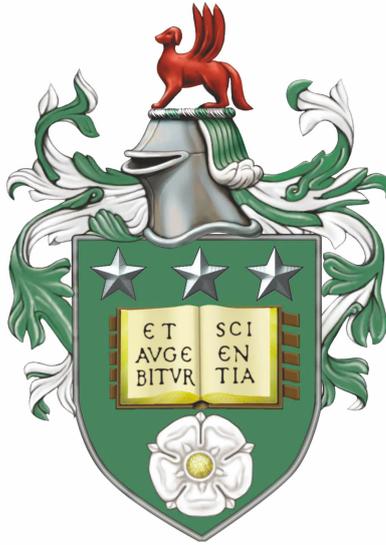# Single molecule studies of protein unfolding in highly saline solutions

Danielle Lauren Walsh

School of Physics and Astronomy
The University of Leeds

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

January 2016

For Mum, Dad, Natalie and Grandma, who have never stopped believing in me.

Declaration

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. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

"Life is not easy for any of us. But what of that? We must have perseverance and above all confidence in ourselves. We must believe that we are gifted for something and that this thing must be attained." – Marie Curie

Acknowledgements

I would like to say thank you to my supervisors Dr. Lorna Dougan and Dr. David Brockwell. Thank you to Lorna for giving me this opportunity and for always believing in my abilities and to David for providing an endless supply of knowledge.

Special thanks go to Dr. Kasia Tych for the vast amount of help and support you have shown me. Also to Megan Hughes, who has been there for every laugh, every tear, every failure and every triumph. For always putting a smile on my face, even through my hardest days.

To all the other members of the Dougan group (Ellen, Jim, Marcelo, Matt, Mike, Natasha, Sam and Toni) and MNP (too many to mention!), thank you for providing the most fun fuelled PhD experience I could have hoped for.

Many thanks to the members of the Hunter, Kendrick, Madho, Newton, Pagden household for allowing me to be a glorified squatter while I wrote my thesis.

Mostly I would like to thank my family and friends as I would not have made it this far with out their support. I hope I have made them proud. To my mum, who is my best friend and my biggest support through the toughest times. For always welcoming me home when I was in need and for always listening to me. To my dad, who has always encouraged me to learn, whether it was science, painting, sports, reading or music. For sitting with me for hours whilst I revised, for learning a subject just to be able to teach me and for always believing in my capabilities. To Natalie, who is the strongest person I know, for always being my inspiration. To my Grandma, for being proud of everything I have achieved and for never wanting me to leave home! Finally to my Nana, who

hopefully would have given this thesis a 12 out of 10.

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.

S. Lenton, **D.L. Walsh**, N.H. Rhys, A. Soper and L. Dougan. Structural evidence for solvent-stabilisation by aspartic acid as a mechanism for halophilic protein stability in high salt concentrations. Paper not yet submitted.

Other publications

D.L. Walsh and L. Dougan, Extremophiles: surviving salt and space. *Chemistry Review*, 23:2–5, 2014.

List of abbreviations

Abbreviations

Å	Ångstroms
AFM	Atomic force microscope
CD	Circular dichroism
IPTG	Isopropyl β–D–1–thiogalactopyranoside
GdmHCl	Guanidinium hydrochloride
Ni-NTA	Nickel – Nitrilotriacetic acid
NMR	Nuclear magnetic resonance spectroscopy
MC	Monte Carlo
MD	Molecular dynamics
PDB	Protein data bank
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SMFS	Single molecule force spectrosccopy
rpm	Revolutions per minute
WLC	Worm-like chain model

Quantities

μg	microgram
mg	milligram
g	gram

kg	kilogram
μL	microlitre
mL	millilitre
L	litre
μM	micromolar
mM	millimolar
M	molar
pN	piconewton
nN	nanonewton
nm	nanometre
mm	millimetre
cm	centimetre
s	seconds
hrs	hours

Abstract

Life on earth has been found thriving in a number of extreme environments, including those of high salinity, high and low temperatures and pH, and high pressure. Organisms which live in the presence of large quantities of salt are known as halophilic (meaning salt–loving), such as in the Dead Sea.

Proteins are fundamental components of all living organisms. They are large, complex molecules that carry out many processes within a cell. Halophilic proteins are of great interest due to their ability to remain soluble, flexible and functional under highly saline conditions. Intriguingly, these proteins are unstable in a low saline environment, suggesting a delicate balance between the intermolecular interactions of the protein, salt and solvent. How have halophilic proteins adapted to survive in highly saline environments?

To probe the effect of salt on the mechanical stability of a protein, a combination of molecular biology and single molecule force spectroscopy (SMFS) was used. Protein engineering was utilised to create chimeric polyprotein constructs including a obligate halophilic and a mesophilic protein. SMFS experiments have been carried out using these polyprotein constructs in 0.5 M and 2 M KCl. The studies suggest that an increase in the hydrophobic interactions of a mesophilic protein cause an increase in its mechanical stability. The results also indicate that an obligate halophilic protein does not have an increased mechanical stability in the increased salt concentration. Further studies in combination with molecular dynamics simulations have the potential to gain atomistic information on the mechanical unfolding behaviour of a halophilic protein.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Life on earth has been found thriving in a number of different extreme environments, including those of high salinity, high and low temperatures, pH and high pressure[1]. Organisms live in the presence of large concentrations of salt, such as in the Dead Sea, which has been reported to have an average overall salt concentration of 322.6 gL^{-1} including individual ion concentrations of Mg^{2+} (40.7 gL^{-1}), Na^+ (39.2 gL^{-1}), Ca^{2+} (17 gL^{-1}), K^+ (7 gL^{-1}), Cl^- (212 gL^{-1}) and Br^- (5 gL^{-1})[2]. These organisms are known as halophilic (meaning salt–loving).

This thesis presents protein engineering experiments to construct halophilic and non–halophilic polyproteins and single molecule force spectroscopy experiments to gain information on the stability and dynamics of halophilic proteins. The first chapter will begin by introducing proteins, their structure and interactions. The study of protein stability will then be outlined. In order to fully understand halophilic proteins it is necessary to describe the co–solutes (proteins and salt) and their interactions with the solvent. To do this, information will be provided on salt, its interactions with water and current literature available on the stability of a protein in salt. A review on the literature available on halophilic proteins will then be presented and the major conclusions summarised. Finally I will introduce the study of proteins using single molecule force spectroscopy, highlighting important studies carried out and where more research needs to be done.

The chapter will finish with a discussion of all of the information given and will outline the aims of this project.

1.1   Protein structure and function

Proteins are naturally occurring, unbranched polymers of subunits called amino acids (figure 1.1). They are a vital component of all living organisms; they are responsible for almost all functions of a cell, its internal organization and the manufacture of other molecules. In order to understand the function of living organisms at the molecular level, it is therefore vital to understand proteins. Proteins are made up of at least 40 amino acids that are joined together by a peptide bond[3]. Proteins can have a wide range of sizes, some are over 10000 amino acids in length. Since the 3D structure of a protein was first documented in 1958 by Kendrew et al. and later improved by Perutz et al., over 115,000 protein structures have now been recorded[4–6].
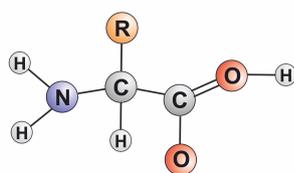
Figure 1.1: The structure of an amino acid. Each amino acid consists of a main–chain backbone structure (The amine nitrogen, a carbonyl oxygen and the two intervening carbons) and a unique side–chain group (R), which give them different chemical properties.

This section will describe the building blocks and hierarchical structures of proteins, introduce the importance of protein folding and the interactions which govern this, and give some examples of their important roles. The section will finish with a description of how protein stability is defined.

1.1.1 Protein structure

The wide variation in the function of proteins is possible due to their precise and complex structural arrangement. An overview of this hierarchical structure is shown in figure 1.2.
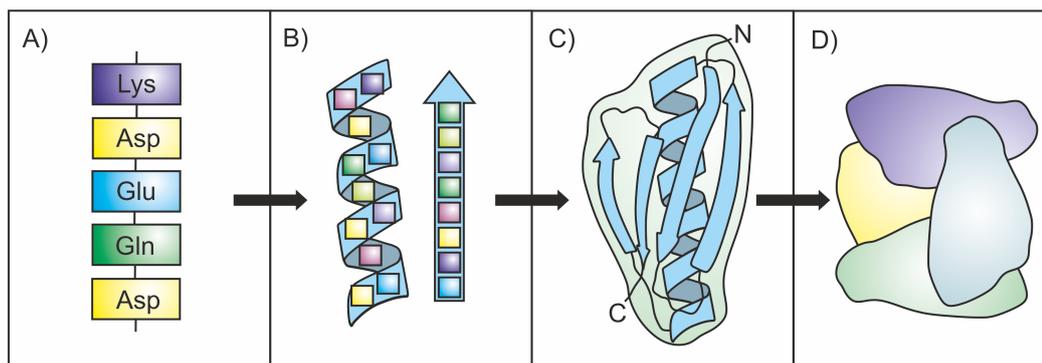
Figure 1.2: The hierarchical structure of a protein. A) The amino acid sequence is referred to as the primary structure of a protein where each amino acid is represented by its three letter abbreviation. B) The secondary structure describes the local conformations of the polypeptide chain. C) The spatial arrangement of these secondary structure elements forms the tertiary structure; this is the 3 dimensional structure of the protein. The C– and N–termini are shown as C and N. D) Many proteins consist of a number of polypeptide chains arranged into a multimeric complex, giving the protein its quaternary structure.

1.1.1.1 Primary structure

Proteins are made up of a sequence of sub–units called amino acids, each linked by a covalent peptide bond, to form unbranched polypeptide chains, represented in figure 1.2 A). Each protein has its own unique amino acid sequence with varying chain length. There are 20 naturally occurring amino acids that have been identified, which differ from each other with respect to one side chain R–group, see figure 1.1. The R–group gives each amino acid its unique properties such as their hydrophobicity or charge. They are usually classified into 6 groups, shown in figure 1.3.
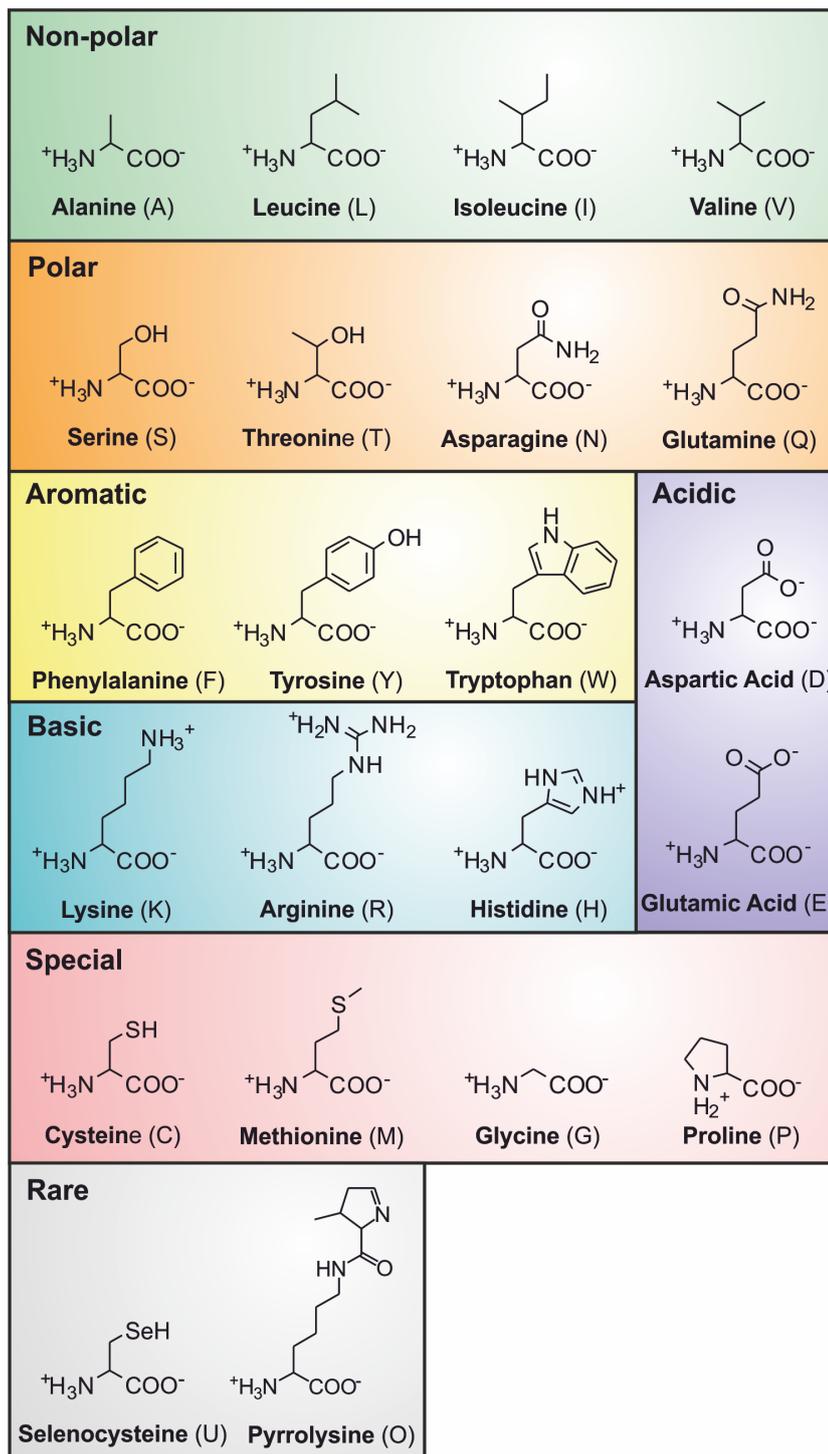
Figure 1.3: The 22 naturally occurring amino acids. The amino acids are grouped according to the properties of their side chains. Non–polar (green), polar (orange), aromatic (yellow), acidic (purple), basic (blue), special cases (pink), which include sulphur–containing- and imino–moieties, and rare (grey). Each amino acid has an L– configuration.

