar solute is "ice-like" and some waters can be immobilized at the surface [99]. This view has also been challenged [104]. Molecular dynamics simulations have shown that, whilst the rotational dynamics of the water molecules are hindered around the nonpolar solute at low temperatures[101], the waters are never immobilised [104].

Since this first model was proposed, others have been proposed to explain the thermodynamics of hydrophobic interactions. Lee [96] and Lucas [105] suggested that hydrophobic hydration is a result of the energetic penalty of finding a cavity large enough to encompass the nonpolar solute. In this model the large entropic cost occurs from the creation of a large cavity for the solute from small pockets of free volume within the

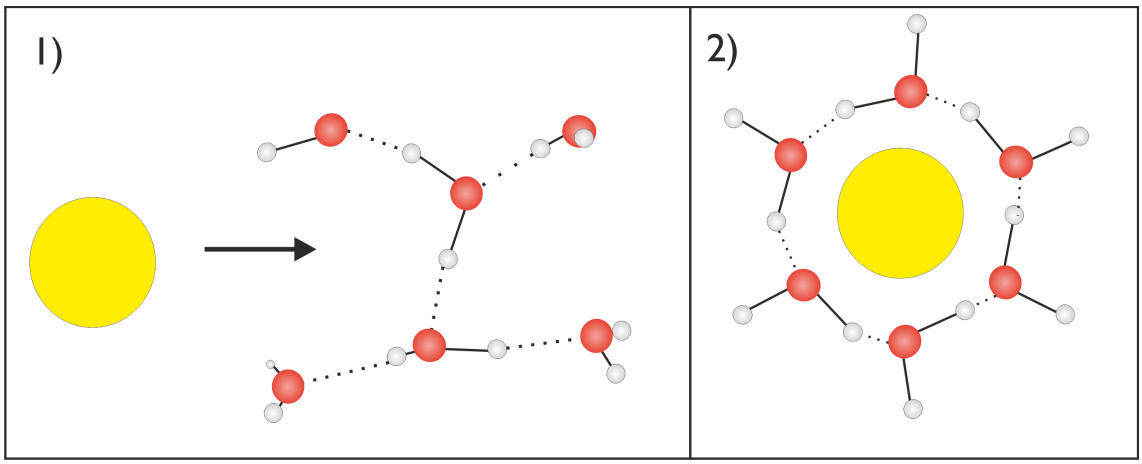


Figure 1.4.2.1: Schematic illustrating the theory behind the "ice-berg" model for hydrophobic interactions. (1) When non polar solute is added to water, the hydrogen bonding network of the water molecules will be disrupted. (2) The waters reorientate themselves at the surface to minimise the loss of hydrogen bonds. This freezes the motion of the water molecules causing the local entropy to increase.

solvent. If the molecules in the solvent are larger, the "pockets" of free volume are larger which would reduce the entropic penalty. On the other hand, the larger the solute, the larger the cavity that is required to encompass the molecule.

Alternatively, Stillinger, in 1973 [106], suggested that the hydrogen bond network of water can be maintained around small nonpolar solutes, however the entropic penalty arises from maintaining these hydrogen bonds. However, these cannot be maintained for large nonpolar solutes. For large non polar solutes, the hydrophobic interaction would be driven by enthalpy, contradicting the classical view that entropy is the underlying driving force. Studies have since highlighted the importance of lengthscale on the hydrophobic interaction when considering the driving force [103, 107–109]. Whilst water may be able to reorientate around small solutes, it would not be geometrically possible for the water to reorientate, as in Figure 1.4.2.1, around large molecules [93, 107]. There appears to be a length crossover of 1 nm between more ordered "hydrophobic" waters with stronger hydrogen bonds around small nonpolar solutes and disordered waters with weaker hydrogen bonding around larger nonpolar solutes [108, 109]. Simulations have indicated that the association between small nonpolar solutes is driven by entropy but the negative enthalpy of the reaction occurs due to the larger number of hydrogen bonds within the water surrounding the solutes [109]. For larger nonpolar solutes the waters are unable to maintain this hydrogen bonding network and therefore association is driven to minimise the loss of hydrogen bonds. More recently, it has been suggested that nonpolar solutes aggregate to replace high energy waters that are not hydrogen bonded [110].

1.4.3 Hydrophobic interactions in biological systems

All water-soluble globular proteins contain a hydrophobic core [111] formed by the hydrophobic residues indicated in Figure 1.1.0.2. This hydrophobic core is thought to play a critical role in protein stability and the of the tertiary, and often quaternary, structure of proteins [2, 111, 112]. Furthermore, hydrophobic interactions are thought to play a key role in protein folding [20, 87] and the evolution and adaptation of proteins [113].

(a) **(b)**

Figure 1.4.3.1: Figure showing the position of hydrophobic residues in two proteins. (a) the protein's hydrophobic residues are shown as green spheres. (b) the hydrophobic residues are shown as yellow spheres. In both proteins the hydrophobic residues are clustered within the centre of the tertiary structure. Figure made using Pymol [10] for PDB IDs [8] 1HZ6 [9] and 1TIT [114].

To help elucidate the contribution of hydrophobic interactions to protein stability, a number of studies have tried to model the hydrophobic core as the transference of nonpolar side chains from water to a nonpolar environment [83, 84]. This is thought to mimic the folding of proteins, where the non polar side chains are no longer in contact in water [84]. By studying the $\Delta G_{\text{water-nonpolar}}$ for amino acids, or chemical derivative, hydrophobicity scales for the 20 naturally occurring amino acids have been determined [115–118]. Studies have used nonpolar environments such as bilayers [119], liquid alkanes or alcohols [117, 120], which are slightly polar, to determine the hydrophobicity (or hydropathy scales). It is hard to determine which environment best mimics that of the protein core [83]. The hydrophobic core is slightly polar due to the enhanced Van der Waal's interactions in the tightly packed core [121]. Therefore there are often large discrepancies between the available hydrophobic scales. This is illustrated in Figure 1.4.3.2 which shows examples of the hydrophobicity rankings of the amino acids determined by different groups and methods. Whilst, there are some general trends in these scales, the inconsistencies make it challenging to elucidate the contribution of hydrophobic interactions to protein stability. Another method to determine the contribution is to

mutate non-polar amino acids in the protein. Experimental evidence from these studies has suggested that the hydrophobicity cyclohexane may best describe a protein core [122].

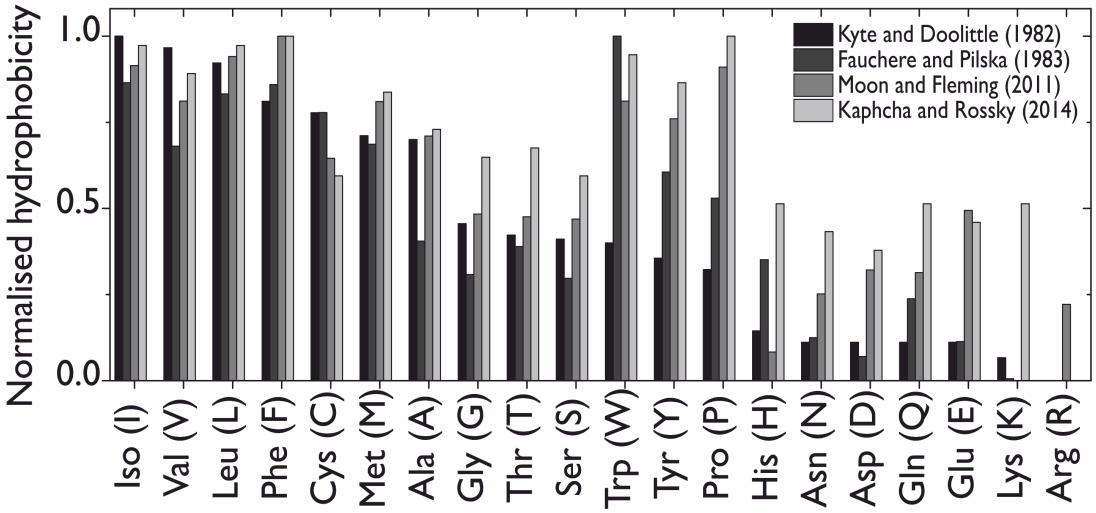


Figure 1.4.3.2: Graph showing the normalised hydrophobic scales from four groups for the 20 naturally occurring amino acids where 1 corresponds to the most hydrophobic value and 0 corresponds to the more hydrophillic amino acid. Kyte-Doolittle scale (black) is taken from ref. [116] is determined from The Frauchere-Pilska scale is taken from [2] (dark grey) and is determined from the transfer of amino acids from water to alcohol[120]. The Moon-Fleming scale from ref. [119] (lighter grey) is derived from water to bilayers. Finally the Kapcha and Rassky scale [115] is determined from a simple coarse grain computational model.

The results of such studies have highlighted the importance of core packing to the thermal, chemical and mechanical stability of proteins[112, 122–126]. It has been estimated, from chemical denaturation experiments, that the burial of a CH_2 group contributes $\sim$ 4.6 ($\pm$ 2.1) kJ/mol to the stability of a protein [122]. In a study by Pace et al. (2011), the hydrophobic interactions within a protein core provide more stability to the folded state of a large protein (containing 341 amino acids protein) than compared with small proteins (containing 36 amino acids) [122]. Furthermore, optimising the hydrophobic core packing within a protein can increase its thermodynamic stabilty [123, 127]. The hydrophobic core therefore must play an important role to the thermodynamic stability of a protein.

Understanding how hydrophobic interactions maintain a protein in the folded state is challenging. Hydrophobic interactions occur over large lengthscales that are larger than the distances that homogeneous molecules affect one another in liquids [93]. Studies have suggested that there is an exponential decrease in the strength of the hydrophobic interaction with increasing distance [128]. Furthermore, whilst single molecule studies have been used to look at hydrophobic interactions in nonpolar molecules [103], characterising the contribution of hydrophobic interactions to protein stability is challenging due to their

heterogeneous environment [129]. Therefore it is not well understood how hydrophobic interactions individually affect the transition barrier to folding. To determine the net contribution of the hydrophobic interaction to protein stability, the strength of these interactions would need to be manipulated. Determining the individual contribution of hydrophobic interactions and hydrogen bonds to the rate limiting step would need careful design of model systems. One method to do this is by using isotopes of hydrogen.

1.5 Deuterium

Isotopic substitution of hydrogen in bio-molecules using deuterium (D or ^{2}H) and deuterium oxide (D_2O) can be used as a method to test the bulk importance of HBs and hydrophobic interactions on the stability of proteins [130–132]. Deuterium was first discovered in 1932 by Urey *et al.*[133], it has a relative abundance of 0.015 % in the atmosphere[134]. It is an isotope of hydrogen with an additional neutron in its nucleus. This causes the molecular weight of deuterium (2.014) to be twice that of hydrogen [135], therefore D_2O is often referred to as "heavy water". This additional mass results in many physical and chemical differences between deuterium and hydrogen. This includes a smaller Van der Waal radius (and therefore small steric requirement) of deuterium and greater bond strength [136, 137]. The differences caused by isotopic substitution has a significant effect on the the solubility of nonpolar molecules and the stability of proteins, which will be discussed in this section.

1.5.1 An introduction to deuterium

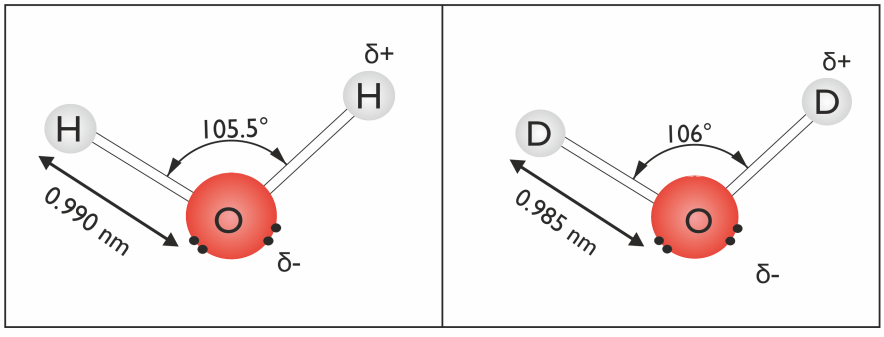


Figure 1.5.1.1: Geometries of the H_2O (left) and D_2O (right) molecules. Hydrogens are coloured white and the oxygen/deuterium atom red. There is slight differences between the lengths (0.005 nm) and angles (0.5 °) of the molecules. Values are taken from a neutron diffraction study comparing the two molecules [138].

Table 1.5.1.1 lists the some of the properties measured for both D_2O and H_2O. The

geometries of the two molecules is not significantly different, these are are shown in Figure 1.5.1.1. There is a 0.005 nm shortening of the O-D bond[138]. Both experimental[139] and theoretical [140] studies have determined that "hydrogen" bonds formed by deuterium are stronger than those in water by $\sim$ 0.4-0.8 kJ/mol. An increase of $\sim$ 0.6 kJ/mol in HB strength was determined from the change in excess enthalpy observed for transferring methanol to water or deuterium oxide[139]. Computational methods determined the increase to be $\sim$ 0.8 kJ/mol based on ab-initio calculations [140]. This latter study showed that the increase in hydrogen bond strength in gas is caused by a decrease in the zero-point vibrational energy of the bond (ZPVE) caused by the additional mass in deuterium [140]. Figure 1.5.1.2 illustrates the ZPVE and a schematic of the differences in the ZPVE for O-D and O-H bond. This lower ZPVE is a consequence of the additional mass of deuterium lowering the frequency of the bending motion of the bridging hydrogen or deuterium atom ($\sim$ 100 cm^{-1} [141]). This, in turn, reduces the frequency of the entire bond. The decrease in ZPVE also has an effect on the rate of cleavage (breaking) of the C-D, O-D and N-D covalent bonds. The rates of cleavage have been calculated to be 7, 10.6 and 8.5 times slower than their protiated homologues respectively [142, 143]. The ZPVE refers to the level with energy E=1/2(hν), where ν is the vibrational frequency and h is Plank's constant [144]. A C-D bond has a vibrational frequency of 2100 cm^{-1} and a C-H bond has a vibrational energy of 2900 cm^{-1} [145]. Therefore more energy would be required to break a C-D bond.

Chemical property	H_2O	D_2O
Melting point (°C)	0	3.81
Boiling point (°C)	100.0	101.4
Viscosity (millipoise, 25 °C)	0.895	1.113
Density (g/cm^3 - 25 °C 101 kPa)	0.997	1.104
Dielectric constant (25 °C)	78.304	77.937
Average number of hydrogen bonds (liquid)	3.62	3.76

Table 1.5.1.1: Properties of liquid H_2O and D_2O. Values for the melting points, boiling points, viscosities and dielectric constant are taken from [146]. Values for the viscosity are taken from [147]. Values for the average number of hydrogen bonds are taken from [148]

Despite the relatively small difference in the hydrogen bond strength, significant differences have been noted in the bulk properties of water (e.g. boiling point +1.8 °C, density $\sim$ +0.1 g/cm^3 [146]), including an increase in the average number of HBs [148]. There is some dispute to the geometry of the HB in D_2O. Studies have suggested that the bond is longer in D_2O (0.181 nm compared with 0.174nm), but more linear than H_2O[148], whilst others suggest that there is no change in hydrogen bond length due to competition of inter- and intra- molecular quantum effects [149].

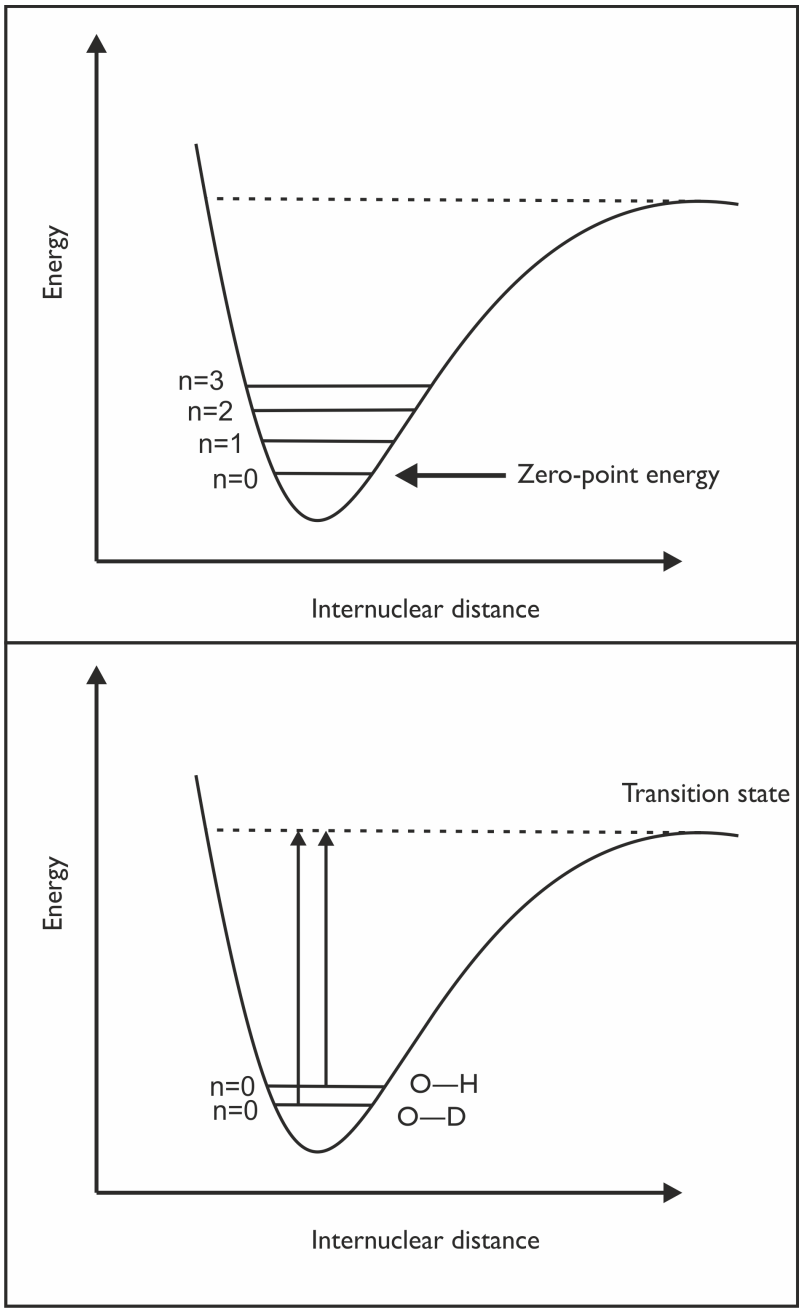


Figure 1.5.1.2: Top: Potential energy well for the bond cleavage of two atoms. The y axis is the energy and the x axis is the internuclear distance between the two atoms. The system has a distribution of vibrational energy levels. To energy required to cleave the bond from a particular energy level is given by: $E = (n + 1/2)\hbar\omega$. Where n is the vibrational quantum number (n=0,1,2,3...), ω is the vibrational frequency of the bond and $\hbar$ is Plank's constant n=0 is the lowest energy system, this energy level is known as the zero-point vibration energy level. Bottom: Differences between the zero-point vibrational levels of the O-H and O-D covalent bonds. The O-D sits deeper in the potential well due to the additional mass in the D nucleus. This causes the bond to vibrate at a lower frequency. Therefore more energy is required to cleave this bond.

1.5.2 Effect of deuterium oxide on hydrophobic interactions

The results for the hydrophobic effect in D_2O are somewhat controversial. An increase in hydrogen bond strength, and more extensively hydrogen bonded network in D_2O would suggest that the hydrophobicity of nonpolar solutes would be enhanced in D_2O. An increase in heat capacity was observed for the transference of nonpolar solutes from H_2O to D_2O [150].

This increase in heat capacity upon transference is typically indicative of a lower solubility of a solute. However, experiments have shown that for simple hydrocarbons and noble gases, heavy water is a better solvent than H_2O [151–153]. This means that a negative $\Delta G_{H_2O\text{-}D_2O}$ is associated with the transfer of a nonpolar solute to D_2O from H_2O. These molecules are therefore more soluble in D_2O than H_2O. This makes it challenging to explain these results in terms of the classical model for the hydrophobic interaction [150]. However, the hydrophobic hydration of small nonpolar solutes in D_2O can be explained by a lower packing density in D_2O. The negative $\Delta G_{H_2O\text{-}D_2O}$ arises due to the lower penalty for forming a cavity in D_2O. The increase in ΔH is observed due to the enthalpy fluctuations from reforming the HBs [150]. As for H_2O, the hydrophobic effect becomes more pronounced in D_2O with increasing temperature[152].

As expected, an increase in the size of the nonpolar solute causes an increase in the hydrophobic effect in D_2O [151, 154]. For larger nonpolar solutes and proteins, the hydrophobic interaction is found to be intensified in D_2O relative to H_2O [151, 154–156]. Studies on polypeptides in D_2O have shown a tendency for them to reduce their SASA by aggregating more readily. Furthermore aggregation in proteins where hydrophobic interactions are thought to be important has also been increased by the presence of D_2O. Some macromolecular complexes have also been shown to form a more compact globular shape due to increased hydrophobic interactions. The hydrophobic effect of amino acids in D_2O is more complex due to the presence of hydrophilic groups. Whilst alanine and phenylalanine are more soluble in D_2O, proline and glycine were found to be less soluble [157]. The solubility in phenylalanine reverses at temperatures greater than 37 °C. The $\Delta G_{H_2O\text{-}D_2O}$ for a number of amino acids with nonpolar solutes were also determined. The hydrophobicity of these amino acids were found to be enhanced in D_2O and are dependent on the length of the hydrophobic side chain [151]. This has consequences for the hydrophobicity of proteins however, the change in hydrophobicity in D_2O on protein sta-

bility are less clear with some studies reporting an increase in stability, whilst others show an increase or little change in stability. These results will be discussed below.

1.5.3 Deuterium and protein stability

1.5.3.1 Solvent isotope effects

Significant differences have been observed in the flexibility and thermodynamic, mechanical and thermal stability of biological systems in D_2O and H_2O [130–132, 158–160]. The thermal stability of a number of proteins have been found to be enhanced in D_2O [160–164]. For example the melting point of the protein tubulin has been found to increase by 3 °C. Furthermore, tubulin was found to remain active in D_2O for 14 hours more than in H_2O at 20 °C. At 37 °C, whilst the protein had lost full activity after 8 hours in H_2O, only 20 % of the activity was lost in D_2O. In another study the thermodynamic stability of two proteins, a "hard" (requires large amount of heat, enthalpy, to unfold per unit mass) protein and a "soft" (conformational adaptable) protein in were compared both solvents [161]. At every pH tested both proteins were found to be more stable in D_2O. Furthermore the increase in stability could be related to an increase in hydrophobicity of the nonpolar amino acids within the proteins [161]. The increase in hydrophobicity has also been observed for other systems. For example polypeptides have been found to reduce their SASA by aggregating more readily or forming more compact globular shapes in D_2O [131, 159, 165, 166]. In addition the concentration at which surfactants aggregate into micelles has been found to be lower in D_2O than H_2O [151, 152, 167]. The tendency for a protein to reduce its SASA in D_2O more readily, is supported by experiments illustrating that the protein forms a more compact form in D_2O [168, 169]. These studies determined that proteins have a smaller radius of gyration in D_2O[168], and a smaller hydrodynamic radius[169]. This would suggest an increase in hydrophobicity for large molecules due to differences in the solvation of polar and nonpolar amino acid side chains [131, 146, 158, 169, 170]. However, others have suggested the increase in stability of proteins might be explained by the increase in hydrogen bond strength within the protein[165, 171] or a reduction in the hydrogen bond dynamics in the solvent [172]. Molecular dynamics simulations of a small β-turn and α-helix have indicated that the rupture time of the water-protein hydrogen bonding is retarded in D_2O [172]. Hydrogen bond rupture in D_2O was coupled with a deepening of the energy well and an increase in the rupture barrier height. There was also a suppression of internal protein-protein hydrogen bonds. Furthermore, the structures were shown to be more compact across the 2 ns of the simulation [172]. This is attributed to the stronger hydrogen bonding

within the solvent compressing the protein into a smaller cavity [172]. Furthermore, fewer solvent-protein hydrogen bonds were formed upon folding of the structures in the D_2O simulation [172]. There are also some results to suggest that proteins are unperturbed in D_2O or largely unchanged [163, 173]. In one study in particular, despite an increase thermal stability, there was little change in the ΔG of unfolding [163]. Furthermore the protein azurin was thermally destabilised in D_2O. It has been suggested that the changes enthalpic and entropic effects in D_2O cancel one another other out [174].

Isotope effects also influence the unfolding and folding rates of proteins[170, 175–177]. The refolding rate of the protein lysozyme is significantly reduced in the presence of D_2O [175]. It was suggested that the lower folding rate was due to the increased C=O$\cdots$DOD hydrogen bonding energy, or due to a change in the hydrophobic interactions within the protein. This result was independent of whether the exchangeable backbone groups were protiated or deuterated [175]. This has also been observed for the protein CD2 [170]. The decrease in the observed rates were attributed to a change in the hydrophobic interaction due to increased solvent-solvent hydrogen bond affinities [175]. Furthermore, the solvent isotope effect is thought to increase the rigidity of a number of proteins[158, 178]. The intrinsic dynamics of a series of proteins was found to be dampened in the presence of D_2O [158]. This suggested an increase in the structural rigidity of the protein. Whilst an increase in intra-peptide hydrogen bond strength could not be ruled out completely, the increase in rigidity was attributed to the propensity for D_2O to form water-water hydrogen bonds [158]. This was due to the rapid response (within 10 minutes) to the change in dynamics of the protein when diluted into D_2O. Full isotopic substitution was assumed to not occur within the 10 minute dead time of the experiment [158].

1.5.3.2 Deuteration of the carbon backbone and side chains

A fully detuerated (sometimes referred to as perdeuterated) protein is a protein containing deuterium instead of hydrogen at all non-exchangeable (C–H bonds) and exchangeable (e.g. amide) groups in the protein backbone and sidechains. The example dipeptide in Figure 1.5.3.1 illustrates the full deuteration of a protein. Fully deuterated proteins are synthesised by bacteria when they are grown on a deuterated carbon source [130, 179]. Deuterated proteins have been found to be biochemically similar to protiated (non-deuterated) proteins [180]. However, they appear to have different physiochemical properties [130, 137]. Studies on deuterated proteins have shed light on the importance of different interactions to the thermodynamic stability of proteins.

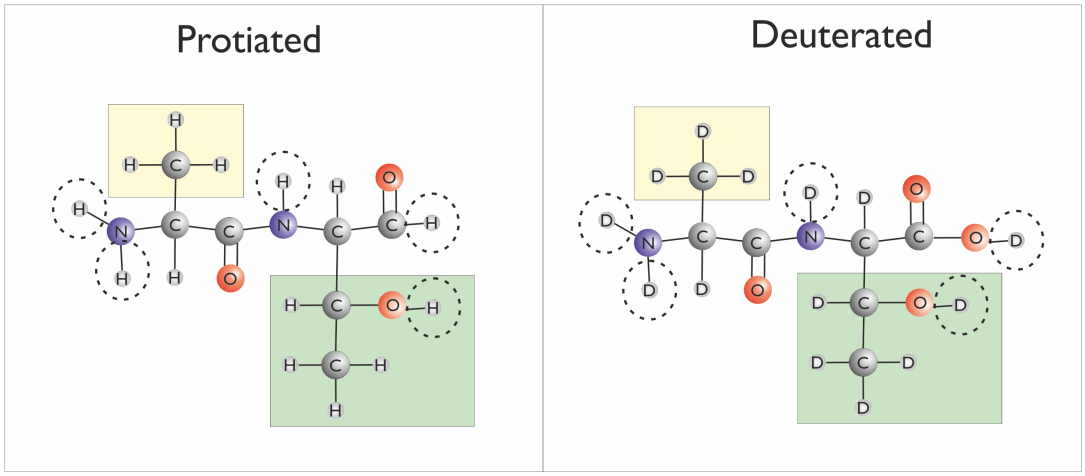


Figure 1.5.3.1: Figure showing a fully protiated and fully deuterated polypeptide. A fully protiated protein contains hydrogen. A fully deuterated protein contains deuterium at all positions. However, partial labelling of the protein can occur at the exchangeable sites, indicated by the dashed circles. This is achieved by saturating a protiated polypeptide in D_2O or a deuterated peptide in H_2O.

A study by Crespi and Katz in 1962 determined the thermal melting temperature (T_m) of fully deuterated DNA [181]. Despite no differences in the T_m of DNA, significant differences have been noted between, fully deuterated, and protiated proteins. X-ray structures of a number of deuterated proteins have been determined. These proteins have been shown to have the same structure as the protiated homologues[180, 182, 183]. Despite this, deuterated proteins proteins have generally been found to be thermodynamically less stable than their protiated analogues in both H_2O and D_2O. This has been determined using chemical denaturation, thermal denaturation and activity assays. The first study to compare the physio and chemical differences of deuterated and protiated proteins was performed in 1963 by Berns *et al.* [184]. In this study the thermal melting temperature of a deuterated protein was compared to its protiated analogue in H_2O. In this study the thermal transition of the deuterated protein was 5 °C lower than the protiated analogue. This change in stability was attributed to differences in stabilities of the side chain interactions.

This study was followed by a number of other studies on deuterated proteins [130, 137, 156, 170, 185]. This included an extensive study comparing the thermodynamic stabilities of deuterated and protiated proteins in both H_2O and D_2O [137]. In this study the least stable system, once again, was the deuterated protein in H_2O [137]. This was followed by the protiated protein in H_2O and the deuterated protein in D_2O [137]. The most stable protein was therefore the protiated protein in D_2O. [137] This destabilisation was more pronounced in the H_2O solvent in comparison to the protontated and deuterated proteins in D_2O. This trend has been observed for a number of deuterated proteins studied

[130, 170, 186]. The deuterated proteins, in both solvents, were therefore found to be less stable than the protiated homologues in the same solvent. The decrease in stability of the deuterated proteins has been attributed to the shorter C-D bond compared to a C-H bond [137]. The C-D bond has been determined to be ~ 0.005 Å[187] shorter than the C-H bond. This reduces the steric requirement of a methyl group within the protein core. The same effect has been observed for the change in activity of a deuterated and protiated enzyme [130, 144].

1.5.3.3 Single molecule force spectroscopy in deuterium oxide

One method to directly test regions of specific interactions within a single proteins is single molecule force spectroscopy (SMFS). Dougan *et al.* (2008) mechanically unfolding protiated protein, called I27, in a solvent environment containing H_2O and D_2O. These experiments reported an increase in the force required to unfold the protein in D_2O [132]. The unfolding force of the protein was reported to be 17.6 % higher than the protein in H_2O. Furthermore the protein was reported to refold more readily at small forces (30 pN) in D_2O than in H_2O. This suggested the driving forces for folding were already present in D_2O but absent in H_2O. This study suggested that this refolding was driven by enhanced hydrophobic interactions and illustrated the role of stronger hydrogen bonds within the solvent.

SMFS experiments of a synthetic peptide (containing a sequence of amino based on the sequence of a protein responsible for the elasticity of wheat) in H_2O and D_2O solvents, have also been performed [188]. In this study, the elastic peptide was incubated in D_2O to allow for full isotopic exchange to take place. The thermal softening of the peptide was enhanced in D_2O suggesting that the protein was less mechanically stable in D_2O [188]. This destabilisation was observed as an increase in the persistence length of the peptide with temperature. An increase in persistence length is indicative of a decrease in the peptides flexibility. This effect was less significant in H_2O. The larger increase of persistence length of the peptide in D_2O was associated with a faster expansion or unfolding of the peptide chain [188]. This is not in agreement with the results discussed previously [132]. However, this peptide chain was highly hydrophillic, leading the authors to hypothesise that it was due to isotopic substitution and hydrophobic interactions were less prominent [188]. In the previously discussed study, the increase in mechanical stability was due to an increase in hydrophobic contacts.

1.5.4 Using deuterium as a model system

This section of the report has reviewed the literature on the importance of hydrogen bond strength to protein stability, function and dynamics. Hydrogen bonds are thought to play a key role in the stabilisation and dynamics of proteins, both intrinsically within the protein and externally between the protein and the solvent. Studies have utilised the increased hydrogen bond strength of deuterium to quantify changes in protein stability. This is done either by performing experiments in D_2O and experiments on deuterated proteins. These studies have found that, in general, an increase in hydrogen bond strength in the solvent increases the thermodynamic stability of proteins and biological systems. This is attributed to the increase in hydrophobic interactions due to an increase in the ordering and strength of hydrogen bonds in the solvent. Perdeuteration, on the other hand, destabilises proteins due to a decrease in hydrophobicity. This highlights the importance of hydrophobic interactions in protein stability. Furthermore, these studies illustrate a delicate balance between the hydrogen bonding within the protein and the solvent.

SMFS has been widely used to characterise proteins mechanically and can determine the forces that are involved in maintaining the stability of the protein. However, deuterated proteins have not previously been characterised using a single-molecule approach. This makes it a powerful tool to determine the importance of hydrogen bond strength and hydrophobic interactions in protein stability. The single molecule approach in this project will be used to determine information about the mechanical and kinetic stability of a protein. This will enable an in-depth study into the effect of solvent-protein hydrogen bonding and internal interactions on the energy landscape of proteins. The contribution of SMFS to the understanding of the interactions governing protein stability will be discussed in the next section.

1.6 The mechanical stability of proteins

One method to study the importance of specific interactions within proteins is by mechanical denaturation. SMFS enables single protein molecules to be unfolded over a well-defined reaction co-ordinate and is able to determine the underlying perturbation in the energy landscape. Furthermore, single molecule experiments are able to determine stable intermediates in protein structures along this single reaction coordinate [189]. For a protein to be mechanically stable it must have a mechanical clamp region, that is, there must be a region of interactions that are able to withstand unfolding forces greater than

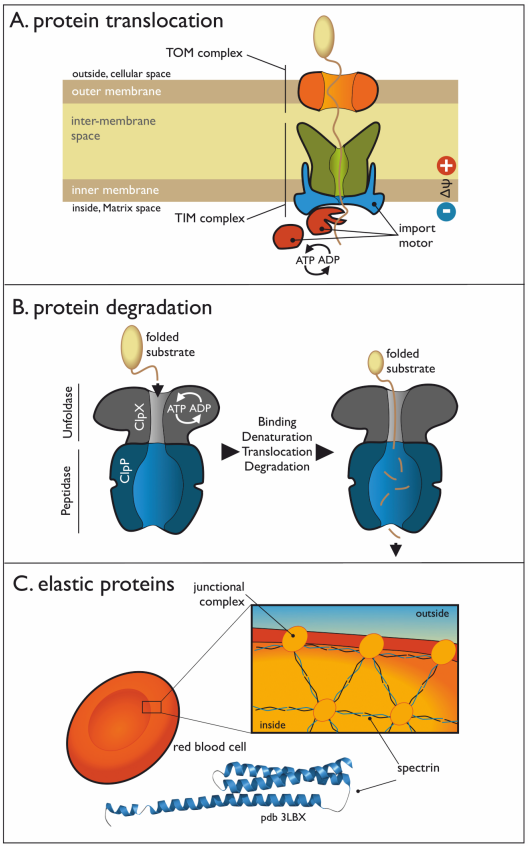


Figure 1.6.0.1: Schematics of the use of forces in biology. A: shows a protein translocating through a cell membrane. This example shows the import of a protein into michondria. For directional transport to occur the peptide must be partially unfolded first. The mechanism is not yet known but one suggestion is that it occurs by a power stroke mechanism where the peptide is actively pulled through the channel. B: Example of a protein being degraded by the hexameric ClpX complex. ATP hydrolysis provide the energy which is used to mechanically unfolded a protein prior to translocation into the chamber for degradation. C: The cytosckeleton of red blood cells is made up of α-helical proteins which form an elastic mesh. This mesh provides the mechanical stability required by the red blood cell due to external forces such as shear flow rates. Reproduced from Ref [189] with permission from The Royal Society of Chemistry.

the noise of an instrument.

Proteins have been found to unfold and refold during biological processes, due to chemical, thermal or even mechanical denaturation [190]. As a result, the significance of force within biological systems has become apparent [189, 191–193]. For a protein that is initially in a folded state to translocate across a membrane in mitochondria, requires forced unfolding. The mechanism of unfolding is not fully understood, however, two hypotheses have been proposed. The first model assumes that the protein propagates through a channel via random sliding movements of the peptide chain; whereas the other model proposes that the protein is actively pulled through the channel by a chaperone[194]. Titin kinase and the Von Willebrand factor are examples of mechanosensor proteins[195–197]. These proteins have been identified as components of cell signalling by converting mechanical forces into biochemical signals. Furthermore, proteins have been identified that require the ability to withstand applied forces[198, 199]. Examples include fibronectin (a), tenascin (b) and titin (c), these proteins extend and contract for cell migration and adhesion ((a) and (b)) and for the passive elasticity to prevent damage to skeletal and cardiac muscle sarcomeres (c). Some more examples of forces in biology are given in Table 1.6.0.1. To help understand more about biological forces within cells, mechanical biosensors have been developed and inserted into cells. These sensors have

directly shown the presence of force in cellular activity [200]. The relevant physiological forces for some proteins have been reported to be >10 pN, and single molecule force extension experiments can measure forces from several fN up to a few nN, which correspond to entropic forces up to the energy required to break covalent bonds [189, 201, 202].

Mechanical process	**Function**	**Ref.**
Protein degradation	In the bacterial cell proteins are degraded by two proteins which combined are known as ClpXP; one protein is an AAA+ATPase known as ClpX and a peptidase called ClpP. To be degraded the protein needs to be unfolded before translocation to the ClpP region for degradation. The energy required to perform the mechanical processes of unfolding and translocation of the protein by ClpX is produced by ATP binding and hydrolysis [203]	[204]
Cell signalling	Mechanical forces can also play a role in cell signalling. For example the ear is subjected to mechanical forces from sound waves and movements of the head. Sound waves provide a mechanical signal which is converted to an electrical signal through hairs in the inner ear.	[205]
Extracellular matrix	The extracellular matrix is the important "shell" surrounding cells in all tissues. This matrix is important for providing mechanical and structural support as well as facilitating biological processes such as cell migration and adhesion. Elastic proteins such as collagen, fibrillin and elastin are often found in the extracellular matrix in tissues. The composition of tissue is dependent on the cellular forces that is is subjected to. Often they are required to withstand high tensile and repetitive stresses.	[206]
Cytoskeletan	Red blood cells survive for 120 days in the human body and are subjected to mechanical stress such as shear flows. The proteins spectrin and F-actin are found in the cytoskeletan of red bloods cells. The proteins interact to form a cross-linked skeletal structure to the cell and maintains the specific structure of the red blood cell under different mechanical perturbations.	[207]
DNA replication, transcription and packaging	Molecular machinery such as DNA helicases and RNA polymerase are required to "unwind", linearise and "unzip" single stranded and double stranded DNA during replication and transcription. Motors involved in packaging of DNA have also been found to be subjected to internal forces of the DNA opposing confinement.	[208]

Table 1.6.0.1: This table describes some of the examples of force in biological systems.

Force is therefore an important tool to study the interactions within proteins on a single molecule scale. This allows the experiment to probe important unfolding intermediates along a single reaction coordinate. Bulk chemical and thermal denaturation of proteins can give information about an averaged ensemble of proteins but do not follow a single reaction coordinate. There are a few experimental techniques available that use force to probe single molecules such as individual protein domains. These include optical traps [209], magnetic beads [210] and the atomic force microscope (AFM) [211]. In these single molecule experiments the force response of a protein is monitored as the separation between two well-defined attachment points (usually between the N- and C- termini) is

increased, thus allowing interactions holding the structure together to be examined. The mechanical stability of the protein is described as the maximum force the protein can withstand before it unfolds, resulting in a fully extended state [189]. Force perturbs the energy landscape by lowering the energy barrier to unfolding. The assumption is that an application of force tilts the energy landscape by $-F \times x$, where x is the reaction coordinate over which the force is being applied [189, 212]. For most proteins this process is kinetic, which means that it depends on the activation energy of the transition barrier. By determining the unfolding rate of a protein at different forces or velocities, information about the unperturbed energy landscape can be extracted, namely: x_u, the distance between the folded state and the unfolded state and $k^0_{u,force}$, the unfolding rate at zero force [189].

Single molecule force spectroscopy (SMFS) has been used to mechanically unfold a diverse range of proteins; some proteins have been found to be mechanically labile, whereas others have been found to withstand forces in the range from tens to hundreds of picoNewtons (pN) [189, 212]. The energy landscape has been found to be altered by changes in temperature, pH and mutations [213–215]. In this section, the origin of mechanical resistance of proteins is discussed along with studies performed using SMFS.

1.6.1 How does protein structure affect the mechanical properties of a protein?

Some proteins are able to withstand forces of above 200 pN, whereas other proteins unfold at forces below the measurable limit of an AFM which is $\sim$ 15 pN [189]. Table 1.6.3.1 demonstrates that protein secondary structure is a key determinant of the mechanical stability of a protein. The most mechanically stable proteins are those from the all-β SCOP class, followed by α/β structures and with the most mechanically labile being proteins containing all-α structure[216, 217]. This appears to be a general trend across all proteins that have been mechanically characterised, when pulled from the N–C terminialigned with the force.

An extensive computational study by Lu *et al.* (1999) compared the unfolding of 10 proteins containing different secondary structures [218]. This study compared the unfolding of 10 proteins with a similar number of amino acids but with different secondary structures. Using molecular dynamic simulations (MDS), a constant velocity was applied to the ends of the proteins and the force response was monitored. This study found that elongation of proteins containing β-sheets resulted in a force peak. This force peak

corresponded to the breaking of hydrogen bonds in the backbone of the protein. For proteins containing α-helices, there were no inter-strand hydrogen bonds to break in this structure and the protein unfolded readily with no force response. This suggested the hydrogen bonds connecting the β-strands in a β-sheet conformation provide more stability than hydrophobic contacts between α-helical proteins. A simple elastic model has also been used to descibe this general trend between unfolding force and type of secondary structure [219].

1.6.2 Importance of the direction of the applied force on the mechanical stability of a protein

Despite the general trend suggesting secondary structural elements are a key determinant of protein mechanical stability, it cannot explain the broad range of force responses observed for proteins with similar structures. For example, the protein TNfn3 unfolds at a force of 120 pN at 600 nm/s [216, 220] which is a significantly lower unfolding force than the, structurally similar, protein I27 (184 pN at 200 nm/s) [214]. Experimental [221] and theoretical studies [218, 222] of different proteins suggested that the number and geometry of hydrogen bonds, with respect to the applied force, could explain the differences between structurally similar proteins. However, it was challenging to directly determine the effect of hydrogen bond geometry in different proteins due to differences in protein sequences and stabilities [223]. This problem was circumvented in two studies both published in 2003. In the study by Carrion-Vazquez *et al.*, the pulling geometry of the protein was changed by pulling the protein ubiquitin between the N- to C- termini or between a lysine (lys48) and cysteine residue [224]. The ubiquitin domains pulled between the two termini had unfolding forces of 203 pN at 400 nm/s [224]. The domains pulled between lys48 and the cysteine residue had an unfolding force of 85 pN at 300 nm/s [224]. The differences in forces were attributed to the direction of the hydrogen bonds with respect the applied force vector. This sensitivity to hydrogen bond geometry was also observed in the study by Brockwell *et al.* on the protein E2lip3 [223]. In this study the protein could be immobilised on a surface in two positions. This enabled the hydrogen bonds to be either "peeled" and "sheared". In the peeling geometry, the hydrogen bonds are parallel to the direction of the applied force. In the shearing geometry the hydrogen bonds are perpendicular to the direction of the applied force. This is illustrated in Figure 1.6.2.1. Whilst a large unfolding force (177 pN at 700 nm/s) was observed for the protein with the shearing geometry, no force unfolding peak was observed for the protein in the peeling geometry [223]. This suggested the protein unfolded below the noise of the instrument. The hypothesis that this was as a consequence of the hydrogen bond

geometry in the termini was strongly supported by MD simulations.

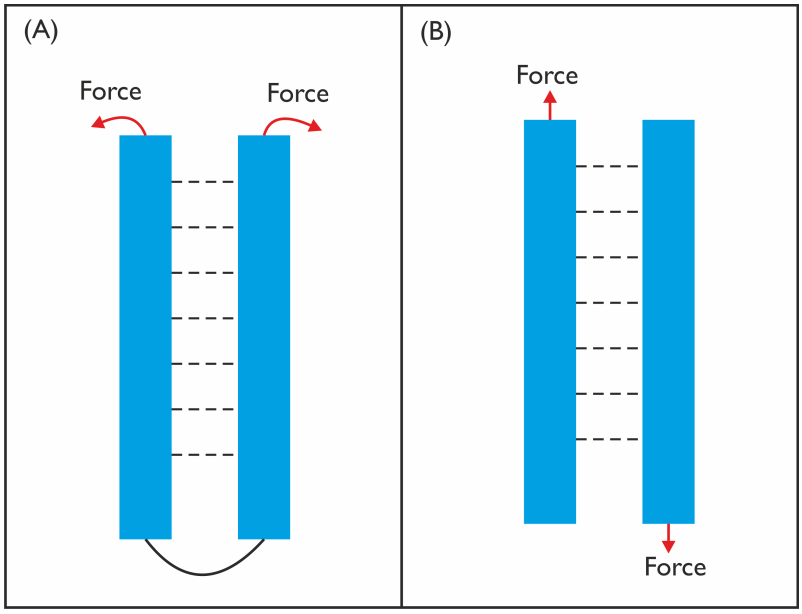


Figure 1.6.2.1: Illustration of the different pulling geometries of hydrogen bonded β-strands on the application of an unfolding force using SMFS. (A) The "peeling" geometry. The force (red arrows) is applied to the N- and C- termini of the protein such that the hydrogen bonds are sequentially broken. In this case the force is applied parallel to the hydrogen bonds. (B) The "shearing" geometry. In this application of force, the hydrogen bonds must be broken simultaneously. This is because the force is applied perpendicularly to the hydrogen bond geometry.

Since then, experimental [225–231] and theoretical [232–236], studies have been performed to determine the importance of pulling geometry to protein mechanical stability. The pulling geometry has been controlled by circular permutations[6] to protein structures or by engineering exposed pairs of cysteine residues. The force response of the green fluorescence protein, GFP, was determined by applying force across five pairs of cysteine residues [231]. A diagram illustrating these results is shown in Figure 1.6.2.2 [231]. A large range of forces were observed between the different unfolding geometries, the forces varied from $\sim$ 100 pN for the least stable geometry and $\sim$ 600 pN for the most stable geometry [231]. Circular permutations on the protein DHFR caused a decrease of $\sim$ 56 pN in the unfolding force c.f. the wild type (w.t.) protein [237]. Alternative unfolding pathways have also been observed by manipulating the pulling geometry of a protein [238]. These studies have identified key structural motifs involving hydrogen bonds that are responsible for the unfolding pathway and therefore force response of the protein. Certain arrangements of hydrogen bonds give rise to much larger unfolding forces than others. The pattern of hydrogen bonds within a protein can help identify the force response of the protein. These regions of hydrogen bonds are known as the mechanical clamp of a

[6] A circular permutation results in an identical topology, but a different location of the N- and C-termini. This causes the secondary structure elements to unfold in an alternative order.

protein.

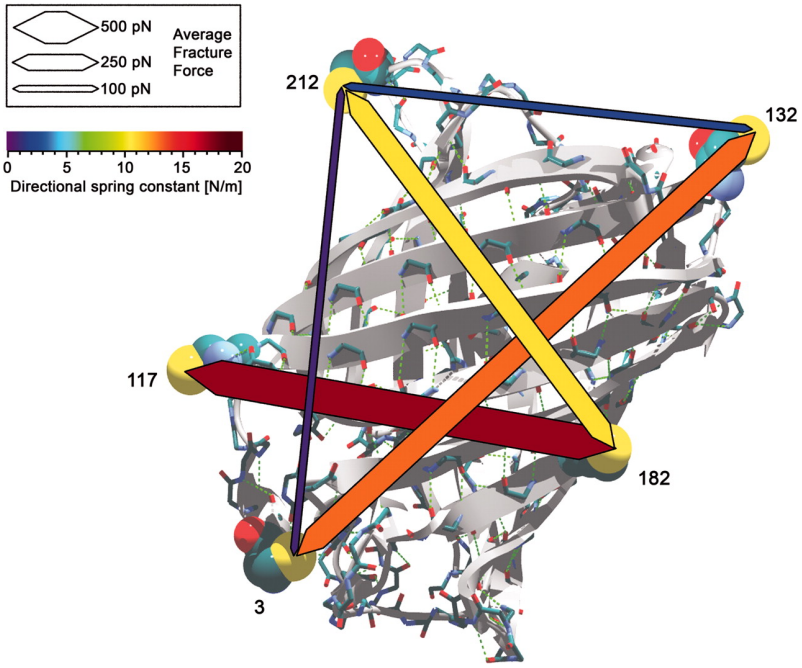


Figure 1.6.2.2: Illustration of the deformation response of the protein GFP (PDB ID 1EMB [239]) when subjected to five different pulling geometries. The position of force application are indicated by space filled residues. The width of the connecting lines indicates the mechanical unfolding force. The colour of the line indicates the rigidity of the protein in the pulling direction. Where red indicated a high directional spring constant (most rigid) and blue a low directional spring constant (less rigid). Figure taken from Ref [231] subject to © 2006 by The National Academy of Sciences of the USA.

1.6.3 Mechanical clamps in proteins

A mechanical clamp is defined as a localised force bearing region within a protein which results in the largest force response of the protein [240–242]. The first mechanical clamp region within a protein was identified by Lu *et al.* 1998 in the protein I27 in 1998 using MDS [243]. The force bearing region was identified as a cluster of six hydrogen bonds. For the protein to unfold, simultaneous rupturing of these hydrogen bonds was required. The existence of this mechanical clamp was experimentally verified in 2000 by Li *et al.* by mutating residues identified in the mechanical clamp region[244]. These mutations severely destabilised the protein. This clamp region consists of two directly, hydrogen bonded β-strands located near the terminal of the protein. Since then other mechanical regions in other proteins have been identified [189, 193, 217, 245, 246]. These mechanical regions are not always directly part of the force bearing region. Two structurally analogous, β-sheet proteins were found to have significantly differing force responses. The protein Top7 could withstand significantly higher forces than Barstar. The differences in force response of Top7 was attributed to supporting neighbouring strands within the force-bearing region. A theoretical study, comparing coarse grain

mechanical unfolding simulations of 17 134 protein structures from the PDB, has identified a series of different mechanical clamp regions within proteins [240, 240]. The mechanical clamps were identified by hydrogen bonding patterns between secondary structure elements within the protein structures. These clamp regions are illustrated in Figure 1.6.3.1.

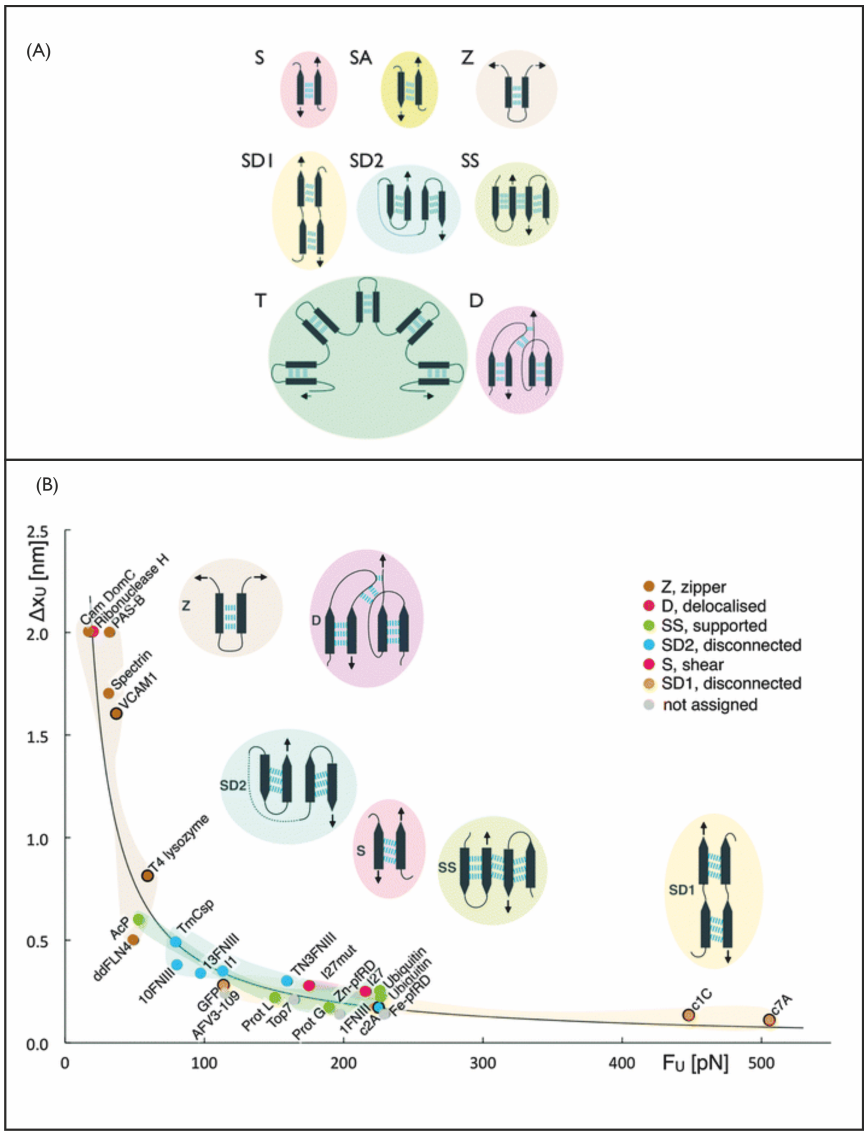


Figure 1.6.3.1: Top: Clamp motifs described by Ref. [241]. S is a shear clamp, SA is a shearing antiparallel, Z is a zipper, SD1 is a shear disconnected clamp, SD2 is a shear disconnected 2, SS is a shear clamp supported by other strands, T is a torsional clamp and D is shear delocalised. Each clamp gives a different force response. Bottom: The relationship between Δx_u and F_U (600 nm/s), proteins are grouped by the clamp motif. The data can be described by a power law function $\Delta x_u = \frac{39.4}{F_{UN}}$ with an R^2=0.91 (goodness of fit). Clamp types are illustrated above. Adapted from Ref [247] with permission of the PCCP Owner Societies.

A correlation between the type of clamp in a protein and its mechanical properties such as: the distance from the unfolded state to the transition state, Δx_U and the unfolding force, has recently been reported [193]. The relative contact order (RCO) is

defined as the average sequence distance between residues that are in contact, normalised to the contour length of the protein. It had previously been suggested that there was a correlation between the RCO and the mechanical unfolding force of the protein. A paper by Tych, Hoffmann *et al.* (2013) showed that, although there was a general trend for all proteins with increasing RCO and unfolding force, a stronger correlation was observed for all β proteins. There was no clear correlation for other protein fold types or certain clamp motifs such as SD2 or zipper. Despite this, this study observed clusters the proteins based on their mechanical clamp motif in a change in x_U-F plot across 25 proteins studied. Typically proteins with the SD1 motif have a smaller x_u and a larger range of unfolding forces. This study demonstrates that similarities in the structural components in proteins results in similarities in the energy landscapes. The relationship between the x_U and F has been shown to follow a power law suggesting proteins with a lower unfolding force have a higher x_U (forces determined at 600 nm/s) [193].

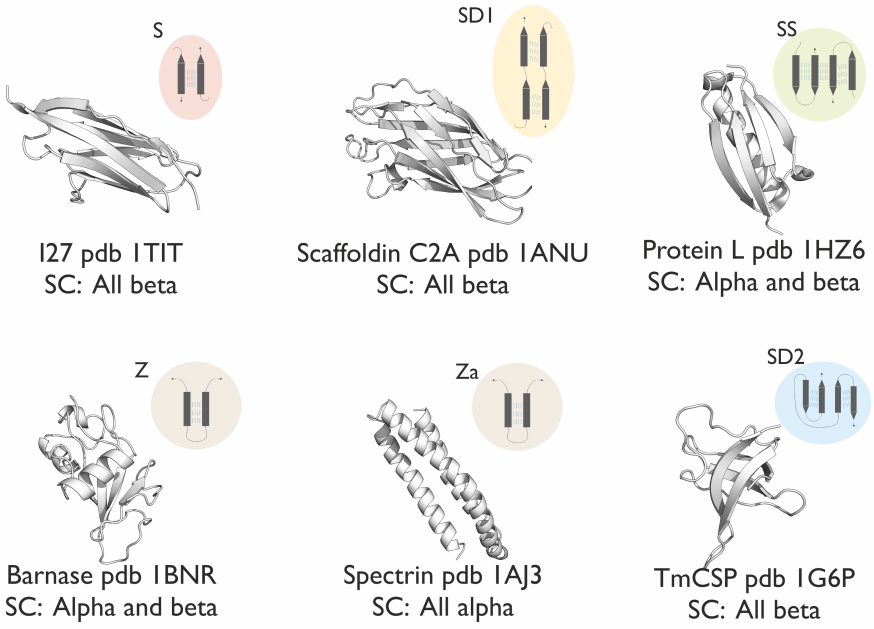


Figure 1.6.3.2: Figure showing the structures of 6 different proteins created using the software PyMOL [10]. Each protein has a unique 3D structure containing α-helices and β-strands. Proteins can contain force bearing regions known as mechanical clamps. The regions within the protein that satisfy the mechanical clamp patterns defined by Sikora *et al.* are indicated to the right of the structure [241, 242]. Structures were obtained from the PDB database [8] and the references are as follows: 1TIT [114], 1HZ6 [9], 1G6P [248],1ANU [249], 1BNR [250], 1AJ3 [251].

These SMFS studies identified the importance of specific hydrogen-bonded regions to the mechanical stability of a protein. Hydrogen bond patterns between secondary structure elements of proteins are responsible for the force response of proteins. Furthermore, there is a correlation between the unfolding force observed and type of mechanical clamp

Protein	Mechanical unfolding force/pN (Velocity/nm/s)	Clamp motif	Protein fold (SCOP)	Reference
I27	204 (400–600)	S	all beta	[252]
I27 (C47S C63S)	172 (600)	S	all beta	[217]
I1	127 (400-600)	SD2	all beta	[103]
1FNIII	220 (600)	SD2	all beta	[253]
10FNIII	74 (600)	SD2	all beta	[253]
13FNIII	89 (600)	SD2	all beta	[253]
C2A	60 (600)	SD2	all beta	[221]
C2B	100 (500)	SD2	all beta	[254]
E2Lip3	15 (700)	Z	all beta	[223]
TmCSP	78 (400)	Z	all beta	[193]
Barnase	70 (100-500)	Z	alpha and beta	[255]
Protein L	152 (700)	SS	alpha and beta	[217]
GB1	184 (400)	SS	alpha and beta	[256]
Barstar	<50 (400)	-	alpha and beta	[257]
Top7	150 (400)	-	alpha and beta	[257]
Ubiquitin	203 (300)	SS	alpha and beta	[224]
AcP	$\sim$ 60 (600)	SS	alpha and beta	[258]
Spectrin	$\sim$50 (80-800)	Za	all alpha	[259]
Calmodulin	<15 pN (600)	-	all alpha	[221]

Table 1.6.3.1: Values, at a given velocity, for the maximum force a protein can withstand before unfolding. The type of structure was obtained from a search of the SCOP class of the protein. The mechanical clamp motif of the protein was determined from the definitions given by the Biomolecular Stretching Database (BMSD) [241] [193]. The table illustrates the effect of protein topology and pulling direction to the mechanical resistance of proteins. Force values depend on the pulling velocity used in the experiment, proteins unfold at lower forces for lower pulling speeds because the protein has more time to sample the energy landscape before mechanical unfolding occurs. Some mechanical clamp motifs have not been defined by the hydrogen bond patterns from [241], these are indicated with a -.

of the protein. However, hydrogen bond patterns cannot explain the differences in the forces observed for proteins containing the same type of mechanical clamp. Mutational studies [124, 245] and simulations [217, 218, 260] have identified other interactions and non-native contacts that can be important in the mechanical responses of proteins. Furthermore combination of mutagenisis, mechanical and chemical unfolding of proteins can reveal information about the importance of specific amino acids to the mechanical transition state of the protein.

1.6.4 Point mutations can significantly alter the mechanical stability of proteins

Regions of hydrogen bonds can be identified as the force bearing region of a protein. Mutations of amino acids within this mechanical clamp region of proteins have been shown to alter the mechanical stability of proteins [124, 244, 245, 261]. These studies have illustrated the importance of hydrogen bonded regions within proteins[244], identified intermediates in the unfolding pathways [261], highlighted the importance of hydrophobic interactions [124, 245] and identified multiple unfolding pathways of a protein [262] within proteins. The first studies combining mutational studies and SMFS were performed on the I27 protein. These studies mutated key residues identified from MD simulations. The protein was severely destabilised by mutations of hydrogen bonded amino acids within the clamp but not altered significantly by changes in the hydrophobic core [263]. However, the hydrophobic core has been shown to play a key role in the mechanical stability of protein L and GB1[124, 245].

Mutational SMFS studies have also identified a "mechanical rheostat" in the core of protein L [124]. A mutation of a single residue, I60V, resulted in a 36 pN decrease in unfolding force campared with the w.t. protein despite no change in the thermodynamical stability of the protein[124]. This decrease in stability was coupled with a large increase in the distance to the transition barrier (0.46 nm compared with 0.26 nm for the w.t.) [124]. Therefore the transition state moved further away from the native state, allowing larger deformations of the protein before unfolding occurred. Two other single mutations were performed on protein L, these mutations were hypothesised to increase the hydrophobic core packing. Whilst one mutation increased the unfolding force by 13 pN, the other mutation increased the force by 72 pN at a single velocity [124]. This increase in force was coupled with an increase in the thermodynamic stability of the protein. The importance of hydrophobic core packing on the mechanical unfolding force of GB1 has also been

shown by mutations [245]. Protein GB1 is structurally similar to protein L. A series of mutations to this protein identified three mutations that caused a significant decrease in unfolding force. These mutations were positioned in the mechanical shearing region between two structural motifs containing hydrophobic contacts. These amino acids were hypothesised to provide stabilising hydrophobic interactions. Mutating these residues disrupts the contacts between the shearing region [245].

1.6.4.1 How is the transition barrier affected by mutations?

The change in the unfolding barrier height of a protein as a result of a mutation with respect to the w.t. protein can be compared to the change in thermodynamical stability of the native state with the wild-type (w.t.) protein. This comparison enables information about the native contacts formed at the transition state upon folding in the mutated protein to be obtained. Φ-value analysis determines how many of the native contacts are formed at the rate limiting step[264]. Best *et al.* have also implemented Φ-value analysis to determine how mutations affected the contacts of the protein at the transition state in mechanical unfolding experiments. Φ-value is given in equation 1.10 [264] and is defined as the ratio of the change in free energy of activation for folding ($\Delta\Delta G_{TS,force}$), to the change in equilibrium free energy ($\Delta\Delta G_{UN}$).

$$\Phi = 1 - \frac{\Delta\Delta G_{TS,force}}{\Delta\Delta G_{UN}} \tag{1.10}$$

A value of 1 indicates that the protein is fully native at the transition state. A partial value in SMFS experiments indicates partial structuring of the protein [264]. Φ-value analysis has been used to determine the role of different amino acids on the transition state of the protein. Best *et al.* determined the regions of the protein that were involved in the mechanical unfolding event in the protein [264]. Furthermore, the Φ-values obtained for the destabilising hydrophobic deletions in both Protein L and GB1 indicated that the mutated proteins were only partially structured at the transition state [124, 245]. This sensitivity of SMFS experiments to the structure of the single molecules at the transition state enables determination of important interactions at the rate limiting step. Whilst in bulk experiments, a partial Φ-value can suggest alternative transition states or partial contacts, in SMFS experiments it can only represent partial structuring of the protein. This analysis has never been used to determine the influence of deuterium on the rate limiting step. It is apparent that manipulating the hydrophobic interactions of a protein can affect the structure of the protein at the transition state.

1.6.5 The role of solvent environment on the mechanical stability of proteins

The solvent environment is believed to be important for the dynamics of proteins [158, 265, 266]. It is therefore useful to understand the role of solvent environment on the stability of proteins [132].

In recent years there has been a large number of studies to understand the role of osmolytes on protein stability. Osmolytes are small organic molecules that can shift the equilibrium between the folded and unfolded states of a protein [267]. Osmolytes, such as glycerol, sorbitol and sucrose, push the equilibrium towards the folded state of the protein. Denaturing osmolytes such as guanidine hydrochloride (GdnHCl) and urea shift the equilibrium to the unfolded state of the protein. Despite an abundance of research, little is known about the role of solvent molecules on the mechanical transition barrier and shape of the folding energy landscapes. Information about these is pivotal to the understanding of protein dynamics. SMFS have been performed in solvent environments containing a number of different osmolytes [268–270]. These molecules enabled the mechanical properties of proteins to be modulated. For example, glycerol has been shown to increase the mechanical stability of the protein GB1 [270]. Furthermore, dentaturing osmolytes such as GdnHCl and urea have been found to decrease the mechanical stability of proteins[269, 271]. In fact the mechanical unfolding of the protein GB1 indicated that the rate of folding and unfolding obtained from SMFS showed similarities to the rates obtained from bulk denaturation experiments [269]. However, this would not be the case for all proteins because of the differences of unfolding pathways between mechanical and chemical denaturation.

The mechanical properties of a protein can alternatively by tuned by a change in pH or ionic strength of a buffer. Electrostatic interactions within a protein can be affected by changes in the environment. Tuning the pH of a solvent has been shown to "switch-on" or "switch-off" electrostatic attractions and repulsions within a protein [272]. Furthermore, these electrostatic interactions can be screened by the introduction of salt [272, 273].

More recently, SMFS has also been used to mechanically probe a protein in a nonpolar environment [274]. Moving the protein from a water-based solvent to an organic nonpolar solvent octyl benzene, resulted in the loss of mechanical stability of the protein. This was equivalent to unfolding the protein in a high (6 M) concentra-

tion of GdnHcl and therefore suggested that the mechanical resistance of the protein was lost due to the loss of the tertiary structure of the protein[274]. However, MD simulations indicated that the loss of mechanical resistance was not due to a loss of tertiary structure, but due to an interface between water molecules in a hydration layer remaining on the surface of the protein resisting transference into the nonpolar environment.

SMFS is therefore sensitive to the changes in environment of a protein. In addition, SMFS can also probe the transition barrier of a protein. This therefore endows the technique with the ability to provide a unique understanding of the role of solvents on the mechanical folding pathway of a protein.

1.7 Discussion and objectives

In this chapter the importance of understanding protein stability was introduced. Protein folds are held together by different interactions such as hydrogen bonds and hydrophobic interactions. Whilst there are an abundance of studies addressing the contribution of these interactions to the stability of a protein, little is known about how they affect the unfolding transition state of the protein. Understanding how they shape the energy landscape can help understand misfolding and protein adaptation. Many studies have used mutations to try and probe this problem. However, it is complex to deconvolute the contribution of hydrogen bond strength from single point mutations due to the complexity of the environment of a HB. Very little is known about the net contribution of hydrogen bond strength and hydrophobic interactions to the unfolding (or folding) pathway of a protein or the important interactions governing the mechanical stability of proteins.

SMFS experiments are sensitive to the key interactions within a proteins mechanical clamp and solvent environment. The mechanical clamps are formed by regions of hydrogen bonds but can also be stabilised by other interactions such as hydrophobic interactions. Because the reaction coordinate probed in these experiments is well defined, the technique is sensitive to small changes within this region. It is, however, still challenging to "bridge the gap" between mechanical and chemical unfolding experiments on proteins. The two techniques probe different unfolding pathways. Mechanical unfolding drives the protein to a fully extended state, whereas chemical unfolding drives a protein to a molten globular state [275]. Simulations, however, have made some progress in connecting mechanical and thermal denaturation[276].

In this thesis the objective was to use SMFS studies to yield information about how hydrogen bond strength, hydrophobic interactions and the solvent affect the mechanical unfolding transition state of proteins. The strength of the interactions will be manipulated by mechanically unfolding fully deuterated and protiated proteins in both H_2O and D_2O.

Chapter 2

Materials and Methods

This chapter will give an overview of the materials and methods used in this project. It will begin by summarising the buffers used in the project and then will describe the biological techniques used to express deuterated and non-deuterated proteins (protiated). It will follow by giving an overview of the techniques used to characterise the proteins.

2.1 Materials

A detailed list of apparatus and chemicals and manufacturers is given in the appendix. The buffers/media were all made with Purite 18 MΩ distilled (MiliQ) water unless containing deuterium oxide (D_2O) or otherwise stated.

2.1.1 Media and Buffers

2.1.1.1 Growth Media

This section gives details about the growth medium used in this project. Luria-Bertani Broth (LB) growth medium was dissolved in Mili-Q water whereas the deuterated growth media, ISOGRO-D, was dissolved in D_2O before sterilisation. Agar is dissolved in the relevent growth media and the correct antiboitics. The details are given in the appendix.

Growth media	Components
LB medium	Premixed :40 % bacto-tryptone, 20 % yeast extract, 40 % NaCl. 25 g of media was dissolved in 1 litre of Mili-Q water.
Deuterated growth medium	ISOGRO-D medium (premixed: 30 % salts, 3 % water, 2 % glucose, 65 % amino acids/peptides), 1.8 g/L K_2HPO_4, 1.4 g/L KH_2PO_4, 1 g/L MgSO 50 g/L $MGSO_4$ 111 $\times$ 10^{-4} g/L of $CaCl_2$, pH 7.0. 1 g of medium is required per 100 ml of D_2O.
Agar plates	15 g/L agar

Table 2.1.1.1: Table giving the list of growth media used in this project and how to make them.

2.1.1.2 Buffers

This section describes the standard buffers used in this project. The buffers were all made with Mili-Q water or D_2O. The pH or pD were corrected with NaOH, HCl or NaOD or DCl respectively unless stated.

Buffer name	**Components**
Gel electrophoresis	
Anode buffer	400 mM Tris.HCl, pH 8.8
Cathode buffer	100 mM Tis.HCl, 100 mM Tricine, 0.1 % sodium dodecyl sulfate SDS, pH 8.3
2 $\times$ Loading buffer	50 mM Tris.HCl pH 6.8, 100 mM DTT, 2 % (*w/v)* SDS, 0.1 % (*w/v*) bromophenol blue, 10 % (*v/v*) glycerol
Protein Purification	
100 % protease inhibitor cocktail	100 mM PMSF, 200 mM benzamidine dissolved in ethanol
Size exclusion chromatography buffer	300 mM NaCl, 25 mM Tris.HCl, 2 mM DTT, 1 mM EDTA, pH 8.0
Lysis buffer	300 mM NaCl, 25 mM Tris.HCl, 20 mM imidazole, 2 mM DTT, 0.5 ml 100 % Triton-X , 2 % Sodium azide, pH 8.0

Wash buffer	300 mM NaCl, 25 mM Tris.HCl, 20 mM imidazole, 2 mM DTT, 2 % sodium azide, pH 8.0
Elute buffer	300 mM NaCl, 25 mM Tris.HCl, 200 mM imidazole, 2 mM DTT, 2 % Sodium azide, pH 8.0
Protein preparation	
Unfolding buffer (I27)	3 M GdnHCl, 25 mM sodium phosphate, 2 mM DTT, 1 mM EDTA, pH or pD 7.4
Unfolding buffer (pL)	3 M GdnHCl, 25 mM sodium phosphate, 1 mM EDTA, pH or pD 7.4
Protein characterisation	
Circular dichroism	63 mM sodium phosphate pH 7.4
Circular dichroism (I27)	25 mM sodium phosphate pH or pD 7.4, 2 mM DTT
Circular dichroism (pL)	50 mM sodium phosphate pH or pD 7.4
Equilibrium denaturation unfolding buffer (I27)	25 mM sodium phosphate, 2 mM DTT, pH or pD 7.4
Equilibrium denaturation folding buffer (I27)	10 M urea, 25 mM sodium phosphate, 2 mM DTT, pH or pD 7.4
Equilibrium denaturation unfolding buffer (pL)	50 mM sodium phosphate, pH or pD 7.4
Equilibrium denaturation folding buffer (pL)	8 M GdnHCl, 50 mM sodium phosphate pH or pD 7.4
Kinetic unfolding folding buffer (I27)	9 M urea, 25 mM sodium phosphate, 2 mM DTT, pH or pD 7.4
Kinetic unfolding folding buffer (pL)	8 M GdnHCl, 50 mM sodium phosphate, pH or pD 7.4
Atomic force microscopy buffer	63 mM sodium phosphate, pH or pD 7.4

2.1.2 Preperation of deuterated buffers

All deuterated buffers (buffer containing deuterium in place of hydrogen) used throughout this project were dissolved in 99.8 % D_2O. To ensure all hydrogen atoms were exchanged to deuterium, buffers were freeze-dried (lyophilised) and then re-dissolved in D_2O. This was repeated twice more. The pH of the buffers were determined using the relationship in Equation 2.1 [277, 278]. All buffer pD were corrected using sodium deuteroxide, NaOD,

or deuterium chloride, DCl.

$$\mathrm{pD} = \mathrm{pH}^* + 0.42 \tag{2.1}$$

2.1.2.1 Sodium phosphate buffer

To ensure that the pD of the deuterated sodium phosphate buffer is the same as the pH of the protiated buffer (buffer containing hydrogen), the apparent $\mathrm{p}K_a^{H*}$ of the buffer can be determined using Equation 2.2 [278]. The Henderson-Hasselbach (Equation 2.3 [279]) can then be used to determine the correct volume of Na_2DPO_4 (di-basic sodium phosphate) and NaD_2PO_4 (mono-basic sodium phosphate) to mix for the correct pH and molarity of sodium phosphate buffer before exchange. The pH can then be determined from Equation 2.1 [277].

$$\mathrm{pK}_a^H = 0.929\mathrm{pK}_a^{H*} + 0.42 \tag{2.2}$$

The $\mathrm{pK_a}^{\mathrm{H}}$ of dibasic sodium phosphate at room temperature is 6.82 [279], the apparent $\mathrm{pK_a}$ of sodium phosphate in D_2O is therefore 6.89. A pH^* of 7.0 would give a pD of 7.4 according to Equation 2.1.

$$\mathrm{pH} = \mathrm{pK}_a + \log \frac{[A^-]}{[HA]} \tag{2.3}$$

2.2 Protein Expression and Purification

This section gives details of the experimental procedures used to express and purify protiated [1] and deuterated [2] protein constructs.

2.2.1 Translation and Transcription for protein expression

In cells there are two steps for protein expression; transcription of DNA and translation of RNA into the protein [280]. Cells contain ribonucleic acid (RNA) polymerases which are enzymes that can build an RNA molecule with a sequence complementary to a deoxyribonucleic acid (DNA) molecule. This is illustrated in Figure 2.2.1.1. DNA molecules encode the genetic information for proteins. The process of making a complementary RNA molecule is known as transcription (see Figure 2.2.1.1 part B for a schematic of this process). Protein production is carried out on the ribosome; ribosomes are common to all cells and consist of proteins and ribosmal RNA (rRNA) molecules. There are three

[1] Proteins containing hydrogen

[2] Protein containing deuterium/D^2 in place of hydrogen atoms

important RNA molecules that are involved in protein production. Messenger RNA (mRNA) is the molecule that carries the genetic coding information for protein synthesis. Ribosomal RNA (rRNA) is a ribosomal component alongside protein ribosomes and acts as a catalyst for the formation of the primary amino acid sequence that makes up a protein. Transfer RNA (tRNA) are molecules that are tethered to amino acids which bind to the mRNA, linking the messenger sequence to the protein sequence.

Translation is the reading of mRNA for production of a protein. Every three codons of the mRNA represent a single amino acid, therefore the mRNA strand contains the information for the order of amino acids in a protein. A stop codon is at the end of a protein sequence and signals for the translation process to terminate. Figure 2.2.1.2 illustrates the process involved. Initially the subunits that make up the ribosome encase the mRNA at the start codon (black and white rectangles) for translation to occur and tRNA molecules provide the correct amino acid. The tRNA molecules bind to the correct codon site on the mRNA inside the ribosome, rRNA molecules then catalyses peptide bond formation between the amino acids in primary sequence. The ribosome moves along the mRNA until it reaches the stop codon.

Bacteria strains have been successfully genetically modified to facilitate the production of recombinant proteins for extraction [281, 282]. In this project modified *Escherichia coli* (*E.coli*) were used for the production of proteins.

2.2.2 The pET vector

For bacteria to produce the proteins which have been studied in this project, the genetic information for production of this protein must be inserted, transcribed and translated within the bacteria. Recombinant proteins can be toxic to the cell, therefore production of protein must be carefully controlled to optimise the amount of protein expressed. The production can be controlled by inserting genetic code for the protein of interest (POI) into a pET vector[3] [283]. A diagram outlining the important components of this bacterial plasmid is shown in Figure 2.2.2.1.

The POI is expressed by insertion, also known as transformation, (see Section 2.2.4) of the pET vector into genetically modified bacteria. The expression of protein is controlled by manipulation of the lac repressor protein. In cells the lac repressor protein

[3]The pET vectors containing the correct inserts for the POI in this project were provided by Dr David Brockwell

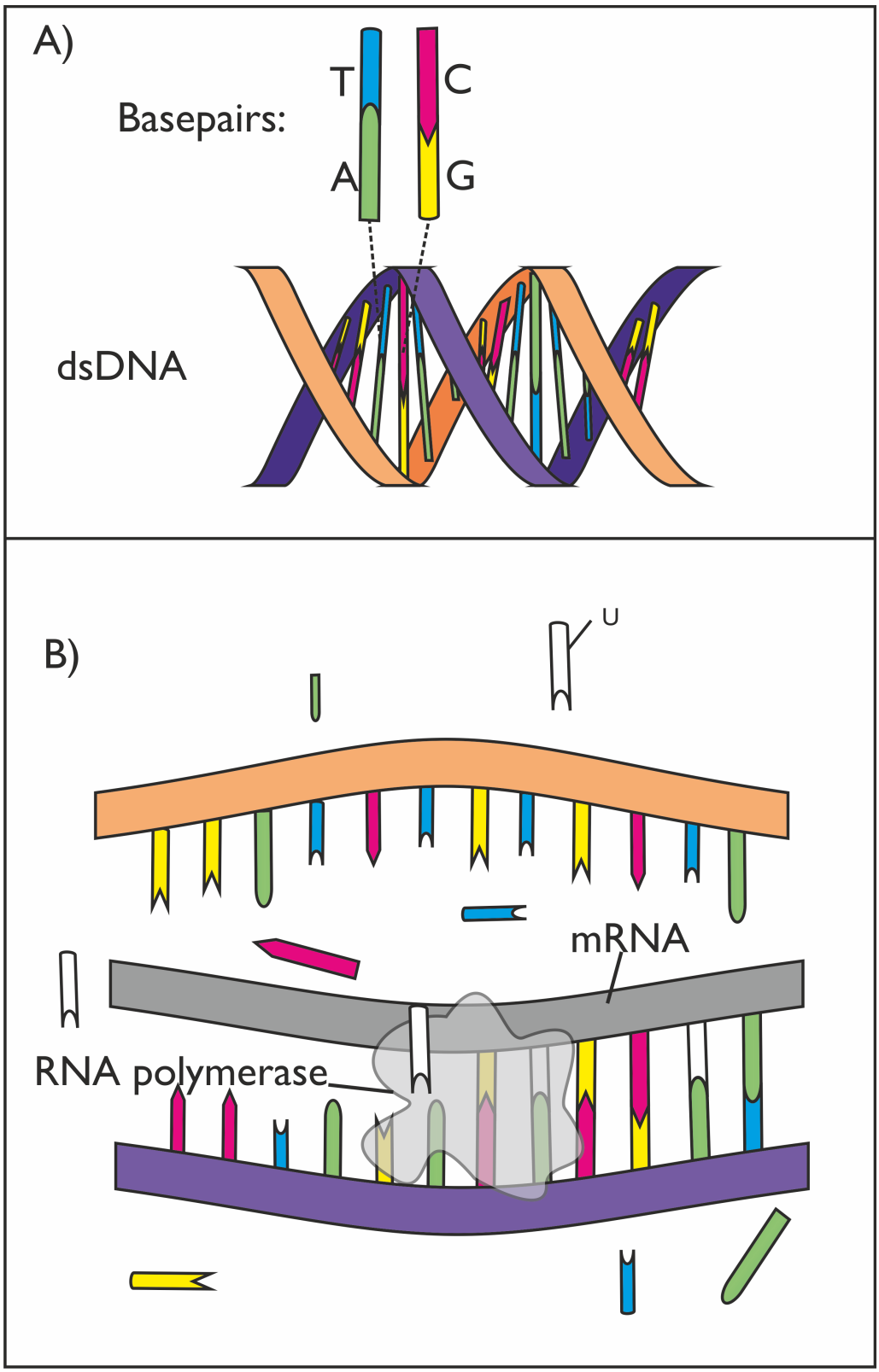


Figure 2.2.1.1: Top: Schematic of a double stranded DNA molecule (dsDNA). DNA molecules are made up of nucleotides. Nucleotides are formed of a phosphate group, sugar group and a nitrogen base. These nucleotides covalently bind to form the phosphate sugar backbone of DNA. There are four nitrogen bases; adenine (A), throemine (T), cytosine (C) and guanine (G). Each base has a complementary base, which bind to form base pairs; A (green) binds to T (blue) and C (pink) in complementary to G (yellow). The base pairs on one strand in double stranded DNA will bond to the complementary base pairs in the other. Bottom: Complementary mRNA molecules are synthesised by RNA polymerases, which translate across a DNA molecule. mRNA molecules differ from the DNA molecules; the base pair T in DNA is replaced by Uracil (U) and RNA molecules are always single stranded. Once the RNA polymerase has finished transcribing the DNA strand, the mRNA are able to carry the genetic information to the ribosome for translation into a protein.

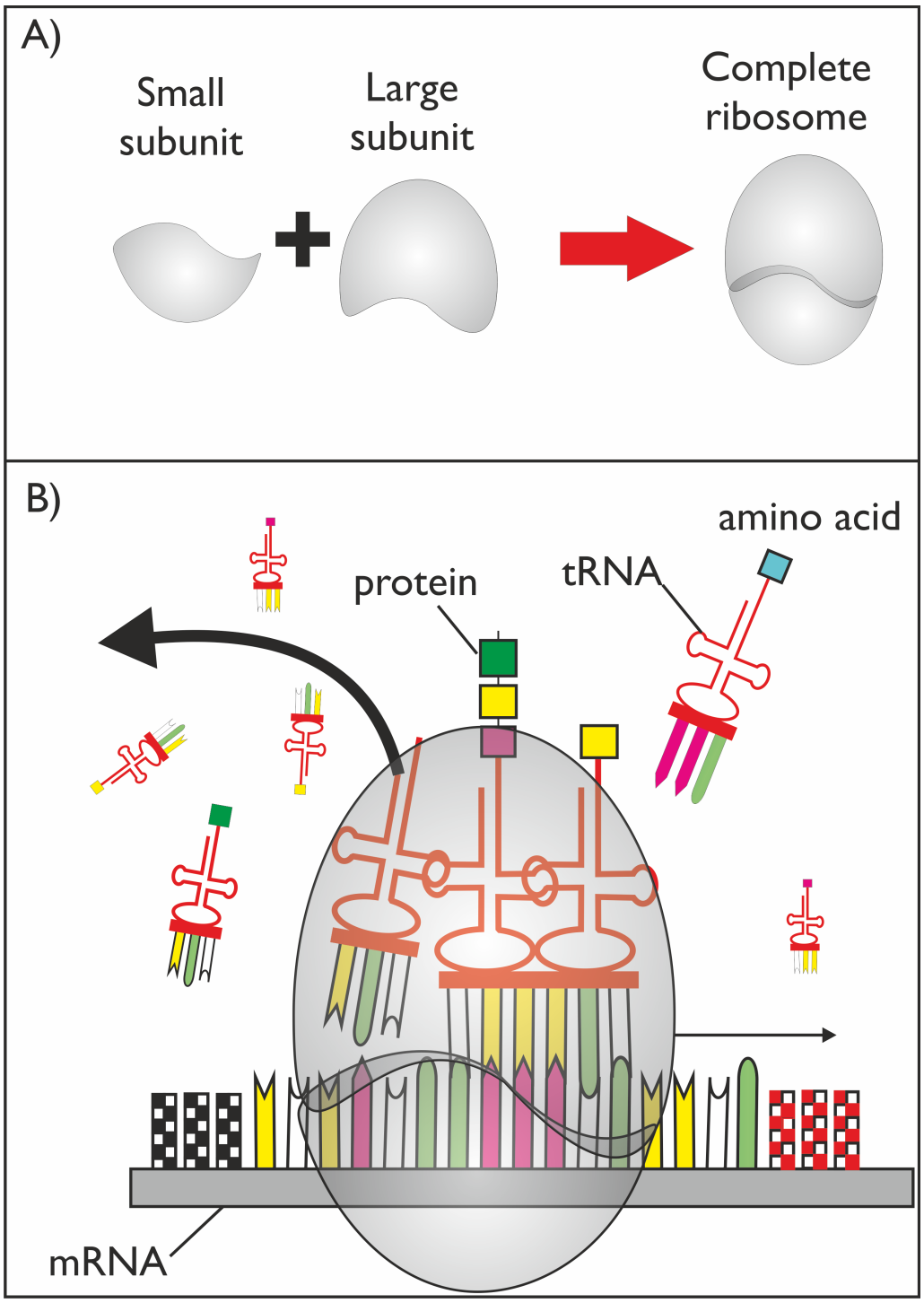


Figure 2.2.1.2: Top: A schematic of the ribosome, the ribosome is made up of two subunits. These subunits encase the mRNA during translation. The ribosome contains ribosomal RNA, the catalyst for polypeptide bond formation. Bottom: Schematic of the transcription process within cells. The ribosome identifies an mRNA to translate by virtue of a start codon (black and white rectangles), it then reads the genetic information encoding a protein (a small polypeptide in this example). The tRNA with the relevant amino acid are attracted to the ribosome and bind to the complementary sequence on the mRNA strand. Once the rRNA has catalysed a peptide bond between the amino acids, the preceding tRNA molecule detaches from the mRNA strand (large arrow). This process continues and a polypeptide of amino acids is synthesised. The ribosome stops translation at a stop codon (red and white rectangles). This codon causes the ribosome to detach from the mRNA strand and release the polypeptide.

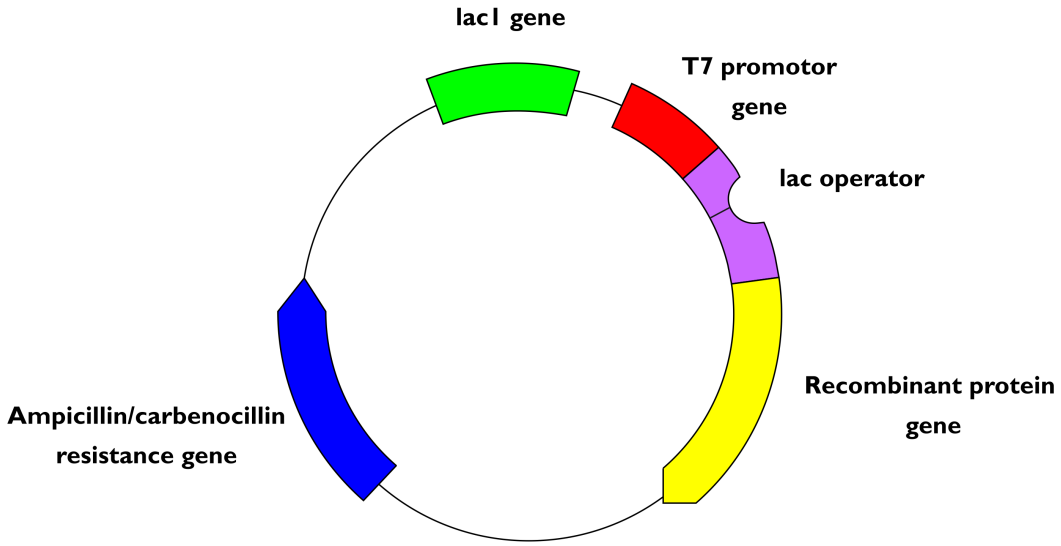


Figure 2.2.2.1: Figure showing pET vector, this vector encodes for specific antibiotic resistance (ampicillin or carbenicillin), a lacI gene that encodes the genetic information for a lac repressor protein, a T7 promotor for T7 RNA polymerase provided by the host bacterial strain, a lac operator and the gene of interest.

is responsible for blocking the transcription of the lac operon, a group of bacterial genes and a promoter gene. These genes encode the genetic information required to synthesise proteins that metabolise allolactose. The binding of the lac repressor protein to DNA is modulated by the presence or absence of allolactose. Typically bacteria use glucose as their energy source, lactose is only used when the presence of glucose is low. The lac repressor protein is continually expressed from the regulatory lacI gene. When the protein is expressed in the absence of lactose, it binds to an operator between the promotor gene preceding structural genes. This binding prevents the transcription of the genes by RNA polymerase. When lactose is present, allolactose binds to the repressor protein causing a conformational change in the protein structure. This change in conformation significantly reduces the affinity of the protein to the binding site. RNA polymerase can therefore transcribe the genes at a high rate due to the presence of the promotor gene, which has a high affinity for RNA polymerase.

The pET vector containing the genetic sequence for the POI contains the lacI gene and a T7 promotor site. The bacteria are genetically modified to contain the gene for expression of the T7 RNA polymerase, which is required to express the POI [284–286]. The transcription of the gene by T7 RNA polymerase is also controlled by binding of the lac repressor protein [287]. Transcription of T7 RNA polymerase, and therefore overexpression of the POI, occurs when isopropyl β-D-1-thiogalactopyranoside (IPTG),

a non-cleavable synthetic structural analogue of allolactose, is present, or an excess of lactose is present in the cell. IPTG displaces the repressor from the lac operator, present on both the pET vector and gene for the T7 RNA polymerase and this activates both genes. The T7 RNA polymerase can then proceed to transcribe the gene for the POI, causing the cell to over express the protein. This process is illustrated in Figure 2.2.2.2. The bacterial strains used in this project were BLR[DE3]pLysS and BL21[DE3]pLysS (Stratagene, UK). BL21 and BLR describe the strain of bacteria, [DE3] indicates that T7 RNA polymerase is present on the bacterial chromosome and pLysS indicated the presence of T7 lysozyme which reduces uninduced ("leaky") expression of proteins [288–290].

The pET vectors used in this project also contain the gene for antibiotic resistance to ampicillin or carbenicillin, allowing for selective growth of bacteria containing the correct DNA sequence for protein production.

2.2.3 Preparation of competent cells

For any bacteria strain to uptake the pET vector for protein production, the cells need to be made competent. This means treating the surface of the bacteria with chemicals so that the bacteria wall becomes more permeable and can uptake manipulated DNA plasmid. Competent cells were made using the calcium chloride method as previously described [291]. A 100 ml stock of 100 mM $CaCl_2$ was made and filter-sterilised using a 0.2 μm membrane[4]. A 5 ml stock of 100 mM $CaCl_2$, 30 % glycerol was also made and filtered through a sterile 0.2 μm filter. Both stocks were chilled at 4 °C before use. The relevant strain of bacteria was grown in 100 ml LB or ISOGRO-D growth media (97 % atom[5]) were made as described in section 2.1.1.1. The optical density at 600 nm (OD_{600}) of the bacteria was monitored during the growth using a UV spectrometer. Measuring the OD_{600} determines the growth of the bacteria; as more bacteria grow, the scattering of light at 600 nm increases and the OD_{600} reading therefore increases. Once the bacteria reached an OD_{600} of 0.4, the cells were spun down in a centrifuge at 4000 rpm (revolutions per minute) for 10 minutes at 4 °C to form a pellet. The remaining liquid, or supernatant, was removed from the pellet which was then re-suspended in 10 ml pre-chilled sterile 100 mM $CaCl_2$. The bacterial culture was then chilled for 10 minutes on ice before being centrifuged at 4000 rpm for 10 minutes at 4 °C. The supernatant

[4]This is typically done by flaming the top of the container with a Bunsen flame
[5]This refers to the percent deuteration of the carbon source

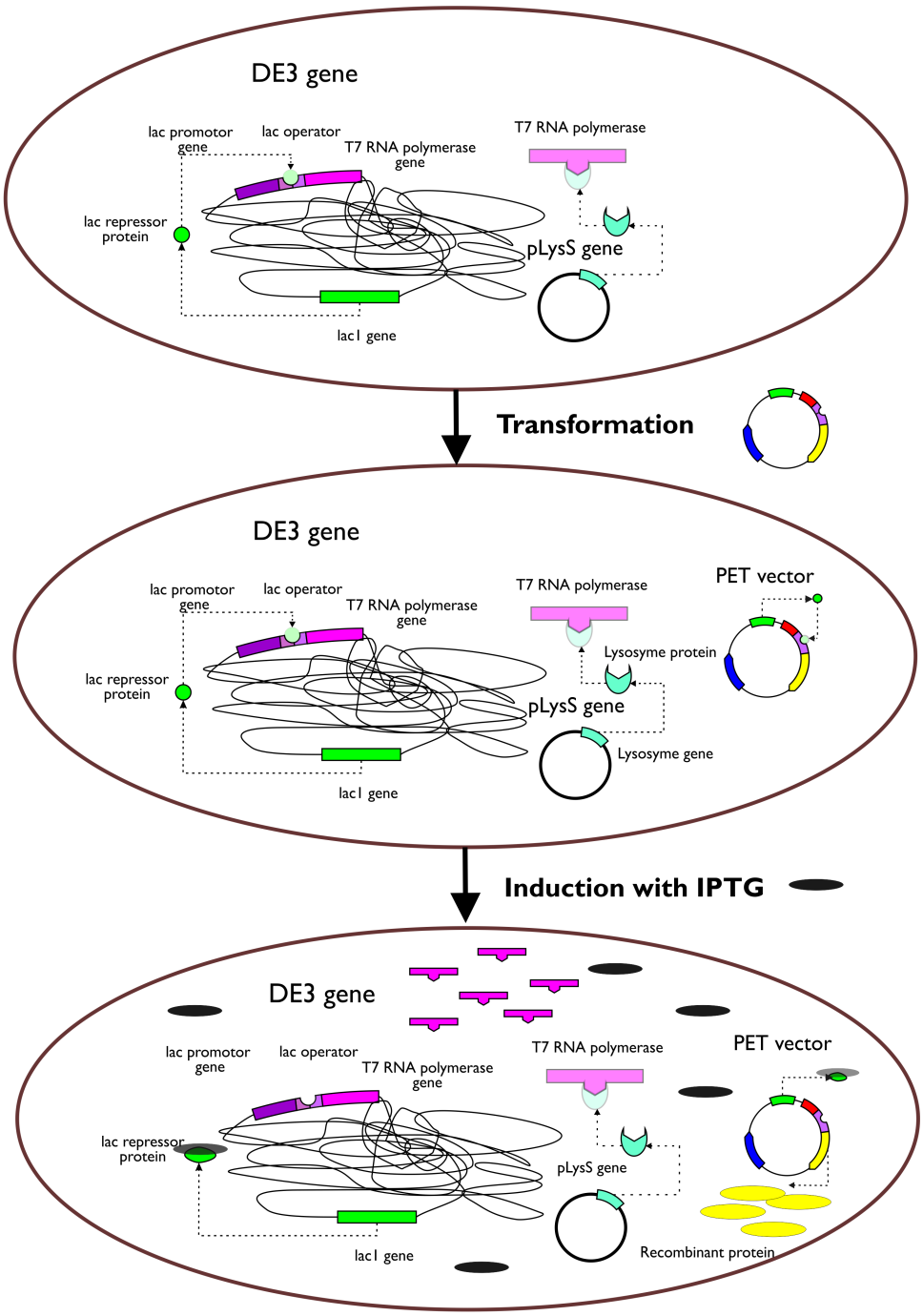


Figure 2.2.2.2: This schematic diagram illustrates how the T7 RNA polymerase (the RNA polymerase used to produce mRNA encoding the genetic information for a recombinant protein) production is controlled within each stage for DNA insertion and protein expression in DE3 *E.coli* bacteria. In this strain the *E.coli* genome has been genetically modified to contain the sequence for expression of T7 RNA polymerase. Top: The expression of the polymerase is controlled by the lac repressor protein. In the absence of lactose or IPTG (a structural analogue of lactose), the lac repressor protein binds to the lac operon preventing transcription of the T7 polymerase gene. These *E.coli* BL21 or BLR [DE3] cells also contain a pLysS gene which encodes the expression of lysozyme. Lysozyme can bind and inhibit T7 RNA polymerase. Middle: A pET vector encoding the genetic information for a recombinant protein can be transformed into the cell. This gene also contains a lac operon gene to prevent premature expression of the protein before induction with IPTG. Bottom: When IPTG is added to the cell, it binds to the lac repressor protein reducing the protein's affinity for the lac operon. This enables the transcription of T7 RNA polymerase and therefore the transcription of the gene for the recombinant protein. The pET vector is colour coded according to Figure 2.2.2.1.

was removed and the remaining pellet was re-suspended in 2 ml sterile 100 mM $CaCl_2$, 30 % glycerol. The solution was then divided into 100 μl aliquots in 0.5 ml eppendorfs that were pre-chilled on dry ice[6]. The competent cells were stored at -80 °C until required.

2.2.4 Transformation of cells

Once competent cells were made as described in section 2.2.3, these could then be transformed with the recombinant DNA. This involved the uptake of the DNA plasmid by heat shocking the bacteria. The bacteria could then be grown in a medium and, once induced with IPTG, can over-express the POI.

Initially 50 μl of competent cells, defrosted on ice, were transferred to a sterile 0.5 ml tube and 2-4 μl of 100 μg/μl DNA plasmid was added. The bacteria and DNA mixture was incubated on ice for 30 minutes with occasional gentle mixing. After 30 minutes the bacteria were subjected to a heat shock by incubation at a temperature of 42 °C in an unstirred water bath for 30 seconds then returned to ice. The higher temperature provides the heat shock required to increase the permeability of the bacteria wall which allows the diffusion of the plasmid into the cell. Returning the bacteria to ice reduces the permeability once more.

After 10 minutes of incubation on ice, 900 μl of sterile LB was added to the eppendorf and the bacteria were grown at 37 °C for 1 hour spinning at 200 rpm. This growth step allows the bacteria to express the antibiotic resistance protein (see Section 2.2.2) that makes the cell resilient to ampicillin or carbenicillin.

After growth for 1 hour the bacteria were plated on an LB or ISOGRO-D Agar plate containing either 100 μg/ml of carbenicillin or ampicillin. The bacteria were plated by pipetting 100 μl of bacteria into the center of an agar plate and then spreading using a sterilised glass rod. The glass rod was sterilised by dipping in 100 % ethanol and then flaming repeatedly. The glass rod was then allowed to cool before spreading the bacteria. The agar plates were incubated at 37 °C overnight. The transformation was successful if colonies of approximately equal size grew.

To check that the transformation was successful a control was carried out. This control involved repeating the procedure with the addition of water instead of DNA. The

[6]Solid CO_2

untransformed bacteria were also plated on an agar plate to check for colonies. There were no colonies for every control performed suggesting that the transformation was successful.