1.1.1.2 Secondary structure

The restricted torsional angles of the polypeptide backbone and interactions between the amino acids give rise to specific secondary structure elements of helices, sheets and loops, see figure 1.4. The most common structures are right–handed –helices and parallel and anti–parallel β–sheets. α–helices are stabilised by hydrogen bonds between the amino acids, β–sheets form rigid structures within the protein and loops connect the other secondary structures together.
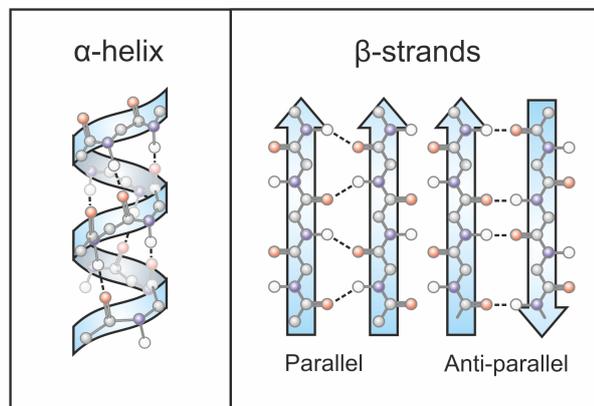
Figure 1.4: Examples of secondary structure elements. The figure shows the atomic structure with intramolecular hydrogen bonds as dashed lines in front of the representative ribbon structure. The carbonyl oxygen is shown in red, the amide nitrogen is shown in blue and the backbone carbons are shown in grey. The figure shows an α–helix as well as parallel and anti–parallel β–sheets.

1.1.1.3 Tertiary structure

The 3D protein structure is governed by the interactions, described in section 1.1.2, between structural elements of the protein. The core of the folded protein is generally rich in non–polar amino acids, while the surface contains mainly polar and charged amino acids. The different folds of proteins can be classified according to the arrangement of the secondary structure elements relative to each other.

1.1.1.4 Quaternary structure

The weak interactions that enable protein folding may also bind together a number of polypeptides to produce larger structures, represented in figure 1.2 D). One example of this is hemoglobin, which binds two α–globin polypeptide chains and two β–globin polypeptide chains to form a quaternary structure. The formation of this structure is crucial to the function of the protein. The 3D structures of proteins are available from the Protein Data Bank (PDB), which is a repository of information about the 3D structures of large biological molecules[6]. Each protein structure is given a unique, 4 digit PDB code, see figure 1.5.
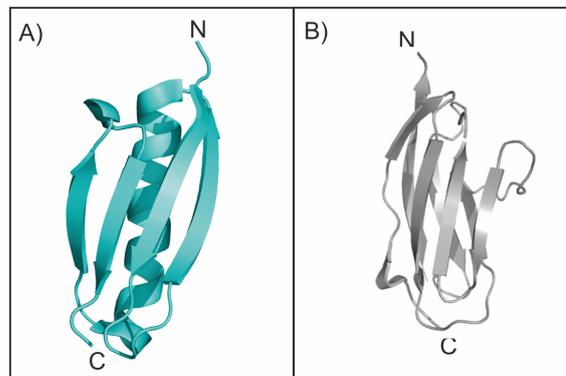
Figure 1.5: The 3D structures of two proteins. Figure A) shows the 3D structure of protein L (PDB code: 1HZ6) and figure B) shows the 3D structure of I27 (PDB code: 1TIT)[6–8]

.

1.1.2 Interactions of protein residues

Proteins are able to form complex 3D structures due to the inter– and intramolecular interactions between the sections of the polypeptide chain and its surrounding environment. The side chain R–groups of amino acids interact in a number of ways, all of which are crucial for protein folding[9]. An overview of the dominant interactions between amino acid residues is shown in figure 1.6. The interactions of the amino acid residues of a protein is a huge subject that has been well researched. A large amount of literature

is available on the subject, a brief description of each interaction is given below[10–12].
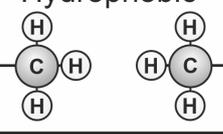
Hydrophobic	Disulphide bond
<40 kJ mol^{-1} 0-10 nm	167 kJ mol^{-1} 0.2 nm
Electrostatic	Hydrogen bond
Environment dependant	1.6-5.3 kJ mol^{-1} 0.25 nm
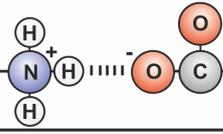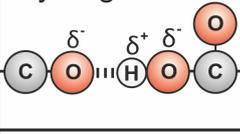
Figure 1.6: The dominant interactions between protein residues. The figure shows atomic representations of hydrophobic interactions, electrostatic interactions, a disulphide bond and a hydrogen bond. The approximate strength of each bond in kJmol^{-1} and the approximate interaction length in nm is shown for each interaction[13–18].

1.1.2.1 Van der Waals interactions

Van der Waals interactions are caused by the disturbance of an electron cloud on an atom when approaching another atom, creating a temporary dipole, see figure 1.7 for a representation. These are weak interactions and non–specific but can become important when molecular surfaces are brought close together. Van der Waals interactions may occur between any atoms but they are relatively weak. Each of these interactions are $<$ 4 kJ mol^{-1} per atom[16]
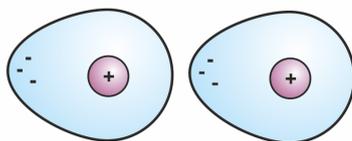
Figure 1.7: The dipoles of van der Waals interactions. The nucleus of the atom is represented in purple and the electron cloud is represented in blue. An interaction is formed between the partially positive and partially negative ends of the atoms.

7

Although Van der Waals forces are weak and transient, they are an important component of protein structure. Due to the close proximity of amino acids residues when a protein is folded, the total number of Van der Waals interactions is very large. Van der Waals forces can play important roles in protein–protein association[19]. When unfolded, the interfaces of the subunits are surrounded by water molecules, all atoms on the surface of the interface are therefore able to form Van der Waals contacts with water. In order to have a favourable association, contacts with the solvent must be replaced by contacts between the protein subunits. This means the proteins need to have good steric complementarity.

1.1.2.2 Covalent interactions

Covalent bonds form when two atoms share one or more pairs of valence electrons. E.g. H_2, each hydrogen atom has one unpaired electron and is therefore unstable. When the two atoms are close they can share their unpaired electrons, i.e. both electrons can orbit both nuclei (see figure 1.8). The molecule formed is stable; it has no net charge and both hydrogen atoms can be considered to have two electrons in their outer energy levels, satisfying the octet rule[20,21].
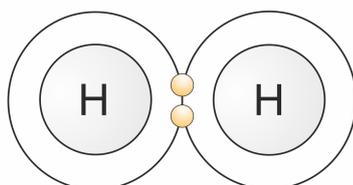
Figure 1.8: Schematic of a covalent bond. The figure shows a schematic of the covalent bond formed between two hydrogen atoms. The electrons are represented in yellow.

Unlike ionic bonds, covalent bonds act between two specific atoms, forming discreet molecules. Covalent interactions form strong bonds and are often irreversible, for example a C–C single bond has a strength of 340–350 kJmol^{-1} [22]. A covalent interaction may be formed between the sulphide atoms of cysteine residues to form a di–sulphide bridge.

1.1.2.3 Electrostatic interactions

Interactions between fully charged groups form ionic bonds, which involves the transfer of a pair of electrons from one atom to another. Interactions may also occur between partially charged (polar) groups. The estimated strength of a charge–charge interaction in a protein is about 87 kJmol^{-1}, these interactions in the presence of water are shielded by water molecules and the interactions are, therefore, experimentally found to be weaker (about 5–20 kJmol^{-1})[16]. There are 3 positively charged and 2 negatively charged naturally occurring amino acids, shown in figure 1.3. These amino acids can interact to stabilise the folded structure of a protein (see figure 1.6).

The importance of electrostatic interactions to the stability of a protein is still under discussion. The effect of modifying the charged amino acids of a protein seem to be highly dependent on context[23]. Proteins from thermophilic organisms have been shown to have an increased number electrostatic interactions compared to their mesophilic counterparts, which are suggested to stabilise the protein[24,25]. The electrostatic interactions of a protein may be studied by changing the pH of its environment. This will either protonate or deprotonate the charged residues. A number of studies have investigated the effect of pH on the thermodynamic stability of a protein[26]. Studies have also changed the salt concentration of the environment to investigate the charge screening of the amino acid residues of a protein[26].

1.1.2.4 Hydrogen bonds

A hydrogen bond is a specific type of electrostatic interaction where one electronegative atom (an atom with a high attraction for an electron), called the hydrogen bond acceptor, interacts with a hydrogen atom (which is highly electropositive) attached to another electronegative atom, called the hydrogen bond donor, shown in figure 1.9. The electron density is attracted to the donor atom, away from the hydrogen, causing a dipole. The partially charged hydrogen atom can then interact with the electron density from the acceptor atom.
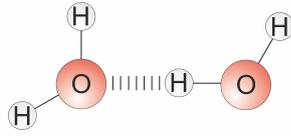
Figure 1.9: Schematic of a hydrogen bond. The figure shows a hydrogen bond between two water molecules where O represents an oxygen and H represents a hydrogen.

Hydrogen bonds are stronger than Van der Waals interactions but weaker than ionic and covalent bonds[16]. They account for a large amount of protein interactions. They are the driving force for the formation of secondary structure elements such as β–sheets and also the interaction of the protein with water molecules[27].

1.1.2.5 Salt bridges

Salt bridges are described as a form of particularly strong hydrogen bonds, formed from the interaction between two charged residues. In a salt bridge interaction, a proton transfers from a carboxylic acid group to an amine or guanidine group. These interactions are generally formed using Lys or Arg and Asp or Glu. Salt bridges have been shown to be important for the stability of thermophilic and hyperthermophilic proteins[28]. Increasing the thermostability of a protein can be achieved by optimising the long–range electrostatic interactions in salt bridges[29].

1.1.2.6 Hydrophobic interactions

Fully charged and partially charged amino acids are able to readily form bonds with water molecules. Amino acids with non–polar residues cannot form hydrogen bonds with water. They tend to aggregate in order minimise their interaction with the water molecules (see figure 1.10)[30].
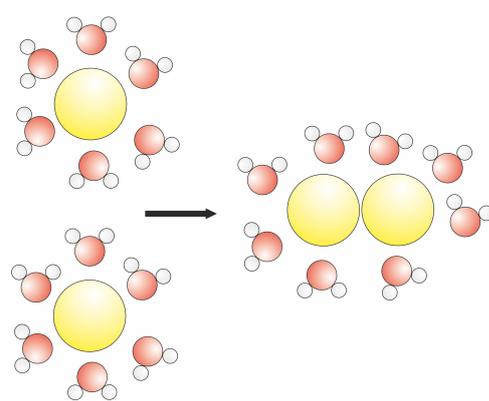
Figure 1.10: A non–polar molecule in water. The non–polar molecule is shown in yellow, oxygen is shown in red and hydrogen is in white. The two non–polar molecules cluster together to minimise the surface contact area with the water.

The hydrophobic effect is described as the driving force for initial folding of proteins where the non–polar residues in a polypeptide chain, group and form a hydrophobic core [31].

1.1.3 Protein stability

In order for most proteins to function, they must remain folded in a unique 3D structure, the native state. A protein can adopt its native state structure by self–assembly. The unique 3D structure of a protein depends on its sequence of amino acids. Although the amino acid sequence of a protein may be easily determined, the process of folding is complex and involves many interactions. Protein folding can often occur through multiple intermediates, which are shown al local minima in figure 1.11 A). In order to understand how a protein remains folded, it is possible to study the stability of its native state.