2.2.4.1 Plasmid purification and sequencing

The pET vectors were sequenced using Beckman Coulter Geonomics after their isolation from 10 – 15 ml of bacterial culture using a PureYeild plasmid miniprep kit (Promega, UK). This involved expressing the plasmid in the bacteria, lysing the bacteria and then purifying the plasmid from the bacteria on a spin column according to the manufacturers instructions. The protein were then eluted using nuclease-free water. The plasmid was diluted to 100 μg/μl for sequencing. The concentration of DNA, C_{DNA} was estimated using a UV spectrophotometer and Equation 2.4[292]. Where A_{260} and A_{320} are the absorbance of the DNA at 260 and 320 respectively, d is the dilution factor (typically 100). The multiplication accounts for the relationship that A_{260}=1 corresponds to 50 μg/ml of pure DNA. The sequenced verified DNA was stored at -20 °C.

$$C_{DNA} = (A_{260} - A_{320}) \times d \times 50 \tag{2.4}$$

2.2.5 Trial Expression

Once the plasmid was transformed into the relevant bacterial strain, test expressions were carried out to determine the optimum time, t_{opt}, of bacteria growth for maximum protein yield. The pET vectors encoding the genetic information for expression of the protein are discussed in Section 2.2.2.

Initially a single colony was picked from the agar plate containing the transformed bacteria. This colony was grown at 37 °C, shaking at 200 rpm, overnight in a flask containing 50 ml of sterile LB or ISOGRO-D medium (made as described in section 2.1.1) containing 100 μg/ml carbenicillin. 100 ml of sterile growth medium was incubated overnight at 37 °C in a 200 ml flask, for an LB test grow, or 500 ml flask for an ISOGRO-D test grow. Before sub-culturing[7] the bacteria, carbenicillin was added to the 100 ml of sterile pre-incubated media to a final concentration of 100 μg/ml. The OD_{600} was recorded of the overnight bacterial culture using a four times dilution. A 30 % glycerol stock was taken of the bacteria by mixing 0.3 ml 50 % glycerol with 0.7 ml of bacterial culture. The overnight culture of bacteria was then sub-cultured into the sterile media

[7]Adding the bacteria from the overnight grow of the transformed colony

to give an initial OD_{600} reading of 0.1 for the trial expression. This was then shaken at 200 rpm at a temperature of 37 °C until an OD_{600} of 0.6 was reached. At an OD_{600}=0.6 the bacteria were in an exponential phase of growth, this means that the bacteria were at optimum growth and would be producing chaperones that prevent misfolding of proteins. At this point a 1 ml sample was taken and protein over-expression was induced by inoculating the media with 1 M filter-sterilised IPTG to a final concentration of 1 mM. The OD_{600} of the 1 ml sample was recorded, OD^0_{600}, and this sample was pelleted by centrifuging the bacteria at 13000 rpm for 10 minutes and the supernatant is removed. The pellet was re-suspended using 100 μl of Bugbuster (Merck Millipore Corporation, Germany) which lyses the cells.

After inoculation with IPTG, 1 ml samples were taken every 1 hour (t=1) for up to 10 hours (t=10), and the OD_{600} of the samples were measured (OD^t_{600}). Each sample was pelleted by centrifugation at 13000 rpm for 10 minutes and the supernatant was removed before storing at -20 °C. The cell pellets were lysed by re-suspension in Bugbuster, the volume, V, of which was determined using Equation 2.5. An equal quantity of 2 x loading buffer was added to the sample.

$$V(\mu l) = \frac{100}{\mathrm{OD}^0_{600}} \times \mathrm{OD}^t_{600} \tag{2.5}$$

These samples were boiled for 2 minutes then centrifuged at 13 000 rpm for 2 minutes before incubation on ice. To determine the t_{opt} a sodium dodecyl sulphate polyacrlamide gel (SDS-PAGE gel) was run (see section 2.2.6). Between 10 - 15 μl of sample was loaded per well and one well contained 10 μl of Mark12 Protein Standard (Invitrogen, UK) which contains a series of proteins with known mass.

2.2.6 Sodium-dodecyl sulphate polyacrylamide gel electrophoresis

Tris-tricine buffered sodium-dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE gel) is the process of separating proteins based on their size. The proteins are separated through a gel mesh under the influence of an applied electric field. Proteins are fully unfolded by boiling in SDS buffer before loading. The SDS fully unfolds proteins to a linear state and coats the unfolded protein in a layer of negative charge. This coating in negative charge masks the charge on the side groups on the amino acids of proteins and also prevents the protein from refolding. This causes the proteins to all have an equal

mass:charge ratio which removes the effect of different migration speeds of proteins based on charge.

When an electric charge is applied across the gel, the proteins enter the gel and migrate towards a positive anode. The rate at which the protein migrates through the gel matrix is inversely proportional to the logarithm of the protein mass. A smaller protein will move quicker through the acrylamide gel matrix than a larger protein. To ensure all protein enters the gel at the same time, and not at a staggered rate, the gel contains a stacking region followed by a separating region.

The SDS-PAGE gel was made by mixing the chemicals given in Table 2.2.6.1, ammonium persulphate (APS) was added last. Once the APS was added to the separating gel mixture, the mixture was quickly mixed then poured into a two-glass plate system with a 1.5 mm spacer, leaving roughly 2.5 cm gap to the top. Glycerol was added to this layer to prevent mixing of the two components of the gel. The APS was then added to the stacking gel buffer and rapidly poured on top. A 14-tooth comb was added to the top of the gel to provide the wells for sample loading. The plates were left to set and typically set after an hour. Once the gel was set, the gel plates were stored in an air tight container at 4 °C after wrapping in a damp cloth.

	Stack/resolving gel	Separating gel
Components	Volume (ml)	
30 % Bis-acrylamide	7.5	0.83
3M Tris.Hcl, 0.3 % SDS, pH 8.45	5	1.55
H_2O	0.44	3.72
Glycerol	2	-
10 % ammonium persulphate (APS)	0.15	0.1
Tetramethylethylenediamine, TEMED	0.015	0.005

Table 2.2.6.1: Table indicating the volume of buffers that formed the resolving and separating components of an SDS-PAGE gel. The APS was omitted until all other components were added.

The SDS buffered gel was subjected electrophoresis in a Dual Slab Gel Electrophoresis Chamber (ATTO, Japan). The anode buffer was poured into the inner reservoir and a cathode buffer was poured into the outer reservoir. The anode and cathode buffers are made according to Section 2.1.1. Samples are made by mixing the sample with 2 × loading buffer (Section 2.1.1) in equal volumes. The appropriate sample volume (typically 10-20 μl) was pipetted into individual wells. Typically one well was used to run a pre-stained protein mass reference ladder (Mark12 Protein Standard, Invitrogen, UK). Once

the samples were loaded a Standard Power Pack P25 (Biometra, Germany) was attached to the gel chamber. A current of 30 mA was initially applied across the gel plates until the sample was suitably stacked. After about 30 minutes the current was increased to 60 mA until the dye reached the bottom of the gel. The gels were then stained using Instant Blue Stain (Expedeon, UK) for at least 1 hour before being transferred to water to remove excess stain. The gels were photographed using either a Syngene InGenius gel documentation system or a flat-bed scanner.

2.2.7 Full-scale Expression

The optimum growth time (t_{opt}) for maximum protein yield was determined as described in Section 2.2.5.

2.2.7.1 LB large scale grows

A flask containing 200 ml of sterile LB, 100 μg/ml carbenicillin and 50 μl of glycerol stock from section 2.2.5 was incubated (37 °C, 200 rpm) the night preceding the large scale grow. 20-25 ml of this overnight culture was sub-cultured into each of the 10 2 litre flasks containing 1 litre of pre-incubated (37 °C) LB medium after the addition of carbenicillin to a final concentration of 100 μg/ml. This typically gave a starting OD_{600} of 0.1. The cultures were incubated (37 °C, 200 rpm) and the growth was monitored using OD_{600}. At an OD_{600}=0.6, 1 M IPTG was added to a final concentration of 1 mM. The cultures were left to grow for t_{opt} before pelleting the cells by centrifugation at 16 000 rpm (Avanti J-E centrifuge, JLA-16.250 rotor, Beckman Coulter, UK) at 4 °C. After centrifugation, the supernatant was removed and disposed and the protein is extracted as described in Section 2.2.8.

2.2.7.2 ISOGRO-D large scale grows

Two flasks containing 100 ml of sterile ISOGRO-D (made according to Section 2.1.1), 100 μg/ml carbenicillin and 50 μl of glycerol stock from section 2.2.5 were incubated (37 °C, 200 rpm) the night preceding the large scale grow. 10–15 ml of this was sub-cultured into each of the 10, 500 ml flasks containing 100 ml of pre-incubated (37 °C) ISOGRO-D media (made according to Section 2.1.1) after the addition of carbenicillin to a final concentration of 100 μg ml^{-1}. The volume added was chosen to provide a starting OD_{600} of 0.1. The cultures were then incubated (37 °C, 200 rpm) and the growth was monitored using OD_{600}. At an OD_{600}=0.6, 1 M IPTG (dissolved in D_2O) was added to give a final concentration of 1 mM. The cultures were left to grow for t_{opt} before pelleting the cells by centrifugation at 16 000 rpm (Avanti J-E centrifuge, JLA-16.250 Beckman Coulter,

UK) at 4 °C. After centrifugation, the supernatant was removed and disposed of and the protein was extracted as described in Section 2.2.8.

2.2.8 Cell lysis

After the bacteria were pelleted, the cells were then lysed to extract the recombinant protein. Protease inhibitor cocktail (Section 2.1.1) was added to the lysis buffer (Section 2.1.1) to give a final concentration of 1 mM PMSF and 2 mM Benzamidine. Complete EDTA-free protease inhibitor tablets (Roche, USA) were also added to the lysis buffer according to the user manual. Protease inhibitor tablets were used to ensure minimal degradation of the protein during the purification process.

Different volumes of lysis buffer were added to the cell pellets depending on the volume of the large scale grow: 100 ml of lysis buffer made as described in Section 2.1.1 was added to the cells pelleted from a 10 litre grow, 20 ml was added to cell pellets from a 1 litre grow and 30 ml was added to cell pellets from a 2 litre grow. Small amounts of DNAase [8], RNAase [9] and lysozyme [10] were added. The solution was stirred until minimal DNA and RNA was left in the sample. The solution was then homogenised using an electronic disperser, to disrupt any solid left. The cells were then subjected to a cell disruptor or tip sonicator to ensure maximum extraction of protein.

The solution was then centrifuged at 16 000 rpm for 30 minutes to separate the insoluble and soluble material of the lysate. The supernatant (which should contain the protein) was separated from the insoluble material, both lysate and pellet were stored at - 20 °C before purification.

2.2.9 Purification of proteins

All proteins were subjected to two stages of purification, all purification steps were performed at room temperature. Both stages of purification are described in this section.

2.2.9.1 Ni-NTA protein purification

Nickel-nitrilotriacetic acid, Ni-NTA, protein purification is a type of immobilised metal-affinity chromatography (IMAC) [293, 294]. In this purification step a transition metal (Ni^{2+} in this case) is immobilised on a matrix and free to bind to the side chain of

[8] Breaks down DNA molecules
[9] Breaks down RNA molecules
[10] Breaks down the cell wall (peptidoglycan layer) of bacteria

histidine. In this study proteins were previously genetically engineered to contain a hexa-histidine[11] tag at the C-terminus of the construct. The imidazole group has a high affinity for Ni^{2+} ions. The electron donor groups on the imidazole side chain of histidine is readily available to bond to Ni^{2+}. As the proteins contain six histidines, the proteins can strongly interact with the IMAC column matrix. Changing the pH or introducing large quantities of imidazole can disrupt this interaction and unbind the protein, allowing it to be eluted. It has been reported that this method of purification can increase the purity of the protein to 95 % whilst minimising the loss of protein sample. The matrix used in this project was Ni^{2+}-NTA, coupled to a solid support resin.

The cells were lysed as described in 2.2.8 and the lysate was defrosted on ice. Lysis, washing and elution buffers were made as described in Section 2.1.1 with the addition of 100 x protein inhibitor cocktail to give final concentrations of 1 mM PMSF and 2 mM Benzamidine. All buffers were then filtered through a 0.4 μm filter using a vacuum pump. Depending on volume of lysate to be loaded, either a, XK 50/20 column containing 30 ml Ni Sepharose resin, 5 ml Ni-NTA column or 1 ml Ni-NTA column (all G.E. Healthcare, UK) was used to perform the IMAC purification. The columns were attached to an ÄKTA Prime [12] which enabled monitoring of the sample A_{280}. The A_{280} is the absorbance of the sample at 280 nm. Proteins absorb light at a wavelength of 280 nm when they contain tryptophan, tyrosine or cysteine disulphide bonds. This allows accurate measurement of protein concentration for a known extinction coefficient and enables protein peaks to be distinguished during the elution stage.

Initially the columns were washed with 10 $\times$ column volume of filtered and de-gassed mili-Q water. The columns were then equilibrated into lysis buffer, from Section 2.2.8, until the A_{280} remained constant. At this point the A_{280} was baselined by subtracting the constant A_{280} reading from the buffer. The lysate was filtered through a 0.2 μm membrane filter before loading on the column to prevent blockages and bacterial contamination. There are three stages to the purification process: loading, washing and elution. All flow-through from the column was collected for testing on SDS-PAGE gels (see section 2.2.6).

Loading: The samples were loaded onto the column at a flow rate according to the manufacturers instructions (typically 1 ml/min or slower to ensure maximum binding).

[11] A chain of six histidines, an amino acid containing an imidazole group
[12] A pump for buffers and sample

Washing: After the lysate was loaded onto the column, the column was washed with column washing buffer until the A_{280} returned to the baseline (approximately).

Elution: Elution buffer was introduced at different concentrations after the washing stage. Initially a concentration of 25 % elution buffer was introduced to the wash buffer. Mixing was automated using the ÄKTA Prime. This was done until the A_{280} was constant. The concentration of elution buffer was then increased to 50 % and 100 % until the A_{280} is constant. During this step, it is assumed that protein is eluted when there is a presence of a peak in the A_{280} reading.

20 μl samples were taken from the loading, washing and elution stages, including any peaks in the absorbance. These samples were mixed with 20 μl of 2 $\times$ loading buffer. These samples were loaded on a SDS-PAGE gel to analyse protein recovery, yield and purity. The purification was repeated for flow-through of the loading and washing steps if necessary. Elution peaks containing protein of similar purity were pooled and either: dialysed (see section 2.2.9.3) and lyophilised into approximately 10 mg aliquots, or concentrated to $\sim$ 10 mg/ml using a Amicon Ultra Centrifugal filter (3 000 MWCO[13]) to be used in the next stage of purification (Section 2.2.9.2).

2.2.9.2 Size-exclusion chromatography

Size-exclusion chromatography (SEC), also known as gel filtration chromatography (GF), is a purification technique which separates proteins according to their size. The gel matrix that the column is comprised of is made up of porous beads which buffers/liquids can flow through or around (>95 % of the column is liquid volume). The rate at which a protein elutes from the column is dependent on how much access the protein has to the pores in the beads. A larger protein will travel through the column faster than a small protein because they are less likely to travel through pores in the beads; smaller proteins can travel through the beads and therefore take a longer path through the column. Aggregates and very large proteins typically elute from the protein very fast and are seen in the void volume[14] [295].

The advantage of this purification technique is that it can be used to remove impurities or desalt proteins. The column used in this project is a Superdex 75 (SUP-75)

[13]Molecular weight cut off
[14]Volume of the column occupied by the beads

gel filtration column (GE Healthcare, UK). The matrix of this column consists of beads made from highly cross-linked dextran and cross linked agarose, with an average particle size of 34 μm. It is functional over a large range of pH (2–13) and is chemically inert in a range of buffers and conditions (according to the manufacturers guidelines).

The SUP-75 column was attached to an ÄKTA prime for purification. Initially the column was rinsed with water at 1 ml/min and then equilibrated in 400 ml of gel filtration buffer (see Section 2.1.1). A 5 ml sample loop was attached to the ÄKTA and flushed through with buffer following the manufacturers instructions. The samples were filtered using a 0.4 μm filter and loaded onto the column. Samples were loaded at concentrations of $\sim$ 10 mg/ml in 3 ml aliquots via the 5 ml sample loop. The protein was loaded onto the column at a rate of 2 ml/min. The flow-through was collected in 2 ml fractions, protein elution was observed by peaks in the A_{280} output of the ÄKTA. To test which peak corresponded to the protein of interest 10 μl samples of fractions across each peak were run on an SDS-PAGE gel for analysis (see section 2.2.6). The relevant fractions containing pure protein were pooled, dialysed and lyopholised as described in Section 2.2.9.3.

2.2.9.3 Sample dialysis and preparation

Dialysis involves the removal of salts and other small impurities from protein samples using osmosis. Dialysis tubing is a semi-permeable material which allows small molecules to pass through the wall, whilst keeping larger molecules inside. Salts are removed by placing the protein and salt solution inside dialysis tubing and introducing it to a large volume of water. The salts will then permeate across the membrane due to the concentration gradient until the gradient no longer exists. Most of the salts can be removed by repeatedly changing the dialysis water.

Dialysis of protein was performed in Snakeskin dialysis tubing (3 kDa MWCO) for volumes larger than 5 ml using Mili-Q water. Addition of 20 mM ammonium bicarbonate was added for proteins with low solubility. For volumes less than 5 ml the proteins were dialysed in Maxi or Midi Protein dialysis kits following the manufacturers instructions. After 3 changes of dialysis water, the protein concentrations were checked using the sample absorbance at A_{280} as discussed in Section 2.3.3.1. Protein purity was determined from the ratio of A_{260} to A_{280}. Pure protein gives a ratio of 0.57 whereas pure DNA or RNA give a ratio of 2.00. If there was contamination in the sample additional purification steps can be run, including desalting columns (if contaminated with small salt molecules), concentrators with a large cut off and analytical gel filtration (not discussed here). Once the purity was

verified, the proteins were aliquoted into the relevant quantities for experiments before being rapidly frozen using liquid nitrogen or dry ice mixed with ethanol. Frozen samples were lyophilised on a freeze dryer.

2.2.10 Storage of proteins

After purification proteins were stored at -20 °C until use. The quantity of protein required for each experiment is given in Table 2.2.10.1.

Experimental method	Sample mass (final concentration)
Atomic force microscope (Section 2.3.4)	0.05mg/0.1mg (0.1 mg/ml)
Circular Dichroism (Section 2.3.2)	0.3 mg (0.3 mg/ml)
Fluorescence (Section 2.3.3)	4 mg (0.5 mg/ml)
Protein preparation (Section 2.2.11)	5-10 mg (5 mg/ml)

Table 2.2.10.1: Table indicating the appropriate sample aliquots for each experimental technique.

2.2.11 Protein preparation

Within this project there were four test systems being characterised for each protein:

p(protein name)$_5$[15]or p(protein name)[16]in H_2O

Fully protiated protein in a H_2O solvent environment

p(protein name)$_5$ or p(protein name) in D_2O

Fully protiated protein in a D_2O solvent environment

d(protein name)$_5$ or d(protein) in H_2O

Fully deuterated protein in a H_2O solvent environment

d(protein name)$_5$ in D_2O

Fully deuterated protein in a D_2O solvent environment

Where p indicates a protein with hydrogen in all non-exchangeable[17] hydrogen positions and d indicates a protein containing deuterium in all non-exchangeable hydrogen positions. To minimize the effects of exchange during characterisation, all proteins were incubated in the appropriate buffer to ensure all other positions were populated with either H or D at the start of each experiment.

[15]for the polyprotein containing five proteins
[16]for the monomer
[17]Carbon backbone and nonpolar side chains such as methyl

Initially the lyophilised protein was resuspended in the relevant unfolding buffer (see Section 2.1.1). The samples were left overnight to ensure the protein was fully exchanged at the relevant pH or pD. The protein was then desalted into pure Mili-Q H_2O or D_2O using a desalting column. The protein was then dialysed in either Mili-Q H_2O or D_2O depending on the system. The samples were then aliquoted into the correct quantities based on their concentration and lyopholised as described in Section 2.2.9.3.

2.3 Characterisation of protein constructs

2.3.1 Mass spectrometry

The mass of all protein constructs was determined by electrospray ionisation mass spectrometry (EI-MS)[18], 0.05 mg of proteins were unfolded in a 50 % (v, v) acetonitrile, 0.1 % (v, v) formic acid pH $\sim$ 2.5 in water. To determine the level of deuteration of protein constructs, the proteins were left to exchange and the mass was measured every 24 hours until the mass remained constant. After purification the mass was typically constant after 24 hours.

2.3.2 Circular dichroism spectroscopy

Circular dichroism spectroscopy uses the interactions of chiral molecules with circularly polarised light to determine information about their structure. Molecules can be chiral in nature, in particular this is seen for all amino acids with the exception of glycine. Chiral molecules have mirror image isomers, which cannot be superimposed on one another and can occur for three reasons:

(a) The chirality is as a consequence of the structure, for example, the C atom with bonds with four different substituents, and the disulphide bonds due to dihedral angles between the atoms.

(b) There is a covalent bond to a chiral centre in the molecule.

(c) The 3-D structure adopted by the molecule causes the molecule to be placed in an asymmetric environment.

The only distinguishing features between these enantiomers are the different absorptions of polarised light and the way they interact with other chiral molecules [296].

[18] Mass spectrometry experiments were performed by Dr James Ault

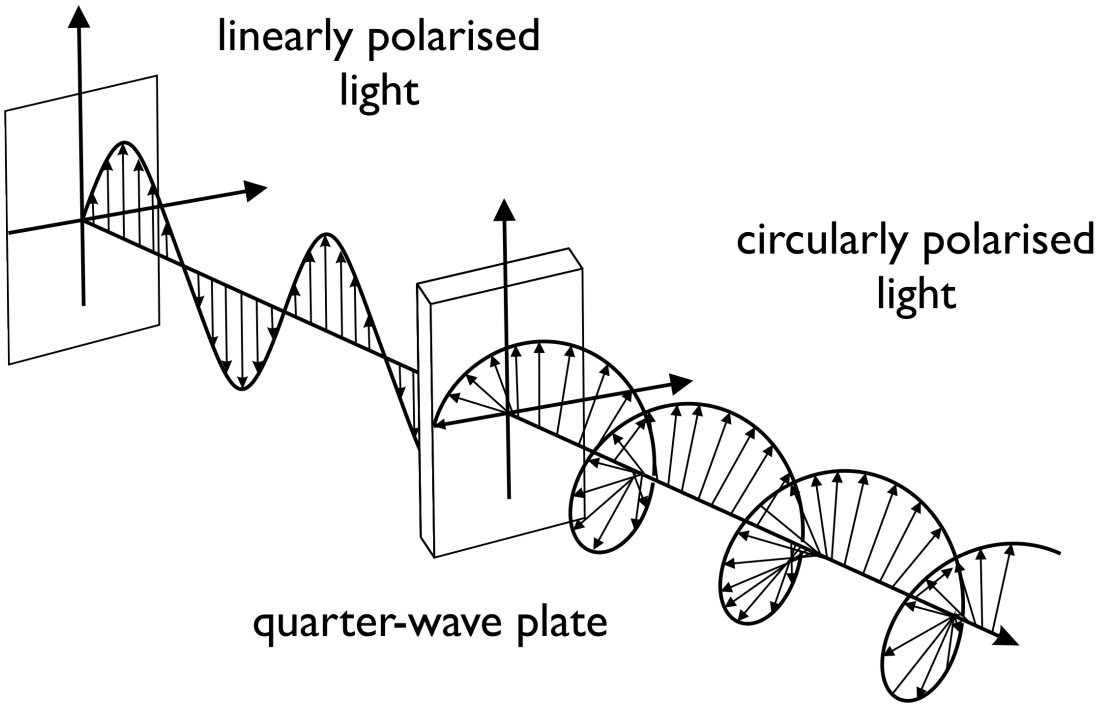


Figure 2.3.2.1: Schematic diagram illustrating the production of circularly polarised light. In linearly (or plane) polarised light, the resultant vectors making up the light wave lie in a single plane. After the light passes through a quarter-wave plate, the the vector is continuously rotating about the axis because the vectors making up the light pass through the plate at different speeds.

Linearly polarised light is light where the oscillations of the electric field are in a single plane. The electric field will be made up of two vector components orthogonal to one another with equal amplitude and phase. Circularly polarised light occurs when linearly polarised light is passed through a quarter-wave plate. The quarter-wave plate is made of a birefringent material which causes the vectors comprising the electric field of the linearly polarised light to be 45° out of phase with one another. This phase difference between the vectors result in a constantly changing electric field causing circularly polarised light (Figure 2.3.2.1). Circular dichroism spectroscopy (CDS) measures the difference in absorption of left (counter clockwise) and right hand circularly polarised light (clockwise), $\Delta A = A_L - A_R$ [296], when is passes through a sample. If the left-hand and right-hand circularly polarised light were not absorbed at all then the two waves would produce the same circularly polarised light that was incident on the sample. However when chiral molecules are present, the left and right polarised light will be absorbed in different amounts and therefore the emitted light will be different to the incident light. CD

instruments typically report the ellipticity, θ, of a sample where $\theta = \arctan(a/b)$ where a and b are the minor and major axis of the elliptical light respectively. The ellipticity can be calculated by $\theta = 32.98\ \Delta A$ [296]. To obtain a CD spectrum, the ellipticity is recorded as a function of wavelength of the radiation propagating through the sample. The set up of a typical CD instrument is shown in Figure 2.3.2.2.

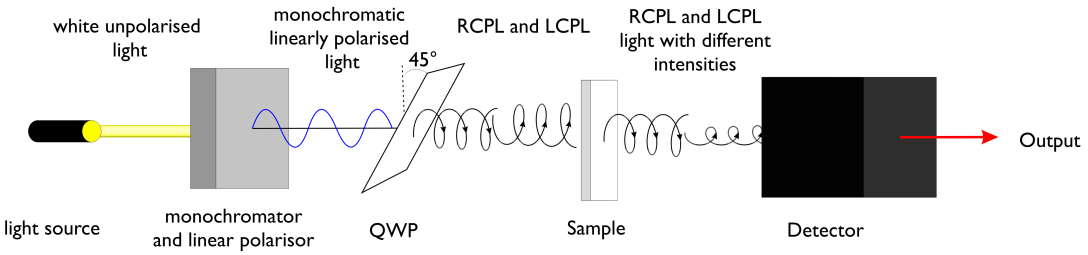


Figure 2.3.2.2: Schematic showing the principles behind a CD experiment. Initially white light is passed through a monochromator, to obtain a single wavelength of light, and a plane polariser, to obtain light where the oscillations of the electronic field are in a single plane. This plane polarised light is then passed through a quarter-wave plate to produce right hand (clockwise) and left hand (anticlockwise) circularly polarised (RCPL and LCPL respectively) light at a rate of 50 kHz. This light then passes through the sample and the change in RCPL and LCPL is determined and converted to ellipticity of the sample.

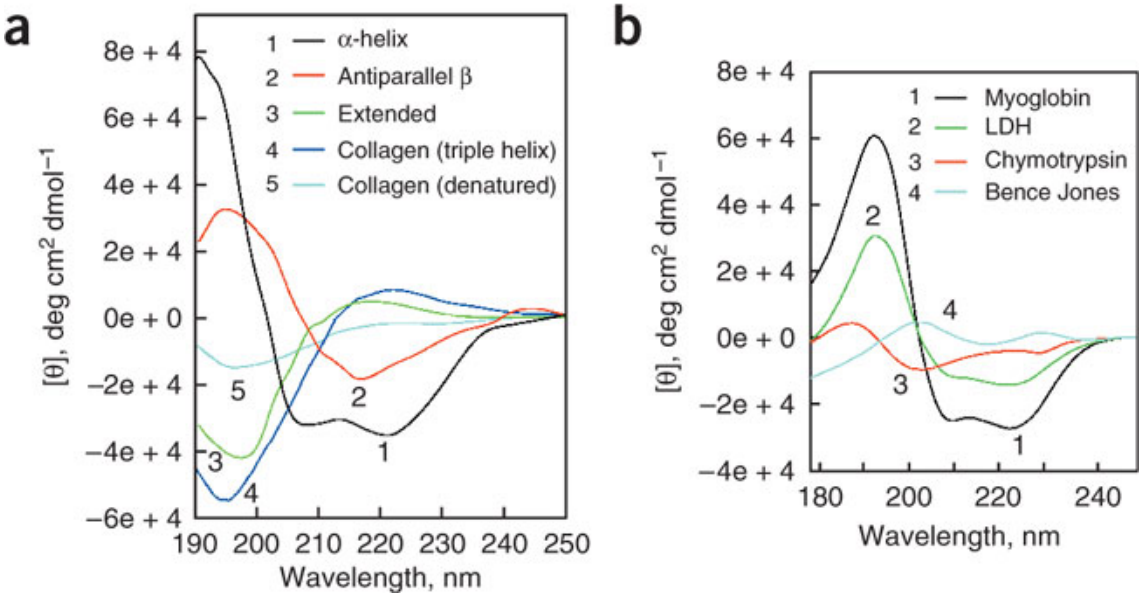


Figure 2.3.2.3: Figure showing examples of CD spectra for different protein secondary structures. a) Example of far UV spectra taken for different types of structure. 1.(black line) shows a typical spectrum for a protein containing α-helices. 2. (red line) shows the spectrum obtained for a β-sheet conformation. 3. (green line) is indicative of a disordered polypeptide. 4. (blue line) shows a triple helical domain and 5. (cyan line) shows a denatured polypeptide. b) Example spectra for proteins with different conformations showing the sensitivity of the technique to different protein structures. Reprinted by permission from Macmillan Publishers Ltd: Nature Protocols ([297]), copyright (2006)

Peptide bonds, aromatic amino acids and disulphide bonds absorb radiation below 240 nm, in the range 260 – 329 nm (resulting in a broad peak which is weak in intensity)

respectively, making it possible to study proteins using CDS. In addition, ligand binding and cofactors[19] can also cause absorption of radiation over a large range of wavelengths [298]. The absorption of the different wavelengths can give information about different properties of protein, including information about the type of secondary structure (see Figure 2.3.2.3)[297], tertiary structure [299], and thermodynamic parameters [300]. For the latter, if the protein is thermally unfolded, the unfolding of the protein must be reversible and the specific heat capacity of the protein must be known or easily estimated. Figure 2.3.2.3 shows the spectra obtained from proteins containing different types of secondary structure. These spectra arise because of the absorption of radiation by the peptide bond below 240 nm. This region is known as the far UV CD spectrum [296]. This is the region that has been probed in this project. The tertiary structure of the protein can be studied using the near UV CD spectrum of the sample [296]. This region probes the absorption of radiation between 260 and 320 nm. In this region the aromatic amino acids such as tryptophan (Trp) give rise to characteristic wavelength profiles depending on the number of aromatic amino acids, their mobility and the nature of their environment (including the H-bonding, polar groups and polarisibility). When proteins are subjected to a change in temperature, for example, they can undergo conformational changes where the type of structure will change. This would result in a different wavelength profile. These changes can therefore be monitored using CDS and information about the melting temperature, T_m can be obtained. T_m is the temperature at which 50 % of the protein domains in the sample are fully unfolded. This value can be obtained by following the change in the ellipticity of the protein at a single wavelength. The wavelength chosen is often the wavelength at which the largest change of ellipticity is observed as a function of temperature. If a protein undergoes a conformational change upon melting, the change in ellipticity, as a function of temperature, will give rise to a sigmoidal curve. The midpoint of this sigmoidal curve will determine the T_m of the protein [300].

CD was performed using on of the polyprotein and monomeric systems studied in this project using a Chirascan CD Spectrometer. Samples for CDS were made up by re-suspending 0.3 mg of lyophilised protein in 1 ml of the relevant sodium phosphate buffer to give a final concentration of 0.3 mg/ml. For the polyprotein constructs the buffer was 63 mM sodium phosphate pH or pD 7.4. For the monomers the buffer was either 25 mM sodium phosphate, 2 mM DTT, pH or pD 7.4 or 50 mM sodium phosphate pH 7.4 or pD 7.4 depending on the protein. Initially a reference was taken for the CDS in air, followed by a CD spectrum of the buffer. This buffer spectrum was taken for every sample and

[19]Cofactors are molecules required for the function of a protein

subtracted from all protein CD spectra. The range of wavelengths measured was 180 - 280 nm using an increment of 1 nm. Initially a spectrum was taken at 23 – 25 °C to determine whether the structures of the systems were correctly folded. Three repeats were taken for each sample. The spectra of the different systems were normalised to account for the different concentrations, $C_{mg/ml}$, of protein samples (measured by A_{280}) using Equation 2.6[301].

$$\Delta [\theta]_{mrw} = \frac{\theta \times MRW}{10 \times C_{mg/ml} \times L} \tag{2.6}$$

Where MRW is the mean residue weight, computed from the molecular weight of the protein divided by the number of residues and L is the path length of the cuvette (0.1 cm).

After spectra were acquired a temperature ramp was performed from 10 °C to 85– °C in 2.5 °C steps for the polyprotein or 1 °C steps for the monomers. The change in ellipticity at a single wavelength (corresponding to the largest change in ellipticity of all samples or a structural feature on the spectra) was followed as a function of temperature. Two-state unfolding pathways were observed for all proteins evidenced by a single transition in the sigmoidal curve. The midpoint of the transition corresponds to the T_m. The sigmoidal curves were fitted by Boltzmann sigmoidal function accounting for slopes in the baselines before (pre-transitional) and after (post-transitional) the protein unfolded. The sigmoidal curve used to fit the data is given in Equation 2.7. θ_{i} and θ_{f} are the initial and final ellipticity of the sample respectively, w is the slope of the transition between the folded and unfolded states and m is the slope of the pre and post translational baselines (assumed the be the same).

$$\theta = \theta_f + \frac{(\theta_i - \theta_f)}{1 + \exp\left(\frac{T_m - T}{w}\right)} + m \times T \tag{2.7}$$

2.3.3 Studying Proteins using Fluorescence

Fluorescence uses the wave-particle duality properties of light to study excited states of electrons [302]. A beam of light can be thought of as a continuous stream of photons with fixed quanta of energy. Each photon may only interact with one atom or molecule - therefore the quantum of energy available to each molecule is directly determined from the energy of the photon. By increasing the intensity, and therefore the flux of incoming photons, the number of molecules that the photons interact with can increase.

Molecules have discrete energy states that are separated by a specific energy difference, $\Delta E = h \times \nu$, where h is Plank's constant and ν is the frequency. For a molecule to be in an excited energy state it must be subject to a specific quantum of energy that equates to ΔE. Molecules will therefore only be excited by light of a certain frequency. A molecule can possess different quantised internal energies such as, rotational, vibrational and electronic energy [302]. Translational energy is not quantised but only accounts for 4 % of the total energy of the molecule [303]. Atoms may only have electronic internal energies - an increase in electronic energy involves the movement of electrons to higher energy orbitals. The type of internal energy excited depends on the frequency (or wavelength) of light incident upon it. It is the electronic transitions in molecules that are probed using fluorescence by using the ultraviolet region of light [302].

For the electronic energy of a molecule to increase, the electron must be excited from a lower energy state to a higher energy state by a photon with energy corresponding to the energy difference between the two electronic states. Electrons can also relax to lower energy states from higher energy states by emission of light with energy equal to the difference between the states. The type of emission from excitation of electrons is dependent on the path the electron takes upon relaxation. A schematic diagram of this process is shown in Figure 2.3.3.1. Stimulated emission is the emission that results from the excitation of the electron by a stimulus. Spontaneous emission can also occur, this is a random process and can occur in the absence of a stimulus. Fluorescence, phosphorescence and chemiluminescence are usually spontaneous. As spontaneous emission is a random process, the rate of emission is proportional to the concentration of excited species and is thus a first order kinetic reaction [302].

The ability of a molecule to fluoresce can be defined by its quantum yield. Quantum yield describes the "*number of molecules of reactant consumed per photon of light absorbed*" (Wayne, 1996). If a system is efficient the quantum yield will be 1 - this would mean that for every photon absorbed, a photon was emitted. Three amino acids have been found to have relatively high quantum yields upon excitation - tryptophan (0.20), tyrosine (0.21) and phenylalanine (0.04) [304]. The quantum yield of these amino acids is dependent on the environment of the residue including the position within a protein. This enables the conformation of a protein to be monitored as it is denatured [305].

In this project the tryptophan emission for protein molecules was monitored as a function of chemical denaturant. Many studies have been performed to determine the

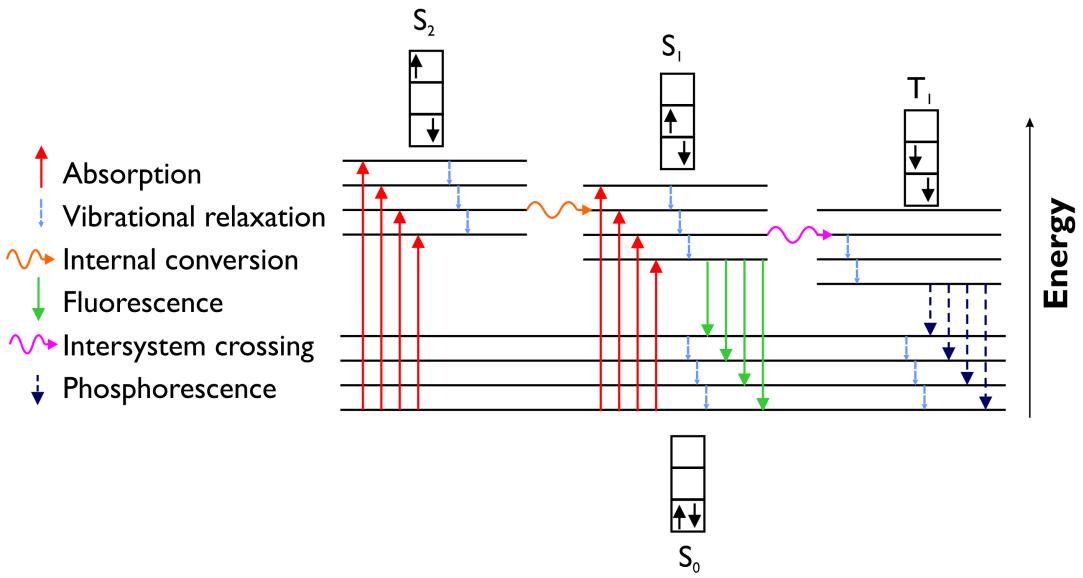


Figure 2.3.3.1: Schematic diagram illustrating the excitation and relaxation of electron which leads to fluorescence. Initially an electron is excited by a photon which has an energy equal to the difference between the energy state the electron originates in and the excited energy state. Relaxation of the excited electron can take different paths and the wavelength of the photon released depends on the path of relaxation.

response of tryptophan emission to the environment of the amino acid [306–309]. In this project protein samples were excited at a wavelength of 280 nm and the emission spectra of the protein was monitored for wavelengths of 300 – 380 nm. Tryptophan fluorescence was used to monitor the protein species present when samples are left to reach equilibrium in different concentrations of denaturant. It was also used to determine the kinetic response of proteins when denaturant was introduced to the system, or, when the protein was rapidly diluted to low concentrations of denaturant. The typical set-up of the equilibrium denaturation experiment is shown in Figure 2.3.3.2.

2.3.3.1 The Beer-Lambert Law

$$\frac{I_t}{I_0} = 10^{-\varepsilon C d} \tag{2.8}$$

The Beer-Lambert law, Equation 2.8 [302], uses the relationship between the fraction of the intensity of transmitted monochromatic light (I_t) compared to incident light (I_0) by a sample with concentration C, through a path length, d with a known extinction coefficient[20], ε. This expression can be used to determine protein concentrations in experiments. This relationship uses the assumption that the absorption of the light is proportional to the concentration of the species [302].

[20]The ability of a molecule to absorb light.

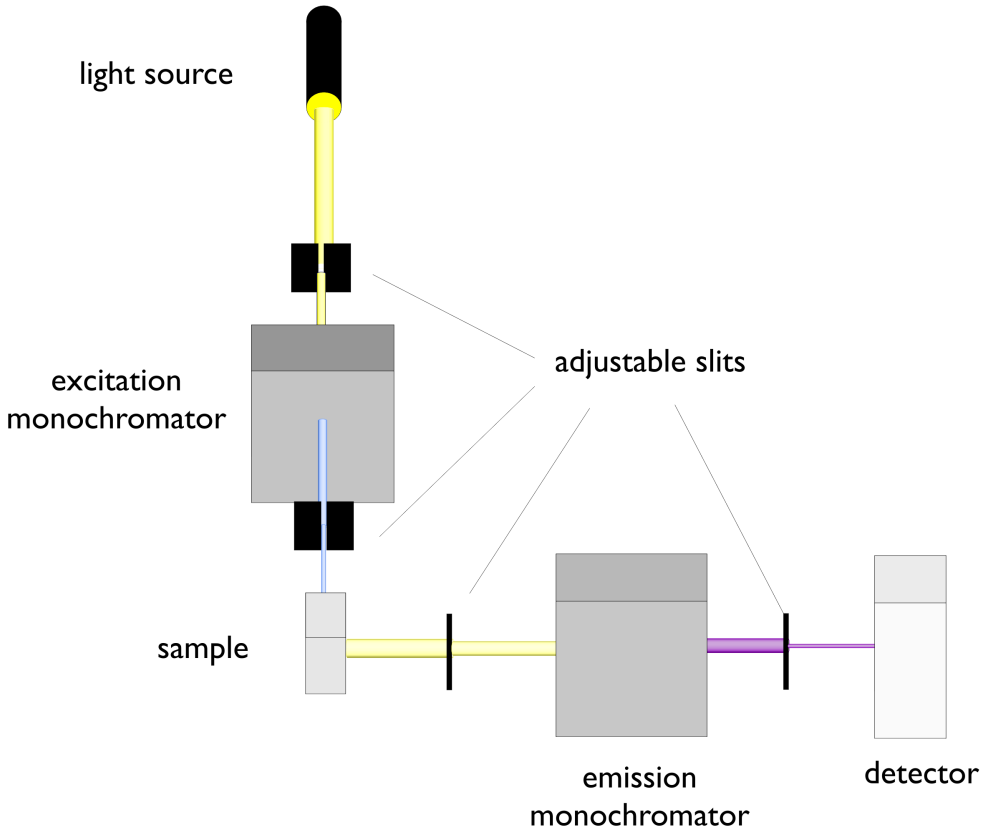


Figure 2.3.3.2: Schematic diagram illustrating the typical set-up in a fluorescence experiment. Initially light, containing a broad spectrum of wavelengths, passes through an excitation monochromator which isolates a single wavelength of light. The light is then passed through the sample and the light emitted from the relaxation of electrons in the sample is passed through another monochromator to allow monitoring of the signal for different wavelengths of light. This passes through a detector which determines the intensity of photons for each wavelength. The amount of light transmitted to the sample and detector is controlled by adjustable slits.

Extinction coefficients[21] for the proteins studied in this project had previously been determined using the method by Gill and Von Hippel explained elsewhere [310]. It was assumed to be unperturbed upon deuteration.

2.3.3.2 Equilibrium denaturation

Equilibrium denaturation measures the change of the fraction of folded protein as a function of denaturant to determine the free energy of unfolding as a function of denaturant concentration. The protein is allowed to reach equilibrium in the denaturant before measuring the fluorescence. In this study the proteins were equilibrated overnight at 25 °C in a waterbath.

Initially 100 ml of equilibrium folding buffer was made according to Section 2.1.1.1. This buffer was used to dissolve the unfolding buffers also given in Section 2.1.1.1. A 10 × stock of protein was made by dissolving 4 mg of lyophilised protein in the folding buffer to a final concentration of 0.5 mg/ml. This was filtered using a 0.2 μm filter to removed any aggregated protein. Substocks were then made by combining the correct amount of folding buffer, unfolding buffer and the 10 × protein stock. Each substock containined a final protein concentration of 0.05 mg/ml. The substocks for urea had final urea concentrations of 0 M, 2 M, 4 M, 6 M and 8 M (maximum denaturant concentration). The substocks of GdnHCl had final GdnHCl concentrations of 0M, 2 M, 4 M, 6 M and 7 M (maximum denaturant concentration). The denaturant concentrations of all substocks were checked using a refractometer. The refractometer is used to determine the refractive index of the 0 M denaturant buffer, n_{0M} and the buffers containing denaturant n_{xM}, where x is the expected concentration of the substock. The change in refractive index Δn can be used to determine [denaturant][22] using the relationships given in Equation 2.10. The true concentrations of denaturant were used to correct for the concentration of denaturant in each sample.

$$[\text{Urea}] = 117.66\Delta n + 29.75(\Delta n)^2 + 185.56(\Delta n)^3 \tag{2.9}$$

$$[\text{GdnHCl}] = 57.15\Delta n + 38.68(\Delta n)^2 - 91.60(\Delta n)^3 \tag{2.10}$$

The substocks were then mixed in the correct ratios to give 1 ml of 0.05 mg/ml protein

[21] Provided by Dr David Brockwell

[22] Square bracket, [x] indicates concentration of x.

samples increasing by 0.2 M of denaturant from 0 M up to the maximum denaturant concentration.

The fluorescence experiments were performed at 25 °C. The denaturation was monitored by following the change of fluorescence for a single wavelength (SW) for 60 seconds (s), by excitation of tryptophan residues at 280 nm. The SW method uses the average of the intensity of fluorescence over 60 s. The SW monitored was chosen based on the largest change in fluorescence between the emission spectra (300 nm – 380 nm) of the fully folded (0 M sample) and fully unfolded sample (maximum denaturant concentration).

The change in intensity, I, as a function of denaturant concentration, [denaturant], was fitted using a two state model in Origin Pro 6.1, using the expression in Equation 2.11. Where m is the change in free energy between the unfolded and native states per mole of denaturant, R is the ideal gas constant, T in the temperature of the experiment in Kelvin and $\Delta\ G_{u,chem}$ is the free energy of unfolding at 0 M denaturant.

$$I = \frac{(A[denaturant] + B)e^{(\frac{\Delta G_{u,chem} - m[denaturant]}{RT})}(C[denaturant] + D)}{1 + e^{(\frac{\Delta G_{u,chem} - m[denaturant]}{RT})}} \tag{2.11}$$

This expression includes parameters that determine linear fits to the pre- and post-transitional baselines; A and C are the slopes of the folded and unfolded transition and B and D given the limits of the equations, i.e. the signal of a totally folded and unfolded protein respectively. The data was corrected for the slopes before and after the transition and normalised to the fraction of folded and unfolded protein using in house written software in Igor Pro (Wavemetrics, USA).

Normalisation of the data and elimination of the pre- and post- transitional slopes were carried out in a single step. The parameters for the straight line fits to the pre and post transition data were determined as described from Equation 2.11. It was assumed that the straight line fit to the data pre-transition corresponded to fully folded protein domains. It was also assumed that the straight line fit to the data after the transition corresponded to the fully unfolded protein and the protein underwent a single unfolding step. Using the straight lines fits, the values for the folded and unfolded proteins were calculated at every denaturant concentration. The data at each [denaturant] was then normalised to the values of the straight lines. Once the normal-

isation was completed, Equation 2.11 was again fitted to the data to check that there were no considerable changes in the fit parameters, except for the slopes pre- and post-transition. This process was automated and carried out using software written in Igor-Pro.

2.3.3.3 Studying the kinetics of protein unfolding

Fluorescence spectroscopy can also be used to determine the rates of unfolding and folding of proteins, and therefore to determine their stability. This is achieved by monitoring the exponential increase or decrease of fluoresence as a function of time when the denaturant concentration of the protein is rapidly changed. In this project, the unfolding and folding rates for the p(protein) in H_2O and d(protein) in H_2O were determined by measuring the change of fluoresence with time by rapidly increasing or decreasing the concentration of denaturant in the sample.

2.3.3.4 Manual mixing experiments

For protein unfolding transients that took over 30 seconds, manual mixing was performed. Initially stocks of equilibrium unfolding and kinetic folding buffers were made as described in Section 2.1.1.2. The concentration of the GdnHCl or Urea in the folding buffers were tested by refractometry as described in Section 2.3.3.2. A 10 × stock (0.5 mg/ml) of protein was made by dissolving 5 mg in unfolding buffer. The unfolding and folding buffers were mixed using a dispenser. The dispenser was programmed to give substocks that would have denaturant steps of 0.2 M when mixed with the protein. The denaturant concentrations ranged from the midpoint of protein unfolding, determined from an equilibrium experiment, to the maximum denaturant concentration (defined in Section 2.3.3.2). The substocks and 10 × stock of protein were incubated overnight at 25 °C.

For the manual mixing experiment 100 μl if protein solution was rapidly added to 0.9 ml of the substocks to give a final protein concentration of 0.05 mg/ml. This was then gently mixed before the fluorescence was measured. The time between the addition of protein and recording of signal was recorded and is known as the dead time of the experiment. The fluorescence was measured by excitation at 280 nm and emission at the SW used in equilibrium denaturation for 900 s or until equilibrium was reached. Equilibrium was reached when the intensity of fluorescence reached a constant value. This was repeated for each 0.2 M step of denaturant three times beginning with the highest concentration of substock. All unfolding experiments were performed at 25 °C

using a sample handling unit connected to a temperature controller.

The emission spectra were fitted by a single exponential function (Equation 2.12) where the rate of decay gives the observed unfolding rate constant, $k_{u,chem}^{obs}$, of the protein at the concentration of denaturant. The fitting was performed using automated code written in Igor Pro. The rate constant at 0 M denaturant, $k_{u,chem}^{0}$, was determined by linear extrapolation to 0 M of a linear fit to the natural logarithm of the observed unfolding rate constants, $\ln(k_{obs})$ against the concentration of denaturant.

$$I = I_0 + A \times \exp\left(-\frac{t}{k_{chem}^{obs}}\right) \tag{2.12}$$

2.3.3.5 Stopped flow fluorescence spectroscopy

A stopped flow fluorimeter was used to determine the folding and/or unfolding rates constant for proteins where the folding and/or unfolding rate was too quick to monitor using manual mixing. The stopped flow instrument rapidly mixes samples within a mixing chamber kept at a constant temperature by a water bath. In this study the temperature was maintained at 25 °C for all unfolding and refolding experiments. Fluorescence emission can be monitored for all wavelengths above 320 nm, after excitation at 280 nm with a 10 nm bandwidth. The deadtime of the instrument was determined to be 2 ms using quenching of N-acetyltryptophanamide (NATA) with different concentrations of N-bromosuccinimide (NBS) [311].

Initially equilibrium unfolding and kinetic folding buffers were made up as described in Section 2.1.1.2. A dispenser was used to mix the buffers to give substocks with varying denaturant concentrations. For protein refolding experiments, lyophilised protein was dissolved in a buffer containing the maximum denaturant concentration being measured (for example, for urea this was 6 M). The protein concentration was typically 0.5 mg/ml. This protein solution was diluted 1:10 into solutions with lower denaturant concentrations down to a final concentration 0.6 – 0.7 M denaturant. The fluorescence was monitored for enough time to observe at least 60 % of a full unfolding transient.

For protein unfolding experiments lyophilised protein was resuspended in unfolding buffer to a final concentration of 0.5 mg/ml. This protein was mixed in a 1:10 ratio with buffer containing higher GdnHCl concentrations. Several spectra were taken for each folding or unfolding experiment. These spectra were fit to the single exponential

given in Section 2.3.3.4 unless the protein contained proline. For proteins that contain proline, $\sim$ 20 % of the amplitude of the spectra was attributed to proline isomerisation [214]. These spectra were fit with a double exponential and the contribution for the proline isomerisation was ignored in the analysis. Data was fitted in Origin Pro or using automated code written in Igro Pro. Both fitting programs yielded the same observed rate constants.

$$\ln\left(\mathrm{k_{chem,obs}}\right) = \ln\left(\mathrm{k^0_{u,chem}e^{(m_u[denaturant])}} + \mathrm{k^0_{f,chem}e^{(-m_f[denaturant])}}\right) \tag{2.13}$$

After unfolding and refolding experiments were performed, the observed rate constants for both reactions were plotted against denaturant concentration to obtain chevron plots. Figure 2.3.3.3 shows example chevron plots in addition to yielding rate constants. The shape of the graph can determine information about the energy landscape (see figure 2.3.3.3). The chevron plots for all proteins tested were fitted using a two-state model given in Equation 2.13. This expression determined the rate constant of unfolding in 0 M buffer, $\mathrm{k^0_{u,chem}}$, the rate of folding in 0 M buffer, $\mathrm{k^0_{f,chem}}$ and the denaturant dependence of the unfolded and folded rates, $\mathrm{m_u}$ and $\mathrm{m_f}$ respectively. R and T were the gas constant and temperature in Kelvin respectively. To determine whether the two-state model agrees with the data, the free energy $\Delta G_{u,chem}$ and the total m value could be determined from Equation 2.14, where $\mathrm{K_{eq}}$ is the equilibrium rate constant, and Equation 2.15 respectively. These values can be compared to those obtained from equilibrium experiments.

$$\Delta G_{u,chem} = RT\ln\left(K_{eq}\right) = RT\ln\left(\frac{\mathrm{k^0_{u,chem}}}{\mathrm{k^0_{f,chem}}}\right) \tag{2.14}$$

$$m = m_u + m_f \tag{2.15}$$

2.3.4 The Atomic Force Microscope

Proteins can be mechanically unfolded by a number of experimental techniques [312] including optical tweezers [209], magnetic tweezers [210, 313] and the atomic force microscope (AFM)[189, 202, 211, 314], the technique used in this study. There are different methods by which a protein can be mechanically unfolded using an AFM. In each mode the protein is typically tethered at one end to a surface or tip leaving the other end free to interact with either the tip, another molecule or a surface. The force response of the protein is then monitored by either unfolding the protein at constant velocity (force-extension

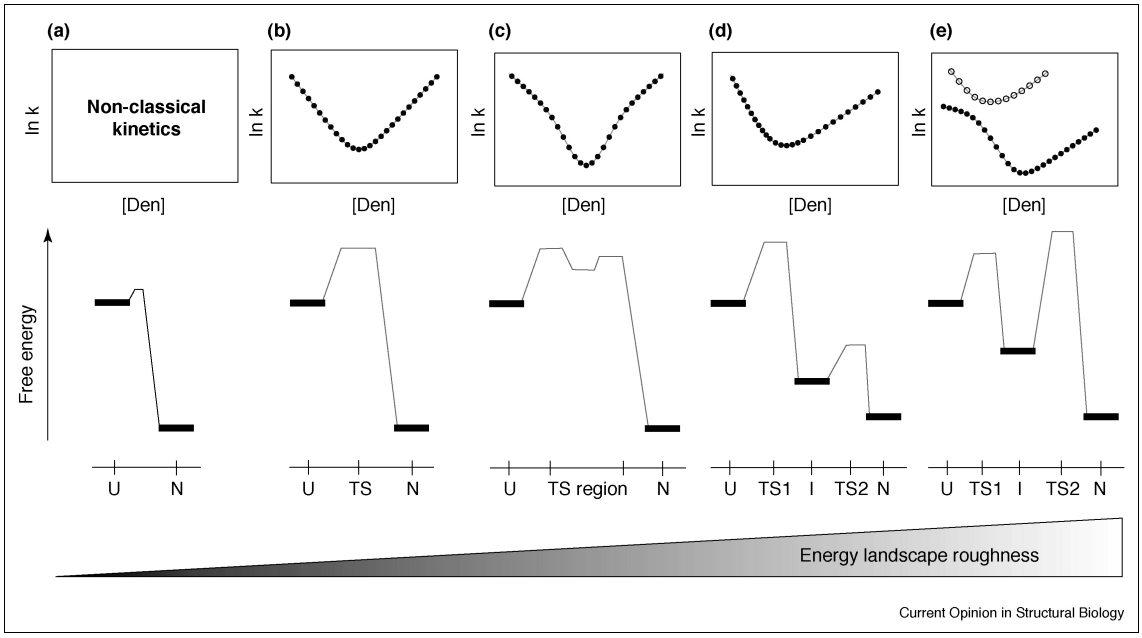


Figure 2.3.3.3: Figure showing example chevron plots (top) for different protein energy landscapes (middle). Chevron plots are graphs of the natural logarithm of the observed rate constant ($\ln(k)$ in this figure) against denaturation concentration ([denaturant]). The unfolding rate constant is determined from measuring the rate of folding (left branch) and unfolding (right branch) of a protein. Differences in the shape of the chevron plots are observed when intermediates are present in the energy landscapes of proteins. In (a) there is no barrier to unfolding and therefore no rates are observed. (b) There is a single barrier to protein folding. This is known as two-state folding. (c) A slight curvature is observed in the folding and unfolding branches of the chevron plots due to a high energy intermediate. (d) A protein with a late stage intermediate. This protein can display two-state unfolding and these late intermediates are inferred from other experiments. (e) A rollover of the folding branch of the chevron plot indicates a rapid folding intermediate (closed circles). This intermediate can be probed by ultra-fast mixing experiments (open circles). This figure was taken from [216]

mode), or under a constant force. In this project, protein constructs were mechanically unfolded using an Asylum MFP-3D atomic force microscope (AFM) in force-extension mode. This section will give a background to AFM experiments and the models used to interpret the data.

2.3.5 Single-molecule force spectroscopy

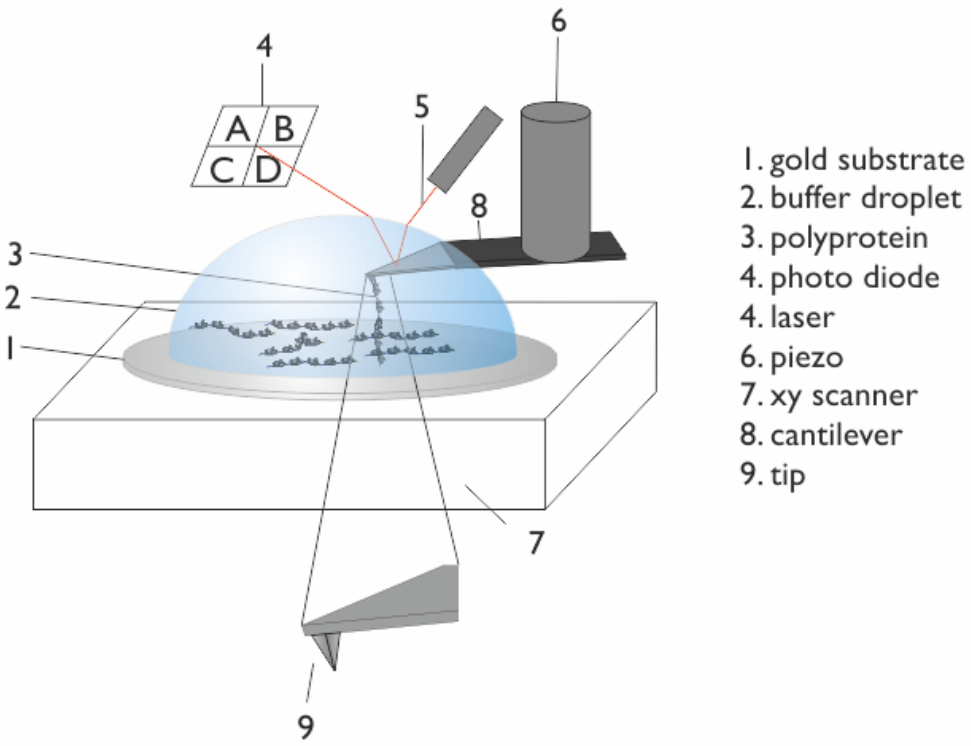


Figure 2.3.5.1: Schematic showing the set up for the Asylum MFP-3D instrument used for SMFS in this project. The protein is bound to a gold substrate, covered in a suitable buffer, which itself is placed on a piezo electric device. A laser beam is deflected off the cantilever tip into a photo-detector which monitors the deflections of the cantilever.

The AFM (Figure 2.3.5.1) is used to extend and refold single protein domains and can measure force responses of proteins from just a few pN up to hundreds of picoNewtons (pN) [315]. It is challenging to study a single protein molecule using single protein molecule force spectroscopy (SMFS). This is because the tip can often interact with the surface causing non-specific peaks in a force-extension trace [316]. For this reason, polyprotein constructs are more commonly used. A polyprotein, shown in Figure 2.3.5.2, is a macromolecule that contains tandem repeats of a protein.

In the AFM experiments presented here, the polyprotein constructs were bound at one end to a gold substrate in solution. Cysteine residues were genetically encoded onto

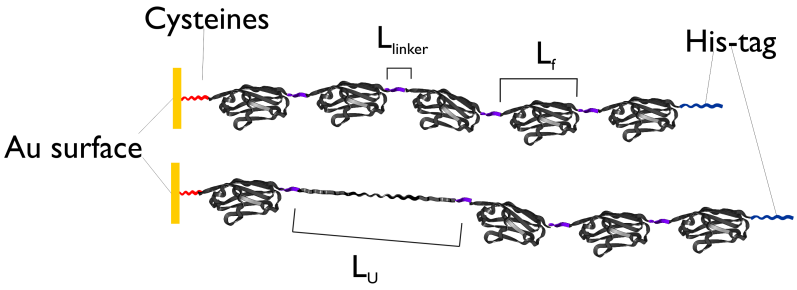


Figure 2.3.5.2: Schematic of a polyprotein chain with all protein domains folded (top) and one protein domain unfolded (bottom). Protein domains are linked together by polypeptide linker regions which have a specific linker length L_{linker}. The folded and unfolded length of the protein is labelled by L_f and L_u respectively. Cysteine residues have been genetically engineered to one end of the chain to allow an Au-S bond to form between the sulphur side chain of the cysteine and the gold surface. The other end of the chain has a hexa-histidine chain which is used for affinity chromatography purification using Ni^{2+}-NTA resin.

one end of the polyprotein chain. This enabled the polyprotein to form a covalent bond between the sulphydryl groups of two cysteine residues at the N-terminus of the protein with the gold residues (an Au-S bond). Initially 0.05 – 0.1 mg protein samples were resuspended in 63 mM sodium phosphate buffer to a final concentration of 0.1 mg/ml and centrifuged at 14 500 × g for 10 minutes. The protein solution was then passed through a 0.45 μm filter. 30 – 50 μl of the protein was incubated on a fresh gold surface for 30 – 60 minutes. Gold substrates were made by gluing glass slides onto a gold templated plate (obtained from Platypus Technologies[23]) using epoxy heat-set glue (Epoxy Technology, UK) following the manufacturers instructions. A new glass slide was freshly cleaved from the plate before each experiment. After the incubation period, excess protein was removed by gently rinsing the surface with 50 - 80 μl of fresh buffer.

A new MLCT cantilever (Bruker,USA), made of a gold coated silicon nitride tip mounted on a flexible cantilever, was used in every experiment. In this AFM setup the cantilever is mounted onto a cantilever holder and the position of the cantilever with respect to the surface is controlled by a piezo-electric device. This device acts as an actuator; when a particular voltage is applied across the piezo crystal it will contract or expand in the direction the voltage is applied. The magnitude of the contraction or expansion is proportional to the voltage applied. This allows the position of the cantilever to be accurately controlled at a nanometer length scale. A laser beam is reflected off the surface of the cantilever into a photodetector. The photodetector is made up of four quadrants and the position of the incident laser beam, and the photodetector, is optimised

[23]This is a custom made product without a titanium adhesion layer under the gold coating

until the reflected laser beam is in the centre of the four quadrants. The deflections of the cantilever can be monitored by movement on the centralised laser spot on the photodiode. The cantilever is an essential component of the AFM, it must be sensitive to the pN forces observed in AFM experiments. In these single molecule experiments the cantilever has a low spring constant, typically between 25 – 50 $pNnm^{-1}$. The spring constant is slightly different for every cantilever and therefore is calibrated before each experiment and recalibrated during long experiments. The method for cantilever calibration in this study was the thermal fluctuation method discussed in Section 2.3.6. The cantilever spring constant, k_c allows the force applied to the cantilever to be calculated using Hooke's law, Equation 2.16, where x is the cantilever deflection.

$$F = k_c x \tag{2.16}$$

After the cantilever was calibrated, it was brought into contact with the surface at a constant velocity, v. Polyproteins can adsorb to the tip when it is brought into contact with the surface; this can occur anywhere along the macromolecule. A contact force of 1500 pN was applied to the surface in these experiments to increase the probability of protein adsorbtion. The tip could then be retracted at a constant velocity. If a polyprotein was attached to the tip, a restoring force on a force-extension trace was observed. This restoring force results from a change in entropy of the system. At a certain force a protein domain will unfold and the restoring force will be removed. This unfolded protein domain will continue to be fully elongated adding to the effective length of the chain. Once fully extended, a force will be applied across the remaining folded protein domains and another one will unfold. This process is repeated until all the protein domains attached to the cantilever tip are unfolded. A detachment peak is then observed which corresponds to the protein detaching from the tip or substrate. This process is illustrated in Figure 2.3.5.3. The force-extension profile therefore results in a saw-tooth pattern, where each peak reports on the unfolding of a protein domain. A single unfolding peak for each protein indicates a two-state unfolding process. The elongation of the protein is observed in the force-extension trace after an unfolding event due to the additional lengthening of the unfolded protein. Figure 2.3.5.4 is a schematic of a typical AFM trace for a polyprotein containing five protein domains. In this figure the unfolding forces of the protein and the detachment peak are labelled. Each unfolding event occurs a regular distance apart if the polyprotein is made up of a protein domains with a single mechanically resistant unit. These sawtooth patterns are characteristic of

polyproteins and provide a mechanical fingerprint in SMFS experiments [221]. These mechanical fingerprints of previously studied proteins have provided a useful tool for mechanical characterisation of previously uncharacterised proteins. This is done by sandwiching uncharacterised proteins between previously characterised proteins [124, 255, 317].

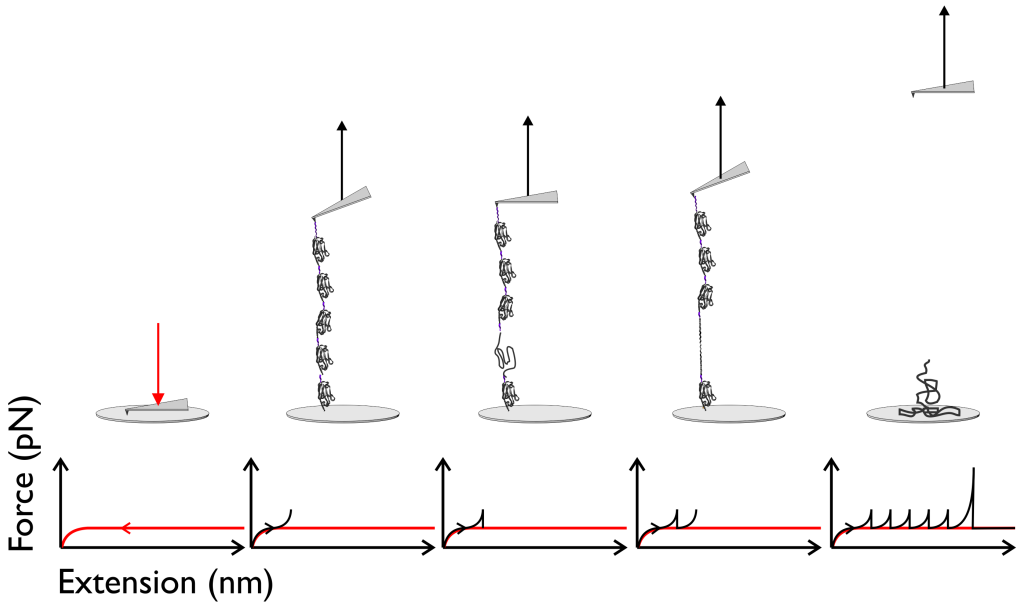


Figure 2.3.5.3: Schematic diagram showing how a force-extension trace is obtained. The red line is the approach of the substrate to the cantilever, the force is negative initially as the cantilever is deflected upwards when in contact with the surface. The black line is the retraction of the cantilever from the substrate at constant velocity. As the distance increases a restoring force results as the protein resists the change in entropy of the chain. At a particular force the bonds within the protein will break. Following this the chain will elongate and a restoring force on the cantilever will be observed again. Once all the domains have unfolded the protein detaches from the tip or substrate.

Figure 2.3.5.3 shows how a force-extension trace is obtained. The red line is the force-measured during the approach of the cantilever to the substrate; typically the contact force on approach is 1500 pN. The black line is the force measured upon the retraction of the cantilever from the substrate at constant velocity. Initially there can be some interactions, either with the surface, or protein-protein interactions. The first peak on the graph corresponds to one of the domains in the construct unfolding, the peak is the point of rupture of the bonds holding the protein together, and the curve rising up to the force of the subsequent protein rupture is the elongation of the protein as the restoring force increases again. In the resulting graph, each peak is an unfolding peak and the last peak is the detachment of the elongated chain from the cantilever and is known as the detachment peak. The protein chain can also dissociate from the cantilever before detachment occurs, this typically results in a smaller detachment peak.

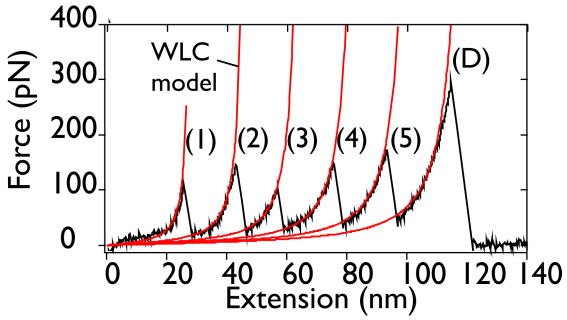


Figure 2.3.5.4: Schematic of a force-extension trace for a polyprotein containing five protein domains. The detachment is labelled. Each peak preceding the detachment corresponds to a protein unfolding event. The unfolding force of the protein is taken as the maximum force observed before the restoring force is removed.

In this project experiments were repeated at least three times at retraction velocities of 160 nm/s, 400 nm/s, 1000 nm/s and 2000 nm/s. Where possible, multiple velocities were performed in a single experiment (i.e. one calibrated cantilever) to ensure that there were no errors due to cantilever calibration.

2.3.6 Calibration of cantilevers

Cantilever spring constants must be soft enough to respond to the pN forces observed for protein unfolding events. The force sensitivity of a cantilever is limited by its thermal noise [318] therefore cantilevers with low spring constants (softer) are more susceptible to noise[319]. Recent advances have been made improve the precision of force measurements. The short-term force precision of a cantilever is sensitive the hydrodynamic drag on the cantilever. By reducing the size of the cantilever, this drag is reduced [320]. The Bio-lever by Olympus Optical Co. reduces the length and thickness of the cantilever which does not significantly increase the spring constant but significantly reduced the noise of the cantilever. Reducing the reflective coating on the cantilever can also increase the long-term force precision of a cantilever [321–323]. Removing some of the gold coating on a cantilever reduced the instrumental drift of soft ($\sim$ 6 pN/nm) cantilevers[321]. To ensure the forces obtained in these SMFS were as accurate as possible, the cantilever was calibrated at the start of each experiment. The cantilevers were also recalibrated during the course of an experiment. Many models have been developed to estimate the spring constant of the cantilever [324]. In this project the cantilever was calibrated using the thermal fluctuation model [325]. The thermal fluctuation model was first derived by Hutter *et al.* (1993). This model assumes that, in equilibrium, the cantilever tip behaves

as a simple harmonic oscillator, with one degree of freedom fluctuating in response to thermal noise. This assumption is a good approximation when the deflections of a cantilever are small. The AFM tip is therefore modelled as an ideal spring with spring constant, k_c, and mass, m [325, 326].

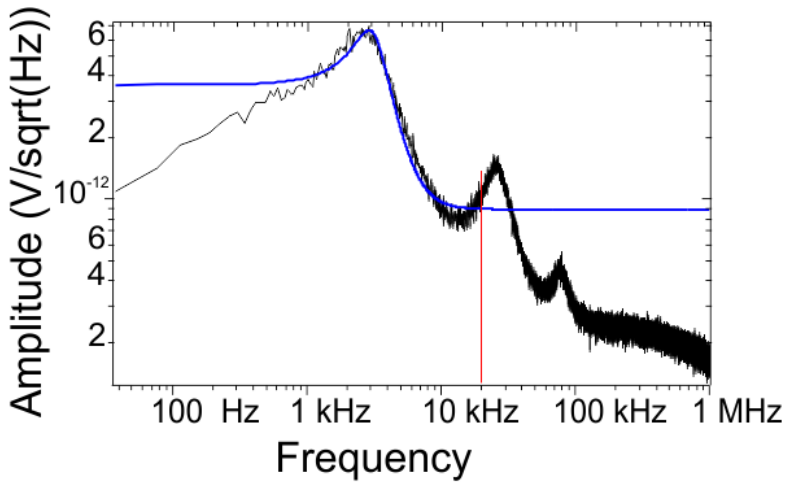


Figure 2.3.6.1: Figure showing an example of a power spectrum of a cantilever in liquid. Where the y axis is the amplitude of the cantilever measured in V/sqrt(Hz), shown on a logarithmic scale, and the bottom axis is the frequency in Hz. The power spectrum is fitted by a Lorenzian curve (blue). The peak position corresponds to the resonant frequency of the cantilever.

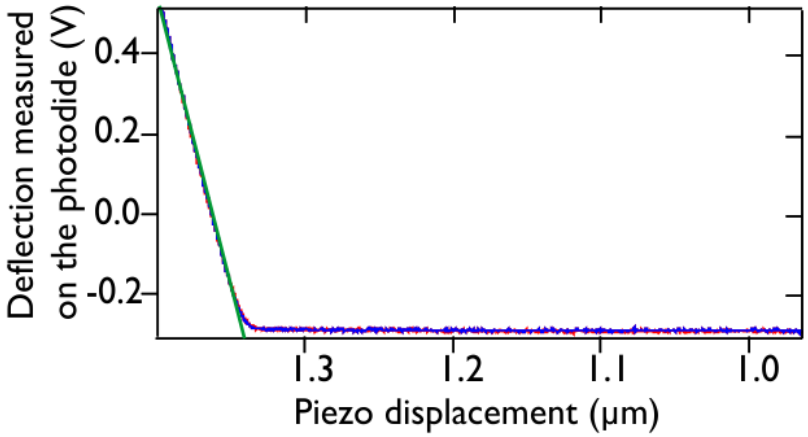


Figure 2.3.6.2: Figure showing an example of a contact slope of a cantilever on a surface. The red line is the approach of the cantilever to the surface and the blue line is the retraction. The graph shows the change in the deflection of the laser beam incident on the cantilever, measured by the photodiode, as a function of sample displacement. The sample displacement is measured by the displacement of the piezoelectric device. The slope (straight line fit shown by the green line) of the contact region of the tip with the sample is used to calibrate the cantilever tip.

When the tip is far from the sample, its motion will be solely due to thermal fluctuations [325]. The frequency of the motion near the resonant frequency will allow

an approximation of the spring constant. The spring will undergo a deflection due to thermal motion in the vertical direction z, will have momentum, p and therefore the system will have an angular resonant frequency, ω_o. This is related to the spring constant by Equation 2.17 [325, 326].

$$\frac{1}{2}m\omega_0^2\langle z^2\rangle = \frac{1}{2}k_c\langle z^2\rangle \tag{2.17}$$

$$k_c\langle z^2\rangle = k_BT \tag{2.18}$$

Initially the cantilever is positioned within the buffer droplet and brought towards the surface. A contact slope, as shown in Figure 2.3.6.1, is taken which measures C, the change in voltage, V, as a function of distance, nm. The cantilever is then withdrawn from the surface and a power spectrum is taken. A schematic of a power spectrum is shown in Figure 2.3.6.2. The power spectrum monitors the oscillation of the tip due to thermal fluctuations. This power spectrum is fast Fourier transformed in Igor Pro to change the data from the time domain to the frequency domain. A Lorentzian[24] is then fitted to the resonant peak of the cantilever. The area under the curve determines, z_v, the mean square displacement of the cantilever due to thermal motion in volts. Equation 2.19 can then be used to determine the displacement in nm [326]. This displacement is then used in 2.17 [326] to determine the cantilever spring constant.

$$z = C^{-1} \cdot z_v \tag{2.19}$$

2.3.7 Analysis of a force-extension trace

Force extension traces from single molecule force experiments were filtered to only include traces where: 2 – 5 unfolding events were observed, minimal surface interactions were present close to the surface and there was a clear detachment peak [214]. Initially the inter-peak distance, x_p, which is labelled in Figure 2.3.7.1 (a), was measured for each protein unfolding domain. If it deviated from the literature value (± 2 nm) significantly the trace was disregarded. Traces were analysed in Igor PRO (Version 6.0, Wavemetrics, USA), using software written in-house. The worm-like chain model (WLC), a model for polymer elasticity (see Section 2.3.7.1), was fitted to each of the curves. The bottom

[24]With functional form $P(f) = \frac{A}{(f-f_0)^2+B} + bg$, where f is the frequency, f_0 is the resonant frequency, bg corresponds to foreign noise with a uniform and constant spectrum, A and B are parameters that be used to determine the quality factor, Q ($Q = \frac{B}{2A^{1/2}}$) height of the peak and L_{peak} ($L_{peak} = \frac{A}{B}$) [325, 327]

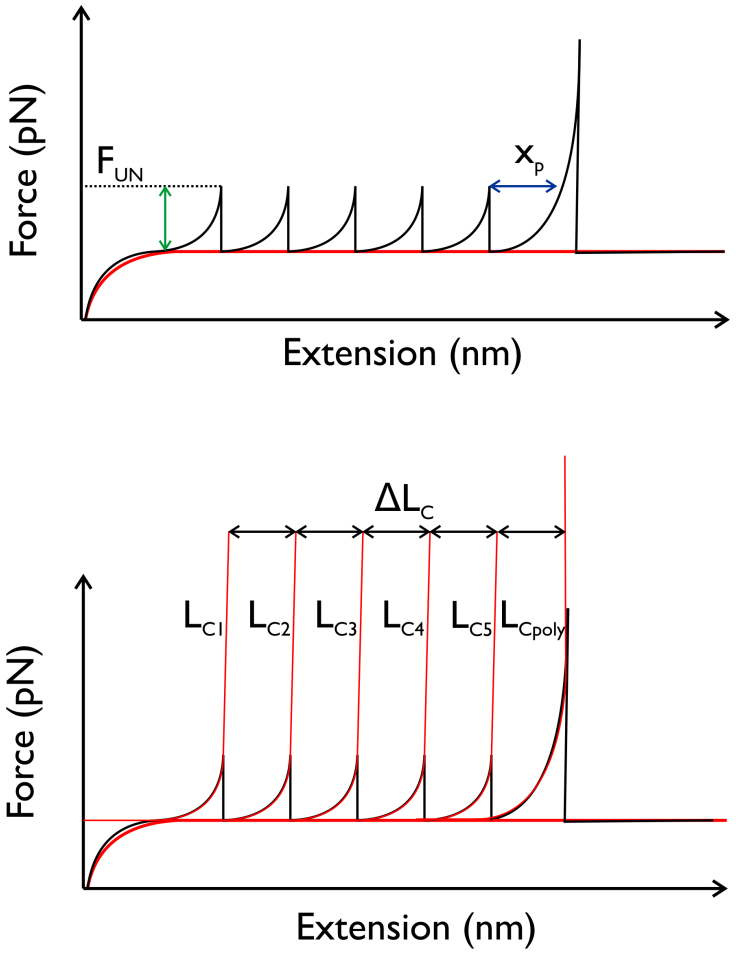


Figure 2.3.7.1: Schematic showing the parameters and analysis of a force-extension trace from a SMFS experiment on a poly-protein containing 5 tandem repeats of a protein. The parameters highlighted are: (a) F_{UN}: the unfolding force of a single protein and x_p: the interpeak distance between peaks. The interpeak distance should correspond to the elongation of a protein domain upon unfolding. (b) Worm-like chain (WLC) fits to each of the the unfolding peaks. Each worm-like chain fit will yield a contour length, the full length of the chain, at each unfolding event. These unfolding events should occur at regular distances apart. This change in contour length, ΔL_c, should correspond to the fully unfolded length of the protein. This can be used to determine between unspecific and events. It can also be used to determine whether a cantilever is properly calibrated in the experiment.

schematic in Figure 2.3.7.1 shows how the WLC is fitted to a force-extension trace. The red lines indicate the WLC fits to the rising edges of protein unfolding events in the saw-tooth pattern.

The peaks can be fitted to each curve simultaneously by setting a change in contour length ΔL_c which corresponds to the additional length added when a protein has unfolded. The ΔL_c should therefore be identical for each unfolding peak in a single force-extension trace when using a polyprotein containing repeats of the same protein. Furthermore the ΔL_c should be consistent across the different experimental force-extension traces for the same polyprotein. This is a characteristic of the mechanical fingerprint provided by polyproteins. The persistence length, L_p, can also be varied in this model. This value affects the gradient, and therefore L_c, of the WLC fit. The persistence length is a measure of the flexibility of a protein and is defined as the length along which correlation in the tangent vector is lost. For proteins it is normally found to be ~ 0.4 nm which approximates to the length of a single amino acid [252].

Once the WLC had been fitted to the unfolding curves, model free analysis of the data was performed. The unfolding force, F_{UN}, of a protein is the peak force reached before the protein unfolds. This gives the mechanical stability of a protein. This is the amount of force the mechanical clamp (see Section 1.6.3) within the protein can tolerate before it unfolds, if it contains a mechanical clamp motif. It should be noted that for these polyprotein constructs, F_{UN} varies depending on how many domains remain unfolded when a protein domain unfolds (an effect known as the unfolding history)[328]. By performing these measurements for the same construct at different speeds, a plot of pulling speed against F_{UN} can be used to extract basic estimates of the 2-dimensional energy landscape of the protein (see Section 2.3.9). F_{UN} increases as the pulling velocity is increased, this can be attributed to a decrease in the time that thermally activated protein unfolding can occur [267].

In this project the pulling velocities tested are 160 nm/s, 400 nm/s, 1000 nm/s and 2000 nm/s for each of the systems. At least three data sets are collected for each speed for each of the four systems. After a WLC fit to the data, fixing $L_p = 0.4$ the F_{UN} was determined for each unfolding event in every accepted force-extension trace. For each experiment the F_{UN} were pooled into a histograms with a bin width of 10 pN. The modal F_{UN} was obtained from a Gaussian fit to the distribution. The average force, F, was then taken as the average of the modal forces for each pulling velocity. These forces were

plotted against the natural logarithm of pulling velocity to extract information about energy landscape of the protein systems as discussed in Section 2.3.9.1.

2.3.7.1 The worm-like chain model for polymer elasticity

The worm-like chain model (WLC) is a model of polymer elasticity used to fit the experimental force-extension data. If the polymer is modelled as a continuous chain of N segments, each with length d, that are at an angle θ to the adjacent segment, the projection of each segment, from the first segment to the last, affects the projection of the adjacent segment along the chain by $l \times \cos(\theta)$. The projection onto the first bond is given by $\langle \cos(\theta_{1N}) \rangle = \cos^{N-1}(\theta)$. The average projection is therefore given by[329]:

$$\langle x \rangle = \frac{1}{d} \sum_{i=1}^{N} \langle \mathbf{d_i} \cdot \mathbf{d_j} \rangle = d \sum_{i=0}^{N-1} \cos^{i}(\theta) = \frac{d(1 - \cos^{N}(\theta)}{1 - cos(\theta)} \tag{2.20}$$

As N tends to infinity this expression tends to [329]:

$$\lim_{N \to \infty} \langle x \rangle = \frac{d}{1 - cos(\theta)} = L_p \tag{2.21}$$

This is the definition of the persistence length which is a measure of polymer stiffness, it determines the distance over which the correlations between the segment are lost. The more flexible and bendy the chain, the quicker these correlations are lost, and therefore L_p is small. A more rigid rod like chain will have a higher L_p as the correlations are lost more slowly. For example the persistence length of a β-sheet protein, determined by fitting the protein unfolding force-extension curve by a WLC, was found to be about 0.4 nm [252] which roughly corresponds to the size of an amino acid. The persistence length of individual actin filaments has been found to be 17.7 μm indicating the structure is much stiffer than a single protein [330]. The reason for a higher persistence length than protein is attributed to their function in the cytoskeleton. These materials are known as semi-flexible materials. Microtubules, hollow cylinders made up of tubulin proteins, are also found in the cytoskeleton. These are found to be very rigid with persistence lengths of 5200 μm [330]. These microtubules therefore form stiff scaffolds and supporting compressive loads within the cell [331].

For small angles $\cos(\theta) \approx 1 - \frac{\theta^2}{2} \equiv \exp(\frac{\theta^2}{2})$ the expression given in Equation 2.26 tends to $L_p = \frac{2d}{\theta^2}$. In the WLC the segment length becomes infinitely small as the number of segments in the chain increases, whilst the contour length, $L_c = Nd$, and the L_p remain

constant. The projection in this limit then becomes [329, 332]:

$$\langle x \rangle = L_p(1 - e^{-\frac{L_c}{L_p}}) \tag{2.22}$$

The exponential term means that the angular correlation decays exponentially along the chain. As $L \to \infty$ this would tend to the persistence length of the chain. The mean square radius of gyration is also given by:

$$\left\langle r^2 \right\rangle = 2L_p \left[1 - \frac{L_p}{L_c}(1 - e^{-\frac{L_c}{L_p}})\right] \tag{2.23}$$

There is an energy cost associated with the bending of a WLC given by a bending elasticity term [333]:

$$E = \frac{\kappa}{2} \int_0^{L_c} \left(\frac{\partial^2 \hat{\mathbf{r}}(s)}{\partial s^2}\right)^2 ds \tag{2.24}$$

This term accounts for the shape of the polymer. Where E is the bending energy, s is the distance along the curve, κ is the bending coefficient and $\frac{\partial^2 \mathbf{r(s)}}{\partial s^2}$ is the tangent vector along the chain [333–335]. To straighten out the chain work needs to be done against the bending energy. The hamiltonian, H, for such a system is given by Equation 2.25 [334, 336]. The $\mathbf{f} \cdot \mathbf{R}$ accounts for the application of a constant force to the ends of the chain.

$$H = \int_0^L ds \frac{\kappa}{2} |\frac{\partial^2 \hat{\mathbf{r}}(s)}{\partial s^2}|^2 - \mathbf{f} \cdot \mathbf{R} \tag{2.25}$$

This model therefore takes into account entropic and some enthalpic contributions to the unfolding of a polymer chain [337]. To determine the resultant force-extension relationship this would have to be computed over all possible shapes of a polymer. An interpolation formula for this model has successfully been used to describe the unfolding of DNA by Marko and Siggia and Bustamente *et al.* and is given in Equation 2.26 [336, 338]. The force (F) is measured as a function of extension (x), L_p is the persistence length, k_B is the Boltzmann constant and L_c is the contour length i.e., the fully extended length of a protein, or full poly-protein construct and T is the temperature of the experiment (assumed to be 298 K). The WLC can also be used to calculate ΔL_c, which is also labelled on Figure 2.3.7.1, this is typically the change in length between the folded and unfolded protein. The WLC model has been used extensively to fit force-extension curves from forced protein unfolding [193, 252, 317, 339, 340]. Despite being able to describe most biomolecular stretching such as the stretching of DNA [338],

this model fails for large forces where over-stretching of the bonds within the polymer occur[334]. Furthermore this model assumes that the persistence length is the same for the folded and unfolded proteins and does not change as a function of distance. This is a crude estimation, the stiffness of a folded domain would differ to that of a fully unfolded domain [341]. Other models have tried to add additional terms to the WLC model to minimise the discrepancies between the experimental data and model. Some of these models are discussed in the next section along with alternatives to the WLC.

$$F(x) = \frac{k_B T}{p}\left[\frac{1}{4}\left(1 - \frac{x}{L_c}\right)^{-2} - \frac{1}{4} + \frac{x}{L_c}\right] \tag{2.26}$$

2.3.7.2 Alternative to the WLC model

As discussed in the previous section, the WLC fails to describe the over-stretching of molecules at high force [334]. To try and account for this there have been a few extensions to the worm-like chain. The work by Wang *et al.* (1997) incorporates a stretching term in the WLC model[342, 343]. This model is known as the WLC model with segment elasticity (or extendable WLC model). In this model the ratio of extension to contour length, x/L_c, is replaced by $x/L_c + F/\kappa_s$, where κ_s is the extension modulus of the protein. Bouchiat *et al.* (1999) have also incorporated additional terms to the WLC model [334]. This study determined an exact solution to the WLC model and added correction terms to the WLC model to eliminate discrepancies. This method successfully reduced errors between the experimental data for DNA stretching and the model.

There are also alternative models that describe the force-extension behaviour of polymer unfolding. A simple model for polymer elasticity is the freely jointed chain model (FJC) [344]. This model describes the chain as a series of rigid segments. An extension to the FJC was also made to include the elasticity of the segments [345]. Another extension of the FJC is heterogeneous FJC model which incorporates the difference in stiffness between a folded and unfolded protein chain [341]. The FJC model is extended to be the sum of the contribution of the number of folded and unfolded domains in a polyprotein. Models have also attempted to incorporate information about the 3D structure of a protein. One example is the thick chain (TC) model [346]. This model is used to infer information about the width of the chain in force-extension relationships.

This section briefly reviews some of the models available for protein unfolding. The WLC model was chosen based on its simplicity and popularity in the SMFS field.

2.3.8 Modelling the energy landscapes of proteins in forced unfolding

Force acts as a denaturant in these experiments, as the force is applied across the protein domains the activation barrier for unfolding is lowered - this increases the probability that the protein will unfold. There are many models that describe how forces affects the rate of unfolding of a protein [347–349]. Although the Bell model has been used extensively to determine the free energy landscape of proteins using single force spectroscopy, other models have emerged for determine the free energy landscape from such experiments. These models incorporate more detail than the simple assumptions made in the Bell model, and can account for the curvature that is sometimes observed in the plot of force against the natural logarithm of the velocity [313, 350]. Figure 2.3.8.1 shows schematics of some of the available models. These models will be described in this section beginning with the most frequently used model, the Bell model. This section will also describe the Monte Carlo simulations used to determine the energy landscape parameters.

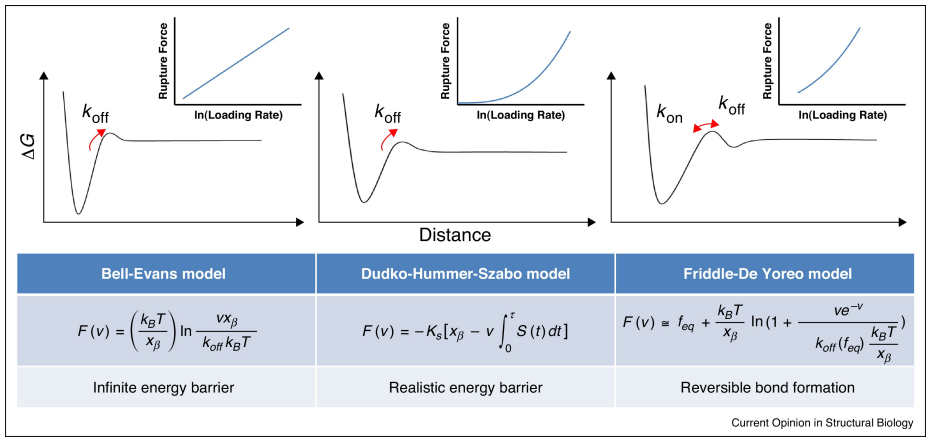


Figure 2.3.8.1: Figure showing a schematic of three available models to describe the effect of force on the unfolding rate of a protein. F(v) is the most probable unfolding force at a loading rate here denoted as v. In these expressions T is the temperature, x_β denotes the distance between the folded and transition states (in the text this is denoted as x_u. The k_{off} corresponds to the rate of unfolding (denoted as $k^0_{u,force}$ in the text). Right: The Bell model describtes the force-ln(loading rate) relationship as a straight line. Middle: The Dudko-Hummer-Szabo Model accounts for some curvature in the F-ln(v) plots. S(t) is the probability of unfolding as a function of time and κ_s is the harmonic force constant. Right: The Friddle-De Yoreo model assumes that the protein is in equilibrium between folding and unfolding before a certain force. Figure taken from [313].

2.3.9 The Bell Model

The Bell Model (BM) [347] is an empirical model, which can be applied to data obtained from SMFS that describes the rate of unfolding of a biological molecule under a mechanical force k_u^F. It is a one-dimensional model that assumes the rate of unfolding is described by a van't Hoff Arrhenius-like equation, where the rate of the reaction is force dependent.

George Bell originally developed this model to describe the rate of adhesion of cells to other cells or surfaces in the presence of hydrodynamic stress [347]. In fact this kinetic model for mechanical strength was first derived by Zhurkov who applied the same principles to solids under mechanical strength [351]. Subsequently, an analytical version of the equation was derived from Kramers reaction rate equation by Evans and Richie and is given by Equation 2.27 [352].

$$k_u^F = A \exp \frac{-(\Delta G^0_{TS,force} - F x_u)}{(k_B T)} = k^0_{u,force} \exp \frac{F x_u}{(k_B T)} \tag{2.27}$$

Where $k^0_{u,force}$ is the rate of unfolding in the absence of force, A is the attempt frequency, F is the force applied along the reaction co-ordinate, kB is the Boltzmann constant, T is the temperature, $\Delta G^0_{TS,force}$ is the height of the unperturbed free energy barrier opposing the mechanical fracture (or activation energy to the transition state) and x_u is the difference in the extension of the native and transition states (or the distance to the unfolding transition state).

If unfolding occurs via a two-state process, there is an energy barrier between the folded and unfolded states of the protein. This 1D energy barrier is defined by the reaction coordinate studied. In SMFS this is the distance between the -N and -C terminus of the protein. By applying a force across the protein, the energy barrier to the unfolded state decreases. This increases the probability that the protein will unfold at a certain force. Using the BM, the rate of a reaction can be related to the force applied. It can be assumed that a protein has two well-defined states: the native state and the transition state. If mechanical work is performed linearly over the reaction coordinate, the effect of force on the reaction can be governed by the van't Hoff Arrhenius-like equation given by the BM (Equation 2.31). If it is assumed that the transition state and native state on the 1D free energy profile are independent of force, then that activation barrier at any force is given by $\Delta G* = \Delta G^0_{TS,force} - F \times x_u$. Force therefore lowers the energy barrier by tilting the energy landscape, (see Figure 2.3.9.1) assuming there is no change in the shape of the energy landscape [350].

Evans and Richie derived the equation for the most probable force of unfolding, F_{UN}, which is given by Equation 2.28 [352]. This describes how the most probable force, F, is affected by the loading rate $R = F'$. For constant pulling velocity experiments, the loading rate is given by $R = k_{eff} v$, where k_{eff} is the effective spring constant of the system and v is the pulling velocity. Previously it was thought that the loading rate could

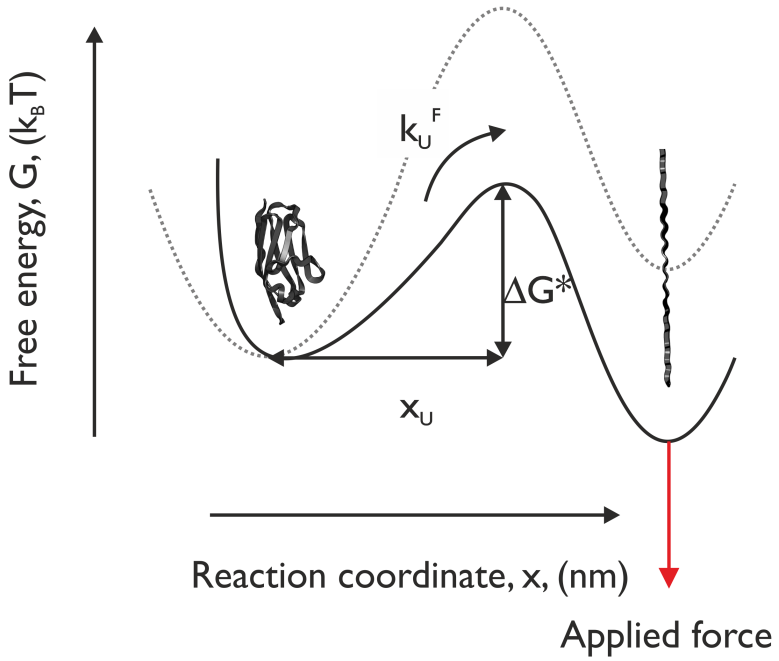


Figure 2.3.9.1: Schematic showing the effect on force on the energy landscape of a protein, where $\Delta G*$ is the change in $\Delta G^0_{TS,force}$ under applied force. The protein is folded in the energy minimum (right hand side) and will be unfolded only when it gets over a barrier to another energy minimum.

be directly calculated from the spring constant of the cantilever ($R = k_c v$). In this case the energy landscape parameters could be obtained directly from a linear fit to the $F - ln(v)$ data. However, it has been shown that, as a protein unfolds in a poly-protein chain, the compliance changes [328]. Linkers also change the compliance of the protein [348]. Therefore polyproteins preclude the direct calculation of unfolding rates. To overcome this problem, Monte Carlo simulations (MCS) are often used to determine the parameters governing the energy landscape [214, 217, 340]. These are described in Section 2.3.9.1.

$$F_{UN} = \frac{k_B T}{x_u} \ln \frac{R x_u}{k_B T \times k^0_{u,force}} \tag{2.28}$$

2.3.9.1 Monte Carlo simulations on poly-proteins to extract energy landscape parameters

The average modal force (F) can be plotted as a function of the natural logarithm of the speed and, according to the BM, the data should be described by a linear fit. The average persistence length, average spring constant, the linker length, the length of a single folded and unfolded I27, contour length and the temperature of the experiment (assumed to be performed at 23 °C) were used in the MCS. The MCS initially calculates the probability that a protein domain will unfold based on the constraints set, including a randomly

selected chain length. In these simulations the MCS can only select chain lengths of 2 – 5 domains in order for the simulated data to correspond to the data collected experimentally.

The software used to run the MCS was written in C programming language following the code previously written in Igor Pro [124]. The code is based on MC simulations that allowed for unfolding and refolding of proteins [340]. The compiler used for writing and running the program was Xcode (Apple, USA). The simulations run using a two state model, where the protein is either folded or unfolded (all or nothing), as described by the BM. It was assumed that the pulling rates used in this experiment were large enough for protein refolding not to occur [217]. The simulation was written to loop over a range of x_u and $k^0_{u,force}$ with given step sizes. Table 2.3.9.2 gives the necessary information required for the simulation and the typical inputs for each protein system. An algorithm has previously been designed [217, 332] to determine the number of proteins unfolded in the simulations. This algorithm is presented and described below and in Figure 2.3.9.2. The parameters typically used are presented in Table 2.3.9.2.

1. The number of proteins to unfold, N^0_f, is initially determined from:

$$N^0_f = (4RN + 2) \tag{2.29}$$

   Where RN is a pseudo-random number generator to obtain a random number (RN) between [0,1)[25]. Random numbers are generated using a Marsenne-Twister algorithm provided by the website of Makoto Matsomoto, Hiroshima University, the numbers generated are from a normal distribution and a period of $2^{19937} - 1$. This number also gives the number of linkers between the protein domains.