1.1.3.1 Thermodynamic stability

If a protein is in a two–state system, the molecules will be in equilibrium between the number of unfolded and folded states. The equilibrium constant, K_U, describes the ratio

of the unfolding and folding rates, $K_U = \frac{k_U}{k_F}$. It can, therefore, be used to determine which state the protein favours[32]. Intermediates structures are generally unstable and have very short lifetimes.

$$\text{Native} \xrightleftharpoons{K_U} \text{Unfolded}$$

The thermodynamic stability of a protein can be defined as the difference in free energy between its folded and unfolded states, ΔG_U, which can be calculated using equation 1.1[33–35], were R is the gas constant and T is the temperature.

$$\Delta G_U = -RT \ln K_U \tag{1.1}$$

Where ΔG_U^o is equal to ΔG_U in standard conditions:

$\Delta G_U^\text{o} > 0$ native state favoured

$\Delta G_U^\text{o} < 0$ unfolded state favoured

$\Delta G_U^\text{o} = 0$ equal mix of N and U

The are two important components of ΔG given in the equation, $\Delta G = \Delta H - T\Delta S$[35]:

i) Enthalpy

The enthalpy, H, of a system is defined $H = U - PV$, where U is the internal energy, P is the internal pressure and V is the internal volume of the system. Under physiological conditions, the pressure and volume are almost negligible so in the majority of biological systems, the enthalpy is equal to the internal energy.

ii) Entropy

The thermodynamic entropy defined $\Delta S = \frac{q_\text{rev}}{T}$, where q_rev is heat transferred between two states and T is absolute temperature. The change in thermodynamic entropy is defined as the measure of the energy dispersed in a process.

12

1.1.3.2 Kinetic stability of a protein

The process of protein folding is often represented using an energy landscape, shown in figure 1.11. Energy landscapes are generally depicted as funnels to show that energy decreases, on average, as the native 3D structure is formed, reaching a global minimum[36]. If the energy landscape was smooth, protein folding would proceed without occupying any folding intermediates[37]. An energy landscape with stable, partially folded intermediates is depicted with local minima, referred to as kinetic traps.
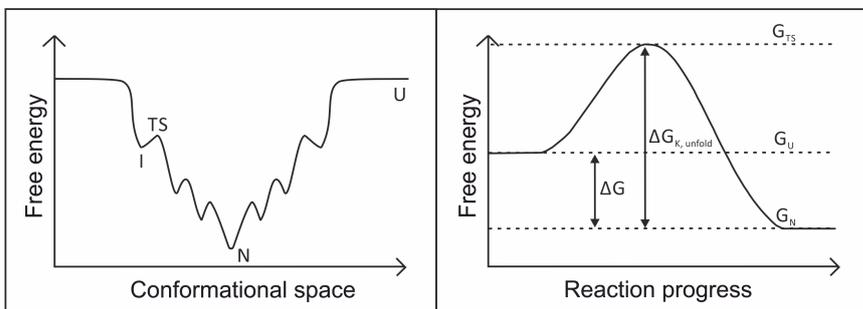
Figure 1.11: A representation of an energy landscape for protein unfolding and a basic free energy diagram A) A representation of an energy landscape showing the unfolded state, U, the native state, N, an intermediate, I, and a transition state, TS. Proteins can have complex energy landscapes where protein folding can occur through many folding intermediates but is biased towards the native state. B) A basic two-state free energy diagram, where the unfolded and folded states are separated by a single barrier. The diagram shows ΔG, the thermodynamic stability, $\Delta G_K, unfold$ the kinetic stability for unfolding, G_{TS}, the free energy of the transition state, G_U the free energy of the unfolded state and G_N the free energy of the native state.

Figure 1.11 B) depicts a two state unfolding reaction. In a two–state unfolding reaction, the folded and unfolded states are separated by a single energy barrier. This barrier is known as the transition state of the protein, TS. The kinetic stability describes the height of the barrier separating the two states, which is also known as the activation energy of this reaction. The kinetic stability can be determined by studying the rate of unfolding of a protein[38]. A kinetically stable protein has a high energy TS, which results in a very slow unfolding rate. Even if ΔG_U is favourable for unfolding, e.g. at high concentrations of denaturant, the high activation energy of the TS will significantly

slow down the unfolding rate of the protein.

The physical basis for kinetic stability is still under debate. Studies have found that the addition of hydrophobic residues, di–sulphide bonds, an increased number of electrostatic ion pairs and co–ordination of metal ions all increase the kinetic stability of a protein[39–42].

1.2 Salt and its effect on proteins

This section will introduce the background information necessary before introducing halophilic proteins. Firstly the section will introduce the chemical structure of a salt. The structure of water and the effect of salt ions on the structure of water will then be described. Lastly the section will introduce the information currently available on the effects of salt ions on proteins.

1.2.1 What is a salt?

A salt is defined as a chemical compound consisting of an assembly of cations and anions[43]. Ionic bonds form when one atom donates an electron to another atom, e.g. sodium chloride (NaCl)[44]. When solid sodium and gaseous chlorine react, each sodium atom gives one electron to a chlorine atom, forming ions (Na^+ and Cl^-). The resulting positively charged atoms or molecules are called cations and the resulting negatively charged atoms or molecules are called anions. The charges on these atoms attract and associate to form a stable crystal lattice.
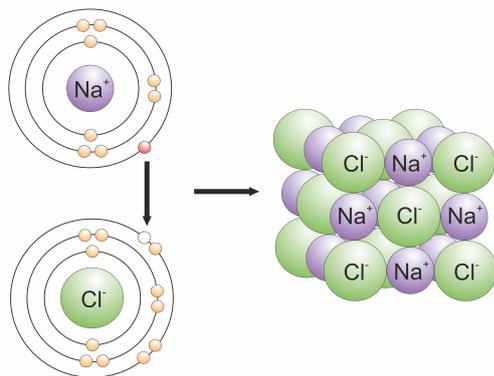
Figure 1.12: Formation of the salt, NaCl. The electrons are represented in yellow, sitting on black circles that represent the energy levels. Chlorine is shown in green and sodium is shown in purple.

1.2.2   Water and its interactions

In vivo, most proteins exist in an aqueous solution. It is therefore important to understand the structure and interactions of water. The oxygen and hydrogen atoms in a water molecule both attract the electrons shared in their covalent bond[44]. The strength of this attraction is named the electronegativity of an atom. The attraction for the electrons is stronger for the oxygen atom, as it is more electronegative. The shared electrons are therefore much more likely to be found orbiting the oxygen atom, causing a partial negative charge on the oxygen and a partial positive charge on the hydrogen. Molecules that have partial charge separation are called polar, figure 1.13 shows a polar water molecule.
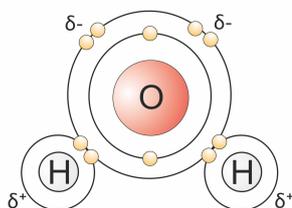
Figure 1.13: A polar water molecule. Electrons are represented in yellow, sitting on black circles that represent energy levels. The partial charges of the atoms are shown.

The partial charge of one polar molecule attracts the partial charge of another, causing an interaction between polar molecules, named dipole–dipole interactions. When a hydrogen atom, bonded to a highly electronegative atom, such as oxygen, interacts with another highly electronegative atom, this is called a hydrogen bond. Hydrogen bonds form between water molecules.

1.2.2.1 Interaction of water with salt ions

Water molecules can be attracted to any substance that has an electrical charge, either full (ionic) or partial (polar). A polar substance is held together by strong intermolecular dipole–dipole interactions. The polar groups of the substance can interact favourably with water molecules allowing spontaneous dissolution. When a salt is placed into water the electrical attraction of the water molecules disrupts the forces holding together the lattice, causing the ions to dissociate, shown in figure 1.14. The crystal lattice dissociates into freely movable ions, increasing the entropy of the solution due to the increased disorder.
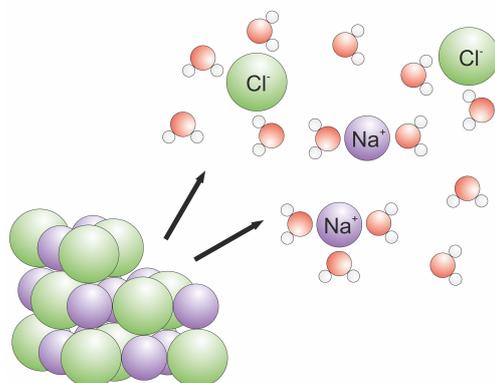
Figure 1.14: Schematic of a dissociation of salt ions in water. The water molecules are represented by red and white circles.

The highly electronegative oxygen of water, with two lone pairs of electrons, is attracted to the positive metal ion. The water molecule rotates so that its polarized charge faces the opposite charge of the ion. As the water molecule moves towards the ion, it

16

breaks its hydrogen bonds with other water molecules. The hydrogens of the water molecules interact with the negative ions. The association, or solvation, of the water molecules around each salt ion creates an effectively larger radius and, therefore, increases the distance between the positive and negative ions and weakens their attractive force. The structure of the water molecules of the hydrated ions is more ordered, decreasing entropy. Absolute values of enthalpy and entropy values for a selection of ions can be found in table 1.1.

The amount of energy needed to break up 1 mole of the salt crystal lattice is the termed the lattice dissociation enthalpy. The greater the force of attraction between salt ions, the greater the energy needed to break the ionic lattice. The radius of the ion and the ionic charges are the most important considerations:

i) The smaller the ion radius or the greater the ion charge, the greater the lattice enthalpy. Smaller ions are able to move closer together causing a larger attractive force

ii) A larger charge on the ion causes a greater attractive force for water molecules.

Hydration enthalpy is a measure of the energy released when attractions are set up between positive or negative ions and water molecules. The overall free energy of the solution depends on the quantities of the lattice enthalpy, the hydration enthalpies and change in entropy of the system.

Ion	Radius (pm)	ΔG°_{hyd} (kJmol^{-1})	ΔH°_{hyd} (kJmol^{-1})	ΔS°_{hyd} (kJmol^{-1})
Na$^+$	102	-371	-404	-110
K$^+$	138	-300	-321	-70
Mg^{2+}	72	-1836	-1931	-320
Ca^{2+}	100	-1517	-1586	-230
Cl$^-$	181	-334	-361	-90
Br$^-$	196	-309	-330	-70

Table 1.1: Absolute values of the change in free energy of hydration, ΔG°_{hyd}, enthalpy of hydration, ΔH°_{hyd} and entropy of hydration ΔS°_{hyd}, along with radius for the salt ions found in abundance in the dead sea[2,45].

1.2.3 The effect of ions on the structure of water

Salt ions have been shown to exert strong electric fields on surrounding water molecules, which disrupts the hydrogen bond network of waters[46]. The water molecules reorient around the ions causing the formation of hydration spheres, see figure 1.15. The effect of ions on water has been generally described as arising from a balance between the water–water interactions and ion–water interactions.
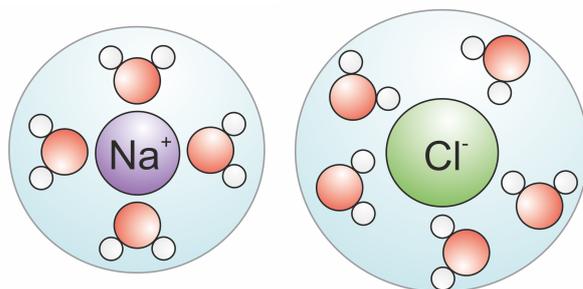
Figure 1.15: The hydration layer of a sodium ion and a chloride ion in water. The water molecules are represented by red and white circles.

A controversial issue still under discussion is the extent to which ions affect the structure and dynamics of surrounding water molecules. Hribar et al. used a simple molecular model to calculate how ions affect neighbouring water molecules. The model predicted that small ions have a larger degree of ordering in water compared to larger ions. This is due to the difference in charge densities of the ions, and the balance of forces that determine the water structure, i.e. electrostatics (dipole interactions) and hydrogen bonding. Small ions have high charge densities therefore causing strong electrostatic ordering of nearby waters, reducing the number of hydrogen bonds. Large ions have low charge densities, so water–water hydrogen bonds dominate[47].

Some experiments have shown little effect on the dynamics of water molecules beyond the first hydration sphere of the ion[48–50]. One study by Cappa et al. used X–ray absorption spectroscopy to examine the effects of cations on water[49]. The results showed only a small influence of monovalent cations on the unoccupied molecular orbitals of water and the distortion of the hydrogen bond network was found to be negligible.

Recent studies, using new spectroscopy techniques accompanied by computer simulations, have now suggested the influence of ions on the structure of water does in fact reach beyond the first hydration shell[51–54]. Teilrooij et al. examined water dynamics around a range of ions using terahertz and femtosecond infrared spectroscopy, which allowed detection of water reorientation dynamics along different molecular axes[53]. The results showed the structure of the hydration sphere of a strongly hydrated ion, e.g. SO_4^{2-}, was greatly dependant on the properties of the counterion. If the counterion was weakly hydrated e.g. Cs^+, the strongly hydrated ion was found to be surrounded by a semi–rigid hydration sphere, where movement of the water molecule was restricted only in a certain direction. If the counterion was strongly hydrated, e.g. Mg^{2+}, the hydrogen–bond network restricted the movement of water in multiple directions.

1.2.4   The effect of salt on a protein

To study halophilic proteins, it is important to highlight firstly the effects of salts on a mesophilic protein, a protein that is not halophilic. The effects of salts on the precipitation of proteins were first examined by Hofmeister in 1988[55,56]. He studied the effects of a number of salts on proteins purified from egg white. By keeping either the cation or the anion constant, he was able to separate the cationic and anionic effects. The anions and cations were then ordered according to their ability to precipitate the proteins, the list is shown in figure 1.16. Protein precipitation was used as a marker for the solubility of a protein in various salts. Later Hofmeister attempted to explain his observations using the hydration of the ion[57].

The terms 'kosmotrope' (order–maker) and 'chaotrope' (disorder–maker) were originally used to describe solutes that either stabilised, or destabilised respectively, proteins and membranes[56]. Kosmotropes stabilise proteins and hydrophobic aggregates in solution and reduce solubility. They were described as having the ability to order water molecules beyond their immediate hydration shells and effectively 'steal' water from the protein, leading to a 'salting–out effect'[47,58]. Chaotropes were described as lacking the ability to order water molecules beyond their immediate hydration shells. They promote
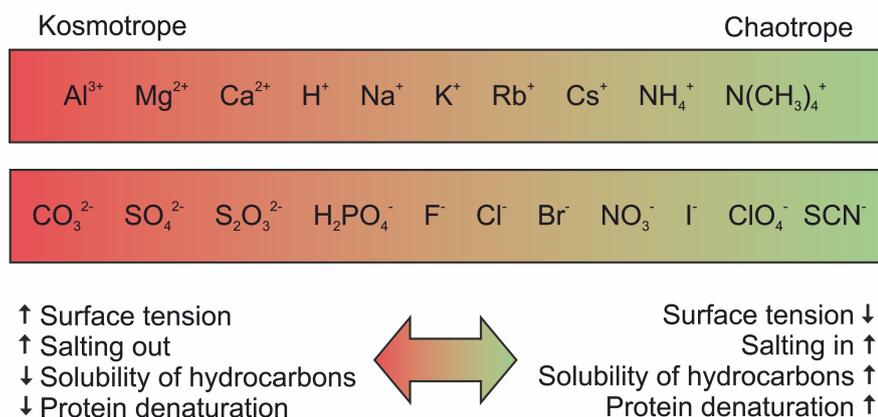
Figure 1.16: The Hofmeister series of ion specific effects on proteins. The series of cations are shown in the top diagram and the series of anions are shown in the bottom diagram. The colours, green to red, represent the kosmotropes going to chaotropes. The properties of the ions are given underneath for kosmotropes and chaotropes.

protein denaturation, destabilise hydrophobic aggregates and increase solubility, leading to a 'salting in' effect. Section 1.2.3 shows some of the flaws in this argument.

The effects of ions on the stability of a protein are now generally ascribed to one of the following mechanisms:

i) screening of electrostatic interactions

ii) the Hofmeister effect

iii) ion specific binding

A vast amount of literature is available that describe the effects of salts on proteins[47,56,58–63]. Here, a few select examples have been presented to describe each mechanism mentioned.

1.2.4.1 Protein stability in low salt concentrations

Oppositely charged ions are attracted to each other. As salt ions are free to move in a solution, anions will likely to be found near cations and vice versa[21]. The overall solution is electrically neutral but close to one ion, there will be an excess of counter ions. The time–averaged sphere surrounding the central ion, which has an excess of counter

ions, will have a net charge equalling that of the central ion but with the opposite charge. This sphere is named the screening layer. The free energy of the central ion is lowered due to the presence of its screening layer. The thickness of this layer effectively cuts off the long–range coulomb interactions. This theory predicts that the measured free energy of a protein should vary linearly with the square root of the ionic strength. The ionic strength of a solution is proportional to the concentration of salt and to the charge of the component ions but is independent of the ion type[64]. This theory is only valid in low concentrations of salt, i.e. lower than about 100 mM. Salt ions at low concentration can exhibit either a stabilizing or a destabilising effect on a protein, this varies with the properties of each protein.

This effect of the screening of surface charge–charge interactions, which is independent of ion type, has been demonstrated in a study by Kumar et al[63]. A range of salts with a constant cation and anions ranging from kosmotropes to chaotropes was examined (Na_2SO_4, $NaCl$, $NaBr$, and $NaNO_3$). The effect of these salts on the stability of the protein diferric–ovotransferrin (Fe_2oTf) as a function of pH and urea was measured. Low concentrations of salt, e.g. $0.1(\pm 0.02)$ M Na_2SO_4 for each of the salts examined, were found to destabilize the Fe_2oTf protein (assessed by equilibrium experiments), independent of salt type. The destabilisation of the protein was suggested to be due to weakening of Coulombic interactions at the protein surface.

1.2.4.2   Protein stability in high salt concentrations

The Hofmeister series is shown in figure 1.16. The species to the left of Cl^- in this figure are referred to as kosmotropes, they bind effectively with water molecules. These strongly hydrated salt ions are excluded from the protein surface, leading to an increase in the hydrophobic interactions within the protein[65–67]. Increased hydrophobic interactions cause an increase in the stability of a protein. The ions to the right of Cl^- are known as chaotropes, the weakly hydrated ion shown preference for binding to the protein, thus destabilising it[68]. These specific binding sites lead to net anion inclusion in the protein surface. Once these sites are saturated, as salt concentration increases, there becomes a

net anion exclusion of salt ions from the surface. This causes 'salting out' behaviour.

An exact mechanistic explanation of effects of various ions on protein stability is still an active area of research[60,62,68]). A study by Rembert et al. used a combination of NMR, thermodynamic measurements and molecular dynamics simulations to identify specific binding sites of salt ions of the Hofmeister series on an uncharged 600–residue polypeptide[68]. They found evidence that most significant binding affinities of the polypeptide backbone were observed for the weakly hydrated anions, e.g. I^- and SCN^-. The binding was observed for both the backbone amide moiety and α–carbon. The interactions of the backbone with cations were found to be less pronounced. The hydrophobic side chains of the polypeptide were reported not to contribute significantly to anion binding or the resulting 'salting–in' behaviour. Both strongly hydrated anions (SO_4^{2-}) and cations (Na^+) were found to be repelled from the polypeptide backbone and hydrophobic side chains. The study suggests that the exclusion of the strongly hydrated ions from the polypeptide/water interface leads to hydrophobic collapse and a 'salting out' effect on entropic grounds (i.e. excluded volume effect). It should be mentioned that strongly hydrated ions could bind to charged side chains, which may lead to salting–in behaviour when these residues are present in the protein sequence.

1.2.4.3 Protein stability effects of ion specific binding

Observations of specific ion effects at low salt concentrations are attributed to specific binding to stabilise a protein. Lytic transglycosylase from *E. coli* (Slt35) was reported to contain a single metal ion–binding site that resembled a structural domain or found in a large family of calcium–binding proteins[66]. The report of this structural feature was described as only the second observation of such a domain in a prokaryotic protein. The crystal sturctue of the protein was determined by van Asselt et al., which showed that both Ca^{2+} ions and Na^+ ions could bind to the domain. Temperature–induced unfolding experiments demonstrated that Ca^{2+} ions were bound preferentially and only Ca^{2+} thermally stabilised Slt35, showing Ca^{2+} binding is important for the stability of the protein.

A vast number of studies are available investigating ion–protein interactions using both theoretical and experimental methods. Very few studies have applied the information gained to the specific problem of how halophilic proteins have adapted to their environment[69].

1.3 Halophilic organisms and their proteins

This section will review the literature on 3 halophilic proteins, which have been studied most extensively, highlighting the key relevant findings. Subsequently the section will review the literature on the protein of interest for this thesis, 2Fe–2S ferredoxin from *Haloarcula marismortui*. Further studies of haloadaptation using designed model halophilic proteins will then be presented. Finally the key findings drawn from the literature will be summarised.

1.3.1 Adaptation of organisms

A major obstacle for halophilic organisms to overcome is the water loss caused by osmosis[70]. Osmosis is the flow of solute, a substance dissolved in a solvent, in high concentration across a semi–permeable membrane to a region of lower concentration[71]. The increase in inorganic ions on the outside of the cell causes a concentration gradient across the cell membrane. Water molecules permeate through the cell membrane to restore concentration equilibrium. In order to overcome this problem, organisms have evolved two approaches:

i) The first is the "salt–out" strategy, balancing the osmotic pressure by increasing the amount of certain organic solutes inside the cell, such as polyols (e.g. glycerol and arabitol), low molecular weight carbohydrates (e.g. sucrose and trehalose), amino acids and their derivatives (e.g. proline and glutamate), unique organic zwitterions and methylamines[72,73].

ii) The second is the "salt–in" strategy, where the organism increases the amount

of inorganic ions inside the cell[74]. This process is energetically cheaper but the intracellular proteins of these organisms must be capable of functioning in the presence of a large amount of ions. These halophilic proteins are of great interest due to their ability to remain soluble, flexible and functional under highly saline conditions.

1.3.2 Halophilic proteins

A number of studies have been carried on halophilic proteins to identify how these proteins are able to function in such high salt concentrations[75–77]. Although many common characteristic adaptations of halophilic proteins have been found, the mechanism of stabilisation by these adaptations is still under investigation.

1.3.2.1 Halophilic Malate Dehydrogenase

The most extensively studied halophilic protein is malate dehydrogenase from the *Haloarcula marismortui* (*Hm*MDH), shown in figure 1.17[78–86]. *Haloarcula marismortui* is a halophilic red archaeon, which was first isolated in the 1960s by Ginzburg et al. from the Dead Sea[87–89]. It is an extremely halophilic organism, i.e. it uses the "salt–in" strategy. The dominant ions within the cytoplasm of *Haloarcula marismortui* have been found at concentrations of 3.7–4 M potassium, 0.5–0.7 M sodium and 2.3–2.9 M chlorine[87].

*Hm*MDH has been shown to be able to function at the high concentrations of KCl/NaCl, similar to its physiological environment[91]. Consequently, most mesophilic proteins would precipitate at such high KCl/NaCl concentrations (see section 1.2.4.2). *Hm*MDH has been found to be unstable at salt concentrations lower than 2.5 M KCl or NaCl[91]. Instability at low salt concentration appears to be a common feature of halophilic proteins. Zaccai et al. measured the stability of *Hm*MDH in varying concentrations of $MgCl_2$ by monitoring the enzymatic activity at different incubation times with various solvents[79]. Mg^{2+} is a salting in ion, usually favouring the folded state of a protein (see section 1.2.4.2) and is also present in high concentrations in the Dead

Figure 1.17: The 3D structure of *Hm*MDH. This figure was created using PDB file 1HLP and Pymol[86,90].

Sea[2]. The study reported that *Hm*MDH was stabilised by concentrations of $0.3 - 1.3$ M MgCl$_2$. This unexpected behaviour, therefore, showed a new model was required to describe halophilic protein stabilisation in various salts.

Halophilic proteins have been found regularly to have weakened hydrophobicity[75]. This is important to increase the solubility of the protein in a high salt environment by allowing greater interaction with the water molecules, although this is not enough to stabilise the structure. The stabilisation model, proposed by Zaccai et al, states that further stabilisation of the protein structure is needed in high salt conditions[79]. Halophilic proteins are thought to have a structure which is able to compete for water ions against the large amount of salt. The increased amount of acidic residues on the surface of the protein is a possible mechanism for achieving this through coordination with hydrated salt ions. This coordination of hydrated salt ions causes a reduction in entropy, which could be compensated for by the favoured formation of hydrated ion bonds to the carboxyl groups. This method of stabilisation is therefore enthalpy driven. Zaccai et al. provided evidence for this model, showing Na ions and Cl ions were included in the hydration shell of *Hm*MDH using small angle neutron scattering. The local

concentrations of Na^+ and Cl^- ions around the protein were found to be higher than concentrations in the bulk[78].

Extensive work on the structure of HmMDH suggested that active protein in solution was a dimer that, at high salt concentrations, had significantly larger solvent interactions than for mesophilic proteins[76]. This structure was thought to be evidence for coordinated hydrated salt ions described by salt stabilisation model for halophilic proteins proposed by Zaccai et al[79]. Further study of the structure of the protein found it was actually tetrameric[80]. As a consequence of this, repeated analysis of experiments using densimetry, X–ray and small angle neutron scattering found that the water binding of HmMDH was now similar to that of a mesophilic homologue. The salt binding, however, remained higher so the stabilisation model proposed by Zaccai et al. of hydrated ion binding remained valid.

The crystal structure of HmMDH was solved by Dym et al in 1995 and the resolution was improved in 2000 by Richard et al., shown in figure 1.17[81,86]. The protein was crystallised in 1.8 M NaCl and resolved to 2.6 Å resolution. The structure of HmMDH was shown to have twice as many acidic residues as the mesophilic homologue and a significant increase in the number of salt bridges[92]. The major interactions connecting each monomer were found to be electrostatic in nature instead of hydrophobic as in the mesophilic homologue, probably arising from the increased number of charged residues. Only two sodium ions were found in the crystal structure for each tetramer, which were bound to two glutamic acid residues that are part of the monomer–monomer interfaces. Two chloride ions were also found in interactions with lysine residues as part of the salt bridge clusters found at the dimer–dimer interface[86]. It was suggested that introducing the protein to low salt concentration would lower the probability of binding the salt ions in these positions, allowing the tetramer to dissociate into monomers, which it does so at 2 M NaCl. The dissociation of the monomers inhibits the protein from functioning.