2. The length of the protein is determined for the protein in a folded and unfolded conformation where L_f is the length of a folded protein i.e, the distance between the -N and -C termini of the folded protein, L_u is the unfolded length of the protein and L_{linker} is the length of each linker region between each protein domain. N_u and N_f is the number of unfolded and folded protein domains at the time-step respectively.

$$L = N_f L_f + N^0_f L_{linker} + N_u L_u \tag{2.30}$$

3. For the first time step, dt, where the extension x is given by $x = Vdt$ where V is the velocity of the simulated experiment, the extension of the protein, x_p is determined. This is done by obtaining the real cubic root of the polynomial from solving

[25]The random number is <1 but can be equal to 0

$F_{WLC}(x_p) = k_c(x - x_p)$ for x_p. The cantilever extension can then be determined from $x = x_c + x_p$.

4. The force, F, at this time point is calculated using Hooke's law for the cantilever $F = k_c x_c$, where k_c is the experimentally determined cantilever spring constant.

5. The unfolding rate under applied force, k_u^F is calculated according to the Bell Model.

$$k_u^f = k_{u,force}^0 \exp \frac{F x_u}{(k_B T)} \tag{2.31}$$

6. The probability a single protein domain will unfold at a defined time step of dt is then calculated by:

$$p(protein) = N_f k_u^f \exp{(F \times x_u)/(k_B T) dt} \tag{2.32}$$

7. The probability is compared to a RN from [0,1)

   - If the probability is greater than the RN then the protein unfolds and N_f decreases by one and N_u increases by one, the force at this point is stored. If the $N_f > 0$ the process continues until all protein domains are unfolded. If unfolding of the protein domain occurs there is an increase of δL_c to the total length of the chain. The calculation can then be repeated for the new contour length of the chain until all domains are unfolded or the probability of unfolding is zero.

   - If the probability is less than the RN then the process continues until the probability is greater than a RN.

This algorithm is repeated until all proteins have unfolded. The whole process is then repeated 1000 times to achieve good statistics [332]. Repeats of this typically at a single velocity give modal forces that vary by less than 1 pN.

The simulations are used to generate a list of forces for each experimental pulling velocity, which can be compared to the experimental distribution. The data is imported into Igor Pro and analysed - this has been automated to loop the analysis over the range of x_u and $k_{u,force}^0$. The analysis works by converting the list of forces for each v for a given x_u and $k_{u,force}^0$ to a histogram. The histograms can be described by a Gaussian and the mean force and width of distribution is obtained. The $F - \ln v$ from the simulated data can be described by a linear fit. The method described by Zinober (2005) was employed to

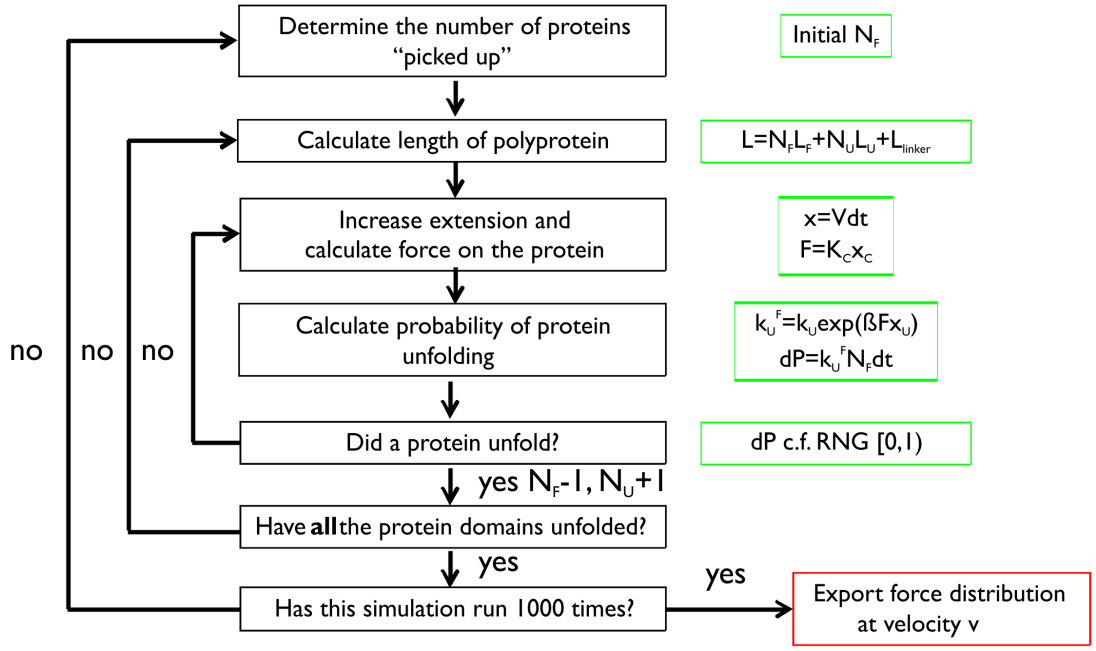


Figure 2.3.9.2: Diagram representing the algorithm used for the MCS written in C programming environment to mimic SMFS experiments for a given x_u and $k^0_{u,force}$. Initially the number folded proteins, N_f is determined using a random number, RN, generator. The number of protein domains can be anywhere between 2–5 to mimic the experimental traces. The full length, L, of the polyprotein can then be calculated based on the user inputs for the length of the folded protein L_f, length of the unfolded protein, L_u and the linker length L_{linker}. The extension across the protein for velocity, v, can be determined for each timestep dt. This allows the force across the protein domains to be calculated. The rate of unfolding under applied force, k_u^F, is then calculated using the BM (in this diagram, k_u is the rate of unfolding in the absence of force and β is $\frac{1}{k_B T}$). This allows the probability, dP, of unfolding to be computed. This probability is compared to a RN from [0,1). If dP is greater than RN the protein unfolds, if dP is less than RN the timestep is increased. If a protein unfolded, the N_f is decreased by 1 and the number of unfolded proteins, N_u increases by 1. If N_f is now zero, there are no more protein domains to unfold, the simulations check if it has run enough times (1000) before exporting the forces. If N_f is greater than zero, the algorithm continues until all proteins have unfolded in the chain.

determine the range of x_U and $k^0_{u,force}$ that "fit" experimental data [332]. In this method, the linear fits to the simulated data, $F(\log(v); x_u, k^0_{u,force})$ [26], were determined for a grid of x_u and $k_{u,force}0$ to determine the best fit parameters to the data. The best fit parameters, $x_{u,best}$ and $k^0_{u,best}$, were the parameters whose $F(\log(v); x_u, k^0_{u,force})$ yielded the best linear least squares fit to the experimental data. TThis $F(\log(v); x_u, k^0_{u,force})$ corresponded to the $F_{bestift}(\ln(v))$. The 95 % confidence interval of this fit to the experimental data was also computed to determine the limits of the straight line fit to the experimental data. The normalised deviation, $E(x_u, k^0_{u,force})$, of the fits to the forces from grid of x_u and $k^0_{u,force}$ was computed using Equation 2.33. In this equation, the linear fit describing the forces

[26] $F(\log(v); x_u, k^0_{u,force}) = m \times \log(v) + c$ where m is the slope (pN) and c is the intercept (pN)

for a pair of x_u and $k^0_{u,force}$ ($F(\log(v); x_u, k^0_{u,force})$), was compared to the $F_{bestift}(\ln(v))$. This was integrated over the range of experimental velocities (160 – 2000 nm/s), where v_1 is the lowest velocity and v_2 is the maximum velocity. The range of acceptable E were determined by computing E_{95}; the normalised deviation between $x_{u,best}$ and $k^0_{u,best}$ and the 95 % confidence intervals. This gave an upper limit to the accepted $E(x_u, k^0_{u,force})$; if a $F(\log(v); x_u, k^0_u)$ gave a value less than or equal to this they were accepted. This method provides a simple way to compute the error in the grid of parameters. Furthermore the method allows the error to be driven by the accuracy of forces computed from simulations.

$$E(x_u, k^0_{u,force}) = \frac{1}{v_2 - v_1} \int_{v_1}^{v_2} \left[\frac{\left(F(\log(v); x_u, k^0_{u,force}) - F_{bestift}(\ln(v)) \right)}{F_{bestift}(\ln(v))} \right]^2 dv \quad (2.33)$$

		Typical value/range	
Parameter	Definition	pL	I27
L_u (nm)	Length of unfolded protein	22.25	28
L_f (nm)	Distance between the -N and -C termini when protein is folded	3.7	4.2
L_{linker} (nm)	Average linker length	2.8	2.8
p (nm)	Persistence length	0.4	0.4
K_c (pN/nm)	Spring constance of cantilever in experiment	40	40
dt (s)	Time step	0.0001	0.0001
v (nm/s)	Pulling speed of experiment	160 - 2000	160 - 2000
T (°C)	Temperature	23	23
n_V	Number of velocities	4	4
$N_{iterations}$	Number of iterations in the Monte Carlo	1000	1000
x_u (nm)	Distance to the unfolding barrier from the folded state	0.1 - 0.4	0.1 - 0.4
$k^0_{u,force}$ (s^{-1})	Rate of unfolding at zero force	0.0005 - 0.3000	0.0005 - 0.3000

Table 2.3.9.2: Table showing a list of the parameters and typical values used in the Monte Carlo code to simulate forced-protein unfolding events.

2.3.10 Alternative models to the Bell Model

2.3.10.1 Friddle-de-Yoreo model

The Friddle-de-Yoreo model, [349] takes into the account the probability that a protein can refold once unfolded, especially in the limit of lower rates. At the lower velocities the protein is assumed to be in an equilibrium phase where reforming of bonds can occur.

At faster pulling speeds the protein is no longer in an equilibrium between folding and unfolding. At this point the process switches to a kinetic phase and the unfolding is irreversible. Fiddle et al. (2012) determined the force, f_{eq}, at which the phase changes from equilibrium to kinetic as:

$$f_{eq} = \sqrt{2k_c \Delta G^0_{u,force}} \tag{2.34}$$

f_{eq} is dependent of the cantilever spring constant, k_c and the free energy difference between the unfolding event at zero force relative to the cantilever ($\Delta G^0_{TS,force}$). In this model the rate of unfolding, k_u and rate of refolding k_f is taken into account. Both are given in Equations 2.35, 2.36 and 2.37.

$$k_u\left(F\right) = k^0_{u,force} \exp\left[\beta\left(Fx_u - \frac{1}{2}k_c x_u^2\right)\right] \tag{2.35}$$

$$k_F\left(F\right) = k^0_{F,force} \exp\left[-\beta\frac{k_c}{2}\left(\frac{F}{k_c} - x_u\right)\right] \tag{2.36}$$

$$k_F\left(F\right) = k^0_{u,force} \exp\left[\beta\left(\Delta G^0_{TS,force} - \frac{F^2}{2k_c}\right)\right] \tag{2.37}$$

Where F is the unfolding force, x_u is the distance between the unfolded state and the barrier to unfolding, k^0_u is the rate of unfolding at zero force and k^0_F is the rate of refolding at zero force. $\exp\left(\beta\Delta G_{TS,force}\right) = k^0_{F,force}/k^0_{u,force}$ is used to define the equilibrium free energy, $\Delta G^0_{TS,force}$. An approximation can be made for the unfolding force, $F(r)$ by:

$$\langle F(r)\rangle = f_{eq} + \frac{k_B T}{x_u}\ln\left(1 + \frac{Re^{-\gamma}}{k_u\left(f_{eq}\right)\frac{k_B T}{x_u}}\right) \tag{2.38}$$

Where γ is Euler's constant. This can be used to fit the force versus loading rate, R, relationships. It is hard to determine the loading rate for polyproteins. This is due to the complications in measuring the loading rate at each protein unfolding event. The loading rate depends on the effective spring constant, which is a combination of the spring constant of the cantilever and protein. As a protein unfolds the effective spring of the chain changes. Due to the stochastic nature of the number of unfolding events, the loading rate is not constant between force-extension traces. For simpler systems such as the rupture of a single protein or interaction between two molecules, it is quite easy to correct for the loading rate as described in Section 2.3.10.2. Friddle and coworkers also extend the method to multiple parallel bonds between the tip and substrate. Due to the number of free parameters we were unable to use this model in this report.

2.3.10.2 Dudko-Hummer-Szabo model

The Dudko-Hummer-Szabo (DHS) model [348, 353] is based on Kramer's theory for the movement of a gas molecule over a diffusive barrier. This model aims to account for the non-linearity in the data observed in some F-log(Pulling speed) plots. The theory behind the model was originally completed by Hummer and Szabo (2001) [354]. In this model the lifetime of a bond is dependent on an arbitrary exponent, a. Equation 2.39 [348, 353] gives the relationship between the change in free energy $\Delta\ G^0_{TS,force}$, the distance between the unfolded and transition state, x_u and the loading force, F. τ_0 is the lifetime of the bond at zero force.

$$\tau(F) = \tau_0 \left(1 - \frac{aFx_u}{\Delta G^0_{TS,force}}\right)^{1-\frac{1}{a}} e^{\Delta G\left[1-\left(\frac{aFx_u}{\Delta G^0_{TS,force}}\right)^{\frac{1}{a}}\right]} \tag{2.39}$$

The equation governing this relationship can be reduced to the widely accepted Bell's model with a=1.

The loading rate, R, during an experiment, for a single molecule, can be approximated by $R = k_c v$, however, when multiple proteins or molecular linkers are present the spring constant must be modified. For the Bell model this is done by accounting for the change in effective spring constant of the protein chain in Monte Carlo simulations. Dudko et al. (2008) showed that, when molecular linkers are present, the rate needs to corrected to Equation 2.40. Where L_p is the persistence length, L_c is the contour length, k_c is the cantilever spring constant, β is $(k_B T)^{-1}$ and v is the pulling speed of the experiment.

$$R(F) = v\left[\frac{1}{k_c} + \frac{2\beta L_c L_p\left(1+\beta F L_p\right)}{3+5\beta F L_p + 8\left(\beta F L_p\right)^{\frac{5}{2}}}\right]^{-1} \tag{2.40}$$

The distribution of lifetimes, $p(F)$, is related to the lifetime by Equation 2.41.

$$p\left(F\right) = \frac{\exp\left(-\int_0^F \left[r\left(F\right)\tau\left(F\right)\right] df\right)}{R\left(F\right)\tau\left(F\right)} \tag{2.41}$$

Dudko et al. (2008) showed that this could be inverted to the lifetime by Equation 2.42.

$$\tau\left(F\right) = \int_F^{\inf} p\left(F\right) df / \left[R\left(F\right) p\left(F\right)\right] \tag{2.42}$$

This expression suggests that the distribution of the forces can readily describe the rate of unfolding of the protein, by approximating the probability of the distribution as a

Gaussian distribution the expression can be approximated by Equation 2.43.

$$\tau\left(F\right)=\left[\frac{\pi}{2}\left(\left\langle F^{2}\right\rangle-\left\langle F\right\rangle^{2}\right)\right]^{1/2}/R\left(\left\langle F\right\rangle\right) \tag{2.43}$$

In an ideal experiment, the true $R(F)$ can be determined for each unfolding event, the histograms could be separated into N bins, where each bin would correspond to a different loading rate. The force distributions can then be used to determine the lifetime of the bond as a function of loading rate (Equation 2.43), which in turn can be fitted by Equation 2.39. This method has the advantage of not requiring approximations of the shape of the energy landscape. However, in this form, it cannot be used for multiple barriers to unfolding, where the barriers are codependent on the preceding and following unfolding events. When a poly-protein unfolds each barrier is dependent on the preceding barrier [328, 355]. Zhang and Dudko (2013) have since developed this model so that it could be used for such a system. In this paper they describe a method for mapping the forces from multiple unfolding events into force dependent rates. This paper assumes that the free energy landscape has an arbritary number of barriers with metastable intermediates between the folded and unfolded state. The population in the system that will move from state i to state j when the force changes by dF (or similarly between time t and dt) is given by:

$$P_{ij}|dF|=P_{ij}(t)dt=k_{ij}(F(t))\mathcal{N}_{i}(t)dt \tag{2.44}$$

Where $\mathcal{N}_i(t)$ is the population of state i at time t and is dependent on the original population of this state ($\mathcal{N}_i^0$) and the number that transitioned to state i from/to the m^{th} state. $k_{ij}(F)$ is the force dependent rate of moving between the two states. This led the authors to the following relationship:

$$k_{ij}(F)=|\dot{F}(F)|_{i}\frac{P_{ij}}{\mathcal{N}_{i}(F)} \tag{2.45}$$

$$\mathcal{N}_{i}(F)=\mathcal{N}_{i}(F(t))=\mathcal{N}_{i}^{0}-\sum_{m}\int_{0}^{t}P_{im}(t')dt'+\sum_{m}\int_{0}^{t}P_{mi}(t')dt' \tag{2.46}$$

Equations 2.45 and 2.46 can be used to map the experimental force to force dependent loading rates. Where P_{im} is the population that moved between i to m and P_{im} is the population that moved between m to i. $P_{ij}(F)=n_{bin}/w_{bin}$ where n_bin is the number of counts in a bin and w_{bin} is the width of each bin. $|\dot{F}(F)|_i$ is the loading rate in state i to correct for the constantly changing loading rate between systems. The loading rate can be read directly from the force-extension profiles as the gradient of the force-time

trajectory. There are two methods for determining $\mathcal{N}_i(F)$: by counting the number of force trajectories in state i at force F or by approximating it as the integrals in Equation 2.46 by the corresponding bins in the histogram. The force-dependent rates for each unfolding event could then be fit by Equation 2.43 as described above. This method would require separation of unfolding events based on the force-extension profile. This can be complicated when the full poly-protein has not unfolded or the chain is picked up elsewhere. To be able to accurately pool unfolding events, one would have to know where the chain was picked up. This method does provide a new procedure for analysing force-extension traces without the need to incorporate the WLC model. However, it is impossible to determine how many protein domains are initially picked up by the cantilever, and therefore how many barriers are preceding an unfolding event. Often the polyprotein can detach from the tip before all unfolding events have occurred. To implement this model accurately a large set of data only allowing unfolding of full constructs should be used to overcome this problem. However, this was not possible in this project and therefore this model was not implemented.

Chapter 3

Expression, purification and characterisation of deuterium labelled proteins

3.1 Introduction

The first research goal of this project was to design and produce model systems to determine the importance of hydrogen bond strength and hydrophobic interactions on protein stability. This chapter will begin by introducing the model proteins in this study. The expression of protiated and deuterated monomeric and polyprotein domains, will then be presented before discussing the characterisation of the proteins.

Figure 3.1.0.1 outlines the four test systems studied in this project:

1. A protiated protein p(protein) in H_2O is the wild-type protein. This serves as the reference system, i.e, the baseline of the native contacts in the protein.

2. A protiated protein p(protein) in D_2O. In this system the protein is fully isotopically exchanged in D_2O.

3. A deuterated protein d(protein) in H_2O. In this system the deuterated protein is fully exchanged in D_2O.

4. A deuterated protein d(protein) in D_2O. This protein will be fully exchanged in D_2O.

Mechanical studies on these proteins systems will determine information about specific regions of the protein which are involved in the mechanical clamp. Therefore studying

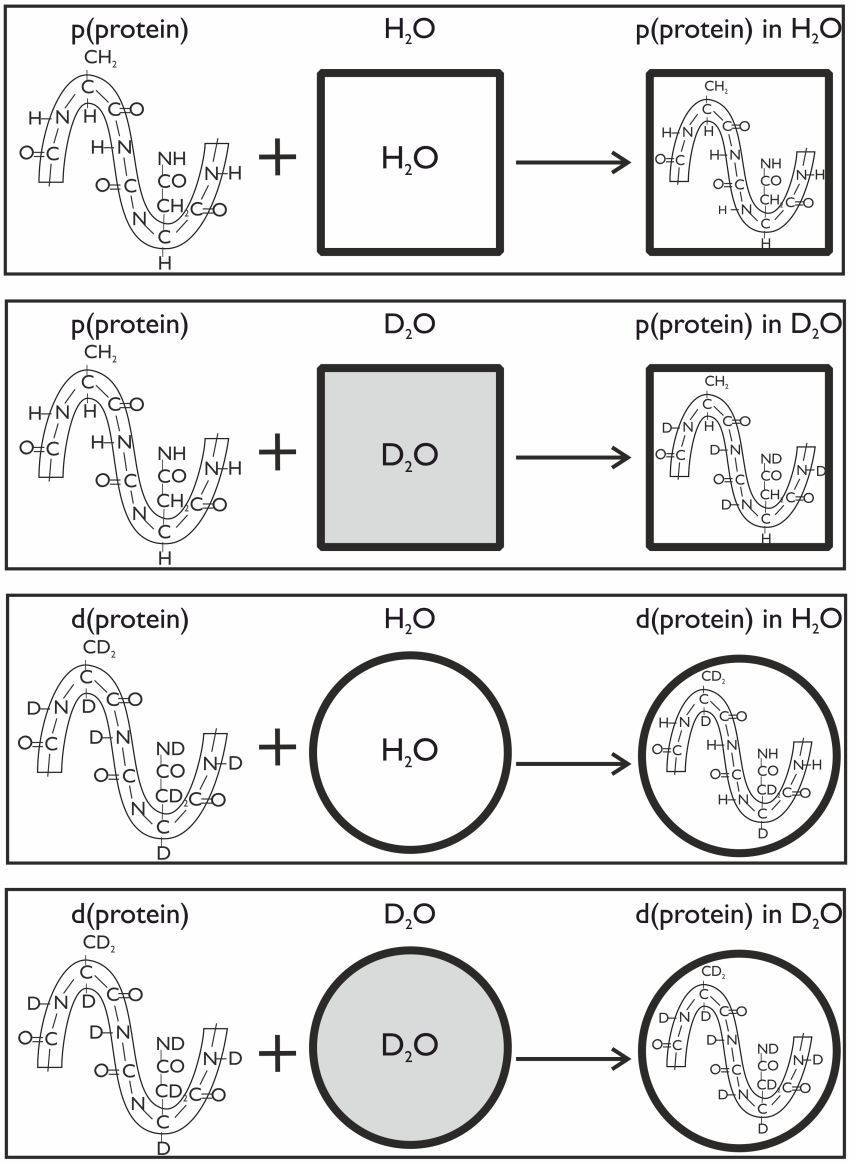


Figure 3.1.0.1: Figures illustrating the four test systems used in this project where the proteins are represented by an arbitrary peptide. The schematics also illustrate the positions of deuterium or hydrogen in the protein.

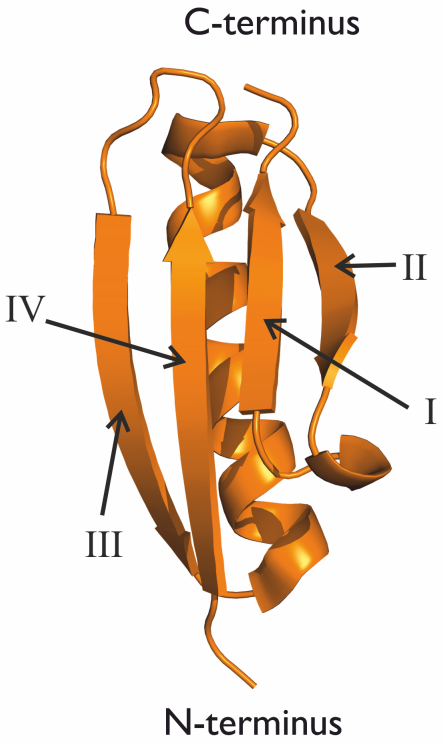


Figure 3.1.1.1: The 3-D structure of the protein pL PDB code: 1HZ6 [8, 9]. The protein contains 72 amino acids and is made up of 4 β-strands in a β-sheet which fold onto an α-helix. Structure made using Pymol [10].

proteins with well-defined mechanical clamp regions is necessary. The two proteins chosen on this basis are I27 and protein L (pL).

3.1.1 Protein L

Since the first mechanical unfolding experiment of a protein in 1996, the number of different proteins studied by force, with different topologies, has increased[189]. These studies indicate the importance of secondary structure, topology and pulling geometry on the mechanical resistance of proteins[216, 247]. One study predicted a small protein, with no known mechanical function, to have significant mechanical resistance based solely on its simple topology. This protein is the pseudo-wild type[1] of the protein L B1 domain (referred to as pL) [217]. This small protein is found in the bacterial cell wall of *Peptostreptococcus magnus* and binds immunoglobulin G (IgG) domains [356]. This protein has been studied both kinetically (by mechanical [217] and chemical denaturation) and thermodynamically (by equilibrium denaturation) [357]. Mechanical studies of this protein will be discussed in more detail in Section 4.1.1.

The crystal structure of the mutated pL has been determined [9]. The protein consists of an α-helix packed onto four-stranded β-sheet (see Figure 3.1.1.1). Thermodynamic and

[1]It contains a tyrosine to tryptophan substitution to aid biological studies

kinetic studies of this protein indicate unfolding can be described by a simple two-state model and no intermediates are observed [357]. The simplicity of the unfolding mechanism of this protein makes it a good model for protein unfolding.

3.1.2 Protein I27

Titin is a giant muscle protein from the human sarcomere and has been the subject of many mechanical unfolding experiments[340, 358–362]. It is the largest single chain protein in the human body made up of 27 000 amino acids [363] and spans over half the length of the sarcomere (Figure 3.1.2.1). The protein is divided into modules called the I band and A band and is responsible for the passive elasticity when the muscle fibre is stretched [363]. The A-band consists of tandem repeats of Immunoglobulin (Ig) and fibronectrin III (FnIII) domains. The I band is also made up tandem repeats of Ig domains but also has a PEVK (proline, aspartic acid, valine and lysine) rich domain and N2B region, the latter of which contains a unique sequence (i.e, not repeated in the rest of the protein) of 572 residues. There are over 300 Ig and FnIII domains in the entire protein. In 1997 three studies were published on the mechanical unfolding of this large protein [340, 358, 359]. Two of these mechanical unfolding studies were performed using an optical trap [358, 359] and one was conducted using an AFM [340]. All studies showed similar results indicating the unfolding of the protein at low and high forces could be modelled as a WLC. The optical trap measurements indicated that titin could unfold at high forces and refold at low forces, but between 5–40 % of the giant protein remained unfolded[358]. However, the protein behaved differently in the two force regimes; At low forces the PEVK domains elongate and at high forces the Ig and FnIII domains unfold [359]. Furthermore the protein displayed hysteresis between the extension and relaxation cycles indicated the different folding kinetics of the chain. The SMFS study using the AFM indicated that the force-extension cycle of the protein gave a characteristic sawtooth pattern of the unfolding forces (Figure 3.1.2.2) indicative of single Ig and FnIII protein domains unfolding in an all or none fashion [340].

Ig and FnIII domains have therefore been extensively studied using force-spectroscopy studies. One protein in particular has been a popular candidate for mechanical unfolding and theoretical studies ([214, 218, 252, 261, 316, 328, 365]); the protein I91. This protein, formally known as I27, consists of 81 residues and is the 27th Ig domain from the I-band of titin and was initially selected for study due to the availability of a high-resolution structure [114]. I91 will be referred to protein I27 for simplicity. I27, shown in Figure 3.1.2.3 is formed of 8 β-strands labelled A',A-G, these strands make up two β-sheets

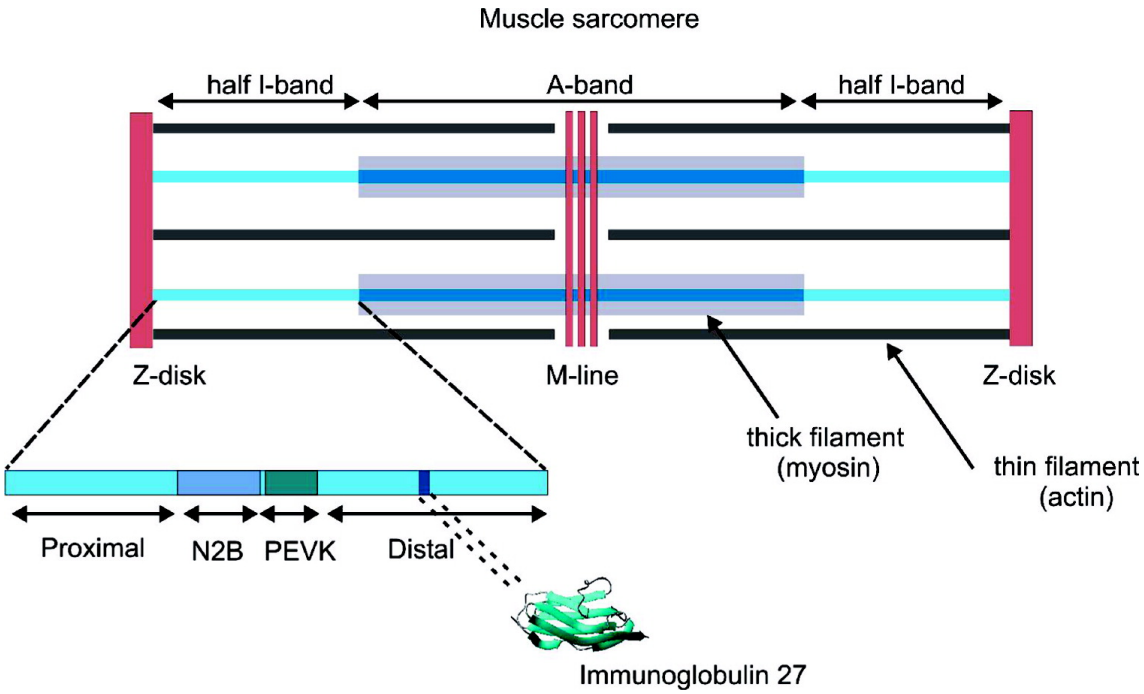


Figure 3.1.2.1: Schematic showing the architecture of the sacromere. Titin spans from the M-line to the Z-disk. Titin is mostly composed of structured Ig and FnIII domains but also has unstructured PEVK and N2B regions. Protein I27 is located in the I-band of the protein. Reprinted with permission from [364]. Copyright 2007 American Chemical Society.

packed onto one another in a β-sandwich conformation. I27 is a good model system for this project due to the detailed understanding of its mechanical properties gained over almost 20 years of study.

A homo-polyprotein of a mutant of I27 (T42A, A78T) was first constructed, and then experimentally unfolded, by Carrion-Vazquez et al. (1999). In this study they compared the mechanical and chemical unfolding rates of the protein [252]. Initially these experiments indicated that the rate of unfolding obtained from mechanical, $k_{u,force}{}^0$, and chemical denaturation, $k_{u,chemical}{}^0$, were similar ($k_{u,force}{}^0$=3.3 $\times$ 10^{-4} s^{-1} and $k_{u,chemical}{}^0$=4.9 $\times$ 10^{-4} s^{-1} , however later experiments on cysteine free mutants indicated they probed different unfolding pathways [214]. It is the cysteine free mutant of I27 (I27 C43S,C63S) that is the focus of this study due to the lack of cysteines in the protein. The chemical denaturation experiments on this protein show it is significantly destabilised by the cysteine mutations compared to the T42A, A78T mutant[214]. The $\Delta\Delta G_{UN,chemical}$, the difference between the $\Delta G_{UN,chemical}$ of the two mutants, is 18 kJ mol^{-1}. The rate of unfolding at 0M guanidine hydrochloride, $k_{u,chemical}{}^0$, has also been found to increase from 4.9 $\times$ 10^{-4} s^{-1} for the T42A, A78T mutant to 10.3 $\times$ 10^{-3} s^{-1} for the C43S, C63S mutant and the rate of folding is lower, $k_{f,chemical}{}^0$, from 30 s^{-1} to 1.9 s^{-1} for the proteins respectively [214]. The transition state placement, β_T, measured by

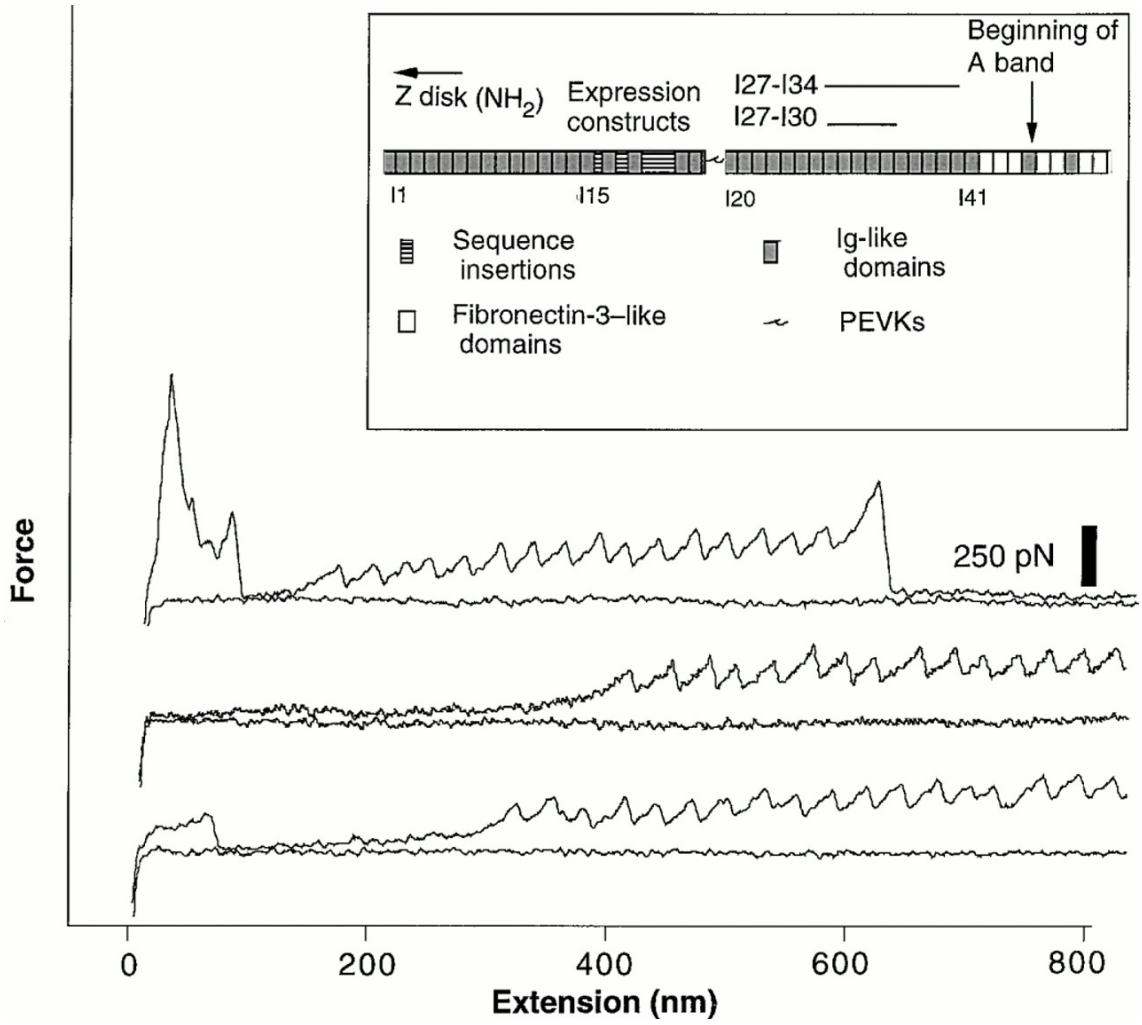


Figure 3.1.2.2: Force extension traces from mechanical unfolding of the entire native titin protein. The traces have sawtooth peaks that are indicative of protein unfolding events. Up to 20 forces peaks were observed in the traces. Many had a long plateau before the sawtooth peaks were observed, these are attributed to unfolding of the PEVK segments. From [340]. Printed with permission from AAAS.

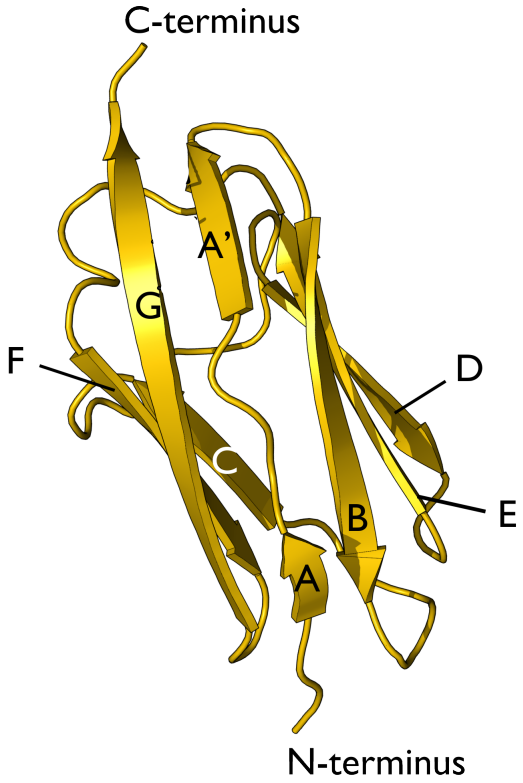


Figure 3.1.2.3: The 3-D structure of the protein I27 PDB code: 1TIT [8, 114]. The protein is 89 amino acids long and is made up of 8 β-strands which fold to form a β-sandwich fold. The strands are labelled A', A-G. This structure was drawn using PyMOL [10].

chemical denaturation, is similar (0.90 kJ $mol^{-1}M^{-1}$ for the T42A, A78T mutant and 0.94 kJ $mol^{-1}M^{-1}$ for the C43S, C63S mutant [214]).

This protein has been extensively studied mechanically [124, 214, 317] and thermodynamically [214] and therefore is a good candidate to study. The mechanical properties of this protein will be discussed in more detail in Chapter 4.1.2. The mutant I27C43SC63S will be referred to as I27 in this thesis.

3.1.3 Objectives

To our knowledge, deuterated I27 and pL have never been expressed or purified, furthermore no deuterated polyprotein construct has been expressed or characterised. The objectives of this chapter are as follows:

1. Determine whether fully deuterated proteins can be successfully expressed and purified.

2. Confirm that the protein is correctly folded in deuterated form.

3. Determine the thermal melting temperature of each of the I27 and pL test systems.

4. Perform chemical denaturation experiments of the four I27 and pL test systems to determine equilibrium free energies and unfolding and folding rate constants.

3.2 Results

3.2.1 Expression of protiated and deuterated monomeric and polyproteins

This section outlines the results from expression and purification of protiated and deuterated monomeric and pentameric pL and I27 proteins. The monomers were expressed for characterisation using thermal and chemical denaturation experiments. The polyproteins were required for mechanical denaturation experiments. Initially an attempt was made to adapt the BLR[DE3]pLysS *E.coli* cells to growth on deuterated carbon source, ISOGRO-D. However after three iterations of the adaptation, no improvement was found in the growth rate of the bacteria. Expression trials were carried out as described in Section 2.2.5 for all proteins except $p(I27)_5$. The optimal time for protein expression for the $p(I27)_5$ protein was previously determined to be 4 hours [214]. SDS PAGE gels were used to determine the level of expression of each protein. Example SDS PAGE gels for the deuterated monomer and polyprotein are shown in Figure 3.2.1.1. These gels were used to determine the optimal growth time for maximal expression of the protein and the maximal expression level determined the scale of growth culture. The bacteria containing the plasmid for $d(I27)_5$ were grown in 1 litre of ISOGRO-D (10 $\times$ 100 ml in 500 ml flasks) for 7 hours after inducing with IPTG. The bacteria containing the plasmids for d(I27) and $d(pL)_5$ were grown in 1 litre (10 $\times$ 100 ml in 500 ml flasks) and d(pL) 2 litres (4 $\times$ 500 ml in 2.5 litre flasks) of ISOGRO-D dissolved in 99.8 % D_2O for 24 hours after induction with IPTG. All protiated proteins were grown for 4 hours in 10 litres (10 $\times$ 1 litre in 2.5 l flasks) of LB after induction with IPTG.

3.2.2 Purification and preparation of proteins

To obtain pure protein samples after expression, two purification steps were performed these were outlined in Section 2.2.9. A second large scale expression and purification of $p(I27)_5$ was carried out by Gael Radou and this protein was also used in this project.

3.2.2.1 Affinity chromatography of His-tagged proteins using Ni-NTA

The bacteria containing the protein were lysed as described in Section 2.2.8. The supernatent, or lysate, was added to a Ni-NTA column. Figure 3.2.2.1 shows the SDS PAGE gel with the steps involved in this purification. The protein was eluted from the column

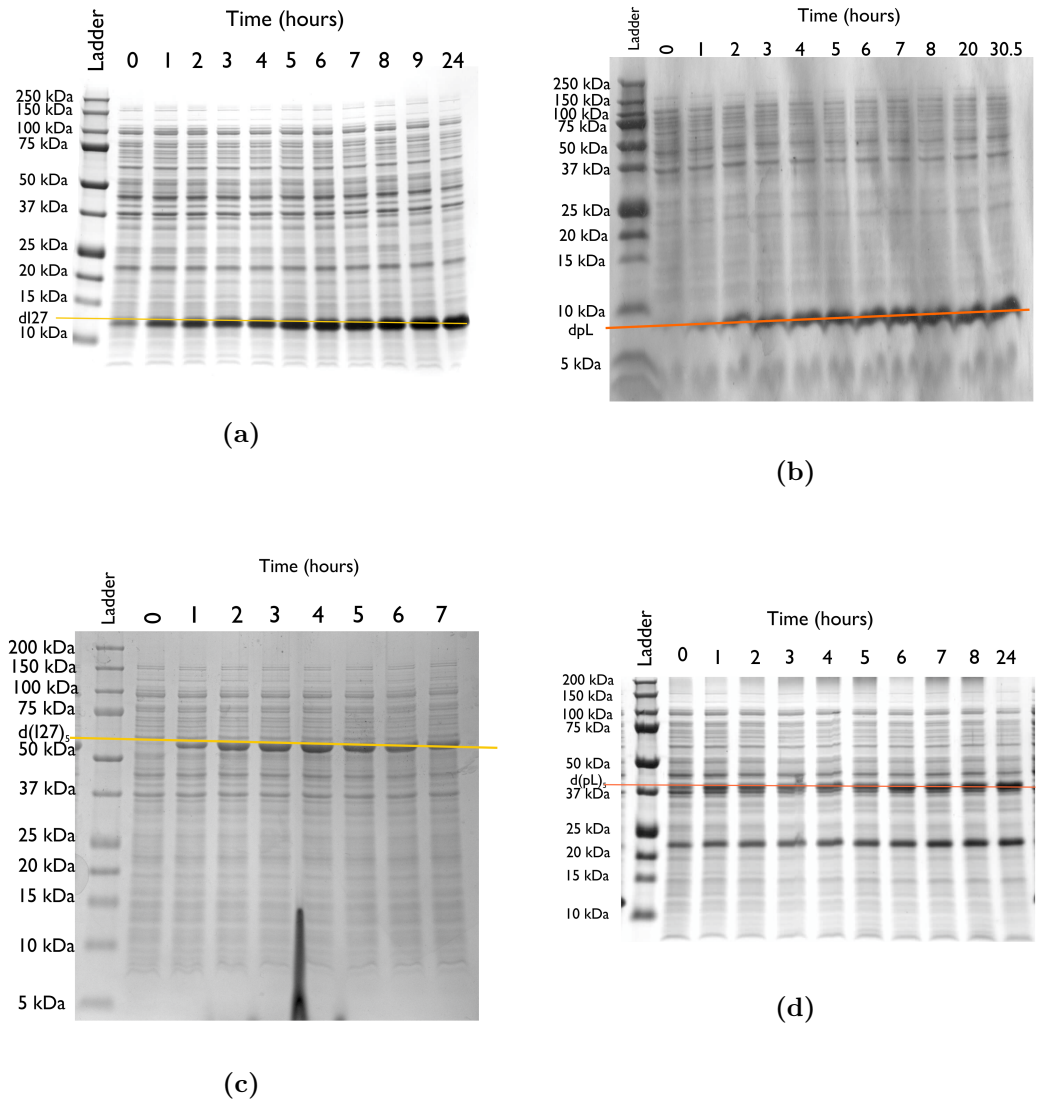


Figure 3.2.1.1: SDS-PAGE gels showing the expression trials of the deuterated monomeric (a) I27 and (b) pL and deuterated pentameric constants (c) $(I27)_5$ and (d) $(pL)_5$ proteins. 0 hours on (a) and (b) indicates before IPTG. The protein sizes of the ladder are labelled along with the coloured band corresponding to the estimated weight of each of the proteins. For the deuterated I27 monomer, growth after 8 hours resulted in a maximal protein expression. For the deuterated protein L monomer, growth after 7 hours resulted in maximal protein expression. The maximum expression was observed after 6 hours for the detuerated $(I27)_5$. Growth for 24 hours of $(pL)_5$ resulted in a protein band with sufficient intensity suggesting maximal protein expression. Note: each sample was normalised for cell density.

Protein	Molar extinction coefficient $M^{-1}cm^{-1}$	Estimated molecular mass Da
p(I27)	6970	10878
d(I27)	6970	11643
$p(I27)_5$	34970	52219
$d(I27)_5$	34970	55935
p(pL)	9970	8039
d(pL)	9970	8577
$p(pL)_5$	47770	39952
$d(pL)_5$	47770	42676

Table 3.2.1.1: Table indicating the predicted molar extinction coefficient and molecular weight based on the amino acid composition of the protein. The deuterated masses are based on a fully deuterated protein construct. The estimated mass was determined using the ExPASy server [366].

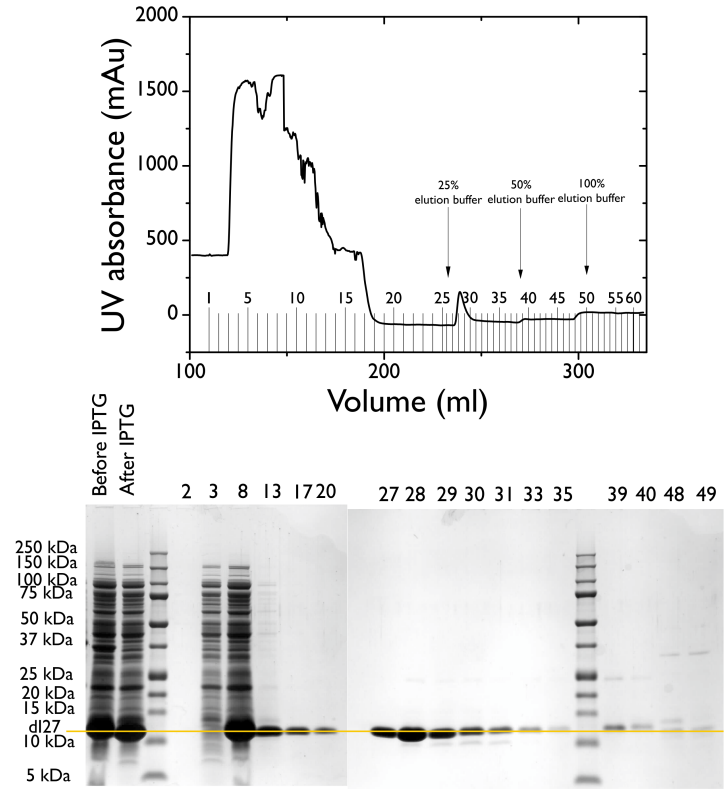
in the 25 % elution step. The SDS PAGE gels were obtained for each protein construct. If protein was observed in the loading stage, it was reapplied to the Ni-NTA column in a second purification step. Protein eluted, in the 25–100 %, elution stages was pooled and dialysed for further purification. SDS PAGE gels indicated impurities in the eluted samples. The typical amounts of protein predicted from this stage of the purification are shown in Table 3.2.2.1. These were estimated from the area underneath the elution peak, the molar extinction coefficient and formula mass of the protiated protein (given in Table 3.2.1.1). After four changes of dialysis, the protein was snap frozen and freeze-dried in 30 mg aliquots (where possible).

Protein	Estimated yield (mg/l of culture medium)
p(I27)	10
d(I27)	60
$p(I27)_5$	5
$d(I27)_5$	20
p(pL)	9
d(pL)	40
$p(pL)_5$	5
$d(pL)_5$	30

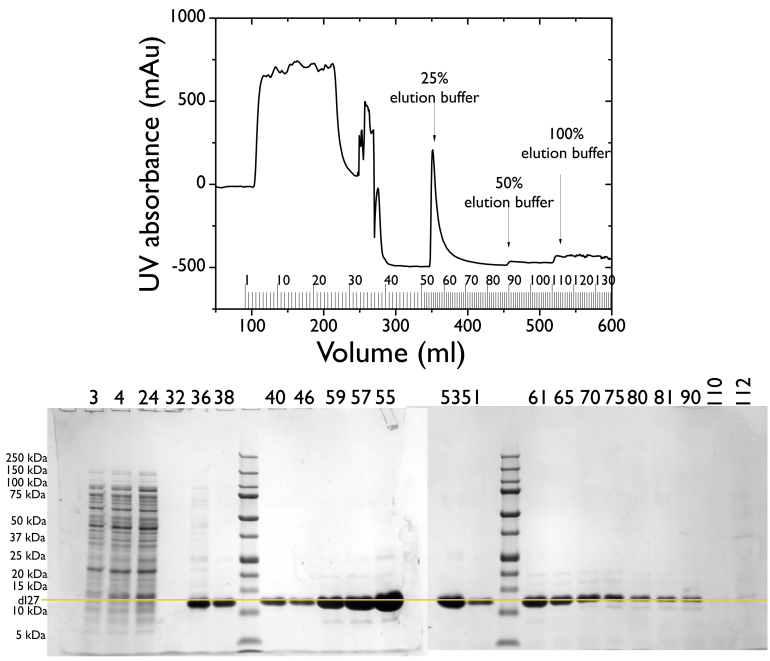
Table 3.2.2.1: Table indicating the estimated yield of the proteins expressed and purified in this project. The estimated yeild was determined from the area under the protein elution peaks from Ni-NTA purification.

3.2.2.2 Size-exclusion chromatography

Protein contaminants were still present in the protein samples after the Ni-NTA stage of purification. These were observed as additional bands on the SDS-PAGE gels from the Ni-NTA chromatography. The protein was purified further using size-exclusion chromatography (SEC) as described in Section 2.2.9.2. Lyophilised protein was resuspended



(a)



(b)

Figure 3.2.2.1: Caption on the following page

Figure 3.2.2.1: SDS-PAGE gels showing the NiNTA for the d(I27). The size of proteins in the reference ladder are shown. The predicted electrophoretic of the target protein is also labelled. The numbers on each gel represent the fraction number on the chromatograms above. (a) and (b) is the NiNTA for dI27. d(I27) underwent two purification steps because a lot of protein was observed in the lysate and wash steps. These fractions were pooled and loaded onto a second NiNTA column with a larger binging capacity. The lysate was loaded over fractions 1-18 (a) and 1-38 (b). The column was then washed with washing buffer over fractions 19-25 (a) and 39-49 (b). Both dI27 gels show protein in the wash steps despite loading on a second NiNTA column. The fractions from (b) were kept and stored at -20 °C. 25 % eluting buffer mixed with 75 % washing buffer was then added to the column. A protein elution peak was observed for each protein. 50 % elution buffer was then mixed with 50 % washing buffer followed by 100 % elution buffer. These steps were to ensure all protein was recovered. Fractions containing protein in the elution steps were pooled for dialysis.

in 3–10 ml of gel-filtration buffer. Only 3 ml of protein could be loaded at a time for each run of the column so multiple runs were performed for all proteins. Figure 3.2.2.3 shows example SDS PAGE gels for the SEC elution profiles of a monomeric and a pentameric protein. The larger polyproteins elute soon after the void volume and the smaller, monomeric proteins elute later due to the smaller size, as expected. For each run, the elution peak was reproducible. The purest samples from the eluted fractions were pooled before dialysis and freeze-drying.

3.2.2.3 Preperation of samples for characterisation

All proteins, excluding p(protein) in H_2O, were isotopically exchanged for full characterisation. This was achieved for all proteins by unfolding the proteins in protiated/deuterated buffer described in Section 2.2.11 before dialysis and lyophilisation. The majority of $d(pL)_5$ was lost using this method (yield <20 %). This suggested that the protein was not stable in the absence of salts within the solvent.

A second method was therefore employed for the preparation of the $d(pL)_5$ proteins. The $d(pL)_5$ protein constructs were exchanged by unfolding them in the relevant buffer whilst immobilised on Ni-NTA beads. The proteins were washed using protiated or deuterated washing buffer before being eluted using protiated or deuterated elution buffer. The samples were dialysed directly into relevant buffers for experiments before snap freezing in liguid N_2. These samples were defrosted before use in each experiment.

3.2.2.4 Determining the level of deuteration of the protein

Mass spectrometry was used to determine the percentage of deuteration of the proteins in non-exchangeable positions[2]. The proteins were unfolded at 50 % acetonitrile, 0.1 % formic acid pH 2.5 in water and the mass of the protein was monitored over 2–3 days to

[2]Mass spectrometry experiments were performed by James Ault.

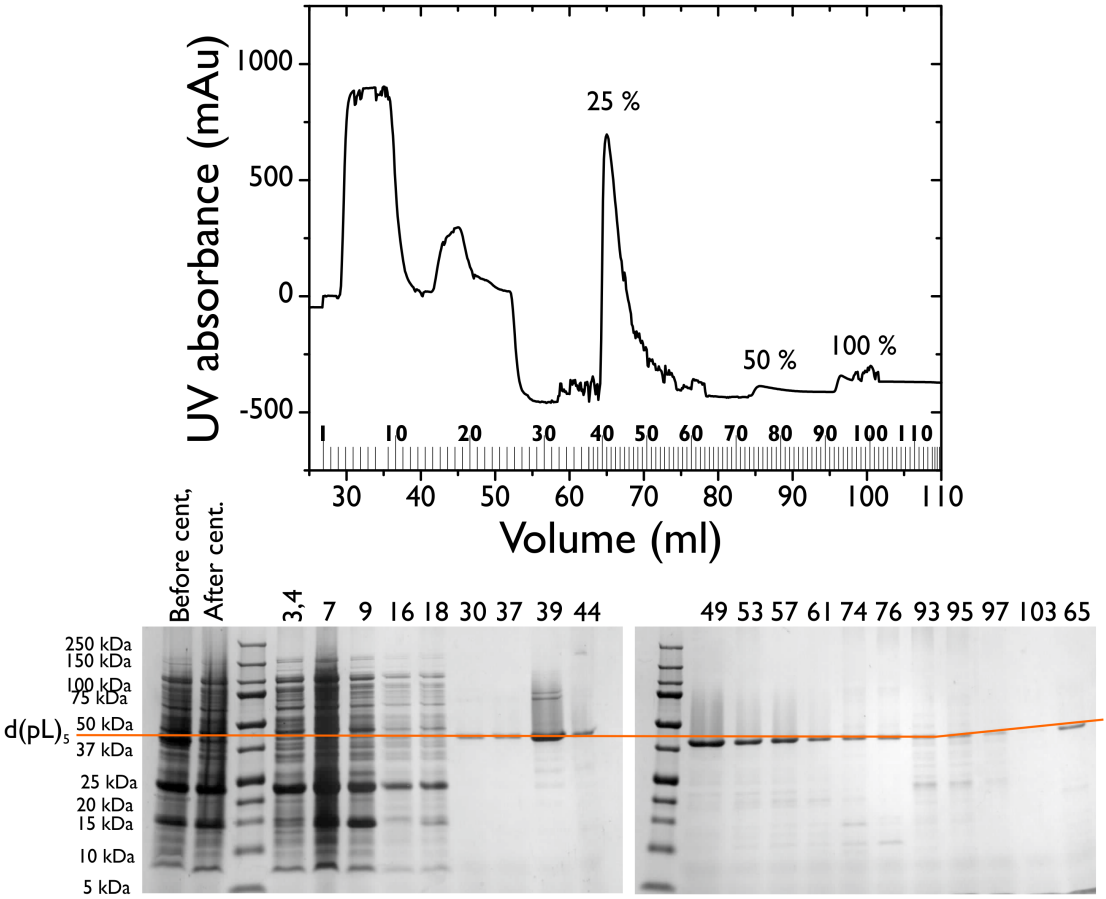


Figure 3.2.2.2: SDS Page gels showing the NiNTA for the $d(pL)_5$. The size of proteins in the reference ladder are shown. The protein band is also labelled. The numbers on each gel represent the fraction number on the chromatograms above. These fractions were pooled and loaded onto a second bigger NiNTA column. The lysate was loaded over fractions 1-26. The column was then washed with washing buffer over fractions 27-37. Some $d(pL)_5$ was observed in the wash step. 25 % eluting buffer mixed with 75 % washing buffer was then added to the column and a protein elution peak was observed. 50 % elution buffer was then mixed with 50 % washing buffer followed by 100 % elution buffer. These steps were to ensure all protein was recovered. Fractions containing protein in the elution steps were pooled for dialysis.

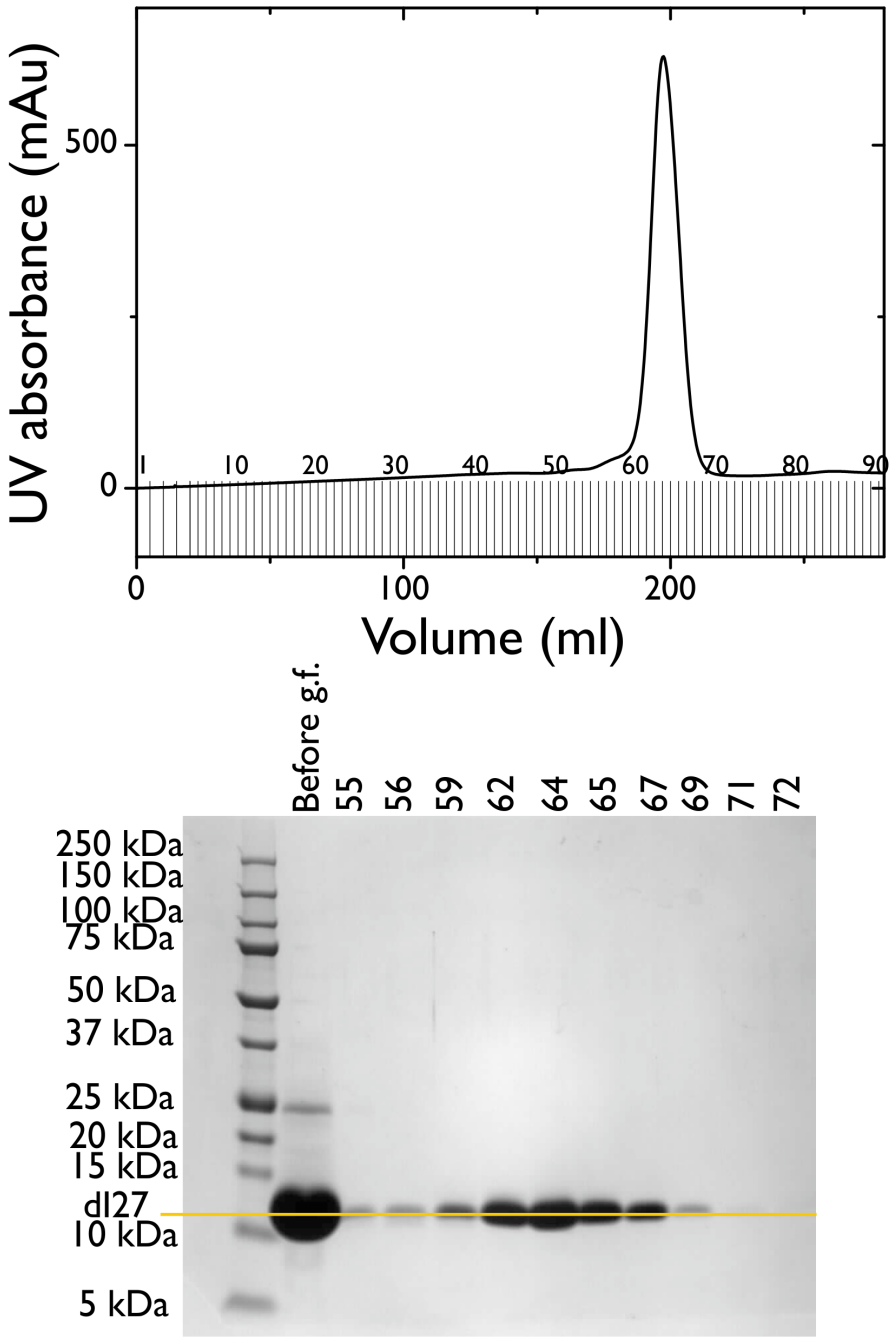


Figure 3.2.2.3: SDS-PAGE gels showing the relative purity of the protein in each fraction from a size-exclusion column for the d(I27). A large peak is observed in the A_{280} peak on the chromatogram when the protein elutes (top).

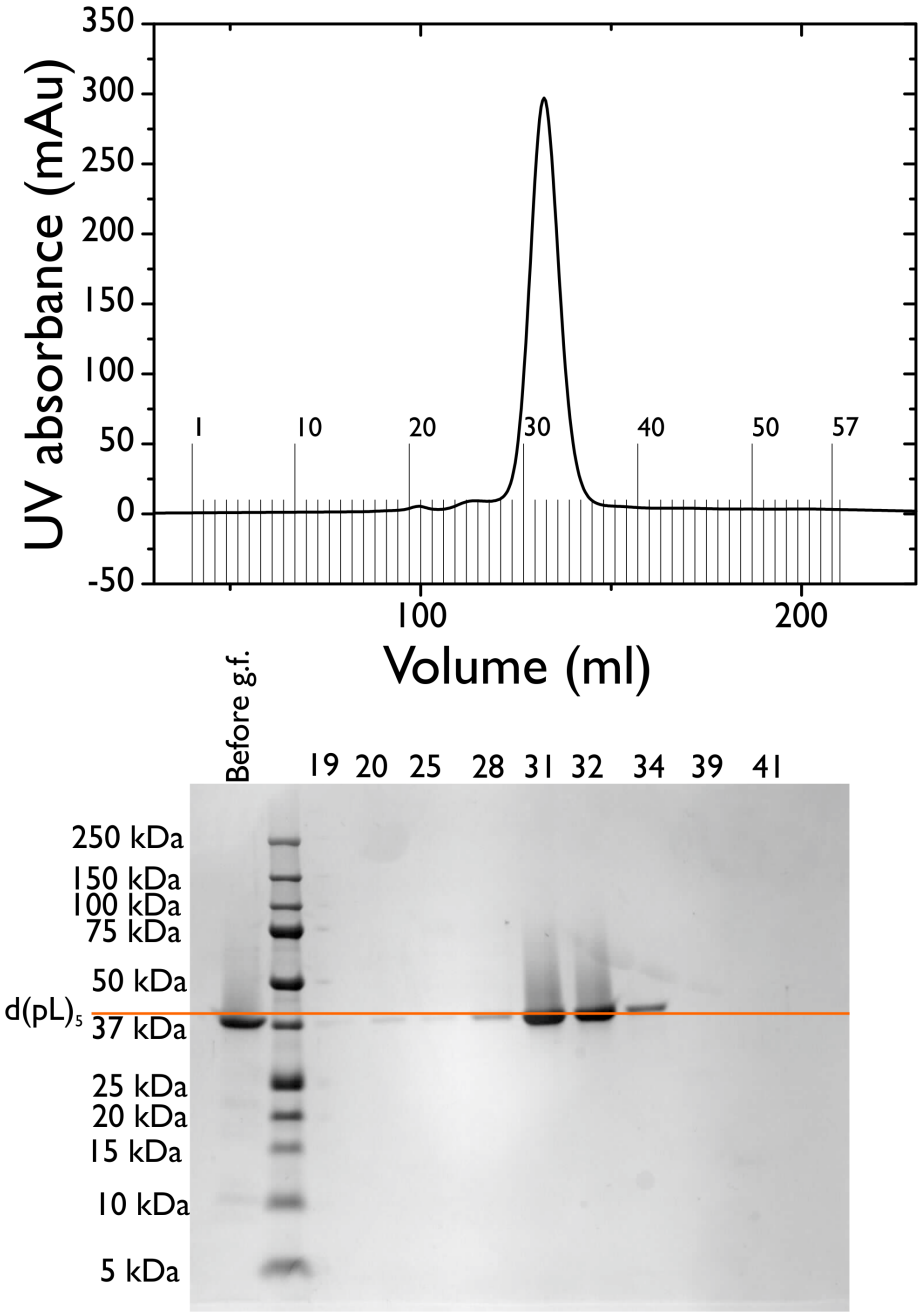
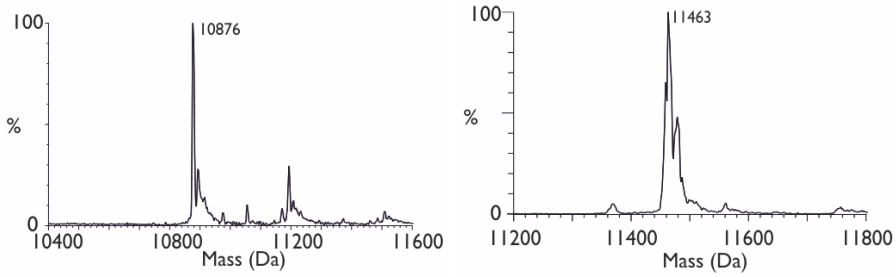


Figure 3.2.2.4: SDS page gels showing the relative purity of the protein in each fraction from the column for the $d(pL)_5$ protein. A large peak is observed in the A_{280} peak on the chromatogram when the protein elutes (top)

determine whether it remained constant. The results indicate that the masses remained stable over the experimental period. The mass spectrometry results are shown in Figures 3.2.2.5–3.2.2.8. The mass spectra monitored for the deuterated proteins over 2–3 days can be found in the Appendix B. To determine the level of deuteration, the number of non-exchangeable hydrogen atoms was calculated for each protein. This mass was compared to the final mass obtained from mass spectrometry to determine the level of deuteration. It was assumed that the protein was fully exchanged to hydrogen in the exchangeable positions during the purification process. Table 3.2.2.2 outlines the calculated mass and percent deuteration for the proteins. The results indicate close agreement between the calculated masses and the experimental masses. This suggested >97 % deuteration for all deuterated protein constructs and is consistent with previous expression of deuterated proteins [130].

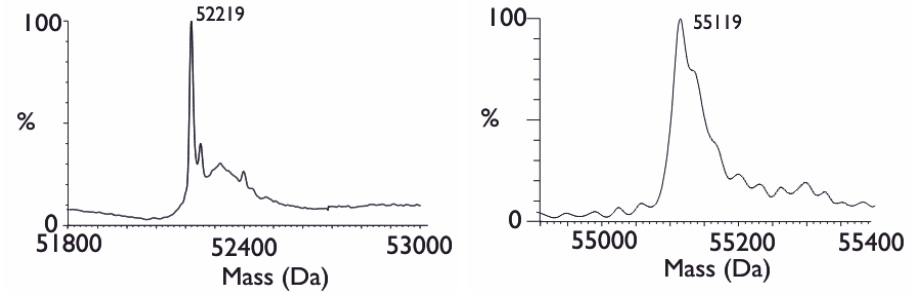
Protein	Mass (Da)	Percent deuteration
p(I27)	10878	n/a
d(I27)	11463	98.6 %
$p(I27)_5$	52218	n/a
$d(I27)_5$	55119	98.9 %
p(pL)	8039	n/a
d(pL)	8437	98.4 %
$p(pL)_5$	39950	n/a
$d(pL)_5$	41994	98.5 %

Table 3.2.2.2: Table indicating the mass of the protiated and deuterated proteins used in this project. The mass of the deuterated protein shown is assumed to be fully exchanged to hydrogen in the exchangeable positions. The percent deuteration of these proteins is obtained based by determining total number of non-exchangeable hydrogen atoms in each protein.



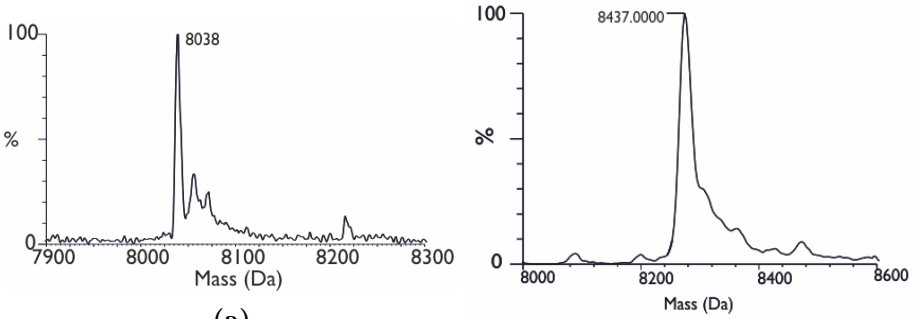
(a) **(b)**

Figure 3.2.2.5: Mass spectra of the purified protein of p(I27) (left) and d(I27) (right). Single peaks can been seen for the mass of each of the proteins. For the d(I27) there is a considerable shift in mass suggesting the protein is deuterated. The mass spectra for this protein were taken every day for two days to check that deuterons at labile sites had fully exchanged to protons.



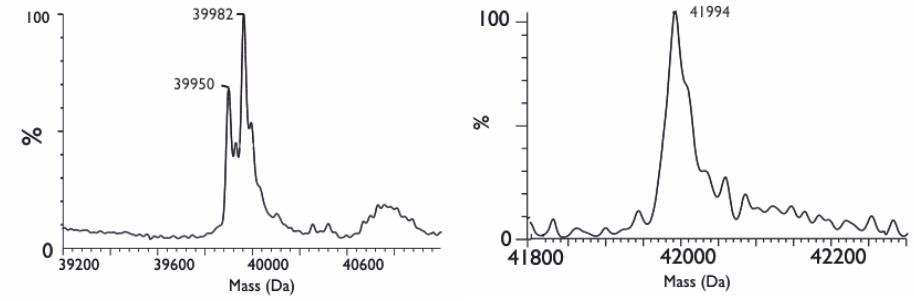
(a) **(b)**

Figure 3.2.2.6: Mass spectra of the purified protein of $p(I27)_5$ (left) and $d(I27)_5$ (right). Single peaks can been seen for the mass of each of the proteins. For the $d(I27)_5$ there is a considerable shift in mass suggesting the protein is deuterated. The mass spectra for this protein were taken every day for three days to check that deuterons at labile sites had fully exchanged to protons.



(a) **(b)**

Figure 3.2.2.7: Mass spectra of the purified protein of p(pL) (left) and d(pL) (right). Single peaks can been seen for the mass of each of the proteins. For the d(pL) there is a considerable shift in mass suggesting the protein is fully deuterated. The mass spectra for this protein were taken every day for three days to check that deuterons at labile sites were fully exchanged to protons.



(a) **(b)**

Figure 3.2.2.8: Mass spectra of the purified protein of $p(pL)_5$ (left) and $d(pL)_5$ (right). Single peaks can been seen for the mass of each of the proteins. For the $d(pL)_5$ there is a considerable shift in mass suggesting the protein is deuterated. The mass spectra for this protein were taken every day for three days to check that deuterons at labile sites had fully exchanged to protons.

3.2.3 Characterisation of proteins using circular dichroism

It has previously been demonstrated that the structure of proteins is unaltered when proteins are deuterated[180, 182, 183]. To determine the effect of deuteration of the secondary structure in the model I27 and pL protein systems studied here, circular dichroism spectroscopy (CDS) was performed as described in Section 2.3.2. CDS can also be used to determine the thermal melting temperature, T_m of proteins. Furthermore, if the folding is reversible, the thermodynamic stability of the proteins can be determined following the protocol outlined in [300].

When protein is unfolded by denaturant it transitions between the folded state (F) and the unfolded state (U). The equilibrium constant, K_{eq}, can be determined from the concentration of the unfolded proteins [U] and the concentration of folded proteins [F] [300]:

$$K_{eq} = \frac{[F]}{[U]} \tag{3.1}$$

This allows the free energy, $\Delta G_{u,temp}$ of unfolding to be computed at every temperature, T:

$$\Delta G_{u,temp} = RT \ln (K_{eq}) \tag{3.2}$$

Where R is the ideal gas constant. The fraction of folded proteins, α is given by:

$$\alpha = \frac{K_{eq}}{(1 + K_{eq})} \tag{3.3}$$

This can be related to the ellipticity of the folded, θ_F and unfolded proteins, θ_U using Equation 3.4. Where θ_T is the ellipticity at temperature T in Kelvin.

$$\alpha = \frac{(\theta_T - \theta_U)}{(\theta_F - \theta_U)} \tag{3.4}$$

The thermodynamic parameters can be determined from the Gibbs-Helmholtz (Equation 3.5). The parameters that can be probed are: the change in enthalpy ΔH, the change in entropy ΔS, the change in free energy of folding, $\Delta G_{f,temp}$ and the melting temperature T_m. The melting temperature is defined as the temperature at which $\alpha = 0.5$[297].

$$\Delta G_{f,temp} = \Delta H(1 - T/T_m) - \Delta C_p((T_m - T) + T \ln (T/T_m)) \tag{3.5}$$

The thermodynamic parameters governing protein folding can be determined by fitting

Equations 3.2–3.5 to the thermal denaturation data. The method outlined by Greenfield (2007) suggests to set change in heat capacity going from the folded to the unfolded state (ΔC_p) to zero for the initial calculations unless it has previously been determined [300]. However, in this study the ΔC_p is estimated from the fitting procedure.

3.2.3.1 Comparison of secondary structures

To determine whether the protein domains were correctly folded, CDS was used to determine the secondary structure of the protein. CD spectra were taken at 23–25 °C for each system. The spectra have previously been determined for the monomeric proteins [357, 367, 368].

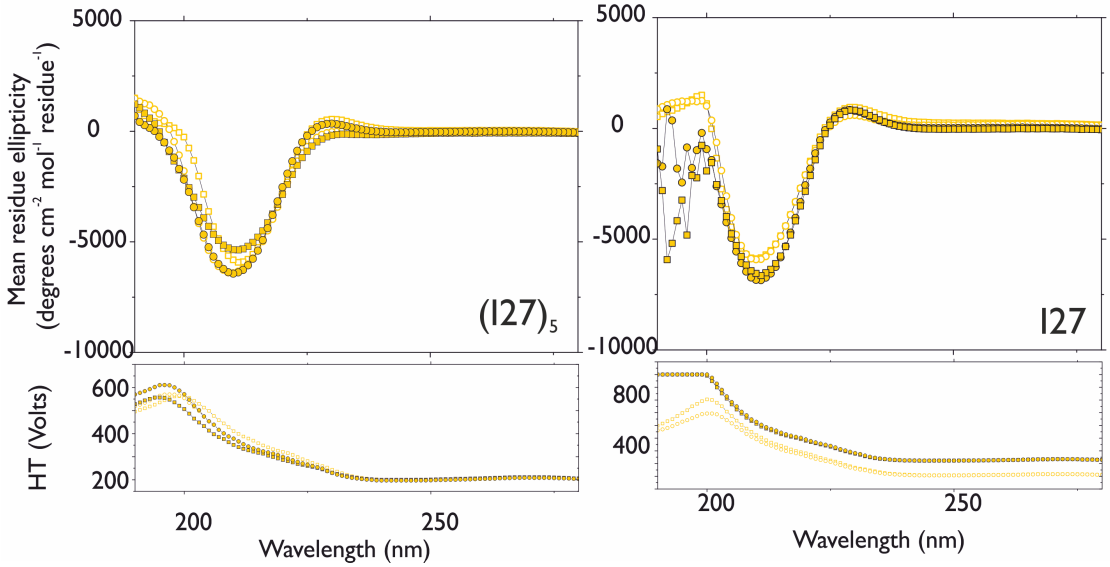


Figure 3.2.3.1: Overlaid circular dichroism spectra at 23 °C for the $(I27)_5$ (left) test systems and the I27 monomer (right) in deuterated or protiated buffers. The high tension (HT) of the instrument is also shown in the inset. A high HT signal indicates a low signal strength from the sample. The detector has to increase its sensitivity to the decrease in the signal intensity from the sample. The units of ellipticity have been corrected for differences in the protein concentration. Left: Open squares are the $p(I27)_5$ in H_2O, open circles are $d(I27)_5$ in H_2O, closed squares are $p(I27)_5$ in D_2O and closed circles are $d(I27)_5$ in D_2O. The spectra should overlay perfectly, however the protein with the lower concentrations has subtle differences in the spectra. Errors could be attributed in the differences in concentration. Right: Open squares are the p(I27) in H_2O, open circles are d(I27) in H_2O, closed squares are p(I27) in D_2O and closed circles are d(I27) in D_2O. The spectra overlay well in this graph apart from at wavelengths <200 nm. At this point the HT for two of the systems is considerably larger and could suggest unreliable data.

The overlaid spectra for the four $(I27)_5$ (left) and I27 monomer (right) test systems are shown in Figure 3.2.3.1. The CD spectra of I27 has a characteristic maximum centred around 230 nm and a minimum around 215 nm [367]. This minimum around 215 nm is indicative of β-sheet structure. The maximum at 230 nm is attributed to β-turns within the protein [369]. The CD spectra all display the characteristics of a folded I27 domain.

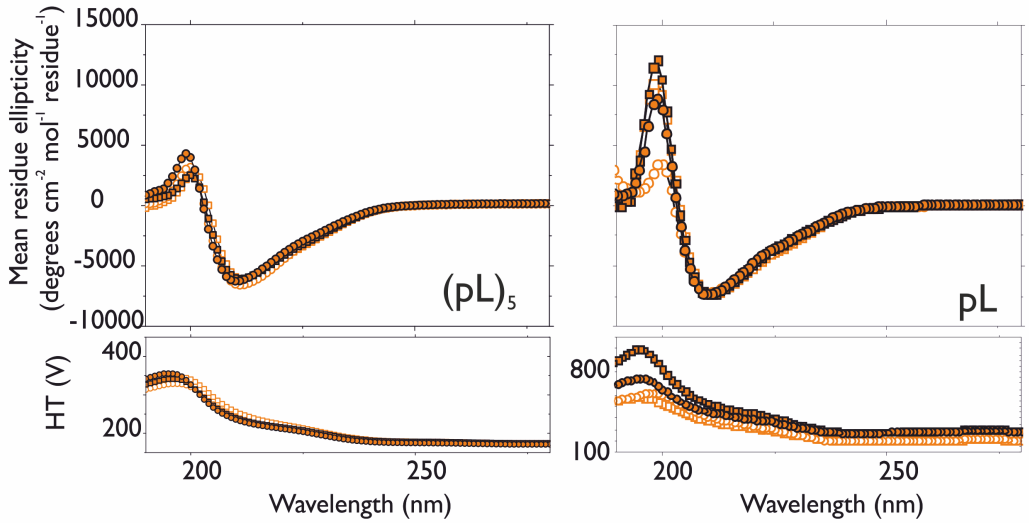


Figure 3.2.3.2: Overlaid circular dichroism spectra at 25 °C for the $(pL)_5$ (left) test systems and the pL monomer test systems. The high tension (HT) of the instrument is also shown in the inset, when the HT is too high, the detector is saturated and the spectra become noisy. The units of ellipticity have been corrected for differences in the protein concentration. Left: Open squares are the $p(pL)_5$ in H_2O, open circles are $d(pL)_5$ in H_2O, closed squares are $p(pL)_5$ in D_2O and closed circles $d(pL)_5$ in D_2O. The spectra are very similar for all systems which suggests the proteins are correctly folded. Right: Open squares are the p(pL) in H_2O, open circles are d(pL) in H_2O, closed squares are p(pL) in D_2O and closed circles are d(pL) in D_2O.

The CD spectra overlay well except the $(I27)_5$ where there were deviations in the spectra in the maximum region at 230 nm. This was attributed to the low concentrations of protein. The two systems that deviated the most were the $p(I27)_5$ in D_2O and $d(I27)_5$ in H_2O. These samples had concentrations <0.3 mg/ml. The spectra of the monomers were also obtained at 25 °C, ensuring the concentrations of the protein samples were $\sim$ 0.35 mg/ml. The spectra for the four I27 monomer test systems overlay well at 25 °C (see Figure 3.2.3.1). This suggests that the protein domains are correctly folded in all systems despite changes to the strength of interactions.

The spectra for protein L also show large minima at 211 nm which are indicative of the α and β structure and agree with the structure determined by Scalley et al. (1997) [357]. The spectra for the $(pL)_5$ and pL systems are shown in Figure 3.2.3.2. The spectra overlay well for all systems in both constructs. This suggests that the proteins are correctly folded in the monomer and polyprotein domains.

3.2.3.2 Thermal denaturation

Thermal denaturation was performed for the polymeric and monomeric protein systems to obtain the midpoint temperature of unfolding, T_m. If thermal denaturation was reversible, that is proteins were able to refold upon cooling, information governing the thermodynamics of the protein folding was also estimated.

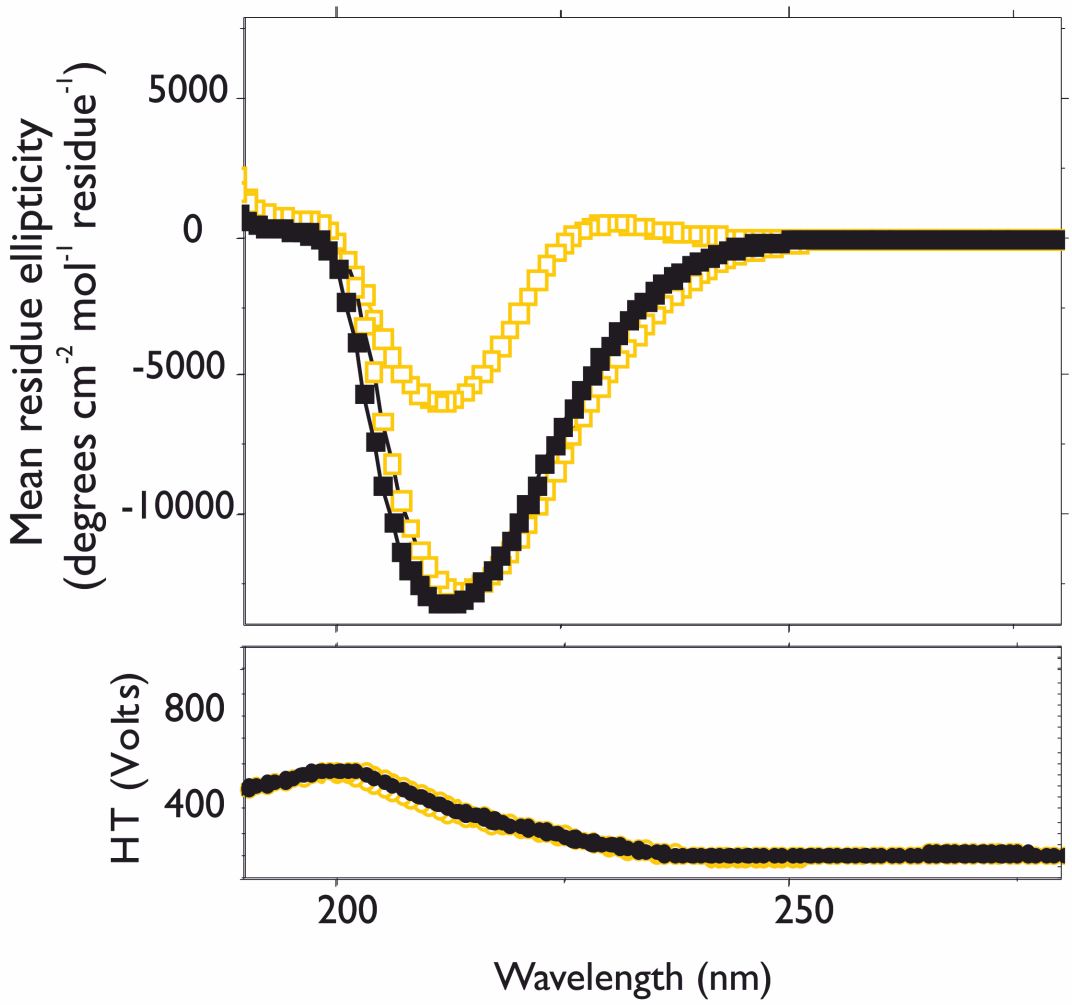


Figure 3.2.3.3: CD spectra of the $p(I27)_5$ in H_2O before and after thermal denaturation. The spectra before denaturation at 23 °C is shown as open squares connected by a solid yellow line. The spectra of the fully unfolded protein is shown in as the open squares connected by the solid black line. Finally the CD spectra of the protein after cooling to 23 °C is shown as the black squares connected by a black line. The spectra indicate that the protein does not refold upon cooling.

Example spectra before and after thermal denaturation of the I27 polyprotein are given in 3.2.3.3. These spectra were the same for the protein solution conditions tested. These demonstrate that the protein does not refold upon cooling to 23 °C. This has previously been observed for polyproteins containing 5 or more tandem repeats of I27

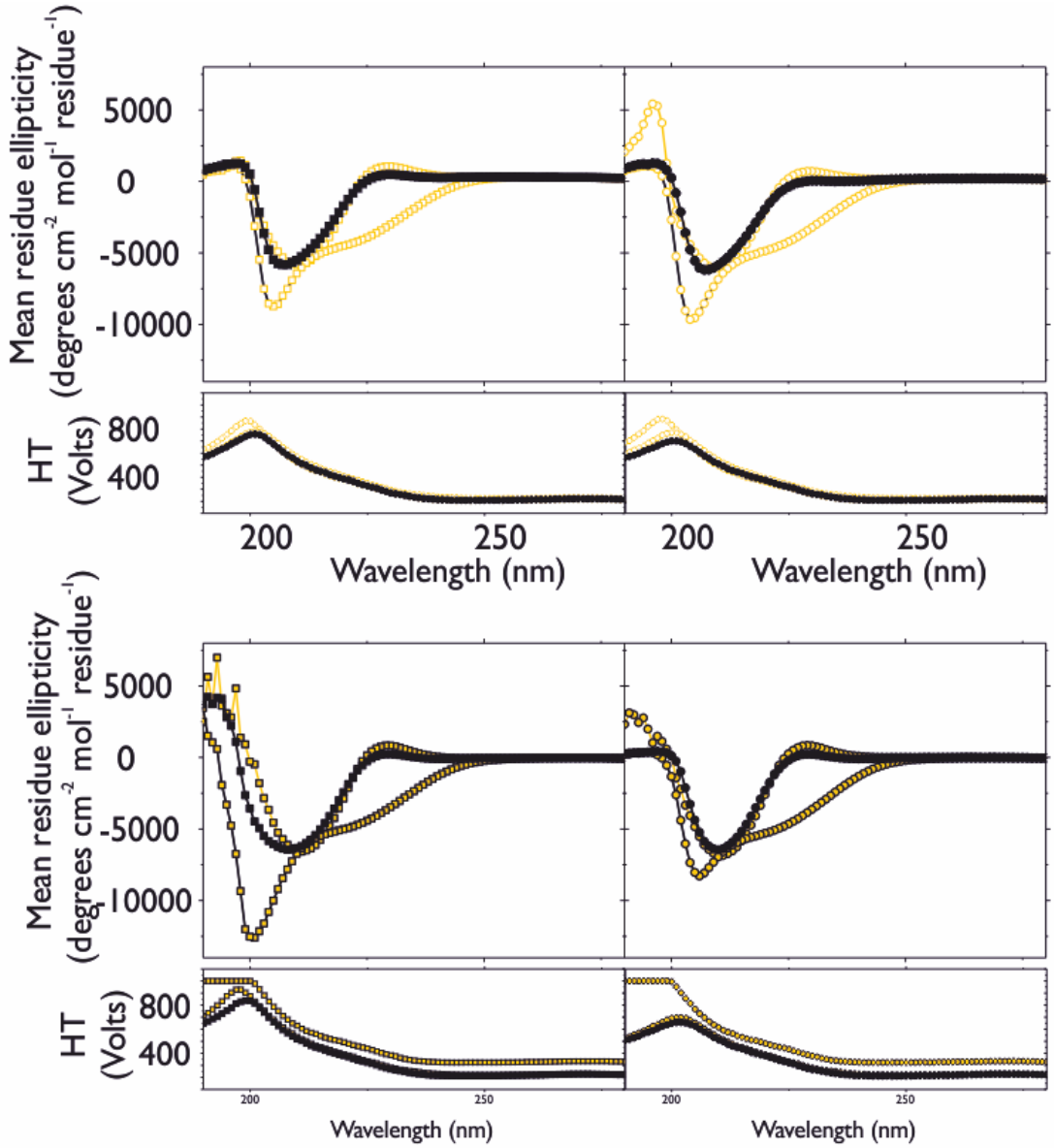


Figure 3.2.3.4: CD spectra of the I27 monomer systems before (25 °C) and after heating to 90 °C. The open yellow squares are the p(I27) in H_2O (top left), the open yellow circles are the d(I27) in H_2O (bottom left), the closed yellow squares is the p(I27) in D_2O (top right) and the d(I27) in D_2O (bottom right) are the closed yellow circles. These markers connected by solid yellow lines indicate the protein before thermal unfolding. The same markers connected by solid black lines are the thermally unfolded spectra of the protein. The black markers connected by a black line is the data corresponding to the protein system after cooling to 25 °C. The spectra show the protein regains the secondary structure upon cooling.

[368]. It has been suggested that this occurs due to subdomain swapping in the unfolded poylpeptide which prevents refolding of the protein domains [368, 370–372]. The spectra before and after heating for the I27 monomers are shown in Figure 3.2.3.4. The spectra show that the majority of structure is reformed upon cooling the protein sample to 25 °C for 1 hour as compared with heating for 5 hours. However, it has been suggested that cooling of the protein sample should be performed at the same rate as heating [300]. Therefore, this could explain some discrepancies between the initial spectra and refolded protein spectra. It was thus assumed that the thermal denaturation of the I27 monomers was reversible.

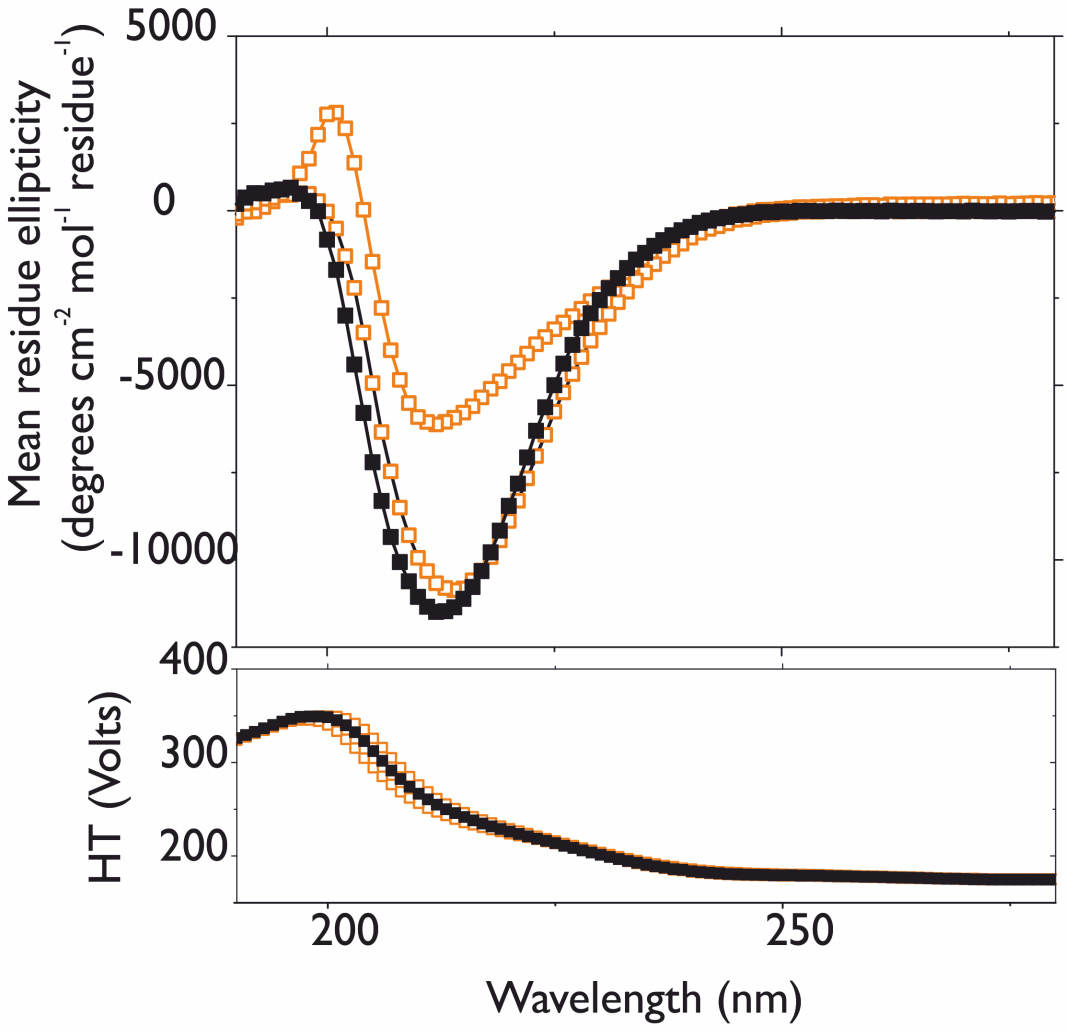


Figure 3.2.3.5: CD spectra of the $p(pL)_5$ in H_2O before and after thermal denaturation. The spectra before denaturation at 23 °C is shown as open organge squares connected by a solid orange line. The spectra of the fully thermally unfolded protein is shown in as the closed black squares connected by the solid black line. Finally the CD spectra of the protein after cooling to 23 °C is shown as the orange squares connected by a black line. The spectra indicate that the protein does not refold upon cooling.

Example spectra before and after thermal denaturation of $(pL)_5$ are given in 3.2.3.5.

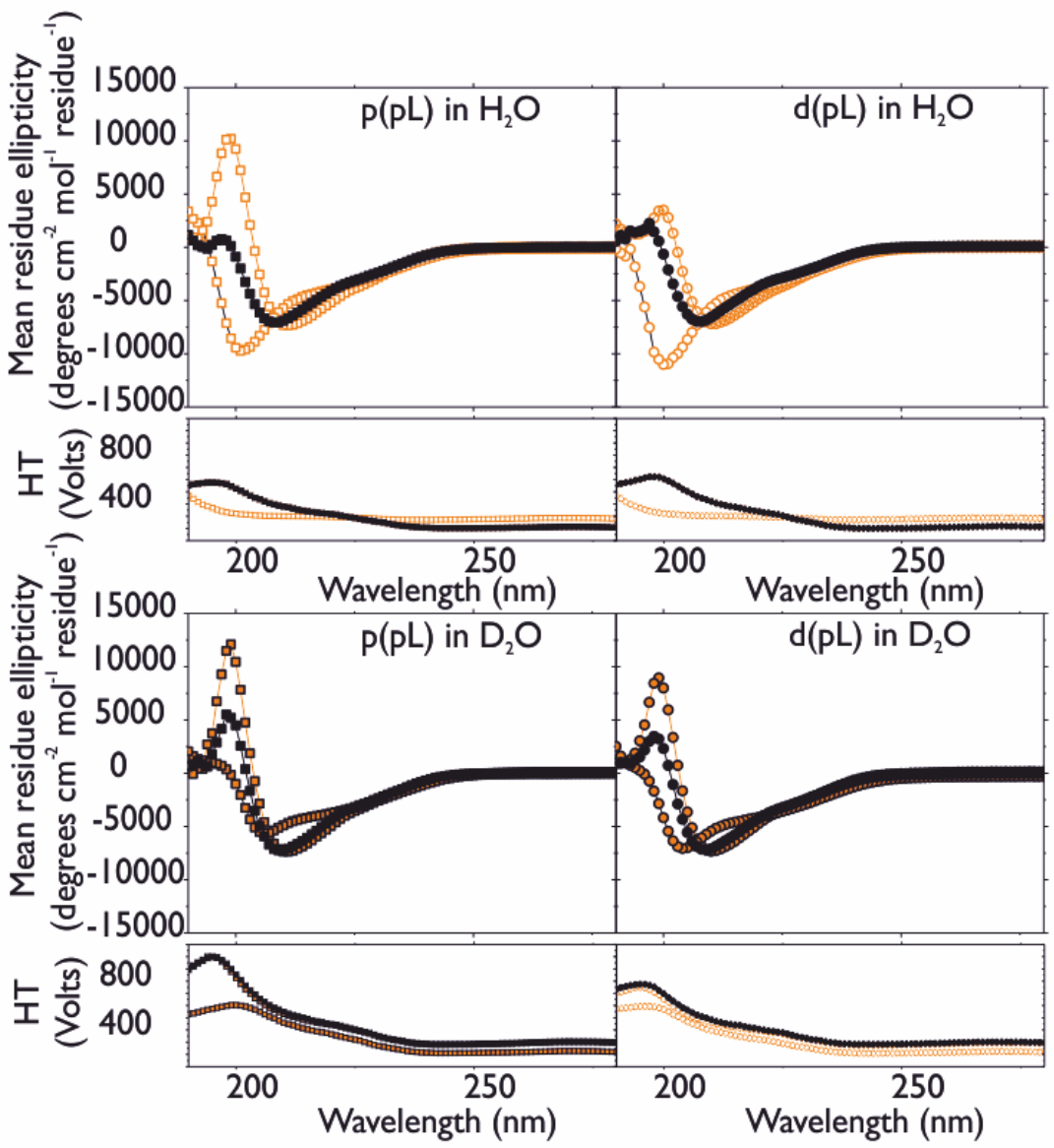


Figure 3.2.3.6: CD spectra of the pL monomer systems before (25 °C) and after heating to 90 °C. The open orange squares are the p(pL) in H_2O, the open orange circles are the d(pL) in H_2O, the closed orange squares are the p(pL) in D_2O and the d(pL) in D_2O are the closed orange circles connected with an orange line are the proteins before thermal unfolding. The same markers connected by solid black lines are the thermally unfolded spectra of the protein and the black closed markers connected by the black line show the respective proteins after cooling to 25 °C. The spectra show the protein regains the secondary structure upon cooling.

These spectra were the same for the isotopic protein/solution systems. This shows that the protein does not refold upon cooling to 23 °C. It has previously been shown that protein refolding is less likely to occur if the neighbouring protein has greater than 42 % sequence identity [370]. By contrast the spectra before and after heating for the pL monomers are shown in Figure 3.2.3.6. The spectra show that the majority of structure is reformed upon cooling to 25 °C.

The thermal denaturation of the polyprotein systems is discussed in Section 3.2.3.3 and the monomers is discussed in Section 3.2.3.4.

3.2.3.3 Thermal denaturation of polyproteins

The thermal denaturation curves for the four $(I27)_5$ and $(pL)_5$ test systems were fitted as described in Section 2.3.2 using Equation 2.7. The denaturation curves for the $(I27)_5$ protein systems are given in Figure 3.2.3.7. The values from the fits to the data for the four $(I27)_5$ test systems are given in Table 3.2.3.1. There is not much differentiation between the melting temperatures of the systems. The least thermally stable is the $d(I27)_5$ in H_2O and the most stable are the $(I27)_5$ systems in D_2O.

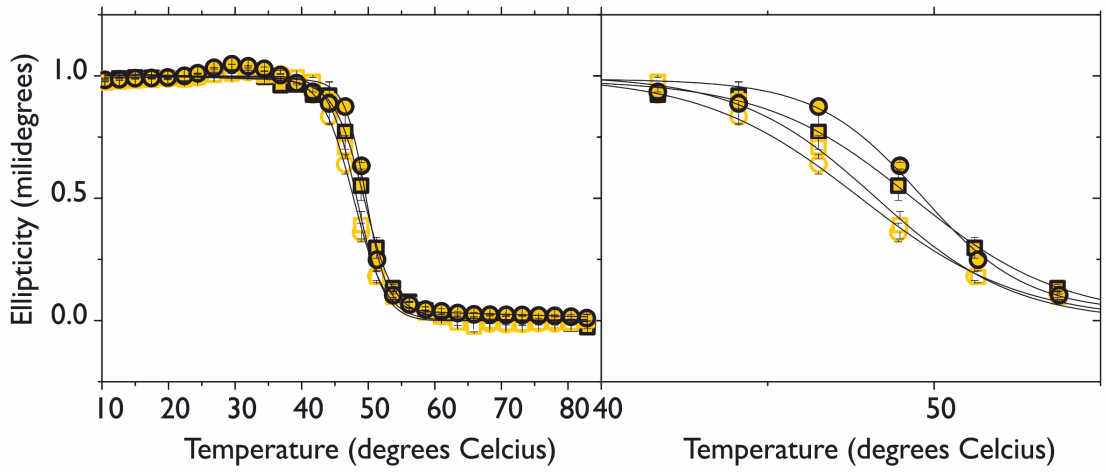


Figure 3.2.3.7: Thermal denaturation spectra for the $(I27)_5$ systems (left) and a close up on the transition (right) followed at 222 nm. Open squares: $p(I27)_5$ in H_2O, open circles: $d(I27)_5$ in H_2O, closed squares: $p(I27)_5$ in D_2O and closed circles:$d(I27)_5$ in D_2O. Solid lines are fits to Equation 2.6. All spectra have a single transition from the folded to unfolded states suggesting two-state unfolding. There is not a significant difference between the melting temperatures of the systems. The error bars are the standard error from two repeats of the thermal denaturation.

The T_m for the $(pL)_5$ are given in Table 3.2.3.2. These were determined by fits to the spectra given in Figure 3.2.3.8. There is not a significant difference between the melting temperatures for each of the systems. The least stable is the $(pL)_5$ systems in H_2O and the most stable is the $(pL)_5$ systems in D_2O.

Protein system	Wavelength (nm)	T_m ± error (°C)
p$(I27)_5$ in H_2O	222	48.3 ± 0.5
p$(I27)_5$ in D_2O	222	49.3 ± 0.7
d$(I27)_5$ in H_2O	222	47.8 ± 0.4
d$(I27)_5$ in D_2O	222	49.6 ± 0.1

Table 3.2.3.1: Table of thermal melting temperatures for each of the $(I27)_5$ systems. Thermal denaturation was followed at 222 nm and the error stated is the standard error from two repeats.

Protein system	Wavelength (nm)	T_m ± error (°C)
p$(pL)_5$ in H_2O	216	56.5 ± 0.4
p$(pL)_5$ in D_2O	216	57.5 ± 0.1
d$(pL)_5$ in H_2O	216	56.5 ± 0.5
d$(pL)_5$ in D_2O	216	57.4 ± 0.4

Table 3.2.3.2: Table of thermal melting temperatures for each of the $(pL)_5$ test systems. Thermal denaturation was followed at 216 nm and the error stated is the standard error from two repeats.

There is a greater difference in the T_m for the I27 systems compared to the pL systems.

3.2.3.4 Thermal denaturation of monomers

The thermal melting of the pL and I27 monomers in all four test systems was reversible. Therefore the equations outlined in Section 3.2.3 were fit to the data in Origin Pro using the method outlined in reference [297]. The thermal denaturation curves were fitted by Equations 3.2–3.5 allowing ΔC_p to be estimated in the fitting. The parameters obtained from these fitting procedures are given in Table 3.2.3.3 and Table 3.2.3.4 for protein I27 and pL respectively. This fitting procedure enabled ΔH and ΔC_p to be obtained and therefore enabled $\Delta G_{f,temp}$ to be determined at every temperature. The gradient from the straight line graph from a plot of $-\Delta G_{f,temp}$ against the temperature, T, therefore enables ΔS to be determined using 1.3 (Gibbs-Helmholtz equation). Additionally, the T_m was determined by a single transition sigmoidal fit (see Section 2.3.2, Equation 2.7) for comparison. The thermal denaturation of I27 was followed at 230 nm and 222 nm. The thermal denaturation of pL was followed at 222 nm. The denatuation curves for I27 and pL are shown in Figures 3.2.3.9 and 3.2.3.10 respectively.

The thermal denaturation of the four I27 monomer test systems indicates larger differences between T_m of the four test systems compared with the T_m obtained for of the four $(I27)_5$ test systems. The melting temperature of the d(I27) in H_2O is $\sim$ 4 °C

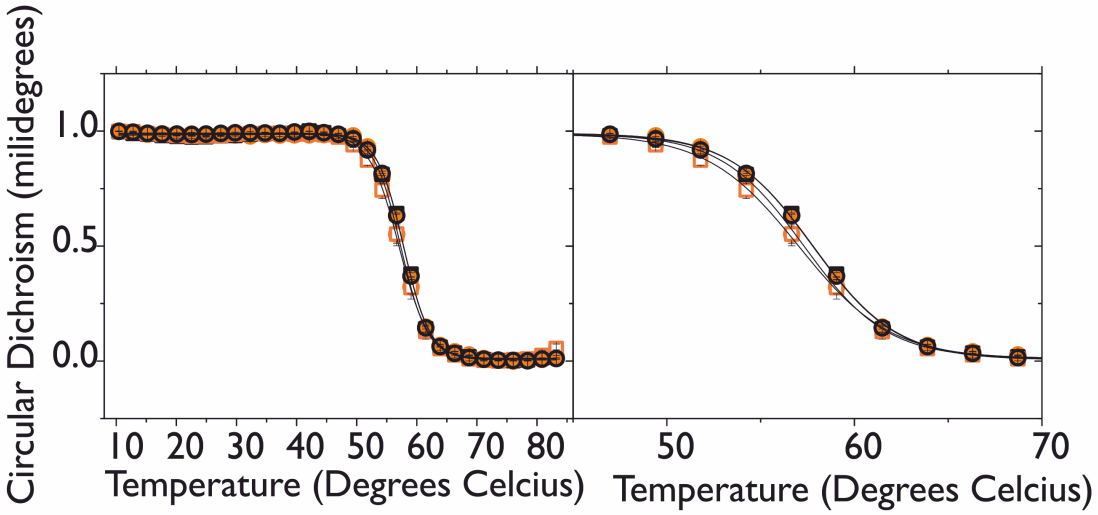


Figure 3.2.3.8: Thermal denaturation spectra for the $(pL)_5$ systems (left) and a close up on the transition (right) followed at 216 nm. Open squares: $p(pL)_5$ in H_2O, open circles: $d(pL)_5$ in H_2O, closed squares: $p(pL)_5$ in D_2O and closed circles:$d(pL)_5$ in D_2O. Solid lines are fits to Equation 2.6. All spectra have a single transition from the folded to unfolded states suggesting two-state unfolding. There is not a significant difference between the melting temperatures of the systems. The error bars are the standard error from two repeats of the thermal denaturation.

Parameters	p(I27) in H_2O	p(I27) in D_2O	d(I27) in H_2O	d(I27) in D_2O
Wavelength 222 nm				
(a)				
T_m (°C)	51.40 ± 0.02	57.89 ± 0.11	47.75 ± 0.10	55.90 ± 0.39
(b)				
T_m (°C)	51.40 ± 0.02	57.9 ± 0.1	47.8 ± 0.1	55.9 ± 0.4
ΔH ($kJmol^{-1}$)	273.3 ± 0.8	319.4 ± 0.4	245.7 ± 9.2	298.0 ± 5.7
ΔS ($Jmol^{-1}K_{-1}$)	823.8 ± 2.9	905.6 ± 0.01	766.0 ± 28.0	860.6 ± 13.8
ΔC_p ($kJmol^{-1}K^{-1}$)	1.68 ± 0.02	1.92 ± 0.01	1.53 ± 0.06	1.81 ± 0.03
Wavelength 230 nm				
(a)				
T_m (°C)	51.18 ± 0.04	57.67 ± 0.10	47.61 ± 0.10	55.67 ± 0.33
(b)				
T_m (°C)	51.18 ± 0.04	57.7 ± 0.1	47.6 ± 0.1	55.8 ± 0.3
ΔH ($kJmol^{-1}$)	272.5 ± 0.4	299.3 ± 1.3	246.5 ± 8.8	290.5 ± 6.7
ΔS ($Jmol^{-1}K_{-1}$)	822.5 ± 0.4	849.7 ± 2.1	769.4 ± 27.3	839.3 ± 15.1
ΔC_p ($kJmol^{-1}K^{-1}$)	1.71 ± 0.01	1.80 ± 0.01	1.54 ± 0.05	1.76 ± 0.04

Table 3.2.3.3: Parameters governing the thermal denturation of the I27 protein test systems. (a) The thermal denaturation curves were fitted with Equation 2.7, as outlined in Section 2.3.2, to obtain T_m. (b) The thermal denaturation curves were fitted using equations 3.2–3.5. The ΔC_p from this fitting procedure will not be accurate but can give an indication of the true heat capacity. The errors stated are the standard error from two independent unfolding curves.

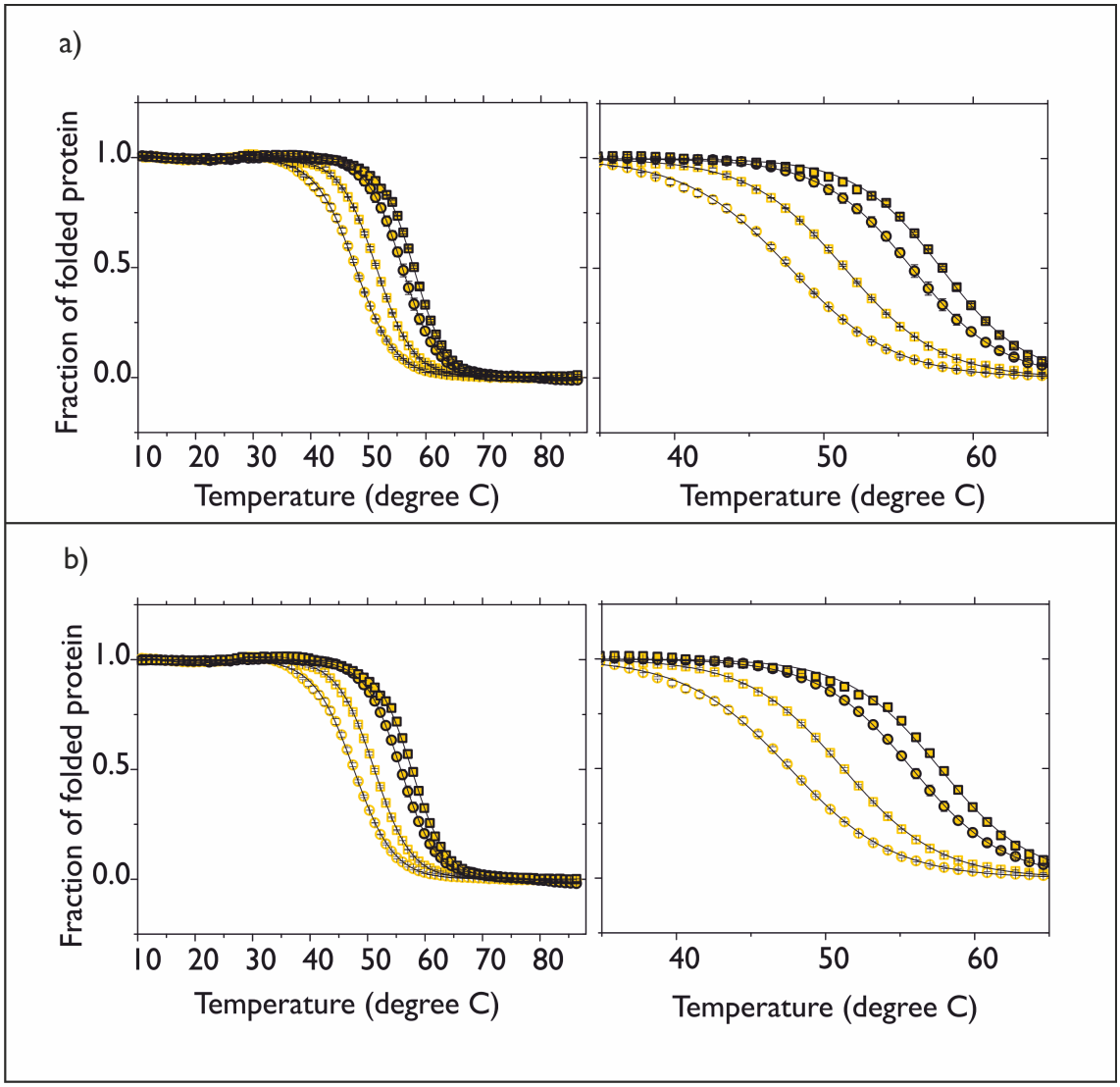


Figure 3.2.3.9: Thermal denaturation spectra for the I27 systems (left) and the transition (right) followed at 222 nm (a) and 230 nm (b). Open squares: p(I27) in H_2O, open circles: d(I27) in H_2O, closed squares: p(I27) in D_2O and closed circles:d(I27) in D_2O. The differences between the curves are significant suggesting a change in the melting temperature of the protein. All spectra have a single transition from the folded to unfolded states suggesting two-state unfolding. Solid lines are fits to Equation 2.6. The error bars are the standard error from two repeats of the thermal denaturation.

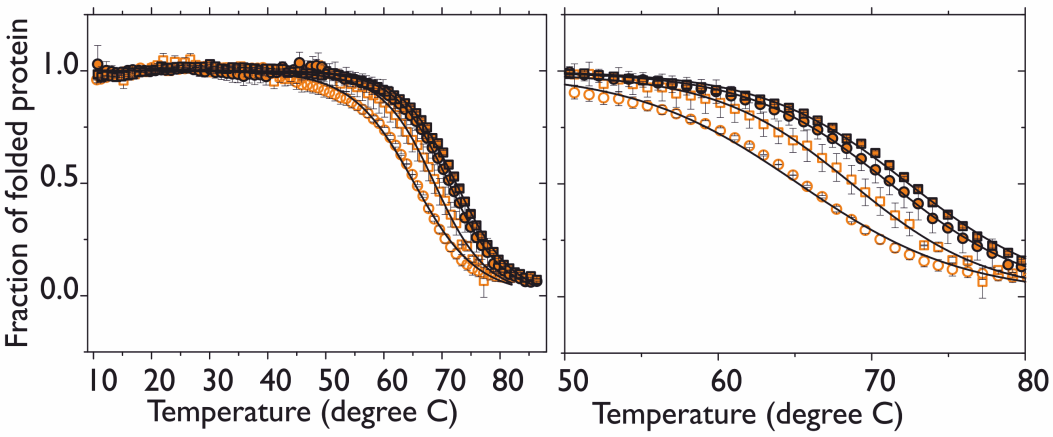


Figure 3.2.3.10: Thermal denaturation spectra for the pL systems (left) and the transition (right) followed at 216 nm. Open squares: p(pL) in H_2O, open circles: d(pL) in H_2O, closed squares: p(pL) in D_2O and closed circles:d(pL) in D_2O. There are significant differences between the melting temperatures of the systems. All spectra have a single transition from the folded to unfolded states suggesting two-state unfolding. Solid lines are fits to Equation 2.6. The error bars are the standard error from two repeats of the thermal denaturation.

	System			
Parameters	p(pL) in H_2O	p(pL) in D_2O	d(pL) in H_2O	d(pL) in D_2O
Wavelength 216 nm				
(a)				
T_m (°C)	68.5 ± 1.0	72.4 ± 0.1	65.24 ± 0.02	71.2 ± 1.3
(b)				
T_m (°C)	68.1 ± 1.2	71.4 ± 1.9	64.6 ± 0.3	72.1 ± 0.2
ΔH ($kJmol^{-1}$)	218.5 ± 38.9	213.9 ± 0.8	174.6 ± 23.4	196.7 ± 6.7
ΔS ($kJmol^{-1}K_{-1}$)	-	-	-	-
ΔC_p ($kJmol^{-1}K^{-1}$)	6.11 ± 1.42	5.37 ± 0.01	4.77 ± 0.71	5.44 ± 0.17

Table 3.2.3.4: Parameters governing the thermal denturation of the pL protein test systems. The parameters were obtained from three different fitting methods (a) and (b). (a) The thermal denaturation curves were fitted with Equation 2.7, as outlined in Section 2.3.2, to obtain T_m. (b) The thermal denaturation curves were fitted using equations 3.2–3.5. The ΔC_p from this fitting procedure will not be accurate but can give an indication of the true heat capacity. The errors stated are the standard error from two independent unfolding curves. The errors stated are the standard error from two independent unfolding curves.

lower than the p(I27) in H_2O and is 8 °C lower than the p(I27) in D_2O. The decrease in melting temperature for the d(I27) in H_2O was also coupled with a decrease in the change in enthalpy and entropy of unfolding. This suggests that the interactions are weaker in the d(I27) in H_2O system, and potentially that the hydrophobic interaction is less prominent (hydrophobic interactions are driven by entropy see Section 1.4.3). The largest change in enthalpy upon unfolding is observed for the I27 proteins in D_2O. This could be attributed to the stronger hydrogen bonds in the solvent. There are some discrepancies between the change in enthalpy determined for the p(I27) in D_2O for the two wavelengths monitored during unfolding. The entropy of the system was determined from the slope of the negative $\Delta G_{f,chem}$ (calculated using Equation 3.5) against temperature. There was a slight curvature in the plots, this graph is shown in the Appendix B. The two I27 systems in D_2O also appear to have a larger change of entropy upon folding. The largest change in the entropy of unfolding is observed for the p(I27) in D_2O. The heat capacities have also been estimated. There is not enough information from a single unfolding curve in one condition [300] to determine an accurate value for the change in heat capacity and therefore they will not be discussed in this project. More accurate values could be obtained from global fitting of thermal denaturation curves in different concentrations of chemical denaturant.

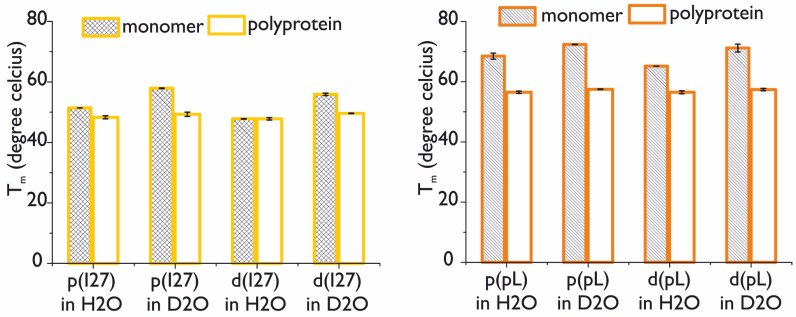


Figure 3.2.3.11: Bar charts showing the thermal melting temperatures, T_m, for the I27 (left, yellow) and pL (right, yellow) four test systems. The patterned bars are the T_m for the protein monomer systems and the plain bars are the T_m for the polyprotein systems.

Table 3.2.3.4 gives the parameters obtained from fitting to the thermal denaturation curves for the four pL monomer test systems in Figure 3.2.3.10. There is only a short post-transitional baseline in the pL unfolding curves due to the heating limits of the instrument and the water buffer. The instrument has a maximum temperature of 90 °C, after this temperature the temperature probe can be damaged. Furthermore around this temperature, bubbles and condensation can build in the samples. This therefore will lead to large uncertainties in the fitting parameters of the data. Despite this similar trends of

T_m were observed between the I27 and pL test systems. These results are summarised in Figure 3.2.3.11. For the pL systems, the d(pL) in H_2O was again found to be the least thermally stable. The melting temperature was $\sim$ 3 °C lower than the p(pL) in H_2O and $\sim$ 7 °C lower than p(pL) in D_2O. These differences are similar to those observed for I27 (4 °C and 8 °C respectively). Despite both systems in D_2O being more thermally stable, the enthalpy obtained from fitting regime (b) suggested the change in enthalpy was negligible. However, ΔS for the protein L could not be accurately obtained from the Gibbs free energy (Equation 1.3) due to curvature in a plot of calculated $\Delta G_{f,temp}$ against ΔH (figure shown in Appendix B). This could suggest that the lack of post transitional baseline causes large errors in the calculated thermodynamic parameters that are not accounted for from the standard error.

Another result to note is that the polyproteins of pL are significantly destabilised compared to the monomers. There is an increase in T_m of $\sim$ 10 °C for the monomer test systems compared with the polymeric constructs. The differences between the T_ms obtained for the monomeric and polymeric I27 proteins systems are less significant. This has been observed in a study using chemical denaturation to compare the stability of the monomer with polyproteins [373]. These studies found that, whilst the I27 polyprotein had a similar to the monomer, fibronectrin domains showed a significant increase in their stability within a polyprotein domain. This study suggested that the domains in the polyprotein were stabilised by inter-domain electrosatic interactions [373]. It appears as though unfavourable interactions are occuring between the pL monomers within the polyprotein.

3.2.4 Chemical denaturation of proteins

In order to explore the importance of the primary and secondary isotope effect on the stability of the folded protein state in these protein systems, equilibrium denaturation was carried out. Kinetic unfolding experiments were also performed to examine the effect of isotope substitution the folding and unfolding rates in the I27 and pL test systems. These experiments were performed on the monomers only. It has previously been shown that chemical denaturation experiments yield similar results for the poly- and monomeric proteins [217]. Therefore it is assumed that the free energy differences and rates of unfolding and folding are the same for the monomer and polyproteins.

3.2.4.1 Equilibrium denaturation

The equilibrium of protein L was carried out in GdnHCl in similar conditions to those described by Scalley et al. [357]. Folded and unfolding spectra of the pL and I27 test systems can be found in the Appendix B. The equilibrium of I27 was carried out in urea using similar conditions as previous studies on this Ig domain [214, 367]. The data obtained from following the change in intrinsic tryptophan fluorescence followed at a single wavelength of 320 nm and 328 nm are shown in Figures 3.2.4.1 and 3.2.4.2 for I27 and pL respectively. The spectra were taken at 25 °C to allow direct comparison of the p(protein) in H_2O from previous studies [214, 357, 367, 368].

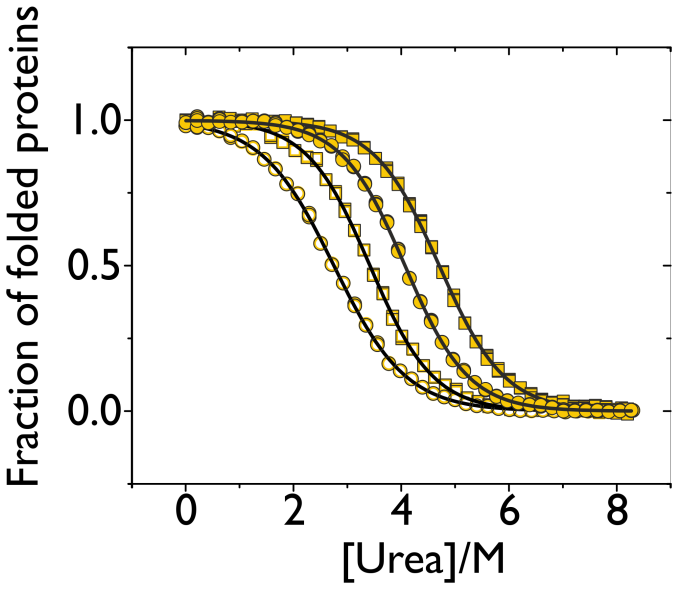


Figure 3.2.4.1: Normalised chemical denaturation spectra for the I27 systems. Denaturation of the protein using urea was followed by intrinsic tryptophan fluorescence excited at 280 nm and followed at 320 nm. Open squares: p(I27) in H_2O, open circles: d(I27) in H_2O, closed squares: p(I27) in D_2O and closed circles:d(I27) in D_2O. The differences between the curves are significant suggesting that $\Delta G_{u,chem}$ has been affected in all systems. All spectra have a single transition from the folded to unfolded states suggesting two-state unfolding. Solid lines are fits to Equation 2.11.

Fits assuming two-state unfolding were used to extract the free energy of unfolding, $\Delta G_{u,chem}$, and the dependence of the free energy of unfolding on denaturant concentration, m. The parameters determined are given in Table 3.2.4.1 for both proteins. The trend in the results are similar as the thermal denaturation. The d(protein) in H_2O is the least stable of all of the systems for both pL and I27. The most stable system in the p(protein) in D_2O for both pL and I27. There are varied results for the d(protein) in D_2O. Whilst the d(I27) in D_2O is significantly more stable than the p(I27) in H_2O, the d(pL) in D_2O does not significantly differ from the p(pL) in H_2O. The value of $\Delta G_{u,chem}$ obtained for

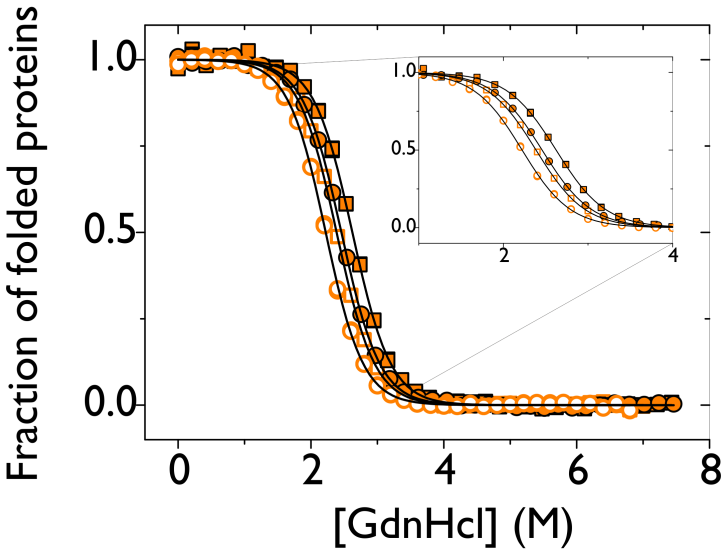


Figure 3.2.4.2: Normalised chemical denaturation spectra for the pL systems. Denaturation of the protein using GdnHcl was followed by intrinsic tryptophan fluorescence excited at 280 nm and followed at 328 nm. Open squares: p(pL) in H_2O, open circles: d(pL) in H_2O, closed squares: p(pL) in D_2O and closed circles:d(pL) in D_2O. The differences between the curves are significant suggesting that $\Delta G_{u,chem}$ has been affected in all systems. All spectra have a single transition from the folded to unfolded states suggesting two-state unfolding. Solid lines are fits to Equation 2.11. The right hand figure allows clear differentiation between the unfolding curves of the four different pL protein systems.

the p(I27) in H_2O of 20.6 ± 0.5 kJmol^{-1} agrees with the previously determined value of 19.3 ± 0.8 kJmol^{-1}. The m value of 8.6 ± 0.2 kJmol^{-1}M^{-1} is also close to the previously determined value of 8.0 ± 0.3 kJmol^{-1}M^{-1}.

Protein system	$\Delta G_{u,chem}$ (kJmol^{-1})	m (kJmol^{-1}M^{-1})
p(I27) in H_2O	14.5 ± 0.4	4.2 ± 0.1
p(I27) in D_2O	19.3 ± 0.3	4.1 ± 0.1
d(I27) in H_2O	10.2 ± 0.3	3.7 ± 0.1
d(I27) in D_2O	17.1 ± 0.2	4.2 ± 0.1
p(pL) in H_2O	20.6 ± 0.5	8.6 ± 0.2
p(pL) in D_2O	22.4 ± 0.6	8.5 ± 0.2
d(pL) in H_2O	19.5 ± 0.4	8.8 ± 0.2
d(pL) in D_2O	21.0 ± 0.5	8.6 ± 0.2

Table 3.2.4.1: Parameters found from fitting equilibrium denaturation curves with a two state model for the pL and I27 test systems. The stated errors are determined from the standard error of the least linear squares fit.

The m value has been found to correlate with the area of protein exposed to the solvent (solvent accessible surface area or SASA upon unfolding) [374]. The m values determined for the proteins in these experiments indicate no change in the SASA of the proteins from deuteration or change in solvent for all proteins. However, there is a slight decrease in the

m value for the d(I27) in H_2O. This may be attributed to the lack of a pre-transitional baseline in protein denaturation curve.

3.2.4.2 The effect of the backbone deuteration on the kinetics of protein folding

To determine whether the difference in the m value of the d(I27) in H_2O is due to the lack of a pre-transitional baseline, it can be compared to the total m value m_{tot}, determined from kinetic unfolding experiments. The m_{tot} can be determined from fitting Equation 2.13 to a chevron plot (see Section 2.3.3.5) where $m_{tot}=m_u+m_f$. Chevron plots were obtained for the p(protein) in H_2O and d(protein) in H_2O for both I27 and protein L (Figures 3.2.4.3 and 3.2.4.4 respectively). The results of these experiments are presented in Tables 3.2.4.2 and 3.2.4.3 for the I27 and pL systems respectively. The results obtained for the unfolding and refolding rate constants for p(I27) in H_2O, 3×10^{-3} s^{-1} and 1.8 s^{-1} respectively, are consistent with previous studies under similar conditions ($k^0_{u,chem}$=10.7 $\times$ 10^{-3} s^{-1} and $k^0_{f,chem}$=1.9 s^{-1})[214]. Similarly the results obtained for the p(pL) in H_2O ($k^0_{u,chem}$=0.05 s^{-1} and $k^0_{f,chem}$=76 s^{-1}) agree with previous studies ($k^0_{u,chem}$=0.02 s^{-1} and $k^0_{f,chem}$=61 s^{-1}) [357].

These results indicated that the rate of folding of the d(protein) in H_2O is significantly reduced for both I27 and pL proteins. Despite the large differences in the folding rates, only small differences were observed in the folding barrier heights. Diagrams of the energy landscapes for the I27 and pL systems can be found in the Appendix B. The rates of unfolding of the two protein systems tested (p(protein) in H_2O and d(protein) in H_2O) are similar for both protein systems. The values for $\Delta G^0_{u,chem}$, obtained from the rate constants of unfolding and refolding at 0 M denaturant, for both I27 and pL proteins, are comparable with those obtained from the equilibrium unfolding experiments except for the d(protein) in H_2O pL and I27 systems. The values calculated for these test systems might be higher in these experiments because they do not depend on a pre-transitional baseline. Furthermore, the m_{tot} are comparable with those obtained for all systems in the equilibrium experiments. This suggests the m value for the d(I27) in H_2O from equilibrium is not accurate, this would lead to an underestimation in the $\Delta G_{u,chem}$.

A slight curvature is observed in the unfolding branch on the chevron plots for the pL protein systems. This can often suggest the presence of stable intermediates [216] however, this has not been observed previously for this protein under similar conditions [367].

Parameters	p(I27) in H_2O	d(I27) in H_2O
$k^0_{u,chem}$ (s^{-1})	$2.92 \pm 0.41 \times 10^{-3}$	$3.67 \pm 0.44 \times 10^{-3}$
m_u ($kJmol^{-1}M^{-1}$)	0.28 ± 0.06	0.26 ± 0.05
$k^0_{f,chem}$ (s^{-1})	1.79 ± 0.09	0.78 ± 0.05
m_f ($kJmol^{-1}M^{-1}$)	4.41 ± 0.08	4.44 ± 0.13
$\Delta G_{u,chem}$ ($kJmol^{-1}$)	15.9 ± 3.2	13.3 ± 1.3
m_{tot} ($kJmol^{-1}M^{-1}$)	4.69 ± 0.10	4.70 ± 0.14

Table 3.2.4.2: Table of the values obtained from the fit of a two-state model to the denaturant dependence on the observed rate for the the: p(I27) in H_2O and d(I27) in H_2O test systems. The standard error is determined from standard error from the fit to the curves.

Parameters	p(pL) in H_2O	d(pL) in H_2O
$k^0_{u,chem}$ (s^{-1})	$50.9 \pm 3.1 \times 10^{-3}$	$65.2 \pm 3.5 \times 10^{-3}$
m_u ($kJmol^{-1}M^{-1}$)	2.13 ± 0.04	1.98 ± 0.04
$k^0_{f,chem}$ (s^{-1})	76.1 ± 10.1	39.4 ± 4.7
m_f ($kJmol^{-1}M^{-1}$)	6.95 ± 0.22	6.68 ± 0.21
$\Delta G_{u,chem}$ ($kJmol^{-1}$)	18.1 ± 2.6	15.9 ± 2.1
m_{tot} ($kJmol^{-1}M^{-1}$)	9.08 ± 0.22	8.66 ± 0.21

Table 3.2.4.3: Table of the values obtained from the fit of a two-state model to the denaturant dependence on the observed rate for the: p(pL) in H_2O and d(pL) in H_2O test systems. The standard error is determined from standard error from the fit to the curves.

3.2.5 Discussion

The results presented here indicate that the deuterated polyprotein and monomers of I27 and pL can be expressed and purified with a good yield of protein. It was not necessary to adapt the bacteria to growth on a deuterated carbon source. The percentage deuteration was above 98 % which is consistent with deuteration of other proteins [130]. The proteins appear to be correctly folded within both polyprotein and monomeric constructs for all systems. Furthermore, thermal and chemical denaturation experiments on the proteins are consistent with results obtained for other proteins suggesting that the secondary isotope effect destabilises proteins whereas the primary isotope effect is mostly stabilising. An interesting result is the differences obtained between the T_m for the pL monomer test systems and polyprotein test systems. It has been suggested previously that the thermal and chemical stability of a protein is not affected by incorporation into a polyprotein [217, 255, 367]. However chemical denaturation studies using urea as a denaturant suggested that incorporation of I27 and TNfn3 into a homo polyprotein stabilises the protein fold [373] by decreasing the rate of unfolding.

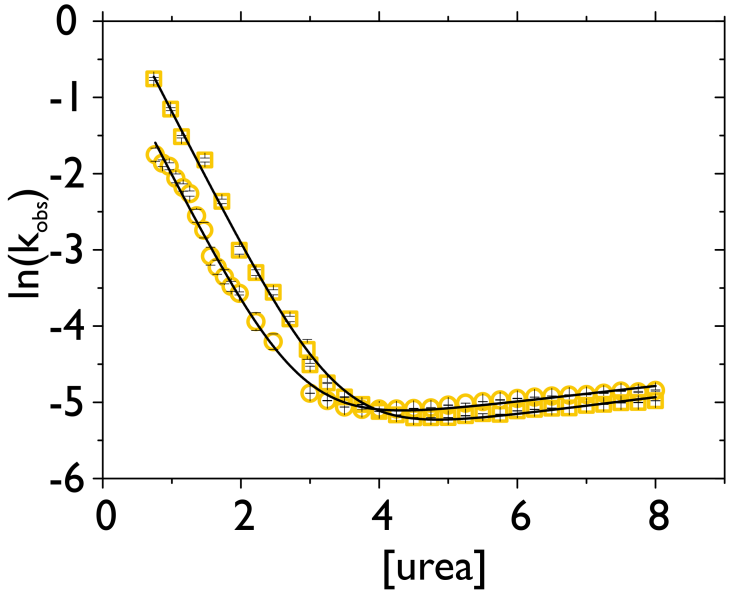


Figure 3.2.4.3: Denaturant dependence of the the natural logarithm of observed rate constant, $\ln(k_{obs})$. The folding branch of the chevron plot (left) was obtained by stopped-flow fluorescence. The unfolding branch of the chevron plot (right) was determined by manual mixing. The open squares are p(I27) in H_2O and the open circles are d(I27) in H_2O. The solid line fit of the kinetic data to a two-state model (Equation 2.13). There are no curves in the branches of the chevron plot which indicates no presence of intermediates. The error bars are calculated as the standard error from the average rate determined from exponential fits to the unfolding or refolding data.

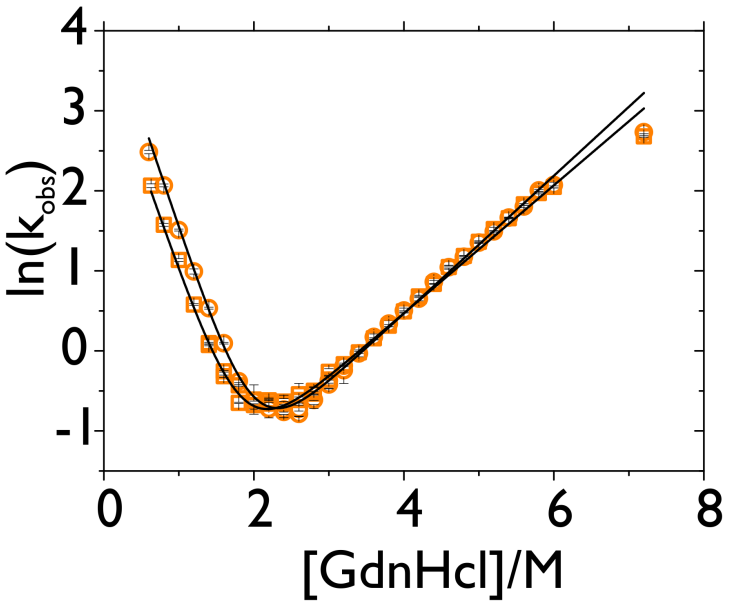


Figure 3.2.4.4: Denaturant dependence of the the natural logarithm of observed rate constant, $\ln(k_{obs})$. The folding branch of the chevron plot (left) was obtained by stopped-flow fluorescence. The unfolding branch of the chevron plot (right) was determined by manual mixing. The open squares are p(pL) in H_2O and the open circles are d(pL) in H_2O. The solid line fit of the kinetic data to a two-state model (Equation 2.13). There is a slight curve in the unfolding branch of the chevron plot which may indicate the presence of an intermediate for both proteins. However, this has not been observed previously for this protein and therefore could be attributed to the large jump in denaturant concentration to this value. The error bars are calculated as the standard error from the average rate determined from exponential fits to the unfolding or refolding data.

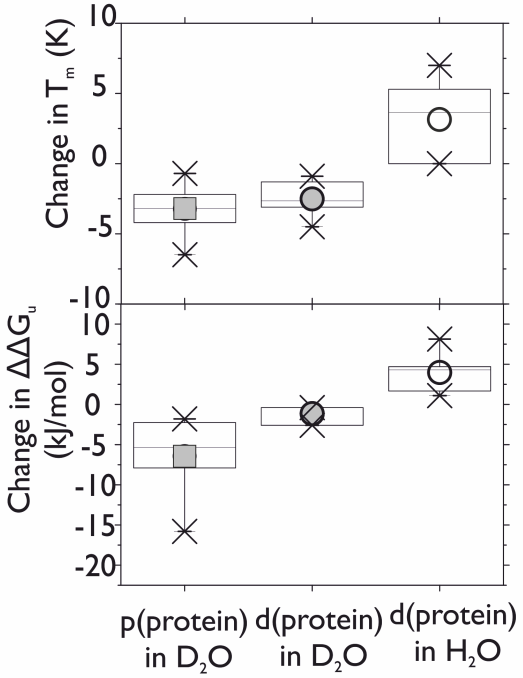


Figure 3.2.5.1: Top: Figure showing box-plots for change in melting temperature in comparison to the p(protein) in H_2O for different proteins. In total there were 6 d(protein) in D_2O studies (filled circle), 13 p(protein) in D_2O (filled square) studies and 10 studies on d(protein) in H_2O (open circle). The markers indicate the average and range of values for the proteins studied. A negative sign indicates an increase in the melting temperature for this system compared to the p(protein) in H_2O and an positive value means the T_m has decreased. The differences for the proteins in D_2O are similar - both systems are stabilised by similar amounts. There is a significant decrease in the T_m for the d(protein) in H_2O. Bottom: Graph showing the change in free energy compared with the p(protein) in H_2O. In total there were only 3 d(protein) in D_2O studies, 7 p(protein) in D_2O studies and 5 studies on d(protein) in H_2O. The markers indicate the average and range of values for the proteins studied. A negative $\Delta\Delta G$ indicates an increase in the barrier to unfolding, a positive value indicates the folded state is destabilised. Generally the p(protein) in D_2O are the most stable, the d(protein) in D_2O are slightly more stable than the p(protein) in H_2O and the d(protein) in H_2O is always >0 kJ/mol. The data were taken from this report and from references:[130, 137, 160–162, 170, 375, 376].

3.2.5.1 Effect of deuterium on the T_m of proteins

Figure 3.2.5.1 (top) shows a summary of some proteins that have been thermally denatured (including those from this study). It is apparent that the least thermally stable system is the d(protein) in H_2O, for both proteins. This is due to a shorter C-D bond resulting from reduced vibrational amplitudes of the bond [137]. This decreases the steric requirement compared with the C-H bond. The most stable are the p(protein) in D_2O and d(protein) in D_2O.

The results from this study are varied and are summarised in the bar charts shown in Figure 3.2.3.11. For the polyproteins systems there are only slight differences between the melting temperatures of the systems. This could be as a result of the presence of linkers between the proteins or due to interactions between the protein domains. For the $(I27)_5$ systems the most significant difference is between the d$(I27)_5$ in D_2O and the d$(I27)_5$ in H_2O (1.8 °C). There is also a difference (outside of error) between the d$(I27)_5$ in D_2O and the p$(I27)_5$ in H_2O (1.3 °C). There is not a significant difference for the other systems. This is surprising and could be attributed to the large error in the CD data. For the pL test systems there is a difference of 1 °C between the two least stable (p$(protein)_5$ in H_2O and d$(protein)_5$ in H_2O) and the two most stable (d$(pL)_5$ in D_2O and the p$(pL)_5$ in D_2O).

The T_m values observed for the I27 monomer test systems are comparable to those obtained in the polyprotein construct. However, there was significant differences between the T_m of the protein in these protein systems that were not observed in the I27 polyprotein systems. This could suggest that using CD to monitor the thermal melting of polyproteins may obscure differences between the melting temperatures. When a protein is unfolded in the polyprotein the spectra tends to that of a random coil with a significant minimum at 215 nm. This minimum is not observed in the monomer test systems. The differences between the polyprotein and monomer test systems are more pronounced in the pL test systems, there is an increase in 10 °C in the monomer melting temperatures. This may suggest that the protein is destabilised in a polyprotein construct or the spectra is dominated by random coil formation which masks the unfolding of the single protein domains.

The trend in T_m values obtained from thermal melting temperatures are consistent with the previous studies. For both proteins there is a decrease of 3-4 °C in the melting

temperature for the d(protein) in H_2O. There is also $\sim$ 3 °C increase for the d(protein) in D_2O compared to the p(protein) in H_2O. The largest increase in T_m is observed for the p(protein) in D_2O. This could suggest that this protein is stabilised by stronger hydrogen bonds in the protein and/or the solvent and enhanced hydrophobic contacts. For I27 this is further supported by an increase in the change in enthalpy for the protein. However, this was not observed for the pL protein but it is challenging to interpret the pL data due to the shorter post-translational baseline. The I27 data also suggests that this increase in enthalpy is coupled with an increase in entropy in D_2O. This change in entropy has been observed previously for proteins [137]. This result suggests that it is entropically favourable for the protein to be unfolded in D_2O. The isotopic substitution in the solvent is therefore thermodynamically stabilising the unfolded state. The folded state is stabilised by enhanced interactions within the protein. The smallest change in enthalpy and entropy is associated with the d(protein) in H_2O. The smaller change in enthalpy suggests that some Van der Waals interactions may be lost in the folded state. The smaller ΔS can be associated with the reduction in steric constraint in this protein [130]. The trends observed in the thermodynamic parameters for I27 are consistent between the two wavelengths followed. For pL the value for the ΔH fluctuates between the fitting methods. Furthermore a curvature is observed when trying to determine the change in entropy, this suggests that the model does not fit the data well. All interpretation of these data should be taken with caution due to the lack of information governing the heat capacity. An accurate heat capacity should be determined by performing multiple experiments at different pH or denaturant concentrations for accurate thermodynamic parameters [297].

3.2.5.2 Effect of deuterium on the free energy of unfolding

Figure 3.2.5.1 (bottom) also shows a summary of the $\Delta\Delta G_{u,chem}$ for some proteins that have been chemically denatured (including those from this study). It is apparent that the least stable are the d(protein) in H_2O. In this chapter, the decrease in stability was more pronounced for the d(I27) in H_2O ($\Delta\Delta G_{u,chem}$=3.8 kJ/mol) than the d(pL) in H_2O system ($\Delta\Delta G_{u,chem}$=0.9 kJ/mol). This could suggest that hydrophobic interactions are more important to the thermodynamical stability of I27 than pL. On the other hand, the $\Delta\Delta G_{u,chem}$, suggests that this difference could be as a consequence of the lack of pre-transitional baseline in the d(I27) in H_2O equilibrium curve. The differences between the two systems for the two proteins are less pronounced when considering free energies obtained from kinetic denaturation experiments. This result suggests that caution should

be taken if the pre- or post- transitional baselines of proteins are short. This could lead to errors not accounted for by the fit to the data.

Figure 3.2.5.1 also shows that generally proteins in D_2O are more stable than those in H_2O. This further suggests that hydrogen bonds stabilise the folded protein. A further increase is observed for the p(protein) in D_2O suggesting that the hydrophobicity is increased further supporting the idea that the stronger hydrogen bonds in the solvent leads proteins to form more compact state.

3.2.5.3 How does deuteration of the carbon backbone affect the kinetics of protein folding?

The results in this study indicate that the unfolding rate of a d(protein) in H_2O is similar to is similar to that of a protiated protein. The most significant difference was observed in the rate of folding. In both pL and I27 test systems the rate of folding was $\sim 2 \times$ slower in the deuterated proteins. Whilst studies have probed the unfolding and refolding rates of proteins in D_2O [170], few studies have probed the kinetics of deuterated proteins. The results presented here suggest that hydrophobic interactions may play an important role in protein folding. Furthermore it has been hypothesised that the hydrophobic effect might be the driving force for changes in protein unfolding [170, 175].There is only a small change between the unfolding rates of the pL and I27 deuterated and protiated proteins in H_2O systems. This suggests that the barrier to unfolding is the same but the barrier to folding is higher. This could be further supported by the increase in folding rate observed for an alternative study of a mutant p(I27) in D_2O using SMFS [132]. Comparing the results of the two proteins yields important information on the kinetics of protein folding. However, other differences may account for the differences observed such as differences of the hydration of the nonpolar regions of the protein, it is an important result for the understanding of the driving force behind protein folding. Exploring the rate of unfolding at a single molecular level could provide more insightful information about the interactions governing the protein fold.

Chapter 4

The effect of hydrogen bond strength on the mechanical stability of proteins

4.1 Introduction

Thermal and chemical denaturation suggest that deuteration of the solvent stabilises both protiated and deuterated proteins. However, deuteration of the protein destabilises the protein with respect to the protiated protein. To test the direct effect of deuteration on a specific region of the protein, SMFS experiments are performed on two model protein systems. To understand the role of the of the different non-covalent interactions, hydrogen bonds and hydrophobic interactions, it is useful to have information about the region involved in the rate limiting step of unfolding. In SMFS, the region, that primarily defines the transition state, is the mechanical clamp region of the two proteins being studied in this project; protein I27 and protein L. The mechanical clamps of each of the proteins will initially be discussed in this chapter. The chapter will then go on to discuss how deuterium effects mechanical stability of proteins in the four protein L and I27 test systems.

4.1.1 The mechanical clamp of protein L

.

Protein L (pL) was first mechanically unfolded by Brockwell et al. (2005) [217]. The protein was chosen due to its simple topology and hydrogen bonded parallel β-strands. Simulations have shown that the protein has a simple unfolding pathway. In constant velocity simulations, there is a steep force response to the extension of its termini followed by a sharp decrease in the force as the protein unfolds. The unfolding occurs after an

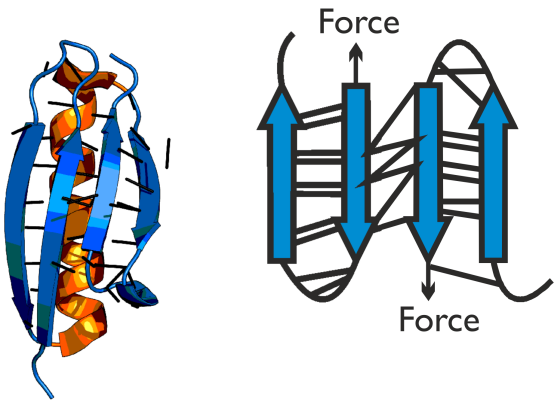


Figure 4.1.1.1: 3D structure of pL (protein L, PDB code 1HZ6) indicating the position of the mechanical clamp of the protein (blue). The clamp consists of a shear-supported clamp motif formed by the four hydrogen bonded β-strands. The hydrogen bonds are illustrated by the black lines. Contacts between this sheet and the α-helix also contribute to the mechanical unfolding as identified by mutational studies.

initial extension of 0.11 nm as a single, sharp transition suggests the protein is brittle [217]. Constant force simulations, where the -N and -C termini are held under constant tension, indicate that the protein occupies a metastable state that is similar to the native state of the folded protein. The only structural rearrangement in this protein, before the main unfolding event, is the rearrangement of the termini in the direction of the force. This disrupts several contacts between the N- terminal, hydrophobic residues and residues in a β-strand. The main unfolding event involves the breaking of contacts between the N- and C- termini β-stands, labelled I and IV in Figure 4.1.1.1. The transitional state of the protein is highly reproducible and therefore is well defined [217]. The identified mechanical clamp of pL is shown in Figure 4.1.1.1 [247], it is a shear supported clamp where two shearing β-strands are sandwiched between additional β-strands.

Mutational studies have also identified a residue isoleucine (I60) that acts a force 'rheostat' in the protein core and mechanically stabilises the protein [124]. Truncating this side chain results in a decrease in the mechanical stability. This study illustrated the importance of hydrophobic contacts to the mechanical clamp region of this protein. Truncation of side chains only has been shown to only have an effect on the force if the residues were part of the force bearing region [124]. Therefore, for protein L, hydrophobic contacts have been shown to be important for the proteins mechanical stability.

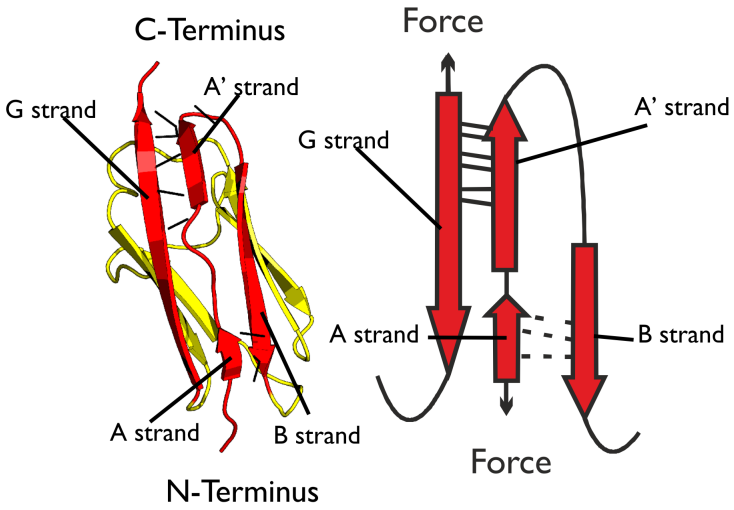


Figure 4.1.2.1: Figure of the 3D structure of I27 (PDB 1TIT) indicating the position of the mechanical clamp region (red) formed by six hydrogen bonds (black lines) between the A' and G strands and hydrophobic contacts with amino acid side-chains. An initial unfolding event is often observed on the force-extension traces of the protein as a deviation from the WLC on the unfolding peak. This is due to the breaking of two to three hydrogen bonds (dashed lines) between the A and B strands.

4.1.2 The mechanical clamp of protein I27

The existence of a mechanical clamp in I27 was first identified using simulations by Lu et al. in 1998 [243]. The characterisation of the force bearing region was experimentally verified by SMFS experiments on the protein. This was achieved by performing mutations to the region of amino acids involved in the clamp region identified by simulations [244]. The clamp consists of 6 hydrogen bonds between the A and G β-strands shown in Figure 4.1.2.1, that are sheared apart when force is applied between the -N and -C termini [218]. Further mutational and computational studies have also highlighted the importance of hydrophobic side-chain interactions between the A and G-stands and the A'-B and E-F loops in the mechanical stability of the protein [263]. These results have also highlighted that contacts between residues in the hydrophobic core are unperturbed at the transition state of the protein and therefore do not play a significant role in the mechanical stability of the protein [263]. Figure 4.1.2.2 highlights the important regions within the I27 protein for mechanical, kinetic and thermodynamic stability. Additional simulations on the w.t. I27 protein also revealed that the breaking of hydrogen bonds in the mechanical clamp was mediated by interactions with the water [377]. These water interactions are found to repeatedly interact with the hydrogen bonds and weaken them.

The mutant I27 protein used in this project maintains the hydrogen bonds in the clamp region of the protein. However, contacts between C47, found in the D strand, and amino acids in the D and E strands are assumed to be disrupted [214]. Furthermore,

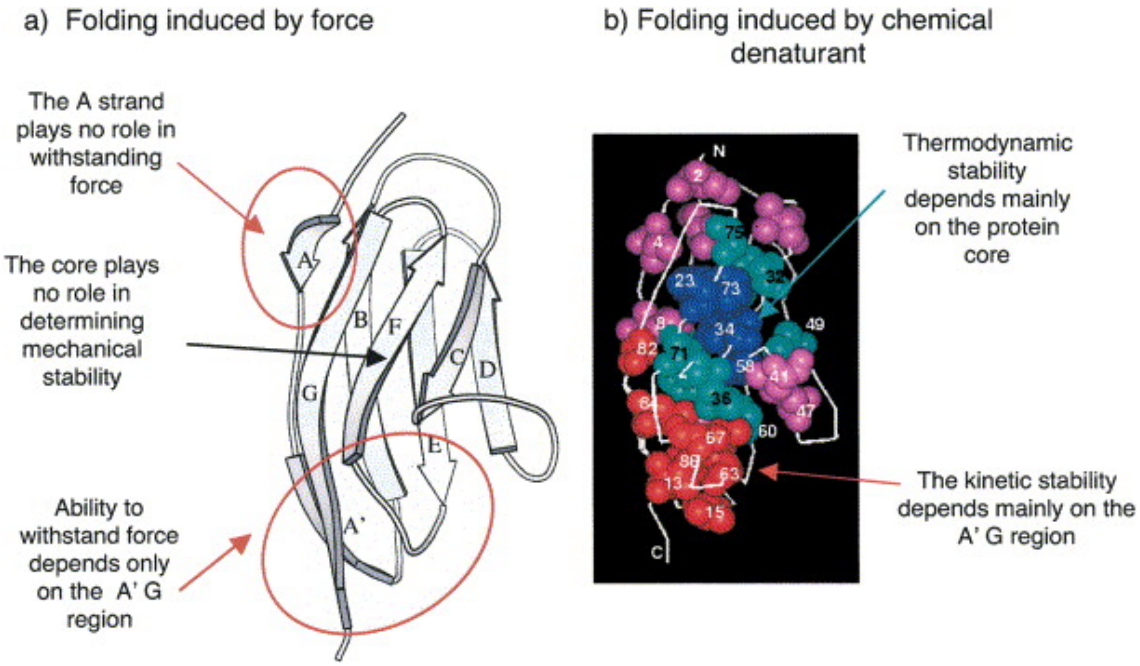


Figure 4.1.2.2: Figure illustrating the regions that are important for the mechanical (left) and thermodynamic (right) stability of protein I27. The figure on the right is coloured according to the Φ analysis determined by chemical denaturation experiments on mutated proteins. Mutations that resulted in low Φ values (~ 0) are coloured red and mutations that had high Φ values (>0.7) are coloured blue. Reprinted from Journal of Molecular Biology, 330, Robert B. Best, Susan B. Fowler, Jos L. Toca Herrera, Annette Steward, Emanuele Paci, Jane Clarke, Mechanical Unfolding of a Titin Ig Domain: Structure of Transition State Revealed by Combining Atomic Force Microscopy, Protein Engineering and Molecular Dynamics Simulations, 867–877, Copyright (2003), with permission from Elsevier. Reference number [263].

the residue C63, in the loop region between D and E, makes contacts with residues V13, found in the A strand, residues in the E strand and L84, V86 in the G strand. A mutation of cysteine to serine should maintain the force bearing contacts however, both A' and G strand are involved in the mechanical clamp region.

Mechanical unfolding of a polyprotein containing the mutant I27 yields a regular sawtooth pattern [214]. A force-extension trace for the mutant I27 is shown in Figure 4.1.2.3. Previous SMFS revealed that each sawtooth peak occurs at a regular distance of 23.1 ± 1.3 nm apart, consistent with previous studies, however the mutant I27 unfolds at lower forces of 173 pN at 1000 nm/s than an alternative mutant of I27 (see Chapter 3 (206 pN at 1000 nm/s) [214].

Simulations also indicate that mechanical unfolding of I27 occurs via a short-lived metastable state [217]. This intermediate is observed by a characteristic deviation in the rising edge of an unfolding peak from a fit of the WLC model to the unfolding curve for the I27 protein as illustrated in Figure 4.1.2.3 (labelled I). This "hump" feature is observed in the first few unfolding events and corresponds to an initial extension of the

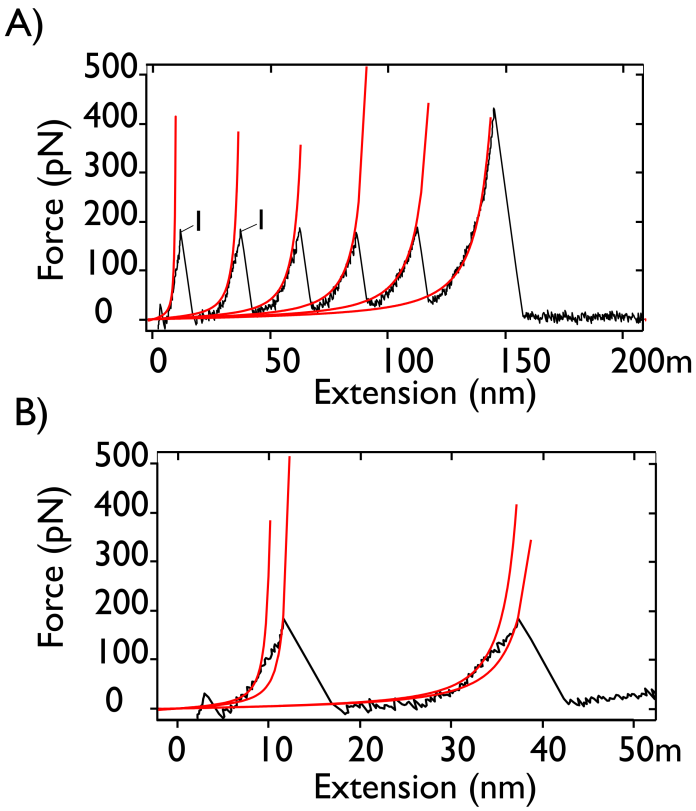


Figure 4.1.2.3: Figure illustrating the initial unfolding event observed in force-extension traces. A) Force extension trace of the $(I27)_5$ C63SC47S mutant. This protein construct shows the characteristic saw-tooth trace that is observed for the other mutant I27 proteins. The WLC fit to the saw-tooth trace is shown as the red solid line. A characteristic deviation from the WLC is observed in the first couple of unfolding peaks. The peak unfolding event in the force-extension trace corresponds to the intermediate (I) unfolding event. B) A close up on the first couple of unfolding peaks highlighting the initial unfolding events. A second WLC has been fit to the intermediate unfolding event - this is typically 3 nm longer than the initial unfolding event. The ΔL_c between the initial and intermediate unfolding force depends on the number of unfolded proteins.

A strand following unfolding from the native fold of I27. This is thought to involve the rupture of two to three hydrogen bonds between the A and B strands in every protein domain [243, 339]. This deviation is most prominent in the first few unfolding events but is also observed in the sequential unfolding events. The distance between the hump and peak unfolding force decreases as more unfolding events occur. This is because, in a polyprotein containing five protein domains, the first hump corresponds to the breaking of the hydrogen bonds in the five protein domains. As the protein unfolds the hydrogen bonds between the four remaining proteins refold. The hump on the second unfolding event corresponds to breaking of the hydrogen bonds in four domains. By the last unfolding event, it is the breaking of the hydrogen bonds in a single folded protein domain [339]. This unfolding event makes this protein an excellent candidate to study. Not only does this protein allow investigations into the affect of deuterium on the mechanical stability of the stable intermediate, it will allow direct testing of the strength of two hydrogen bonds in the native states in the protein.

4.1.3 Objectives

SMFS experiments to probe the force bearing regions of the two proteins will enable the importance of different interactions to be investigated. Protein L has a mechanical clamp motif that involves hydrogen bonds and hydrophobic contacts. I27 has a clamp motif that involves the shearing of 6 hydrogen bonds and interactions between the side chains in the A' and G stands but also requires interactions with water molecules to weaken the hydrogen bonds before unfolding. Furthermore an initial unfolding event, involving just a few hydrogen bonds, occurs before the unfolding of the stable intermediate. The force of this intermediate can be measured [365]. The objectives of this chapter are as follows:

- Determine the change in mechanical stability for the four pL test systems to determine the importance of hydrogen bonds and hydrophobic interaction on the unfolding force.
- Determine the change in mechanical stability for the four I27 test systems to determine the importance of hydrogen bonds and solvent on the unfolding force of the stable intermediate.
- Determine whether a change in unfolding force is observed for the initial breaking of two or three hydrogen bonds in the I27 protein systems.

For each mechanical unfolding experiment, traces containing 2–5 unfolding forces were

accepted. Examples of such traces are shown in Figure 4.1.3.1.

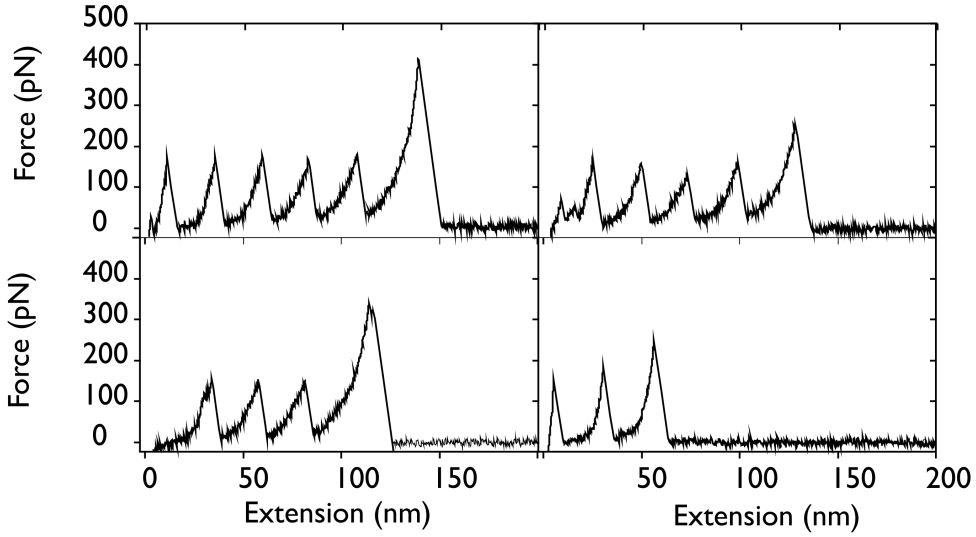


Figure 4.1.3.1: Typical force extension traces for I27 accepted in mechanical unfolding experiments in this project. Between two (bottom,right) and five (top, left) unfolding events were accepted for analysis.

4.2 The mechanical stability of pL

In order to assess the effect of deuterium on the mechanical stability of pL, mechanical unfolding experiments were carried out on the four polyprotein test systems. These are the $p(pL)_5$ in H_2O, $p(pL)_5$ in D_2O, $d(pL)_5$ in H_2O and $d(pL)_5$ in D_2O. Proteins mechanically unfolded in D_2O and H_2O were isotopically exchanged before each experiment to ensure isotope atoms in the exchangeable hydrogen positions were identical to the solvent. Each system was mechanically unfolded in 63 mM sodium phosphate buffer, pH 7.4 at four velocities (160 nm/s, 400 nm/s, 1000 nm/s and 2000 nm/s). Unfolding of the full polyprotein construct was observed for each system at every velocity. Examples of these are shown in Figures 4.2.0.1–4.2.0.4. Observation of the maximum number of possible unfolding events suggested the protein constructs were fully folded and did not degrade. An increase in the peak-to-peak distance, x_p, with velocity was observed for each system. This increase in x_p has been observed previously. For p(pL) in H_2O the x_p increases from 15.8 to 16.1 nm. The p(pL) in D_2O increases from 16.1 to 16.5 nm. The d(pL) in H_2O increases from 16.2 to 16.6 nm and the d(pL) in D_2O increases from 16.2 to 16.6 nm. The average $x_p \pm$ standard deviation for 160 nm/s and 2000 nm/s for each of the systems was: 16.1 ± 0.1 nm and 16.5 ± 0.1 nm; 16.2 ± 0.1 nm and 16.6 ± 0.1 nm; 15.9 ± 0.1 nm and 16.1 ± 0.1; 16.1 ± 0.1 and 16.6 ± 0.1, for $p(pL)_5$ in H_2O, $p(pL)_5$ in D_2O, $d(pL)_5$ in H_2O and $d(pL)_5$ in D_2O respectively. This is consistent with previous

Pulling velocity/nms^{-1}	n	Median unfolding force/pN (SD)	Modal unfolding force/pN (SD)	Average/pN (± SE)
$p(pL)_5$ in H_2O				
160	256	123.2 (31.0)	121.7 (39.2)	
	77	126.8 (29.1)	121.7 (36.6)	121.6 (± 1.4)
	76	124.4 (34.0)	123.9 (45.1)	
400	286	142.0 (35.3)	138.0 (45.4)	
	50	142.9 (29.2)	141.7 (43.1)	139.3 (± 1.2)
	118	141.7 (31.8)	138.3 (43.4)	
1000	165	157.1 (38.8)	154.0 (43.5)	
	340	157.2 (37.1)	154.0 (42.2)	158.7 (± 3.0)
	136	166 (33.3)	158.8 (42.9)	
	115	167.5 (36.4)	167.1 (38.6)	
2000	230	171.4 (38.6)	168.9 (44.7)	
	96	174.9 (39.0)	174.2 (37.4)	173.6 (± 2.5)
	116	179.8 (43.7)	177.6 (43.7)	

Table 4.2.0.1: Summary of unfolding statistics for $p(pL)_5$ in H_2O in 63 mM sodium phosphate buffer, pH 7.4 at room temperature.
n is the number of data points
The modal unfolding force is obtained from a Gaussian fit to the histogram distribution of the forces SD is the standard deviation or the full width at half height of a Gaussian fit to the histogram
The average is the average of the modal forces
SE is the standard error in the average

studies where the x_p increases from 15.2 ± 0.3 nm to 16.9 ± 0.3 nm [217]. The peaks were all fit by an average change in contour length, L_c=19.0 and persistence length of 0.4 nm which is also consistent with previous studies [217].

Force-frequency histograms were obtained for at least three repeats of each velocity for every system. Each histogram was fit by a Gaussian to obtain the modal unfolding force for each experiment. The histograms for each of the systems are shown in Figures 4.2.0.5–4.2.0.8 and the unfolding statistics are given in Tables 4.2.0.1–4.2.0.4. The histograms appear to have similar widths, therefore suggesting that the distance to the transition step is not perturbed. The force - ln(v) graph is shown in Figure 4.2.0.9. The forces for the $p(pL)_5$ in H_2O agree with those previously obtained by Brockwell et al. (2002) under similar conditions. Furthermore, the data for all systems can be described by a linear fit. The graph showing the percent increase between each of the systems and the p(pL) in H_2O is shown in Figure 4.2.0.10. This graph indicates that the systems in D_2O have a higher force than the p(pL) in H_2O. The most mechanically stable is the p(pL) in D_2O. This is attributed to an increase in the hydrophobic interaction for this protein. The average percent increase for this system is $\sim$ 14 % compared with the p(pL) in H_2O. As expected, a decrease in unfolding force was observed for the d(pL) in H_2O. Deuterated proteins in H_2O are typically found to be destabilised as a result of a

Pulling velocity/nms^{-1}	n	Median unfolding force/pN (SD)	Modal unfolding force/pN (SD)	Average/pN (± SE)
p(pL)$_5$ in D_2O				
160	81	149.2 (41.5)	143.9 (38.7)	
	58	145.9 (38.8)	143.9 (28.5)	143.5 (± 0.5)
	67	148.1 (32.8)	142.5 (44.7)	
400	128	165.1 (30.0)	160.9 (43.0)	
	180	167.3 (31.3)	160.4 (33.3)	160.2 (± 0.4)
	60	164.1 (41.9)	159.4 (54.8)	
1000	232	182.9 (36.7)	178.5 (36.9)	
	105	184.0 (39.2)	178.5 (48.4)	178.3 (± 0.2)
	108	177.3 (36.8)	177.9 (48.4)	
2000	82	202.8 (47.8)	199.0 (44.5)	
	100	199.5 (41.7)	197.2 (33.4)	191.1 (± 4.1)
	70	185.1 (30.5)	182.2 (25.2)	
	178	187.8 (36.0)	186.6 (50.0)	

Table 4.2.0.2: Summary of unfolding statistics for p(pL)$_5$ in D_2O in 63 mM sodium phosphate buffer, pH 7.4 at room temperature.
n is the number of data points
The modal unfolding force is obtained from a Gaussian fit to the histogram distribution of the forces SD is the standard deviation or the full width at half height of a Gaussian fit to the histogram
The average is the average of the modal forces
SE is the standard error in the average

Pulling velocity/nms^{-1}	n	Median unfolding force/pN (SD)	Modal unfolding force/pN (SD)	Average/pN (± SE)
d(pL)$_5$ in H_2O				
160	158	116.8 (31.9)	110.9 (42.8)	
	84	116.8 (28.3)	115.1 (42.5)	111.8 (± 1.7)
	107	114.6 (41.9)	109.5 (42.5)	
400	159	135.3 (35.2)	130.9 (44.1)	
	105	133.0 (44.3)	128.2 (45.0)	128.8 (± 1.1)
	90	131.3 (32.2)	127.3 (39.9)	
1000	89	155.0 (35.4)	150.4 (49.4)	
	86	152.1 (40.6)	150.1 (34.4)	148.4 (± 1.8)
	91	148.3 (32.5)	144.8 (42.4)	
2000	183	168.2 (38.3)	163.1 (44.1)	
	110	167.4 (44.7)	167.8 (46.7)	163.5 (± 2.4)
	131	162.6 (38.1)	159.6 (54.0)	

Table 4.2.0.3: Summary of unfolding statistics for d(pL)$_5$ in H_2O in 63 mM sodium phosphate buffer, pH 7.4 at room temperature.
n is the number of data points
The modal unfolding force is obtained from a Gaussian fit to the histogram distribution of the forces SD is the standard deviation or the full width at half height of a Gaussian fit to the histogram
The average is the average of the modal forces
SE is the standard error in the average

Pulling velocity/nms^{-1}	n	Median unfolding force/pN (SD)	Modal unfolding force/pN (SD)	Average/pN (± SE)
$d(pL)_5$ in D_2O				
160	88	143.4 (34.4)	136.1 (35.8)	
	120	140.2 (38.3)	135.6 (36.6)	133.2 (± 2.6)
	204	133.0 (29.9)	128.0 (39.2)	
400	124	155.0 (33.5)	150.2 (40.2)	
	111	156.4 (32.3)	152.7 (39.0)	150.8 (± 1.0)
	201	153.8 (31.1)	149.4 (38.6)	
1000	89	171.5 (36.1)	168.8 (52.2)	
	311	175.3 (34.4)	170.3 (39.6)	169.8 (± 0.6)
	192	173.3 (36.3)	170.6 (37.3)	
2000	408	195.3 (43.9)	190.6 (51.7)	
	104	191.1 (31.2)	186.4 (43.3)	186.3 (± 2.5)
	129	185.4 (38.1)	181.1 (37.3)	

Table 4.2.0.4: Summary of unfolding statistics for $d(pL)_5$ in D_2O in 63 mM sodium phosphate buffer, pH 7.4 at room temperature.
n is the number of data points
The modal unfolding force is obtained from a Gaussian fit to the histogram distribution of the forces SD is the standard deviation or the full width at half height of a Gaussian fit to the histogram
The average is the average of the modal forces
SE is the standard error in the average

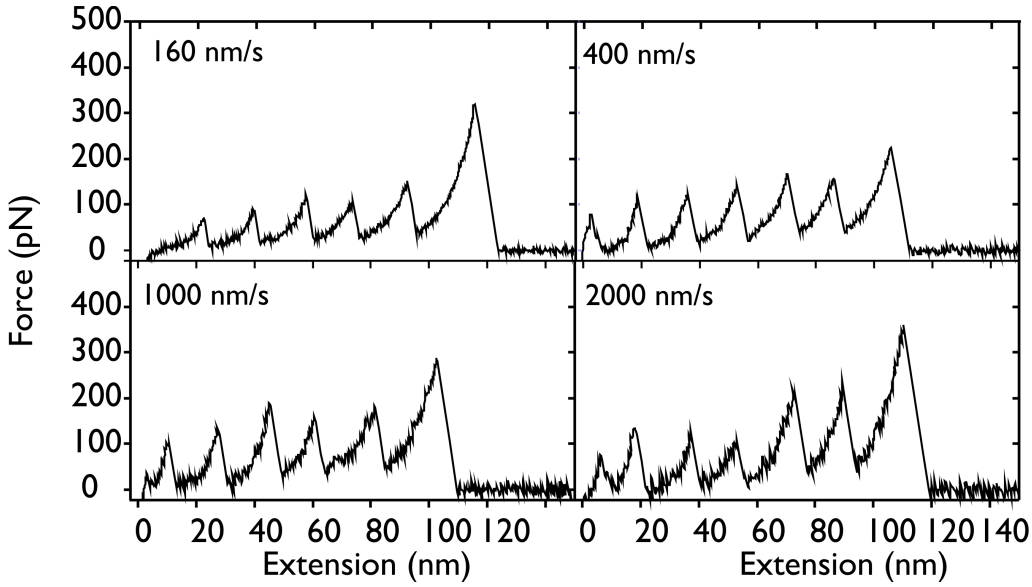


Figure 4.2.0.1: Force-extension traces showing mechanical unfolding events for all five domains within the polyprotein in the $p(pL)_5$ in H_2O. Force-extension traces were obtained at four different velocities. The traces show equally spaces peaks for each of the constructs suggesting the proteins were correctly folded in the polyprotein domain.

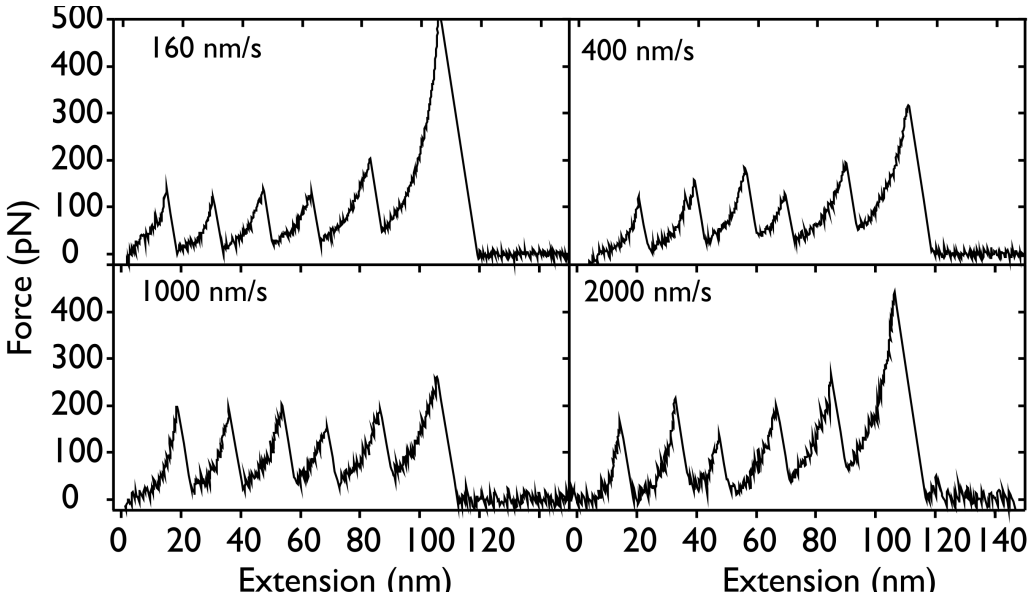


Figure 4.2.0.2: Force-extension traces showing mechanical unfolding events for all five domains within the polyprotein in the $p(pL)_5$ in D_2O systems. Force-extension traces were obtained at four different velocities. The traces show equally spaces peaks for each of the constructs suggesting the proteins were correctly folded in the polyprotein domain.

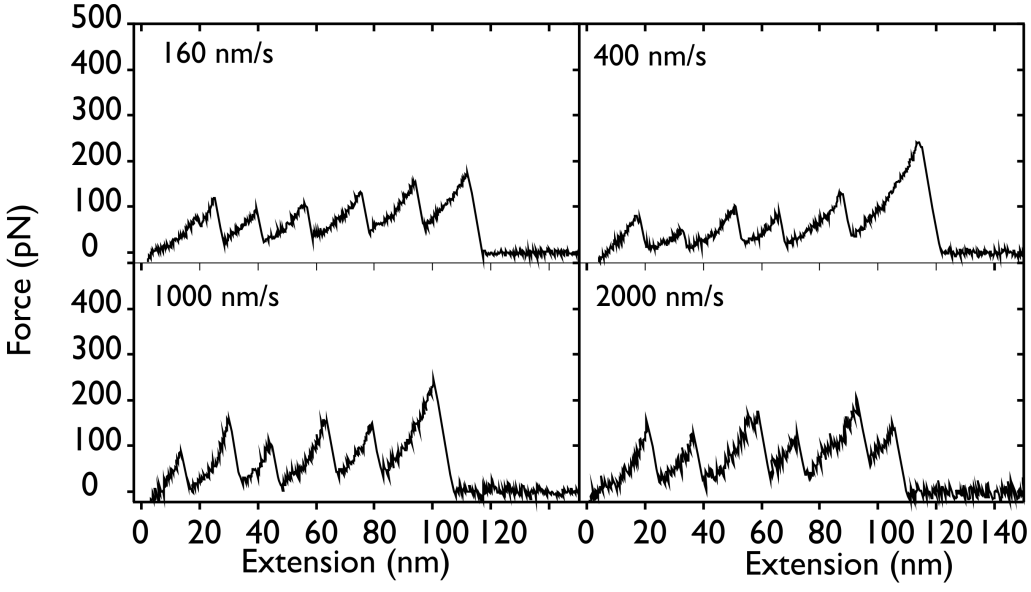


Figure 4.2.0.3: Force-extension traces showing mechanical unfolding events for all five domains within the polyprotein in the $d(pL)_5$ in H_2O systems. Force-extension traces were obtained at four different velocities. The traces show equally spaces peaks for each of the constructs suggesting the proteins were correctly folded in the polyprotein domain.

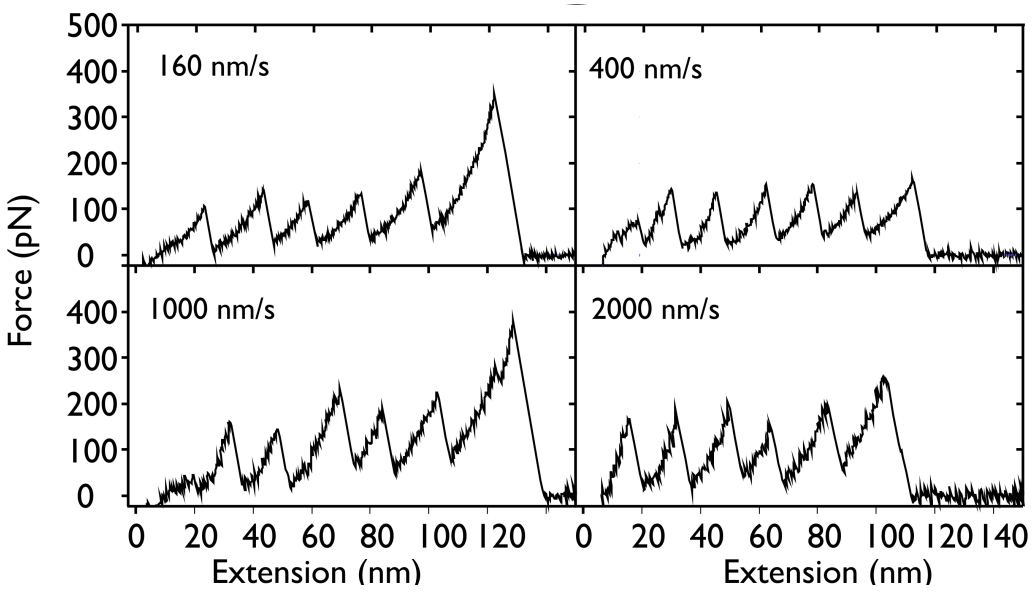


Figure 4.2.0.4: Force-extension traces showing mechanical unfolding events for all five domains within the polyprotein in the $d(pL)_5$ in D_2O test systems. Force-extension traces were obtained at four different velocities. The traces show equally spaces peaks for each of the constructs suggesting the proteins were correctly folded in the polyprotein domain.

reduction in hydrophobicity of proteins [130, 137]. Another interesting result is that a decrease in 7 % is observed for this system and an increase of 7 % is observed for the d(pL) in D_2O. This suggests that there is an increase in unfolding force of 7 % for an increase in the hydrophobic interactions for the p(pL) in D_2O which is consistent with the decrease of 7 % for reducing the strength of the hydrophobic interactions.

4.3 The mechanical stability of I27

In order to determine the importance of hydrogen bond strength on the mechanical stability of I27, mechanical unfolding experiments were also carried out on the four I27 polyprotein test systems. These are the $p(I27)_5$ in H_2O[1], $p(I27)_5$ in D_2O, $d(I27)_5$ in H_2O and $d(I27)_5$ in D_2O. Each system was mechanically unfolded in 63 mM sodium phosphate buffer, pH 7.4, room temperature at four velocities (160 nm/s, 400 nm/s, 1000 nm/s and 2000 nm/s). Unfolding of the full polyproteins (five protein domains) was observed for each system at every velocity. Examples of these are shown in Figures 4.3.0.1–4.3.0.4. Observation of the maximum number of possible unfolding events suggested that the protein constructs were fully folded and did not degrade. The average peak-to-peak distance, $x_p \pm$ standard deviation for each of the systems was 24.0 ± 1.4 nm, 24.2 ± 1.0 nm, 24.1 ± 0.8 nm and 23.9 ± 1.0 for $p(I27)_5$ in H_2O, $p(I27)_5$ in D_2O, $d(I27)_5$ in H_2O and $d(I27)_5$ in D_2O respectively. This is consistent with previous studies where the x_p

[1]Dr Katarzyna Tych obtained a set of data for this system.

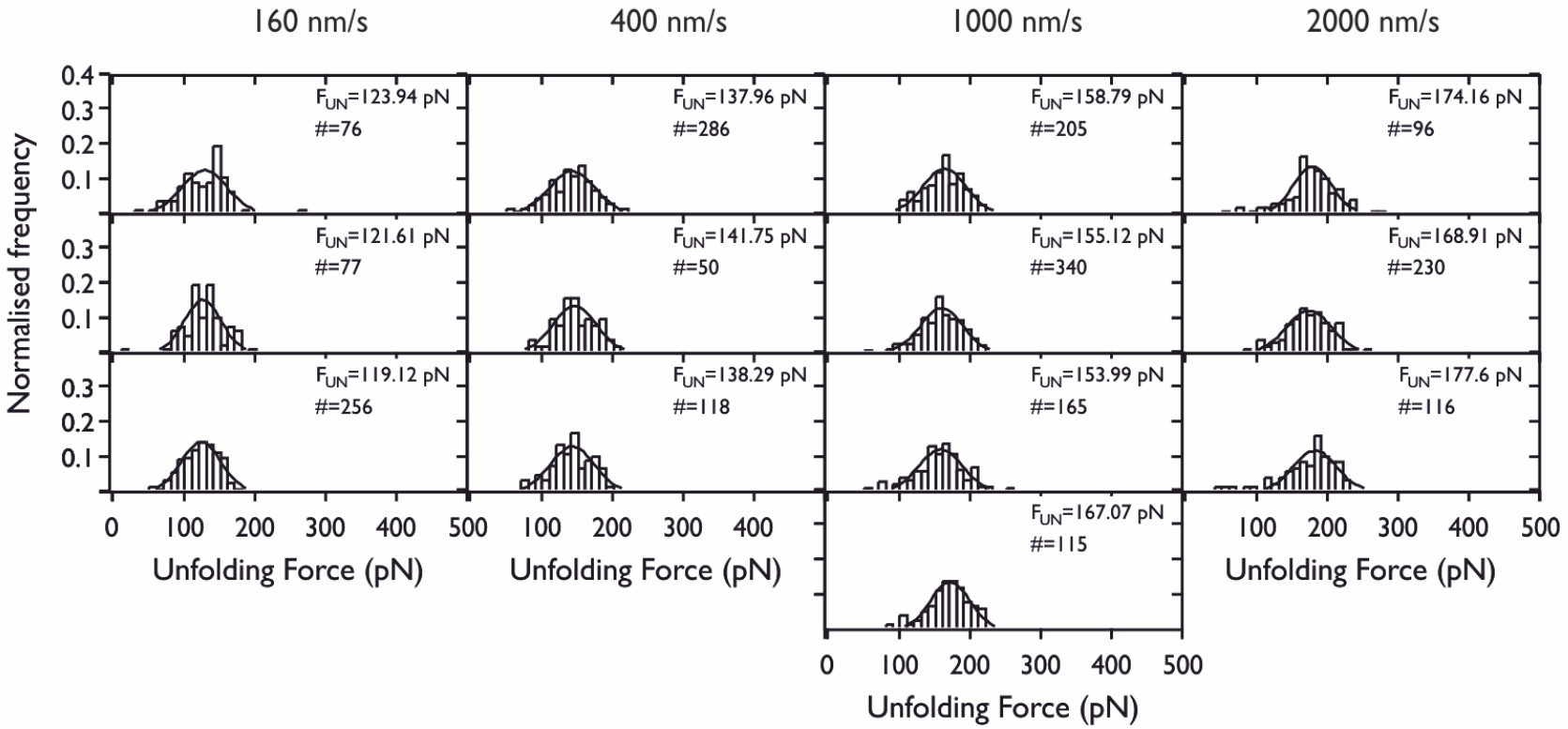


Figure 4.2.0.5: Histograms showing the unfolding force distributions obtained from SMFS experiments on the $p(pL)_5$ in H_2O system. The modal force from a Gaussian fit to the distribution of forces, and number of data points are displayed on each histogram.

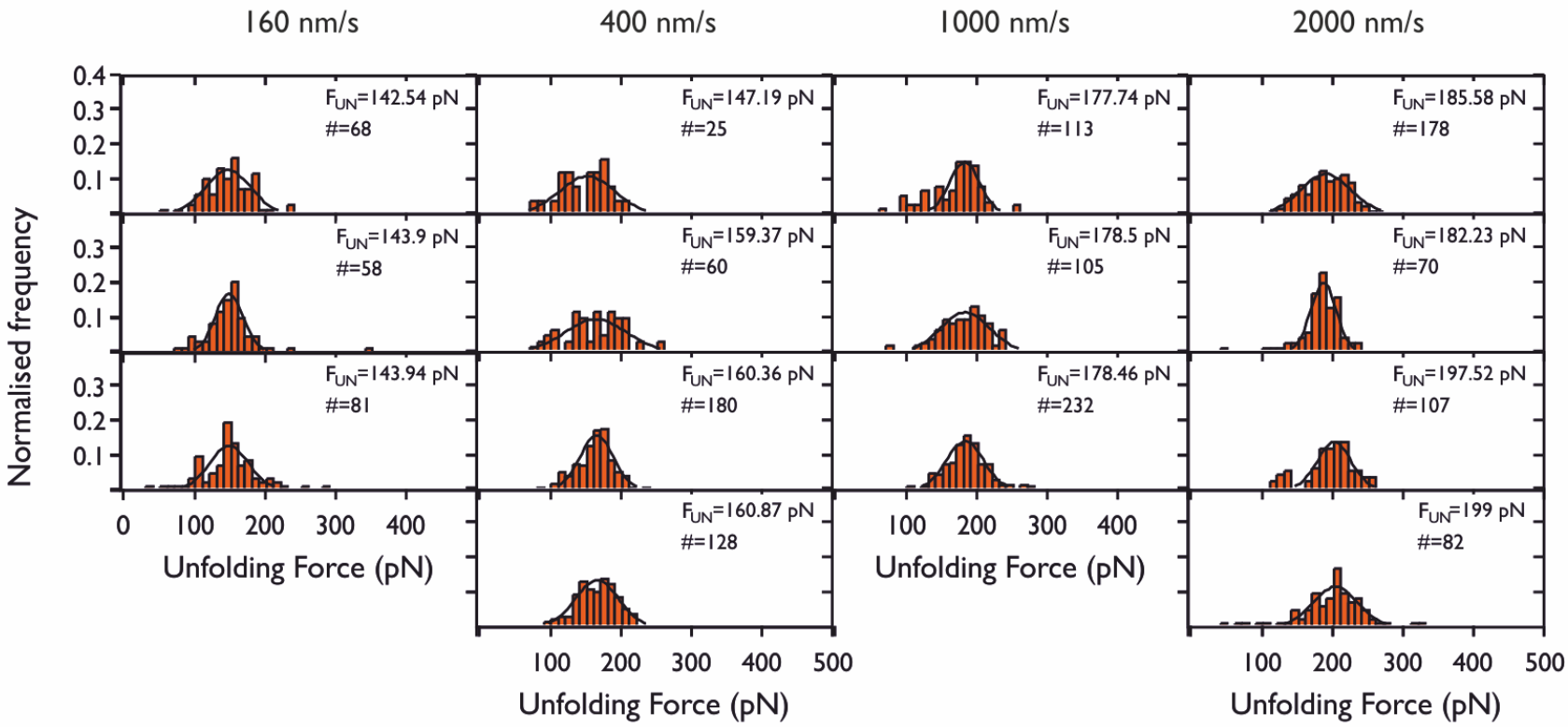


Figure 4.2.0.6: Histograms showing the unfolding force distributions obtained from SMFS experiments on the $p(pL)_5$ in D_2O system. The modal force from a Gaussian fit to the distribution of forces, and number of data points are displayed on each histogram.

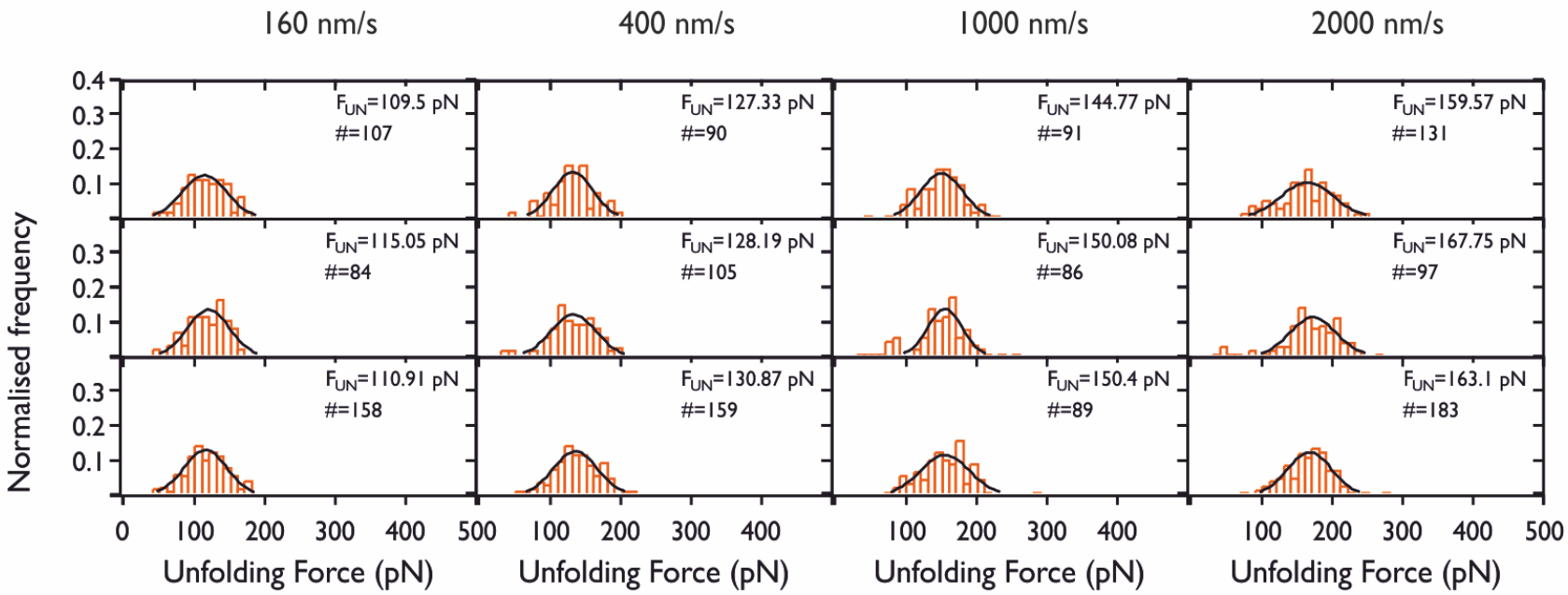


Figure 4.2.0.7: Histograms showing the unfolding force distributions obtained from SMFS experiments on the $d(pL)_5$ in H_2O system. The modal force from a Gaussian fit to the distribution of forces, and number of data points are displayed on each histogram.

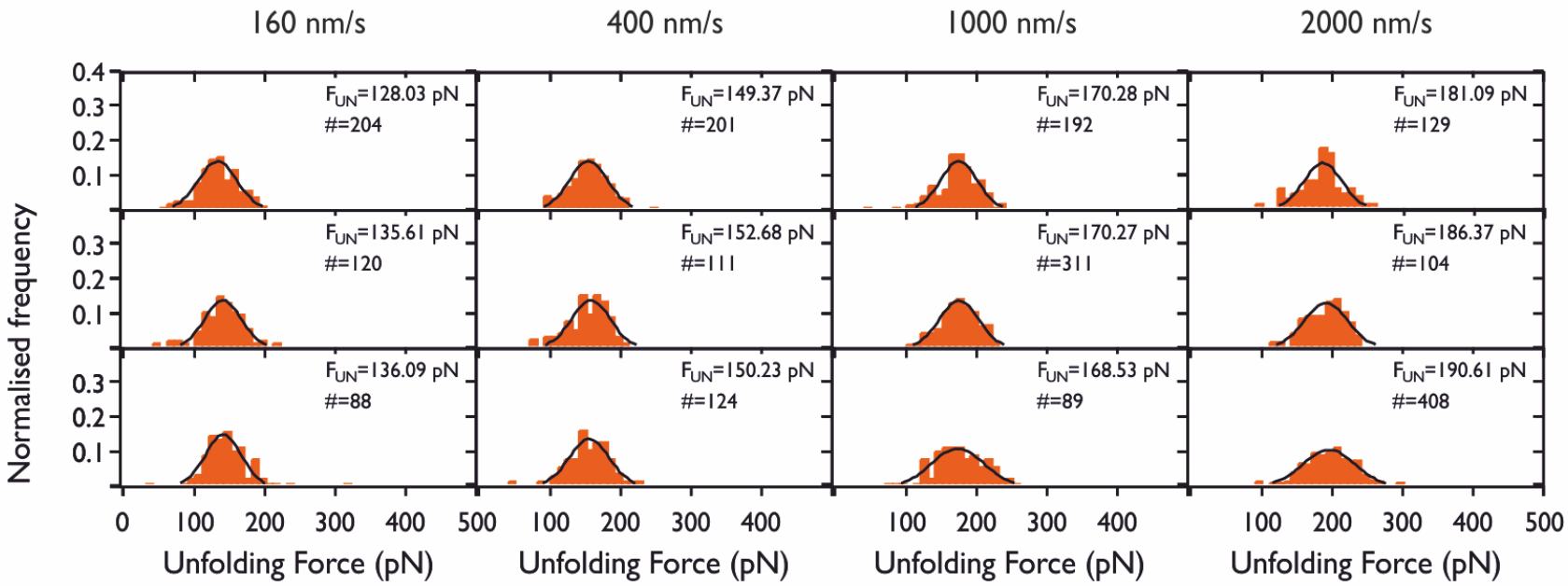


Figure 4.2.0.8: Histograms showing the unfolding force distributions obtained from SMFS experiments on the $d(pL)_5$ in D_2O system. The modal force from a Gaussian fit to the distribution of forces, and number of data points are displayed on each histogram.

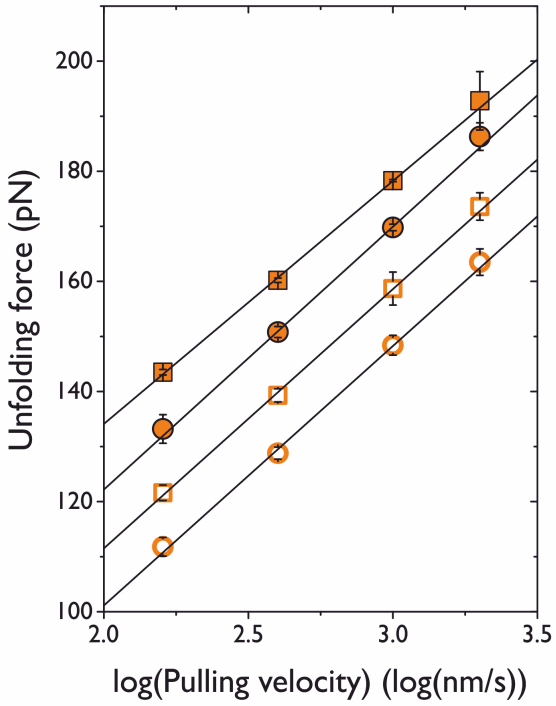


Figure 4.2.0.9: The dependence of force on the logarithm of pulling speed (log(v)). Comparison of the force-log(v) graphs for each of the pL test systems: $p(pL)_5$ in H_2O (open squares), $p(pL)_5$ in D_2O (closed squares), $d(pL)_5$ in H_2O (open circles) and $d(pL)_5$ in D_2O (closed circles). All are described well by a linear fit.

was found to be 23.1 ± 1.3 nm [214]. The peaks were all fit by an average change in contour length, L_c=28.0 nm and persistence length of 0.4 nm. This is in agreement with previous studies where the persistence length was 0.4 nm and ΔL_c was 28 ± 1 nm [214].