A review of the studies on HmMDH, written by Mevarech et al., concludes by modifying the solvation–stabilisation mechanism proposed by Zaccai et al.[85]. Two roles are given for the high amount of acidic residues found in halophilic proteins, maintain-

ing the solubility of the protein through interactions with the water and counteracting the increased hydrophobic effect from high salt concentration through repulsion of like charges. The high salt dependence of the protein was suggested to be due to a few, specific, low affinity binding sites on the protein. Large amounts of salt would therefore be needed to overcome the low affinity. The differing stability of halophilic proteins in low salt conditions could then be explained by the hydrophobicity of a given protein and the affinity of the specific ion sites.

1.3.2.2 Halophilic Dihydrofolate Reductase

Another well studied halophilic protein is dihydrofolate reductase from *Haloferax volcanii*, (*Hv*DHFR), shown in figure 1.18[93–95]. *Haloferax volcanii* is a halophilic archaeon, which was also isolated from the Dead Sea (initially named *Halobacterium volcanii*)[96,97].

Figure 1.18: The 3D structure of *Hv*DHFR. This figure was created using PDB file 2ITH and Pymol[90,94].

A study carried out by Bohm and Janike examined protein instability in low salt conditions,instead of protein stability, in high salt conditions[93]. *Hv*DHFR was found to be active in KCl concentrations as low as 0.5 M, although has optimal activity at 3–4 M. The increased amount of acidic residues on the surface, noted as a regular feature of halophilic proteins, were viewed as a destabilising effect. Repulsion of these groups in a low salt environment would lead to instability and denaturing of the protein. Cir-

cular Dichroism (CD) spectroscopy studies showed that when HvDHFR is in low salt conditions, about 0.25 M, the protein was completely unfolded. When the protein was put in a low pH environment, however, salt ions were not needed for stability. Decreasing the pH provides protons, which are able to substitute for the salt ions. HvDHFR was further investigated by computational modelling[93]. A 3D structure of HvDHFR was produced using the amino acid sequence, modelled on the known 3D structure of a mesophilic homologue. The pKa value of the ionisable groups that cause the pH dependence, mentioned above, was estimated to be around 5.3. This value is higher than any single amino acid. It was therefore postulated that clusters of charged residues give to the high pKa and this effect is important for halophilic adaptation.

The crystal structure of HvDHFR was reported in 1998 (figure 1.18), showing a slight increase in the number of acidic residues compared to the mesophilic homologues[94]. The positions of acidic residues predicted by the Bohm model (described above) were shown to be inaccurate; the negative charges tend to orient themselves in opposite directions. Compared to that of a mesophilic homologue from *E. coli*, the crystal structure also revealed there was no significant increase in the number of salt bridges (28 for HvDHFR and 27 for EcDHFR), as was found for HmMDH[92,98].

A study carried out by Wright et al. used CD spectroscopy and fluorescence studies to look at the effect of salt concentration on HvDHFR and a mesophilic homologue[95]. The study pointed out the importance of the hydrophobic residues in the stabilisation of HvDHFR. The weaker hydrophobicity of HvDHFR compared to mesophilic homologues, was shown to be brought about by the apparent preference of small hydrophobic residues (Ala and Val) over large ones (Phe, Tyr, Trp, Ile, Leu, and Met). It was suggested that this allows the protein to have the flexibility necessary to function despite the increase in the hydrophobic effect due to the high concentration of salt. At low salt concentration, the protein is thought to unfold because of the destabilisation caused by the decrease in the hydrophobic effect.

1.3.2.3 2Fe–2S Ferredoxin from *Halobacterium salinarum* (*Hs*Fd)

2Fe–2S ferredoxin from *Halobacterium salinarum* (*Hs*Fd) is a halophilic protein that has also been studied, shown in figure 1.19. *Halobacterium salinarum*, despite its name is also a halophilic archeon, which was first isolated from salted fish about 80 years ago[99,100].
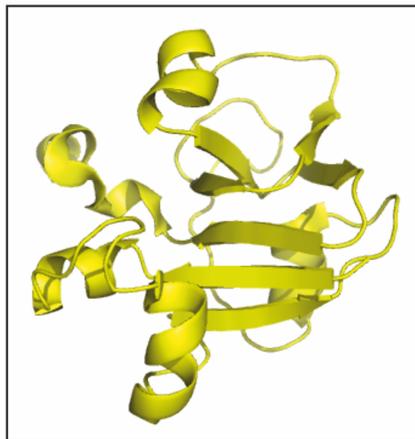
Figure 1.19: The 3D structure of *Hs*Fd. This figure was created using PDB file 1E0Z and Pymol[90,101].

The halophilic protein, *Hs*Fd, was studied by Bandyopadhyay et al. by investigating the structure of the protein in low salt concentrations[102]. A time–dependant loss of secondary structure was shown using CD Spectroscopy when the protein was diluted from high ionic strength solutions (4.5 M NaCl) into solutions with less than 1 M NaCl. Similar results were also found in the region of the 2Fe–2S cluster using absorbance spectroscopy and for the tryptophan residues using fluorescence measurements. This study points out an inserted domain with many acidic residues in the amino acid sequence of *Hs*Fd compared to the mesophilic ferredoxins (discussed in section 1.3.2.4 for *Hm*Fd). Stabilisation of the protein is suggested to be most likely due to extensive solvation of theses acidic residues, where solvation is defined as the binding of water and/or salt. It is suggested that the screening of electrostatic charges through solvation is insufficient to explain the stability of halophilic proteins in very high salt conditions. The increased hydrophobic effect in high salt conditions provides extra stabilisation.

Further investigations by Bandyopadhyay revealed the structure of *Hs*Fd in low salt conditions still contained significantly more secondary structure than *Hs*Fd, which had been completely unfolded in the presence of 8 M urea[103,104]. This suggested that *Hs*Fd is not completely unfolded in low salt but is an intermediate structure. At pH 5 it was found that a high concentration of salt was not needed to stabilise the structure, supporting the need for charge stabilisation to stabilise the structure mentioned previously. Bandyopadhyay also found that *Hs*Fd could be stabilised by the presence of small amounts of polylysine, a polycation[103]. It is expected that polylysine forms salt bridges with the carboxyl groups of the surface acidic residues, again supporting the argument for the increased stability through charge screening. It was shown in CD spectroscopy and fluorescence studies that polylysine was not able to fully substitute for high salt concentration, showing that the screening of electrostatic charges is insufficient for full stabilisation.

The NMR structure of *Hs*Fd was solved in 1995 by Marg et al., shown in figure 1.19[101]. *Hs*Fd was found exhibit greater kinetic stability upon a decrease in salt concentration (0.05 M NaCl) than other halophilic proteins, such as *Hm*MDH[85]. Surprisingly, the NMR structure was found to be unaltered when *Hs*Fd was in 1.5 M and 0.45 M NaCl, suggesting this salt concentration has no effect on the overall structure of the protein. Marg et al. conclude that "the decreased protein stability observed at lower salt concentrations does not reflect structural differences, but is rather due to a lowered energy barrier of unfolding".

1.3.2.4 Introduction to 2Fe–2S Ferredoxin from *Haloarcula marismortui* (*Hm*Fd)

The protein of interest for this thesis is 2Fe–2S ferredoxin from *Haloarcula marismortui* (*Hm*Fd), shown in figure 1.20. The organism, *Haloarcula marismortui*, was described in section 1.3.2.1. Experimental results using this protein are presented in chapter 3.

The X–ray crystal structure of *Hm*Fd was reported in 1996 by Frolow et al., shown in figure 1.20[105]. *Hm*Fd was found to contain a 4–stranded β sheet, distorted into a barrel–
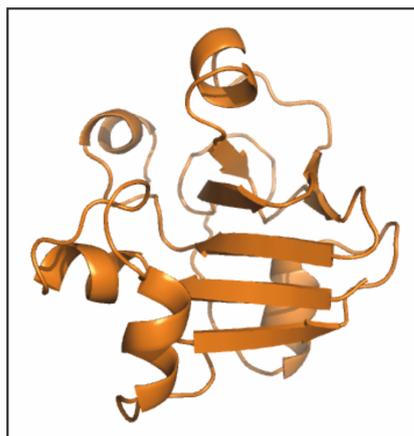
Figure 1.20: The 3D structure of *Hm*Fd This figure was created using PDB file 1DOI and Pymol[90,105].

like structure with 5 surrounding α helices. Four sulphur atoms, from cysteine residues, bind Fe atoms that are involved in a 2Fe–2S cluster, see figure 3.1. The 2Fe–2S cluster is located in the hydrophobic core of the protein. In accordance with the commonly observed features of halophilic proteins, *Hm*Fd has a large number of acidic residues predominantly located on the surface. The protein contains 34 carboxylate groups, which constitute 26.6% of the protein and only six basic residues, located in the C–terminal half of the structure, making *Hm*Fd the highest density of negatively charged residues in the Protein Data Bank[105]. Mesophilic homologues of *Hm*Fd, *Ab*Fd and human Fd have 10% and 19% acidic residue content respectively[106,107]. *Hm*Fd appears to have an extra 22 residue domain inserted into the loop region between two anti–parallel beta strands near the N–terminus when compared to the structures of homologous mesophilic ferredoxins[105–107]. The extra domain contains 15 negatively charged residues and no positively charged ones, and folds into two α–helices. *Hm*Fd only contains five salt bridges in the monomer structure whereas other halophilic proteins, such as *Hm*MDH which contains many more[81]. It is suggested that *Hm*Fd is stabilised by its 2Fe–2S cluster and does not need further stabilisation from extra salt bridges. *Hm*Fd has been reported to be active at salt concentrations as low as 0.4 M NaCl, although no experimental data has currently been published about the protein stability in varying salt concentrations[108].

A study by Elcock and McCammon carried out a theoretical analysis of the contribution of electrostatic interactions to the stability of HmFd[109]. They initially investigate the salt dependence of the pK_a values of the acidic residues aspartate and glutamate, and the basic residues arginine and lysine in HmFd and a mesophilic homologue. The pK_a is defined by equation 1.2.

$$pK_a = -\log_{10}K_a \tag{1.2}$$

Where:

$$K_a = \frac{[A^-][H^+]}{[HA]} \tag{1.3}$$

HA is an acid which dissociates in the conjugate base of the acid A^- and a hydrogen ion. A change in value of pK_a indicates a change in the environment of a residue within a protein[110]. For example, when a protein is unfolded the residue is surrounded by solvent, when a protein is folded the residue may form interactions with other residues. For the acidic residues, the study reports an upward shift in the pK_a values compared to the mesophilic homologue[109]. This is due to the greater contribution of unfavourable interactions upon deprotonation from the higher number of acidic residues in HmFd than in the mesophilic homologue[106,107]. The pK_a shift was found to be larger in low salt conditions than in high salt conditions, so it was suggested less counter–ions were available to shield the electrostatic charges of the acidic residues. It was estimated, using these findings, between pH 5–7 the destabilisation is due to deprotonation of the acidic residues. Increasing the salt concentration allows shielding for the charges of the acidic residues. The increased amount of salt also destabilises the salt bridges. At pH 11, HmFd would be stable in 1 M salt or above. At higher pH values, a larger salt concentration is needed to shield the larger amount of deprotonation caused by the increased pH. It was estimated in this study that HmFd would be increasingly stable in 0.04 M NaCl as pH is lowered. Experimental results do not show this as the protein in pH 5 is shown to aggregate. That calculation is based on a system that only exists in 2 states, folded and unfolded, therefore the possibility of aggregation of the protein is not taken into account in this calculation. Although these calculations only take into account the electrostatic contribution to stability, excellent agreement was shown between their theoretical results

and experimental data to determine the stability of HmMDH at various pH and salt conditions.

It was concluded that the increased amount of charged residues in the halophilic HmFd are not to provide stabilisation at higher salt concentrations, but simply to prevent aggregation. Evidence was provided for this theory by comparing the salt dependence of halophilic HmFd with a mesophilic homologue at pH 7. It was concluded that both proteins are stabilised with increasing salt concentration, although the effect is more pronounced with the halophilic protein.

1.3.3 Using model halophilic proteins to study the adaptations of halophilic proteins

Tadeo et al. carried out a large scale systematic mutation study, followed by a thermodynamic and structural characterisation on three different proteins: the halophilic DNA ligase, Hv1ALigN, a mesophilic homolgue, Ec1ALigN, and the mesophilic protein L from *Streptococcus magnus* (ProtL)[111].

The first step of the investigation was to examine the effect of salt on the stability of the three wild type proteins. CD and fluorescence experiments showed the halophilic protein, Hv1ALigN, was stabilized by salt (3.2 M KCl or NaCl) while the stability of both mesophilic proteins were independent of ionic strength.