Force-frequency histograms were obtained for at least three separate repeats of the four velocities for every protein system. Each histogram of unfolding forces was fit by a Gaussian to obtain the modal force of unfolding, F_{UN}. The histograms for each of the systems are shown in Figures 4.3.0.5–4.3.0.8 and the unfolding statistics are given in Tables 4.3.0.1–4.3.0.4. Comparing the unfolding force histograms it can be seen that the distributions are slightly wider for the systems in D_2O (Figures 4.3.0.6 and 4.3.0.8). This could suggest a change in the distance between the folded and transition state. The average width at half height of Gaussian fits to the histograms for the I27 test systems are: 31.1 ± 7.6 pN, 36.6 ± 8.3 pN, 33.5 ± 5.4 pN and 36.4 ± 6.5 pN for the p(I27) in H_2O, p(I27) in D_2O, d(I27) in H_2O and d(I27) in D_2O respectively. The force - ln(v) graph is shown in Figure 4.3.0.9. The forces for the $p(I27)_5$ in H_2O agree with those previously obtained by Brockwell et al. (2002) under similar conditions (see Appendix C). The pulling speed dependence data, in Figure 4.3.0.9, for all I27 protein-solvent systems can be described by a linear fit. Inspection of the data shows that the gradients

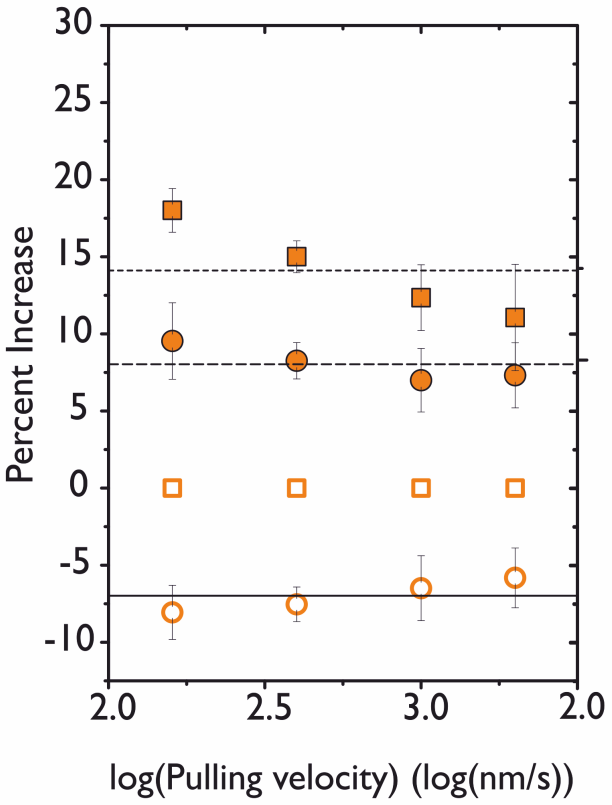


Figure 4.2.0.10: The percent increase in the forces for the test systems compared to p)$_5$ in H_2O. p(pL)$_5$ in H_2O (open squares), p(pL)$_5$ in D_2O (closed squares), d(pL)$_5$ in H_2O (open circles) and d(pL)$_5$ in D_2O (closed circles). All are described well by a linear fit. The percent increases were determined from the experimental forces at each velocity. The dotted line corresponds to the average percent increase for the p(pL)$_5$ in D_2O system, the dashed line corresponds to the average percent increase for the d(pL)$_5$ in D_2O system and the solid line indicates the position of the average percent increase in the d(pL)$_5$ in H_2O system.

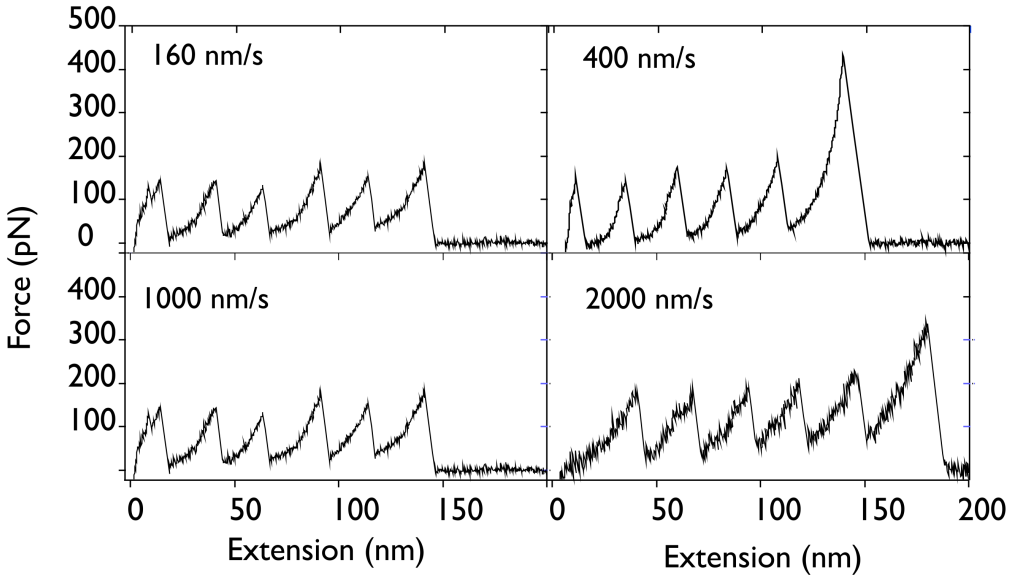


Figure 4.3.0.1: Force-extension traces showing the full construct unfolding for the p(I27)$_5$ in H_2O test system. Force-extension traces were obtained at four different velocities. The traces show equally spaces peaks for each of the constructs suggesting the proteins were correctly folded in the polyprotein domain.

Pulling velocity/nms^{-1}	n	Median unfolding force/pN (SD)	Modal unfolding force/pN (SD)	Average/pN (± SE)
$p(I27)_5$ in H_2O				
160	45	159.8 (35.9)	155.6 (24.9)	
	34	153.3 (43.0)	152.4 (29.1)	154.3 (± 2.7)
	80	162.0 (34.0)	161.4 (35.8)	
	24	153.4 (18.7)	147.8 (23.2)	
400	82	170.0 (35.3)	166.2 (24.9)	
	81	177.8 (33.3)	170.5 (36.6)	168.1 (± 1.1)
	165	173.8 (21.3)	167.7 (24.3)	
1000	75	183.2 (26.6)	179.1 (31.0)	
	139	181.0 (25.1)	175.8 (30.3)	180.0 (± 2.7)
	47	187.2 (32.4)	185.1 (50.0)	
2000	43	195.5 (34.2)	194.0 (24.1)	
	62	192.5 (29.1)	190.9 (29.3)	190.2 (± 1.4)
	59	194.5 (27.5)	188.0 (38.1)	
	29	191.1 (34.7)	188.0 (37.8)	

Table 4.3.0.1: Summary of unfolding statistics for $p(I27)_5$ in H_2O in 63 mM sodium phosphate buffer, pH 7.4 at room temperature.
n is the number of data points
The modal unfolding force is obtained from a Gaussian fit to the histogram distribution of the forces SD is the standard deviation or the full width at half height of a Gaussian fit to the histogram
The average is the average of the modal forces
SE is the standard error in the average

Pulling velocity/nms^{-1}	n	Median unfolding force/pN (SD)	Modal unfolding force/pN (SD)	Average/pN (± SE)
$p(I27)_5$ in D_2O				
160	69	186.1 (34.9)	179.7 (32.6)	
	58	184.3 (23.8)	178.7 (34.6)	180.6 (± 0.9)
	47	187.8 (29.0)	182.3 (33.8)	
	73	180.9 (37.2)	181.9 (25.1)	
400	93	191.9 (34.3)	188.1 (29.8)	
	73	191.0 (30.6)	188.1 (29.1)	195.2 (± 4.5)
	48	200.8 (39.6)	197.9 (27.5)	
	51	210.8 (30.1)	206.7 (29.3)	
1000	84	214.5 (36.3)	208.1 (40.8)	
	63	215.9 (33.2)	212.3 (32.1)	215.1 (± 2.9)
	80	221.8 (44.4)	219.8 (46.9)	
	115	221.3 (40.1)	220.1 (40.5)	
2000	179	224.1 (34.0)	219.4 (38.3)	
	49	230.2 (41.3)	227.6 (53.4)	224.8 (± 2.3)
	94	234.5 (38.9)	229.5 (41.0)	
	66	221.1 (52.5)	222.6 (50.1)	

Table 4.3.0.2: Summary of unfolding statistics for $p(I27)_5$ in D_2O in 63 mM sodium phosphate buffer, pH 7.4 at room temperature.
n is the number of data points
The modal unfolding force is obtained from a Gaussian fit to the histogram distribution of the forces SD is the standard deviation or the full width at half height of a Gaussian fit to the histogram
The average is the average of the modal forces
SE is the standard error in the average

Pulling velocity/nms^{-1}	n	Median unfolding force/pN (SD)	Modal unfolding force/pN (SD)	Average/pN (± SE)
$d(I27)_5$ in H_2O				
160	60	161.0 (31.5)	156.1 (34.6)	
	64	169.1 (26.9)	166.3 (32.8)	161.2 (± 2.2)
	58	166.9 (28.1)	162.8 (26.2)	
	96	164.9 (25.3)	159.7 (31.4)	
400	78	176.6 (34.8)	172.2 (25.8)	
	37	183.2 (25.4)	176.2 (39.5)	175.8 (± 1.9)
	89	178.0 (26.7)	174.0 (38.7)	
	62	181.9 (32.8)	180.8 (37.5)	
1000	82	192.1 (30.2)	187.1 (30.7)	
	71	196.4 (29.0)	191.3 (32.1)	188.7 (± 1.3)
	107	191.6 (29.5)	187.6 (29.0)	
2000	134	208.4 (33.7)	205.3 (39.9)	
	152	204.3 (37.0)	200.3 (36.3)	203.0 (± 1.5)
	115	207.4 (41.8)	203.4 (44.8)	

Table 4.3.0.3: Summary of unfolding statistics for $d(I27)_5$ in H_2O in 63 mM sodium phosphate buffer, pH 7.4 at room temperature.
n is the number of data points
The modal unfolding force is obtained from a Gaussian fit to the histogram distribution of the forces SD is the standard deviation or the full width at half height of a Gaussian fit to the histogram
The average is the average of the modal forces
SE is the standard error in the average

Pulling velocity/nms^{-1}	n	Median unfolding force/pN (SD)	Modal unfolding force/pN (SD)	Average/pN (± SE)
$d(I27)_5$ in D_2O				
160	62	174.8 (45.3)	171.2 (32.6)	
	89	174.9 (40.2)	172.3 (36.8)	173.5 (± 1.2)
	43	179.5 (30.6)	176.5 (26.1)	
	53	176.0 (29.1)	174.1 (30.2)	
400	83	185.4 (30.1)	180.2 (27.4)	
	75	190.7 (33.0)	188.0 (32.6)	188.0 (± 4.5)
	62	200.4 (51.2)	195.7 (34.6)	
1000	80	204.2 (32.7)	203.0 (33.7)	
	130	210.8 (32.2)	202.9 (39.8)	204.5 (± 1.6)
	132	212.2 (41.1)	207.6 (44.3)	
2000	90	221.5 (37.0)	221.6 (48.3)	
	81	227.9 (52.4)	223.4 (41.5)	219.0 (± 2.1)
	39	221.9 (28.6)	214.5 (41.0)	
	113	219.1 (41.7)	216.7 (40.8)	

Table 4.3.0.4: Summary of unfolding statistics for $d(pL)_5$ in D_2O in 63 mM sodium phosphate buffer, pH 7.4 at room temperature.
n is the number of data points
The modal unfolding force is obtained from a Gaussian fit to the histogram distribution of the forces SD is the standard deviation or the full width at half height of a Gaussian fit to the histogram
The average is the average of the modal forces
SE is the standard error in the average

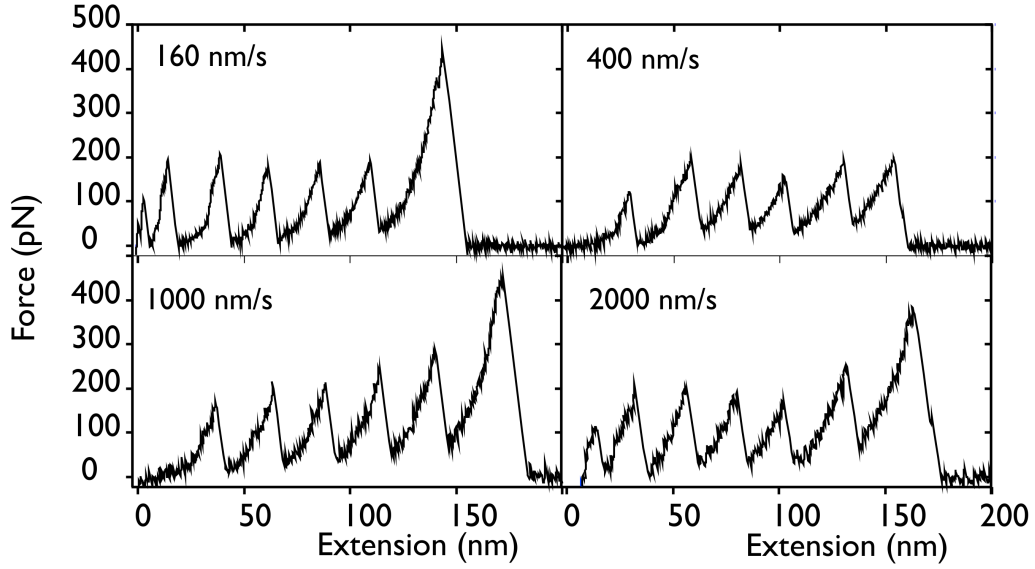


Figure 4.3.0.2: Force-extension traces showing the full construct unfolding for the $p(I27)_5$ in D_2O test systems. Force-extension traces were obtained at four different velocities. The traces show equally spaces peaks for each of the constructs suggesting the proteins were correctly folded in the polyprotein domain.

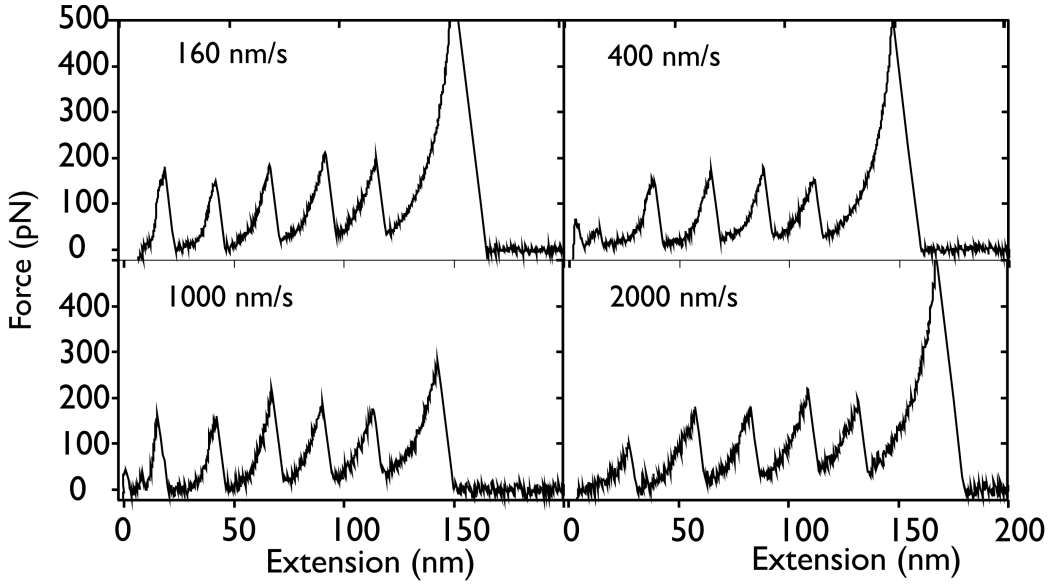


Figure 4.3.0.3: Force extension traces showing mechanical unfolding of the full construct unfolding for the $d(I27)_5$ in H_2O test system. Force-extension traces were obtained at four different velocities. The traces show equally spaces peaks for each of the constructs suggesting the proteins were correctly folded in the polyprotein domain.

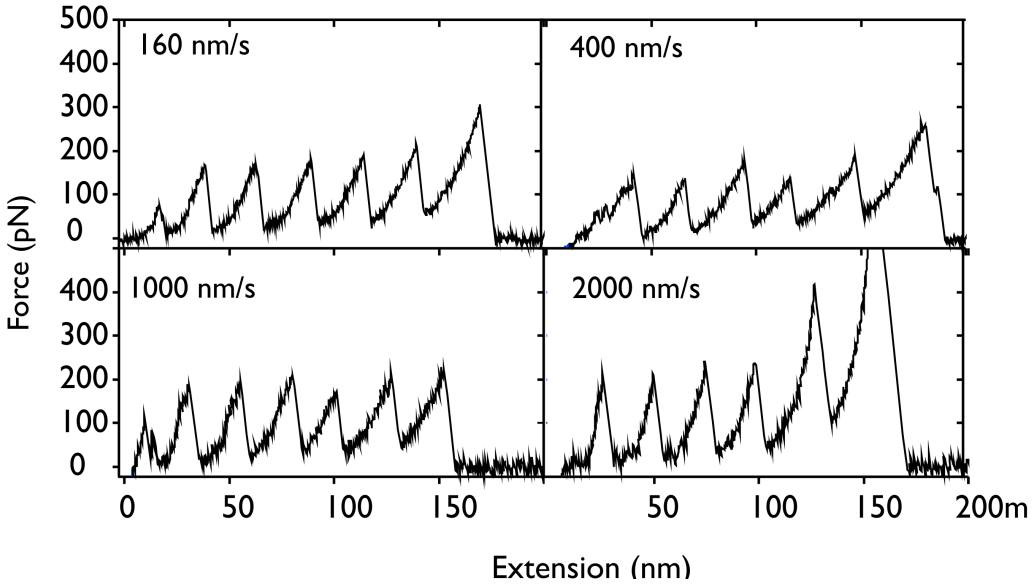


Figure 4.3.0.4: Force extension traces showing the full construct unfolding for the $d(I27)_5$ in D_2O test system. Force-extension traces were obtained at four different velocities. The traces show equally spaces peaks for each of the constructs suggesting the proteins were correctly folded in the polyprotein domain.

are not significantly different. The percentage increase in the forces, compared to the $p(I27)_5$ in H_2O show a stepwise increase in the unfolding force for the d(I27) in D_2O, p(I27) in D_2O and, most suprisingly, d(I27) in H_2O (Figure 4.3.0.10). This result is surprising for the d(I27) in H_2O where a decrease in the hydrophobicity was expected to eliminate hydrophobic contact made by the side chains and decrease the stability of the protein. However, the increase in force for this system is only marginal ($\sim$ 5%). This is similar to the percentage difference observed between the d(I27) in D_2O and p(I27) in D_2O (4 %). The mechanical stability is most significantly changed as a consequence of increasing the hydrogen bond strength in the solvent in the p(protein) in D_2O ($\sim$ 18 %) and d(protein) in D_2O ($\sim$ 13 %). The mechanical clamp of I27 is made up of six hydrogen bonds. Therefore the increase in mechanical stability of the systems in D_2O is not surprising. The average percent increase for the p(I27) in D_2O, is $\sim$ 18 %. This agrees well with the increase in force observed by Dougan et al. (2008) (17.6 %) [132].

Populations of the intermediate, evidenced by a "hump" in rising edge of early protein unfolding events, were observed on traces for all I27 test systems. The force of the intermediate should correspond to force required to break two–three hydrogen bonds.

4.3.0.1 Effect of deuterium on the initial unfolding event of I27

To determine the effect of isotopic substitution on the relevant strength of the native state of I27 is was necessary to quantify the force at which the deviation to the WLC occurs. A

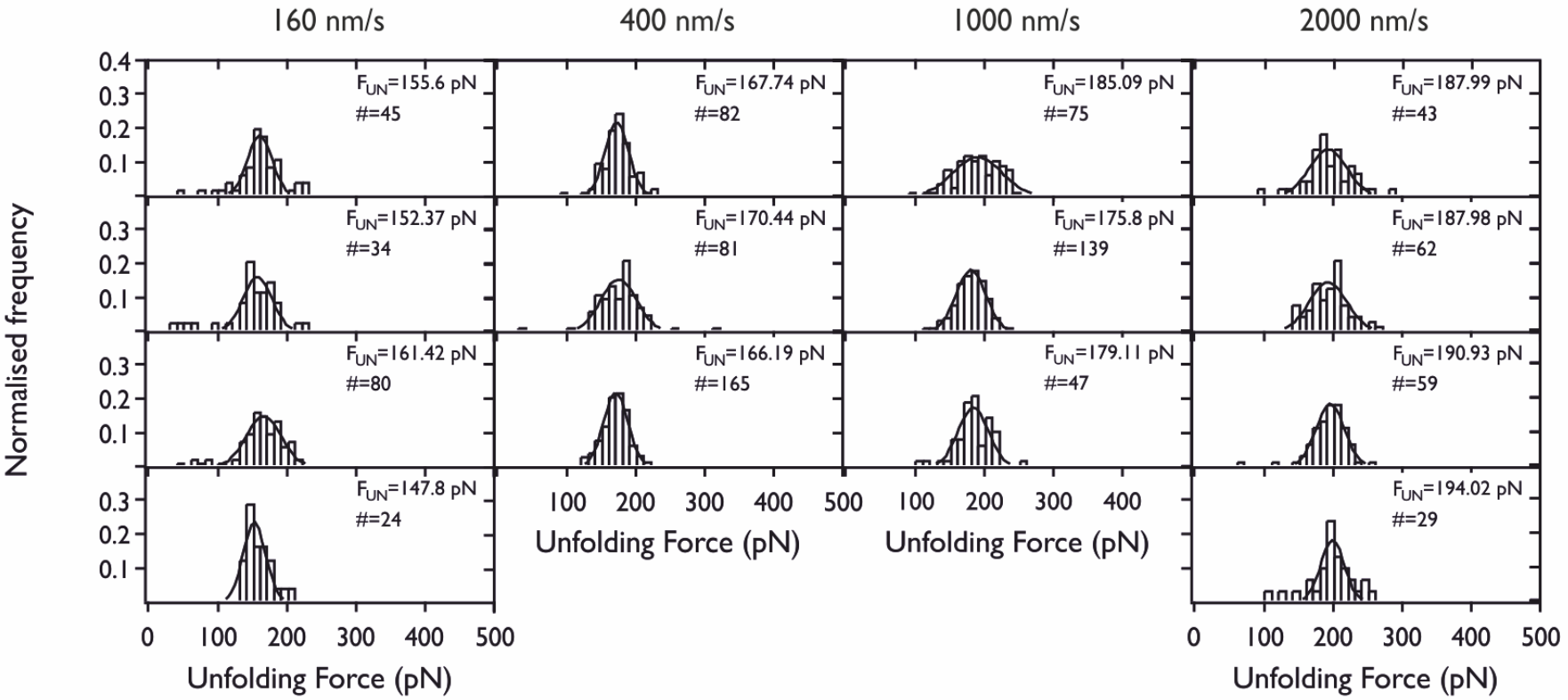


Figure 4.3.0.5: Histograms showing the unfolding force distributions obtained from SMFS experiments on the $p(I27)_5$ in H_2O system. The modal force from a Gaussian fit to the distribution of forces, and number of data points are displayed on each histogram.

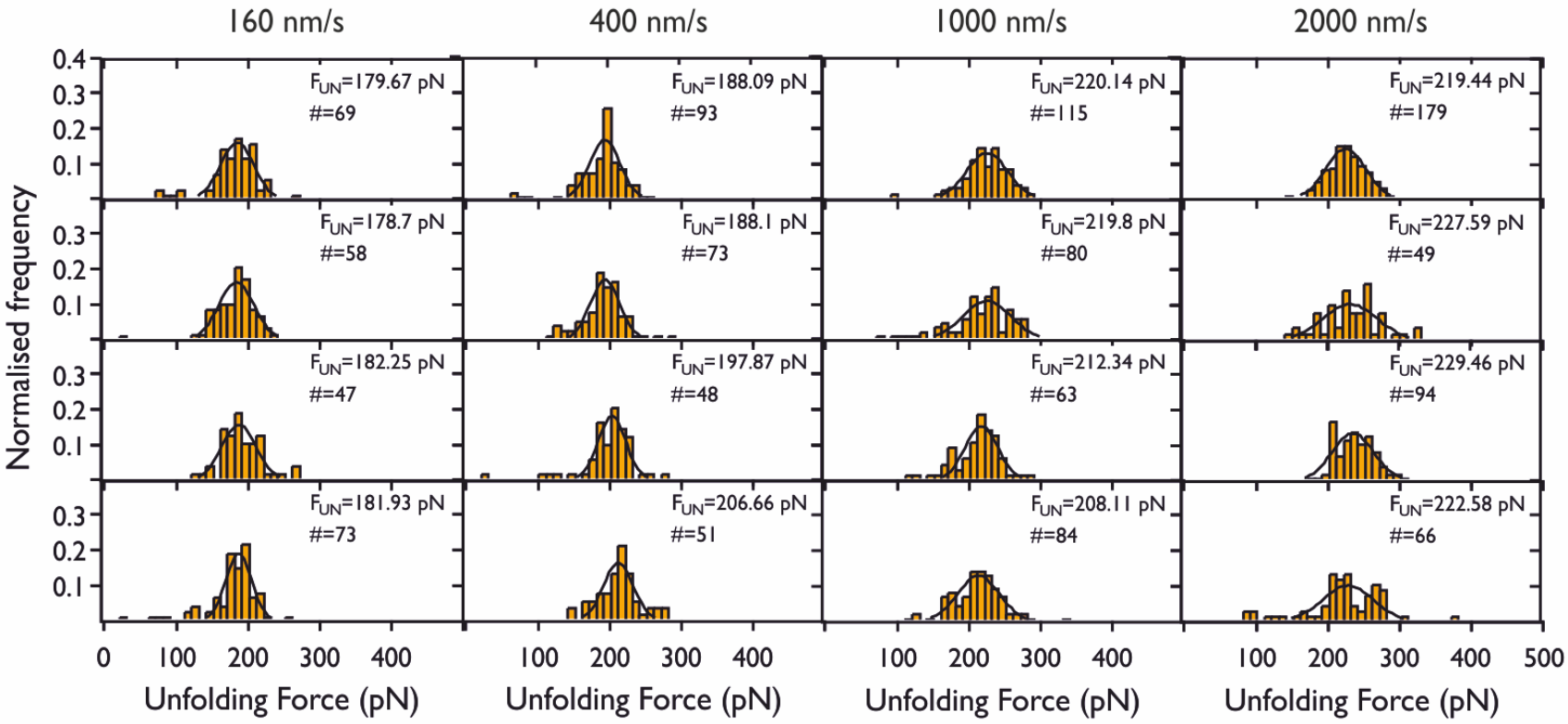


Figure 4.3.0.6: Histograms showing the unfolding force distributions obtained from SMFS experiments on the $p(I27)_5$ in D_2O system. The modal force from a Gaussian fit to the distribution of forces, and number of data points are displayed on each histogram.

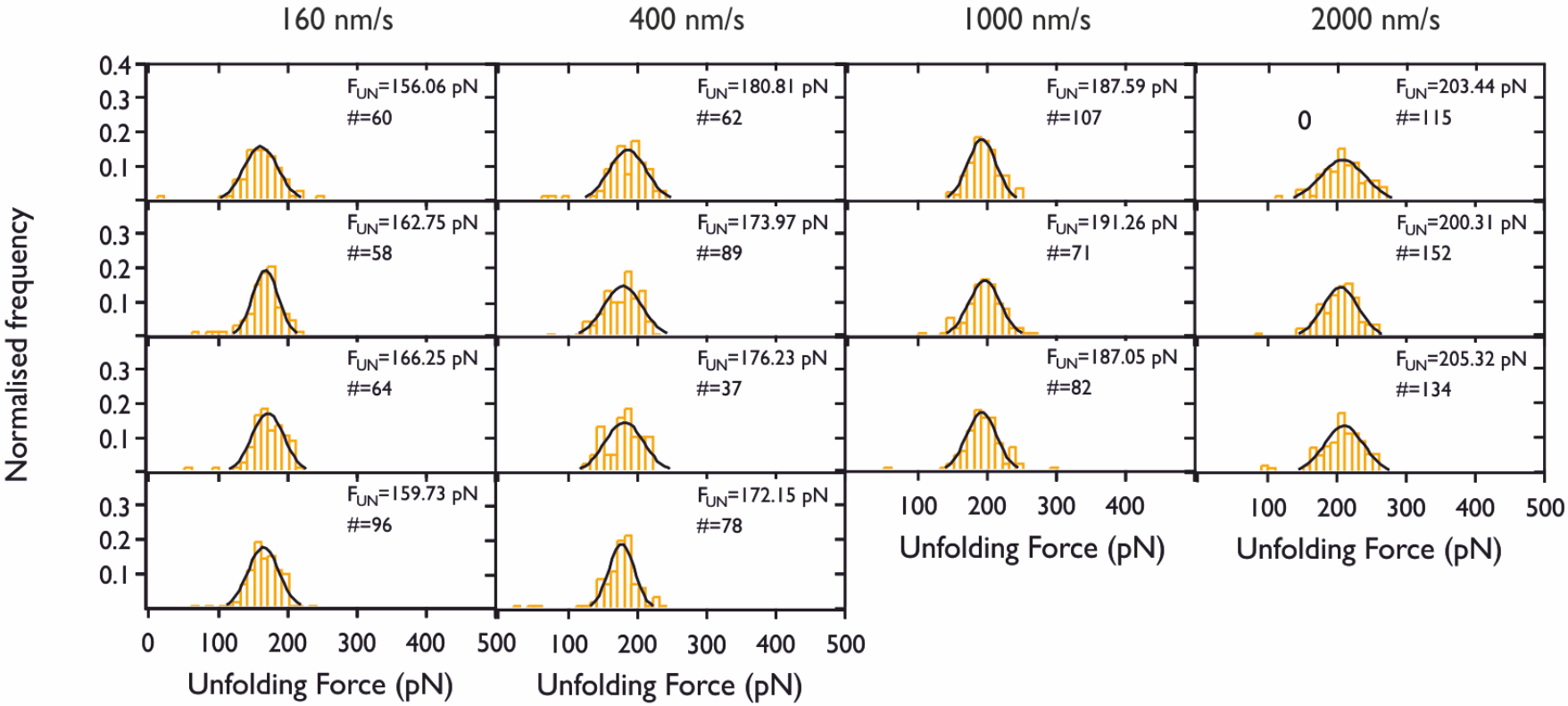


Figure 4.3.0.7: Histograms showing the unfolding force distributions obtained from SMFS experiments on the $d(pL)_5$ in H_2O system. The modal force from a Gaussian fit to the distribution of forces, and number of data points are displayed on each histogram.

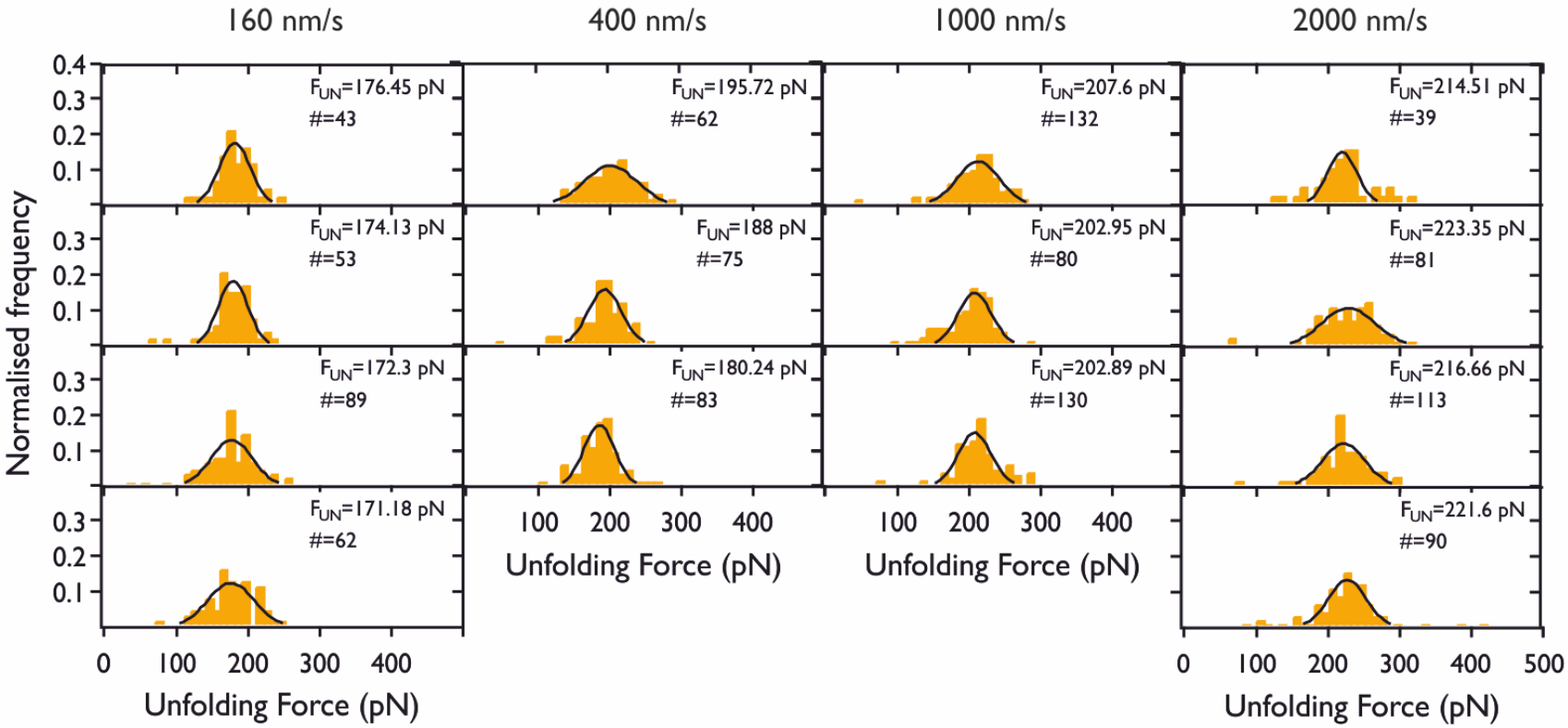


Figure 4.3.0.8: Histograms showing the unfolding force distributions obtained from SMFS experiments on the $d(pL)_5$ in H_2O system. The modal force from a Gaussian fit to the distribution of forces, and number of data points are displayed on each histogram.

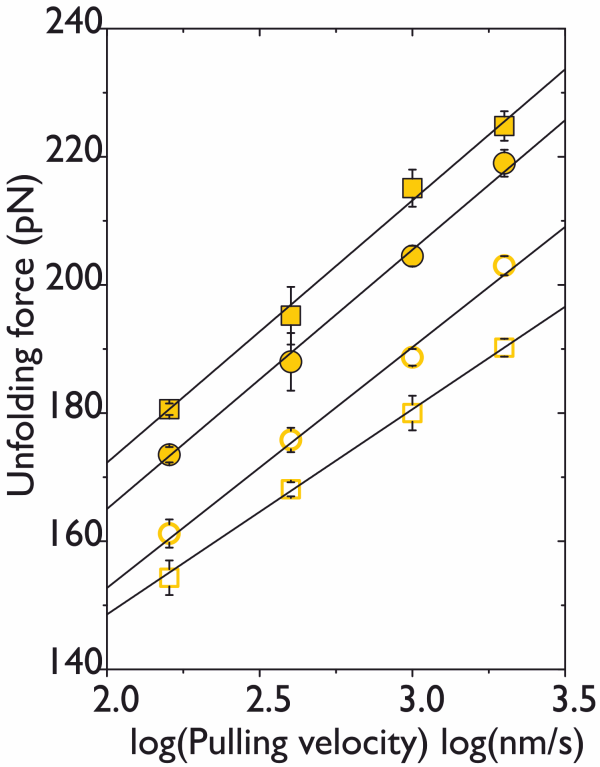


Figure 4.3.0.9: The dependence of force on the logarithm of pulling speed (log(v)). Comparison of the force-log(v) graphs for each of the I27 test systems: $p(I27)_5$ in H_2O (open squares), $p(I27)_5$ in D_2O (closed squares), $d(I27)_5$ in H_2O (open circles) and $d(I27)_5$ in D_2O (closed circles). All are described well by a linear fit.

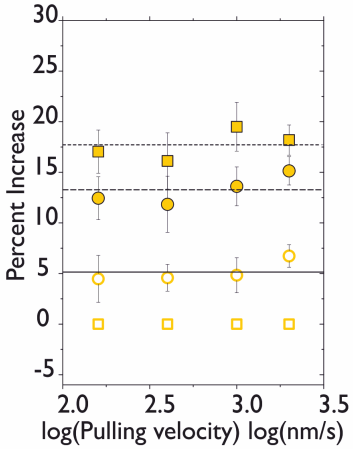


Figure 4.3.0.10: The percent increase in the forces for the test systems compared to I27 where: $p(I27)_5$ in H_2O (open squares), $p(I27)_5$ in D_2O (closed squares), $d(I27)_5$ in H_2O (open circles) and $d(I27)_5$ in D_2O (closed circles). All are described well by a linear fit. The percent increases are determined from the experimental forces. The dotted line corresponds to the average percent increase for the $p(I27)_5$ in D_2O system, the dashed line corresponds to the average percent increase for the $d(I27)_5$ in D_2O system and the solid line indicates the position of the average percent increase in the $d(I27)_5$ in H_2O system.

two state WLC model has previously been employed to determine the position of the force [365]. In this project, the change in gradient was determined along the force-extension profile of an unfolding event. The initial unfolding event should correspond to the minimum gradient along the force-extension profile. This method was easy to employ and did not require selection of the unfolding event by eye. The ratio of the initial unfolding force, F_I, to the intermediate unfolding, F_{UN} was also determined to see whether the hydrophobic interaction was involved in the unfolding of the stable intermediate of the protein. Figure 4.3.0.11 illustrates the change in gradient along the unfolding curve of the protein. The minimum closest to the peak unfolding event was taken as the position of F_I. The gradient was taken over 20 points for each increment along the force-extension trace. The F_I was taken to be the average of the 20 points corresponding to the minimum in the gradient. The initial unfolding force was taken for the first two peaks on the trace (when present) only, unless the third unfolding event had a well defined minimum. The initial unfolding events resulted in better resolution in the position of F_I for 160 nm/s, 400 nm/s and 1000 nm/s therefore 2000 nm/s was omitted from the analysis.

System	F_I*[pN]/ratio$^+$ 160 nm/s	n	F_I[pN]/ratio 400 nm/s	n	F^{F_I}[pN]/ratio 1000 nm/s	n
p(I27) in H_2O	103.7(33.3)/0.75(0.12)	24	116.2(15.8)/0.71(0.10)	88	116.0(27.3)/0.67(0.11)	18
	117.5(32.3)/0.81(0.08)	16	125.2(23.4)/0.73(0.12)	68	121.7(19.9)/0.68(0.11)	68
	125.8(24.5)/0.78(0.10)	38	120.5(21.1)/0.69(0.10)	40	118.6(21.4)/0.66(0.11)	40
	101.8(25.3)/0.70(0.11)	13				
Average (±SE)	112.2(±5.7)/0.76(±0.02)		120.6 (±2.6)/0.71(±0.01)		118.1(±1.8)/0.67(±0.01)	
p(I27) in D_2O	134.5(30.0)/0.77(0.11)	38	139.7(18.9)/0.73(0.15)	48	135.8(31.8)/0.68(0.12)	53
	137.1(21.6)/0.76(0.11)	35	136.4(29.0)/0.77(0.12)	33	137.0(18.5)/0.65(0.11)	23
	135.8(31.9)/0.67(0.13)	34	135.9(27.6)/0.76(0.11)	19	137.0(36.0)/0.65(0.14)	50
	138.3(23.4)/0.75(0.13)	32	140.7(29.3)/0.66(0.14)	34	141.8(25.1)/0.65(0.11)	55
Average (±SE)	136.4(±0.8)/0.74(0.02)		138.2(±1.2)/0.73(±0.03)		137.9(±1.3)/0.66(±0.01)	
d(I27) in H_2O	114.8(29.3)/0.77(0.13)	33	120.3(28.0)/0.74(0.09)	44	130.8(23.8)/0.73(0.11)	57
	123.6(23.9)/0.77(0.10)	32	124.8(14.4)/0.75(0.12)	21	128.1(23.2)/0.69(0.12)	40
	120.0(22.6)/0.77(0.10)	32	121.0(24.0)/0.74(0.11)	54	132.8(29.0)/0.73(0.16)	60
	127.0(27.5)/0.79(0.12)	47	130.4(31.4)/0.72(0.13)	32		
Average (±SE)	121.3(±2.6)/0.77(±0.01)		124.1(±2.3)/0.73(±0.01)		130.6 (±1.4)/0.71(±0.01)	
d(I27) in D_2O	134.2(24.4)/0.78(0.12)	30	136.4(25.7)/0.73(0.13)	53	143.3(27.7)/0.68(0.11)	60
	136.3(30.9)/0.80(0.13)	61	136.2(30.3)/0.77(0.11)	39	133.4(25.5)/0.71(0.12)	43
	135.0(20.6)/0.76(0.10)	19	143.1(25.3)/0.73(0.10)	34	132.2(34.8)/0.64(0.12)	75
	138.9(26.2)/0.76(0.13)	30				
Average (±SE)	136.1(±1.0)/0.77(±1.0)		138.6(±2.3)/0.74(±0.01)		137.3(±3.1)/0.67(±0.02)	

Table 4.3.0.5: Summary of unfolding statistics for the initial unfolding event for all I27 systems.
* F_I refers to median initial unfolding force found by determined from n initial unfolding forces. The force is determined from a minimum in the gradient along the protein unfolding curve in a force-extension trace. The value in brackets is the standard deviation of the median. $^+$ is determined as the median ratio of the initial unfolding force/ peak unfolding force (initial:peak) for every unfolding event. The value in brackets is the standard deviation of the n values. SE is the standard error in the average initial unfolding force or intial:peak ratio.

The results from the analysis on the initial unfolding event for all four I27 systems are shown in Figure 4.3.0.12. The forces obtained for the $p(I27)_5$ in H_2O are lower than those

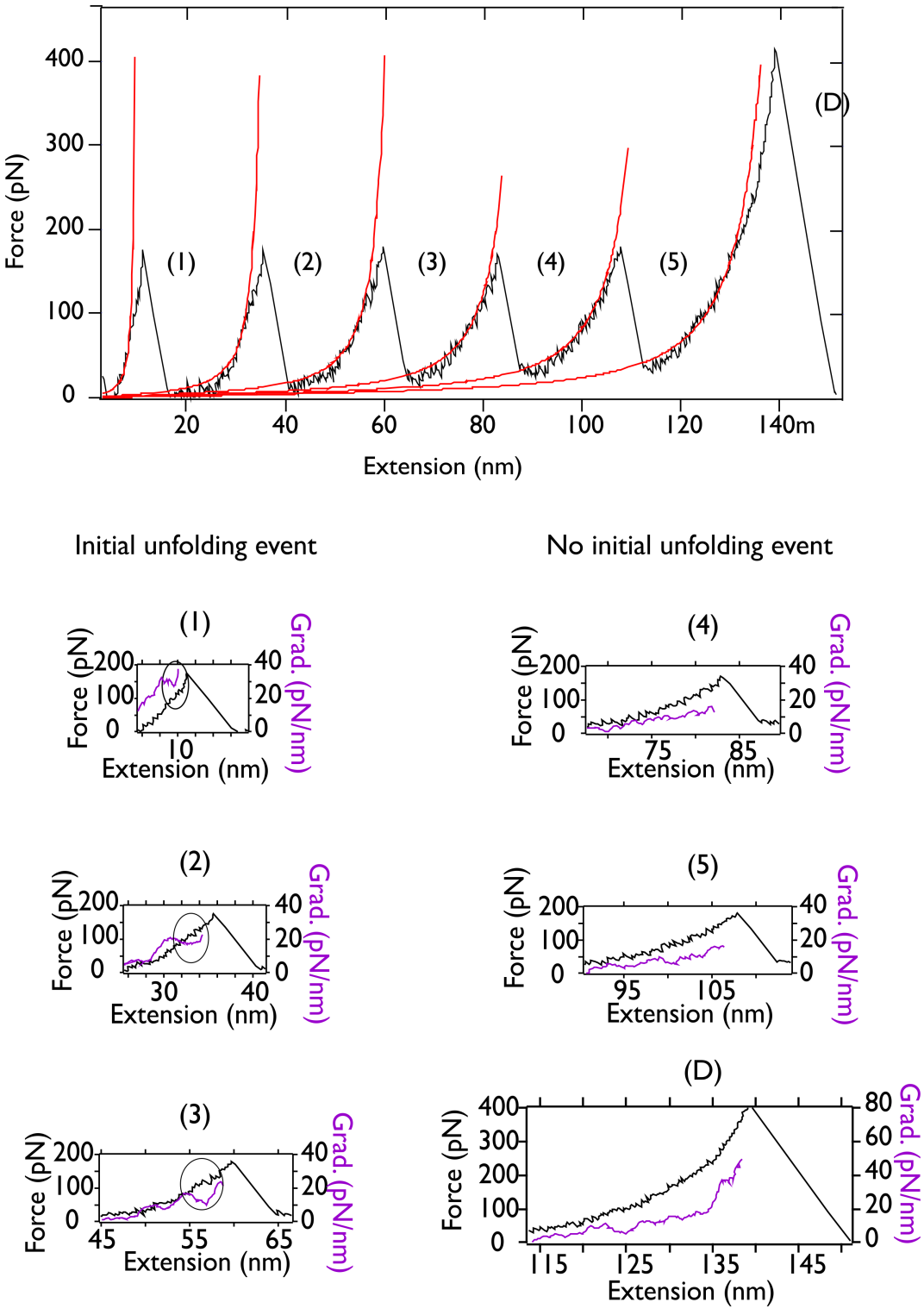


Figure 4.3.0.11: Figure illustrating how the initial unfolding event is determined on a force-extension trace. The force-extension is graph is shown in black, the WLC is shown in red and the change in gradient as a function of extension is shown in purple. The first three unfolding peaks have a clear minimum in the gradient which is taken from initial rise in the force peak, to the unfolding peak (peaks (1)–(3)). The minima in the gradient are circled. Peaks (4), (5) and the detachment (for comparison) do no appear to have this minimum in gradient.

previously obtained ($\sim$ 130 pN [365]). This could be a consequence of using an alternative method to determine the F_I. There is, however, a significant increase ($\sim$ 17 pN compared with the p(I27) in H_2O) in the F_I observed for the proteins in D_2O, this is attributed to an increase in hydrogen bond strength. If one assumes the energetic cost for breaking a hydrogen bond is $\sim$ 4.5 kJ/mol [46] ($\approx$ 6.5 pNnm, see Section 1.3.2 on hydrogen bonds in proteins) to the protein stability, then the estimated force to simultaneously rupture 3 hydrogen bonds $\sim$ 2 $\mathring{A}$ apart would be 114 pN. If the hydrogen bonds within the backbone of the protein involved deuterium instead of hydrogen then the bond strength is expected to increase by $\sim$ 0.4–0.8 kJ/mol [139, 140]. This would lead to a force of $\sim$ 122–132 which is an increase of $\sim$ 15 pN. This is in excellent agreement with the results presented here. The forces for the p(protein) in H_2O, d(protein) in D_2O and p(protein) in D_2O do not fluctuate with velocity, this suggests this unfolding event is in equilibrium at the experimental velocities probed. There appears to be a force-log(v) dependence for the $d(I27)_5$ in H_2O. This could suggest that this event is loading rate dependent and therefore in not in equilibrium. The forces for the other systems are approximately the same within error for all four velocities. This is because the process is reversible and has been observed previously [365]. No change was observed for the ratios for each test system within error.

4.3.1 Discussion

The effect of deuteration on the protein and the solvent has been investigated in this chapter to determine how different interactions affect the mechanical unfolding force. This is the first time a deuterated protein construct has been mechanically unfolded using SMFS. An alternative protiated I27 mutant has been mechanically unfolded in D_2O previously. This protein was found to unfold at a force 17.6 % higher than the protein in H_2O [132]. The results indicate differences in the mechanical response of the proteins in the presence of deuterium. The mechanical stability is most significantly changed as a consequence of increasing the hydrogen bond strength in the solvent in p(protein) in D_2O. The results presented may relate to the differences the mechanical clamp regions of the proteins being probed.

4.3.2 How does deuteration of the solvent affect the mechanical stability of proteins?

The results presented here suggests that isotopically substituting the hydrogen atoms in the exchangeable positions within both I27 and pL, by saturation of the protein in D_2O, results in an increase in unfolding force of a protein. The increase in mechanical unfolding

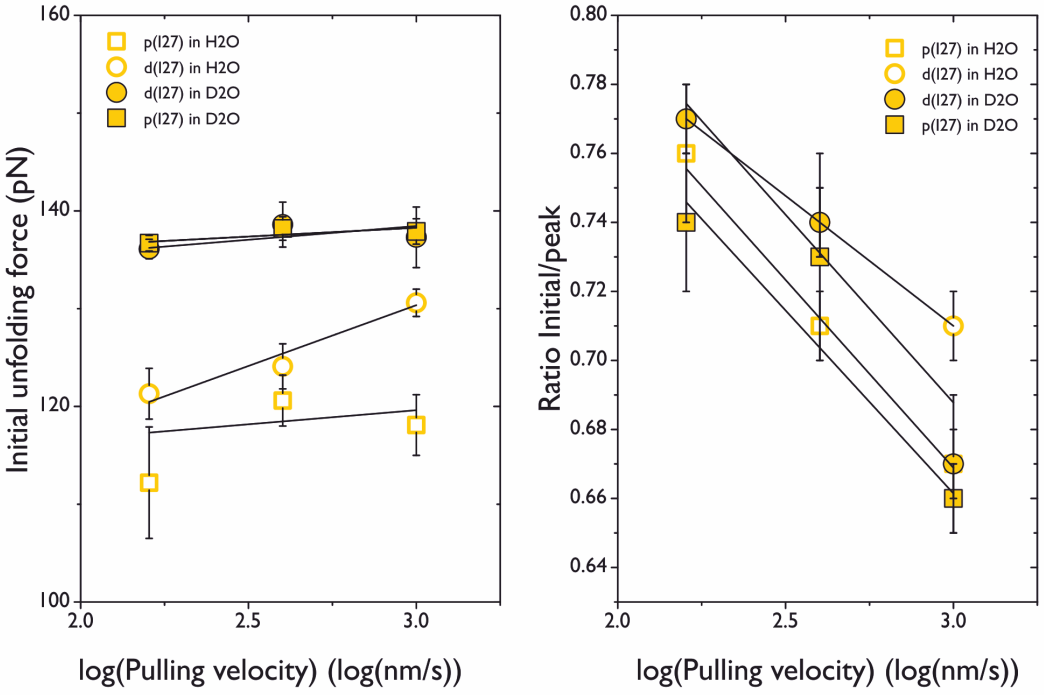


Figure 4.3.0.12: Graph showing the change in force with log(v), where v is the pulling velocity, for the initial unfolding event in a force-extension trace Left: and Right: the change ratio of the initial unfolding force to the peak unfolding force for each test system. The open squares correspond to the $p(I27)_5$ in H_2O, the open circles are the $d(I27)_5$ in H_2O, the closed squares are $p(I27)_5$ in D_2O and the closed circles are the $d(I27)_5$ in D_2O. The error bars for the points are the standard error from multiple experiments.

force for the p(protein) in D_2O is consistent with previous findings [132]. This suggests that it is a result of an increase in hydrogen bond strength and hydrophobicity of the protein. This is consistent with the thermodynamic data obtained for both proteins. The folded state of the protein was more stable for both constructs. The percent increase in the I27 protein was higher than that of protein L ($\sim$ 18 % compared with $\sim$ 14 % respectively) suggesting the effect was more pronounced in I27.

The d(protein) in D_2O was also more mechanically stable for both pL and I27 test systems. This was less significant that the increase in mechanical stability observed for the p(protein) in D_2O suggesting that hydrophobic interactions mechanically stabilise both pL and I27 proteins. An increase in mechanical unfolding force for this protein system, however, could suggest that simply an increase in hydrogen bond strength results in a 7% increase in the F_{UN} for both proteins.

The results from the analysis on the initial unfolding event of I27 provide more evidence that the increase in F_{UN} in D_2O is due to an increase in the strength of intramolecular hydrogen bonds. This increase has been observed previously for an alternative mutant, I27 p(protein) in H_2O, but was more significant than the increase observed here (104 pN – 150 pN, compared to 115 pN – 135 pN). The ratio between the peak height and the initial unfolding event did not lead to conclusive information about the different interactions involved in the unfolding of the intermediate. It is hard to deconvolute this data due to large error bars, this suggests this method for determining the unfolding intermediate may not be sensitive enough to these changes in the system. Furthermore, the initial unfolding event was included in this analysis and it is unknown whether these initial unfolding events will affected by interactions between the protein and the substrate.

This chapter has provided evidence that deuteration of the solvent increases the mechanical stability of the protein. This increase is observed independently of the type of clamp of the protein being studied and the function of the protein. Previous studies have also suggested an increase in rigidity of proteins in D_2O [158]. No significant changes were observed in the WLC fits to the protein unfolding events, however more information can be obtained about the stiffness of the chain from extraction of the energy landscape parameters x_u and $k^0_{u,force}$ for simulations. The x_u has been related to the malleability of protein domains [317] and can indicate changes in the number of degrees of freedom of the protein [378]. This parameter will be determined in Chapter 5.

4.3.3 How does deuterating the C-H bonds affect the mechanical stability of a protein?

Full deuteration of a protein involves deuteration of the protein backbone and sidechains. Typically the deuteration of such hydrogen atoms leads to a destabilised protein in comparison to the protiated proteins [130, 137]. The results here indicate that the effect of protein deuteration to the mechanical stability of proteins is complex.

A decrease in unfolding force was observed for the d(pL) in H_2O system, whereas, an increase in force was observed for the d(I27) in H_2O. The increase in stability observed for the d(I27) in H_2O could be a result of fewer interactions between the water and hydrogen bonds in the I27 protein, or an elevation of a frustrated (not optimally packed) hydrophobic core. However, if it was the latter, the increase of mechanically stability of p(I27) in D_2O might not be as prominent, furthermore the hydrophobic interactions are not expected to play a dominant role in the mechanical stability of I27[263]. A combination of mutational studies with SMFS have provided evidence that the hydrophobic core plays no part in the mechanical stability of I27, whereas chemical denaturation of this protein does depend on the hydrophobic core [263]. The average increase in force of the d(I27) in H_2O, across the four experimental velocities, is only marginal (~ 5 %) therefore changes to the interactions within this system does not appear to be the major contribution to the mechanical stability of the protein. This suggests that the contribution of hydrophobic interactions to the mechanical clamp regions of proteins is context dependent. The pL clamp involved hydrophobic contacts for the mechanical stability [124] which would explain the decrease in unfolding force. The mechanical clamp region of I27, however, does not appear to involve hydrophobic contacts. This has been observed previously for I27. The mechanical unfolding force was not significantly perturbed in TFE or sodium sulphate. TFE solubolises hydrophobic side-chains, reducing the hydrophobicity of proteins, and sodium sulphate reduces the rate of unfolding by stabilising the native contacts in the protein. Despite this, no change was observed when mechanically unfolding the protein in these two solvents [367].

The results presented here suggest that the type of mechanical clamp is sensitive to the solvent environment. If hydrophobic interactions are involved in a mechanical clamp motif it is hypothesised that similar results as d(pL) in H_2O would be observed. However, if the mechanical clamp region is dominated by hydrogen bonds, then a protein is expected to behave more similarly to I27. In mutational studies, the structure of the

transition state can be determined by ϕ-analysis [263, 264]. This will be explored in chapter 5 to determine if hydrophobic contacts have a significant effect on the structure of the transition state for the pL protein but not for I27 as hypothesised.

Chapter 5

Effect of deuterium on the mechanical unfolding pathway of proteins

5.1 Introduction

In this chapter theoretical models and computational modelling are used to extract information from the experimental SMFS data obtained in Chapter 4. The MC simulations are discussed in Section 2.3.9.1. The four systems are the p(protein) in H_2O, p(protein) in D_2O, d(protein) in D_2O and d(protein) in H_2O. These were described in more detail in Chapter 3. The two model systems chosen to study, and their mechanical properties, were described in Chapters 3 and 4 respectively. The MC simulations will be used to obtain an estimate of the x_U, the distance between the folded and the transition state, and $k^0_{u,force}$, the unfolding rate at zero force. In the initial MC simulations, it is assumed that there is an equal probability of observing two, three, four or five unfolding events in experiments[124]. However, in SMFS it has been shown that this is an oversimplification[379–382]. An additional model is considered which also takes this unequal unfolding probability into account.

When a polyprotein, containing five repeats of a protein, is "picked-up" during an SMFS experiment, between zero and five unfolding events can be observed in the force-extension trace. Figure 5.1.0.1 shows four example scenarios of why this might occur [379]. The stochasticity of this pick-up event in the experiment results in the total number of traces containing zero, one, two, three, four or five protein unfolding events to differ in each individual SMFS experiment. The most frequent trace ob-

Figure 5.1.0.1: Schematic diagram illustrating why unfolding of the full construct might not always be observed during a SMFS experiment. When the cantilever tip is initially brought into contact with a surface, a protein may adsorb to the tip. This figure illustrates different possible ways a protein may be adsorbed onto the tip. (From left to right) In the first diagram, all five protein unfolding events may be observed because the tip has adsorbed to one end of the polyprotein, while the other end is attached to the protein surface. In the next diagram, two protein domains have non-specifically adsorbed to the surface. Therefore a maximum of three protein unfolding events will be observed. In the third diagram, the tip has adsorbed the protein along the polypeptide chain, therefore only three unfolding events can be observed. In the final schematic, the polyprotein detaches from the tip before full unfolding has occurred. Reprinted figure with permission from ([379]) as follows: Katarzyna M. Tych, Megan L. Hughes, James Bourke, Yukinori Taniguchi, Masaru Kawakami, David J. Brockwell, and Lorna Dougan, Optimizing the calculation of energy landscape parameters from single-molecule protein unfolding experiments, 91, 012710 and 2017. Copyright (2015) by the American Physical Society.

served contains no protein unfolding events. As these force-extension traces contain no information about protein unfolding events, they are not considered for analysis. However, the number of these traces can be used as a guide to optimise the experiment and increase the number of protein unfolding events observed. Force-extension traces in the experiments reported here are only accepted when they contain two or more protein unfolding events. This therefore results in an unequal unfolding probability distribution for traces containing two, three, four or five unfolding events. In most experiments, the number of observed traces containing two unfolding events is substantially higher than observing five unfolding events [379]. In an earlier study on I27, the probability of observing more than two unfolding events decreased by $\sim$ 40 % of the total per unfolding event for a pentameric polyprotein [379]. These unequal probability distribution of events has also been observed in force-clamp spectroscopy experiments [381, 382]. In this experiment, for a polyprotein consisting of 12 domains, the dwell time of the polyprotein chain was dependent on the number of unfolding events "picked-up" [381].

If the forces observed in experimental traces were independent of the number of do-

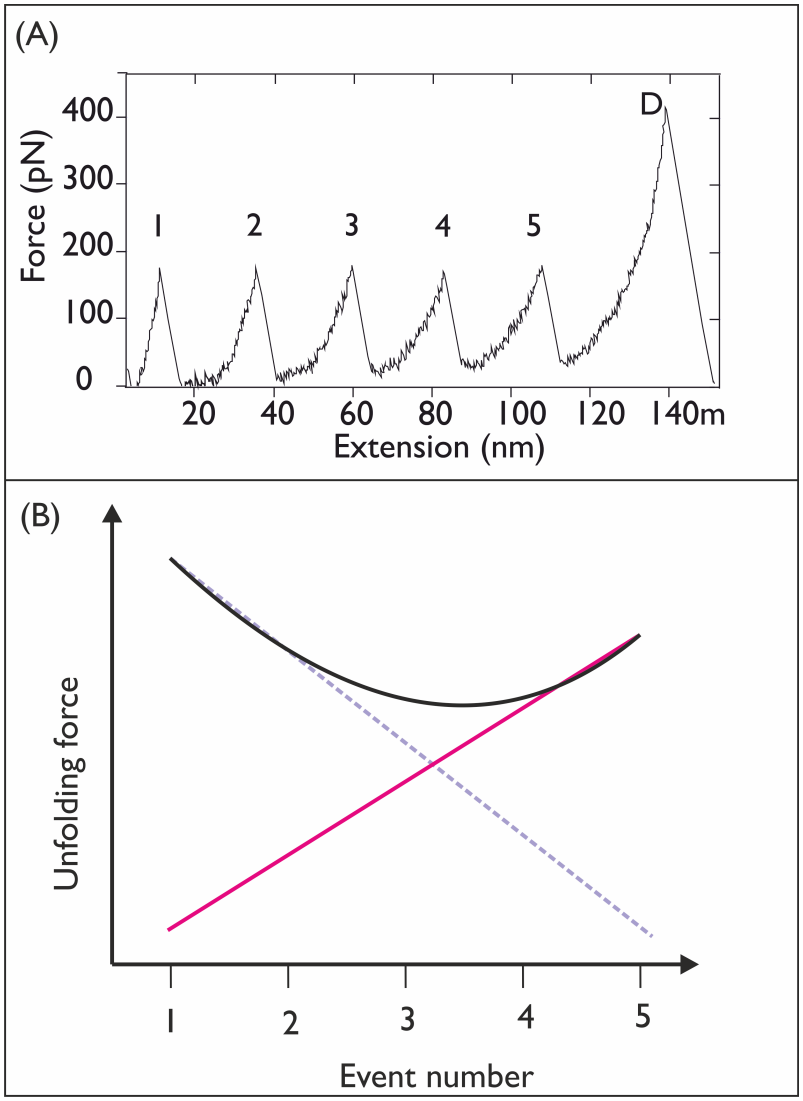


Figure 5.1.0.2: Shematic illustrating the unfolding history effect on the unfolding force of protein domains within a polyprotein. (A) Figure showing a force extension trace for a polyprotein containing five I27 domains. Each unfolding event is numbered from 1–5, 1 indicates the first unfolding protein event observed, and 5 is the unfolding event corresponding to the fifth protein unfolding event. D corresponds to the detachment of the chain from the tip or substrate. (B) Schematic showing the expected unfolding force – event number relationship due to two competing effects illustrated as the pink and blue straight lines. Pink - as protein domains unfolding, the number of folded domains decreases. This decreases the probability of observing an unfolding event at a given force and therefore acts to increase the unfolding force observed. Blue: As the protein domains unfold, the compliance of the chain increases. This decreases the effective spring constant of the chain and therefore acts to decrease the unfolding force of the chain. The combination of the two effects results in a non-linear relationship between the unfolding force and event number [328].

mains picked up by the cantilever, an equal unfolding distribution would suffice to describe the data. However, it has previously been shown that the unfolding force of a single domain depends on the "unfolding history" of a polyprotein [328]. Each individual force peak can be given an event number as shown in Figure 5.1.0.2 (A). The force of each individual unfolding event (i.e, 1, 2, 3, 4, 5) can be measured for a number of traces at a particular pulling velocity. A minimum has been observed in the graph of modal unfolding force against event number (Figure 5.1.0.2). This suggested that the unfolding force of a single protein domain within a polyprotein construct was dependent on the preceding unfolding events [328]. The minima in the forces occurs due to the competition between two effects:

1. The number of protein domains that have unfolded previously and therefore, the number of protein domains which remain folded

2. The stiffness of the supramolecular scaffold: This depends on the cantilever spring constant and the polyprotein domain (i.e. the compliance of the chain)

Figure 5.1.0.2 (B) illustrates these effects. These individually influence the loading rate applied to each folded protein domain; 1. acts to increase the loading rate and 2. acts to decrease it [328, 379]. As the number of unfolding domains increases, the probability of a protein unfolding at a given force decreases. This suggests that the unfolding force would increase with event number. However, as more protein domains unfold, the length, and therefore compliance of the chain, increases. The increase in compliance decreases the effective spring constant of the polyprotein and therefore decreases the force required to unfold a protein domain. The MC simulations used in SMFS experiments incorporate this effect by measuring the effective spring constant of the protein at every timestep [328]. Using the MC simulations this unfolding history has been shown to be sensitive to the number of domains in a polyprotein [328, 383]. Therefore, the unfolding force will be different if two protein domains were picked up as opposed to five. Incorporating the unequal unfolding probability observed in the SMFS experiments into the analysis could be important.

An MC simulation has been developed to include the experimentally observed unfolding probability distrubution (Figure 5.1.0.1) [379]. In the study by Tych et al. simulations were used to obtain the energy landscape parameters for an $(I27)_5$ protein using two different simulations. In the simulations using an equal unfolding probability, the simulations used a random number generator to determine whether the simulations observed two, three, four or five proteins unfolding events. This allowed for the observation two, three, four

or five events to be statistically equally probable. The unequal probability distribution was computed by averaging the percent observation of each trace containing two, three, four or five protein unfolding events in each experiment. These percentages were used to directly determine the number of two, three, four and five unfolding events observed in the simulation with the assumption that the polyprotein, in the experiment, did not detach from the tip before all unfolding events were observed. This study showed that the changes in energy landscape parameters observed using both simulations for the I27 protein were small in these simulations [379]. The equal unfolding distribution resulted in an x_u (distance between the folded state and the transition state) of 0.29 nm and a $k^0_{u,force}$ (the rate of unfolding at zero force) of 0.0011 s^{-1}[379]. The simulations with an unequal unfolding distribution resulted in an x_u of 0.28 nm and $k^0_{u,force}$ of 0.0021 s^{-1}[379]. In this study the incorporation of the unequal unfolding probability did not change the energy landscape parameters obtained from the simulations. However, the spring constant has also been shown to affect the unfolding history of the polyprotein [328]. The simulations in this chapter aim to mimic the experimental uncertainty for each experiment independently. Furthermore, it will incorporate the individual experimental spring constant.

5.1.1 Objectives

The aim of this chapter was to explore a number of different theoretical and computational models to determine the effect of deuterium on the mechanical properties of pL and I27. Specifically the objectives of this chapter are as followed:

1. Obtain the energy landscape parameters of the mechanical unfolding energy landscape for the I27 and pL proteins from MC simulations.

2. Determine whether the experimentally observed unequal unfolding distribution affects the the energy landscape parameters from the MC simulations.

3. Perform Φ-value analysis on the data to determine whether the structure of the protein is affected at the transition barrier.

5.2 Obtaining the mechanical unfolding energy landscape parameters from MC simulations

MC simulations, using the Bell model, (see Section 2.3.9.1) were performed by computing the unfolding force for each of the four experimental velocities (160 nm/s, 400 nm/s, 1000 nm/s and 2000 nm/s) over incremental steps of x_u and $k^0_{u,force}$ as described in Section 2.3.9.1. These simulations were performed for both I27 and protein L. The log(v) against

unfolding forces (from the MC) were described well by a linear fit F(log(v):x_u, k_u^0) for all pairs of x_u and $k_{u,force}^0$. As described in Section 2.3.9.1 the error from fitting the linear fits from the simulated data to the observed forces was calculated using Equation 5.1[332]:

$$E(x_u, k_u^0) = \frac{1}{v_2 - v_1} \int_{v_1}^{v_2} \left[\frac{\left(F(\log(v); x_u, k_u^0) - F_{bestift}(\ln(v))\right)}{F_{bestift}(\ln(v))} \right]^2 dv \qquad (5.1)$$

Two methods were used to determine the line of best fit to the data and the errors associated with x_u, $k_{u,force}^0$:

1. **Method 1:** The residuals between the experimental data and F(log(v):x_u, k_u^0) were computed. The function that best fit the data, F(log(v):best fit), was the F(log(v):x_u, k_u^0) that minimised the squared sum of residuals. The 95 % confidence interval of the F(log(v):best fit) to the data was then computed. This enabled the straight lines that span the 95 % confidence interval to be computed, F(log(v):95,min) and F(log(v):95,max). The range of accepted error was determined by:

$$E(x_u, k_u^0) = \frac{1}{v_2 - v_1} \int_{v_1}^{v_2} \left[\frac{(F(\log(v); 95, min[ormax]) - F_{bestift}(\ln(v)))}{F_{bestift}(\ln(v))} \right]^2 dv \quad (5.2)$$

   The errors F(log(v):95,min) and F(log(v):95,max) were identical. The F(log(v):x_u, k_u^0) were then compared to F(log(v):best fit) using Equation 5.1. Values of x_u, k_u^0 were accepted if they lay within the error determined from the 95 % confidence interval.

2. **Method 2:** A fixed x_u was assumed for the second method. The x_u was taken as the average x_u that determined the F(log(v):best fit) for the four protein L or I27 test systems. The F(log(v):best fit) for this method was taken as the $k_{u,force}^0$ at the given x_u that minimised squared sum of residuals. The error was determined as above.

Once the $k_{u,force}^0$ was obtained, the $\Delta G_{u,force}^0$ could be computed from Equation 5.3[216]. The prefactor is assumed to be A =1×10^6 s^{-1} for all I27 and protein L test systems in H_2O [384]. The prefactor has been determined to be inversely proportional to the viscosity of the solvent. The prefactor, for the proteins mechanically unfolded in D_2O, was corrected for the change in viscosity of D_2O, η_{D_2O} (1.113 millipoise [146]), in comparison to H_2O, η_{H_2O} (0.896 millipoise [146]), by scaling by using ($A_{D_2O} = A\eta_{D_2O}/\eta_{H_2O}$) [132, 385].

$$\Delta G^0_{u,force} = -RT \ln \left(\frac{k^0_{u,force}}{A} \right) \tag{5.3}$$

If the energy well of the folded protein is assumed to be parabolic and the shape of the curve unpertubed by force (see Figure 2.3.9.1), the spring constant of the protein can also be determined using Equation 5.4[213]. The spring constant will indicate whether the malleability of the protein has been significantly perturbed by deuteration. It has previously been shown that the rigidity of the protein is affected by D_2O [158].

$$D = \frac{2\Delta G^0_{TS,force}}{x_u^2} \tag{5.4}$$

5.2.1 The four protein L test systems

x_u^i (nm)	x_u^f (nm)	Δx_u (nm)	$k^i_{u,force}$ s^{-1}	$k^f_{u,force}$ s^{-1}	$\Delta k^0_{u,force}$ s^{-1}
0.18	0.30	0.01	0.00005	0.17000	0.00005
0.18	0.30	0.0025	0.00005	0.20000	varied

Table 5.2.1.1: The grid of parameters used in the MC simulations of the four pL test systems. x_u is the distance between the folded and transition states and $k^0_{u,force}$ is the rate of unfolding at zero force. In this table i indicates the initial value and f indicates the final value. The step size for the second row of the table was varied logarimically (0.01 logarithmic steps between $\log(k^i_{u,force})$ and $\log(k^f_{u,force})$. This was done to optimise the simulation time. The unfolded length of the protein was taken to be 22.25 nm [217], the folded length was 3.7 nm [217], the average linker lengths were 2.8 nm [217], the temperature was 23 °C and the number of iterations was 1000.

The parameters presented in Table 5.2.1.1 were used in the MC simulations. The MC algorithm was iterated 1000 times for each pair of x_u and $k^0_{u,force}$ parameters to obtain unfolding forces at each of the four experimental velocities.

Method 1

The best fit parameters for each of the protein L systems using the different method are presented in Table 5.2.1.2. The shaded regions in Figures 5.2.1.1–5.2.1.4 show the distribution of accepted values of x_u and $k^0_{u,force}$ using Method 1. The error distribution is elliptical in x_u and $k^0_{u,force}$ for all pL test systems which agrees with previous studies [217, 264]. The reason the error is not continuous is due to the step size used for x_u. When the step size of x_u is reduced, the contours between the error distribution become more smooth. The best fit parameters for the p(pL) in H_2O, x_u=0.22 nm and $k^0_{u,force}$=0.0551 s^{-1}, agree well with the literature values of x_u=0.22 nm and $k^0_{u,force}$=0.05 s^{-1} [217]. Furthermore similar parameters were obtained by running the simulations with

a smaller stepsize (x_u=0.22 nm, $k^0_{u,force}$=0.0575 s^{-1}) suggesting the stepsize used for this system is valid.

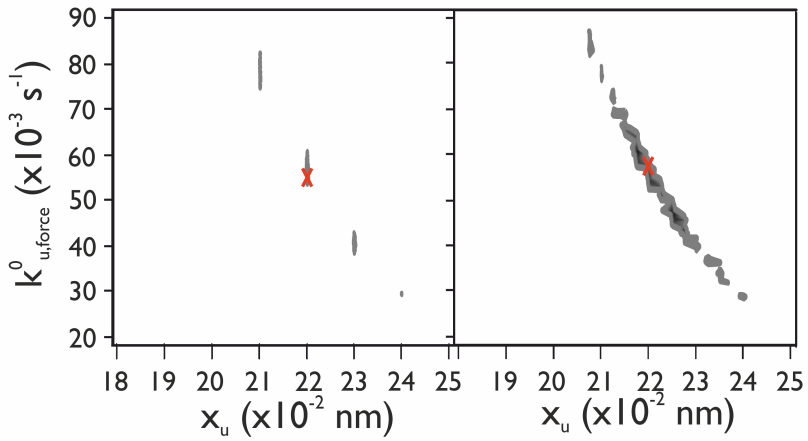


Figure 5.2.1.1: Graph of the accepted $k^0_{u,force}$ against x_u for the p(pL) in H_2O system using an x_u stepsize of 0.1 nm (left) and 0.025 nm (right). The stepsize on the left for $k^0_{u,force}$ is 0.00005 s^{-1}. The $k^0_{u,force}$ were varied on a logarithmic scale with a logarithmic step-size of 0.01. These parameters have errors within the 95 % confidence interval defined by the best fit parameters (red cross) obtained from MC simulations. Darker grey regions of the contour graph indicated values with errors close to 0.

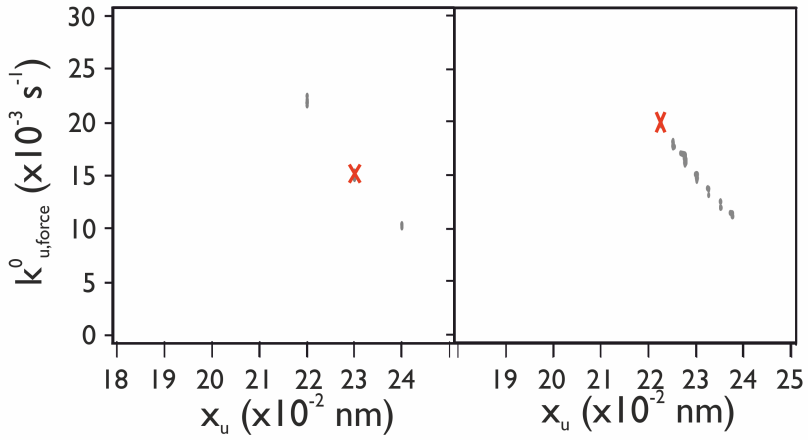


Figure 5.2.1.2: Graph of the accepted $k^0_{u,force}$ against x_u for the p(pL) in D_2O system using an x_u stepsize of 0.1 nm (left) and 0.025 nm (right). The stepsize on the left for $k^0_{u,force}$ is 0.00005 s^{-1}. The $k^0_{u,force}$ were varied on a logarithmic scale with a logarithmic step-size of 0.01. These parameters have errors within the 95 % confidence interval defined by the best fit parameters (red cross) obtained from MC simulations. Darker grey regions of the contour graph indicated values with errors close to 0.

The x_u of the four test systems are the same within error. Therefore the transition state has not moved with respect to the p(pL) in H_2O by deuteration of the protein and solvent. The rate of unfolding at zero force has changed significantly for the p(pL) in D_2O. The rate is three-fold smaller that the p(pL) in H_2O. This results in an increase in barrier height of 3.4 kJ/mol. Despite the forces of the d(pL) in H_2O being 7 % lower than

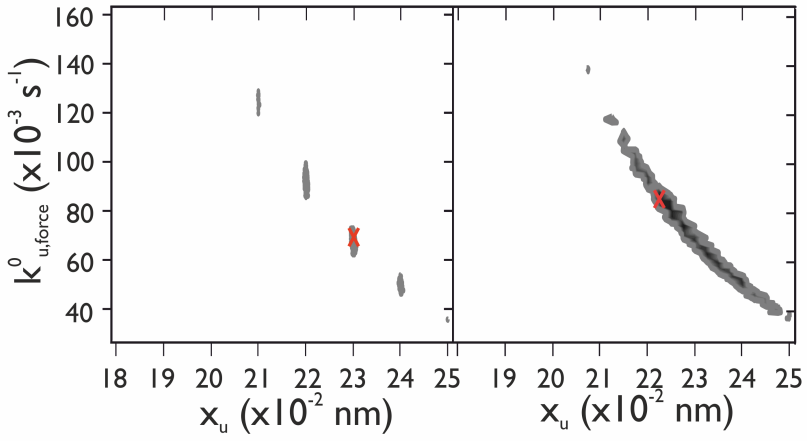


Figure 5.2.1.3: Graph of the accepted $k^0_{u,force}$ against x_u for the d(pL) in H_2O system using an x_u stepsize of 0.1 nm (left) and 0.025 nm (right). The stepsize on the left for $k^0_{u,force}$ is 0.00005 s^{-1}. The $k^0_{u,force}$ were varied on a logarithmic scale with a logarithmic step-size of 0.01. These parameters have errors within the 95 % confidence interval defined by the best fit parameters (red cross) obtained from MC simulations. Darker grey regions of the contour graph indicated values with errors close to 0.

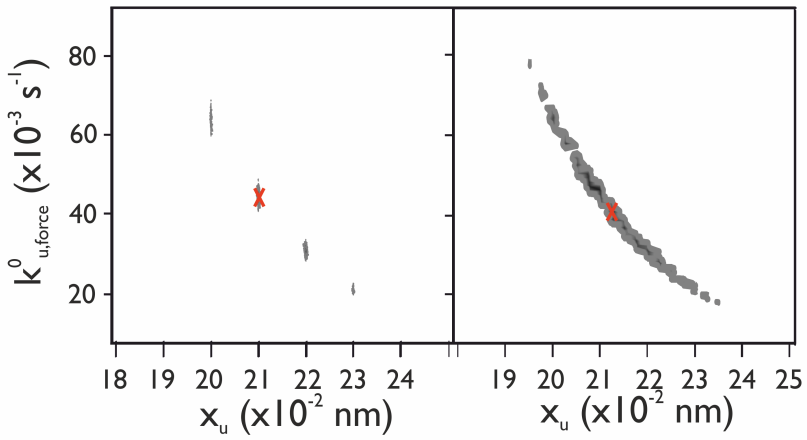


Figure 5.2.1.4: Graph of the accepted $k^0_{u,force}$ against x_u for the d(pL) in D_2O system using an x_u stepsize of 0.1 nm (left) and 0.025 nm (right). The stepsize on the left for $k^0_{u,force}$ is 0.00005 s^{-1}. The $k^0_{u,force}$ were varied on a logarithmic scale with a logarithmic step-size of 0.01. These parameters have errors within the 95 % confidence interval defined by the best fit parameters (red cross) obtained from MC simulations. Darker grey regions of the contour graph indicated values with errors close to 0.

p(pL) in H_2O, the barrier height is not significantly perturbed. In fact the $\Delta\Delta G^0_{TS,force}$ is only 0.26 kJ/mol (where $\Delta\Delta G^0_{TS,force}$ = p(pL) in H_2O $\Delta G^0_{TS,force}$ - d(pL) in H_2O $\Delta G^0_{TS,force}$). Similarly the barrier height is not significant perturbed, within error, for the d(pL) in D_2O system in comparison to the p(pL) in H_2O ($\Delta\Delta G^0_{TS,force}$=0.02 kJ/mol). It should be noted that the changes are small between the $\Delta\Delta G^0_{TS,force}$ but more significant between the $k^0_{u,force}$ due to the magnitude of the prefactor. The spring constants (presented in Table 5.2.1.2) using the parameters obtained from this method indicate an insignificant change in the stiffness of the pL test systems.

Variable	System			
Method 1	p(pL) in H_2O	p(pL) in D_2O	d(pL) in H_2O	d(pL) in D_2O
B.f. x_u (nm)	0.22	0.23	0.23	0.21
Min x_u (nm)	0.21	0.22	0.22	0.20
Max x_u (nm)	0.24	0.24	0.25	0.23
Average x_u (nm)	0.22	0.23	0.23	0.21
Sd x_u (nm)	0.01	0.01	0.01	0.01
B.f. $k^0_{u,force}$ (s^{-1})	55.1×10^{-3}	15.2×10^{-3}	69.1×10^{-3}	44.4×10^{-3}
Max $k^0_{u,force}$ (s^{-1})	82.3×10^{-3}	22.5×10^{-3}	129.0×10^{-3}	68.9×10^{-3}
Min $k^0_{u,force}$ (s^{-1})	22.7×10^{-3}	10.1×10^{-3}	34.5×10^{-3}	20.1×10^{-3}
Average $k^0_{u,force}$ (s^{-1})	56.6×10^{-3}	15.8×10^{-3}	76.5×10^{-3}	43.9×10^{-3}
Sd $k^0_{u,force}$ (s^{-1})	13.4×10^{-3}	4.1×10^{-3}	20.5×10^{-3}	14.8×10^{-3}
Av. $\Delta G^0_{TS,force}$ (kJ/mol)	41.1	43.8	40.4	41.3
Sd $\Delta G^0_{TS,force}$ (kJ/mol)	0.6	0.7	0.7	0.9
Av. D (N/m)	2.8	2.8	2.6	3.1
Sd D (N/m)	0.1	0.1	0.2	0.2
Method 2	p(pL) in H_2O	p(pL) in D_2O	d(pL) in H_2O	d(pL) in D_2O
B.f. x_u (nm)	0.22	0.22	0.22	0.22
B.f. $k^0_{u,force}$ (s^{-1})	55.1×10^{-3}	22.5×10^{-3}	93.0×10^{-3}	32.7×10^{-3}
Max $k^0_{u,force}$ (s^{-1})	60.5×10^{-3}	25.0×10^{-3}	99.9×10^{-3}	34.5×10^{-3}
Min $k^0_{u,force}$ (s^{-1})	52.2×10^{-3}	20.6×10^{-3}	82.4×10^{-3}	28.7×10^{-3}
Average $k^0_{u,force}$ (s^{-1})	56.5×10^{-3}	22.7×10^{-3}	91.7×10^{-3}	31.6×10^{-3}
Sd $k^0_{u,force}$ (s^{-1})	1.8×10^{-3}	1.0×10^{-3}	3.9×10^{-3}	1.5×10^{-3}
Av. $\Delta G^0_{TS,force}$ (kJ/mol)	41.07	42.79	39.88	41.98
Sd $\Delta G^0_{TS,force}$ (kJ/mol)	0.08	0.11	0.11	0.12
Av. D (N/m)	2.82	2.93	2.74	2.88
Sd D (N/m)	0.01	0.01	0.01	0.01

Table 5.2.1.2: Statistics obtained from MC simulations for the four pL test systems: p(pL) in H_2O, p(pL) in D_2O, d(pL) in H_2O, d(pL) in D_2O. Two methods were used to obtain the parameters x_u; the distance between the unfolded and transition state in a two-state unfolding energy landscape, $k^0_{u,force}$; the rate of unfolding at zero force, $\Delta G^0_{TS,force}$; the height of the transition barrier to unfolding and D; the spring constant of the protein. Method 1 determined the best fit to the experimental data, allowing x_u to vary. Method 2 determined the best fit to the experimental data assuming a fixed x_u. B.f. indicates the best fit parameters. Av. is the average of the accepted x_u and $k^0_{u,force}$, Sd is the standard deviation of the values. The Av. and Sd $\Delta G^0_{TS,force}$ was the average and standard deviation taken from calculating $\Delta G^0_{TS,force}$ for every accepted $k^0_{u,force}$.

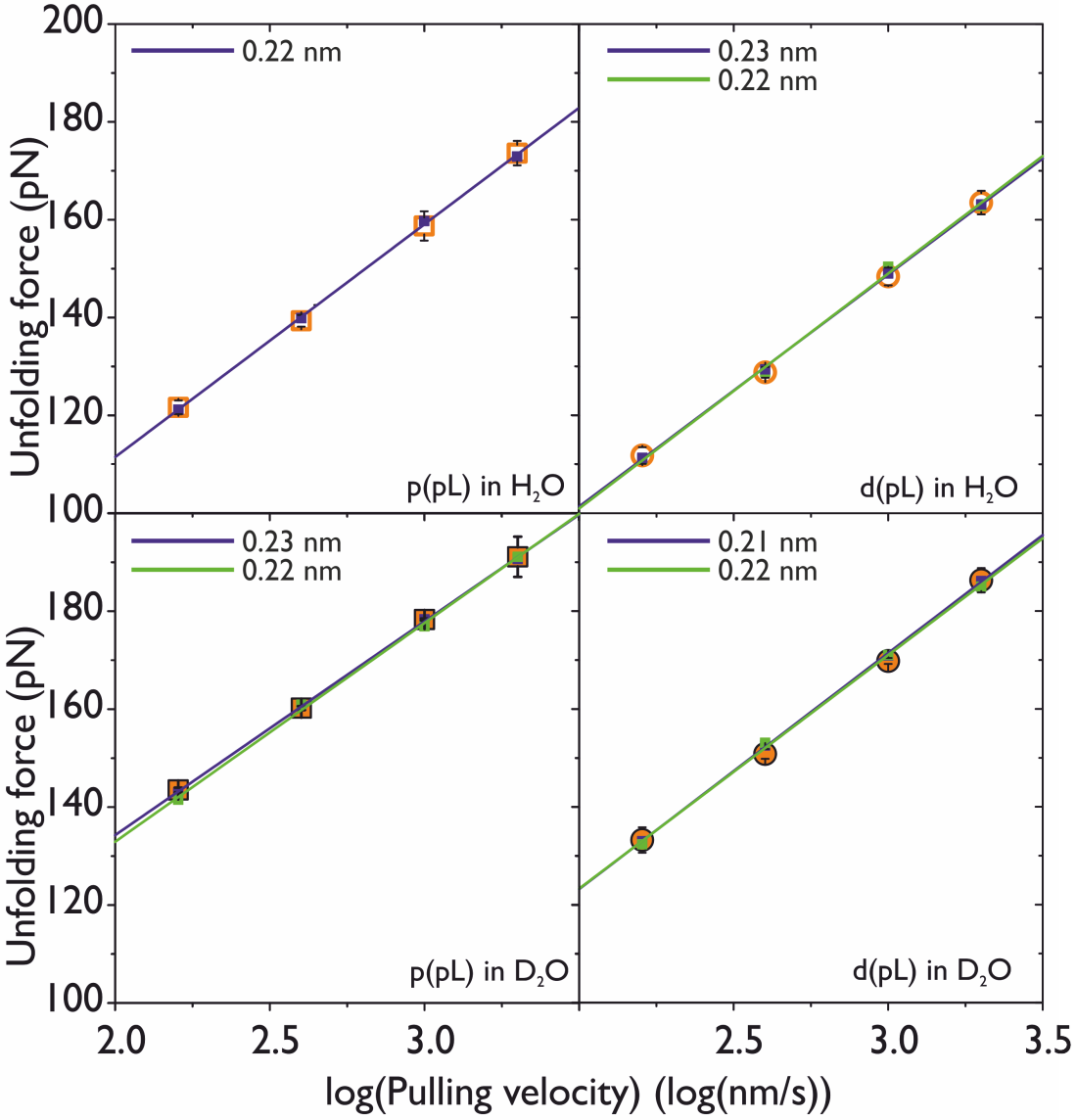


Figure 5.2.1.5: Graph of Unfolding force against log(Pulling velocity) for the four pL test systems (as labelled) comparing the experimental and simulated data determined by two methods. The numbers next to the line indicate the associated x_u and the method used to determine the parameters. Blue: The straight line fit to the simulated data for pair of x_u and $k^0_{u,force}$ from method 1 that minimises the error to the experimental data. These are the best fit parameters to the experimental data. This line minimises the least squares of a weighted fit to the experimental data. The associated x_u are labelled. Green: Straight line fits to the simulated forces that minimise the error to the experimental data for using method 2 with a fixed x_u of 0.22 nm. The associated best fit $k^0_{u,force}$ are presented in Table 5.2.1.2. The graph indicates close agreement between the experimental and simulated data for both methods.

Method 2

Table 5.2.1.2 give the parameters obtained using the second method described above where the x_u was fixed to 0.22 nm. This significantly reduced the error on the $k^0_{u,force}$ and therefore, the $\Delta G_{TS,force}$. This results in a large decrease in error for each of the systems. In fact for the d(pL) in D_2O the decrease is nearly 10 fold. This results in larger differences in the barrier heights of the four protein L test systems. However, it is only possible to fix x_u if the gradients are similar. A significant difference in gradient can suggest a change in x_u[264]. In the protein L test systems the values of x_u were similar for both methods implemented. Figure 5.2.1.5 illustrate the similarities observed between the best fit parameters obtained by the two methods and the experimental data. There is close agreement between the observed unfolding forces in the experiment and simulation. The spring constants calculated from this method are presented in Table 5.2.1.2. The results suggest a small increase of $\sim$ 0.1 N/m in the spring constants of both protiated and dueterated proteins in D_2O.

5.2.2 The four I27 test systems

x_u^i (nm)	x_u^f (nm)	Δx_u (nm)	$k^i_{u,force}$ s^{-1}	$k^f_{u,force}$ s^{-1}	$\Delta k^0_{u,force}$ s^{-1}
0.16	0.34	0.01	0.00005	0.02000	0.00005
0.185	0.325	0.005	0.00005	0.01500	0.000005

Table 5.2.2.1: The grid of the x_u and $k^0_{u,force}$ parameters used in the MC simulations of the I27 test systems. Where i indicates the initial value, f indicates the final value and Δ is the incremental step size used. The unfolded length of the protein was taken to be 28 nm [214], the folded length was 4.2 nm (determined from the PDB structure in PyMOL [10]), the average linker lengths were 2.4 nm, the temperature was 23 °C and the number of iterations was 1000.

The parameters in Table 5.2.2.1 were used in the MC simulations and 1000 iterations of the simulation were run for I27. The best fit parameters for each method are presented in Table 5.2.2.2.

Method 1

The shaded region in Figures 5.2.2.1–5.2.2.4 show the distribution of accepted values of x_u and $k^0_{u,force}$ for the I27 test systems using method 1. The error distribution is approximately elliptical in x_u and $k^0_{u,force}$ for all I27 test systems which agrees with previous studies [217, 264]. This ellipse is narrower in the x_u and spans over a large range of $k^0_{u,force}$. Therefore, the errors associated with the $k^0_{u,force}$ are typically the same

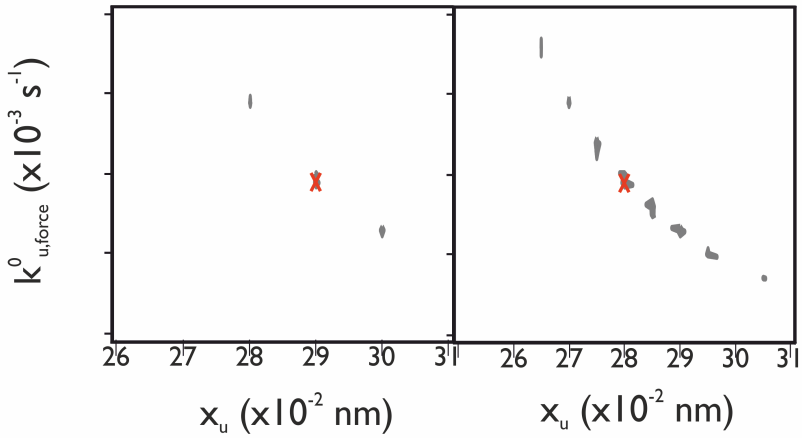


Figure 5.2.2.1: Graph of the accepted $k^0_{u,force}$ against x_u for the p(I27) in H_2O system using an x_u stepsize of 0.1 nm (left) and 0.05 nm (right). The stepsize of $k^0_{u,force}$ is 0.00005 s^{-1} in both graphs. These parameters have errors within the 95 % confidence interval defined by the best fit parameters (red cross) obtained from MC simulations. Darker grey regions of the contour graph indicated values with errors close to 0.

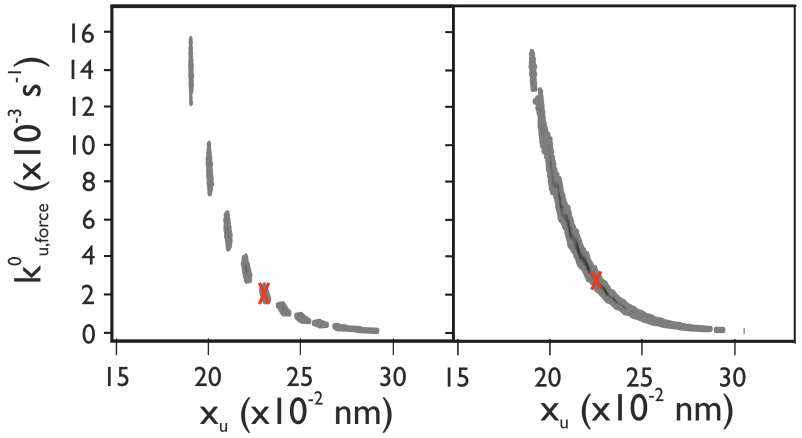


Figure 5.2.2.2: Graph of the accepted $k^0_{u,force}$ against x_u for the p(I27) in D_2O system using an x_u stepsize of 0.1 nm (left) and 0.05 nm (right). The stepsize in both graphs for $k^0_{u,force}$ is 0.00005 s^{-1}. These parameters have errors within the 95 % confidence interval defined by the best fit parameters (red cross) obtained from MC simulations. Darker grey regions of the contour graph indicated values with errors close to 0.

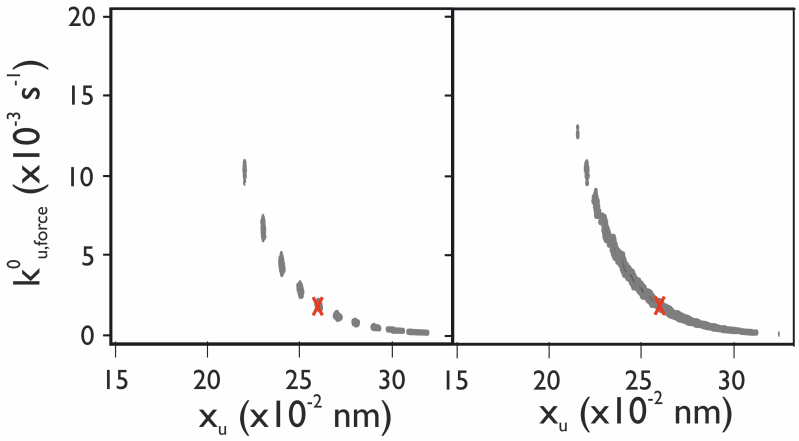


Figure 5.2.2.3: Graph of the accepted $k^0_{u,force}$ against x_u for the d(pL) in H_2O system using an x_u stepsize of 0.1 nm (left) and 0.05 nm (right). The stepsize on for $k^0_{u,force}$ is 0.00005 s^{-1} in both graphs. These parameters have errors within the 95 % confidence interval defined by the best fit parameters (red cross) obtained from MC simulations. Darker grey regions of the contour graph indicated values with errors close to 0.

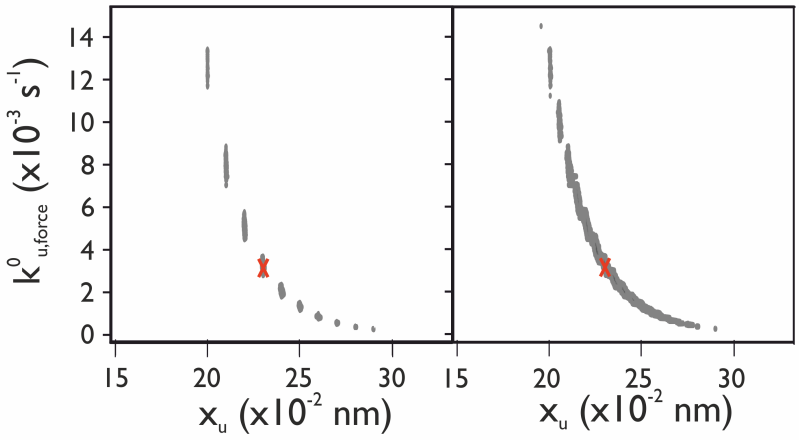


Figure 5.2.2.4: Graph of the accepted $k^0_{u,force}$ against x_u for the d(pL) in D_2O system using an x_u stepsize of 0.1 nm (left) and 0.05 nm (right). The stepsize for the $k^0_{u,force}$ is 0.00005 s^{-1} in both graphs. These parameters have errors within the 95 % confidence interval defined by the best fit parameters (red cross) obtained from MC simulations. Darker grey regions of the contour graph indicated values with errors close to 0.

order of magnitude as the $k^0_{u,force}$[264]. The reason the error distributions are not a continuous contour in Figures 5.2.2.1–5.2.2.4 (left) is due to the step size chosen for x_u. These figures show that the contours of the error distribution become more continuous for the smaller Δx_u of 0.05 nm as opposed to 0.1 nm. However, increasing the number of x_u or $k^0_{u,force}$ steps, increases the computational cost of the simulation. The figures indicate that decreasing the step size of the simulation did not significantly alter the best fit parameters obtained.

The agreements of the F(log(v):best fit) to the experimental data are shown in Figure 5.2.2.5. The F(log(v):best fit) for each I27 system describe the experimental data well. The best fit parameters for the p(I27) in H_2O system, using method 1, (x_u=0.29 nm and $k^0_{u,force}$=9.5 $\times 10^{-4}$ s^{-1}) agree well with the values obtained by Brockwell *et al.* of x_u=0.29 nm and $k^0_{u,force}$=20.0 $\times 10^{-4}$ s^{-1} [214]. The slight differences are attributed to the different methods used to obtain the best fit parameters. However, the values for p(I27) in D_2O are not comparable for those obtained for another mutant of I27 [132]. In this study, the $k^0_{u,force}$ was found to decrease in deuterium oxide therefore increasing the transition barrier height. Method 1 indicates an increase in $k^0_{u,force}$ for the protein used in this project. Furthermore, the study by Dougan *et al.* 2008, noted no change in the x_u. In this study, the best fit x_u has reduced by 0.07 nm. However, the range of accepted x_u and $k^0_{u,force}$ is large, this causes large errors associated with the $k^0_{u,force}$. The spring constants determined from the values obtained from Method 1 are in Table 5.2.2.2. These results suggests an increase in $\sim$ 1.5 N/m for both the protiated and deuterated I27 proteins due to isotopic substitution in the solvent.

Method 2

Method 2 was also implemented to determine the best fit parameters to the experimental data for the four I27 test systems. The range of accepted $k^0_{u,force}$ within the 95 % confidence interval of the best fit to the data for the fixed average x_u of 0.25 nm was determined. Whilst fixing x_u reduced the error in three of the I27 test systems, it increased the error in the p(I27) in H_2O. The standard deviation of the $k^0_{u,force}$ for this system increase from 0.1 $\times$ 10^{-4} s^{-1} to 10.2 $\times$ 10^{-4}. This is a ten-fold increase in the associated error. Furthermore the simulated forces deviate from the experimental data (Figure 5.2.2.5). This suggests that fixing x_u for the p(I27) in H_2O to 0.25 nm is not valid. Previous studies have determined the x_u to be 0.29 nm for this system [214]. To determine whether this x_u would adequately describe the four I27 test systems, the

Variable	System			
Method 1	p(I27) in H_2O	p(I27) in D_2O	d(I27) in H_2O	d(I27) in D_2O
B.f. x_u (nm)	0.29	0.23	0.26	0.23
Min x_u (nm)	0.28	0.19	0.22	0.20
Max x_u (nm)	0.30	0.29	0.31	0.29
Av. x_u (nm)	0.29	0.21	0.24	0.22
Sd x_u (nm)	0.01	0.02	0.02	0.02
B.f. $k^0_{u,force}$ (s^{-1})	9.5×10^{-4}	20.5×10^{-4}	18.5×10^{-4}	31.0×10^{-4}
Max $k^0_{u,force}$ (s^{-1})	14.5×10^{-4}	157.5×10^{-4}	109.0×10^{-4}	134.0×10^{-4}
Min $k^0_{u,force}$ (s^{-1})	6.5×10^{-4}	1.0×10^{-4}	2.0×10^{-4}	2.0×10^{-4}
Av. $k^0_{u,force}$ (s^{-1})	10.1×10^{-4}	75.1×10^{-4}	49.0×10^{-4}	63.8×10^{-4}
Sd $k^0_{u,force}$ (s^{-1})	3.3×10^{-4}	48.3×10^{-4}	31.5×10^{-4}	40.2×10^{-4}
Av. $\Delta G^0_{TS,force}$ (kJ/mol)*	51.1	46.4	47.8	46.6
Sd $\Delta G^0_{TS,force}$ (kJ/mol)	0.8	2.5	2.2	2.2
Av.D (N/m)	2.02	3.53	2.70	3.21
Sd (N/m)	0.08	0.45	0.31	0.37
Method 2a	p(I27) in H_2O	p(I27) in D_2O	d(I27) in H_2O	d(I27) in D_2O
B.f. x_u (nm)	0.25	0.25	0.25	0.25
B.f. $k^0_{u,force}$ (s^{-1})	50.0×10^{-4}	8.0×10^{-4}	27.5×10^{-4}	13.0×10^{-4}
Max $k^0_{u,force}$ (s^{-1})	69.5×10^{-4}	10.5×10^{-4}	34.5×10^{-4}	16.5×10^{-4}
Min $k^0_{u,force}$ (s^{-1})	35.0×10^{-4}	6.0×10^{-4}	23.0×10^{-4}	10.5×10^{-4}
Av. $k^0_{u,force}$ (s^{-1})	52.3×10^{-4}	8.3×10^{-4}	28.8×10^{-4}	13.5×10^{-4}
Sd $k^0_{u,force}$ (s^{-1})	10.2×10^{-4}	1.5×10^{-4}	3.3×10^{-4}	1.9×10^{-4}
Av. $\Delta G^0_{TS,force}$ (kJ/mol)	47.0	51.0	48.4	49.8
Sd $\Delta G^0_{TS,force}$ (kJ/mol)	0.5	0.5	0.3	0.4
Av. D (N/m)	2.51	2.75	2.58	2.68
Sd (N/m)	0.03	0.03	0.02	0.02

Table 5.2.2.2:]

Statistics obtained from MC simulations for the four I27 test systems: p(I27) in H_2O, p(I27) in D_2O, d(I27) in H_2O, d(I27) in D_2O. Two methods were used to obtain the parameters x_u; the distance between the unfolded and transition state in a two-state unfolding energy landscape, $k^0_{u,force}$; the rate of unfolding at zero force, $\Delta G^0_{TS,force}$; the height of the transition barrier to unfolding and D; the spring constant of the protein. Method 1 determined the best fit to the experimental data, allowing x_u to vary. Method 2 determined the best fit to the experimental data assuming a fixed x_u. B.f. indicates the best fit parameters. Av. is the average of the accepted x_u and $k^0_{u,force}$, Sd is the standard deviation of the values. The Av. and Sd $\Delta G^0_{TS,force}$ was the average and standard deviation taken from calculating $\Delta G^0_{TS,force}$ for every accepted $k^0_{u,force}$. * 1 $k_BT \approx 2.46$ kJ/mol at 23 °C.

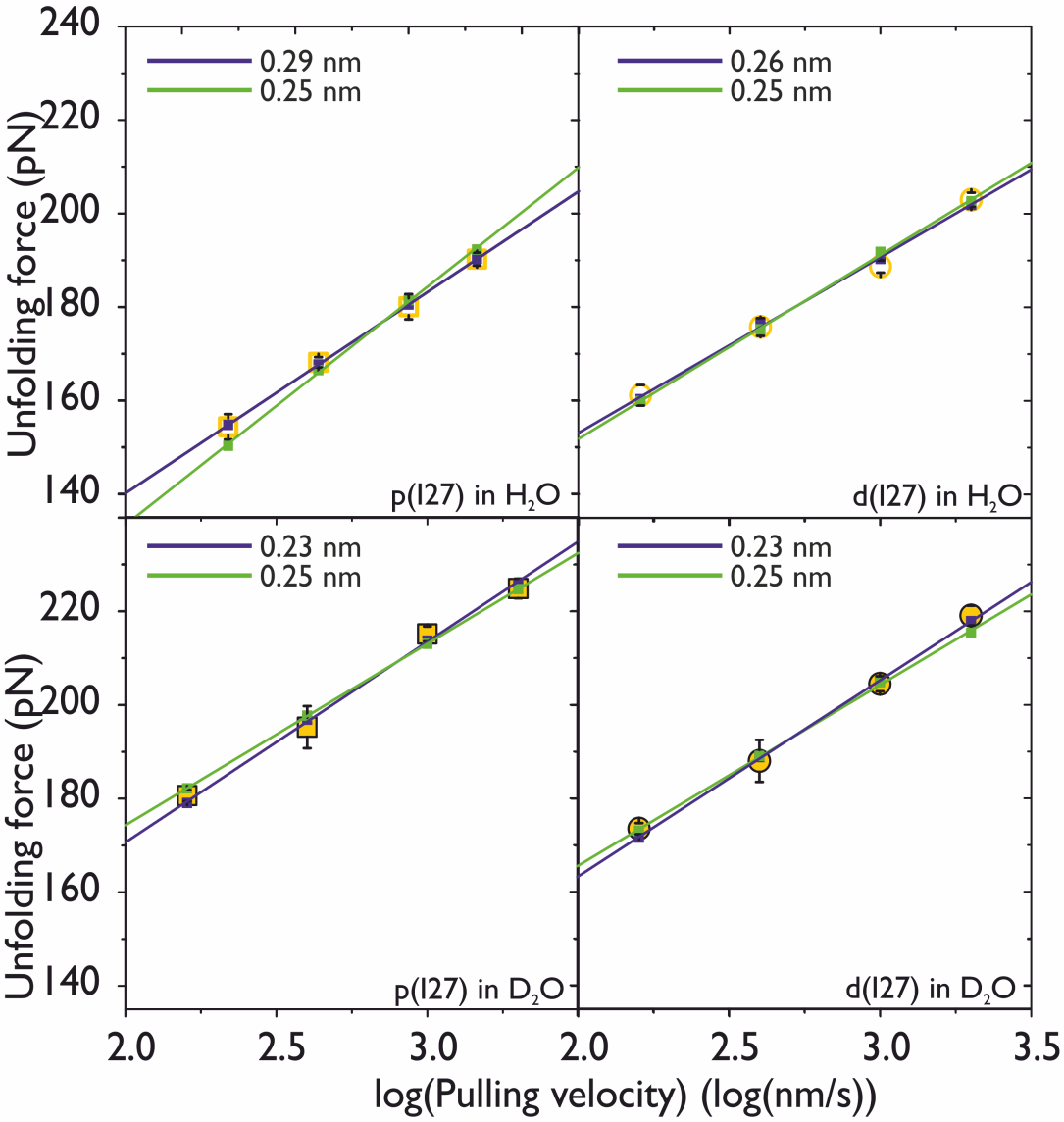


Figure 5.2.2.5: Graph of Unfolding force against log(Pulling velocity) for the four I27 test systems (as labelled) comparing the experimental and simulated data determined by two methods. The numbers next to the associated lines in the key indicate the x_u and the method used to determine the parameters. Blue: The straight line is the fit to simulated forces from method 1 corresponding to the best fit parameters to the experimental data of x_u and $k^0_{u,force}$. This line minimises the least squares of a weighted fit to the experimental data. The associated x_u are labelled. Green: Straight line fits to the simulated forces from method 2 which minimise the error to the experimental data for a fixed x_u of 0.25 nm. The associated best fit $k^0_{u,force}$ are presented in Table 5.2.2.2. The graph indicates close agreement between the experimental and simulated data for the d(I27) in H_2O, d(I27) in D_2O and p(I27) in D_2O using both methods. However, it is apparent the green simulated data deviates from the experimental data for p(I27) in H_2O.

value of x_u was then fixed to the 0.29 nm. The error reduced for all systems. However, comparisons between the experimental and simulated forces showed that large deviations are observed between the straight line fit determined from using the best fit parameters obtained and the I27 systems in D_2O. The associated Figure and Table for fixing x_u to 0.29 nm can be found in D.

An additional step was performed to reduce the error associated with the $k^0_{u,force}$ for all four systems. It was noted that the best fit x_u for the d(I27) in H_2O in method 1 was 0.26 nm, which is similar to p(I27) in H_2O. The two I27 systems in D_2O had smaller x_u than the two I27 systems in H_2O. It was therefore assumed that the x_u for the two I27 systems in H_2O was 0.28 nm. The x_u of the two I27 systems in D_2O was fixed to 0.23 nm. The values obtained from fixing the x_u for the two different solvents are presented in Table 5.2.2.3. The error is significantly reduced for all test systems and the fits generated by the best fit parameters agree well with the experimental data (Figure D.1.0.1. The figure comparing the experimental and simulated forces can be found in the Appendix D. The spring constants associated with the test systems from this second iteration of Method 2 are given in Table 5.2.2.3. The data suggests an increase of $\sim$ 1.0 N/m in the spring constants of the deuterated and protiated I27 proteins in D_2O.

Variable	System			
Method 2	p(I27) in H_2O	p(I27) in D_2O	d(I27) in H_2O	d(I27) in D_2O
B.f. x_u (nm)	0.28	0.23	0.28	0.23
B.f. $k^0_{u,force}$ (s^{-1})	14.5×10^{-4}	20.5×10^{-4}	8.0×10^{-4}	31.0×10^{-4}
Max $k^0_{u,force}$ (s^{-1})	17.0×10^{-4}	25.5×10^{-4}	10.0×10^{-4}	36.0×10^{-4}
Min $k^0_{u,force}$ (s^{-1})	12.5×10^{-4}	17.0×10^{-4}	6.0×10^{-4}	27.5×10^{-4}
Average $k^0_{u,force}$ (s^{-1})	14.8×10^{-4}	21.3×10^{-4}	8.0×10^{-4}	31.8×10^{-4}
Sd $k^0_{u,force}$ (s^{-1})	1.5×10^{-4}	2.7×10^{-4}	1.4×10^{-4}	2.7×10^{-4}
Av. $\Delta G^0_{TS,force}$ (kJ/mol)	50.1	48.6	51.6	47.6
Sd $\Delta G^0_{TS,force}$ (kJ/mol)	0.3	0.3	0.4	0.2
Av D (N/m)	2.11	3.10	2.18	3.04
Sd D (N/m)	0.01	0.02	0.02	0.02

Table 5.2.2.3: Statistics obtained from MC simulations for the four I27 test systems: p(I27) in H_2O, p(I27) in D_2O, d(I27) in H_2O, d(I27) in D_2O. Method 2 was used to obtain the parameters x_u; the distance between the unfolded and transition state in a two-state unfolding energy landscape, $k^0_{u,force}$; the rate of unfolding at zero force, $\Delta G^0_{TS,force}$; the height of the transition barrier to unfolding and D; the spring constant of the protein. Method 2 determines the best fit to the experimental data assuming a fixed x_u. B.f. indicates the best fit parameters. Av. is the average of the accepted x_u and $k^0_{u,force}$, Sd is the standard deviation of the values. The Av. and Sd $\Delta G^0_{TS,force}$ was the average and standard deviation taken from calculating $\Delta G^0_{TS,force}$ for every accepted $k^0_{u,force}$.

Figure 5.2.2.6 shows a comparison of all of the straight line fits from the best fit x_u and $k^0_{u,force}$ parameters determined from each of the methods described. The

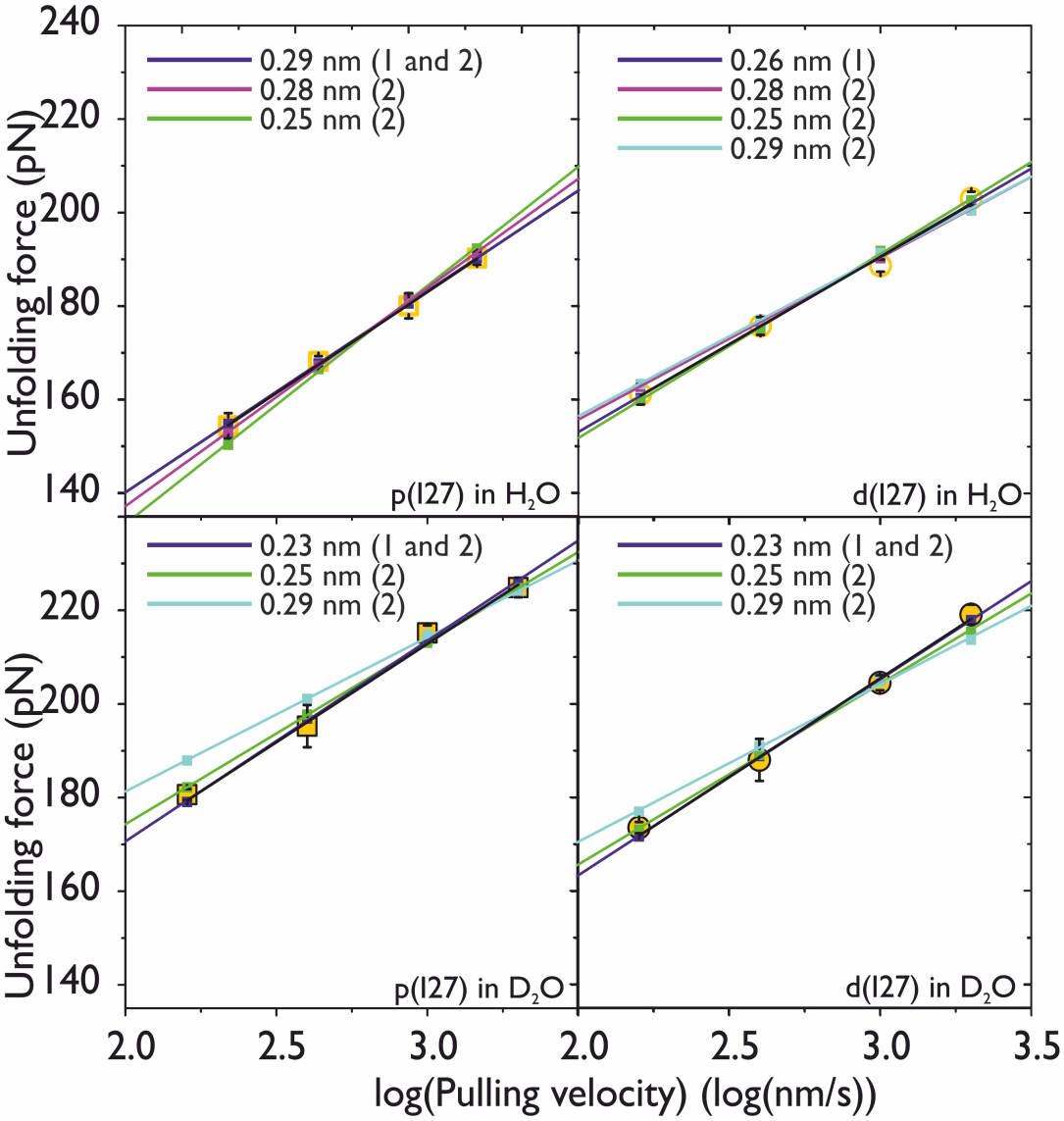


Figure 5.2.2.6: Graph of Unfolding force against log(Pulling velocity) for the four I27 test systems (as labelled) comparing the experimental and simulated data determined by two methods. The values in the key indicate the value of the best fit x_u and the method used to obtain the best fit parameters. Blue: The straight line is the fit to simulated forces corresponding to the best fit parameters to the experimental data of x_u and $k^0_{u,force}$ using method 1. In some cases this x_u was also fixed this is indicated by "and 2" in the key. These line minimises the least squares of a weighted fit to the experimental data. Green line: The line fit to the simulated forces for the best fit parameter $k^0_{u,force}$ using a fixed x_u of 0.25 nm. Pink: Straight line fits to the simulated forces that minimise the error to the experimental data for a fixed x_u of 0.28 nm, for p(I27) in H_2O and d(I27) in H_2O, and $x_u = 0.23$ nm for p(I27) in D_2O and d(I27) in D_2O. The associated best fit $k^0_{u,force}$ are given in Table 5.2.2.2. The graph indicates discrepancies between the experimental and simulated data for some of the fixed x_u used in method 2.

figure shows that the best fit parameters from method 1 describe the data the best in each case (dark blue line). The fit to the best fit parameters obtained by fixing the x_u to 0.25 nm (green) for p(I27) in H_2O deviates from the experimental data points. For the other three I27 systems, the straight line fits from setting x_u to 0.25 nm describe the data well. It is apparent that fixing the x_u to 0.29 nm (cyan) results in significant deviations from the experimental data in the d(I27) in D_2O and p(I27) in D_2O systems. This suggests that the x_u has reduced in these systems and therefore the transition state has moved toward the folded state of the protein. All methods appear to fit the d(I27) in H_2O experimental data well. The largest deviations from the experimental data in this system arise from setting the x_u to 0.28 (pink) and 0.29 nm (cyan).

5.3 Incorporating the unequal probability distributions in MC simulations

Before incorporating the unequal probability distribution into the MC simulations, the effect of deuterium on the unfolding history effect was first determined. This was to see whether the forces were significantly affected for an individual unfolding event in the polyprotein for each of the pL and I27 test systems.

5.3.1 How does deuterium affect the unfolding history of the protein?

The unfolding history is the consequence of the competition between the number of domains unfolded/remaining folded and stiffness of the supramolecular scaffold. The compliance of the chain therefore will influence this unfolding history, either due to a change in the cantilever spring constant or the compliance of the polypeptide [328]. It is expected, for the four I27 and protein L test systems, that the minima in a force-event number graph should be in the same position across the four protein test systems. This is because the same number of domains, of equal length, are in the polyprotein and the cantilevers had similar spring constants (average spring constant: 41.5 ± 0.5 pN/nm protein L and 41.6 ± 0.4 pN/nm I27). Forces were obtained for each unfolding event from traces containing 5 unfolding peaks and a detachment. Each unfolding peak was labelled as shown in Figure 5.1.0.2 (A). Forces for each unfolding event, in a single force-extension trace, were normalised to the sum of all unfolding forces observed in the trace. This was to alleviate errors associated with the spring constant of a fresh cantilever in each experiment. In a previous study, over 104 traces containing the full unfolding profiles of the polyprotein were obtained and analysed [328]. The forces for each event (1,2,3,4,5) were pooled into

histograms and the weighted mean and standard error were obtained. However, it was not possible to obtain this number of full unfolding profiles for each of the four I27 and protein L test systems in this thesis. Therefore, the average normalised force and standard error have been determined for the I27 and protein L test systems. The change of force with event number will be discussed in this section.

5.3.2 The four protein L test systems

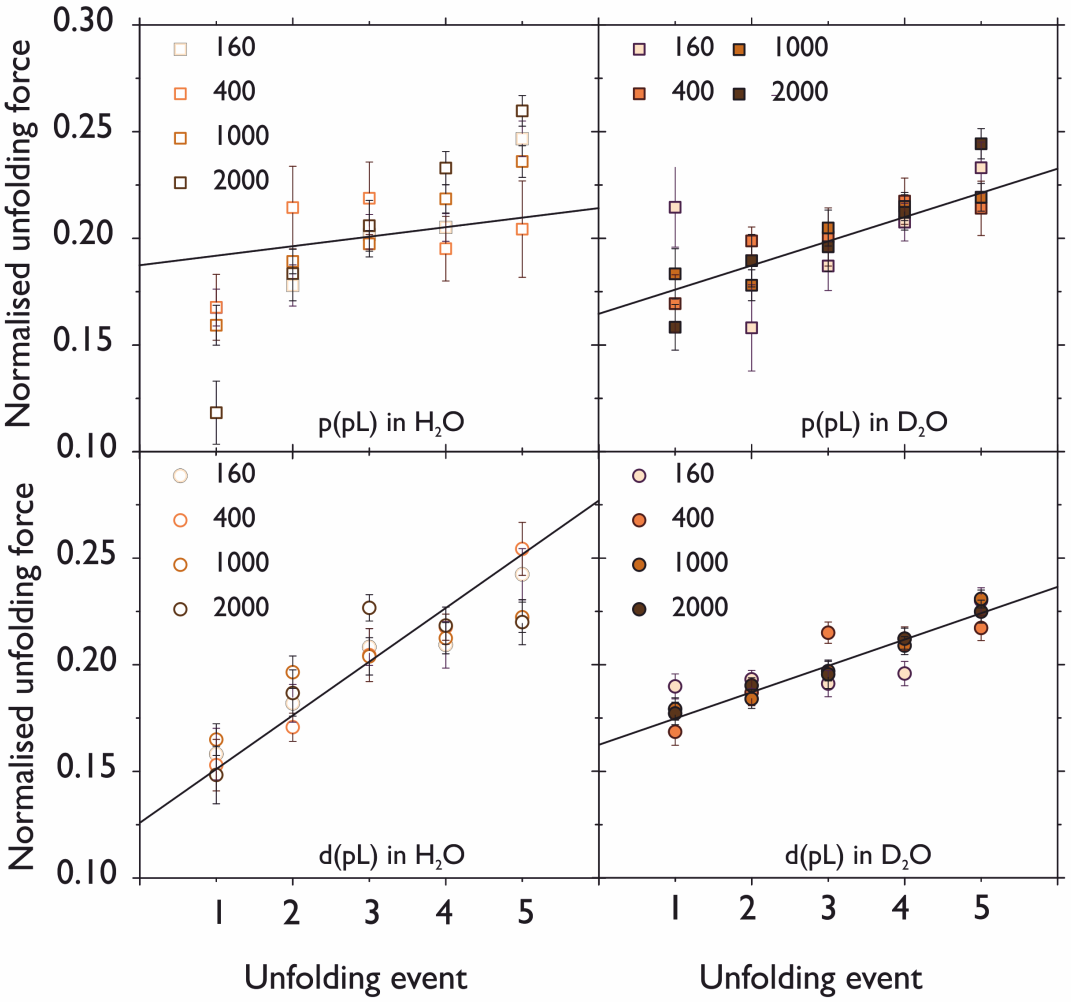


Figure 5.3.2.1: Figure showing how the unfolding force changes with event number for the pL test systems at each experimental velocity tested. Each event number corresponds to the order of protein unfolding peaks observed in a force-extension trace for the unfolding of a full pentameric polyprotein. For example, 5 corresponds to the fifth protein unfolding event observed before the detachment of a protein from the cantilever. The black line is a linear fit to the change in experimental normalised forces at 400 nm/s is also plotted to determine whether there are obvious deviations from a linear increase in force with event number. A minimum is typically observed in an unfolding force–event number graph[328]. For the p(pL) in H_2O, d(pL) in H_2O and p(pL) in D_2O it appears that the forces vary linearly with event number. However, there are large errors associated with the unfolding forces which may obscure the unfolding history of the polyprotein chain. Despite this, the unfolding forces from the d(pL) in D_2O system show a small deviation from a linear increase in the unfolding force between event numbers 3 and 4.

Figure 5.3.2.1 shows the experimental force-event number graphs obtained for the four protein L test systems. The force appears to increase linearly with event number for this protein. The experimental normalised forces for the four different experimental velocities do not overlay well which could suggest compliance affects with the changing velocity. However, for most unfolding events, the associated error is too large to differentiate the normalised forces. For comparison, the best fit x_u and $k^0_{u,force}$ from Section 5.2, using method 1, were used to simulate 1000 perfect traces at each velocity for each protein L test system. A comparison of the simulated and experimental data is shown in Figure 5.3.2.2. Whilst the experimental normalised forces agree with the simulated normalised forces, the minima observed in the simulations was not observed in the experimental traces. The best agreement between the simulated and experimental forces, across all velocities, was with the d(pL) in D_2O system. This system had the highest number of full unfolding traces observed than the other protein L test systems. The total number of traces containing the five unfolding events, i.e. the full polyprotein construct can be found in the Appendix D.

5.3.3 The four I27 test systems

Figure 5.3.3.1 illustrates the change in force with event number graphs obtained for the four I27 test systems. Whilst there appear to be some deviations from a linear increase in force for the p(I27) in H_2O, p(I27) in D_2O and d(I27) in D_2O, it is less clear in the d(I27) in H_2O. A clear deviation may not be observed because are larger errors associated with the normalised forces for the d(I27) in H_2O. It appears as though the minima in the normalised forces, observed for p(protein) in H_2O, occurs later in the I27 systems in D_2O. This could be indicative of an increase in stiffness of the protein due to a change in the solvent. For comparison, the best fit x_u and $k^0_{u,force}$ from Section 5.2, using Method 1, were used to simulate 1000 perfect traces at each velocity. The agreement between the normalised simulated forces with the experimental data suggests that the apparent shift in minimum for the systems in D_2O may be an artefact of the small number of traces containing five protein unfolding events (information given in the Appendix D). However this could also suggest the x_u and $k^0_{u,force}$ determined from the Bell model does not adequately describe these systems, better statistics would have to be obtained to make a fair comparison. Furthermore, the data was simulated from the best fit data collected in Section 5.2. This MCS assumes an equal unfolding probability distribution. Due to the lack of ideal traces, this could suggest that the parameters obtained are not the true "best fit" to the data.

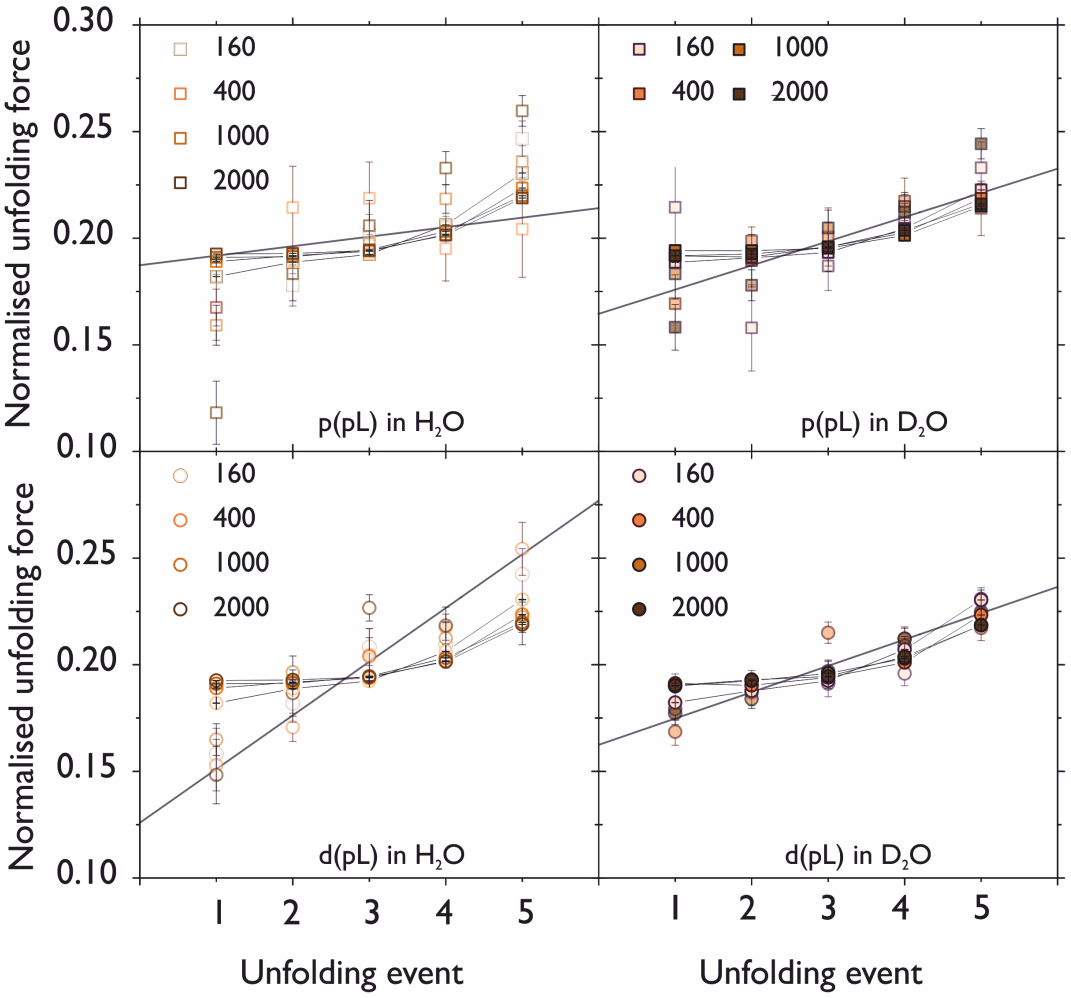


Figure 5.3.2.2: Figure showing how the unfolding force changes with event number for the simulated unfolding forces, from 1000 force-extension traces. Simulations were run using the best fit x_u and $k^0_{u,force}$ for each of the pL test systems at each experimental velocity tested. Each event number corresponds to the order of protein unfolding peaks observed in a force-extension trace for the unfolding of a full pentameric polyprotein. For example, 5 corresponds to the fifth protein unfolding event observed before the detachment of a protein from the cantilever. The simulated data does not overlay well with the experimental data (faint symbols) for the p(pL) in H_2O, d(pL) in D_2O and p(pL) in D_2O but agrees well with the data obtained for the d(pL) in D_2O. The simulated data, for all systems, suggests a deviation from a linear increase in the unfolding force occurs at unfolding event number 3.

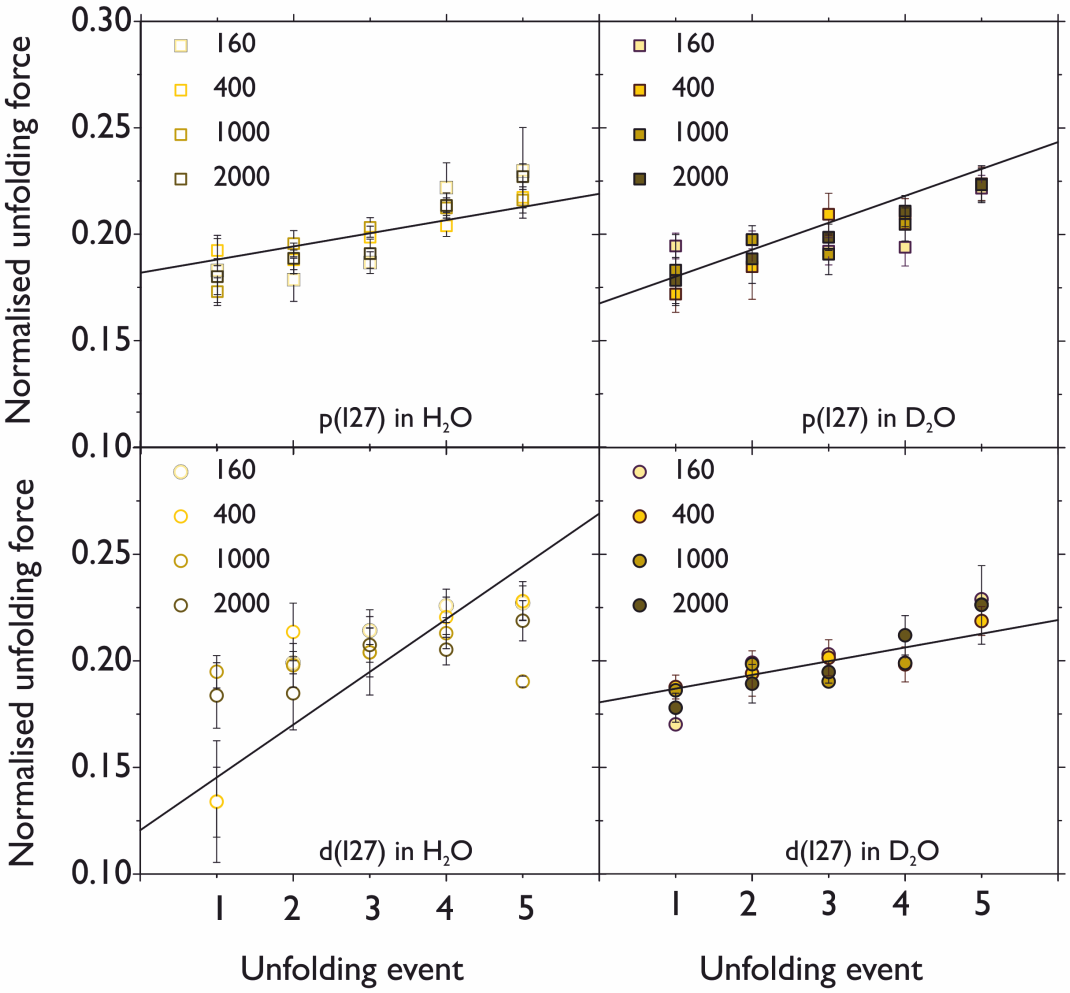


Figure 5.3.3.1: Figure showing how the unfolding force changes with event number for the I27 test systems at each experimental velocity tested. Each event number corresponds to the order of protein unfolding peaks observed in a force-extension trace for the unfolding of a full pentameric polyprotein. For example 5 corresponds to the fifth protein unfolding event observed before the detachment of a protein from the cantilever. The black line is a linear fit to the change in experimental normalised forces at 400 nm/s is also plotted to determine whether there are obvious deviations from a linear increase in force with event number. A minimum is typically observed in an unfolding force–event number graph[328]. For the p(I27) in H_2O and d(I27) in H_2O a minimum in the unfolding force seems to occur around event number 3. This minimum seems to shift to event number 4 for the p(I27) in D_2O and d(pL) in D_2O systems.

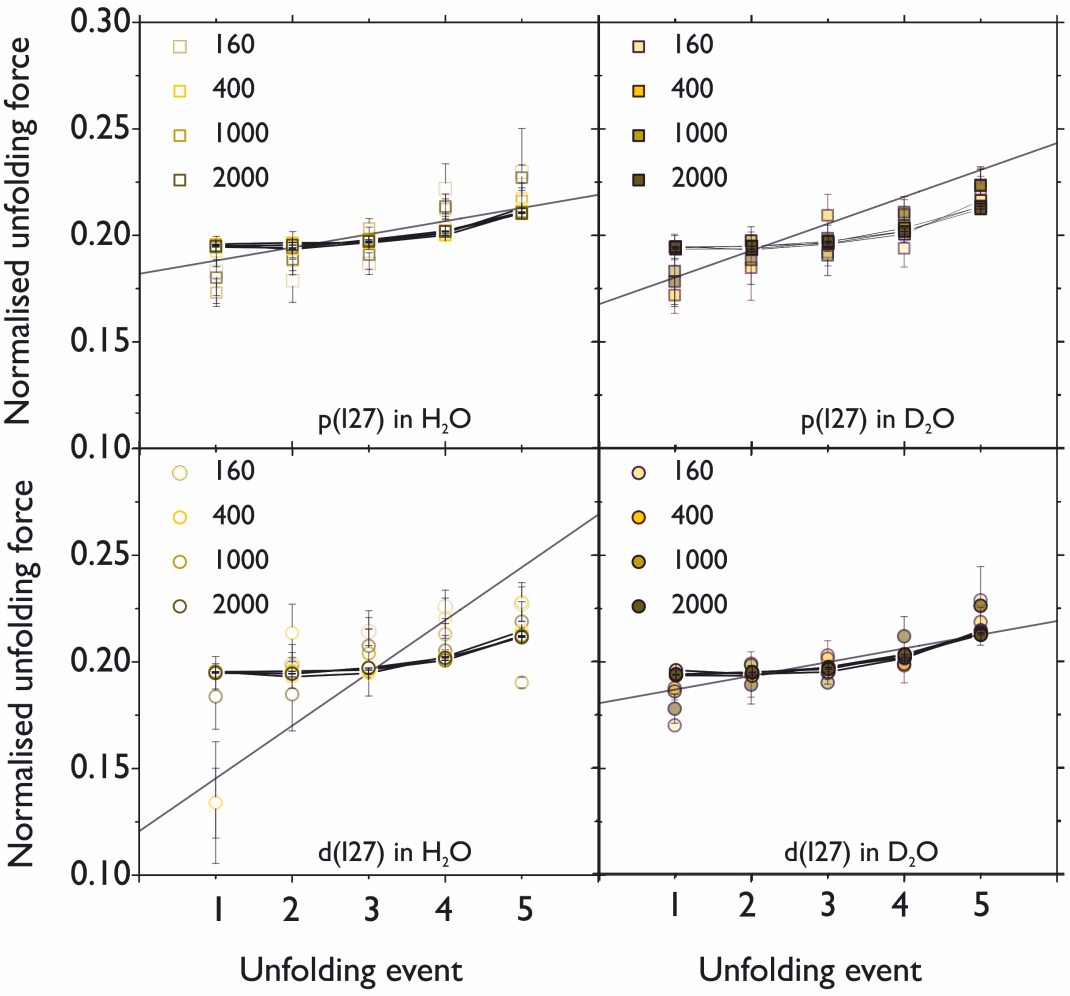


Figure 5.3.3.2: Figure showing how the unfolding force changes with event number for the simulated unfolding forces, from 1000 force-extension traces of the I27 proteins. Simulations were run using the best fit x_u and $k^0_{u,force}$ for each of the I27 test systems at each experimental velocity tested. Each event number corresponds to the order of protein unfolding peaks observed in a force-extension trace for the unfolding of a full pentameric polyprotein. For example 5 corresponds to the fifth protein unfolding event observed before the detachment of a protein from the cantilever. The simulated data does overlays well with the experimental data for all systems (faint symbols). The simulated data, for all systems, suggests a minimum in the unfolding force occurs at unfolding event number 3.

5.3.4 MC simulations with an unequal unfolding probability distribution

It has been shown the probability of observing two unfolding events is higher than observing the full protein construct in SMFS experiments [379, 381]. Furthermore, previous studies have shown that the mean unfolding force changes with the number of protein domains in the construct [214, 328, 383]. As the number of protein domains in the polyprotein construct increases, the mean unfolding force has been shown to decrease [214, 328, 383]. The MC simulations, unfolding five protein domains, assumes an equal probability of observing two, three, four or five unfolding domains. Therefore if only two protein domains were picked up in an experiment the average unfolding force would be overestimated[379].

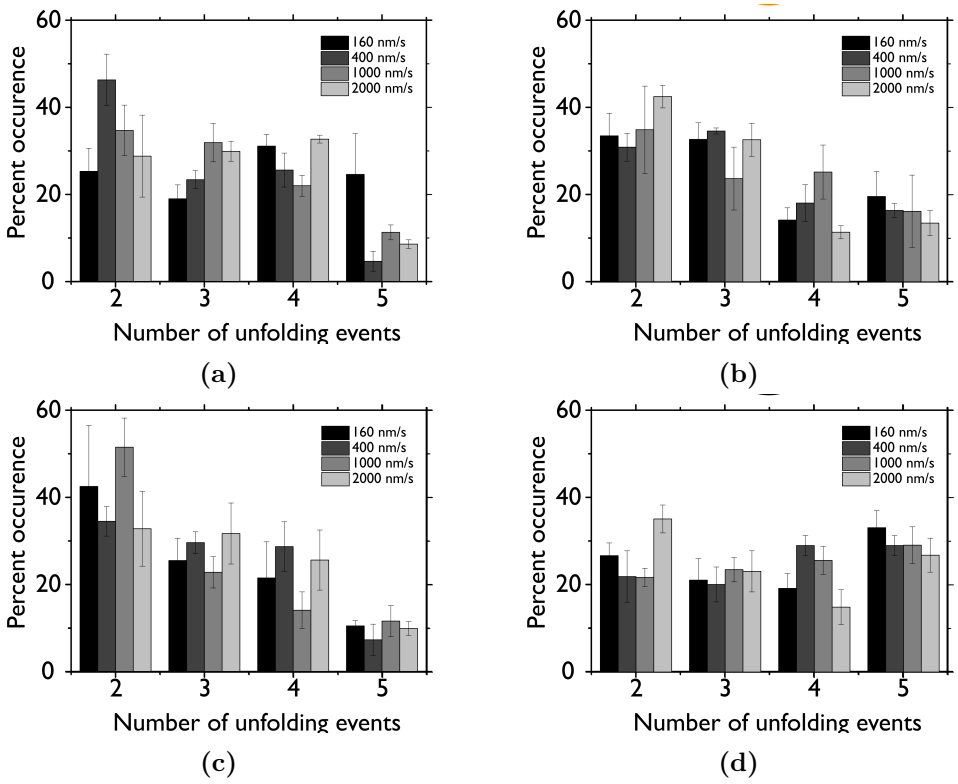


Figure 5.3.4.1: Figure showing the percent occurrence of force-extension traces containing 2, 3, 4 or 5 pL unfolding events for a pentameric polyprotein for each pL test system. (a) p(pL) in H_2O, (b) d(pL) in H_2O, p(pL) in D_2O and (d) d(pL) in D_2O. The percent occurrence generally decreases with the number of unfolding events for (a), (b) and (c). In (d) the percent occurrence does not appear to change with the number of unfolding events.

Figures 5.3.4.1 - 5.3.4.2 show the probability distributions of observed traces containing: two, three, four or five protein unfolding events. There appears to be a bias to observing two unfolding events than five unfolding events. For the d(pL) in D_2O and p(pL) in H_2O experiments, the probability of observing 2, 3, 4 or 5 unfolding events is $\sim$ 25 %. For these two systems the average number of points per experiment was 173 and 150 respectively. The other protein L systems, p(pL) in H_2O, d(pL) in H_2O and p(pL)

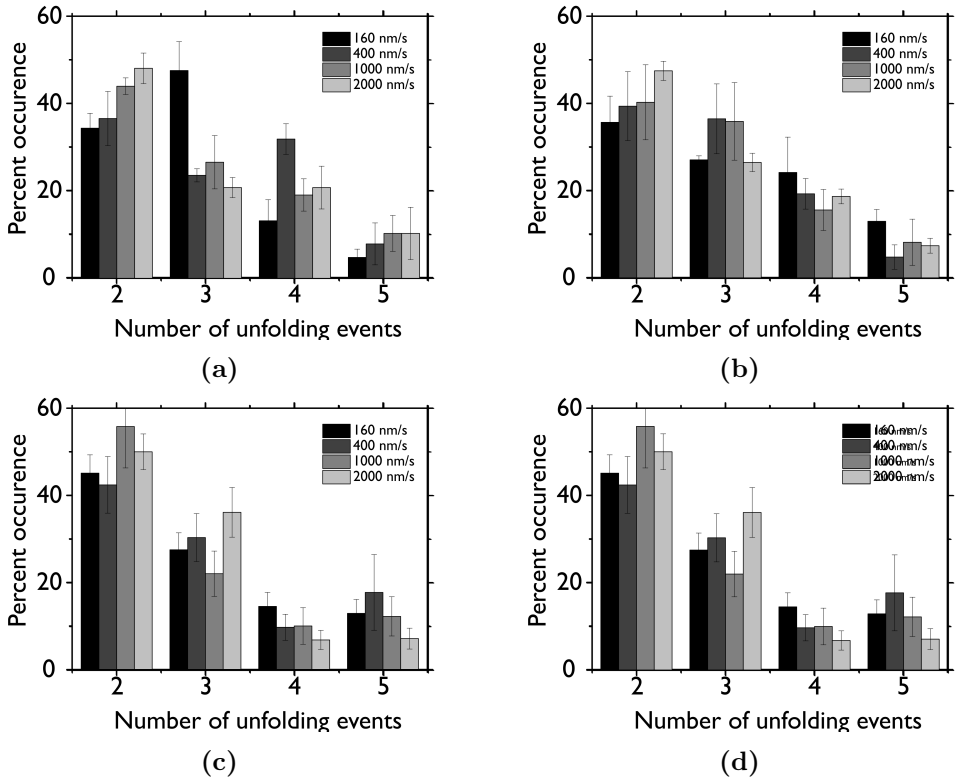


Figure 5.3.4.2: Figure showing the percent occurrence of force-extension traces containing 2, 3, 4 or 5 I27 unfolding events for a pentameric polyprotein for each I27 test system. (a) p(I27) in H_2O, (b) d(I27) in H_2O, p(I27) in D_2O and (d) d(I27) in D_2O. The percent occurrence generally decreases with the number of unfolding events for all protein test systems.

in D_2O, show a bias towards observing traces containing two unfolding events across all velocities. However, the average number of points per experiment for these protein L systems were: 150, 116 and 111 for d(pL) in H_2O and p(pL) in D_2O respectively. This could suggest that the stochasticity observed in experiments reduces with the number of traces observed. To determine whether the same effect is observed in the MC simulations, simulations of 1000, 100 and 30 traces were run to determine the % observation of each trace. The results of these simulations are shown in Figure 5.3.4.3. As the number of traces increases, the % occurrence of 2, 3, 4 and 5 unfolding events tends to 25 % as expected. This suggests that the stochasticity observed in our experiments might not be an artefact of memory in the system.

To determine whether the x_u and $k^0_{u,force}$ determined from MC simulations are sensitive to the unequal probability distribution observed in experiments, the MC was adapted to include the experimental stochasticity. Furthermore, the experimental spring constant was used to mimic the experiments. Both methods outlined in Section 5.2 were used to determine the x_u and $k^0_{u,force}$.

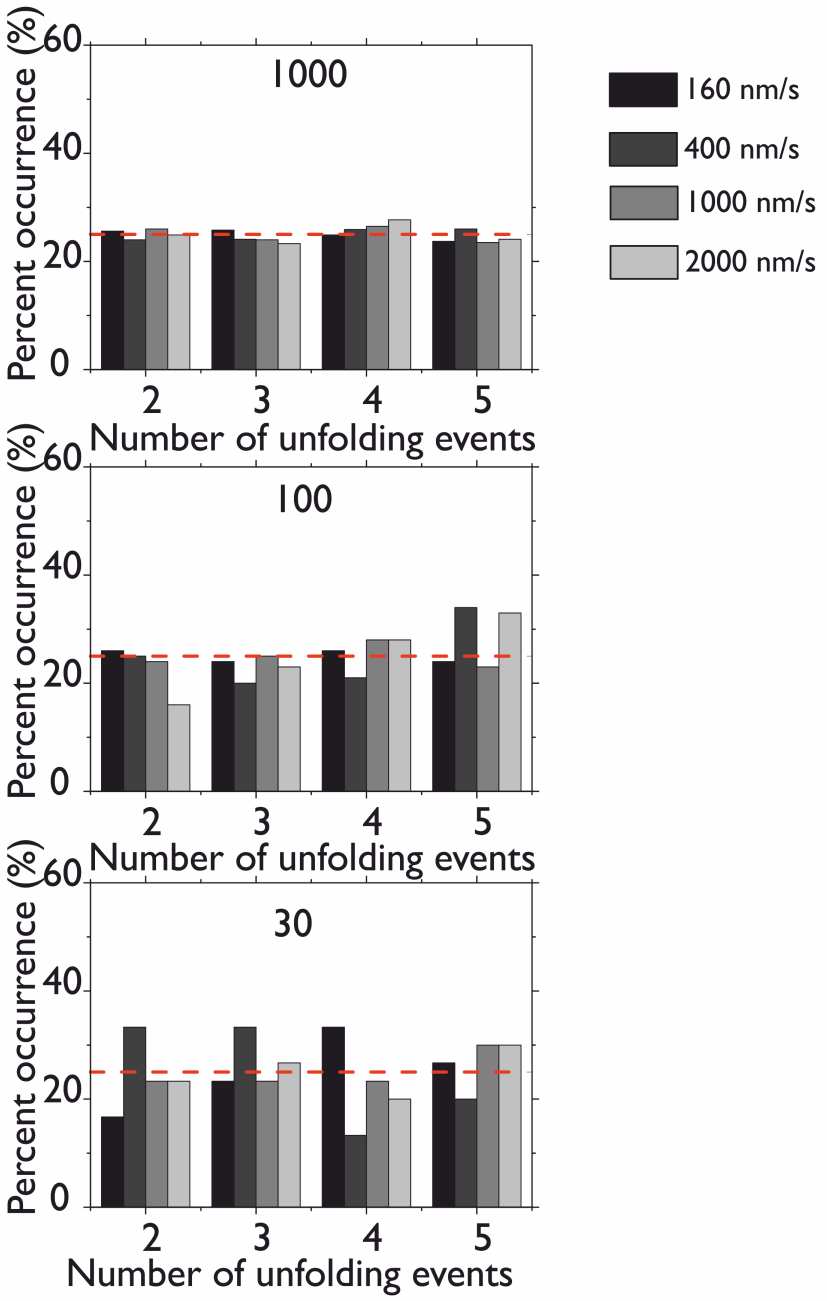


Figure 5.3.4.3: Bar charts showing the percent occurrence of force-extension traces containing 2, 3, 4 or 5 unfolding events from simulations of 1000 traces (top), 100 traces (middle) and 30 traces (bottom). Simulations were performed at 160 nm/s, 400 nm/s, 1000 nm/s and 2000 nm/s. When the number of simulated traces increases, the percent increase of observed traces increases to 25 %. The dashed red line indicates 25 %.

5.3.5 The four protein L test systems

The parameters used in the MCS simulations for each of the four pL systems can be found in the Appendix D. The percentages illustrated in Figure 5.3.4.1 were used in the simulation. This allowed the number of iterations in the simulation of two, three, four and five unfolding events to be determined.

Method1

Table 5.3.5.1 outlines the energy landscape parameters obtained using method 1 on the forces generated by these simulations. A comparison of the agreement between the simulated forces from the MC simulations incorporating equal, and unequal probability distributions and the experimental data for the p(pl) in H_2O system is given in Figure 5.3.4.1. There appears to be negligible difference between the best fits to the simulated data and therefore their agreement with the experimental data. This was the same for the other three pL systems. The figures comparing simulated and experimental data for these systems can be found in the Appendix D. For method 1, described in Section 5.2, the x_u corresponding to the best linear fit to the experimental data increased by 0.01 nm for the p(pL) in H_2O and d(pL) in D_2O. This results in a change in the best fit $k^0_{u,force}$ of -4.7 $\times 10^{-3}$ s^{-1} and -13.5 $\times 10^{-3}$ s^{-1} for the two systems respectively. These values are still comparable to the values obtained previously due to the large associated error with the $k^0_{u,force}$. The x_u corresponding with the best linear fit to the experimental data for the p(pL) in D_2O and d(pL) in H_2O was unperturbed in these simulations. Furthermore, only small changes were observed in the $k^0_{u,force}$ best fit parameter (+4.2 $\times 10^{-3}$ s^{-1} and +8.4 $\times 10^{-3}$ s^{-1} for p(pL) in D_2O and d(pL) in H_2O respectively). The associated error associated with these values is the same order of magnitude as the previous simulations. The spring constants for each of the protein systems were also determined to see how an unequal unfolding probability distribution affected the computed stiffness of the protein and are given in Table 5.3.5.1. Interestingly the average spring constants decreased by 0.1 – 0.4 N/m for all systems when compared to those obtained from an equal unfolding probability distribution. Whilst inclusion of the unequal unfolding probability distribution does not alter the energy landscape parameters obtained substantially, differences may occur between the spring constants. However, there is a large error associated with the values from method 1 are large therefore fixing x_u may provide a better comparison between the methods.

Variable	System			
Method 1	p(pL) in H_2O	p(pL) in D_2O	d(pL) in H_2O	d(pL) in D_2O
B.f. x_u (nm)	0.23	0.23	0.23	0.22
Min x_u (nm)	0.22	0.22	0.22	0.20
Max x_u (nm)	0.24	0.24	0.25	0.23
Average x_u (nm)	0.23	0.23	0.23	0.22
Sd x_u (nm)	0.01	0.01	0.01	0.01
B.f. $k^0_{u,force}$ (s^{-1})	49.4 $\times 10^{-3}$	19.4 $\times 10^{-3}$	77.5 $\times 10^{-3}$	30.9 $\times 10^{-3}$
Max $k^0_{u,force}$ (s^{-1})	69.9 $\times 10^{-3}$	28.8 $\times 10^{-3}$	113.3 $\times 10^{-3}$	64.9 $\times 10^{-3}$
Min $k^0_{u,force}$ (s^{-1})	36.6 $\times 10^{-3}$	13.2 $\times 10^{-3}$	39.8 $\times 10^{-3}$	20.1 $\times 10^{-3}$
Average $k^0_{u,force}$ (s^{-1})	48.8 $\times 10^{-3}$	20.6 $\times 10^{-3}$	77.8 $\times 10^{-3}$	35.0 $\times 10^{-3}$
Sd $k^0_{u,force}$ (s^{-1})	11.7 $\times 10^{-3}$	4.7 $\times 10^{-3}$	22.2 $\times 10^{-3}$	9.7 $\times 10^{-3}$
Av. $\Delta G^0_{TS,force}$ (kJ/mol)	41.5	43.1	40.4	41.8
Sd $\Delta G^0_{TS,force}$ (kJ/mol)	0.6	0.6	0.7	0.7
Av. D (N/m)	2.6	2.7	2.5	2.9
Sd D (N/m)	0.1	0.1	0.2	0.2
Method 2	p(pL) in H_2O	p(pL) in D_2O	d(pL) in H_2O	d(pL) in D_2O
B.f. x_u (nm)	0.23	0.23	0.23	0.23
B.f. $k^0_{u,force}$ (s^{-1})	49.4 $\times 10^{-3}$	19.4 $\times 10^{-3}$	77.5 $\times 10^{-3}$	22.4 $\times 10^{-3}$
Max $k^0_{u,force}$ (s^{-1})	51.6 $\times 10^{-3}$	20.3 $\times 10^{-3}$	85.0 $\times 10^{-3}$	24.6 $\times 10^{-3}$
Min $k^0_{u,force}$ (s^{-1})	46.2 $\times 10^{-3}$	18.8 $\times 10^{-3}$	72.6 $\times 10^{-3}$	20.0 $\times 10^{-3}$
Average $k^0_{u,force}$ (s^{-1})	48.9 $\times 10^{-3}$	19.4 $\times 10^{-3}$	79.0 $\times 10^{-3}$	22.1 $\times 10^{-3}$
Sd $k^0_{u,force}$ (s^{-1})	1.4 $\times 10^{-3}$	0.4 $\times 10^{-3}$	3.2 $\times 10^{-3}$	1.2 $\times 10^{-3}$
Av. $\Delta G^0_{TS,force}$ (kJ/mol)	41.43	43.19	40.25	42.85
Sd $\Delta G^0_{TS,force}$ (kJ/mol)	0.07	0.04	0.10	0.14
Av. D (N/m)	2.599	2.710	2.526	2.677
Sd D (N/m)	0.004	0.006	0.006	0.002

Table 5.3.5.1: Statistics obtained from the MC simulations for the four pL test systems: p(pL) in H_2O, p(pL) in D_2O, d(pL) in H_2O, d(pL) in D_2O. The MC simulations incorporate the unequal probability distributions observed in SMFS experiments. Two methods were used to obtain the parameters x_u; the distance between the unfolded and transition state in a two-state unfolding energy landscape, $k^0_{u,force}$; the rate of unfolding at zero force, $\Delta G^0_{TS,force}$; the height of the transition barrier to unfolding and D; the spring constant of the protein. Method 1 determined the best fit to the experimental data, allowing x_u to vary. Method 2 determined the best fit to the experimental data assuming a fixed x_u. B.f. indicates the best fit parameters. Av. is the average of the accepted x_u and $k^0_{u,force}$, Sd is the standard deviation of the values. The Av. and Sd $\Delta G^0_{TS,force}$ was the average and standard deviation taken from calculating $\Delta G^0_{TS,force}$ for every accepted $k^0_{u,force}$.

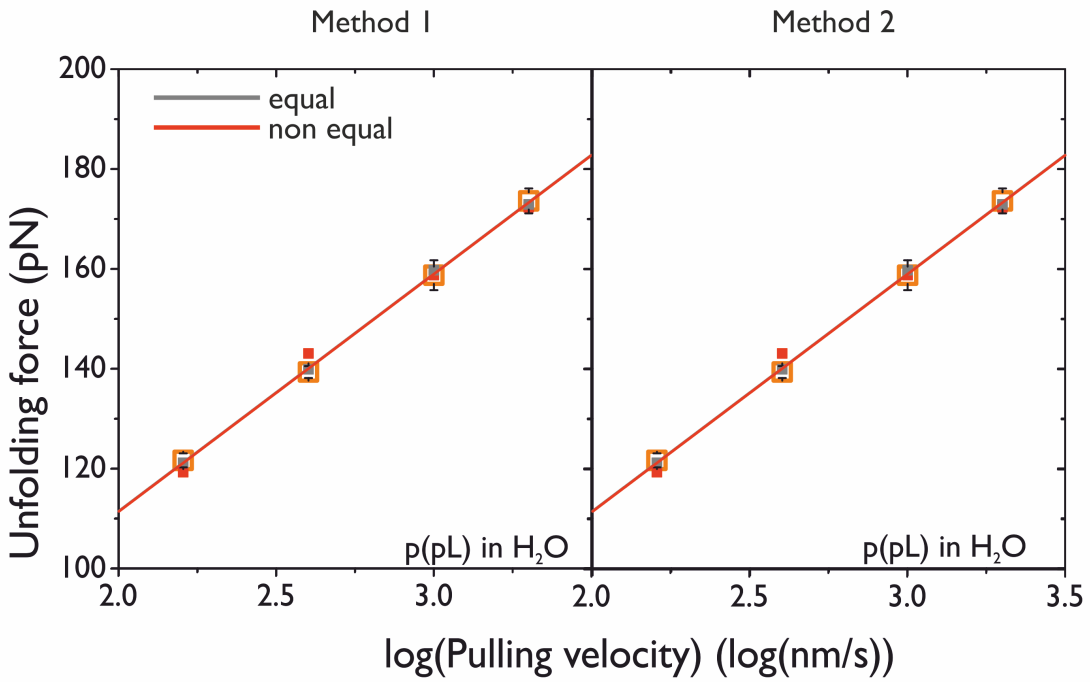


Figure 5.3.5.1: Graph of unfolding force against the logarithm of pulling velocity for the p(pL) in H_2O. The graph shows the experimental forces (open squares) and the simulated forces from the simulation with an equal unfolding probability distribution (grey) or unequal unfolding probability distribution (red).

Method 2

Method 2, described in Section 5.2, was also implemented for this data and the results from these simulations are given in Table 5.3.5.1. The average x_u for the unequal simulations increased to 0.23 nm from 0.22 nm for the equal simulations. This small change in x_u did not significantly perturb the height of the energy barrier. The change in $\Delta G^0_{TS,force}$ was +0.36 kJ/mol, +0.36 kJ/mol, +0.37 kJ/mol and +0.12 kJ/mol for the p(pL) in H_2O, p(pL) in D_2O, d(pL) in H_2O and d(pL) in D_2O respectively. The spring constants for the p(pL) in H_2O, p(pL) in D_2O, d(pL) in H_2O and d(pL) in D_2O are given in Table 5.3.5.1. The spring constants for the four pL systems suggests an increase of 0.1 N/m in the stiffness of the deuterated and protiated protein in D_2O. The values obtained from these unequal probability distribution simulations are $\sim$ 0.2 N/m lower than those obtained from the equal unfolding probability distributions. This suggests that the stiffness of the chain may be sensitive to the unfolding history. This is expected because it will affect the compliance of the chain which, in turn, would change the effective loading rate [328]. This result could suggest that the simulations using an equal unfolding probability distribution may mask some of the changes in the stiffness of a polyprotein.

5.3.6 The four I27 test systems

The parameters used for each I27 system in these simulations can be found in the Appendix D. The percentages illustrated in Figures 5.3.4.2 were inputed into the simulation. This allowed the number of iterations in the simulation of two, three, four and five unfolding events to be determined.

Method 1

The best fit parameters from method 1 for this method agree well with the values obtained in Section 5.2.2. For example the p(I27) in H_2O system gave an x_u = 0.29 nm and $k^0_{u,force}$ = 9.5 $\times 10^{-4}$ s^{-1} for the MC simulations assuming an equal unfolding probability distribution. In these simulations the p(I27) in H_2O system gave an x_u = 0.29 nm and $k^0_{u,force}$ = 12.0 $\times 10^{-4}$ s^{-1}. Therefore there is only a small difference between the $k^0_{u,force}$ (2.5 $\times 10^{-4}$ s^{-1}). Furthermore negligible changes were observed between the agreement of the simulated unfolding with the experimental unfolding forces for the equal and unequal unfolding probability distributions (figures are shown in the Appendix D for all I27 systems). Similarly the change in $k^0_{u,force}$ for the p(I27) in D_2O, d(I27) in H_2O and d(I27) in D_2O systems were comparable between the two different simulations: -3.5 $\times 10^{-4}$, -3.0 $\times 10^{-4}$ and +12.5 $\times 10^{-4}$ respectively. The spring constant for the p(I27) in H_2O was calculated to be 2.0 $\pm$ 0.2 N/m which in comparable with the spring constant obtained from the simulations using an equal probability distribution (2.02 N/m). Similar results were observed for the other systems. These results suggested that the computed stiffness for the I27 systems was not affected by the unequal unfolding probability distribution. The spring constants for the p(I27) in D_2O, d(I27) in H_2O and d(I27) in D_2O (3.3 $\pm$ 0.4 N/m, 2.5 $\pm$ 0.3 N/m and 3.1 $\pm$ 0.3 N/m respectively) protein systems were similar to those obtained by the previous simulations.

Method 2

The values of x_u were also fixed for the I27 test systems to reduce the error associated with the $k^0_{u,force}$. The values were fixed using 0.28 nm for the p(I27) in H_2O and d(I27) in H_2O, and 0.24 nm for the p(I27) in D_2O and d(I27) in D_2O. The average x_u for the p(I27) in D_2O and d(I27) in D_2O was 0.01 nm higher than previously. This did not significantly affect the best fit $k^0_{u,force}$ obtained (-3.5 $\times 10^{-4}$ s^{-1} for both systems). Therefore the heights of the barrier obtained from these simulations was not significantly affected (+0.6 and +0.4

Variable	System			
Method 1	p(I27) in H_2O	p(I27) in D_2O	d(I27) in H_2O	d(I27) in D_2O
B.f. x_u (nm)	0.29	0.24	0.27	0.23
Min x_u (nm)	0.28	0.20	0.23	0.21
Max x_u (nm)	0.32	0.31	0.33	0.28
Average x_u (nm)	0.29	0.22	0.25	0.23
Sd x_u (nm)	0.02	0.02	0.02	0.02
B.f. $k^0_{u,force}$ (s^{-1})	12.0×10^{-4}	17.0×10^{-4}	15.5×10^{-4}	43.5×10^{-4}
Max $k^0_{u,force}$ (s^{-1})	28.0×10^{-4}	132.5×10^{-4}	94.5×10^{-4}	122.0×10^{-4}
Min $k^0_{u,force}$ (s^{-1})	3.5×10^{-4}	0.5×10^{-4}	1.0×10^{-4}	4.0×10^{-4}
Average $k^0_{u,force}$ (s^{-1})	13.4×10^{-4}	65.2×10^{-4}	45.5×10^{-4}	67.4×10^{-4}
Sd $k^0_{u,force}$ (s^{-1})	7.8×10^{-4}	40.5×10^{-4}	29.4×10^{-4}	36.6×10^{-4}
Av. $\Delta G^0_{TS,force}$ (kJ/mol)	50.7	46.7	48.1	46.3
Sd $\Delta G^0_{TS,force}$ (kJ/mol)	1.6	2.5	2.4	1.9
Av. D (N/m)	2.0	3.3	2.5	3.1
Sd D(N/m)	0.2	0.4	0.3	0.3
Method 2	p(I27) in H_2O	p(I27) in D_2O	d(I27) in H_2O	d(I27) in D_2O
B.f. x_u (nm)	0.28	0.24	0.28	0.24
B.f. $k^0_{u,force}$ (s^{-1})	18.0×10^{-4}	17.0×10^{-4}	10.0×10^{-4}	27.5×10^{-4}
Max $k^0_{u,force}$ (s^{-1})	19.5×10^{-4}	20.0×10^{-4}	12.0×10^{-4}	32.0×10^{-4}
Min $k^0_{u,force}$ (s^{-1})	14.0×10^{-4}	5.0×10^{-4}	8.5×10^{-4}	23.5×10^{-4}
Average $k^0_{u,force}$ (s^{-1})	18.0×10^{-4}	17.0×10^{-4}	10.3×10^{-4}	27.8×10^{-4}
Sd $k^0_{u,force}$ (s^{-1})	1.1×10^{-4}	2.0×10^{-4}	1.2×10^{-4}	2.7×10^{-4}
Av. $\Delta G^0_{TS,force}$ (kJ/mol)	49.6	49.2	51.0	48.0
Sd $\Delta G^0_{TS,force}$ (kJ/mol)	0.2	0.3	0.3	0.2
Av. D (N/m)	2.10	2.83	2.16	2.74
Sd D(N/m)	0.01	0.02	0.01	0.01

Table 5.3.6.1: Statistics obtained from the MC simulations for the four I27 test systems: p(I27) in H_2O, p(I27) in D_2O, d(I27) in H_2O, d(I27) in D_2O. The MC simulations incorporate the unequal probability distributions observed in SMFS experiments. Two methods were used to obtain the parameters x_u; the distance between the unfolded and transition state in a two-state unfolding energy landscape, $k^0_{u,force}$; the rate of unfolding at zero force, $\Delta G^0_{TS,force}$; the height of the transition barrier to unfolding and D; the spring constant of the protein. Method 1 determined the best fit to the experimental data, allowing x_u to vary. Method 2 determined the best fit to the experimental data assuming a fixed x_u. B.f. indicates the best fit parameters. Av. is the average of the accepted x_u and $k^0_{u,force}$, Sd is the standard deviation of the values. The Av. and Sd $\Delta G^0_{TS,force}$ was the average and standard deviation taken from calculating $\Delta G^0_{TS,force}$ for every accepted $k^0_{u,force}$.

kJ/mol for the p(I27) in D_2O and d(I27) in D_2O respectively). The barrier heights were also similar for the p(I27) in H_2O and d(I27) in H_2O systems (-0.5 kJ/mol and -0.6 kJ/mol respectively). Similarly the spring constants were not significantly perturbed. The spring constant for four I27 test systems are given in Table 5.3.6.1. The differences observed between the spring constants for the I27 proteins in D_2O and H_2O is smaller than obtained in the equal unfolding probability distributions (≈ 0.7 N/m compared with 1.0 N/m). This may suggest that the equal unfolding probability distribution may overestimate the spring constants in D_2O. Therefore this method may be more sensitive to the small differences observed in these systems.

5.4 How is the protein structure at the transition state affected by deuterium?

As discussed in Section 1.6.4.1, Φ-value analysis can be performed for proteins that have been mutated. The Φ-value is defined as:

$$\Phi = 1 - \frac{\Delta\Delta G^0_{TS,force}}{\Delta\Delta G^0_{u,force}} \tag{5.5}$$

Which compares the change in barrier height of the transition barrier, $\Delta\Delta G^0_{TS,force}$, compared to the change in the free energy difference between the folded and unfolded state of the protein, $\Delta\Delta G^0_{u,force}$ (assuming the folded states have similar energies in chemical and mechanical denaturation experiments). It measures how much structure is formed at the transition state relative to the denatured state. A value of 1 indicates that the structure is fully formed at the transition state and a value of 0 means it is fully unstructured [264, 386]. Values outside of $0 \leqslant \Phi \leqslant 1$ are typically disregarded as an artefact of the experiment [387]. However, it has been suggested, for chemical denaturation Φ-value analysis, that the Φ-value can be indicative of the a change in the speed of folding. In fact, for protein folding kinetic, if the Φ-value is less than 0 that the mutation destabilises the native state of the protein and increases the speed of folding [387]. The Φ-value analysis has been implemented for each of the four protein systems, I27 and protein L. It is assumed that deuterium can be treated as a mutation of hydrogen atoms within the protein. Best *et al.* (2002) have shown that Φ-value can be determined from three methods [264].

Method A:

$$\Phi = 1 + RT\frac{\ln\left(\frac{k_u^{0,mut}}{k_u^{0,w.t.}}\right)}{\Delta\Delta G_{u,chem}} \tag{5.6}$$

The Φ-value analysis can be obtained from the unfolding rates obtained from the Monte Carlo simulations. Where $k_u^{0,w.t.}$ is the rate of unfolding of the p(protein) in H_2O and $k_u^{0,mut}$ is one of the other test systems. Whilst this method is simple and robust, the errors in the unfolding rate are large. Therefore the error associated with the Φ-value analysis is large. Alternatively, by assuming a fixed x_u, the error in the unfolding rate can be greatly reduced.

Method B:

$$\Phi = 1 + RT\frac{\ln\left(\frac{v_{0,mut}}{v_{0,w.t.}}\right)}{\Delta\Delta G_{u,chem}} \tag{5.7}$$

This Φ-value is determined from the velocities at a given force. Where $v_{0,mut}$ is the velocity for the test system involving deuterium and $v_{0,w,t}$ is the velocity for the wild type protein.

Method C:

$$\Phi = 1 - \frac{RT\frac{F_{UN,w.t.} - F_{UN,mut}}{m}}{\Delta\Delta G_{u,chem}} \tag{5.8}$$

This Φ-value is determined from the force of each system at a given velocity. Where $F_{UN,mut}$ is the velocity for the test system involving deuterium and $F_{UN,w,t}$ is the velocity for the wild type protein and m is the average slope from the four I27 or pL test systems.

Methods B and C are, in principle, the same, as long as the x_u does not change, and can be determined from a line fit with an average slope or directly from the experimental forces [264]. Methods 1 and 3 are implemented in this chapter on the experimental fit, using the rate determined from varying x_u and the rate obtained from fixing x_u. The averages of the energy landscape parameters obtained from the MC simulations incorporating the unequal probability were used to determine Φ-values for both pL and I27 protein systems to determine how deuterium affects the structure of the protein at the mechanical unfolding transition state.

5.4.1 The four pL test systems

Table 5.4.1.1 presents the results from the Φ analysis on the pL test systems. There are obvious deviations between the different methods. The value for the d(pL) in D_2O is not within 0 and 1 for all three methods employed. There is also a large error associated with this value due to the large error associated with $\Delta\Delta G^0_{u,chem}$. It is therefore difficult interpret these results. The values of Φ for the p(pL) in D_2O and d(pL) in H_2O, are taken to be ∼0. It should also be noted that, whilst the Φ-value determined from the MC

System		Method A	Method C			
			Line fit[c]		Direct fit[d]	
	$\Delta\Delta G_{u,chem}$[a] (kJ/mol)	MC 2[b] Φ	$F_{UN}^{1000\ nm/s}$ (pN)	Φ	$F_{UN}^{1000\ nm/s}$ (pN)	Φ
p(pL) in H_2O	–	–	158.8 (0.3)	–	158.7 (3.0)	–
p(pL) in D_2O	-1.8 (0.8)	0.03 (0.05)	178.9 (0.3)	-0.35 (0.15)	178.3 (0.2)	-0.32 (0.14)
d(pL) in H_2O	1.1 (0.6)	0.07 (0.07)	148.7 (0.4)	-0.12 (0.07)	148.4 (1.8)	-0.13 (0.24)
d(pL) in D_2O	-0.4 (0.8)	-2.55 (4.51)	170.8 (0.6)	-2.58 (4.56)	169.8 (0.6)	-2.31 (4.09)

Table 5.4.1.1: Table illustrating the Φ-values obtained for the four pL test systems using two different methods. The Φ-value is calculated using $\Phi = 1 - \frac{\Delta\Delta G^0_{TS,force}}{\Delta\Delta G^0_{u,force}}$. Method A uses the unfolding rates obtained from simulations to determine the $\Delta\Delta G^0_{TS,force}$, where $\Delta\Delta G^0_{TS,force} = -RT\frac{k_u^{0,mut}}{k_u^{0,w.t.}}$. Method B uses forces at a single velocity (1000 nm/s) to determine the $\Delta\Delta G^0_{TS,force}$ using:$\Delta\Delta G^0_{TS,force} = RT\frac{F_{UN,w.t.}-F_{UN,mut}}{m}$. a $\Delta\Delta G_{u,chem} = \Delta G^{p(pL)inH_2O}_{u,chem} - \Delta G^{system}_{u,chem}$ values are taken from the equilibrium chemical denaturation experiments outlined in Chapter 4 b $k_u^{0,w.t.}$ from MC simulations where the parameters were obtained using method 2 c Forces from a line fit to the experimental data with average slope of 20.28 pN d Experimental forces

simulatons is corrected for the change in viscosity of D_2O, it was not corrected when using the line and direct fits. This would result in an over-estimation of the $\Delta\Delta G^0_{TS,force}$ for the pL systems in D_2O.

5.4.2 The four I27 test systems

System		Method A	Method C			
			Line fit[c]		Direct fit[d]	
	$\Delta\Delta G_{u,chem}$[a] (kJ/mol)	MC 2[b] Φ	$F_{UN}^{1000\ nm/s}$ (pN)	Φ	$F_{UN}^{1000\ nm/s}$ (pN)	Φ
p(I27) in H_2O	–	–	182.0 (1.5)	–	180.0 (2.7)	–
p(I27) in D_2O	-4.8 (0.5)	1.08 (0.90)	212.5 (1.1)	0.62 (0.08)	215.1 (2.9)	0.56 (0.07)
d(I27) in H_2O	4.3 (0.5)	1.32 (0.34)	191.0 (0.7)	1.13 (0.24)	188.7 (1.3)	1.12 (0.39)
d(I27) in D_2O	-2.6 (0.5)	1.62 (0.40)	204.6 (0.9)	0.48 (0.09)	204.5 (1.6)	0.44 (0.06)

Table 5.4.2.1: Table illustrating the Φ-values obtained for the four I27 test systems using two different methods. The Φ-value is calculated using $\Phi = 1 - \frac{\Delta\Delta G^0_{TS,force}}{\Delta\Delta G^0_{u,force}}$. Method A uses the unfolding rates obtained from simulations to determine the $\Delta\Delta G^0_{TS,force}$, where $\Delta\Delta G_{TS,force} = -RT\frac{k_u^{0,mut}}{k_u^{0,w.t.}}$. Method B uses forces at a single velocity (1000 nm/s) to determine the $\Delta\Delta G^0_{TS,force}$ using:$\Delta\Delta G^0_{TS,force} = RT\frac{F_{UN,w.t.}-F_{UN,mut}}{m}$. a $\Delta\Delta G_{u,chem} = \Delta G^{p(pL)inH_2O}_{u,chem} - \Delta G^{system}_{u,chem}$ values are taken from the equilibrium chemical denaturation experiments outlined in Chapter 4 b $k_u^{0,w.t.}$ from MC simulations where the parameters were obtained using method 2 c Forces from a line fit to the experimental data with average slope of 20.28 pN d Experimental forces

Table 5.4.2.1 presents the results from Φ-value analysis on the data from mechanical unfolding experiments on the I27 systems. The data presented again shows large deviations between the different methods for data interpretation. As the $\Delta\Delta G_{TS,force}$ computed from the experimental results was not corrected for a change in viscosity, the results may over-estimate the computed $\Delta\Delta G^0_{TS,force}$ for the I27 systems in D_2O. Furthermore, there is

evidence to suggest the x_u of the p(I27) in D_2O and d(I27) in D_2O does change. Therefore Method C is an oversimplification for these systems. However, it can be noted that the value of Φ for the d(I27) in H_2O is ≈ 1. This suggests the contacts of the protein at the transition state are native-like, therefore a reduction in the hydrophobic interaction does not affect these contacts.

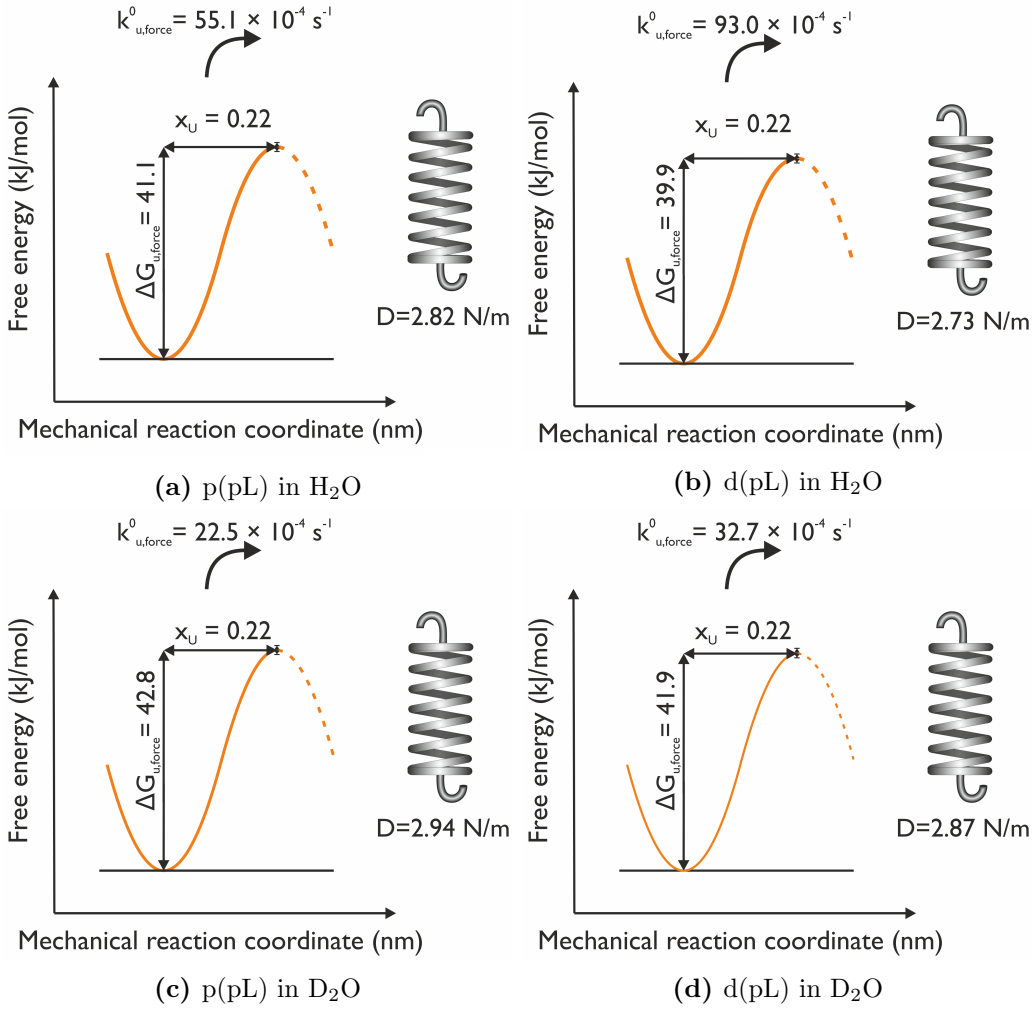
5.5 Discussion

The aims of this chapter included obtaining accurate energy landscape parameters for the pL and I27 test systems to determine how isotopic substitutions within the protein or solvent affected the energy landscapes of the proteins. Once the parameters were obtained, Φ-value analysis was implemented to determine how the structure of the protein was affected by deuterium at the transition state. This can often yield information about contacts that are important in the mechanical unfolding pathway of a protein [124, 245, 255]. However, it appears as though the values may be hard to interpret in this project due to the small changes in the $\Delta G_{u,chem}$[263].

5.5.1 How were the energy landscapes of the proteins affected by isotopic substitution?

The initial MC simulations incorporated an equal unfolding probability distributions. This means there was an equal probability of picking up and unfolding 2, 3, 4 and 5 protein unfolding events in the simulation. The simulations were initially used to obtained the energy landscape parameters, namely x_u and $k^0_{u,force}$, for the pL and I27 test systems. Figures 5.2.2.5 and 5.2.1.5, illustrate that there was a close agreement between the simulated data, yielding the best fit parameters (x_u and $k^0_{u,force}$), and the experimental data.

It has previously been shown that the $k^0_{u,force}$, obtained from MC simulations has large errors associated with it[264, 328]. This error is often the same order magnitude as the unfolding rate [264, 328]. On the other hand, the error in x_u is lower. Two methods were employed to minimise the errors associated with the energy landscape parameters. Figures 5.5.1.2 a–d illustrate the energy landscape parameters using method 2 for the pL test systems. For this method the $x_u = 0.22$ nm was determined from the average obtained from method 1 allowing x_u to vary. Method 2 resulted in a significant decrease in the error in the average $k^0_{u,force}$ without significantly perturbing the associated average $\Delta G^0_{TS,force}$. Despite the large differences observed between the unfolding forces and the



(a) p(pL) in H_2O **(b)** d(pL) in H_2O

(c) p(pL) in D_2O **(d)** d(pL) in D_2O

Figure 5.5.1.1: Schematic illustrating the 1-D energy landscape for each of the protein test systems. The $k^0_{u,force}$, $\Delta G^0_{TS,force}$ (calculated using $\Delta G^0_{TS,force} = RTln\left(\frac{k^0_{u,force}}{A}\right)$) x_u and D (spring constant) have been obtained from MC simulations using a fixed x_u for all pl test systems. Fixing x_u reduced the error associated with the $k^0_{u,force}$ (and therefore $\Delta G^0_{TS,force}$). Despite the large differences in $k^0_{u,force}$ and D, the differences in $\Delta G^0_{TS,force}$ are small. Errors are the standard deviation of the average values.

$k^0_{u,force}$ determined for each of the pL test systems, small changes were observed in the height of the transition barriers and spring constants of the test systems.

For example the rate of unfolding d(pL) in H_2O increased by $\sim$ 62 % compared to the p(pL) in H_2O. This translated to a 3 % decrease in the barrier height of the protein. The small percentage change is due to the prefactor used in the calculation of the barrier height. However, these results do indicate that the transition state is destabilised with respect to the folded state for the d(pL) in H_2O by $\sim$ 1 kJ/mol. This could be associated with a decrease in steric requirement of a C-D bond compared to a C-H bond therefore lowering the activation energy of protein unfolding [130, 137]. The large increase in the rate of unfolding could be associated with a reduction in Van der Waals interactions between side chains therefore causing the protein to favour the unfolded state over the native state. Furthermore, pL has been shown to contain a force rheostat in the core, therefore a reduction in the unfolding rate and force was observed [124].

The average rate of unfolding of the p(pL) in D_2O decreased by $\sim$ 60 %. This was coupled with in increase of barrier height of just $\sim$ 1.7 kJ/mol. This suggests that more favourable interactions must be overcome for the protein to unfolding. The decrease in unfolding rate, and increase in the height of the transition barrier, is consistent with previous studies on a mutant of I27 [132]. In this study, the decrease in unfolding rate was attributed to an enhancement of the hydrophobic interaction due to the increase strength of hydrogen bonds within the solvent.

A $\sim$ 44 % decrease in unfolding rate of the d(pL) in D_2O was also observed in comparison to the p(pL) in H_2O. This could suggest that the decrease observed for the p(pL) in D_2O is also attributed to the increase in the hydrogen bond strength within the protein or the solvent. However, the $\Delta\Delta G^0_{TS,force}$ for the d(pL) in D_2O is $\approx$ 0.9 kJ/mol using the fixed x_u method but only 0.2 kJ/mol using the variable x_u method so this decrease may be an artefact of varying x_u.

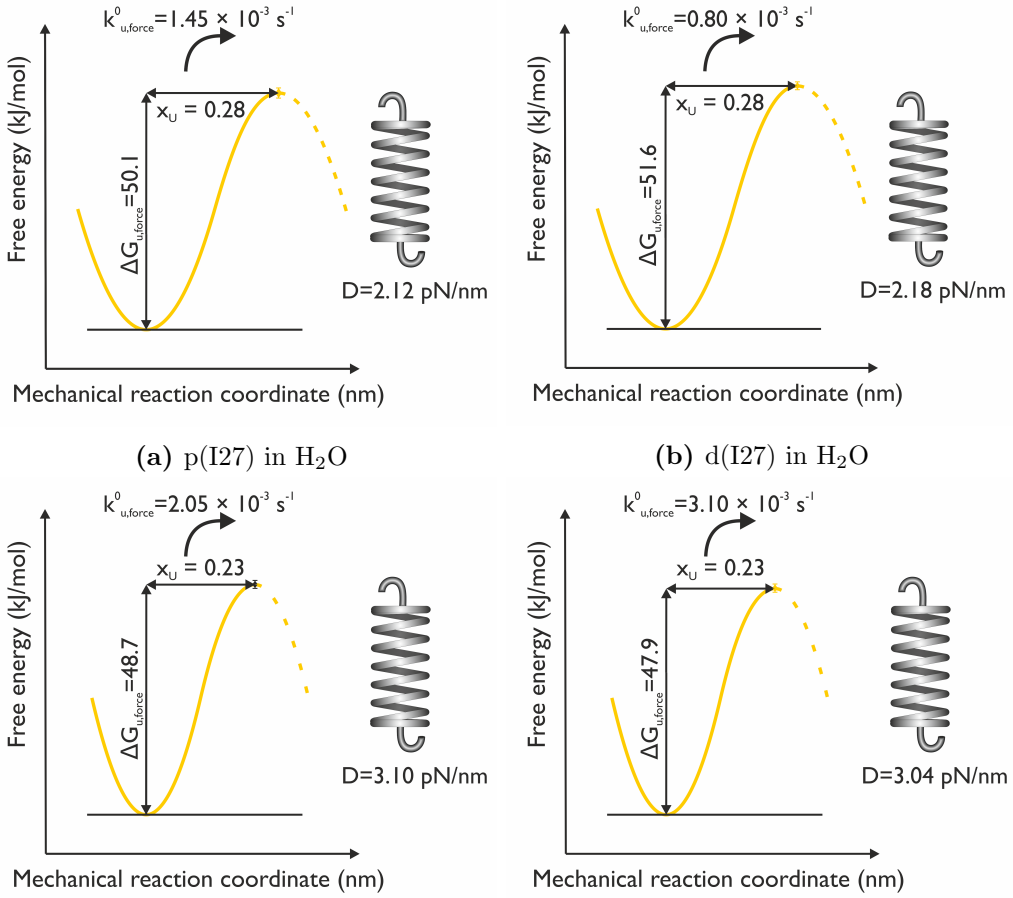
The increase in transition barrier height between the d(pL) in H_2O and d(pL) in D_2O ($\approx$ 2.1 kJ/mol) is larger than the difference observed between the p(pL) in H_2O and p(pL) in D_2O ($\approx$ 1.7 kJ/mol). This suggests that the decrease in hydrophobicity in the d(pL) in H_2O c.f in D_2O is more destabilising that the increase in hydrophobicity observed for the p(pL) in D_2O compared to H_2O.

It has previously been shown that the destabilisation observed in proteins due to isotopic substitution at the non-exchangeable hydrogen sites is more prominent in H_2O than D_2O. This means that typically the $\Delta\Delta G_u$ between the p(protein) and d(protein) in H_2O is greater than the $\Delta\Delta G_u$ between the p(protein) and d(protein) in D_2O. Interestingly, this was not observed in the chemical denaturation experiments on the pL test systems (see Section 3.2.4.1). However, the difference between the $\Delta\Delta G^0_{TS,force}$ p(protein) and d(protein) in H_2O is $\sim$ 1.2 kJ/mol and the difference between the $\Delta\Delta G^0_{TS,force}$ p(protein) and d(protein) in D_2O is $\sim$ 0.8 kJ/mol. This suggests the enhanced destabilisation is observed at the transition states of pL.

It is also interesting to note that the rates of unfolding for the p(pL) in H_2O and d(pL) in H_2O, 56.5 $\times 10^{-3}$ s^{-1} and 91.7 $\times 10^{-3}$ s^{-1} respectively, are similar to those obtained by chemical denaturation experiments, 50.9 $\times 10^{-3}$ s^{-1} and 65.2 $\times 10^{-3}$ s^{-1} respectively (Section 3.2.4.2). This similarity of mechanical and chemical unfolding rates also been observed previously for structurally homologous protein GB1 [269] and could be indicative of the interactions important to the protein unfolding. For these proteins, it has been shown that hydrogen bonds and hydrophobic interactions contribute to the mechanical and chemical unfolding of the proteins [124, 245]. Therefore, whilst for some proteins, chemical and mechanical denaturation probe different unfolding pathways [214, 275], we can hypothesise that proteins involving residues in hydrophobic core along with hydrogen bonds in their mechanical clamps, may have unfolding rates homologous to those obtained from chemical denaturation.

This is also supported by the differences observed between the rates of unfolding for the I27 test systems. It has been shown by Best et al. (2003) that the hydrophobic core of I27 does not contribute to the mechanical unfolding pathway of the protein but does contribute to the chemical unfolding rate [263]. The unfolding rate constants for the p(I27) in H_2O and d(I27) in H_2O (9.5 $\times 10^{-4}$ s^{-1} and 18.5 $\times 10^{-4}$ s^{-1} respectively) are $\sim 3\times$ and $\sim 2\times$ smaller, respectively, than that obtained from chemical denaturation experiments (29.2 $\times 10^{-4}$ s^{-1} and 36.7 $\times 10^{-4}$ s^{-1} respectively). This suggests different unfolding mechanisms probed by the different protein deunaturation techniques and has been observed for this mutant of I27 previously [214]. This further provides evidence that the unfolding mechanism of I27 is dominated by hydrogen bonds, as suggested in Chapter 4, whereas chemical denaturation involves hydrogen bonds and hydrophobic interactions.

Whilst it appears that fixing x_u in the pL systems appears to not significantly alter



(a) p(I27) in H_2O **(b)** d(I27) in H_2O

(c) p(I27) in D_2O **(d)** d(I27) in D_2O

Figure 5.5.1.2: Schematic illustrating the 1-D energy landscape for each of the I27 protein test systems. The $k^0_{u,force}$, $\Delta G^0_{TS,force}$ (calculated using $\Delta G^0_{TS,force} = RTln\left(\frac{k^0_{u,force}}{A}\right)$), x_u and spring constant D, were obtained from best fit parameters using Method 2 in the MC simulations. Method 2c (yellow line) uses a fixed x_u. For the I27 test systems, x_u significantly reduced the error (black and yellow arrows plotted at the transition state) associated with the $\Delta G^0_{TS,force}$ (and therefore $k^0_{u,force}$) furthermore there are more significant differences between the $\Delta G^0_{TS,force}$. Errors are the standard deviation of the average values.

the transition barrier height obtained, the same cannot be said for the I27 test systems. The average x_u across the four I27 protein systems was 0.25 nm (method 2a). This is 0.4 nm smaller than the best fit for the p(I27) in H_2O and reduced the barrier height by $\sim$ 4.1 kJ/mol. Furthermore, the standard deviation of the accepted $k^0_{u,force}$ increased an order of magnitude. Fixing the x_u for all I27 systems to 0.29 nm, the best fit x_u for p(I27) in H_2O in method 2, resulted in large deviations in the fits to the experimental data for the I27 systems in D_2O. Therefore, it appeared as though the differences in x_u observed for the systems in D_2O were significant. Therefore the I27 systems in the homologous solvent environments (H_2O and D_2) were "coupled". The x_u for the p(I27) in H_2O and d(I27) in H_2O was fixed to 0.28 nm. The x_u for the p(I27) in H_2O and d(I27) in D_2O was fixed to 0.23 nm. This decreased the errors associated with all $k^0_{u,force}$ compared with method 1. Furthermore this suggests that D_2O moves the transition state closer to the native state in mechanical denaturation experiments. It has previously been suggested that the movement of the x_u is coupled with a change in the internal dynamics of the protein [378]. However, a change in the x_u for the p(I27) in D_2O with respect to the p(I27) in H_2O, does not agree with a previous SMFS study on an alternative I27 mutant in D_2O [132]. The mutant used in this study, however, has been shown to be mechanically destabilised compared to this alternative pseudo-wild type I27 protein due to lost side-chain contacts of the mutated cysteine residues [214].

One reason for the discrepancies observed between the two different methods, used to obtain the energy landscape parameters, is attributed to the large range of $k^0_{u,force}$ accepted for the p(I27) in D_2O (156.5×10^{-4}), d(I27) in H_2O (106.5×10^{-4}) and d(I27) in D_2O (132.0×10^{-4}) systems. The range of accepted $k^0_{u,force}$ for p(I27) in H_2O is much smaller (8×10^{-4} s^{-1}) therefore, less error is associated with the $\Delta\Delta G^0_{TS,force}$ in method 1. The comparison of these values is therefore not viable. For example, comparing the average transition barrier height from Method 1 suggests that all isotopically labelled I27 systems are destabilised with regard to the p(I27) in H_2O. This results in an increase in the rate of unfolding of all of the systems. Whilst this is still observed for the deuterated systems in method 2, compared to the p(I27) in H_2O, fixing the x_u for the d(I27) in H_2O results in a 1.5 kJ/mol increase in the unfolding barrier of the protein. Similarly, the $\Delta\Delta G^0_{TS,force}$ between the p(I27) in D_2O and d(I27) in D_2O was negligible in method 1. On the other hand, in method 2, the $\Delta\Delta G^0_{TS,force}$ was 1 kJ/mol.

This result is interesting because if the protein unfolding was dominated solely by hydrogen bonds, it would be expected that the barrier height would increase in the

isotopically exchange proteins in D_2O. This is because hydrogen bonds formed by deuterium are stronger than hydrogen bonds formed by hydrogen. In these systems all labile hydrogen positions are fully substituted to D_2O. This could suggest that the unfolding mechanism of the protein is altered as a consequence of the solvent environment. This change due to the solvent, rather than intra-molecular hydrogen bonds, has been noted previously for changes in the internal dynamics of proteins [158] and the refolding kinetics of a protein [170, 175]. The proteins in D_2O have higher spring constants, increasing by $\sim$ 1 N/m compared to the I27 proteins in H_2O. This change was not observed in the pL test systems where the change was almost negligible ($\approx$ 0.1 N/m). This suggests the I27 proteins are more brittle in D_2O. The increased rigidity in a number of proteins in D_2O has been observed previously [158]. The increase in rigidity has been observed as a consequence of a change in the internal dynamics of the protein due to the increased strength of solvent-solvent bonding [140]. This increase in intra-solvent hydrogen bonds often results in a increase in the hydrophobicity in long chain molecules [131, 146, 159]. The hydrophobic core does not play an important role in the unfolding mechanism of I27 [263]. However, simulations suggest that water molecules mediate the unfolding of the protein [377]. An enhanced interaction of the water molecules with the protein upon unfolding could reduce the height of the transition barrier. In fact it has been suggested that the folding kinetics of a protein is slowed due to increased C=O$\cdots$DOD hydrogen bond energies [175]. The increase in the mechanical unfolding force could therefore be attributed to the internal dynamics of the protein [158] or due to the reduction in the hydrogen bond dynamics within the water [172].

The increase in strength of intra-solvent hydrogen bonds in D_2O can also account for the increase in $\Delta G^0_{TS,force}$ of 1 kJ/mol observed between the p(I27) in D_2O and d(I27) in D_2O. Whilst the hydrophobic core does not play a significant role in the mechanical unfolding pathway of I27 [263], side chain contacts can significantly affect the mechanical unfolding of the protein [214, 263]. The interactions are expected to be enhanced in D_2O. However, the side chain interactions fail to explain the increase in barrier height of 1.5 kJ/mol in the d(I27) in H_2O. Whilst this could suggest unfavourable side chain contacts in the p(I27) in H_2O, this is unlikely due to the increased stability of p(I27) in D_2O compared to d(I27) in D_2O. It could, however, be an artefact of fixing the x_u of the protein. In fact, fixing the x_u of the d(I27) in H_2O to 0.26 nm, which is the best fit parameter from method 1, results in an average $\Delta G^0_{TS,force}$=49.4 $\pm$ 0.2 kJ/mol and an average $k^0_{u,force}$=19.0 $\pm$ 1.7 s^{-1}. This suggests a large difference with the associated spring constant (2.43 $\pm$ 0.01 N/m compared with 2.18 $\pm$ 0.02 N/m) but no change in the barrier height of the protein.

Therefore fixing x_u to an average between systems may cause incorrect interpretation of experimental data.

5.5.2 Does incorporating the stochasticity observed in SMFS experiments in the MCS affect the energy landscape parameters?

Monte Carlo simulations typically incorporate an equal probability of picking up 2, 3, 4 or 5 unfolding events [379]. Due to an effect known as the unfolding history [328] of a polyprotein, this can obscure subtle differences between the forces in the protein. Initially the affect of isotopic substitution within the protein and/or solvent on the unfolding history of the unfolding of the full polyprotein domain was reported. It was difficult to accurately observe any changes in the unfolding history effect due to isotopic substitution using the experimental unfolding forces for both pL and I27 test systems. This is due to the low numbers of force-extension traces obtained with the experiments. However, it was apparent that a deviation from a linear increase in force occurred in the p(I27) in H_2O, d(I27) in D_2O and p(I27) in D_2O systems. This minima appeared to have shifted in the systems containing D_2O. This shift was not apparent in the pL test systems. Comparisons of the MC data, using the best fit parameters derived from method 1 in Section 5.2, with the experimental data did not show this apparent shift in the unfolding history behaviour. This could be attributed to subtle changes in the spring constants in the experimental data. In these simulations a spring constant of 40 pN/nm was used. This may be an oversimplification of the systems and lead to in-accuracy in determining the energy landscape parameters from the simulated data. It could also be attributed to changes in compliances of the system that are not accounted for by the simulations which would not be taken into account in the simulations.

We have previously reported on simulations to incorporate an unequal probability distribution into the MC simulations[379]. This unequal probability distribution was determined from the average number of each trace observed across all experiments [379]. To further the investigations into the sensitivity of the MC simulations to the probability of observing 2, 3, 4 or 5 unfolding events, the information from each individual SMFS experiment was used within the simulations. The results of these simulations have been presented in Section 5.3.4. Both methods implemented in Section 5.2 were also used to analyse the data from these simulations.

For the pL systems, the best fit x_u, determined from method 1, increased by 0.01 nm for the p(pL) in H_2O and d(pL) in D_2O. This was coupled with a decrease in the

average $k^0_{u,force}$ by 7.8 $\times 10^{-3}$ s^{-1} for the former system and 8.9 $\times 10^{-3}$ s^{-1} for the latter. Small changes in the $k^0_{u,force}$ were also observed for the p(pL) in D_2O (+4.8 $\times 10^{-3}$ s^{-1}) and d(pL) in H_2O (+1.3 $\times 10^{-3}$ s^{-1}) systems. These changes were not reflected in the associated average $\Delta G^0_{TS,force}$ for all four pL systems; The $\Delta G^0_{TS,force}$ were all the same within error. An increase in the best fit x_u from method 1 for two systems led to a 0.01 nm increase in the average x_u across the four test systems. Fixing the x_u to this average value using method 2 did not significantly perturb the associated $k^0_{u,force}$ (p(pL) in H_2O $\Delta k^0_{u,force}$ = -7.6 $\times 10^{-3}$ s^{-1}, p(pL) in D_2O $\Delta k^0_{u,force}$ = +3.4 $\times 10^{-3}$ s^{-1}, d(pL) in H_2O $\Delta k^0_{u,force}$ = -12.7 $\times 10^{-3}$ s^{-1} and d(pL) in D_2O $\Delta k^0_{u,force}$ = -9.5 $\times 10^{-3}$ s^{-1}). This therefore did not result in changes with the associated transition barrier heights but resulted in a decrease with the associated spring constants of the system. This is expected because the calculation of the compliance of the protein (the inverse of the spring constant) would be most sensitive to the unfolding history effect.