Next the following mutations were carried out including:

i) charge preservation mutations to modify chain length (E to D, DN to EQ, K to R, and R to K – where XY to WZ represents a mutation from X or Y to W or Z)

ii) size preserving mutations to modify protein charge (NQ to DE and DE to NQ)

iii) mutations that modify both size and charge (K to QES and DES to K).

The change in free energy between the unfolded and the folded state of the protein in 3.2 M salt, $\Delta G^{\text{U-F}}_{3.2 \text{ M salt}}$, was determined by equilibrium denaturation experiments. This was carried out, for each variation of each protein. For example Hv1ALigN ExnD,

$\Delta G^{\text{U-F}}_{3.2\text{ M salt}}$ was calculated for each mutation, i.e. Hv1ALigN ExD where only one glutamic acid has been mutated to an aspartic acid, Hv1ALigN Ex2D where two glutamic acids has been mutated to aspartic acids up to Hv1ALigN Ex8D. $\Delta G^{\text{U-F}}_{3.2\text{ M salt}}$ was then plotted in a graph against the number of substituted residues. An increase in the value of $\Delta G^{\text{U-F}}_{3.2\text{ M salt}}$ with the number of substituted residues (i.e. a positive slope) represents a protein becoming increasingly salt stabilised.

Mutation type i) – Reducing one or more amino acids in the protein sequence by one methylene group (E and Q to D and N) introduced a salt stabilising effect for the mesophilic proteins and improved the halophilic adaptation of the halophilic protein. This was shown by an increase in $\Delta G^{\text{U-F}}_{3.2\text{ M salt}}$ with the number of substituted residues, the black circles in figures 1.21 A0 to C). The opposite mutation, increasing one or more amino acids by one methylene group (D and N to E and Q), was found to have reverse effect, see the white squares in figures 1.21 A) to C).

Mutation type ii) – No changes were observed in the salt stabilisation values for the modification of charge in the halophilic protein (figure 1.21 D) black circles) and only a small increase in the halophilic character was observed for the mesophilic proteins (figure 1.21 E) and F) black circles). The results showed charge modification has little effect on the modulation of the halophilicity of the three proteins. As a check, the solubility of wild type Prot L was recorded along with that of Prot L with 4 lysines mutated to glutamines (Kx4Q) and ProtL with 4 lysines mutated to glutamic acids (Kx4E). A dramatic increase in solubility of the ProtL Kx4E mutant protein was found compared to the solubility of the wild type Prot L, shown in table 1.2. The results suggested that acidic residues are favoured in halophilic proteins due increase the solubility of protein in an environment of reduced water activity.

ProtL variant	Solubility (mg/mL)
WT	50 ± 8
Kx4Q	56 ± 7
Kx4E	$>85 \pm 8$

Table 1.2: The measured solubility of ProtL variants in 20 mM phosphate, 2M ammonium sulphate at pH 6.0, and 25 °C[111].
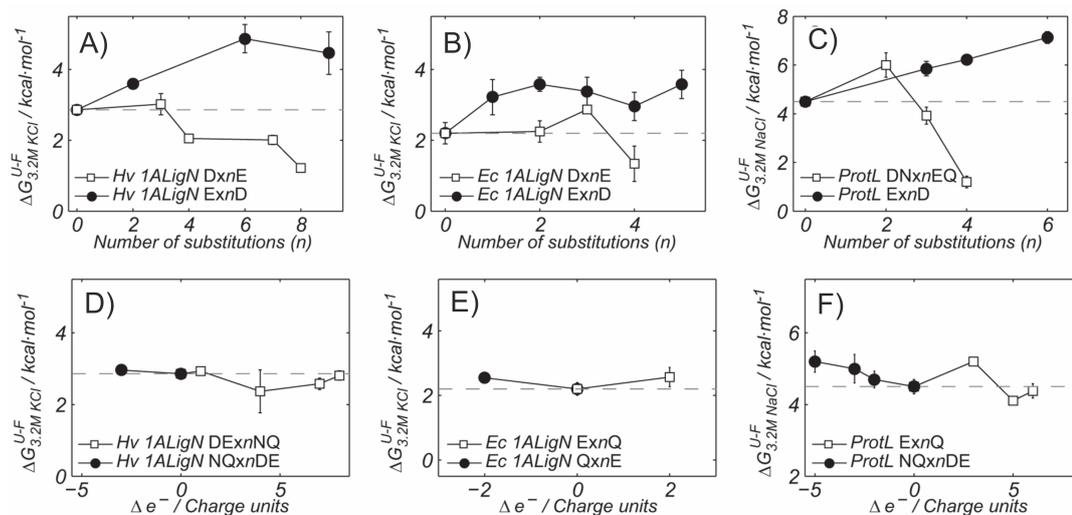
Figure 1.21: The effect of charge and side chain length in protein haloadaptation. The effect of the chain size (figures A) to C)) and charge (figures D) to F)) in haloadaptation of ProtL, Hv1ALigN, and Ec1ALigN was investigated by measuring the free energy at 3.2 M salt (KCl or NaCl) versus the number of substituted residues in the multiple mutations (n in XYxnWZ). Δe is defined as the residual theoretical charge upon mutation (mutant minus wild type). Error bars result from propagation of the experimental uncertainties in the T_m values, by Monte–carlo analysis. Dashed lines represent the $\Delta G_{3.2 \text{ M salt}}^{\text{U-F}}$ values for wild type proteins. Figures B) and C) show the introduced a salt stabilising effect for the mesophilic proteins and figure A) shows the improved the halophilic adaptation of the halophilic protein. Figures D) to F) show the charge modification had little effect on the value of $\Delta G_{3.2 \text{ M salt}}^{\text{U-F}}$ for each mutation for all the three proteins. Figure taken from [111].

Mutation type iii) – Changing K for R, Q, S, or E in ProtL leads to salt stabilization for all replacements, independent of their charge, shown as white shapes in figure 1.22 A). ProtL was shown to be destabilized by NaCl if a sufficient number of lysines were introduced (white squares in figure 1.22 B)). When three lysines are introduced into the surface of the halophilic protein, a decrease in the stability was observed at high salt concentration (black circles in figure 1.22) It was suggested that the low prevalence of lysine residues in a halophilic protein can be explained by destabilising effect of long lysine side chain at high salt concentration.
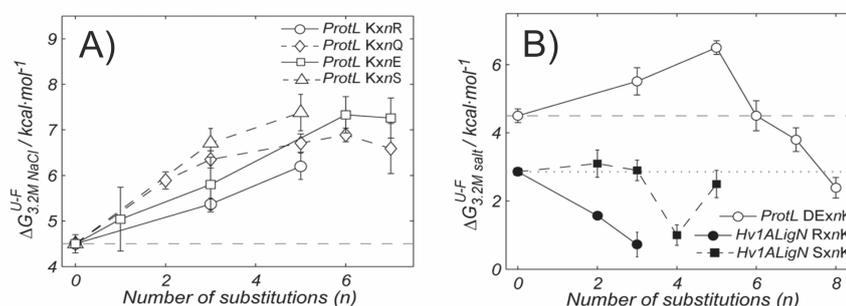
Figure 1.22: The low prevalence of lysines in the amino acid composition of halophilic proteins. The effect of lysines on protein haloadaptation was investigated by measuring the change in stability induced by salt ($\Delta G^{\text{U-F}}_{3.2 \text{ M salt}}$) versus the number of substituted residues in the lysine involving mutations (n in XYxnWZ). Error bars result from propagation of the experimental uncertainties in the T_m values. Dashed and dotted lines represent the $\Delta G^{\text{U-F}}_{3.2 \text{ M salt}}$ for wild type ProtL, Hv1ALigN, and Ec1ALigN. Figure A) shows the increased salt stabilisation when changing K for R, Q, S, or E. Figure B) shows the decrease in stabilisation on the halophilic protein when three lysines are introduced. Figure taken from [111].

To study the effect of changes in surface area, the NMR structures of two mutants of protein L, ProtL Kx5Q and ProtLKx6E, were determined. ProtL Kx6E was found to be an obligate halophile (i.e. unfolded at low salt concentrations), the structure of ProtL Kx6E is shown in figure 1.23. The structures of the mutants were shown to be very similar to the wild type. The data was used to calculate the changes in the accessible solvent area in the folded protein upon mutation. High correlations were found between the slope of the salt stabilisation of the protein and the slope of the decrease in the accessible surface area. The authors suggest that surface packing may, therefore, be important for halophilicity. Proteins with a decreased solvent accessible surface are

preferred in an environment where water molecules also have to solvate the ions.
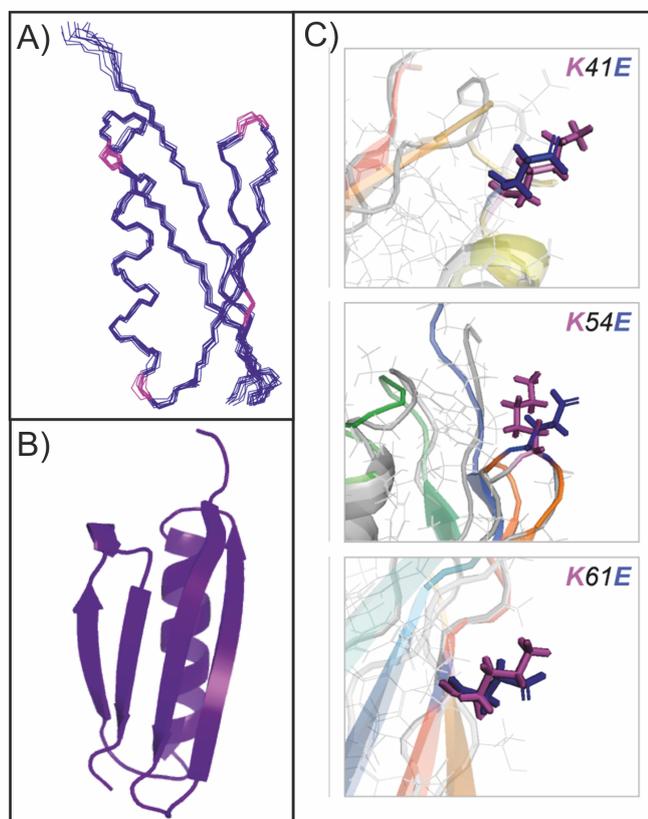
Figure 1.23: The 3D structure of ProtL Kx6E, PDB code: 2KAC. Figure A) shows the 10 lowest energy–refined conformers. Figure B) shows the 3D cartoon ribbon structure of the protein. Figure C) shows close–up views for a representative selection of ProtL WT and ProtL Kx6E side chains that have been aligned. Lysine side chains are coloured in magenta whereas glutamic acid side chains are coloured in blue. Figure taken and edited from [111].

.

The stability of these ProtL mutants was determined in the absence of salt by chemical denaturation experiments. A reduction in the solvent accessible area showed a progressive destabilisation of the protein, which was suggested to be due to a reduction in the protein's hydrophobicity. As a consequence, mutations increasing salt stabilisation also cause a destabilisation of the protein in the absence of salt. This effect would convert a mesophilic protein into an obligate halophile.

A later study by the same group examined the dynamics of water molecules around the halophilic model protein, ProtL Kx6E, to determine whether the introduction of charged residues into a protein gives rise to 'exceptional' hydration, which has been reported for a number of halophilic proteins [112]. Using frequency– and temperature–dependent ^{17}O magnetic relaxation, the study found dynamics of the water molecules of ProtL Kx6E do not differ from the wild type protein L or from other previously studied mesophilic proteins.

Recently, the same group carried out a study in attempt to emulate the process of evolution of halophilic proteins using a range of natural and designed model halophilic polypeptides [69]. They used NMR in combination with other biophysical techniques to separate and quantify the energetic contributions of each residue to the overall haloadaptation. They pointed out a refined model for haloadaptation was required to describe quantitatively the contribution of specific amino acids to the stabilisation of the protein in both unfolded and native states.

The results showed:

i) The native state is stabilized by a number of weak electrostatic interactions between cations in the solute and the negatively–charged residues but not between anions in the solute and positively charged Lys residues.

ii) No such ion–protein interactions were shown by the unfolded protein, which resulted in exclusion of ions from the hydration layer of the protein.

In the absense of ions, only intramolecular electrostatic interactions in the native state are important for differentiating between mesophilic and halophilic proteins. The free energies of their unfolded states can be assumed to be similar. The NMR relaxation data collected also showed that the dynamics were very similar. In a high salt concentration, the native state of a halophilic protein is stabilized by a number of intermolecular electrostatic interactions, causing inclusion of the ions into the proteins hydration layer. In a high salt concentration, the unfolded state of a halophilic protein is destabilized due to its inability to form such electrostatic interactions as the carboxylate groups lack the preorientation required to interact. The ions, therefore, compete with the protein for

hydration. This competition is more exaggerated for halophilic proteins with their large amount of hydrophilic negatively charged residues. This creates an energetic penalty on the unfolded state that increases with the salt concentration, which is larger for halophilic than for mesophilic proteins.