Similar results were obtained for the I27 test systems using method 1. However, method 2 suggested that the difference between the barrier heights of the p(I27) in H_2O and p(I27) in D_2O was only 0.4 kJ/mol. This is a 1.1 kJ/mol decrease in the difference observed in Section 5.2. The results indicate only minor differences in the energy landscape parameters obtained by both simulations but more significant changes in the calculated stiffness of the protein.

Whilst these results suggest that the simulations are insensitive to the probability distribution observed experimentally, they may yield more information about the stiffness of the protein chain. The subtle change in energy landscape parameters obtained is in agreement with the previous study [379] and is important because it justifies the selection criteria used previously in SMFS experiments. However, more accurate information may be obtained by mimicking the experiments more closely.

5.5.3 Which interactions are important at the transition state of the protein?

Φ-value analysis can help identify interactions which are important in denaturation of a protein. It is often used to interpret the chemical denaturation data of mutated proteins. In this study it has been utilised to determine whether specific interactions are important at the transition state of a protein. The results for the protein L test systems suggest that Φ-value is $\approx$ 0 for the p(pL) in D_2O and d(pL) in H_2O. This suggests that the native contacts at the transition state are disrupted in these systems. This suggests

that substitution of deuterium within the protein and in the solvent disrupts the native contacts upon folding. Therefore the protein would not be native like at the transition state. The methods implemented indicate a large negative Φ-value for the d(pL) in D_2O. Typically this value is disregarded as an artefact, however it has been suggested it could be indicative of the acceleration of protein folding [387]. The results presented here are hard to interpret due to the large errors associated with the Φ-value. Furthermore, it has previously been suggested that Φ-values should only be determined for systems where the $\Delta\Delta G_{u,chem}$ is greater than $\sim$ 8 kJ/mol [263]. This would mean it is not suitable to use this analysis in this project.

For the d(I27) in H_2O it is assumed that the Φ=1 within the experimental error and is similar across the MC method, line fit and direct fit. This suggests all native contacts are maintained within this protein. This is expected because the hydrogen bonds are not expected to have been perturbed in this system. It appears as though the Φ-value for the d(I27) in D_2O is >1 which could suggest a deceleration in the protein folding [387]. It appears as though the Φ-value for the p(I27) in D_2O is $\approx$ 1. This would also suggest that all native contacts are formed at the transition state. However, the error associated with the p(I27) in D_2O is too large to be certain of this. The differences observed between the Φ-value computed from the MC simulations, and the experimental forces are expected for the p(I27) in D_2O and d(I27) in D_2O due to the change in x_u of the system.

5.5.4 Summary

The results presented in this chapter indicate that the effect of deuterium on the energy landscape of proteins may depend on the mechanical unfolding pathway of the protein and the interactions involved within the mechanical clamp region. The error associated with the $k^0_{u,force}$ can be substantially reduced by fixing x_u, however, this should be done with caution. These results suggest that fixing the x_u may lead to misinterpretation of the experimental data. Therefore, we suggest an alternative method for reducing the error in $k^0_{u,force}$. Instead of using an average x_u across the four protein systems in the analysis using method 2, the x_u should be fixed using the best fit x_u determined from method 1 allowing the x_u to fluctuate by the x_u increment used in the simulation. In this experiment this would allow for $x_u \pm 0.01$ nm. This chapter also illustrated, and further supported, the insensitivity of the MC simulations using the Zhurkov-Bell model to the unequal unfolding distribution observed in experiments [379]. This shows that, despite obtaining a higher number of traces containing fewer protein unfolding events, the energy landscape parameters obtained from simulations are not significantly perturbed. However,

using an equal unfolding probability distribution may mask differences in the stiffness of the proteins which can be related to protein dynamics [158].

Chapter 6

Discussion and future work

Investigating protein folding, unfolding and stability has been an important endeavour for over 50 years [18, 21, 388]. Since the introduction of SMFS experiments a wealth of information has emerged regarding the mechanical stability of proteins [202, 212, 312]. This has enabled key interactions in the mechanical stability of proteins to be identified, and differentiated from those required for protein stability against other denaturants such as chemicals [214]. It has further allowed different unfolding pathways of proteins to be mapped and intermediate states to be identified[231, 238, 261, 389, 390]. Point mutations enable the role of specific residues in the mechanical, or chemical denaturation of proteins to be elucidated [124, 220, 245, 263]. However, little is known about the net contribution of each type of interaction (i.e, hydrophobic interactions, hydrogen bonds, electrostatic interactions) to the mechanical stability of proteins. Hydrogen bond patterns within protein structures are often used to predict the proteins mechanical stability [240, 241, 247]. This ignores the contribution of hydrophobic interactions within the protein and hydrogen bonding within the solvent which could be important [132, 391].

Here we have presented a novel method to determine the key contributions of different interactions to the mechanical stability of proteins. Isotopic substitution of hydrogen for deuterium within the protein and/or the solvent has been used to address this question in two model proteins, I27 and pL.

6.1 How does deuterium affect the stability of the two model proteins?

6.1.1 Thermodynamic stability

In Chapter 3, the properties of deuterated monomeric and pentameric (polyprotein) constructs were determined and compared to the protiated homologues in H_2O and D_2O. We determined that structures of the poly and monomeric proteins appeared to be largely unperturbed. However, significant differences were noted in the thermodynamic properties of the monomeric I27 and pL protein systems.

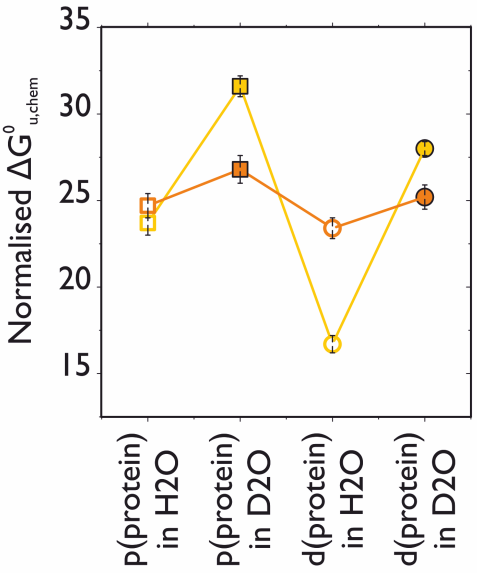


Figure 6.1.1.1: Normalised $\Delta G_{u,chem}$ for each of the I27 (yellow) and pL (orange) protein test systems: p(protein) in H_2O (open square), p(protein) in D_2O (closed square), d(protein) in H_2O (open circle) and d(protein) in D_2O (closed circles). The $\Delta G_{u,chem}$ for each system are normalised to the sum of all four protein $\Delta G_{u,chem}$ to show the relative differences between each of the systems. If all systems had the same $\Delta G_{u,chem}$ then the values would center around 25. More significant differences are observed for the I27 protein systems than the pL protein systems.

The normalised $\Delta G_{u,chem}$ for both the pL and I27 protein test systems are shown in Figure 6.1.1.1. These results indicated that the thermodynamic stability of the I27 protein was more sensitive to isotopic substitution within the protein and the solvent environment than the pL protein.

For both I27 and pL d(protein) in H_2O systems, it was predicted that there would be a reduction in the strength of the hydrophobic interaction compared to the p(protein) in H_2O, whereas the hydrogen bond strength and ionic interactions would be maintained [130, 137, 151]. This is attributed to the shorter C-D bond, reducing the steric requirement of the side chain and therefore decreasing the hydrophobic interaction within the

core [130, 137]. As predicted, for both pL and I27, the thermodynamic stability in the d(protein) in H_2O system decreased. However, the decrease in the normalised $\Delta G_{u,chem}$ with respect to the p(I27) in H_2O was much more substantial for the d(I27) in H_2O (-7.9) than the for the d(pL) protein in H_2O (-2.2). This could be attributed to the differences in hydrophobic core sizes, due to the size of the proteins. It has previously been shown that hydrophobic interactions contribute less to the thermodynamic stability of small proteins compared to large proteins [122]. In this study, the protein I27 is 89 amino acids long, whereas protein L contains 67 amino acids this corresponds to an estimated radius of gyration (R_G) of 13.2 and 12.4 $\AA$ respectively estimated using [392]. These changes are small, in fact there is a 28.2 % difference in the number of amino acids and a 6.3 % difference in R_G compared to a 88.5 % different in the normalised $G_{u,chem}$. It should be noted however, that the $\Delta G_{u,chem}$ for the d(I27) protein was taken from the equilibrium denaturation experiments; The lack of pre-transitional baseline in the equilibrium curve may result in an underestimation of the $\Delta G_{u,chem}$ determined for this system. Using the values obtained from the kinetic unfolding data for p(I27) in H_2O and d(I27) in H_2O, the difference for I27 became -3.4 which was still marginally larger than the value for pL.

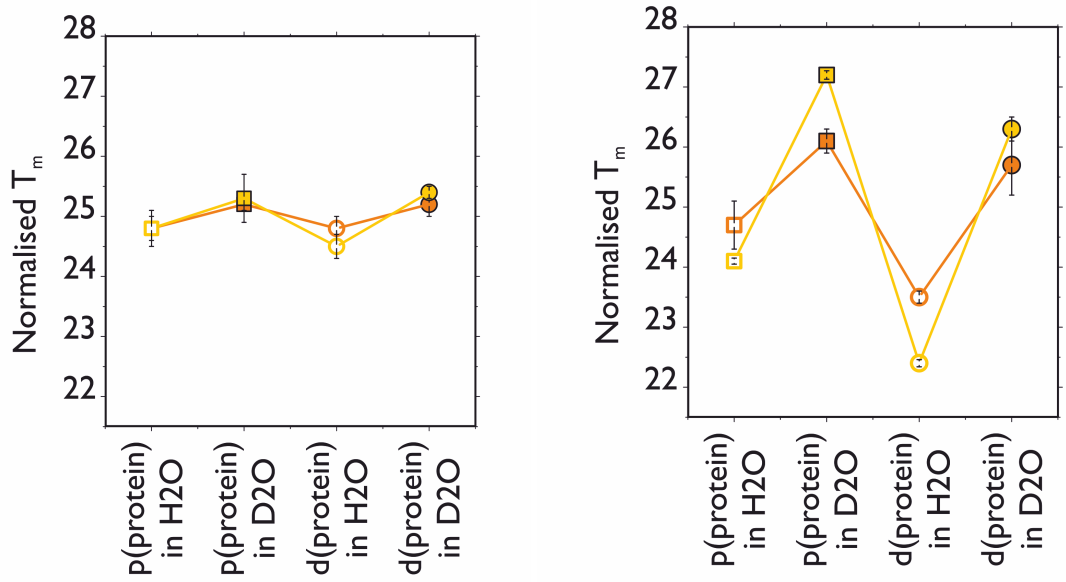
The results also suggest that the increase in stability of the p(protein) in D_2O, compared with the p(protein) in H_2O, for both I27 and pL, could be attributed to an increase in the hydrophobicity. A increase of 4.8 was observed for the protiated I27 and 1.3 for pL in the normalised $\Delta G_{u,chem}$ in D_2O, compared to the values obtained in H_2O. This is at odds with the results for hydrocarbons and ethanol molecules [151, 393]. However, it has been suggested that the approximations used to determine the free energy of transfer in some methods are not sensitive to the changes in hydrophobicity between H_2O and D_2O [394]. It has been shown that hydrophobic interactions are enhanced for nonpolar amino acids with long side chains [151]. An increase in hydrophobic interaction has been observed in studies of the critical micelle concentration [151, 394, 395], protein aggregation and association [165, 396, 397], and protein stability [137, 398]. In fact the rate of protein aggregation in D_2O was only significantly peturbed if hydrophobic interactions were involved in the process[396]. It has been suggested that the increase in stability of protiated proteins in D_2O is due to isotopic substitution within the solvent rather than at the exchangeable sites within the protein [137]. However, it has also been suggested that deuterium bonds do form longer and stronger hydrogen bonds within polypeptides [146, 399]. This may account for the increase in stability of pL and I27 deuterated proteins in D_2O compared to the protiated proteins in H_2O.

The differences observed between the p(I27) in H_2O and d(I27) in H_2O (4.3) were also greater than those observed between d(I27) in D_2O and p(I27) in D_2O (2.2). This is in accord with previous studies [137]. This suggests that the deuteration of methyl groups is more destabilising than the increased hydrogen bond strength in D_2O. However, the opposite was observed for pL. This again could be attributed to the smaller size of the protein and could indicate that, whilst hydrophobic interactions dominate the thermodynamic stability of I27, they are less important for pL. Comparison the differences between the $\Delta G_{u,chem}$ of the p(protein) in H_2O and p(protein) in D_2O and, perhaps,between the d(protein) in H_2O and d(protein) in D_2O for pL and I27 also support this. There is a larger change in the I27 systems (4.8 and 6.9 respectively) than the pL systems (1.8 and 1.5 respectively).

6.1.2 Thermal stability

In Chapter 3, the thermal stability of the monomeric and pentameric proteins were determined. It has previously been suggested that the thermal stability of proteins increases with the number of hydrogen bonds, electrostatic interactions and improved hydrophobic packing within the core of the protein [400, 401]. Therefore significant differences were expected to occur between the systems. The differences in the normalised melting temperature for the monomeric and pentameric pL and I27 systems are shown in Figure 6.1.2.1. Interestingly, the differences between the systems were more apparent between the monomers than the polyproteins. Differences in the stability between monomeric and polyproteins have been noted before for equilibrium of TNfnIII domains [373]. This could be due to the size of the polyprotein chain. More random coil states will form as the polyprotein is unfolded and the signal from these states may dominate in the spectra obtained from CD spectroscopy. In fact the thermal denaturation of these polyproteins is not reversible.

It is apparent that there is a significant decrease in the midpoint temperature of unfolding for the d(protein) in H_2O in comparison to all other protein systems. The decrease between the d(protein) in H_2O and p(protein) in H_2O was greater in I27 (3.1) than pL (1.4). This difference is greater than the difference observed between the d(protein) in D_2O and p(protein) in H_2O for both I27 (0.9) and pL (0.4). This substantial decrease for both proteins suggested that different interactions are important for the thermal and chemical denaturation of these proteins. The thermal denaturation of the proteins was reversible. This enabled thermodynamic parameters to be computed for both proteins. The normalised ΔH_u for pL and I27 and ΔS_u for I27, are shown in Figure 6.1.2.2. The



(a) Pentameric polyprotein **(b)** Monomeric

Figure 6.1.2.1: Normalised melting temperature, the midpoint temperature of unfolding, T_m for each of the I27 (yellow) and pL (orange) polyprotein (left) and monomeric (right) polyprotein test systems. The p(protein) in H_2O are shown with open squares, the p(protein) in D_2O are closed squares, the d(protein) in H_2O are open circles and the d(protein) in D_2O are closed circles. The T_m for each system are normalised to the sum of all four protein T_m to show the relative differences between each of the systems. If all systems had the same T_m then the values would center around 25. Significant differences are observed between the T_m of the protein systems, particularly for the monomeric systems.

results suggested a loss of interactions within the d(protein) in H_2O for both pL and I27 due to a decrease in the ΔH_u compared to the other test systems. Furthermore, the study suggested an increase in the enthalpy for both deuterated and protiated pL and I27 in D_2O compared to the proteins in H_2O. This suggests an increase in the interactions within the systems at some point. This could be attributed to additional intramolecular bonds in the solvent due to the additional propensity for D_2O to form solvent-solvent hydrogen bonds and additional compactness of the protein[140, 158, 402]. This is supported by the decrease in Δ H_u for the deuterated pL and I27 proteins compared to the protiated proteins in both solvents. Changes in the entropy of folding were also noted which were similar to those previously reported for phycocyacins [137].

6.1.3 Mechanical stability

In Chapter 4, the mechanical unfolding forces at four velocities were determined for each of the test systems. The normalised forces, along with the associated normalised spring constant, for the $(I27)_5$ and $(pL)_5$ systems are shown in Figure 6.1.3.1. The forces for pL test systems showed similar trends to those observed in thermal and chemical dentauration experiments. Interestingly, the normalised differences in the forces between

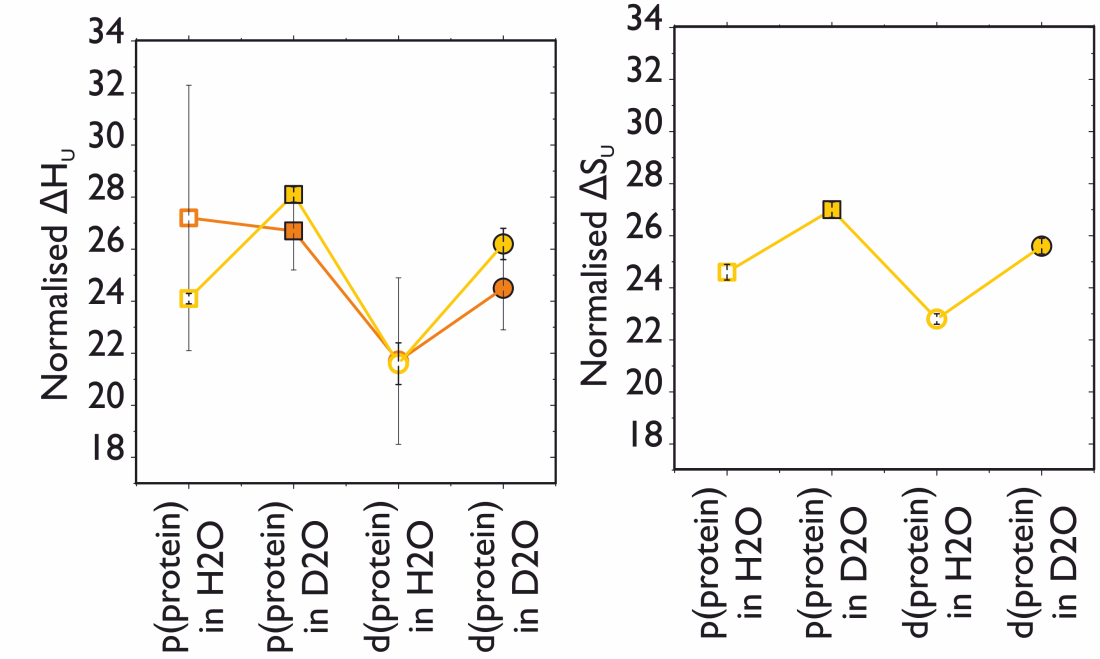
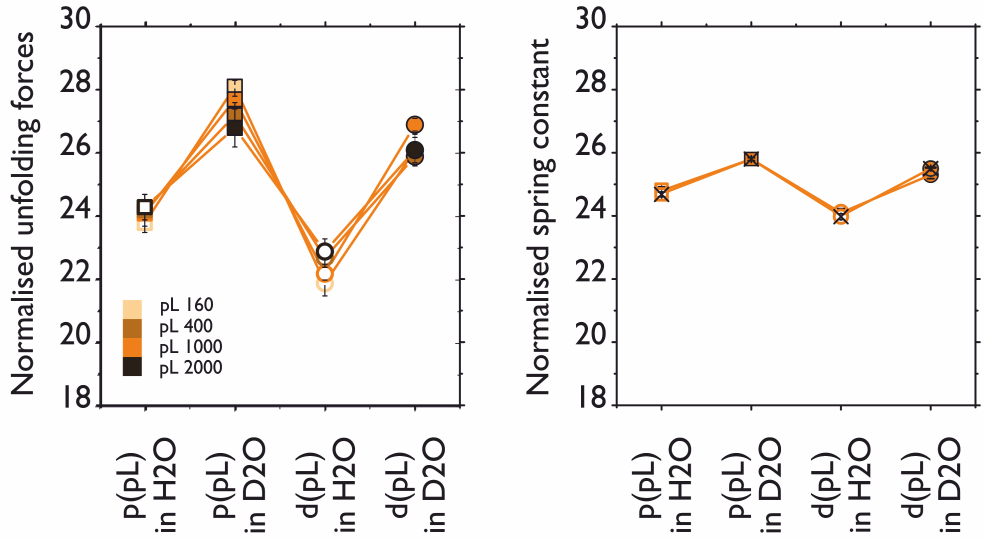


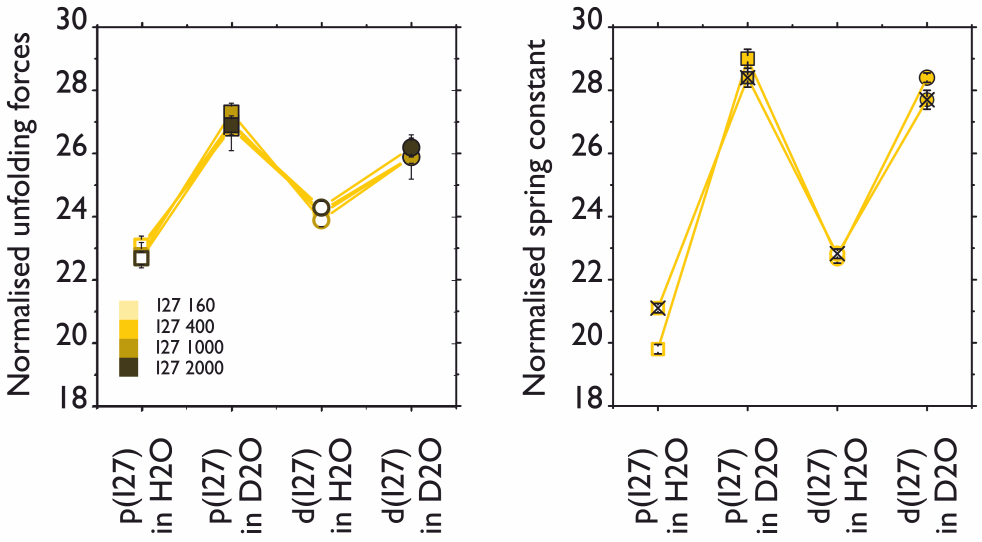
Figure 6.1.2.2: Normalised ΔH_u and ΔS_u for each of the I27 (yellow) and pL (orange) protein test systems: p(protein) in H_2O (open squares), p(protein) in D_2O (closed squares), d(protein) in H_2O (open circles) and d(protein) in D_2O (closed circles). The ΔH_u and ΔS_u for each system are normalised to the sum of ΔH_u and ΔS_u (respectively) of all of the four protein systems to show the relative differences between each of the systems. If all systems had the same ΔH_u and ΔS_u then the values would center around 25. It was not possible to calculate the normalised ΔS_u for the pL test systems.

the pL systems involving deuterium, compared with the p(pL) in H_2O, are similar to those observed between the normalised T_m and $\Delta G_{u,chem}$ for the monomer protein systems.

Figure 6.1.3.1 shows that in the pL systems the force increases for the protiated and deuterated pL in D_2O, with the p(pL) in D_2O having the highest mechanical stability. These differences suggest that there is an increase in intra-molecular hydrogen bonding, in both d(pL) and p(pL) in D_2O, and hydrophobic interactions, in p(pL) in D_2O, due to the isotopic substitution exchangeable positions in the protein and increased hydrogen bond strength in the solvent. The mechanical unfolding force decreases for the d(protein) in H_2O in comparison to the other pL systems. This was expected due to the decrease in hydrophobicity in the protein. However, the force of the d(I27) in H_2O increased with respect to the p(I27) in H_2O. This supports the results by Best et al. (2003), whom suggested that the hydrophobic core plays no role in the mechanical stability of the I27 protein[263] and therefore the force was not significantly reduced by mutations within the core. However, the increase in force for the deuterated I27 is surprising and was coupled with an increase in the spring constant of the protein (Figure 6.1.3.1) due to the movement of the transition barrier to the native state. This movement of the transition



(a) pL test systems



(b) I27 test systems

Figure 6.1.3.1: Normalised forces and associated spring constants for each of the pL (orange, top) and I27 (yellow, bottom) protein test systems: p(protein) in H_2O (open squares), p(protein) in D_2O (closed squares), d(protein) in H_2O (open circles) and d(protein) in D_2O (closed circles). The forces (presented in Chapter 4) and spring constants (presented in Chapter 5) for each system are normalised to the sum of all four protein forces and spring constants respectively to show the relative differences between each of the systems. The crosses in the symbols on the graph corresponding to the normalised spring constant indicate that the parameters were obtained from the unequal probability distribution simulations. If all systems had the same forces, and spring constants, then the values would center around 25. For the d(I27) in H_2O the spring constant was determined for a fixed x_u of 0.26 nm because this best described the data. The x_u for all other systems were presented in Chapter 5.

state has also been observed upon the removal of side chain contacts within the clamp region of the protein [214, 263]. In fact, removal of two amino acids in contact with the G strand of the protein resulted in an increase in force and decrease in x_u of the protein [263]. This could suggest that the d(protein) in H_2O removes some of these side-chain contacts. One hypothesis is that the side chain interactions maintain the malleability of the protein. The wild-type I27 requires mechanical stability to perform its biological function[244]. Therefore, side chain interactions may govern the malleability of the I27 molecule, whilst hydrogen bonds provide the dominant resistance to mechanical unfolding.

The dominance of hydrogen bonding to the mechanical resistance of I27 was also observed by the increase in unfolding force observed in the initial unfolding event of both protiated and deuterated I27 in D_2O without a change in the ratio of the initial:intermediate unfolding event. Increasing the strength of the hydrogen bonds within the solvent also moved the position of the protiated and deuterated I27 protein transition barrier closer to the native state. This is also surprising because it was not observed for another mutant of I27 [132]. However this may be attributed to the loss of key side chain interactions made by the two cysteine residues within the protein: the mutations of these cysteine residues to serine significantly altered the mechanical properties of the protein [214]. The increase in spring constant associated with this movement of the transition state may suggest a change in the interaction of the solvent with the protein. It appears as though, by changing the stiffness of the chain, the effective loading rate upon the proteins is increased which acts to increase the observed mechanical unfolding force. This is supported by the small movement observed in the minima in the unfolding history effect for the d(protein) in D_2O [328]. It has previously been shown that the interaction between the solvent and protein is reduced in D_2O, this has been shown to be coupled with an increase with the rigidity of the protein due to change in the internal dynamics of the protein [158, 178]. The differences observed in Figure 6.1.3.1 between the forces and spring constants of the I27 protein systems in D_2O and H_2O could be indicative of a change in the internal dynamics of the protein. The internal dynamics are thought to be inversely proportional to the thermal stability of proteins [403], therefore this could explain the larger differences observed in the T_m shown in Figure 6.1.2.1. Although, mechanical and thermal denaturation is expected to occur via different pathways [404].

It could also be argued that the movement of the transition state could be a ground state effect [405]. That is, structural changes in the ground state of the protein, causing the unfolded state to move closer to the transition state[405]. Furthermore, the movement

of the transition barrier can also be related to the Hammond effect. The Hammond postulate dictates that the transition state represents the least stable species of the protein[405]. This least stable species may move closer to the native state in the protein. This Hammond behaviour has previously been observed in temperature studies in mechanical denaturation experiments of I27 [213]. It has also been suggested that the movement of the transition state, in mechanical unfolding experiments, is attributed to a change in the internal dynamics of a protein [378]. When a transition state moves closer to the native state fewer states are available to the protein[378]. This again could be attributed to solvent-protein interactions.

6.1.4 Kinetic stability

To determine the effect of isotopic substitution within proteins and the solvent on the transition barrier of the protein, MC simulations were implemented in Chapter 5. These simulations enabled determination of $k^0_{u,force}$ and therefore the $\Delta G_{TS,force}$. Graphs of the normalised values for the pL and I27 test systems are shown in Figure 6.1.4.1. It appears as though the rate of unfolding was was only marginally perturbed by the isotopic substitution for the d(I27) in H_2O and p(I27) in D_2O. The rate of unfolding was more substantially affected in the d(I27) in D_2O. This suggests a combination of deuterium in exchangeable positions and deuteration within the solvent actually destabilises the transition barrier to unfolding. This could suggest that additional side chain interactions are in the d(protein) in D_2O, compared with the d(protein) in H_2O, that mechanically destabilise the protein. These interactions would also be enhanced in the p(I27) in D_2O.

It is also interesting to note that the $k^0_{u,chem}$ for the pL protein was similar to that obtained for $k^0_{u,force}$ for both p(protein) in H_2O and d(protein) in H_2O. This is at odds with mutational studies on pL [124], however the results for $k_{u,chem}$ in this study were taken at 2 M denaturant and therefore were much lower than the ones obtained from the mechanical denaturation experiments. Chemical and mechanical unfolding experiments have been shown to occur via different pathways [214, 275]. This is because chemical denaturants act globally on the protein, lowering the height of the transition and unfolded states of the protein relative to the native state of the protein. Whereas, mechanical denaturation affects the transition barrier in an anisotropic way [223, 352] over a well-defined reaction coordinate. However, this might suggest that, if the same interactions are disturbed at the rate limiting step in the chemical denaturation of pL as in mechanical denaturation experiments, similar unfolding rates would be obtained. That is, the hydrophobic core and hydrogen bonds are equally disturbed in both techniques, then both techniques may probe

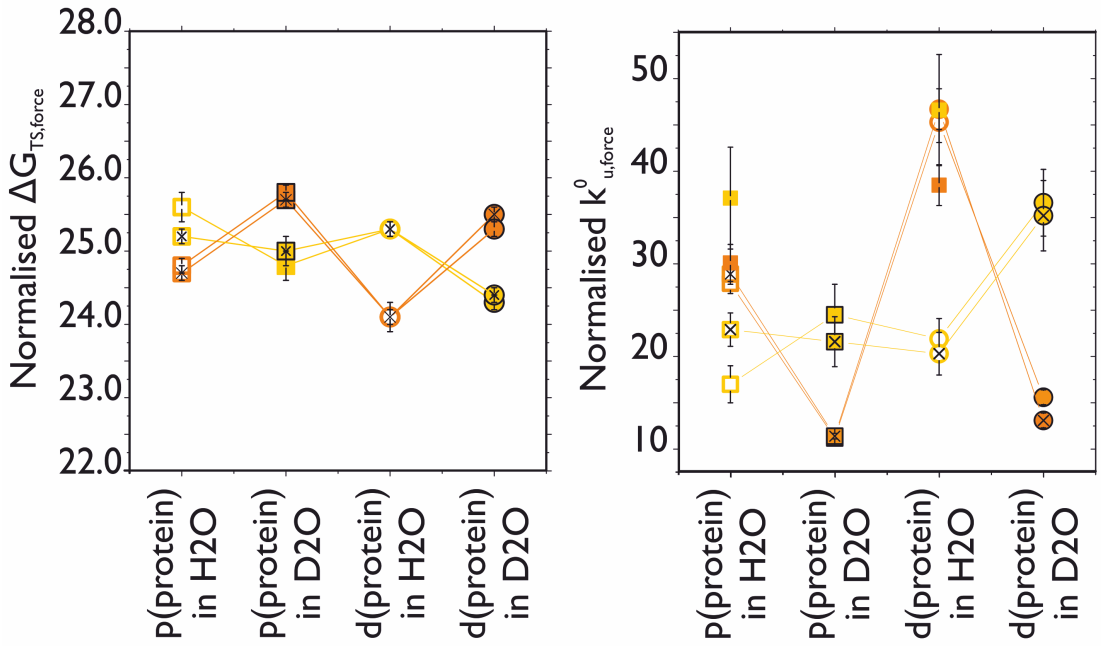


Figure 6.1.4.1: Normalised $\Delta G_{TS,force}$ (left) and corresponding $k^0_{u,force}$ (right) for each of the I27 (yellow) and pL (orange) protein test systems: p(protein) in H_2O (open squares), p(protein) in D_2O (closed squares), d(protein) in H_2O (open circles) and d(protein) in D_2O (closed circles). The $\Delta G_{TS,force}$ and $k^0_{u,force}$ for each system are normalised to the sum of all four protein $\Delta G_{TS,force}$ and $k^0_{u,force}$ respectively to show the relative differences between each of the systems. If all systems had the same $\Delta G_{TS,force}$ and $k^0_{u,force}$ then the values would center around 25. More significant differences are observed for the I27 protein systems than the pL protein systems.

similar transition barriers. In fact, the tryptophan (W47) in pL, probed by fluorescence experiments, is located on the alpha helix which packs onto the β-sheets in the hydrophobic core. This will not be the case for all proteins. For I27, the tryptophan side chain (W43) makes contacts (4 $\AA$ or less) with residues: F30, G41, Q42, K44, I58, L67, V80, S81, F82 and Q83. None of these residues are found within the A or G strands within the protein. This could suggest that the position of tryptophan with respect to the mechanical clamp could determine whether the unfolding rates determined by chemical and mechanical denaturation are similar.

6.2 Future work

How does deuterium affect the interactions in the transition state of a protein?

To determine the effect of deuterium on the structure of the mechanical transition barrier of the proteins, full atomistic data is required. This could be achieved by mechanically unfolding the proteins in *ab initio* MD simulations to take into account the quantum differences between the solvents[148, 149]. To reduce the computational cost, regions

hypothesised to be important to the mechanical stability of proteins could be probed to see how the interactions are affected by the change in solvent. The differences observed in $\Delta G_{u,chem}$ for both proteins were small. Therefore interpretation of the Φ-value analysis is challenging. If one assumes the d(protein) in H_2O only changes the hydrophobic interaction and p(protein) in D_2O changes the hydrogen bonds within the protein: In the pL systems the Φ-value suggested that hydrogen bonds and hydrophobic interactions are fully dissociated at the transition state. On the other hand, the results for I27 suggest that the hydrogen bonds and hydrophobic interactions are fully formed at the transition state. However, this appears to be nonsensical when comparing with the Φ analysis provided by mutational studies [124, 261, 263]. Furthermore MD simulations could also yield information about the movement of the transition state of the protein. Simulations have been performed on small model peptides in D_2O [402]. These suggested an increase in the number of intramolecular hydrogen bonds formed in D_2O and could further explain the differences in forces observed between both pL and I27 systems. However, limited data is available concerning the deuteration of the carbon backbone of the protein. This information could provide useful information to help understand the differences in unfolding rates observed by each of the protein systems.

Is it a universal property for the rates obtained from SMFS experiments to differ from chemical denaturation experiments?

Chemical unfolding of the both the p(protein) and d(protein) should also be performed in D_2O to elucidate any further similarities or differences between the mechanical and chemical denaturation of proteins. Furthermore, it would be useful to compare the rates of folding obtained from chemical denaturation experiments and mechanical refolding experiments. Refolding experiments could also be performed to determine the $k^0_{f,force}$. This would also enable the $\Delta G^0_{u,force}$ to be determined. There are different SMFS techniques that can obtain the rate of folding of proteins [340, 364, 406–409]. Recently, refolding experiments were performed on an alternative mutant of I27, These experiments have been used to probe the refolding and unfolding rates of the protein in equilibrium. A constant force of 4.1 pN was applied across the I27 polyprotein [313]. It was observed at this force that unfolding and refolding was equally probable. This allowed the change free energy between the unfolded and folded states to be determined.

Are the differences observed between the mechanical unfolding forces attributed to the residues involved in the mechanical clamp region?

To verify that the results presented here are indicative of the residues involved in the mechanical clamp, another model protein system should also be probed. It has been noted that fibronectrin protein TNfn3 has a similar structure to I27 domains [220], however, TNfn3 unfolds at significantly lower forces. Studies have indicated that more residues within TNfn3 contribute to the mechanical unfolding pathway than I27. The mechanical unfolding pathway of TNfn3 probed by MD simulations indicated rearrangements of the protein before the mechanical unfolding event within the core and loss of interactions between the A and G strands[220]. Mutations to this protein within the hydrophobic core significantly decreased the unfolding force of the protein [220]. Therefore, we hypothesise this protein will display differences in the mechanical properties of the four test systems similar to those observed in pL. An alternative model to test would be the ankyrin repeat protein N6C. This proteins consists of tandem repeats of alpha helices packed onto one another by extensive hydrophobic interactions and hydrogen bond networks [410].

Could alternative models for effect of force on the 1-D energy landscape reduce the errors associated with the energy landscape parameters?

During this project, one aim was to increase the sensitivity of the MC simulations to subtle changes in the x_u and $k^0_{u,force}$ by using an experimental input. It was found that the simulations, and therefore the Zhurkov-Bell [347] model, was insensitive to the stochasticity in the number of pick-up events in the simulations. Therefore, it would be advantageous to incorporate an alternative model to deconvolute the $k^0_{u,force}$ and x_u directly from the experimental data. One method to do this would be the use of the Dudko-Hummer-Szabo model with the adaptation to multiple barriers by Zhang and Dudko [348, 355]. This would require the acquisition of additional full unfolding trajectories for each protein system.

6.3 Deuterium as a probe for important contacts for protein stability

Whilst the use of force on a protiated protein in H_2O can determine whether a protein is mechanically resilient, in a certain pulling direction, it cannot differentiate between the specific interactions that are important within the mechanical clamp (Section 1.6.3). In this thesis isotopic substitution was used to probe the mechanical properties of two model

proteins: I27 and pL. The aim was to differentiate the key interactions required for a proteins mechanical stability without using point mutations. By using the two model, pL and I27, the origin of the differences in the mechanical properties of the two proteins was possible. This was due to changes in the different interactions, hydrophobic bonds and hydrogen bonds, within the mechanical clamp region. These differences were attributed to changes in strength of hydrophobic interactions and intra- and inter molecular hydrogen bonds in each of the different protein systems. Therefore, the dominant interactions involved in the mechanical clamp region could be determined. Isotopic substitution coupled with SMFS experiments could also be used to investigate the key interactions involved with protein aggregation, ligand–protein binding and mechanical function of biological molecules. The method could also be used to provide more information about how a protein arrives at its unique 3-D structure during protein folding. The changes observed in the mechanical properties of these proteins may suggest differences in the internal dynamics of the protein due to isotopic substitution. This causes the two proteins to behave differently under the application of force and in the presence of chemical denaturants. These results may have implications for experimental techniques that use deuterium and hydrogen in the solvent and within proteins interchangeably such as neutron diffraction, nuclear magnetic resonance, and Raman spectroscopy. Therefore caution should be taken in the interpretation in the conformational changes and dynamics of proteins in these experiments.

Appendix A

List of chemicals and experimental apparatus

The section gives a list of the apparatus and kits and manufacturers used in this project.

ÄKTAprime Plus	GE Healthcare, UK
Asylum MFP-3D Atomic Force Microscope	Oxford Instruments Asylum Research, UK
Avanti J-26 XP Centrifuge	Beckman Coulter, USA
Autoclave	Prestige Medical, UK
BUCHI Vac V-500 vacuum pump	Sigma Aldrich, USA
Cell disruptor	Constant Cell Disruption Systems, UK
Chirascan CD Spectrometer	Applied photophysics, UK
Centrifuge 5810R	Eppendorf, Germany
GenFuge 24D Centrifuge	Progen Scientific, UK
HANNA pH 20 pH Meter	HANNA Instruments Ltd, UK
Grant JB1 Unstirred Waterbath	Grant Instruments, UK
T 18 digital ULTRA-TURRAX homogenisor	IKA, UK
Nickel sepharose	GE Healthcare, UK
Microlab 500 series Hamilton	Hamilton, USA
HisTrap HP 5 ml column	GE Healthcare, UK
HisTrap HP 1 ml column	GE Healthcare, UK
Photon Technology International Fluorimeter	Ford, UK
Snakeskin Pleated Dialysis Tubing, 3,500 MWCO	Thermo Scientific, UK

Superdex 74 gel filtration column	GE Healthcare, UK
Tip sonicator, Sonics vcx-130PB Vibracell	Sonics & Materials Inc, USA
UltroSpec 2100 Pro UV/Visible Spectrophotometer	GE Healthcare, UK
VWR pH Electrode	VWR International, USA
Neslab waterbath RTE-300	Thermo Scientific, UK
Zeba desalting column (5ml and 2ml)	Thermo Scientific, UK
500 ml bottle top filter (0.22 μm)	Coming, USA

A.0.1 List of chemicals

This section lists the chemicals used within this project and recipes for all of the buffers. The buffers/media were all made with Purite 18 MΩ distilled (MiliQ) water unless containing deuterium or stated.

Chemical name	**Company**
A	
Acrylamide 30 % (w/v): 0.8 % (w/v) bisacrylamide	Severn Biotech, UK
Agar	Melford Laboratories, UK
Ammonium Bicarbonate/CH_3NO_3	Sigma Life Sciences, USA
Ammonium Persulphate (APS)	Sigma Life Sciences, USA
Ampicillin	Formedium, UK
B	
Benzamidine	Sigma Aldrich, USA
Bromophenol Blue	Sigma Aldrich, USA
C	
Calcium chloride/$CaCl_2$	Melford Laboratories, UK
Carbenicillin disodium salt	Formedium, UK
Chloroamphenicol	Sigma Life Sciences, USA
D	
Deoxyribonuclease, DNAase	Sigma Aldrich, USA
Deuterium chloride/DCl	Sigma Aldrich, USA
Deuterium Oxide/D_2O 99.9%	Cambridge Isotope Laboratories, USA

1,2-Dithiothreitol, DTT	Formedium, UK
E	
Ethanol	Fisher Scientific, UK
Ethylenediaminetetraacetic acid, EDTA	Acros Organics, Belgium
G	
Glycerol	Fisher Scientific,, UK
Guanidine Hydrochloride (GnHcl)/CH_6CIN_3	Sigma Life Science, USA
H	
Hydrochloric acid/HCl	Fisher scientific, UK
I	
Imidazole	Sigma Aldrich, USA
InstantBlue stain	CBS Scientific, UK
ISOGRO-D growth media 97 atom % D	Sigma Aldrich, USA
Isopropyl β-1-D-thiogalactopyranoside, IPTG	Formedium, UK
L	
Luria-Bertani Broth, LB broth (ready mixed)	Melford Laboratories, UK
Lysosyme	Sigma Aldrich, USA
M	
Magnesium sulphate/$MgSO_4$	Sigma Life Sciences, USA
Mark12 Protein Standard	Promega, UK
N	
Ni Sepharose 6 Fast Flow	GE Healthcare, UK
Nuclease free water	Promega, UK
P	
Phenylmethylsulphonyl fluoride, PMSF	Sigma Aldrich, USA
Phosphoric acid/H_3PO_4	Alfe Aesar, USA
Potassium phosphate dibasic/K_2HPO_4	Acros Organic, USA
Potassium phosphate monobasic/KH_2PO_4	Acros Organic, USA
S	
Sodium Azide	Sigma Aldrich, USA
Sodium Chloride/NaCl	Fisher Scientific, UK
Sodium Deuteroxide/NaOD	Sigma Aldrich, USA

Sodium Dodecyl Sulphate, SDS	Severn biotech Ltd., UK
Sodium Hydroxide/NaOH	Fisher Scientific, UK
Sodium phosphate dibasic/Na_2HPO_4	Sigma Aldrich, USA
Sodium phosphate monobasic/NaH_2PO_4	Sigma Aldrich, USA
T	
Tetramethylethylenediamine, TEMED	Sigma Life Sciences, USA
Tris-(hydroxymethyl)-aminomethane, Tris	Fisher Scientific, UK
Triton-X	VWR (USA)
U	
Urea/$CO(NH_2)_2$	Sigma Life Sciences, USA

Appendix B

Supplementary information for chapter 3

B.1 Mass spectra for the deuterated proteins taken as a function of time

The mass spectra shown in Figures B.1.0.1–B.1.0.4 are of the deuterated proteins taken as a function of time. Little variance is observed in the measured mass and therefore it is assumed that the proteins are fully isotopically exchanged.

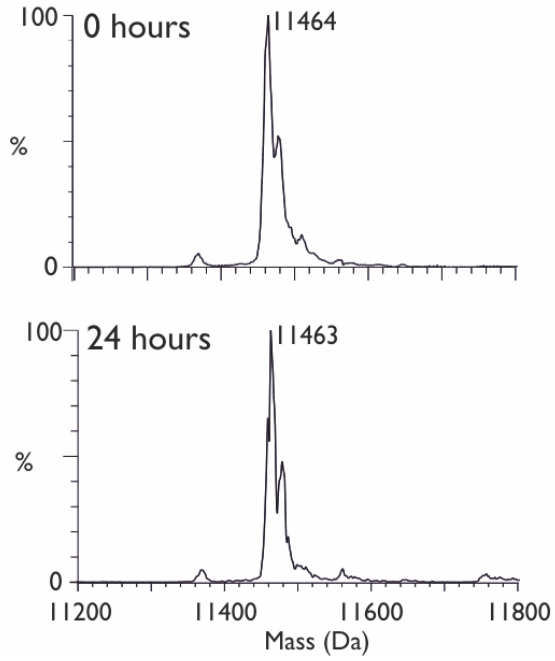


Figure B.1.0.1: Mass spectra of the purified protein of d(I27) taken as a function of time. Single peaks can been seen for the mass of the protein. The mass spectra for this protein were taken at 0 hours and 48 hours to check that deuterons at labile sites had fully exchanged to protons. The mass did not change significantly over the time period monitored.

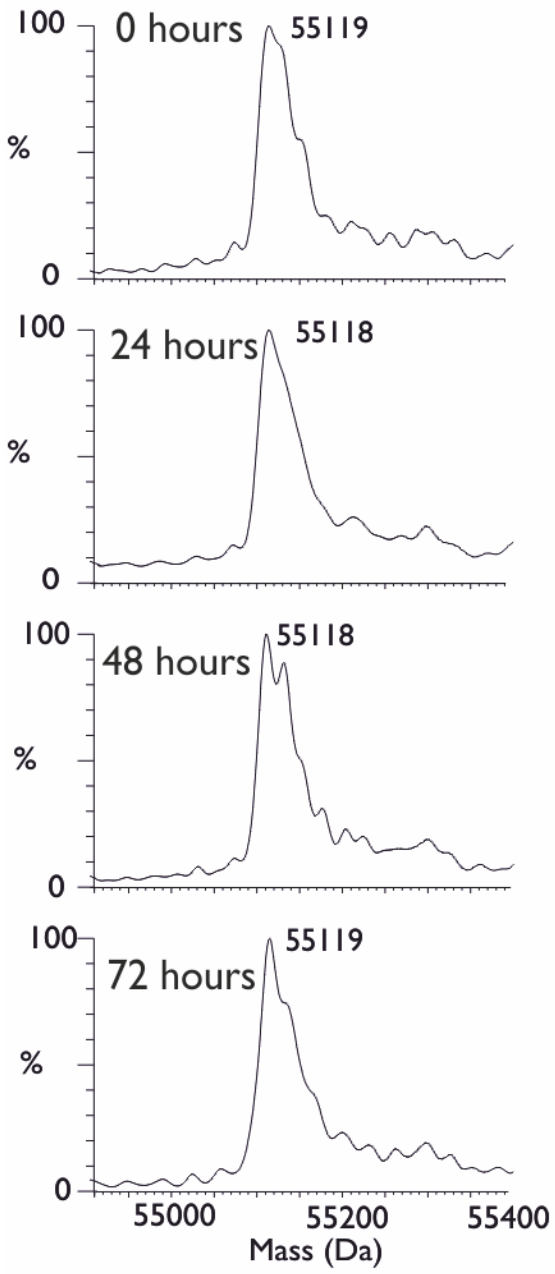


Figure B.1.0.2: Mass spectra of the purified protein of $d(I27)_5$ taken as a function of time. Single peaks can been seen for the mass of the protein. The peaks are quite broad due to the broad range of masses attributed to the size of the polyprotein. The mass spectra for this protein were taken every day for three days to check that deuterons at labile sites had fully exchanged to protons. The mass did not change significantly over the time period monitored.

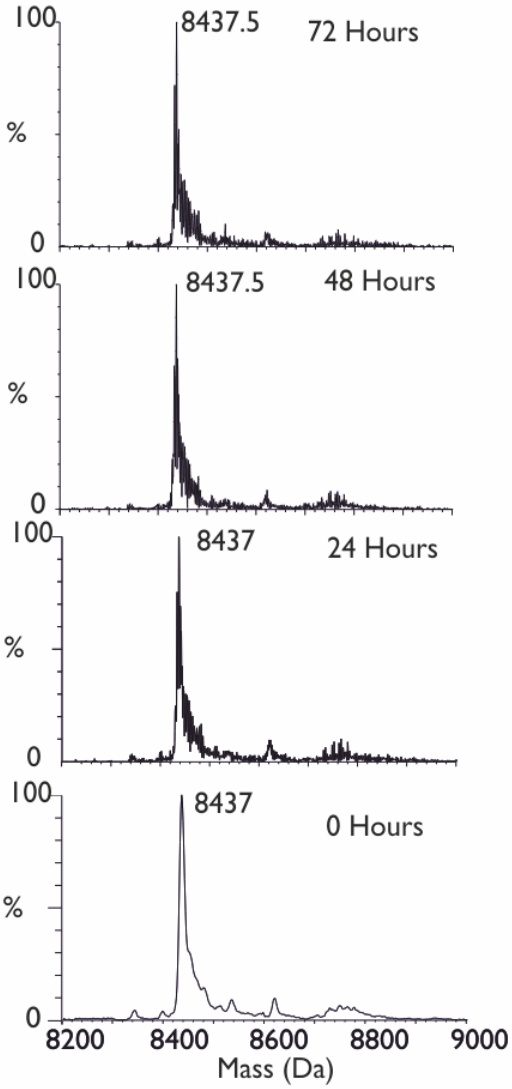


Figure B.1.0.3: Mass spectra of the purified protein of d(pL) taken as a function of time. Single peaks can been seen for the mass of the protein. The mass spectra for this protein were taken every day for three days to check that deuterons at labile sites had fully exchanged to protons. The mass did not change significantly over the time period probed.

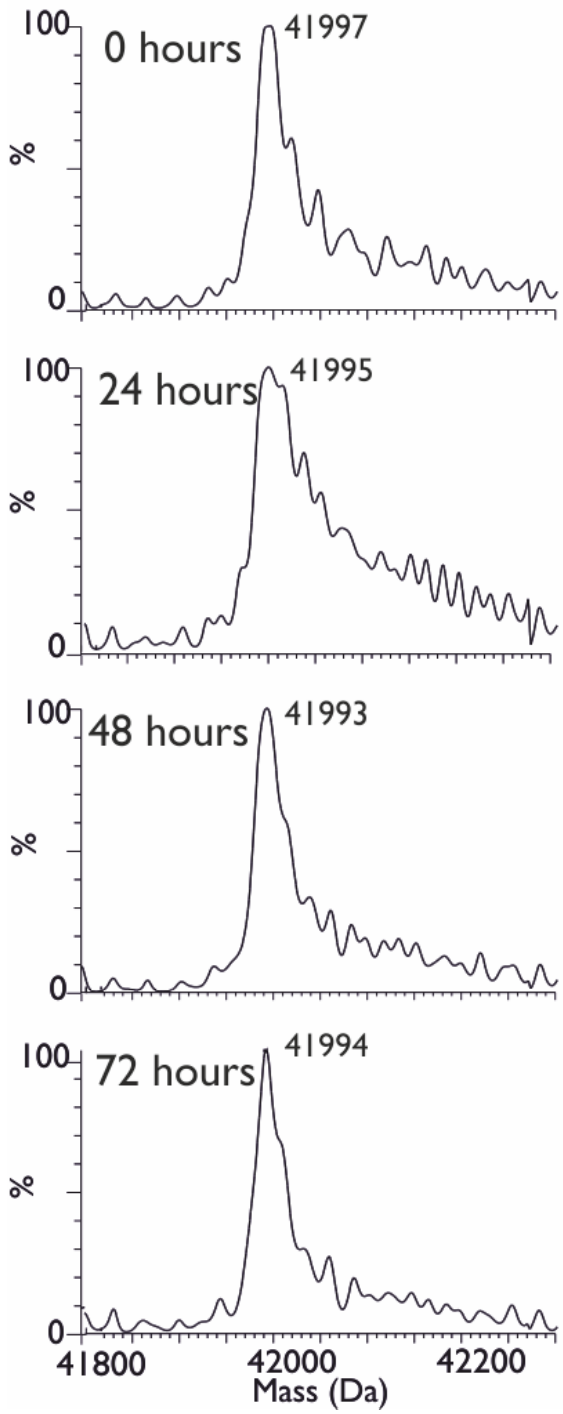


Figure B.1.0.4: Mass spectra of the purified protein of $d(pL)_5$ taken as a function of time. Single peaks can been seen for the mass of the protein. The peaks are quite broad due to the broad range of masses attributed to the size of the polyprotein. The mass spectra for this protein were taken every day for three days to check that deuterons at labile sites had fully exchanged to protons. The mass did not change significantly over the time period monitored.

B.2 The calculation of ΔS

Figures B.2.0.1–B.2.0.2 show the change in $\Delta G_{f,temp}$ with T for the I27 and protein L monomers. The $\Delta G_{f,temp}$ were calculated using Equation 3.5 with the parameters obtained in Section 3.5 for the I27 and pL monomers. The slope of these plots should determine the ΔS of protein folding. The pL plot indicates a non-linear relationship between the $\Delta G_{f,temp}$ and T therefore the ΔS could not be determined.

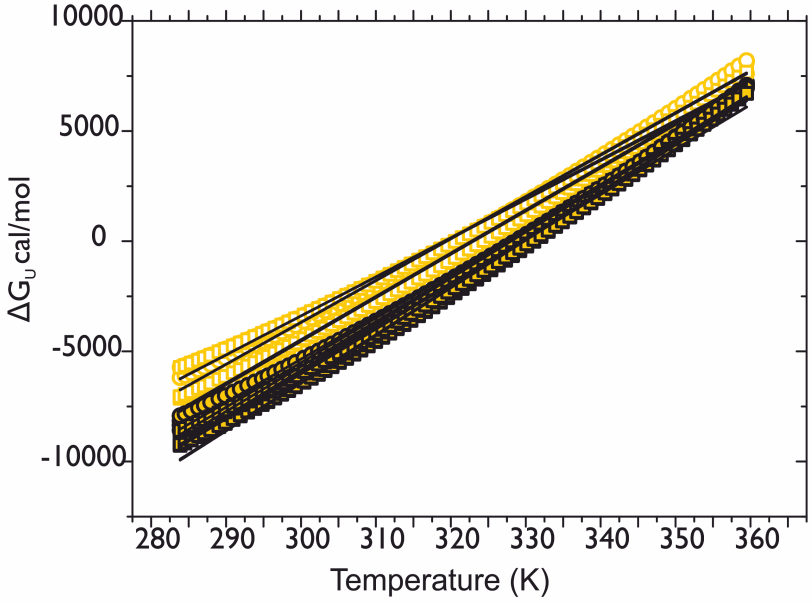


Figure B.2.0.1: $\Delta G_{f,temp}$ against T determined from the thermal melting parameters for the I27 systems. The slope of this plot determined the entropy of folding.

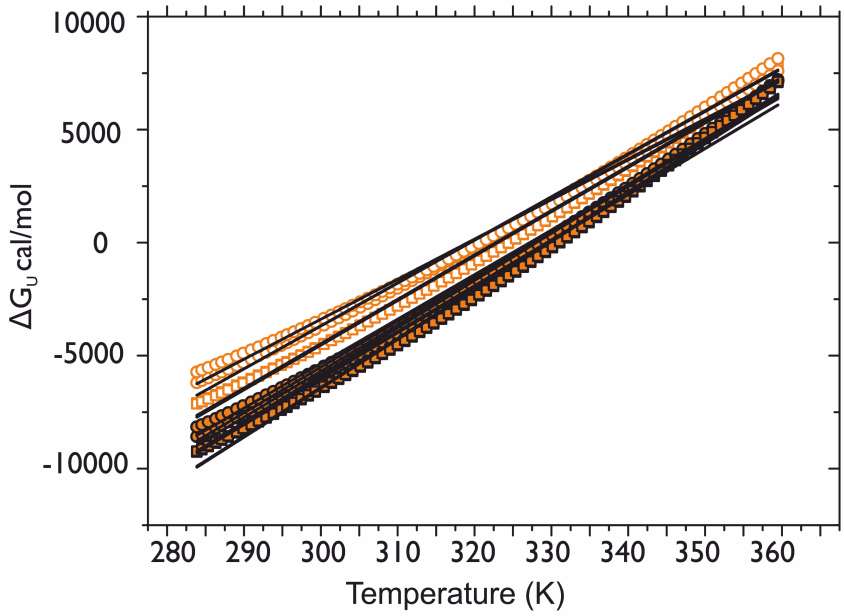


Figure B.2.0.2: $\Delta G_{f,temp}$ against T determined from the thermal melting parameters for the pL systems. The slope of this plot determined the entropy of folding.

B.3 Fluorescence spectra of the folded and unfolding proteins

The folded and unfolded protein fluorescence spectra for the four test systems of I27 and protein L are given in Figures B.2.0.1 and B.2.0.2 respectively.

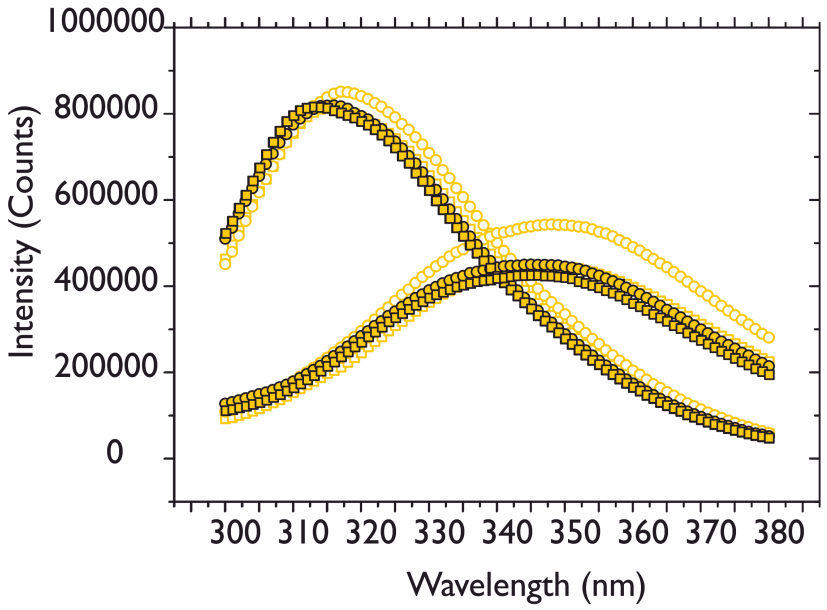


Figure B.3.0.1: Fluorescence spectra for the four I27 protein systems. Open squares are the p(I27) in H_2O, open circles are the d(I27) in H_2O, closed squares are the p(I27) in D_2O and closed circles are the d(I27) in D_2O. The folded spectra of the proteins in 0 M urea have a peak at $\sim$ 310 nm. This peak shifts to $\sim$ 345 nm in 8 M urea.

Comparisons of barrier heights

Negligible differences were observed in the folding and unfolding barrier heights for both deuterated and protiated I27 and pL systems. This can be observed in Figures B.3.0.3–B.3.0.4.

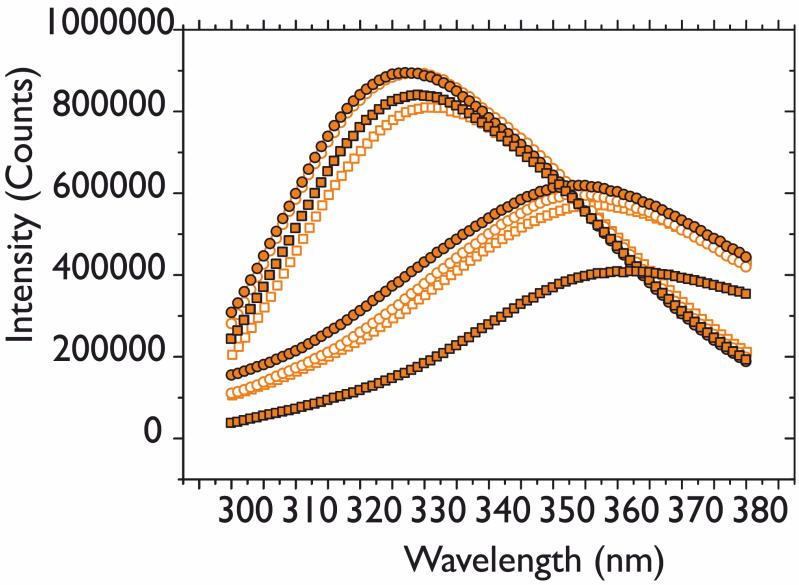


Figure B.3.0.2: Fluorescence spectra for the four pL protein systems. Open squares are the p(pL) in H_2O, open circles are the d(pL) in H_2O, closed squares are the p(pL) in D_2O and closed circles are the d(pL) in D_2O. The folded spectra of the proteins in 0 M urea have a peak at $\sim$ 310 nm. This peak shifts to $\sim$ 345 nm in 7 M GdnHCl. There is a significant decrease in the height of the fluorescence curve for the p(pL) in H_2O systems. This may be indicative to a change in the quantum yield, or reduction in the concentration of protein due to protein instability.

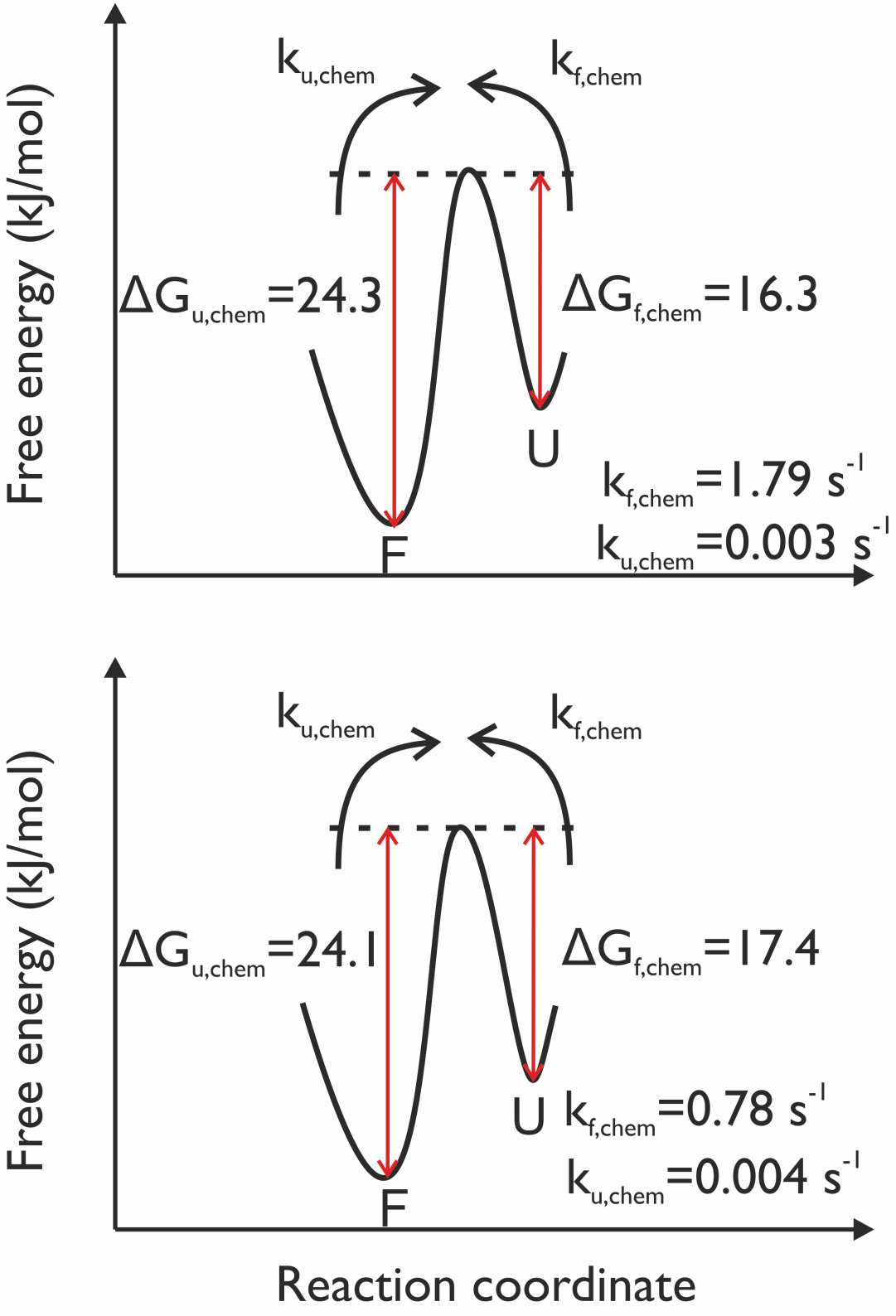


Figure B.3.0.3: Figure showing the differences in in the folding and unfolding barrier heights for the deuterated and protiated I27 proteins, obtained assuming Arrhenius behaviour i.e., $\Delta G^0_{uorf,chem} = -RT \ln\left(\frac{k^0_{uorf,chem}}{A}\right)$, where A is the prefactor (assumed to be 10_6[384]. Despite large differences in the folding rates of the proteins, only small differences were observed in the barrier heights.

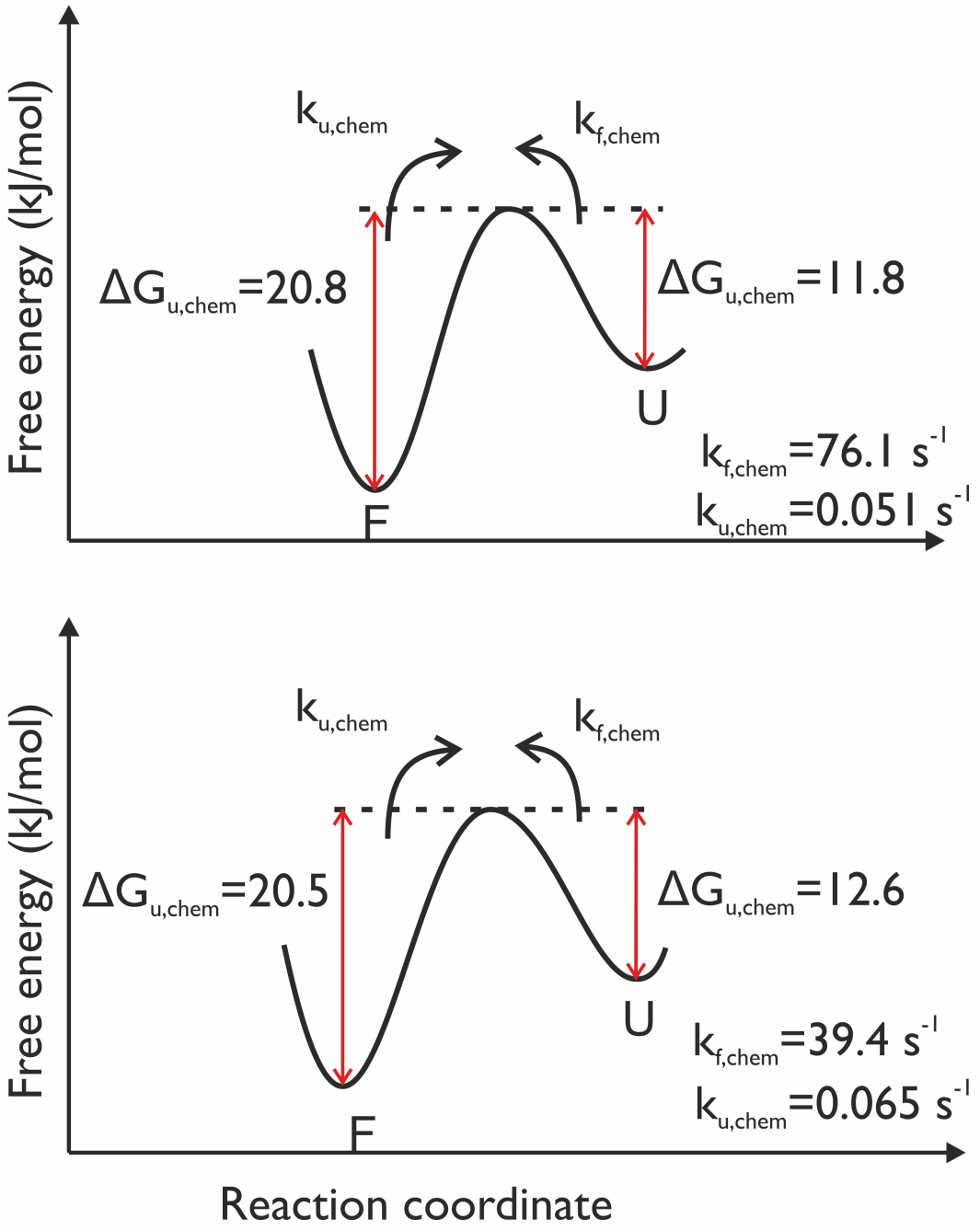


Figure B.3.0.4: Figure showing the differences in in the folding and unfolding barrier heights for the deuterated and protiated protein L systems, obtained assuming Arrhenius behaviour i.e., $\Delta G^0_{uorf,chem} = -RT \ln\left(\frac{k^0_{uorf,chem}}{A}\right)$, where A is the prefactor (assumed to be 10_6[384]. Despite large differences in the folding rates of the proteins, only small differences were observed in the barrier heights.

Appendix C

Supplementary information for chapter 4

The figures presented in this section present a comparison of the experimental obtained in this study and previous studies for: the p(pL) in H_2O [217], Figure C.0.0.1 and p(pL) in H_2O [214], Figure C.0.0.2.

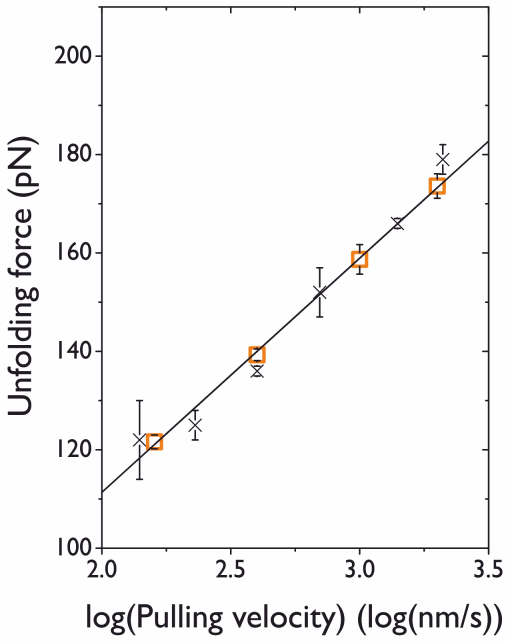


Figure C.0.0.1: Figure comparing the experimental unfolding forces obtained for the p(pL) in H_2O in this study and the forces obtained in the study by Brockwell et al. (2005)[217].

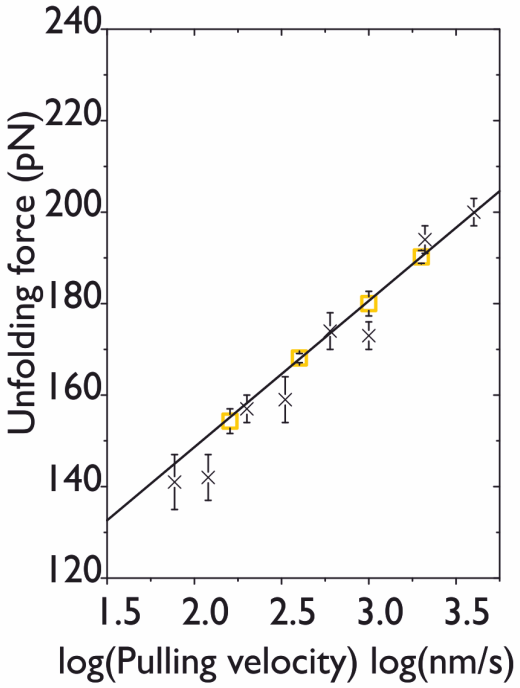


Figure C.0.0.2: Figure comparing the experimental unfolding forces obtained for the p(I27) in H_2O in this study and the forces obtained in the study by Brockwell et al. (2002)[214].

Appendix D

Supplementary information for the MC simulations on the protein test systems

D.1 Equal unfolding simulations on the I27 test systems

Table D.1.0.1 gives the energy landscape parameters obtained from fixing the x_u to 0.29 nm. This corresponds to the best fit parameter for the p(I27) in H_2O system using Method 1 as described in Chapter 5. A comparison of the simulated forces obtained from fixing x_u to 0.29 nm the the experimental unfolding forces is shown in Figure D.1.0.2. Figure D.1.0.1 compares the simulated and experimental forces from fixing the x_u to 0.28 nm for p(I27) in H_2O and d(I27) in H_2O and 0.23 nm form p(I27) in D_2O and d(I27) in D_2O.

D.2 Information about the number of traces containing five unfolding events

The number of traces containing the five unfolding events, corresponding to the full unfolding of the polyprotein, are given in Table D.2.0.1

D.3 Parameters used in the MC simulations with unequal probability distribution

Tables D.3.0.1 and D.3.0.2 list the parameters used for each of the four pL and I27 systems using the unequal unfolding probability distributions respectively. Comparison of the simulated forces with the experimental data and equal unfolding distribution MC sim-

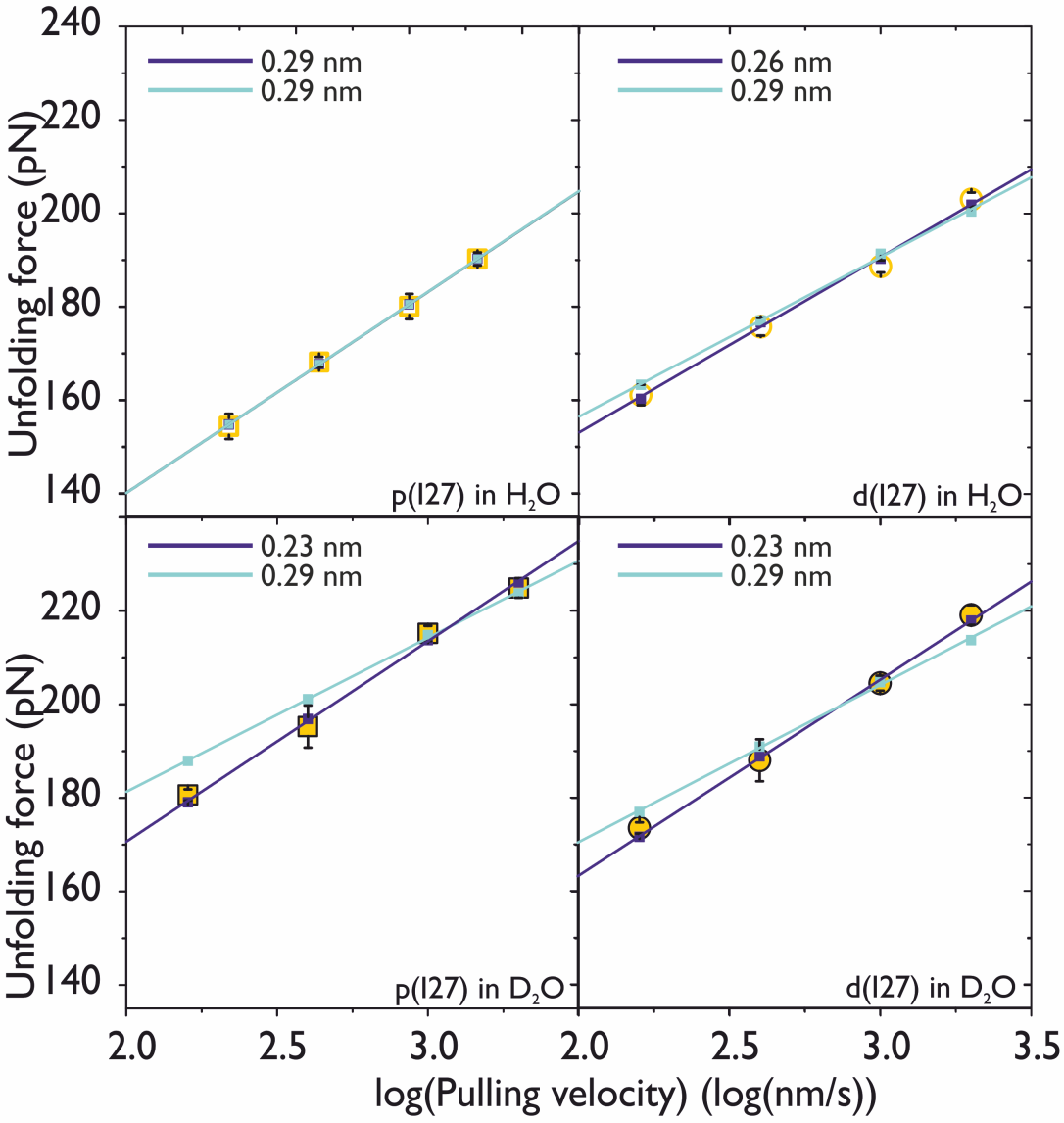


Figure D.1.0.1: Graph of Unfolding force against log(Pulling velocity) for the four I27 test systems (as labelled) comparing the experimental and simulated data determined by two methods. Blue: The straight line is the fit to simulated forces corresponding to the best fit parameters to the experimental data of x_u and $k^0_{u,force}$. This line minimises the least squares of a weighted fit to the experimental data. The associated x_u are labelled. Cyan: Straight line fits to the simulated forces that minimise the error to the experimental data for a fixed x_u of 0.29 nm. The associated best fit $k^0_{u,force}$ are given in Table 5.2.2.2. The graph indicates deviations between the simulated and experimental data for the p(I27) in D_2O and d(I27) in D_2O.

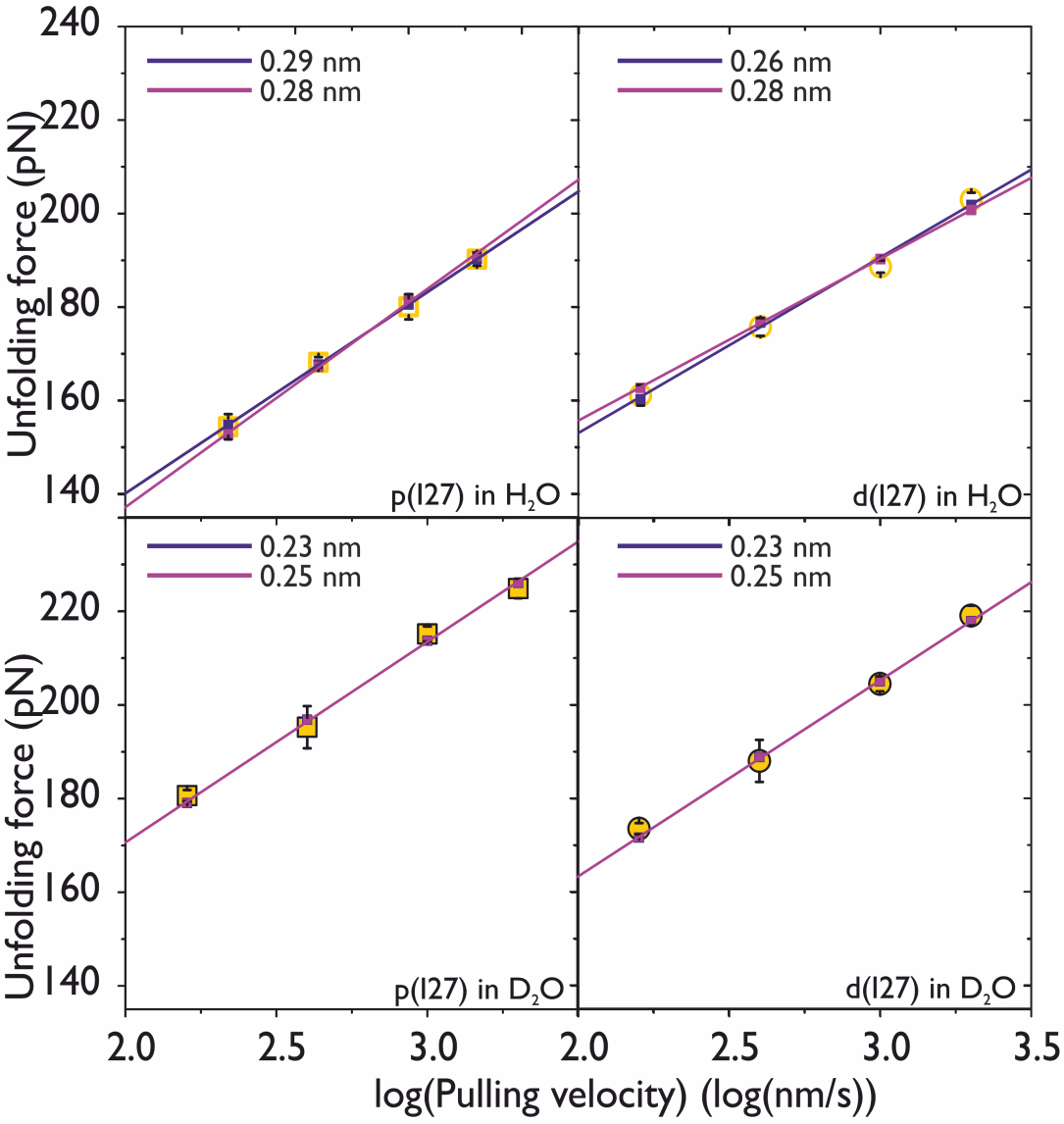


Figure D.1.0.2: Graph of Unfolding force against log(Pulling velocity) for the four I27 test systems (as labelled) comparing the experimental and simulated data determined by two methods. Blue: The straight line is the fit to simulated forces corresponding to the best fit parameters to the experimental data of x_u and $k^0_{u,force}$. This line minimises the least squares of a weighted fit to the experimental data. The associated x_u are labelled. Pink: Straight line fits to the simulated forces that minimise the error to the experimental data for a fixed x_u of 0.28 nm, for p(I27) in H_2O and d(I27) in H_2O, and $x_u = 0.23nm$ for p(I27) in D_2O and d(I27) in D_2O. The associated best fit $k^0_{u,force}$ are given in Table 5.2.2.2. The graph indicates close agreement between the experimental and simulated data for both methods.

Variable	System			
Method 2b	p(I27) in H_2O	p(I27) in D_2O	d(I27) in H_2O	d(I27) in D_2O
B.f. x_u (nm)	0.29	0.29	0.29	0.29
B.f. $k^0_{u,force}$ (s^{-1})	9.5×10^{-4}	1.0×10^{-4}	5.0×10^{-4}	2.0×10^{-4}
Max $k^0_{u,force}$ (s^{-1})	10.0×10^{-4}	2.0×10^{-4}	6.5×10^{-4}	3.5×10^{-4}
Min $k^0_{u,force}$ (s^{-1})	9.5×10^{-4}	0.5×10^{-4}	4.0×10^{-4}	1.5×10^{-4}
Average $k^0_{u,force}$ (s^{-1})	9.8×10^{-4}	1.3×10^{-4}	5.3×10^{-4}	2.5×10^{-4}
Sd $k^0_{u,force}$ (s^{-1})	0.4×10^{-4}	0.7×10^{-4}	1.0×10^{-4}	0.8×10^{-4}
Av. $\Delta G^0_{TS,force}$ (kJ/mol)	51.1	55.9	52.6	54.0
Sd $\Delta G^0_{TS,force}$ (kJ/mol)	0.1	1.5	0.4	0.8
Av. D (N/m)	2.024	2.235	2.085	2.161
Sd D (N/m)	0.004	0.058	0.017	0.032

Table D.1.0.1: Statistics obtained from MC simulations for the four I27 test systems: p(I27) in H_2O, p(I27) in D_2O, d(I27) in H_2O, d(I27) in D_2O. Method 2 was used to obtain the parameters x_u; the distance between the unfolded and transition state in a two-state unfolding energy landscape, $k^0_{u,force}$; the rate of unfolding at zero force and $\Delta G^0_{TS,force}$ the height of the transition barrier to unfolding. Method 2 determines the best fit to the experimental data assuming a fixed x_u. B.f. indicates the best fit parameters. Av. is the average of the accepted x_u and $k^0_{u,force}$, Sd is the standard deviation of the values. The Av. and Sd $\Delta G^0_{TS,force}$ was the average and standard deviation taken from calculating $\Delta G^0_{TS,force}$ for every accepted $k^0_{u,force}$.

	Pulling velocity nm/s			
System	160 nm/s	400 nm/s	1000 nm/s	2000 nm/s
p(pL) in H_2O	12	34	5	35
p(pL) in D_2O	11	9	24	15
d(pL) in H_2O	21	18	16	19
p(I27) in H_2O	3	12	8	17
p(I27) in D_2O	10	11	7	12
d(I27) in H_2O	11	6	3	10
d(I27) in D_2O	6	6	18	6

Table D.2.0.1: Table showing the number of experimental traces containing five protein unfolding events for pL and I27.

ulations, using the forces corresponding to the best fit parameters, are shown in Figures D.3.0.1–D.3.0.7.

System	x_u^i	x_u^f	Δx_u	$k_{u,force}^i$	$k_{u,force}^f$	$\Delta k_{u,force}^0$
p(pL) in H_2O	0.18	0.25	0.01	0.01	0.09	0.00005
p(pL) in D_2O	0.18	0.25	0.01	0.00005	0.04	0.00005
d(pL) in H_2O	0.18	0.26	0.01	0.02	0.15	0.00005
d(pL) in D_2O	0.18	0.25	0.01	0.02	0.07	0.00005

Table D.3.0.1: The grid of parameters used in the MC simulations of the pL test systems. Where i indicates the initial value and f indicates the final value. The unfolded length of the protein was taken to be 22.25 nm , the folded length was 3.7 nm, the average linker lengths were 2.8 nm (values from based on [217]), the temperature was 23 °C and the number of iterations was 1000.

System	x_u^i	x_u^f	Δx_u	$k_{u,force}^i$	$k_{u,force}^f$	$\Delta k_{u,force}^0$
p(I27) in H_2O	0.20	0.34	0.01	0.00005	0.01	0.00005
p(I27) in D_2O	0.18	0.34	0.01	0.00005	0.017	0.00005
d(I27) in H_2O	0.20	0.34	0.01	0.00005	0.012	0.00005
d(I27) in D_2O	0.18	0.34	0.01	0.00005	0.015	0.00005

Table D.3.0.2: The grid of parameters used in the MC simulations of the I27 test systems. Where i indicates the initial value and f indicates the final value. The unfolded length of the protein was taken to be 28 nm [214], the folded length was 4.2 nm (determined from the PDB structure in PyMOL [10]), the average linker lengths were 2.4 nm, the temperature was assumed to be 23 °C and the number of iterations was 1000.

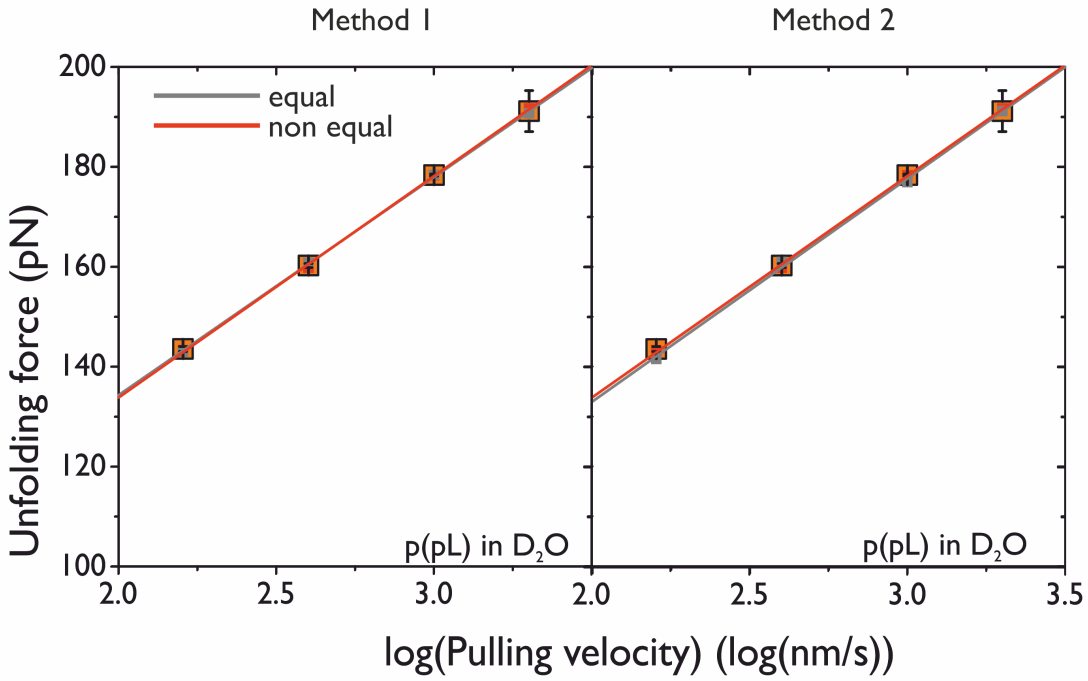


Figure D.3.0.1: Graph of unfolding force against the logarithm of pulling velocity for the p(pL) in D_2O. The graph shows the experimental forces (open squares) and the simulated forces from the simulation with an equal unfolding probability distribution (grey) or unequal unfolding probability distribution (red).

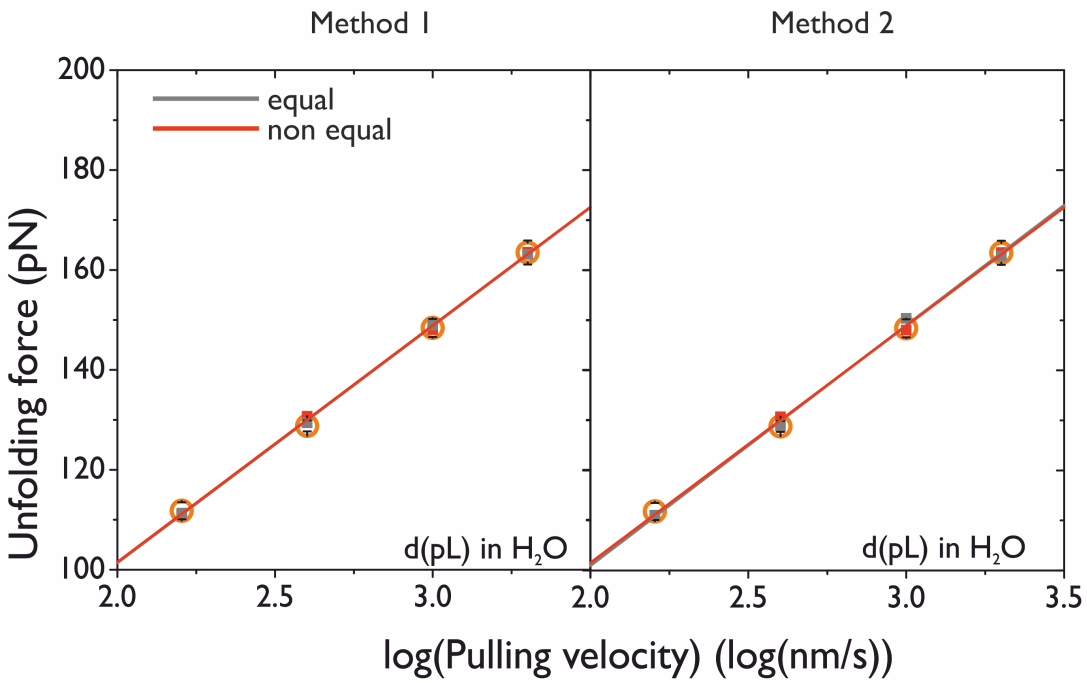


Figure D.3.0.2: Graph of unfolding force against the logarithm of pulling velocity for the d(pL) in H_2O. The graph shows the experimental forces (open squares) and the simulated forces from the simulation with an equal unfolding probability distribution (grey) or unequal unfolding probability distribution (red).

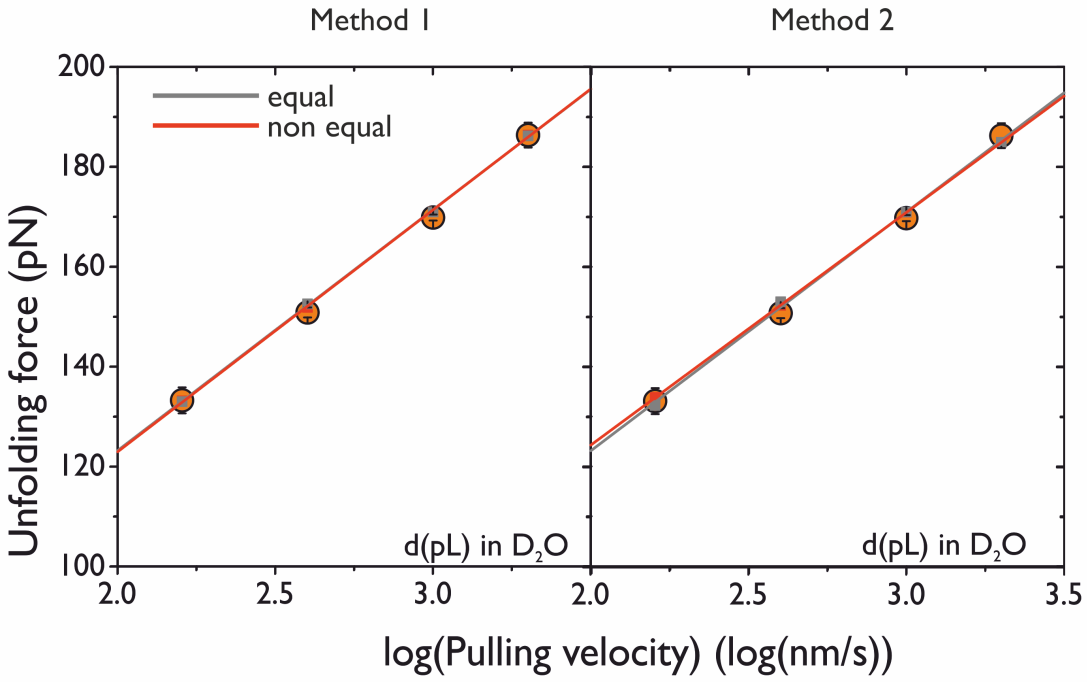


Figure D.3.0.3: Graph of unfolding force against the logarithm of pulling velocity for the d(pL) in D_2O. The graph shows the experimental forces (open squares) and the simulated forces from the simulation with an equal unfolding probability distribution (grey) or unequal unfolding probability distribution (red).

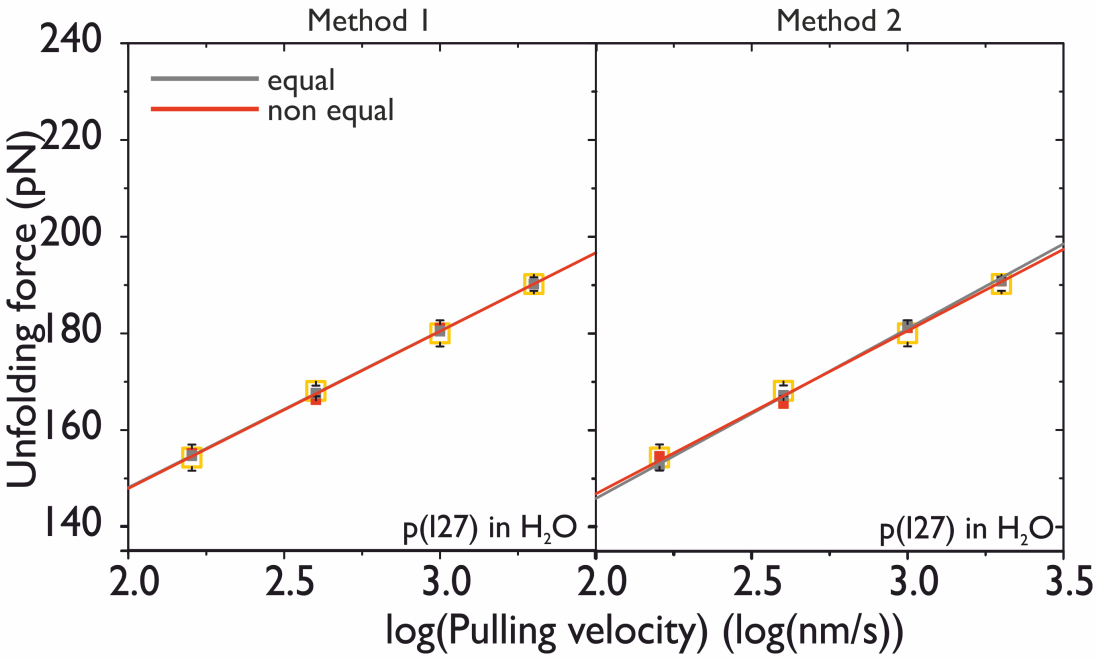


Figure D.3.0.4: Graph of unfolding force against the logarithm of pulling velocity for the p(I27) in H_2O. The graph shows the experimental forces (open squares) and the simulated forces from the simulation with an equal unfolding probability distribution (grey) or unequal unfolding probability distribution (red).

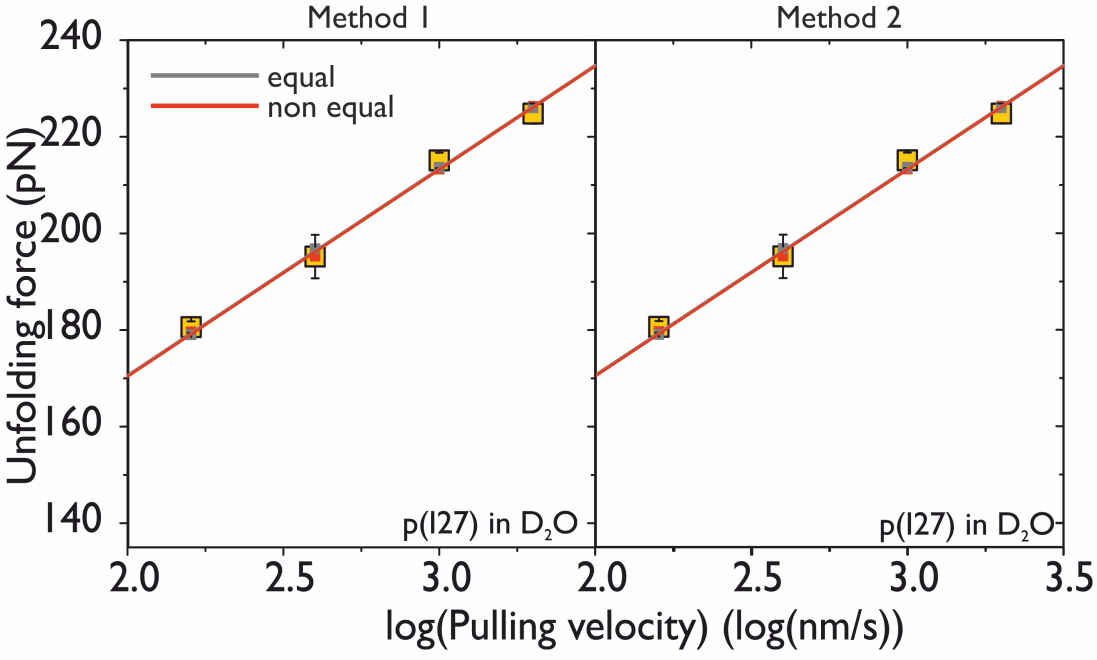


Figure D.3.0.5: Graph of unfolding force against the logarithm of pulling velocity for the p(I27) in D_2O. The graph shows the experimental forces (open squares) and the simulated forces from the simulation with an equal unfolding probability distribution (grey) or unequal unfolding probability distribution (red).