1.3.4 Summary of the literature review on halophilic proteins

To be able to function in such high salt conditions, a halophilic protein needs to be soluble in salting out conditions and needs to have sufficient flexibility to function. A decrease in hydrophobicity is commonly found in halophilic proteins. It is likely this is to counteract the increased hydrophobicity from high salt, shown by the studies on HvDHFR[95]. A number of studies, not described above, have discussed the importance of the decrease in hydrophobicity for haloadaptation[113,114].

The increase in amount of acidic residues is the most common method reported for the halophilic adaptation of proteins. The role of these acidic residues in the stabilisation of proteins in high salt environments is not yet clear. Three separate theories for the role of these acidic residues have emerged:

i) The large number of acidic residues in halophilic proteins are destabilising, i.e. the charges repel and favour the unfolded state of a protein. Salt ions interact with the acidic residues, screening the repulsion from like charges and increasing the stability[93,109]. This theory is supported by experiments on a number of halophilic proteins which show that lowering pH can stabilise a halophilic protein when in low salt conditions[93,104]. The charges are instead screened by the increased amount of hydrogen ions. The presence of salt in solution must, therefore, have a significant effect on the stability of the protein through some form of interaction with the charged residues, i.e. they cannot be stabilised by interaction with water alone.

ii) The halophilic protein stabilisation is described through the coordination of hydrated salt ions, although structural evidence of this also has not been found[79,80]. X–ray crystallography studies seem to show fewer interacting salt ions than suggested by this theory. This led to a modification of the model where salt ions

are needed to bind to only a few specific sites on the protein to allow stabilisation[85]. The destabilising effect brought about by the repulsion of the acidic residues increases the flexibility of the protein, again counteracting the increased hydrophobicity from high salt. This model fits with the structural data, X–ray crystal structures of halophilic proteins do not appear to have many bound salt ions. The model does not account for the lowering of pH stabilising a halophilic protein when in low salt conditions. It is possible that specific ion binding sites are important for specific halophilic proteins, e.g. *Hm*MDH, which shows evidence of salt ions binding through complex salt bridges.

iii) This model describes cumulative weak interactions between hydrated cations and negative residues of the protein in the native state of a halophilic protein, in line with the previous halophilic model by Zaccai el al[79]. But they show these weak interactions only explain one part of the picture; the other driving force behind haloadaptation is the simultaneous destabilization of the unfolded state[69]. Evidence of the interaction of hydrated cations with the carboxylic group of the negative residues is shown using a novel NMR methodology.

1.4 Probing protein stability and dynamics using Single Molecule Force Spectroscopy (SMFS)

The haloadaptation of a protein, discussed in section 1.3, is very complex and has many contributing factors. A tool is needed to decipher the importance of specific interactions that contribute to to the stability of a halophilic protein. One such approach is Single Molecule Force Spectroscopy (SMFS) which is a technique used to investigate the mechanical stability of proteins. Protein unfolding is typically studied on bulk systems using a range of denaturants e.g. chemically or thermally. Measurements using SMFS allow examination of the range of states that may be experienced by a single protein, e.g. subpopulations of proteins that unfold by different pathways and rare unfolding intermediates, which would be overlooked in bulk experiments.

This section will describe a few examples of how force is utilised in nature. The use of polyproteins for the collection of SMFS data will then be explained, along with the use of SMFS data to extract details about the energy landscape of a protein. Factors affecting the mechanical stability of a protein will then be introduced, giving descriptions of studies carried out to investigate each example. The section will finish with a discussion on the importance of the electrostatic interactions on the stability of a protein, including experiments performed to date using SMFS.

1.4.1 Force in nature

Mechanical force commonly occurs in biological systems and is utilised in many activities. A number of processes require the forced unfolding of proteins[115]. One example of the use of mechanical force in nature is the translocation of proteins across a membrane[116–118]. If a protein is initially folded, mechanical force may be used to unfold the protein to allow translocation, shown in figure 1.24 A. The details of this unfolding mechanism are still not fully understood but two possible models have been proposed[119,120]. The first is the Brownian ratchet mechanism, which assumes that the protein chain randomly slides through the channel. The protein binds to an intramitochondrial protein, the chaperone mtHsp70, which traps any unwanted motion in the peptide chain. The second is the power stroke mechanism, which proposes that mtHsp70 is used to pull the protein through the channel by utilising a conformational change upon ATP hydrolysis.

Mechanical force is also used in cell signalling[121,122]. By changing their conformation and interactions, proteins can send signals. Cells are able to sense and respond to these forces. The large muscular protein, titin, contains a protein kinase domain at its C–terminal end, which has been shown to be activated using mechanical force[123].

The function of a number proteins is to withstand mechanical force, e.g. fibronectin, which is found in the extracellular matrix[124,125]. Fibronectin fibrils can stretch under the influence of cell traction forces, allowing for better cell migration and adhesion. Another example is titin, a large fibrous protein found in the muscle[126]. These proteins possess elastic properties to act as shock absorbers to prevent damage of skeletal and
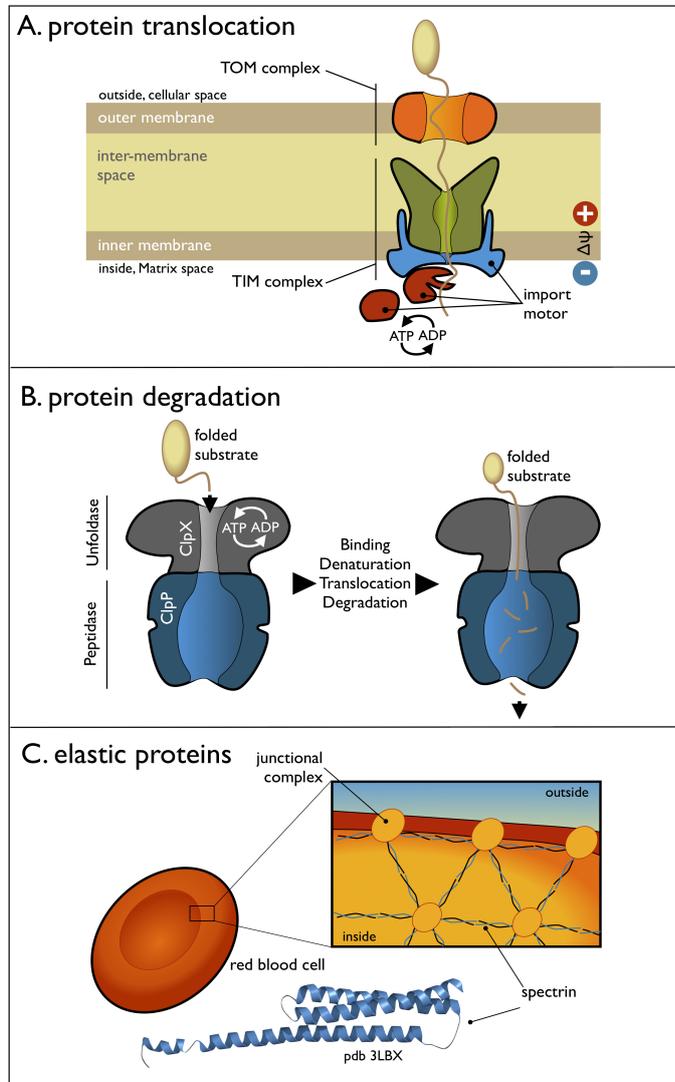
Figure 1.24: Schematics of examples of force used in nature. Figure A. shows the translocation of a protein into a mitochondrion through the membrane. Figure B. shows the chaperone, ClpX, which is involved in protein degradation to remove unwanted or misfolded proteins. Figure C. shows an example of protein that functions due to its capacity to withstand mechanical force, the α–helical protein spectrin. It makes up the cytoskeleton of red blood cells to form an elastic mesh to withstand external forces. Figure taken and edited from [115].

cardiac muscle sarcomeres.

1.4.2   Introduction to SMFS experiments

The Atomic Force Microscope (AFM) was first invented in 1986[127]. It is an instrument that was designed to probe the interaction forces between a sharp tip and the surface of a sample with subnanometre resolution. The AFM has been mostly utilised in biophysics for the imaging of biomolecules[128,129]. The use of the AFM was developed to measure the unfolding force of a protein a decade after the invention[130].

The AFM uses a sharp tip at the end of a flexible cantilever to probe the unfolding properties of a protein, shown in figure 1.25. The protein solution is applied to a freshly cleaved gold surface, the height of which can be controlled using a piezo. During an SMFS experiment, the gold surface is brought into contact with the tip. At this point a protein may be adsorbed on the cantilever tip. An external force is applied across the protein by increasing the distance between two points, stretching the protein into an unfolded, extended state. The displacement of the cantilever, with respect to the extension of the protein, is measured using a photodetector. Further details of an SFMS experiment can be found in section 2.7.3. The technique allows individual proteins to be unfolded and refolded in order to study mechanical stability, folding pathways and features of the transition states.

1.4.3   Advantages of the use of poly–proteins for SMFS studies

Polyproteins are comprised of a single polypeptide chain of repeating single protein domains. Naturally occurring polyproteins, such as titin, were the first to be analysed using SMFS[131]. Natural polyproteins are, however heterogenous[132]. Each of the protein domains are similar but are not identical, therefore, their mechanical properties may differ. The development of techniques to engineer artificial polyproteins has made it possible to examine repeating domains of proteins, where each domain is identical. Thus allowing unambiguous identification of the mechanical properties of the protein.
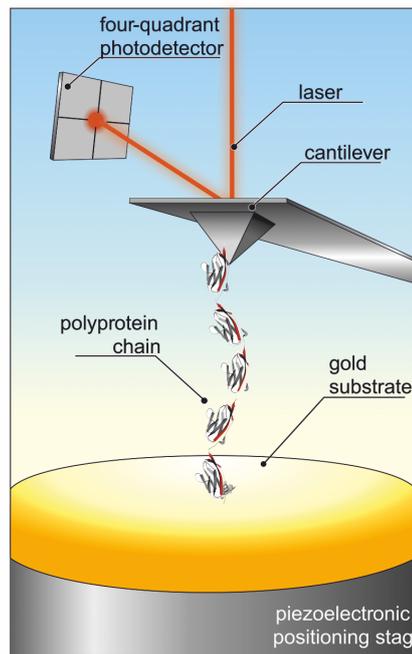
Figure 1.25: Overview of the the SMFS experiment. Protein solution is applied to a freshly cleaved gold surface, the height of which is controlled using a piezo. The protein is adsorbed on the cantilever tip and an external force is applied across the protein. The displacement of the cantilever, with respect to the extension of the protein, is measured using a photodetector. Figure taken and edited from [115].

During an SMFS experiment the cantilever tip comes into contact with a gold surface, which contains many protein constructs. A number of these constructs could interact at any point along each protein. Rarely will the tip interact with the end of a folded protein construct. A resulting force is exerted on the tip by these 'non–specific' interactions, which typically occurs when the tip is close to the substrate. Non–specific interactions produce a large amount of 'noise' which may mask the force peaks from proteins with small force peaks. They also may be difficult to distinguish these peaks from *bona fide* interactions with the protein of interest. Engineered poly–proteins contain multiple identical tandem repeats, which produce periodical sawtooth pattern force–extension curves. The repetitive pattern creates a "fingerprint" that allows unambiguous quantification of the mechanical properties of the protein of interest.

1.4.4 What determines the mechanical stability of a protein?

One of the main aims of SMFS studies is to understand the protein structure characteristics, e.g. topology, hydrogen bonding, hydrophobic core, that determine the mechanical properties of a protein[133–135].

1.4.4.1 The secondary structure of a protein is highly important for mechanical stability

A pioneering SMFS study was carried out in 1997 on the muscle protein titin[131]. Titin was found to have a high mechanical stability and could withstand high forces (150-350 pN). A computational study by Lu et al. used molecular dynamics simulations to stretch the single titin Ig domains[136]. The results showed the peak unfolding force was due to the shearing of hydrogen bonds between two sets of β–strands. The region of importance for mechanical force in a protein structure has since been referred to as the 'mechanical clamp'[137,138]. A number of mechanical clamps have since been identified in many other proteins and ranked according to their unfolding force[133,135].