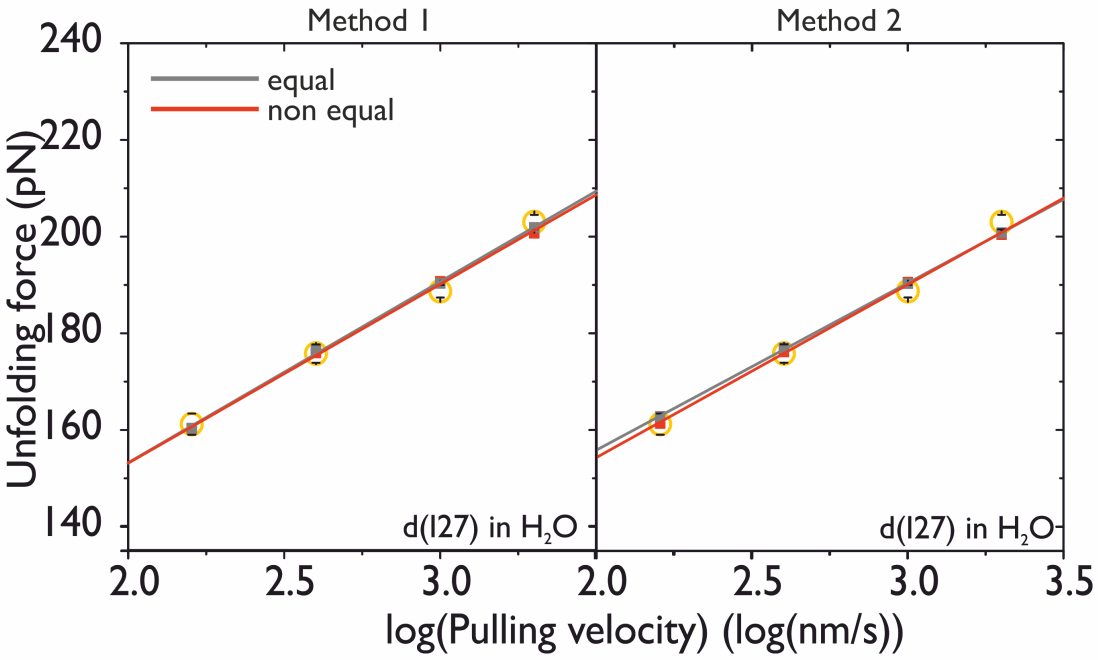


Figure D.3.0.6: Graph of unfolding force against the logarithm of pulling velocity for the d(I27) in H_2O. The graph shows the experimental forces (open squares) and the simulated forces from the simulation with an equal unfolding probability distribution (grey) or unequal unfolding probability distribution (red).

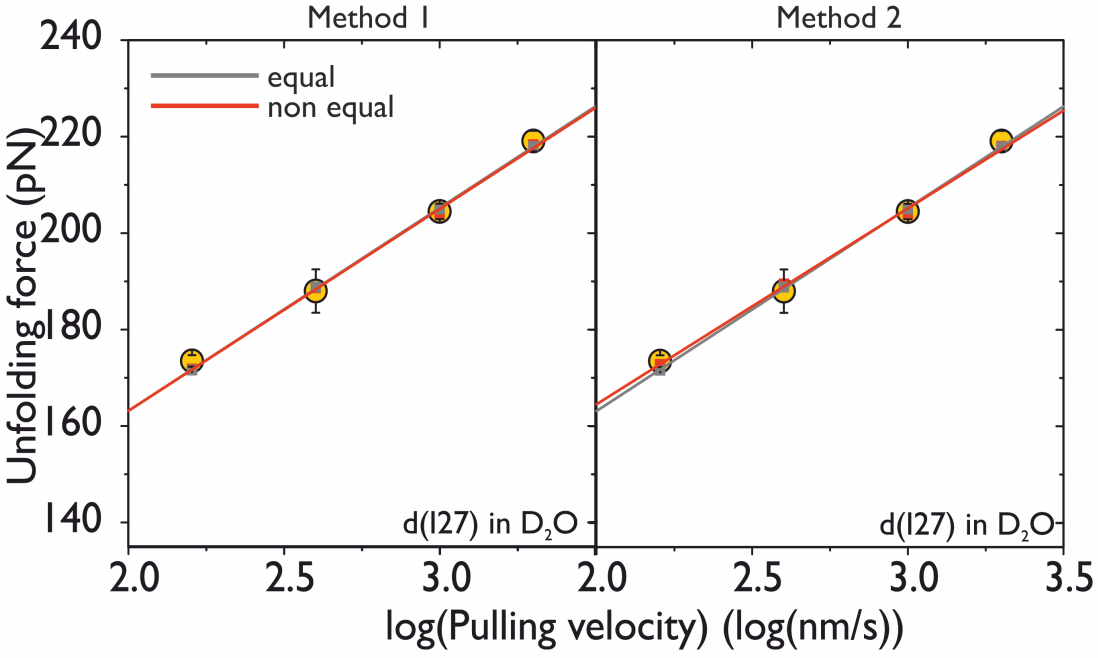


Figure D.3.0.7: Graph of unfolding force against the logarithm of pulling velocity for the d(I27) in D_2O. The graph shows the experimental forces (open squares) and the simulated forces from the simulation with an equal unfolding probability distribution (grey) or unequal unfolding probability distribution (red).

Bibliography

[1] A. J. Cozzone. Proteins: fundamental chemical properties. In *eLS*. John Wiley & Sons, Ltd, 2001. doi:10.1002/9780470015902.a0001330.pub2.

[2] A. V. Finkelstein and O. B. Ptitsyn. *Protein Physics*. Elsevier, Amsterdam, 2002. doi:10.1016/B978-012256781-0/50017-2.

[3] M. Lieberman, A. D. Marks, C. M. Smith, and D. B. Marks. *Marks' Essential Medical Biochemistry*. Lippincott Williams & Wilkins, Philadelphia, 2007.

[4] A. S. Edison. Linus Pauling and the planar peptide bond. *Nat. Struct. Biol.*, 8(3):201–202, 2001. doi:10.1038/84921.

[5] J. M. Berg, J. L. Tymoczko, and L. Stryer. *Biochemistry Fifth Edition*. W H Freeman, Amsterdam, 2002.

[6] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.*, 7:95–99, 1963. doi:10.1016/S0006-3495(65)86759-5.

[7] L. Pauling and R. B. Corey. The pleated sheet, a new layer configuration of polypeptide chains. *Proc. Natl. Acad. Sci. USA*, 37(5):251–256, 1951. doi:10.1073/pnas.37.5.251.

[8] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 112(3):535–542, 1977. doi:10.1111/j.1432-1033.1977.tb11885.x.

[9] J. W. O'Neill, D. E. Kim, D. Baker, and K. Y. Zhang. Structures of the B1 domain of protein L from Peptostreptococcus magnus with a tyrosine to tryptophan substitution. *Acta Crystallogr. D. Biol. Crystallogr.*, 57:480–487, 2001. doi:10.1107/S0907444901000373.

[10] L. L. C. Schrodinger. The PyMOL Molecular Graphics System, Version 1.7.4. 2010.

[11] C. N. Pace. Conformational stability of globular proteins. *Trends Biochem. Sci.*, 15(1):14–17, 1990. doi:10.1016/0968-0004(90)90124-T.

[12] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes. Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.*, 48:545–600, 1997. doi:10.1146/annurev.physchem.48.1.545.

[13] M. Schlierf and M. Rief. Single-molecule unfolding force distributions reveal a funnel-shaped energy landscape. *Biophys. J.*, 90(4):L33–L35, 2006. doi:10.1529/biophysj.105.077982.

[14] C. N. Pace, G. R. Grimsley, J. M. Scholtz, and K. L. Shaw. Protein Stability. *Encycl. life Sci.*, pages 1–5, 2014. doi:10.1002/9780470015902.a0003002.pub3.

[15] J. B. Thompson, H. G. Hansma, P. K. Hansma, and K. W. Plaxco. The Backbone Conformational Entropy of Protein Folding: Experimental Measures from Atomic Force Microscopy. *J. Mol. Biol.*, 322(3):645–652, 2002. doi:10.1016/S0022-2836(02)00801-X.

[16] A. J. Doig and M. J. Sternberg. Side-chain conformational entropy in protein folding. *Protein Sci.*, 4(11):2247–2251, 1995. doi:10.1002/pro.5560041101.

[17] B. W. Matthews, H. Nicholson, and W. J. Becktel. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc. Natl. Acad. Sci. USA*, 84(19):6663–6667, 1987. doi:10.1073/pnas.84.19.6663.

[18] C. N. Pace, B. A. Shirley, M. McNutt, and K. Gajiwala. Forces contributing to the conformational stability of proteins. *FASEB J.*, 10(1):75–83, 1996.

[19] C. N. Pace, H. Fu, K. L. Fryar, J. Landua, S. R. Trevino, D. Schell, R. L. Thurlkill, S. Imura, J. M. Scholtz, K. Gajiwala, J. Sevcik, L. Urbanikova, J. K. Myers, K. Takano, E. J. Hebert, B. A. Shirley, and G. R. Grimsley. Contribution of hydrogen bonds to protein stability. *Protein Sci.*, 23(5):652–661, 2014. doi:10.1002/pro.2449.

[20] K. A. Dill. Dominant forces in protein folding. *Biochemistry*, 29(31):7133–7155, 1990. doi:10.1021/bi00483a001.

[21] F. M. Richards. Protein stability: still an unsolved problem. *Cell. Mol. Life Sci. C.*, 53(10):790–802, 1997. doi:10.1007/s000180050100.

[22] B. K. Shoichet, W. A. Baase, R. Kuroki, and B. W. Matthews. A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci.*, 92(2):452–456, 1995. doi:10.1073/pnas.92.2.452.

[23] J. Mukherjee and M. N. Gupta. Increasing importance of protein flexibility in designing biocatalytic processes. *Biotechnol. Reports*, 6:119–123, 2015. doi:10.1016/j.btre.2015.04.001.

[24] S. Idicula-Thomas and P. V. Balaji. Correlation between the structural stability and aggregation propensity of proteins. *In Silico Biol.*, 7(2):225–237, 2007.

[25] R. Myers. *The Basics of Chemistry*. Basics of the hard sciences. Greenwood Press, London, 2003.

[26] J. M. Berg, J. L. Tymoczko, and L. Stryer. *Biochemistry*. W.H. Freeman, New York, 5 edition, 2002.

[27] M. Nishio, Y. Umezawa, J. Fantini, M. S Weiss, and P. Chakrabarti. CH-π hydrogen bonds in biological macromolecules. *Phys. Chem. Chem. Phys.*, 16(25):12648–12683, 2014. doi:10.1039/c4cp00099d.

[28] P. Needham. Hydrogen bonding: Homing in on a tricky chemical concept. *Stud. Hist. Philos. Sci. Part A*, 44(1):51–65, 2013. doi:10.1016/j.shpsa.2012.04.001.

[29] T. S. Moore and T. F. Winmill. The state of amines an aqueous solution. *J. Chem. Soc.*, 101:1635–1676, 1912. doi:10.1039/CT9120101635.

[30] Linus Pauling. *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry*. Cornell University Press, 1960.

[31] P. Goymer. 100 years of the hydrogen bond. *Nat. Chem.*, 4(11):863–864, 2012. doi:10.1038/nchem.1482.

[32] W. M. Latimer and W. H. Rodebush. Polarity and ionization from the standpoint of the Lewis theory of valence. *J. Am. Chem. Soc.*, 42(7):1419–1433, 1920. doi:10.1021/ja01452a015.

[33] L. Pauling and R. B. Corey. Configuration of polypeptide chains. *Nature*, 168(4274):550–551, 1951. doi:doi:10.1038/168550a0.

[34] 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(4):205–211, 1951. doi:10.1073/pnas.37.4.205.

[35] R. M. Lynden-Bell, S. C. Morris, J. D. Barrow, J. L. Finney, and C. Harper. *Water and Life: The Unique Properties of H_2O.* CRC Press, 2010.

[36] P. L. Silvestrelli and M. Parrinello. Structural, electronic, and bonding properties of liquid water from first principles. *J. Chem. Phys.*, 111(8):3572, 1999. doi:10.1063/1.479638.

[37] J. N. Israelachvili. *Intermolecular and Surface Forces.* Elsevier Science, Waltham, 2 edition, 2010.

[38] G. C. Pimental and A. L. McClellan. *The Hydrogen Bond.* Freeman, San Fransisco, 1960.

[39] K. Modig, B. G. Pfrommer, and B. Halle. Temperature-dependent hydrogen-bond geometry in liquid water. *Phys. Rev. Lett.*, 90(7):075502, 2003. doi:10.1103/PhysRevLett.90.075502.

[40] R. C. Dougherty. Temperature and pressure dependence of hydrogen bond strength: A perturbation molecular orbital approach. *J. Chem. Phys.*, 109(17):7372, 1998. doi:10.1063/1.477343.

[41] J. K. Myers and C. N. Pace. Hydrogen bonding stabilizes globular proteins. *Biophys. J.*, 71(4):2033–2039, 1996. doi:10.1016/S0006-3495(96)79401-8.

[42] N. Eswar and C. Ramakrishnan. Deterministic features of side-chain main-chain hydrogen bonds in globular protein structures. *Protein Eng. Des. Sel.*, 13(4):227–238, 2000. doi:10.1093/protein/13.4.227.

[43] D. F. Stickle, L. G. Presta, K. A. Dill, and G. D. Rose. Hydrogen bonding in globular proteins. *J. Mol. Biol.*, 226(4):1143–1159, 1992. doi:10.1016/0079-6107(84)90007-5.

[44] N. Baker and E. Hubbard. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.*, 44(2):97–179, 1984. doi:10.1016/0079-6107(84)90007-5.

[45] G. J. Lesser and G. D. Rose. Hydrophobicity of amino acid subgroups in proteins. *Proteins*, 8(1):6–13, 1990. doi:10.1002/prot.340080104.

[46] C. N. Pace, J. M. Scholtz, and G. R. Grimsley. Forces stabilizing proteins. *FEBS Lett.*, 588(14):2177–2184, 2014. doi:10.1016/j.febslet.2014.05.006.

[47] B. A. Shirley, P. Stanssens, U. Hahn, and C. N. Pace. Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry*, 31(3):725–732, 1992. doi:10.1021/bi00118a013.

[48] M. P. Byrne, R. L. Manuel, L. G. Lowe, and W. E. Stites. Energetic contribution of side chain hydrogen bonding to the stability of staphylococcal nuclease. *Biochemistry*, 34(42):13949–13960, 1995. doi:10.1021/bi00042a029.

[49] J. A. Schellman. The stability of hydrogen-bonded peptide structures in aqueous solution. *C. R. Trav. Lab. Carlsberg. Chim.*, 29(14-15):230–259, 1955. doi:10.1021/j150570a005.

[50] Y. W. Chen, A. R. Fersht, and K. Henrick. Contribution of buried hydrogen bonds to protein stability. The crystal structures of two barnase mutants. *J. Mol. Biol.*, 234(4):1158–1170, 1993. doi:10.1006/jmbi.1993.1667.

[51] C. N. Pace, G. Horn, E. J. Hebert, J. Bechert, K. Shaw, L. Urbanikova, J. M. Scholtz, and J. Sevcik. Tyrosine hydrogen bonds make a large contribution to protein stability. *J. Mol. Biol.*, 312(2):393–404, 2001. doi:10.1006/jmbi.2001.4956.

[52] S. Deechongkit, H. Nguyen, E. T. Powers, P. E. Dawson, M. Gruebele, and J. W. Kelly. Context-dependent contributions of backbone hydrogen bonding to beta-sheet folding energetics. *Nature*, 430(6995):101–105, 2004. doi:10.1038/nature02611.

[53] B. Honig and A. S. Yang. Free energy balance in protein folding. *Adv. Protein Chem.*, 46:27–58, 1995. doi:10.1016/S0065-3233(08)60331-9.

[54] T. Lazaridis, G. Archontis, and M. Karplus. Enthalpic contribution to protein stability: insights from atom-based calculations and statistical mechanics. *Adv. Protein Chem.*, 47:231–306, 1995. doi:0.1016/S0065-3233(08)60547-1.

[55] M. J. Sippl, M. Ortner, M. Jaritz, P. Lackner, and H. Flöckner. Helmholtz free energies of atom pair interactions in proteins. *Fold. Des.*, 1(4):289–298, 1996. doi:10.1016/S1359-0278(96)00042-9.

[56] N. Ben-Tal, D. Sitkoff, I. A. Topol, A.-S. Yang, S. K. Burt, and B. Honig. Free energy of amide hydrogen bond formation in vacuum, in water, and in liquid alkane Solution. *J. Phys. Chem. B*, 101(3):450–457, 1997. doi:10.1021/jp961825r.

[57] H. Susi and J. S. Ard. Hydrophobic interactions and hydrogen bonding of .epsilon.-caprolactam in aqueous solution. *J. Phys. Chem.*, 73(7):2440–2441, 1969. doi:10.1021/j100727a062.

[58] P. J. Fleming and G. D. Rose. Do all backbone polar groups in proteins form hydrogen bonds? *Protein Sci.*, 14(7):1911–1917, 2005. doi:10.1110/ps.051454805.

[59] J. Gao, D. A. Bosco, E. T. Powers, and J. W. Kelly. Localized thermodynamic coupling between hydrogen bonding and microenvironment polarity substantially stabilizes proteins. *Nat. Struct. Mol. Biol.*, 16(7):684–690, 2009. doi:10.1038/nsmb.1610.

[60] C. N. Pace. Polar group burial contributes more to protein stability than nonpolar group burial. *Biochemistry*, 40(2):310–313, 2001. doi:10.1021/bi001574j.

[61] D. R. Livesay, D. H. Huynh, S. Dallakyan, and D. J. Jacobs. Hydrogen bond networks determine emergent mechanical and thermodynamic properties across a protein family. *Chem. Cent. J.*, 2:17, 2008. doi:10.1186/1752-153X-2-17.

[62] K. Tompa, M. Bokor, T. Verebélyi, and P. Tompa. Water rotation barriers on protein molecular surfaces. *Chem. Phys.*, 448:15–25, 2015. doi:10.1016/j.chemphys.2014.12.008.

[63] W. Doster and M. Settles. *Hydration Processes in Biology: Experimental and Theoretical Approaches*, chapter The dynamical transition in proteins: the role of hydrogen bonds, pages 177–191. IOS Press, Amsterdam, 1999.

[64] M. Tarek and D. J. Tobias. Role of protein-water hydrogen bond dynamics in the protein dynamical transition. *Phys. Rev. Lett.*, 88(13):138101–138104, 2002. doi:10.1103/PhysRevLett.88.138101.

[65] M. Nakasako. Water-protein interactions from high-resolution protein crystallography. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 359(1448):1191–1206, 2004. doi:10.1098/rstb.2004.1498.

[66] M. Chaplin. Do we underestimate the importance of water in cell biology? *Nat. Rev. Mol. Cell Biol.*, 7(11):861–866, 2006. doi:10.1038/nrm2021.

[67] F. Franks. Protein stability: the value of old literature. *Biophys. Chem.*, 96(2-3):117–127, 2002. doi:10.1016/S0301-4622(02)00014-5.

[68] T. Ikura, Y. Urakubo, and N. Ito. Water-mediated interaction at a proteinprotein interface. *Chem. Phys.*, 307(2-3):111–119, 2004. doi:10.1016/j.chemphys.2004.05.010.

[69] I. Brovchenko and A. Oleinikova. Which properties of a spanning network of hydration water enable biological functions? *Chemphyschem*, 9(18):2695–2702, 2008. doi:10.1002/cphc.200800662.

[70] L. Zhang, L. Wang, Y.-T. Kao, W. Qiu, Y. Yang, O. Okobiah, and D. Zhong. Mapping hydration dynamics around a protein surface. *Proc. Natl. Acad. Sci. U. S. A.*, 104(47):18461–18466, 2007. doi:10.1073/pnas.0707647104.

[71] R. K. Eppler, E. P. Hudson, S. D. Chase, J. S. Dordick, J. A. Reimer, and D. S. Clark. Biocatalyst activity in nonaqueous environments correlates with centisecond-range protein motions. *Proc. Natl. Acad. Sci. U. S. A.*, 105(41):15672–15677, 2008. doi:10.1073/pnas.0804566105.

[72] S. Busch, C. D. Bruce, C. Redfield, C. D. Lorenz, and S. E. McLain. Water mediation is essential to nucleation of β-turn formation in peptide folding motifs. *Angew. Chem. Int. Ed. Engl.*, 52(49):13091–13095, 2013. doi:10.1002/anie.201307657.

[73] Y. Levy and J. N. Onuchic. Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.*, 35:389–415, 2006. doi:10.1146/annurev.biophys.35.040405.102134.

[74] M. S. Cheung, A. E. García, and J. N. Onuchic. Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. U. S. A.*, 99(2):685–690, 2002. doi:10.1073/pnas.022387699.

[75] S. K. Pal and A. H. Zewail. Dynamics of water in biological recognition. *Chem. Rev.*, 104(4):2099–2123, 2004. doi:10.1021/cr0206891.

[76] D. Zhong, S. K. Pal, and A. H. Zewail. Biological water: A critique. *Chem. Phys. Lett.*, 503(1-3):1–11, 2011. doi:10.1016/j.cplett.2010.12.077.

[77] T. Yokomizo, M. Nakasako, T. Yamazaki, H. Shindo, and J. Higo. Hydrogen-bond patterns in the hydration structure of a protein. *Chem. Phys. Lett.*, 401(4-6):332–336, 2005. doi:10.1016/j.cplett.2004.11.071.

[78] T. Yokomizo, J. Higo, and M. Nakasako. Patterns and networks of hydrogen-bonds in the hydration structure of human lysozyme. *Chem. Phys. Lett.*, 410(1-3):31–35, 2005. doi:10.1016/j.cplett.2005.04.072.

[79] M. Koizumi, H. Hirai, T. Onai, K. Inoue, and M. Hirai. Collapse of the hydration shell of a protein prior to thermal unfolding. *J. Appl. Crystallogr.*, 40(s1):s175–s178, 2007. doi:10.1107/S0021889807003354.

[80] I. Brovchenko, A. Krukau, N. Smolin, A. Oleinikova, A. Geiger, and R. Winter. Thermal breaking of spanning water networks in the hydration shell of proteins. *J. Chem. Phys.*, 123(22):224905, 2005. doi:10.1063/1.2121708.

[81] N. Smolin, A. Oleinikova, I. Brovchenko, A. Geiger, and R. Winter. Properties of spanning water networks at protein surfaces. *J. Phys. Chem. B*, 109(21):10995–20005, 2005. doi:10.1021/jp050153e.

[82] C. N. Pace, S. Treviño, E. Prabhakaran, and J. M. Scholtz. Protein structure, stability and solubility in water and other solvents. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 359(1448):1225–1235, 2004. doi:10.1098/rstb.2004.1500.

[83] R. L. Baldwin. *Weak Interactions in Protein Folding: Hydrophobic Free Energy, van der Waals Interactions, Peptide Hydrogen Bonds, and Peptide Solvation*, pages 127–162. Wiley-VCH Verlag GmbH, 2008. doi:10.1002/9783527619498.ch6.

[84] R. L. Baldwin. The new view of hydrophobic free energy. *FEBS Lett.*, 587(8):1062–1066, 2013. doi:10.1016/j.febslet.2013.01.006.

[85] J. A. V. Butler. The energy and entropy of hydration of organic compounds. *Trans. Faraday Soc.*, 33:229, 1937. doi:10.1039/tf9373300229.

[86] J. D. Bernal. Structure of Proteins. *Nature*, 143:663–667, 1939. doi:10.1038/143663a0.

[87] W. Kauzmann. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.*, 14:1–63, 1959. doi:10.1016/S0065-3233(08)60608-7.

[88] E. E. Meyer, K. J. Rosenberg, and J. Israelachvili. Recent progress in understanding hydrophobic interactions. *Proc. Natl. Acad. Sci. U. S. A.*, 103(43):15739–15746, 2006. doi:10.1073/pnas.0606422103.

[89] N. T. Southall, K. A. Dill, and A. D. J. Haymet. A view of the hydrophobic effect. *J. Phys. Chem. B*, 106(3):521–533, 2002. doi:10.1021/jp015514e.

[90] W. Blokzijl and J. B. F. N. Engberts. Hydrophobic effects. Opinions and facts. *Angew. Chemie Int. Ed. English*, 32(11):1545–1579, 1993. doi:10.1002/anie.199315451.

[91] B. Widom, P. Bhimalapuram, and K. Koga. The hydrophobic effect. *Phys. Chem. Chem. Phys.*, 5(15):3085, 2003. doi:10.1039/b304038k.

[92] T. Lazaridis. Hydrophobic Effect. In *Encycl. life Sci.* John Wiley & Sons, Ltd, 2013. doi:10.1002/9780470015902.a0002974.pub2.

[93] D. Chandler. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437(7059):640–647, 2005. doi:10.1038/nature04162.

[94] K. A. Sharp, A. Nicholls, R. Friedman, and B. Honig. Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry*, 30(40):9686–9697, 1991. doi:10.1021/bi00104a017.

[95] M. H. Abraham. Free energies, enthalpies, and entropies of solution of gaseous nonpolar nonelectrolytes in water and nonaqueous solvents. The hydrophobic effect. *J. Am. Chem. Soc.*, 104(8):2085–2094, 1982. doi:10.1021/ja00372a001.

[96] B. Lee. Solvent reorganization contribution to the transfer thermodynamics of small nonpolar molecules. *Biopolymers*, 31(8):993–1008, 1991. doi:10.1002/bip.360310809.

[97] H. S. Frank and M. W. Evans. Free volume and entropy in condensed systems III. Entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes. *J. Chem. Phys.*, 13(11):507, 1945. doi:10.1063/1.1723985.

[98] D. T. Bowron, A. Filipponi, M. A. Roberts, and J. L. Finney. Hydrophobic hydration and the formation of a clathrate hydrate. *Phys. Rev. Lett.*, 81(19):4164–4167, 1998. doi:10.1103/PhysRevLett.81.4164.

[99] Y. L. A. Rezus and H. J. Bakker. Observation of immobilized water molecules around hydrophobic groups. *Phys. Rev. Lett.*, 99(14):148301, 2007. doi:10.1103/PhysRevLett.99.148301.

[100] N. Galamba. Water's structure around hydrophobic solutes and the iceberg model. *J. Phys. Chem. B*, 117(7):2153–2159, 2013. doi:10.1021/jp310649n.

[101] C. Petersen, K.-J. Tielrooij, and H. J. Bakker. Strong temperature dependence of water reorientation in hydrophobic hydration shells. *J. Chem. Phys.*, 130(21):214511, 2009. doi:10.1063/1.3142861.

[102] P. Buchanan, N. Aldiwan, A. K. Soper, J. L. Creek, and C. A. Koh. Decreased structure on dissolving methane in water. *Chem. Phys. Lett.*, 415(1-3):89–93, 2005. doi:10.1016/j.cplett.2005.08.064.

[103] I. T. S. Li and G. C. Walker. Signature of hydrophobic hydration in a single polymer. *Proc. Natl. Acad. Sci. U. S. A.*, 108(40):16527–16532, 2011. doi:10.1073/pnas.1105450108.

[104] D. Laage, G. Stirnemann, and J. T. Hynes. Why water reorientation slows without iceberg formation around hydrophobic solutes. *J. Phys. Chem. B*, 113(8):2428–2435, 2009. doi:10.1021/jp809521t.

[105] M. Lucas. Size effect in transfer of nonpolar solutes from gas or solvent to another solvent with a view on hydrophobic behavior. *J. Phys. Chem.*, 80(4):359–362, 1976. doi:10.1021/j100545a004.

[106] F. H. Stillinger. Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory. *J. Solution Chem.*, 2(2-3):141–158, 1973. doi:10.1007/BF00651970.

[107] K. Lum, D. Chandler, and J. D. Weeks. Hydrophobicity at small and large length scales. *J. Phys. Chem. B*, 103(22):4570–4577, 1999. doi:10.1021/jp984327m.

[108] J. G. Davis, K. P. Gierszal, P. Wang, and D. Ben-Amotz. Water structural transformation at molecular hydrophobic interfaces. *Nature*, 491(7425):582–585, 2012. doi:10.1038/nature11570.

[109] R. Zangi. Driving force for hydrophobic interaction at different length scales. *J. Phys. Chem. B*, 115(10):2303–2311, 2011. doi:10.1021/jp1090284.

[110] F. Biedermann, W. M. Nau, and H.-J. Schneider. The hydrophobic effect revisited–studies with supramolecular complexes imply high-energy water as a noncovalent driving force. *Angew. Chem. Int. Ed. Engl.*, 53(42):11158–11171, 2014. doi:10.1002/anie.201310958.

[111] J. Kuriyan, B. Konforti, and D. Wemmer. *The Molecules of Life: Physical and Chemical Principles*. Garland Science, 2012.

[112] M. Munson, S. Balasubramanian, K. G. Fleming, A. D. Nagi, R. O'Brien, J. M. Sturtevant, and L. Regan. What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. *Protein Sci.*, 5(8):1584–1593, 1996. doi:10.1002/pro.5560050813.

[113] G. Saelensminde, Ø . Halskau, and I. Jonassen. Amino acid contacts in proteins adapted to different temperatures: hydrophobic interactions and surface charges play a key role. *Extremophiles*, 13(1):11–20, 2009. doi:10.1007/s00792-008-0192-4.

[114] S. Improta, A. S. Politou, and A. Pastore. Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure*, 4(3):323–337, 1996. doi:10.1016/S0969-2126(96)00036-6.

[115] L. H. Kapcha and P. J. Rossky. A simple atomic-level hydrophobicity scale reveals protein interfacial structure. *J. Mol. Biol.*, 426(2):484–498, 2014. doi:10.1016/j.jmb.2013.09.039.

[116] J. Kyte and R. F. Doolittle. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, 157(1):105–132, 1982. doi:10.1016/0022-2836(82)90515-0.

[117] H. S. Chan. Amino acid side-chain hydrophobicity. *Handb. proteins Struct. Funct. methods*, pages 1–7, 2002. doi:10.1038/npg.els.0003005.

[118] W. C. Wimley and S. H. White. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.*, 3(10):842–848, 1996. doi:10.1038/nsb1096-842.

[119] C. P. Moon and K. G. Fleming. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proc. Natl. Acad. Sci. U. S. A.*, 108(25):10174–10177, 2011. doi:10.1073/pnas.1103979108.

[120] J. L. Fauchere and V. E. Pliska. Hydrophobic parameters II of amino acid side-chains from the partitioning of N-acetyl-amino acid amides. *Eur. J. Med. Chem.*, 18(4):369–375, 1983.

[121] C. N. Pace. Contribution of the hydrophobic effect to globular protein stability. *J. Mol. Biol.*, 226(1):29–35, 1992. doi:10.1016/0022-2836(92)90121-Y.

[122] C. N. Pace, H. Fu, K. L. Fryar, J. Landua, S. R. Trevino, B. A. Shirley, M. M. Hendricks, S. Iimura, K. Gajiwala, J. M. Scholtz, and G. R. Grimsley. Contribution of hydrophobic interactions to protein stability. *J. Mol. Biol.*, 408(3):514–528, 2011. doi:10.1016/j.jmb.2011.02.053.

[123] S. De Vos, J. Backmann, M. Prévost, J. Steyaert, and R. Loris. Hydrophobic core manipulations in ribonuclease T1. *Biochemistry*, 40(34):10140–10149, 2001. doi:10.1021/bi010565n.

[124] D. P. Sadler, E. Petrik, Y. Taniguchi, J. R. Pullen, M. Kawakami, S. E. Radford, and D. J. Brockwell. Identification of a mechanical rheostat in the hydrophobic core of protein L. *J. Mol. Biol.*, 393(1):237–248, 2009. doi:10.1016/j.jmb.2009.08.015.

[125] M. Karpusas, W. A. Baase, M. Matsumura, and B. W. Matthews. Hydrophobic packing in T4 lysozyme probed by cavity-filling mutants. *Proc. Natl. Acad. Sci.*, 86(21):8237–8241, 1989. doi:10.1073/pnas.86.21.8237.

[126] D. E. Anderson, J. H. Hurley, H. Nicholson, W. A. Baase, and B. W. Matthews. Hydrophobic core repacking and aromatic-aromatic interaction in the thermostable mutant of T4 lysozyme ser 117–¿phe. *Protein Sci.*, 2(8):1285–1290, 1993. doi:10.1002/pro.5560020811.

[127] L. P. Gerk, O. Leven, and B. Müller-Hill. Strengthening the dimerisation interface of Lac repressor increases its thermostability by 40 deg. C. *J. Mol. Biol.*, 299(3):805–812, 2000. doi:10.1006/jmbi.2000.3706.

[128] J. N. Israelachvili, R. M. Pashley, E. Perez, and R. K. Tandon. Forces between hydrophobic surfaces in aqueous-electrolyte and surfactant solutions containing common airborne impurities. *Colloids and Surfaces*, 2(3):287–291, 1981. doi:10.1016/0166-6622(81)80016-9.

[129] S. Garde and A. J. Patel. Unraveling the hydrophobic effect, one molecule at a time. *Proc. Natl. Acad. Sci. U. S. A.*, 108(40):16491–16492, 2011. doi:10.1073/pnas.1113256108.

[130] D. Brockwell, L. U. Yu, S. Cooper, S. McCleland, A. Cooper, D. Attwood, S. J. Gaskell, and J. Barber. Physicochemical consequences of the perdeuteriation of glutathione s-transferase from s . japonicum. *Protein Sci.*, 10(3):572–580, 2001. doi:10.1110/ps.46001.exploited.

[131] G Chakrabarti, S Kim, M L Gupta, J S Barton, R H Himes, and M L Gupta Jr. Stabilization of tubulin by deuterium oxide. *Biochemistry*, 38(10):3067–3072, 1999. doi:10.1021/bi982461r.

[132] L. Dougan, A. S. R. Koti, G. Genchev, H. Lu, and J. M. Fernandez. A single-molecule perspective on the role of solvent hydrogen bonds in protein folding and chemical reactions. *ChemPhysChem.*, 9(18):2836–2847, 2008. doi:10.1002/cphc.200800572.

[133] H. C. Urey, F. G. Brickwedde, and G. M. Murphy. A hydrogen isotope of mass 2. *Phys. Rev.*, 39(1):164–165, 1932. doi:10.1103/PhysRev.39.164.

[134] J. G. Stark and H. G. Wallace. *Chemistry Data Book*. Murray, 1988.

[135] Anil Kumar De. *A Text Book of Inorganic Chemistry*. New Age International, 2007.

[136] K. A. Connors. *Chemical Kinetics: The Study of Reaction Rates in Solution*. John Wiley & Sons, 1990.

[137] A. Hattori, H. L. Crespi, and J. J. Katz. Effect of side-chain deuteration on protein stability. *Biochemistry*, 4(7):1213–1225, 1965. doi:10.1021/bi00883a002.

[138] A. Zeidler, P. S. Salmon, H. E. Fischer, J. C. Neuefeind, J. M. Simonson, and T. E. Markland. Isotope effects in water as investigated by neutron diffraction and path integral molecular dynamics. *J. Phys. Condens. Matter*, 24(28):284126, 2012. doi:10.1088/0953-8984/24/28/284126.

[139] L. Benjamin and G. C. Benson. A deuterium isotope effect on excess enthalpy of methanol-water solutions. *J. Phys. Chem.*, 67(4):858–861, 1963. doi:10.1021/j100798a034.

[140] S. Scheiner and M. Cuma. Relative stability of hydrogen and deuterium bonds. *J. Am. Chem. Soc.*, 118(6):1511–1521, 1996. doi:10.1021/ja9530376.

[141] S. Scheiner. *Hydrogen Bonding : A Theoretical Perspective*. Oxford University Press, USA, 1997.

[142] K. B. Wiberg. The deuterium isotope effect. *Chem. Rev.*, 55(4):713–743, 1955. doi:10.1021/cr50004a004.

[143] D. Wade. Deuterium isotope effects on noncovalent interactions between molecules. *Chem. Biol. Interact.*, 117(3):191–217, 1999. doi:10.1016/S0009-2797(98)00097-0.

[144] D. J. Brockwell. *The inhibition of elongation factor Tu by kirromycin*. PhD thesis, University of Manchester, 1995.

[145] F. H. Westheimer. The magnitude of the primary kinetic isotope effect for compounds of hydrogen and deuterium. *Chem. Rev.*, 61(3):265–273, 1961. doi:10.1021/cr60211a004.

[146] G. Nemethy and H. A. Scheraga. Structure of water and hydrophobic bonding in proteins. 4. Thermodynamic properties of liquid deuterium oxide. *J. Chem. Phys.*, 41(3):680, 1964. doi:10.1063/1.1725946.

[147] M. Nakamura, K. Tamura, and S. Murakami. Isotope effects on thermodynamic properties: mixtures of x(D_2O or H_2O) + (1 − x)CH_3CN at 298.15 k. *Thermochim. Acta*, 253:127–136, 1995. doi:10.1016/0040-6031(94)02086-4.

[148] A. K. Soper and C. J. Benmore. Quantum differences between heavy and light water. *Phys. Rev. Lett.*, 101(6):65502, 2008. doi:10.1103/PhysRevLett.101.065502.

[149] A. Zeidler, P. S. Salmon, H. E. Fischer, J. C. Neuefeind, J. M. Simonson, H. Lemmel, H. Rauch, and T. E. Markland. Oxygen as a site specific probe of the structure of water and oxide materials. *Phys. Rev. Lett.*, 107(14):145501, 2011. doi:10.1103/PhysRevLett.107.145501.

[150] G. Graziano. On the solvent isotope effect in hydrophobic hydration. *J. Phys. Chem. B*, 104(39):9249–9254, 2000. doi:10.1021/jp001461+.

[151] G. C. Kresheck, H. Schneider, H. A. Scheraga, and H. Schneide. The effect of D_2O on the thermal stability of proteins. Thermodynamic parameters for the transfer of model compounds from H_2O to D_2O. *J. Phys. Chem.*, 69(9):3132–3144, 1965. doi:10.1021/j100893a054.

[152] G. C. Kresheck. A calorimetric determination of the standard enthalpy and heat capacity changes that accompany micelle formation for four long chain alkyldimethylphosphine oxides in h_2o and d_2o solution from 15 to 79 degc. *J. Am. Chem. Soc.*, 120(42):10964–10969, 1998. doi:10.1021/ja982156t.

[153] Y. Marcus and A. Ben-Naim. A study of the structure of water and its dependence on solutes, based on the isotope effects on solvation thermodynamics in water. *J. Chem. Phys.*, 83(9):4744–4759, 1985. doi:10.1063/1.449000.

[154] M. V. C. Cardoso, L. V. C. Carvalho, and E. Sabadini. Solubility of carbohydrates in heavy water. *Carbohydr. Res.*, 353:57–61, 2012. doi:10.1016/j.carres.2012.03.005.

[155] C. H. Luan and D. W. Urry. Solvent deuteration enhancement of hydrophobicity: Dsc study of the inverse temperature transition of elastin-based polypeptides. *J. Phys. Chem.*, 95(20):7896–7900, 1991. doi:10.1021/j100173a063.

[156] D. S. Berns, J. J. Lee, and E. Scott. Effect of deuterium oxide in protein aggregation in deuterio and protio phycocyanin and other proteins. *Adv. Chem. Ser.*, 84:21–30, 1968. doi:10.1021/ba-1968-0084.ch003.

[157] M. Jeliska-Kazimierczuk and J. Szydowski. Isotope effect on the solubility of amino acids in water. *J. Solution Chem.*, 25(12):1175–1184, 1996. doi:10.1007/BF00972645.

[158] P. Cioni and G. B. Strambini. Effect of heavy water on protein flexibility. *Biophys. J.*, 82(6):3246–3253, 2002. doi:10.1016/S0006-3495(02)75666-X.

[159] H. Omori, M. Kuroda, H. Naora, H. Takeda, Y. Nio, H. Otani, and K. Tamura. Deuterium oxide (heavy water) accelerates actin assembly in vitro and changes microfilament distribution in cultured cells. *Eur. J. Cell Biol.*, 74(3):273–280, 1997.

[160] A. Das, S. Sinha, B. R. Acharya, P. Paul, B. Bhattacharyya, and G. Chakrabarti. Deuterium oxide stabilizes conformation of tubulin: a biophysical and biochemical study. *BMB Rep.*, 41(1):62–67, 2008. doi:10.5483/BMBRep.2008.41.1.062.

[161] Y. M. Efimova, S. Haemers, B. Wierczinski, W. Norde, and A. A. van Well. Stability of globular proteins in H_2O and D_2O. *Biopolymers*, 85(3):264–273, 2007. doi:10.1002/bip.

[162] K. Mizuno and H. P. Bächinger. The effect of deuterium oxide on the stability of the collagen model peptides H-(Pro-Pro-Gly)$_{10}$-OH, H-(Gly-Pro-4(R)Hyp)$_9$-OH, and Type I collagen. *Biopolymers*, 93(1):93–101, 2010. doi:10.1002/bip.21305.

[163] G. I. Makhatadze, G. M. Clore, and A. M. Gronenborn. Solvent isotope effect and protein stability. *Nat. Struct. Biol.*, 2(10):852–855, 1995. doi:10.1038/nsb1095-852.

[164] C. H. Chen, F. Tow, and D. S. Berns. Solvent isotope effect on the differences in structure and stability between normal and deuterated proteins. *Biopolymers*, 23(5):887–896, 1984. doi:10.1002/bip.360230506.

[165] P. A. Baghurst, W. H. Sawyer, and L. W. Nichol. The effect of D_2O on the association of beta-lactoglobulin A. *J. Biol. Chem.*, 247(10):3199–3204, 1972.

[166] F. Bonneté, D. Madern, and G. Zaccai. Stability against denaturation mechanisms in halophilic malate dehydrogenase ''adapt" to solvent conditions. *J. Mol. Biol.*, 244(4):436–447, 1994. doi:10.1006/jmbi.1994.1741.

[167] M. F. Emerson and A. Holtzer. Hydrophobic bond in micellar systems. Effects of various additives on the stability of micelles of sodium dodecyl sulfate and of n-dodecyltrimethylammonium bromide. *J. Phys. Chem.*, 2720(1962):3320–3330, 1967. doi:10.1021/j100869a031.

[168] D. I. Svergun, S. Richard, M. H. Koch, Z. Sayers, S. Kuprin, and G. Zaccai. Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci. U. S. A.*, 95(5):2267–2272, 1998. doi:10.1073/pnas.95.5.2267.

[169] P. Sasisanker, A. Oleinikova, H. Weingrtner, R. Ravindra, and R. Winter. Solvation properties and stability of ribonuclease A in normal and deuterated water studied by dielectric relaxation and differential scanning/pressure perturbation calorimetry. *Phys. Chem. Chem. Phys.*, 6(8):1899, 2004. doi:10.1039/b314070a.

[170] M. J. Parker and A. R. Clarke. Amide backbone and water-related H/D isotope effects on the dynamics of a protein folding reaction. *Biochemistry*, 36(19):5786–5794, 1997. doi:10.1021/bi9629283.

[171] R. H. Maybury and J. J. Katz. Protein denaturation in heavy water. *Nature*, 177(4509):629–630, 1956. doi:10.1038/177629a0.

[172] S. Y. Sheu, E. W. Schlag, H. L. Selzle, and D. Y. Yang. Molecular dynamics of hydrogen bonds in protein-D_2O: The solvent isotope effect. *J. Phys. Chem. A*, 112(5):797–802, 2008. doi:10.1021/jp0771668.

[173] D. Kern, G. Zaccai, and R. Giege. Effect of heavy water substitution for water on the $tRNA^{Val}$-Valyl-tRNA synthetase system from yeast. *Biochemistry*, 19(14):3158–3164, 1980. doi:10.1021/bi00555a007.

[174] B. K. Schowen and R. L. Schowen. Solvent isotope effects on enzyme systems. *Methods Enzymol.*, 87:551–606, 1982. doi:10.1016/S0076-6879(82)87031-6.

[175] L. S. Itzhaki and P. A. Evans. Solvent isotope effects on the refolding kinetics of hen egg-white lysozyme. *Protein Sci.*, 5(1):140–146, 1996. doi:10.1002/pro.5560050117.

[176] B. A. Krantz, L. B. Moran, A. Kentsis, and T. R. Sosnick. D/H amide kinetic isotope effects reveal when hydrogen bonds form during protein folding. *Nat. Struct. Biol.*, 7(1):62–71, 2000. doi:10.1038/71265.

[177] Z. Shi, B. A. Krantz, N. Kallenbach, and T. R. Sosnick. Contribution of hydrogen bonding to protein stability estimated from isotope effects. *Biochemistry*, 41(7):2120–2129, 2002. doi:10.1021/bi011307t.

[178] M. Tehei, D. Madern, C. Pfister, and G. Zaccai. Fast dynamics of halophilic malate dehydrogenase and BSA measured by neutron scattering under various solvent conditions influencing protein stability. *Proc. Natl. Acad. Sci. U. S. A.*, 98(25):14356–14361, 2001. doi:10.1073/pnas.251537298.

[179] X. Chen, K. L. Wilde, H. Wang, V. Lake, P. J. Holden, A. P.J. Middelberg, L. He, and A. P. Duff. High yield expression and efficient purification of deuterated human protein galectin-2. *Food Bioprod. Process.*, 90(3):563–572, 2012. doi:10.1016/j.fbp.2011.12.004.

[180] X. Liu, B. L. Hanson, P. Langan, and R. E. Viola. The effect of deuteration on protein structure: a high-resolution comparison of hydrogenous and perdeuterated haloalkane dehalogenase. *Acta Crystallogr. D. Biol. Crystallogr.*, 63(9):1000–1008, 2007. doi:10.1107/S0907444907037705.

[181] H. L. Crespi and J. J. Katz. Thermal denaturation of ordinary and fully deuteriated DNA in H_2O and D_2O. *J. Mol. Biol.*, 4(2):65–68, 1962. doi:10.1016/S0022-2836(62)80037-0.

[182] T. R. Gamble, K. R. Clauser, and A. A. Kossiakoff. The production and X-ray structure determination of perdeuterated Staphylococcal nuclease. *Biophys. Chem.*, 53(1-2):15–25, 1994. doi:10.1016/0301-4622(94)00072-7.

[183] S. J. Cooper, J. Raftery, J. R. Helliwell, D. J. Brockwell, D. Attwood, and J. Barber. The X-ray crystal structures of perdeuteriated and protiated enzyme elongation factor Tu are very similar. *J. Chem. Soc. Chem. Commun.*, pages 1063–1064, 1998. doi:10.1039/a801152d.

[184] D. S. Berns. Studies of completely deuteriated proteins. I. The immunochemistry of the deuteriated protein and its hydrogen analog. *J. Am. Chem. Soc.*, 85(11):1676–1678, 1963. doi:10.1021/ja00894a031.

[185] S. Rokop, L. Gajda, S. Parmerter, H.L. Crespi, and J.J. Katz. Purification and characterization of fully deuterated enzymes. *Biochim. Biophys. Acta - Enzymol.*, 191(3):707–715, 1969. doi:10.1016/0005-2744(69)90365-9.

[186] G. Piszczek, J. C. Lee, N. Tjandra, C.-R. Lee, Y.-J. Seok, R. L. Levine, and A. Peterkofsky. Deuteration of escherichia coli enzyme I(Ntr) alters its stability. *Arch. Biochem. Biophys.*, 507(2):332–342, 2011. doi:10.1016/j.abb.2010.12.022.

[187] F. S. Bates, H. D. Keith, and D. B. McWhan. Isotope effect on the melting temperature of nonpolar polymers. *Macromolecules*, 20(12):3065–3070, 1987. doi:10.1021/ma00178a021.

[188] S. J. Haward, P. R. Shewry, J. Marsh, M. J. Miles, and T. J. Mc Master. Force spectroscopy of an elastic peptide: Effect of D_2O and temperature on persistence length. *Microsc. Res. Tech.*, 74(2):170–176, 2011. doi:10.1002/jemt.20888.

[189] T. Hoffmann and L. Dougan. Single molecule force spectroscopy using polyproteins. *Chem. Soc. Rev.*, 41(14):4781–4796, 2012. doi:10.1039/c2cs35033e.

[190] A. Matouschek. Protein unfolding an important process in vivo? *Curr. Opin. Struct. Biol.*, 13(1):98–109, 2003. doi:10.1016/s0959-440x(03)00010-1.

[191] M. Shtilerman, G. H. Lorimer, and S. W. Englander. Chaperonin function: Folding by forced unfolding. *Science*, 284(5415):822–825, 1999. doi:10.1126/science.284.5415.822.

[192] S. Labeit and B. Kolmerer. Titins: giant proteins in charge of muscle ultrastructure and elasticity. *Science*, 270(5234):293–296, 1995. doi:10.1126/science.270.5234.293.

[193] T. Hoffmann, K. M. Tych, D. J. Brockwell, and L. Dougan. Single-molecule force spectroscopy identifies a small cold shock protein as being mechanically robust. *J. Phys. Chem. B*, 117(6):1819–1826, 2013. doi:10.1021/jp310442s.

[194] K. Wagner, D. U. Mick, and P. Rehling. Protein transport machineries for precursor translocation across the inner mitochondrial membrane. *Biochim. Biophys. Acta*, 1793(1):52–59, 2009. doi:10.1016/j.bbamcr.2008.05.026.

[195] E. M. Puchner and H. E. Gaub. Force and function: probing proteins with AFM-based force spectroscopy. *Curr. Opin. Struct. Biol.*, 19(5):605–614, 2009. doi:10.1016/j.sbi.2009.09.005.

[196] E. M. Puchner and H. E. Gaub. Single-molecule mechanoenzymatics. *Annual Review of Biophysics*, 41(1):497–518, 2012. doi:10.1146/annurev-biophys-050511-102301.

[197] X. Zhang, K. Halvorsen, C.-Z. Zhang, W. P. Wong, and T. A. Springer. Mechanoenzymatic cleavage of the ultralarge vascular protein von willebrand factor. *Science*, 324(5932):1330–1334, 2009. doi:10.1126/science.1170905.

[198] M. Gautel. The sarcomeric cytoskeleton: who picks up the strain? *Curr. Opin. Cell Biol.*, 23(1):39–46, 2011. doi:10.1016/j.ceb.2010.12.001.

[199] S. Improta, J. K. Krueger, M. Gautel, R. A. Atkinson, J-F. Lefvre, S. Moulton, J. Trewhella, and A. Pastore. The assembly of immunoglobulin-like modules in titin: implications for muscle elasticity. *J. Mol. Biol.*, 284(3):761–777, 1998. doi:10.1006/jmbi.1998.2028.

[200] K. Austen, C. Kluger, A. Freikamp, A. Chrostek-Grashoff, and C. Grashoff. Generation and analysis of biosensors to measure mechanical forces within cells. *Methods Mol. Biol.*, 1066:169–184, 2013. doi:10.1007/978-1-62703-604-7_15.

[201] A. Rio, R. Perez-jimenez, R. Liu, P. Roca-Cusachs, J. M. Fernandez, and M. P. Sheetz. Stretching single talin rod molecules activates vinculin binding. *Science*, 323(5914):638–641, 2009. doi:10.1126/science.1162912.

[202] H. Clausen-Schaumann, M. Seitz, R. Krautbauer, and H. E. Gaub. Force spectroscopy with single bio-molecules. *Curr. Opin. Cell Biol.*, 4(5):524–530, 2000. doi:10.1016/S1367-5931(00)00126-5.

[203] T. A. Baker and R. T. Sauer. ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim. Biophys. Acta*, 1823(1):15–28, 2012. doi:10.1016/j.bbamcr.2011.06.007.

[204] R. A. A. Maillard, G. Chistol, M. Sen, M. Righini, J. Tan, C. M. Kaiser, C. Hodges, A. Martin, and C. Bustamante. ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell*, 145(3):459–469, 2011. doi:10.1016/j.cell.2011.04.010.

[205] T. Stöver and M. Diensthuber. Molecular biology of hearing. *GMS Curr. Top. Otorhinolaryngol. Head Neck Surg.*, 10:Doc06, 2011. doi:10.3205/cto000079.

[206] L. D. Muiznieks and F. W. Keeley. Molecular assembly and mechanical properties of the extracellular matrix: A fibrous protein perspective. *Biochim. Biophys. Acta - Mol. Basis Dis.*, 1832(7):866–875, 2013. doi:10.1016/j.bbadis.2012.11.022.

[207] M. Saito, T. Watanabe-Nakayama, S. Machida, T. Osada, R. Afrin, and A. Ikai. Spectrin-ankyrin interaction mechanics: A key force balance factor in the red blood cell membrane skeleton. *Biophys. Chem.*, 200-201:1–8, 2015. doi:10.1016/j.bpc.2015.03.007.

[208] C. Bustamante, Z. Bryant, and S. B. Smith. Ten years of tension: single-molecule DNA mechanics. *Nature*, 421(6921):423–427, 2003. doi:10.1038/nature01405.

[209] A. Ashkin, J. M. Dziedzic, and T. Yamane. Optical trapping and manipulation of single cells using infrared laser beams. *Nature*, 330(6150):769–771, 1987. doi:10.1038/330769a0.

[210] S. B. Smith, L. Finzi, and C. Bustamante. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*, 258(5085):1122–1126, 1992. doi:10.1126/science.1439819.

[211] J. Zlatanova, S .M. Lindsay, and S. H. Leuba. Single molecule force spectroscopy in biology using the atomic force microscope. *Prog. Biophys. Mol. Biol.*, 74(1-2):37–61, 2000. doi:10.1016/S0079-6107(00)00014-6.

[212] D. J. Brockwell. Force denaturation of proteins-an unfolding story. *Curr. Nanosci.*, 3(1):3–15, 2007. doi:10.2174/157341307779940634.

[213] M. Schlierf and M. Rief. Temperature softening of a protein in single-molecule experiments. *J. Mol. Biol.*, 354(2):497–503, 2005. doi:10.1016/j.jmb.2005.09.070.

[214] D. J. Brockwell, G. S. Beddard, J. Clarkson, R. C. Zinober, A. W. Blake, J. Trinick, P. D. Olmsted, D. A. Smith, and S. E. Radford. The effect of core destabilization on the mechanical resistance of I27. *Biophys. J.*, 83(1):458–472, 2002. doi:10.1016/S0006-3495(02)75182-5.

[215] C.-L. Chyan, F.-C. Lin, H. Peng, J.-M. Yuan, C.-H. Chang, S.-H. Lin, and G. Yang. Reversible mechanical unfolding of single ubiquitin molecules. *Biophys. J.*, 87(6):3995–4006, 2004. doi:10.1529/biophysj.104.042754.

[216] D. J. Brockwell and S. E. Radford. Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol.*, 17(1):30–37, 2007. doi:10.1016/j.sbi.2007.01.003.

[217] D. J. Brockwell, G. S. Beddard, E. Paci, D. K. West, P. D. Olmsted, D. A. Smith, and S. E. Radford. Mechanically unfolding the small, topologically simple protein L. *Biophys. J.*, 89(1):506–519, 2005. doi:10.1529/biophysj.105.061465.

[218] H. Lu and K. Schulten. Steered molecular dynamics simulation of conformational changes of immunoglobulin domain I27 interprete atomic force microscopy observations. *Chem. Phys.*, 247(1):141–153, 1999. doi:10.1016/S0301-0104(99)00164-0.

[219] A. M. Gabovich and M. S. Li. Mechanical stability of proteins. *J. Chem. Phys.*, 131(2):024121, 2009. doi:10.1063/1.3170940.

[220] S. P. Ng, R. W. S. Rounsevell, A. Steward, C. D. Geierhaas, P. M. Williams, E. Paci, and J. Clarke. Mechanical unfolding of TNfn3: the unfolding pathway of a fnIII domain probed by protein engineering, AFM and MD simulation. *J. Mol. Biol.*, 350(4):776–789, 2005. doi:10.1016/j.jmb.2005.04.070.

[221] M. Carrion-Vazquez, A. F. Oberhauser, T. E. Fisher, P. E. Marszalek, H. B. Li, and J. M. Fernandez. Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Prog. Biophys. Mol. Biol.*, 74(1-2):63–91, 2000. doi:10.1016/S0079-6107(00)00017-1.

[222] D. K. Klimov and D. Thirumalai. Native topology determines force-induced unfolding pathways in globular proteins. *Proc. Natl. Acad. Sci.*, 97(13):7254–7259, 2000. doi:10.1073/pnas.97.13.7254.

[223] D. J. Brockwell, E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, and S. E. Radford. Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nat. Struct. Biol.*, 10(9):731–737, 2003. doi:10.1038/nsb968.

[224] M. Carrion-Vazquez, H. B. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez. The mechanical stability of ubiquitin is linkage dependent. *Nat. Struct. Biol.*, 10(9):738–743, 2003. doi:10.1038/nsb965.

[225] H. Dietz and M. Rief. Elastic bond network model for protein unfolding mechanics. *Phys. Rev. Lett.*, 100(9):098101, 2008. doi:10.1103/PhysRevLett.100.098101.

[226] E. A. Shank, C. Cecconi, J. W. Dill, S. Marqusee, and C. Bustamante. The folding cooperativity of a protein is controlled by its chain topology. *Nature*, 465(7298):637–640, 2010. doi:10.1038/nature09021.

[227] B. Jagannathan, P. J. Elms, C. Bustamante, and S. Marqusee. Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein. *Proc. Natl. Acad. Sci. U. S. A.*, 109(44):17820–17825, 2012. doi:10.1073/pnas.1201800109.

[228] M. Bertz, A. Kunfermann, and M. Rief. Navigating the folding energy landscape of green fluorescent protein. *Angew. Chemie Int. Ed.*, 47(43):8192–8195, 2008. doi:10.1002/anie.200802987.

[229] M. Bertz and M. Rief. Mechanical unfoldons as building blocks of maltose-binding protein. *J. Mol. Biol.*, 378(2):447–458, 2008. doi:10.1016/j.jmb.2008.02.025.

[230] H. Dietz and M. Rief. Protein structure by mechanical triangulation. *Proc. Natl. Acad. Sci. U. S. A.*, 103(5):1244–1247, 2006. doi:10.1073/pnas.0509217103.

[231] H. Dietz, F. Berkemeier, M. Bertz, and M. Rief. Anisotropic deformation response of single protein molecules. *Proc. Natl. Acad. Sci. U. S. A.*, 103(34):12724–12728, 2006. doi:10.1073/pnas.0602995103.

[232] P. C. Li and D. E. Makarov. Simulation of the mechanical unfolding of ubiquitin: Probing different unfolding reaction coordinates by changing the pulling geometry. *J. Chem. Phys.*, 121(10):4826–4832, 2004. doi:10.1063/1.1778152.

[233] W. Zheng and P. Glenn. Probing the folded state and mechanical unfolding pathways of T4 lysozyme using all-atom and coarse-grained molecular simulation. *J. Chem. Phys.*, 142(3):035101, 2015. doi:10.1063/1.4905606.

[234] Y. D. Li, G. Lamour, J. Gsponer, P. Zheng, and H. Li. The molecular mechanism underlying mechanical anisotropy of the protein GB1. *Biophys. J.*, 103(11):2361–2368, 2012. doi:10.1016/j.bpj.2012.10.035.

[235] G. Radou, M. Enciso, S. Krivov, and E. Paci. Modulation of a protein free-energy landscape by circular permutation. *J. Phys. Chem. B*, 117(44):13743–13747, 2013. doi:10.1021/jp406818t.

[236] Z. Zhuang, A. I. Jewett, P. Soto, and J.-E. Shea. The effect of surface tethering on the folding of the src-SH3 protein domain. *Phys. Biol.*, 6(1):015004, 2009. doi:10.1088/1478-3975/6/1/015004.

[237] A. J. Wilcox, J. Choy, C. Bustamante, and A. Matouschek. Effect of protein structure on mitochondrial import. *Proc. Natl. Acad. Sci. U. S. A.*, 102(43):15435–15440, 2005. doi:10.1073/pnas.0507324102.

[238] Q. Peng and H. Li. Atomic force microscopy reveals parallel mechanical unfolding pathways of T4 lysozyme: evidence for a kinetic partitioning mechanism. *Proc. Natl. Acad. Sci. U. S. A.*, 105(6):1885–1890, 2008. doi:10.1073/pnas.0706775105.

[239] K. Brejc, T. K. Sixma, P. A. Kitts, S. R. Kain, R. Y. Tsien, M. Ormö, and S. J. Remington. Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.*, 94(6):2306–2311, 1997.

[240] M. Sikora, M. Cieplak, and J. I. Sułkowska. Bio-molecule stretching database, 2010. Accessed: 2015-09-14. URL: http://jowisz.ifpan.edu.pl/BSDB/.

[241] M. Sikora, J. I. Sułkowska, B. S. Witkowski, and M. Cieplak. BSDB: the biomolecule stretching database. *Nucleic Acids Res.*, 39(suppl 1):D443–D450, 2011. doi:10.1093/nar/gkq851.

[242] M. Sikora, J. I. Sułkowska, and M. Cieplak. Mechanical strength of 17,134 model proteins and cysteine slipknots. *PLoS Comput. Biol.*, 5(10):e1000547, 2009. doi:10.1371/journal.pcbi.1000547.

[243] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.*, 75(2):662–671, 1998. doi:10.1016/S0006-3495(98)77556-3.

[244] H. B. Li, M. Carrion-Vazquez, A. F. Oberhauser, P. E. Marszalek, and J. M. Fernandez. Point mutations alter the mechanical stability of immunoglobulin modules. *Nat. Struct. Biol.*, 7(12):1117–1120, 2000. doi:10.1038/81964.

[245] T. Bu, H.-C. E. Wang, and H. Li. Single molecule force spectroscopy reveals critical roles of hydrophobic core packing in determining the mechanical stability of protein GB1. *Langmuir*, 28(33):12319–12325, 2012. doi:10.1021/la301940g.

[246] M. Sikora and M. Cieplak. Cystine plug and other novel mechanisms of large mechanical stability in dimeric proteins. *Phys. Rev. Lett.*, 109(20):208101, 2012. doi:10.1103/PhysRevLett.109.208101.

[247] T. Hoffmann, K. M. Tych, M. L. Hughes, D. J. Brockwell, and L. Dougan. Towards design principles for determining the mechanical stability of proteins. *Phys. Chem. Chem. Phys.*, 15(38):15767–15780, 2013. doi:10.1039/C3CP52142G.

[248] W. Kremer, B. Schuler, S. Harrieder, M. Geyer, W. Gronwald, C. Welker, R. Jaenicke, and H. R. Kalbitzer. Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium Thermotoga maritima. *Eur. J. Biochem.*, 268(9):2527–2539, 2001. doi:10.1046/j.1432-1327.2001.02127.x.

[249] L. J. Shimon, E. A. Bayer, E. Morag, R. Lamed, S. Yaron, Y. Shoham, and F. Frolow. A cohesin domain from Clostridium thermocellum: the crystal structure provides new insights into cellulosome assembly. *Structure*, 5(3):381–390, 1997. doi:10.1016/S0969-2126(97)00195-0.

[250] M. Bycroft, S. Ludvigsen, A. R. Fersht, and F. M. Poulsen. Determination of the three-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy. *Biochemistry*, 30(35):8697–8701, 1991. doi:10.1021/bi00099a030.

[251] J. Pascual, M. Pfuhl, D. Walther, M. Saraste, and M. Nilges. Solution structure of the spectrin repeat: a left-handed antiparallel triple-helical coiled-coil. *J. Mol. Biol.*, 273(3):740–751, 1997. doi:10.1006/jmbi.1997.1344.

[252] M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez. Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl. Acad. Sci. U. S. A.*, 96(7):3694–3699, 1999. doi:10.1073/pnas.96.7.3694.

[253] A. F. Oberhauser, C. Badilla-Fernandez, M. Carrion-Vazquez, and J. M. Fernandez. The mechanical hierarchies of fibronectin observed with single-molecule AFM. *J. Mol.Biol.*, 319(2):433–447, 2002. doi:0.1016/S0022-2836(02)00306-6.

[254] K. L. Fuson, L. Ma, R. B. Sutton, and A. F. Oberhauser. The c2 domains of human synaptotagmin 1 have distinct mechanical properties. *Biophys. J.*, 96(3):1083–1090, 2009. doi:10.1016/j.bpj.2008.10.025.

[255] R. B. Best, B. Li, A. Steward, V. Daggett, and J. Clarke. Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation. *Biophys. J.*, 81(4):2344–2356, 2001. doi:10.1016/s0006-3495(01)75881-x.

[256] Y. Cao and H. B. Li. Polyprotein of GB1 is an ideal artificial elastomeric protein. *Nat. Mater.*, 6(2):109–114, 2007. doi:Doi10.1038/Nmat1825.

[257] D. Sharma, G. Feng, D. Khor, G. Z. Genchev, H. Lu, and H. Li. Stabilization provided by neighboring strands is critical for the mechanical stability of proteins. *Biophys. J.*, 95(8):3935–3942, 2008. doi:10.1529/biophysj.108.134072.

[258] G. Arad-Haase, S. G. Chuartzman, S. Dagan, R. Nevo, M. Kouza, B. K. Mai, H. T. Nguyen, M. S. Li, and Z. Reich. Mechanical unfolding of acylphosphatase studied by single-molecule force spectroscopy and MD simulations. *Biophys. J.*, 99(1):238–247, 2010. doi:10.1016/j.bpj.2010.04.004.

[259] M. Rief, J. Pascual, M. Saraste, and H. E. Gaub. Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *J. Mol. Biol.*, 286(2):553–561, 1999. doi:10.1006/jmbi.1998.2466.

[260] J. R. Forman, Z. T. Yew, S. Qamar, R. N. Sandford, E. Paci, and J. Clarke. Non-native interactions are critical for mechanical strength in PKD domains. *Structure*, 17(12):1582–1590, 2009. doi:10.1016/j.str.2009.09.013.

[261] S. B. Fowler, R. B. Best, J. L. Toca Herrera, T. J. Rutherford, A. Steward, E. Paci, M. Karplus, J. Clarke, and J. L. T. Herrera. Mechanical unfolding of a titin Ig domain: Structure of unfolding intermediate revealed by combining AFM, molecular dynamics simulations, NMR and protein engineering. *J. Mol. Biol.*, 322(4):841–849, 2002. doi:10.1016/S0022-2836(02)00805-7.

[262] L. W. Li, H. H.-L. L. Huang, C. L. Badilla, and J. M. Fernandez. Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module. *J. Mol. Biol.*, 345(4):817–826, 2005. doi:10.1016/j.jmb.2004.11.021.

[263] R. B. Best, S. B. Fowler, J. L. T. Herrera, A. Steward, E. Paci, and J. Clarke. Mechanical unfolding of a titin Ig domain: Structure of transition state revealed by combining atomic force microscopy, protein engineering and molecular dynamics simulations. *J. Mol. Biol.*, 330(4):867–877, 2003. doi:10.1016/S0022-2836(03)00618-1.

[264] R. B. Best, S. B. Fowler, J. L. Toca-Herrera, and J. Clarke. A simple method for probing the mechanical unfolding pathway of proteins in detail. *Proc. Natl. Acad. Sci. U. S. A.*, 99(19):12143–12148, 2002. doi:10.1073/pnas.192351899.

[265] P. W. Fenimore, H. Frauenfelder, B. H. McMahon, and F. G. Parak. Slaving: Solvent fluctuations dominate protein dynamics and functions. *Proc. Natl. Acad. Sci. U. S. A.*, 99(25):16047–16051, 2002. doi:10.1073/pnas.212637899.

[266] G. Caliskan, D. Mechtani, J. H. Roh, A. Kisliuk, A. P. Sokolov, S. Azzam, M. T. Cicerone, S. Lin-Gibson, and I. Peral. Protein and solvent dynamics: how strongly are they coupled? *J. Chem. Phys.*, 121(4):1978–1983, 2004. doi:10.1063/1.1764491.

[267] K. Tych and L. Dougan. *A Single-Molecule Approach to Explore the Role of the Solvent Environment in Protein Folding*, pages 315–334. John Wiley & Sons, Inc., 2013. doi:10.1002/9781118523063.ch15.

[268] S. Garcia-Manyes, L. Dougan, and J. M. Fernandez. Osmolyte-induced separation of the mechanical folding phases of ubiquitin. *Proc. Natl. Acad. Sci. U. S. A.*, 106(26):10540–10545, 2009. doi:10.1073/pnas.0902090106.

[269] Y. Cao and H. Li. How do chemical denaturants affect the mechanical folding and unfolding of proteins? *J. Mol. Biol.*, 375(1):316–324, 2008. doi:10.1016/j.jmb.2007.10.024.

[270] D. Aioanei, M. Brucale, I. Tessari, L. Bubacco, and B. Samorì. Worm-like Ising model for protein mechanical unfolding under the effect of osmolytes. *Biophys. J.*, 102(2):342–350, 2012. doi:10.1016/j.bpj.2011.12.007.

[271] L. Ma, M. Xu, and A. F. Oberhauser. Naturally occurring osmolytes modulate the nano-mechanical properties of polycystic Kidney disease (PKD) domain. *J. Biol. Chem.*, 285(49):38438–38443, 2010. doi:10.1074/jbc.M110.183913.

[272] P. Zheng and H. Li. Direct measurements of the mechanical stability of zinc-thiolate bonds in rubredoxin by single-molecule atomic force microscopy. *Biophys. J.*, 101(6):1467–1473, 2011. doi:10.1016/j.bpj.2011.08.021.

[273] P. Zheng, Y. Cao, T. J. Bu, S. K. K. Straus, and H. Li. Single molecule force spectroscopy reveals that electrostatic interactions affect the mechanical stability of proteins. *Biophys. J.*, 100(6):1534–1541, 2011. doi:10.1016/j.bpj.2011.01.062.

[274] B. Cheng, S. Wu, S. Liu, P. Rodriguez-Aliaga, J. Yu, and S. Cui. Protein denaturation at a single-molecule level: the effect of nonpolar environments and its implications on the unfolding mechanism by proteases. *Nanoscale*, 7(7):2970–2977, 2015. doi:10.1039/c4nr07140a.

[275] G. Stirnemann, S. Kang, R. Zhou, and B. J. Berne. How force unfolding differs from chemical denaturation. *Proc. Natl. Acad. Sci. U. S. A.*, 111(9):3413–3418, 2014. doi:10.1073/pnas.1400752111.

[276] L. Sun, J. K. Noel, J. I. Sulkowska, H. Levine, and J. N. Onuchic. Connecting thermal and mechanical protein (un)folding landscapes. *Biophys. J.*, 107(12):2950–2961, 2014. doi:10.1016/j.bpj.2014.10.021.

[277] A. K. Covington, M. Paabo, R. A. Robinson, and R. G. Bates. Use of the glass electrode in deuterium oxide and the relation between the standardized pD (paD) scale and the operational pH in heavy water. *Anal. Chem.*, 40(4):700–706, 1968. doi:10.1021/ac60260a013.

[278] A. Krezel and W. Bal. A formula for correlating pKa values determined in D_2O and H_2O. *J. Inorg. Biochem.*, 98(1):161–166, 2004. doi:10.1016/j.jinorgbio.2003.10.001.

[279] F. H. Stephenson. *Calculations for Molecular Biology and Biotechnology: A Guide to Mathematics in the Laboratory 2e.* Academic Press, 2010.

[280] E. B. Ruth. Replication, transcription, and translation. *Am. Biol. Teach.*, 46(8):470–472, 1984. doi:10.2307/4447915.

[281] K. Terpe. Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.*, 72(2):211–222, 2006. doi:10.1007/s00253-006-0465-8.

[282] G. L. Rosano and E. A. Ceccarelli. Recombinant protein expression in Escherichia coli: Advances and challenges. *Front. Microbiol.*, 5(172):1–17, 2014. doi:10.3389/fmicb.2014.00172.

[283] Novagen. pET System Manual Novagen. *Biosystems*, pages 1–68, 2003. arXiv:1007.2303v3.

[284] F. W. Studier and B. A. Moffatt. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.*, 189(1):113–130, 1986. doi:10.1016/0022-2836(86)90385-2.

[285] F. W. Studier, A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, 185(1986):60–89, 1990. doi:10.1016/0076-6879(90)85008-C.

[286] F. W. Studier. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.*, 219(1):37–44, 1991. doi:10.1016/0022-2836(91)90855-Z.

[287] J. W. Dubendorff and F. W. Studier. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.*, 219(1):45–59, 1991. doi:10.1016/0022-2836(91)90856-2.

[288] J. Huang, J. Villemain, R. Padilla, and R. Sousa. Mechanisms by which T7 lysozyme specifically regulates T7 RNA polymerase during different phases of transcription. *J. Mol. Biol.*, 293(3):457–475, 1999. doi:10.1006/jmbi.1999.3135.

[289] B. A. Moffatt and F. W. Studier. T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell*, 49(2):221–227, 1987. doi:10.1016/0092-8674(87)90563-0.

[290] X. Zhang and F. W. Studier. Mechanism of inhibition of bacteriophage T7 RNA polymerase by T7 lysozyme. *J. Mol. Biol.*, 269(1):10–27, 1997. doi:10.1006/jmbi.1997.1016.

[291] M Mandel and A. Higa. Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.*, 53(1):159–162, 1970. doi:10.1016/0022-2836(70)90051-3.

[292] C. F. Barbas, D. R. Burton, J. K. Scott, and G. J. Silverman. Quantitation of DNA and RNA. *CSH Protoc.*, 2007(11), 2007. doi:10.1101/pdb.ip47.

[293] J. A. Bornhorst and J. J. Falke. Purification of proteins using polyhistidine affinity tags. *Methods Enzymol.*, 326:245–254, 2000. doi:0.1016/S0076-6879(00)26058-8.

[294] P. Cutler. Affinity chromatography. *Methods Mol. Biol.*, 244(2):139–149, 2004. doi:10.1385/0-89603-336-8:157.

[295] P. Cutler. Size-exclusion chromatography. *Methods Mol. Biol.*, 59:255–267, 2004. doi:10.1385/1-59259-655-X:239.

[296] S. M. Kelly, T. J. Jess, and N. C. Price. How to study proteins by circular dichroism. *Biochim. Biophys. Acta*, 1751(2):119–139, 2005. doi:10.1016/j.bbapap.2005.06.005.

[297] N. J. Greenfield. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.*, 1(6):2876–2890, 2006. doi:10.1038/nprot.2006.202.

[298] N. J. Greenfield. Applications of circular dichroism in protein and peptide analysis. *TrAC Trends Anal. Chem.*, 18(4):236–244, 1999. doi:10.1016/S0165-9936(98)00112-5.

[299] S. Y. Venyaminov and K. S. Vassilenko. Determination of protein tertiary structure class from circular dichroism spectra. *Anal. Biochem.*, 222(1):176–184, 1994. doi:10.1006/abio.1994.1470.

[300] N. J. Greenfield. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.*, 1(6):2527–2535, 2007. doi:10.1038/nprot.2006.204.

[301] S. R. Martin and M. J. Schilstra. Circular dichroism and its application to the study of biomolecules. *Methods Cell Biol.*, 84:263–293, 2008. doi:10.1016/S0091-679X(07)84010-6.

[302] C. E. Wayne and R. P. Wayne. *Photochemistry*. Oxford University Press, 1996.

[303] J. Singh. *Photochemistry And Pericyclic Reactions*. New Age International, 2005.

[304] F. W. Teale and G. Weber. Ultraviolet fluorescence of the aromatic amino acids. *Biochem. J.*, 65(3):476–482, 1957.

[305] F. W. Teale. The ultraviolet fluorescence of proteins in neutral solution. *Biochem. J.*, 76:381–388, 1960. doi:10.1042/bj0760381.

[306] J. T. Vivian and P. R. Callis. Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys. J.*, 80(5):2093–2109, 2001. doi:10.1016/S0006-3495(01)76183-8.

[307] Y. Chen and M. D. Barkley. Toward understanding tryptophan fluorescence in proteins. *Biochemistry*, 37(28):9976–9982, 1998. doi:10.1021/bi980274n.

[308] E. A. Burstein, N. S. Vedenkina, and M. N. Ivkova. Fluorescence and the location of tryptophan residues in protein molecules. *Photochem. Photobiol.*, 18(4):263–279, 1973. doi:10.1111/j.1751-1097.1973.tb06422.x.

[309] D. B. Calhoun, J. M. Vanderkooi, G. R. Holtom, and S. W. Englander. Protein fluorescence quenching by small molecules: protein penetration versus solvent exposure. *Proteins*, 1(2):109–115, 1986. doi:10.1002/prot.340010202.

[310] S. C. Gill and P. H. Von Hippel. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.*, 182(2):319–326, 1989.

[311] B. F. Peterman. Measurement of the dead time of a fluorescence stopped-flow instrument. *Anal. Biochem.*, 93:442–444, 1979. doi:10.1016/S0003-2697(79)80176-1.

[312] K. C. Neuman and A. Nagy. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods*, 5(6):491–505, 2008. doi:10.1038/nmeth.1218.

[313] Y. Chen, S. E. Radford, and D. J. Brockwell. Force-induced remodelling of proteins and their complexes. *Curr. Opin. Struct. Biol.*, 30:89–99, 2015. doi:10.1016/j.sbi.2015.02.001.

[314] K. Mitsui, M. Hara, and A. Ikai. Mechanical unfolding of alpha(2)-macroglobulin molecules with atomic force microscope. *FEBS Lett.*, 385(1-2):29–33, 1996. doi:10.1016/0014-5793(96)00319-5.

[315] P. E. Marszalek and Y. F. Dufrêne. Stretching single polysaccharides and proteins using atomic force microscopy. *Chem. Soc. Rev.*, 41(9):3523–3534, 2012. doi:10.1039/c2cs15329g.

[316] S. Garcia-Manyes, J. Brujić, C. L. Badilla, and J. M. Fernández. Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of I27 and ubiquitin. *Biophys. J.*, 93(7):2436–2446, 2007. doi:10.1529/biophysj.107.104422.

[317] K. M. Tych, T. Hoffmann, D. J Brockwell, and L. Dougan. Single molecule force spectroscopy reveals the temperature-dependent robustness and malleability of a hyperthermophilic protein. *Soft Matter*, 2013. doi:10.1039/c3sm51439k.

[318] A. Labuda, J. R. Bates, and P. H. Grütter. The noise of coated cantilevers. *Nanotechnology*, 23(2):025503, 2012. doi:10.1088/0957-4484/23/2/025503.

[319] A. Noy, editor. *Handbook of Molecular Force Spectroscopy*. Springer US, 2007. doi:10.1007/978-0-387-49989-5.

[320] M. B. Viani, T. E. Schaäffer, A. Chand, M. Rief, H. E. Gaub, and P. K. Hansma. Small cantilevers for force spectroscopy of single molecules. *J. Appl. Phys.*, 86(4):2258, 1999. doi:10.1063/1.371039.

[321] Z. Schumacher, Y. Miyahara, L. Aeschimann, and P. Grütter. Improved atomic force microscopy cantilever performance by partial reflective coating. *Beilstein J. Nanotechnol.*, 6(1):1450–1456, 2015. doi:10.3762/bjnano.6.150.

[322] M. S. Bull, R. M. A. Sullan, H. Li, and T. T. Perkins. Improved single molecule force spectroscopy using micromachined cantilevers. *ACS Nano*, 8(5):4984–4995, 2014. doi:10.1021/nn5010588.

[323] R. M. A. Sullan, A. B. Churnside, D. M. Nguyen, M. S. Bull, and T. T. Perkins. Atomic force microscopy with sub-picoNewton force stability for biological applications. *Methods*, 60(2):131–141, 2013. doi:10.1016/j.ymeth.2013.03.029.

[324] N. A. Burnham, X. Chen, C. S. Hodges, G. A. Matei, E. J. Thoreson, C. J. Roberts, M. C. Davies, and S. J. B. Tendler. Comparison of calibration methods for atomic-force microscopy cantilevers. *Nanotechnology*, 14(1):1–6, 2003. doi:10.1088/0957-4484/14/1/301.

[325] J. L. Hutter and J. Bechhoefer. Calibration of atomic-force microscope tips. *Rev. Sci. Instrum.*, 64(7):1868, 1993. doi:10.1063/1.1143970.

[326] R. Levy and M. Maaloum. Measuring the spring constant of atomic force microscope cantilevers: thermal fluctuations and other methods. *Nanotechnology*, 13(1):33–37, 2002. doi:10.1088/0957-4484/13/1/307.

[327] I. M. Malovichko. Measuring AFM cantilever stiffness from a thermal noise spectrum. *Bull. Russ. Acad. Sci. Phys.*, 77(8):972–974, 2013. doi:10.3103/S1062873813080248.

[328] R. C. Zinober, D. J. Brockwell, G .S. Beddard, A. W. Blake, P. D. Olmsted, S. E. Radford, and D. A. Smith. Mechanically unfolding proteins: the effect of unfolding history and the supramolecular scaffold. *Protein Sci.*, 11(12):2759–2765, 2002. doi:10.1110/ps.0224602.

[329] J. Mark. *Physical Properties of Polymers*. Cambridge University Press, 2004.

[330] F. Gittes, B. Mickey, J. Nettleton, and J. Howard. Flexural rigidity of microtubules and actin filaments measured from thermal fluctuations in shape. *J. Cell Biol.*, 120(4):923–934, 1993. doi:10.1083/jcb.120.4.923.

[331] C. P. Brangwynne. Microtubules can bear enhanced compressive loads in living cells because of lateral reinforcement. *J. Cell Biol.*, 173(5):733–741, 2006. doi:10.1083/jcb.200601060.

[332] R. C. Zinober. *Elasticity and mechanical unfolding of globular protein domains*. PhD thesis, University of Leeds, 2005.

[333] T. Kawakatsu. *Statistical Physics of Polymers: An Introduction*. Springer Science & Business Media, 2013.

[334] C. Bouchiat, M. D. Wang, J. Allemand, T. Strick, S. M. Block, and V. Croquette. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys. J.*, 76(1 Pt 1):409–413, 1999. doi:10.1016/S0006-3495(99)77207-3.

[335] G. C. K. Roberts, editor. *Encyclopedia of Biophysics*. Encyclopedia of Biophysics. Springer Berlin Heidelberg, 2012.

[336] J. F. Marko and E. D. Siggia. Stretching DNA. *Macromolecules*, 28(26):8759–8770, 1995. doi:10.1021/ma00130a008.

[337] P. J. Flory and M. Volkenstein. Statistical mechanics of chain molecules. *Biopolymers*, 8(5):699–700, 1969. doi:10.1002/bip.1969.360080514.

[338] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith. Entropic elasticity of lambda-phage DNA. *Science (80-.).*, 265(5178):1599–1600, 1994. doi:10.1126/science.8079175.

[339] P. E. Marszalek, H. Lu, H. B. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J M Fernandez. Mechanical unfolding intermediates in titin modules. *Nature*, 402(6757):100–103, 1999. doi:10.1038/47083.

[340] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 276(5315):1109–1112, 1997. doi:10.1126/science.276.5315.1109.

[341] T. Su and P. K. Purohit. Mechanics of forced unfolding of proteins. *Acta Biomater.*, 5(6):1855–1863, 2009. doi:10.1016/j.actbio.2009.01.038.

[342] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block. Stretching DNA with optical tweezers. *Biophys. J.*, 72(3):1335–1346, 1997. doi:10.1016/S0006-3495(97)78780-0.

[343] D. W. Urry, T. Hugel, M. Seitz, H. E. Gaub, L. Sheiba, J. Dea, J. Xu, and T. Parker. Elastin: a representative ideal protein elastomer. *Philos. Trans. R. Soc. B Biol. Sci.*, 357(1418):169–184, 2002. doi:10.1098/rstb.2001.1023.

[344] M. Fixman. Polymer conformational statistics. II. Characteristic function of the rotational-isomeric model. *J. Chem. Phys.*, 58(4):1559, 1973. doi:10.1063/1.1679395.

[345] S. B. Smith, Y. Cui, and C. Bustamante. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, 271(5250):795–799, 1996. doi:10.1126/science.271.5250.795.

[346] N. M. Toan, D. Marenduzzo, and C. Micheletti. Inferring the diameter of a biopolymer from its stretching response. *Biophys. J.*, 89(1):80–86, 2005. doi:10.1529/biophysj.104.058081.

[347] G. I. Bell. Models for specific adhesion of cells to cells. *Science*, 200(4342):618–627, 1978. doi:10.1126/science.347575.

[348] O. K. Dudko, G. Hummer, and A. Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci. U. S. A.*, 105(41):15755–15760, 2008. doi:10.1073/pnas.0806085105.

[349] R. W. Friddle, A. Noy, and J. J. De Yoreo. Interpreting the widespread nonlinear force spectra of intermolecular bonds. *Proc. Natl. Acad. Sci. U. S. A.*, 109(34):13573–13578, 2012. doi:10.1073/pnas.1202946109.

[350] Z. T. Yew, P. D. Olmsted, and E. Paci. Free energy landscapes of proteins: Insights from mechanical probes. In *Single-Molecule Biophys.*, pages 395–417. John Wiley & Sons, Inc., 2011. doi:10.1002/9781118131374.ch14.

[351] S. N. Zhurkov. Kinetic concept of the strength of solids. *Int. J. Fract. Mech.*, 1(4):311–322, 1965. doi:10.1007/BF00962961.

[352] E. Evans and K. Ritchie. Dynamic strength of molecular adhesion bonds. *Biophys. J.*, 72(4):1541–1555, 1997. doi:10.1016/S0006-3495(97)78802-7.

[353] O. K. Dudko, G. Hummer, and A. Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.*, 96(10):1–4, 2006. doi:10.1103/PhysRevLett.96.108101.

[354] G. Hummer and A. Szabo. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci.*, 98(7):3658–3661, 2001. doi:0.1073/pnas.071034098.

[355] Y. J. Zhang and O. K. Dudko. A transformation for the mechanical fingerprints of complex biomolecular interactions. *Proc. Natl. Acad. Sci. U. S. A.*, 110(41):16432–16437, 2013. doi:10.1073/pnas.1309101110.

[356] M. Wikström, U. Sjöbring, T. Drakenberg, S. Forsén, and L. Björck. Mapping of the immunoglobulin light chain-binding site of protein L. *J. Mol. Biol.*, 250(2):128–133, 1995. doi:10.1006/jmbi.1995.0364.

[357] M. L. Scalley, Q. Yi, H. Gu, A. McCormack, J. R. Yates, and D. Baker. Kinetics of folding of the IgG binding domain of peptostreptococcal protein L. *Biochemistry*, 36(11):3373–3382, 1997. doi:10.1021/bi9625758.

[358] M. S. Z. Kellermayer, S. B. Smith, H. L. Granzier, and C. Bustamante. Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science*, 276(5315):1112–1116, 1997. doi:10.1126/science.276.5315.1112.

[359] L. Tskhovrebova, J. Trinick, J. A. Sleep, and R. M. Simmons. Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature*, 387(6630):308–312, 1997. doi:10.1038/387308a0.

[360] A. Minajeva, M. Kulke, J. M. Fernandez, and W. A. Linke. Unfolding of titin domains explains the viscoelastic behavior of skeletal myofibrils. *Biophys. J.*, 80(3):1442–1451, 2001. doi:10.1016/S0006-3495(01)76116-4.

[361] H. P. Erickson. Stretching single protein molecules: titin is a weird spring. *Science*, 276(5315):1090–1092, 1997. doi:10.1126/science.276.5315.1090.

[362] M. S. Z. Kellermayer, S. B. Smith, C. Bustamante, and H. L. Granzier. Complete unfolding of the titin molecule under external force. *J. Struct. Biol.*, 122(1-2):197–205, 1998. doi:10.1006/jsbi.1998.3988.

[363] S. Labeit, B. Kolmerer, and W. A. Linke. The giant protein titin: Emerging roles in physiology and pathophysiology. *Circ. Res.*, 80(2):290–294, 1997. doi:10.1161/01.RES.80.2.290.

[364] L. Dougan and J. M. Fernandez. Tandem repeating modular proteins avoid aggregation in single molecule force spectroscopy experiments. *J. Phys. Chem. A*, 111(49):12402–12408, 2007. doi:10.1021/jp076095+.

[365] Y. Taniguchi, D. J. Brockwell, and M. Kawakami. The effect of temperature on mechanical resistance of the native and intermediate states of I27. *Biophys. J.*, 95(11):5296–5305, 2008. doi:10.1529/biophysj.108.141275.

[366] M. R. Wilkins, E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel, and D. F. Hochstrasser. *2-D Proteome Analysis Protocols*, volume 112, chapter Protein identification and analysis tools in the ExPASy server., pages 531–552. Humana Press, Totowa, NJ, 1999. doi:10.1385/1-59259-584-7:531.

[367] A. W. Blake. *The force resistance of proteins revealed by single molecule atomic force microscopy*. PhD thesis, University of Leeds, 2004.

[368] M. Wolny, M. Batchelor, P. J. Knight, E. Paci, L. Dougan, and M. Peckham. Stable single α-helices are constant force springs in proteins. *J. Biol. Chem.*, 289(40):27825–27835, 2014. doi:10.1074/jbc.M114.585679.

[369] M. Crisma, G. D. Fasman, H. Balaram, and P. Balaram. Peptide models for beta-turns. A circular dichroism study. *Int. J. Pept. Protein Res.*, 23(4):411–419, 1984. doi:10.1111/j.1399-3011.1984.tb02739.x.

[370] M. B. Borgia, A. Borgia, R. B. Best, A. Steward, D. Nettels, B. Wunderlich, B. Schuler, and J. Clarke. Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature*, 474(7353):662–665, 2011. doi:10.1038/nature10099.

[371] A. F. Oberhauser, P. E. Marszalek, M. Carrion-Vazquez, and J. M. Fernandez. Single protein misfolding events captured by atomic force microscopy. *Nat. Struct. Biol.*, 6(11):1025–1028, 1999. doi:10.1038/14907.

[372] C. F. Wright, S. A. Teichmann, J. Clarke, and C. M. Dobson. The importance of sequence diversity in the aggregation and evolution of proteins. *Nature*, 438(7069):878–881, 2005. doi:10.1038/nature04195.

[373] R. W. S. Rounsevell, A. Steward, and J. Clarke. Biophysical investigations of engineered polyproteins: implications for force data. *Biophys. J.*, 88(3):2022–2029, 2005. doi:10.1529/biophysj.104.053744.

[374] J. K. Myers, C. N. Pace, and J. M. Scholtz. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.*, 4(10):2138–2148, 1995. doi:10.1002/pro.5560041020.

[375] R. Guzzi, L. Sportelli, C. La Rosa, D. Milardi, and D. Grasso. Solvent isotope effects on azurin thermal unfolding. *J. Phys. Chem. B*, 102(6):1021–1028, 1998. doi:10.1021/jp9724380.

[376] C. H. Chen, I. W. Liu, R. MacColl, and D. S. Berns. Differences in structure and stability between normal and deuterated proteins (phycocyanin). *Biopolymers*, 22(4):1223–1233, 1983. doi:10.1002/bip.360220414.

[377] H. Lu and K. Schulten. The key event in force-induced unfolding of titins immunoglobulin domains. *Biophys. J.*, 79(1):51–65, 2000. doi:10.1016/S0006-3495(00)76273-4.

[378] D. K. West, E. Paci, and P. D. Olmsted. Internal protein dynamics shifts the distance to the mechanical transition state. *Phys. Rev. E*, 74(6):061912, 2006. arXiv:0612183, doi:10.1103/PhysRevE.74.061912.

[379] K. M. Tych, M. L. Hughes, J. Bourke, Y. Taniguchi, M. Kawakami, D. J. Brockwell, and L. Dougan. Optimizing the calculation of energy landscape parameters from single-molecule protein unfolding experiments. *Phys. Rev. E*, 91(1):012710, 2015. doi:10.1103/PhysRevE.91.012710.

[380] Z. N Scholl and P. E. Marszalek. Improving single molecule force spectroscopy through automated real-time data collection and quantification of experimental conditions. *Ultramicroscopy*, 136:7–14, 2014. doi:10.1016/j.ultramic.2013.07.020.

[381] J. Brujić, R. I. Z. Hermans, S. Garcia-Manyes, K. A. Walther, and J. M. Fernandez. Dwell-time distribution analysis of polyprotein unfolding using force-clamp spectroscopy. *Biophys. J.*, 92(8):2896–2903, 2007. doi:10.1529/biophysj.106.099481.

[382] R. I. Hermans. Probability of observing a number of unfolding events while stretching polyproteins. *Langmuir*, 30(29):8650–8655, 2014. doi:10.1021/la501161p.

[383] D. E. Makarov, P. K. Hansma, and H. Metiu. Kinetic Monte Carlo simulation of titin unfolding. *J. Chem. Phys.*, 114(21):9663, 2001. doi:10.1063/1.1369622.

[384] M. S. Li, D. K. Klimov, and D. Thirumalai. Thermal denaturation and folding rates of single domain proteins: size matters. *Polymer (Guildf).*, 45(2):573–579, 2004. doi:10.1016/j.polymer.2003.10.066.

[385] D. Aioanei, I. Tessari, L. Bubacco, B. Samorì, and M. Brucale. Observing the osmophobic effect in action at the single molecule level. *Proteins Struct. Funct. Bioinforma.*, 79(7):2214–2223, 2011. doi:10.1002/prot.23045.

[386] A. R. Fersht and S. Sato. Φ-value analysis and the nature of protein-folding transition states. *Proc. Natl. Acad. Sci. USA.*, 101(21):7976–7981, 2004. doi:10.1073/pnas.0402684101.

[387] S. B. Ozkan, I. Bahar, and K. A. Dill. Transition states and the meaning of Φ-values in protein folding kinetics. *Nat. Struct. Biol.*, 8(9):765–769, 2001. doi:10.1038/nsb0901-765.

[388] K. A. Dill, S. B. Ozkan, M. S. Shell, and T. R. Weikl. The protein folding problem. *Annu. Rev. Biophys.*, 37:289–316, 2008. doi:10.1146/annurev.biophys.37.092707.153558.

[389] C. F. Wright, K. Lindorff-Larsen, L. G. Randles, and J. Clarke. Parallel protein-unfolding pathways revealed and mapped. *Nat. Struct. Biol.*, 10(8):658–662, 2003. doi:10.1038/nsb947.

[390] E. J. Guinn, B. Jagannathan, and S. Marqusee. Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein. *Nat. Commun.*, 6:6861–6869, 2015. doi:10.1038/ncomms7861.

[391] J. R. Forman and J. Clarke. Mechanical unfolding of proteins: insights into biology, structure and folding. *Curr. Opin. Struct. Biol.*, 17(1):58–66, 2007. doi:10.1016/j.sbi.2007.01.006.

[392] Prof B. Jayaram. Radius of Gyration - SCFBio, 2015. Accessed: 2015-10-01. URL: http://www.scfbio-iitd.res.in/software/proteomics/rg.jsp.

[393] J. Wilf and A. Ben-Naim. Intramolecular hydrophobic interaction in light and heavy water. *J. Chem. Phys.*, 70(6):3079–3081, 1979. doi:10.1063/1.437793.

[394] D. G. Oakenfull and D. E. Fenwick. Hydrophobic interaction in deuterium oxide. *Aust. J. Chem.*, 28(4):715–720, 1975. doi:10.1071/CH9750715.

[395] P. Mukerjee, Paz K., and H. G. Meyer. Micelle formation and hydrophobic bonding in deuterium oxide. *J. Phys. Chem.*, 70(3):783–786, 1966. doi:10.1021/j100875a029.

[396] J. J. Lee and D. S. Berns. Protein aggregation. The effect of deuterium oxide on large protein aggregates of C-phycocyanin. *Biochem. J.*, 110(3):465–470, 1968. doi:10.1042/bj1100465.

[397] C. Eginton and D. Beckett. A large solvent isotope effect on protein association thermodynamics. *Biochemistry*, 52(38):6595–6600, 2013. doi:10.1021/bi400952m.

[398] R. Crainic, R. Wu, D. Otelea, M. M. Georgescu, F. Delpeyroux, S. Guillot, J. Balanant, and M. Tardy-Panit. The replacement of water with deuterium oxide significantly improves the thermal stability of the oral poliovirus vaccine. *Dev. Biol. Stand.*, 87:161–166, 1996.

[399] K. Tomita, A. Rich, C. de Lozé, and E.R. Blout. The effect of deuteration on the geometry of the α-helix. *J. Mol. Biol.*, 4(2):83–92, 1962. doi:10.1016/S0022-2836(62)80040-0.

[400] G. Vogt and P. Argos. Protein thermal stability: hydrogen bonds or internal packing? *Folding and Design*, 2(4):S40–S46, 1997. doi:10.1016/S1359-0278(97)00062-X.

[401] S. M. Malakauskas and S. L. Mayo. Design, structure and stability of a hyperthermophilic protein variant. *Nat. Struct. Biol.*, 5(6):470–475, 1998. doi:10.1038/nsb0698-470.

[402] R. Guzzi, C. Arcangeli, and A. R. Bizzarri. A molecular dynamics simulation study of the solvent isotope effect on copper plastocyanin. *Biophys. Chem.*, 82(1):9–22, 1999. doi:10.1016/S0301-4622(99)00097-6.

[403] A. M. Tsai, T. J. Udovic, and D. A. Neumann. The inverse relationship between protein dynamics and thermal stability. *Biophys. J.*, 81(4):2339–2343, 2001. doi:10.1016/S0006-3495(01)75880-8.

[404] E. Paci and M. Karplus. Unfolding proteins by external forces and temperature: the importance of topology and energetics. *Proc. Natl. Acad. Sci. USA*, 97(12):6521–6526, 2000. doi:10.1073/pnas.100124597.

[405] I. E. Sanchez and T. Kiefhaber. Hammond behavior versus ground state effects in protein folding: Evidence for narrow free energy barriers and residual structure in unfolded states. *J. Mol. Biol.*, 327(4):867–884, 2003. doi:10.1016/S0022-2836(03)00171-2.

[406] J. M. Fernandez and H. Li. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science*, 303(5664):1674–1678, 2004. doi:10.1126/science.1092497.

[407] P. Zheng, Y. Wang, and H. Li. Reversible unfolding-refolding of rubredoxin: a single-molecule force spectroscopy study. *Angew. Chem. Int. Ed. Engl.*, 53(51):14060–14063, 2014. doi:10.1002/anie.201408105.

[408] M. T. Woodside and S. M. Block. Reconstructing folding energy landscapes by single-molecule force spectroscopy. *Annu. Rev. Biophys.*, 43:19–39, 2014. doi:10.1146/annurev-biophys-051013-022754.

[409] R. K. Ainavarapu, J. Brujić, H. H. Huang, A. P. Wiita, H. Lu, L. W. Li, K. A. Walther, M. Carrion-Vazquez, H. B. Li, and J. M. Fernandez. Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophys. J.*, 92(1):225–233, 2007. doi:10.1529/biophysj.106.091561.

[410] L. W. Li, S. Wetzel, A. Pluckthun, and J. M. Fernandez. Stepwise unfolding of ankyrin repeats in a single protein revealed by atomic force microscopy. *Biophys. J.*, 90(4):L30–L32, 2006. doi:10.1529/biophysj.105.078436.