An all–α–helical protein, calmodulin, was found to have a low mechanical unfolding

force of 15 pN at a pulling velocity of 600 nms^{-1}. An α–helix has a homogeneous distribution of intrahelix hydrogen bonds. Molecular dynamics simulations predicted that the intra–helix hydrogen bonds of an α–helix were broken sequentially, a process that required little force[139]. Other all α–helical proteins have shown a similar behaviour[140,141] (see figure 1.26).
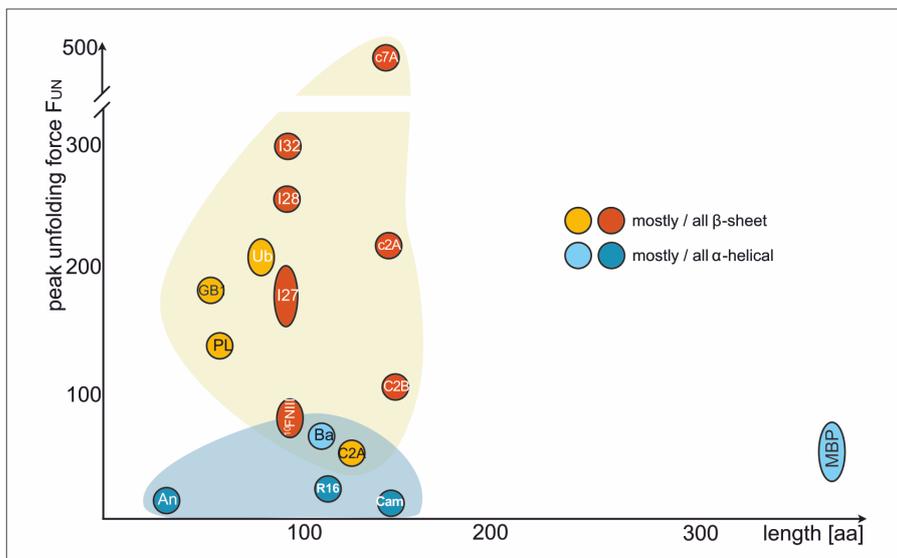
Figure 1.26: A selection of proteins that have been mechanically characterised, comparing unfolding force, structure and extension. The approximate unfolding force (at pulling velocities of about 500 ± 200 nms^1) is plotted against the length of the protein. Structural information is colour–coded; mostly/all β–sheet is shown in orange/red respectively, mostly/all α–helical is shown in light/dark blue respectively. Figure taken and edited from [115].

1.4.4.2 The direction of the applied force affects the mechanical stability of a protein

The pulling direction has also been shown to affect the mechanical properties of proteins[142]. A study by Brockwell at al. used protein immobilisation to probe the mechanical properties of the protein, E2lip3, in two different pulling directions[143]. When the protein was immobilised between residues 1 and 41, the hydrogen bonds between its β–strands were sheared with a mechanical unfolding force of 177 ± 3 pN at a pulling velocity of 700 nm/s, shown in figure 1.27 A). When the protein was immobilised between

residues 1 and 75, the hydrogen bonds between its β–strands were broken one by one and E2lip3 was shown to unfold without showing any mechanical resistance, shown in figure 1.27 B). The results indicated that the mechanical properties of a protein could not simply be defined by amino acid sequence, topology or kinetic stability.
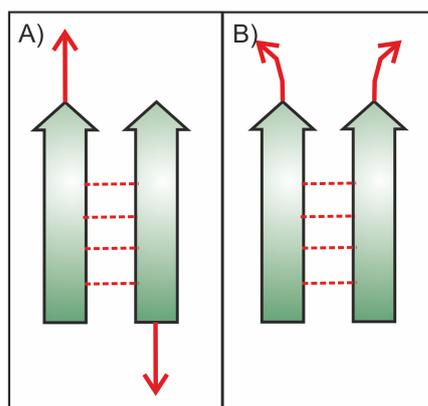
Figure 1.27: A schematic of a change of pulling direction on two β–strands. Figure A) shows the shearing of the hydrogen bonds (depicted as red dashed lines) between two β–strands. Figure B) shows the peeling of the hydrogen bonds between two β–strands.

A later study by Dietz et al. measured the mechanical properties of a folded protein structure when extended in defined directions[144]. This was possible by adding two cysteine residues at precise points in the GFP amino acid sequence, to create di–sulphide linkages and form a polyprotein. The positions of the cysteines in the protein define where the force is applied to the structure. The results from 5 chosen pulling directions showed a wide range of unfolding forces from 100 pN to 600 pN, see figure 1.28. It was concluded that the GFP structure must contain two unfolding pathways of similar stability while pulling in one direction. The study also identified an intermediate structure that was pulling direction–dependant.

1.4.5 The use of chimeric polyproteins to study mechanically uncharacterised proteins

There are two types of polyprotein used for SFMS experiments:
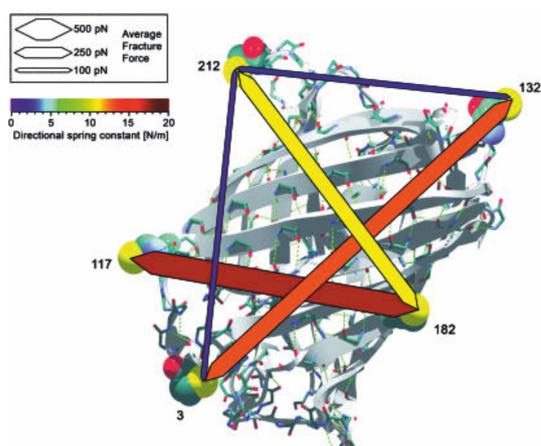
Figure 1.28: Directional deformation response of the GFP fold. The width of the lines, shown connecting the points of force application, represents the average unfolding force in that particular direction. The colour of the lines show the directional spring constant, i.e., protein rigidity in the respective direction. The colours and widths are described in the key. Figure taken and edited from [144].

   i) Homo-polyproteins, containing tandem repeats of one protein

  ii) Hetero-polyproteins, containing repeats of two types of protein.

Hetero-polyproteins may contain a protein that has been previously well studied using SMFS. The advantage of this is the identification of a well defined fingerprint, a marker. Previously unstudied proteins may be added to the polyprotein to allow identification of their mechanical properties. A number of proteins have now been well characterised and which could be used as marker proteins, such as I27.

   I27 is the 27th Ig domain from the I–band region of the human cardiac muscle protein, titin. It is regularly used in hetero–polyprotein constructs for SMFS to provide an easily identifiable fingerprint[145–148]. I27 has a known structure, has a high mechanical stability and has been previously well characterised using SMFS with a simple unfolding pathway[132,134,136,147,149–154]. The protein is 89 amino acids in length and has a β–sandwich topology consisting of 7 β–strands that are folded into 2 β–sheets. SMFS experiments combined with molecular dynamics simulations have been used to identify the unfolding pathway of the protein[136,155,156]. Initially, the A strand (shown in figure 1.29) breaks away from the rest of the protein causing a small deformation in the unfolding peak of the sawtooth pattern. Subsequently, the hydrogen bonds between the A

and G β–strands rupture and the rest of the protein unfolds with little resistance. The A–G hydrogen bond breakage has the highest energy barrier and is therefore the rate limiting step of the unfolding process.
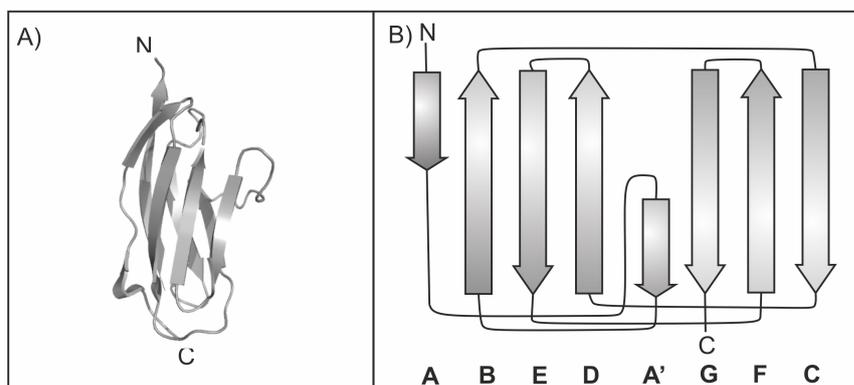
Figure 1.29: The structure of I27. A) shows the 3D topology of the protein. This figure was created using PDB file 1TIT and Pymol[8,90]. B) shows the secondary structure of the protein where the arrows represent β–sheets.

A range of proteins have now been characterised using the chimeric polyprotein approach[29,145,148,157–161]. Chimeric polyproteins have been utilised for the study of the intrinsically disordered protein, human α–Synuclein[157]. SMFS experiments were shown to be able to differentiate between three different classes of conformations in an intrinsically disordered domain including random coil, mechanically weakly folded and β– like by adding specific agents to cause an imbalance in the conformational equilibrium. A featureless mechanical unfolding fingerprint was assigned to the stretching of αSyn moiety having a random coil conformation. The fingerprint signature of mechanically weak confirmations showed single or multiple small peaks. Finally unfolding curves, which showed a worm–like chain behaviour, were observed for β–like conformations. A study by Hoffmann et al. provides a methodology for the use of Gibson assembly cloning to generate a recombinant polyprotein rapidly[161]. The technique was exploited to facilitate the mechanical characterization of the second polypeptide transport associated domain from BamA of *E. coli* (*Ec*POTRA2) by assembling the chimeric polyprotein (I27-*Ec*POTRA2)$_3$-I27.

1.4.6 The use of SMFS to study specific interactions

SMFS experiments have provided insight into the impact of specific structural interactions on the mechanical stability of a protein. The technique has been shown to be a very useful tool to study the importance of hydrogen bonding for protein mechanical stability, many examples are described in section 1.4.4[132,135,136,143,151,162].

The role of solvent hydrogen bonds on the mechanical stability of I27 has been studied by Dougan at al.[151]. The strength of hydrogen bonds with the solvent was controlled by the substitution of water (H_2O) for deuterium oxide (D_2O). SMFS experiments found that D_2O was a worse solvent for protein I27 than H_2O, where a bad solvent is one which favours chain–chain interactions over chain–solvent reactions, promoting chain contraction and a good solvent is one which favours chain–solvent interactions, promoting chain expansion[163]. D_2O was shown to be a worse solvent for protein I27 by the increase in the unfolding force of the protein in D_2O[151]. They reported that the hydrophobic interactions of I27 are enhanced in D_2O, which caused an increase in mechanical strength of the protein. The results of this study provided the first single–molecule level measurement of the influence of D_2O on the hydrophobic effect during protein folding. These results are in contrast with the behaviour of simple hydrocarbons in D_2O, which suggests D_2O as a better solvent due to its larger compressibility[164]. Proteins are complex molecules that contain both hydrophobic and hydrophilic residues, therefore, the differing result is not surprising. The study showed that even though the increase in hydrogen bond strength of the solvent, upon substitution with D_2O, is small, SFMS experiments were able to differentiate the changes.

1.4.6.1 The effect of electrostatic interactions on the mechanical stability of a protein

Only a limited number of studies have been carried out to investigate the role of other interactions on the mechanical properties of a protein. The first evidence of the investigation of electrostatic effects of the mechanical stability of a protein was shown in a

study by Chyan et al[165]. Polymers of the protein ubiquitin were synthesised and characterised using SMFS in a range of PBS buffers with adjusted pH, ranging from 2–10. It was found that in a pH range (around 6-10) the unfolding force of the protein was consistent, 230–240 pN, see figure 1.30. As the pH of the solution was lowered below 6, the force required to unfold the protein was reduced. Unfolding events were not observed at a pH above 10, which was suggested to be due to aggregation of the polymers. Chyan et al. suggested that the charge–charge interactions in the native state of the protein have a stabilising effect on the protein at neutral pH but become destabilising when the pH is lowered. Although the study by Chyan et al showed the first investigation into effect of electrostatics on the stability of a protein, the results reported the average effect of all charged residues and no information on the contribution of individual residues was obtained.
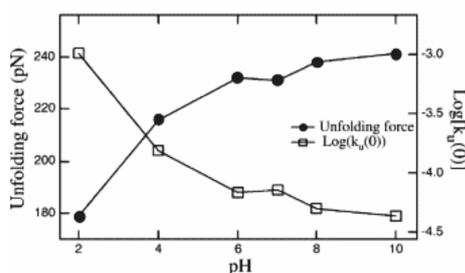
Figure 1.30: The unfolding force and the unfolding rate at zero force of Ubiquitin as a function of pH. Each point in the force curve is the av. unfolding force of all points measured from the measured force–extension traces. The zero force unfolding rate data were obtained from Monte Carlo simulation, by fitting the simulated data to the experimental data at a specific pH value. Figure taken and edited from [165]

A molecular dynamics (MD) study by Craig et al. on the unfolding of the fibronectin, 10FnIII, predicted that the mechanical stability of the protein could be tuned by the protonation of a few amino acids in the force bearing β–strands[166]. An increase in the unfolding force was predicted by lowering the pH from 7 to 4.7 due to protonation of three carbonyl amino acid side chains. The MD simulations showed that two of the protonated carbonyl groups formed a new hydrogen bond between their side chains. A later experimental investigation, to test the prediction from Craig et al. found that decreasing the pH did not cause a significant increase the mechanical stability of the protein, suggesting that some simulations may overestimate the contribution of electrostatic interactions to

the mechanical stability of proteins[167]. It should be noted, however, that the protein 10FnIII contains a number of charged residues, not just the three investigated. Varying the pH of the solvent will affect all of these charged residues. The hypothesis of the MD simulation study was, therefore, not directly examined by experiment.

In order to measure the effect of electrostatic interactions on the mechanical stability directly, a later study carried out by Zheng et al. used protein GB1, which was previously shown to be thermodynamically stable in a wide range of pH[168]. A bi–histidine motif was engineered into the force–bearing region of GB1. By varying the pH of the solvent from 8.5 to 3, the histidine residues could switch between protonated and deprotonated states. This was used to modify the electrostatic interactions between the two histidine residues. The results showed the measured unfolding force of the engineered protein decreased by 34% by decreasing the pH from 8.5 to 3. It was suggested this was due to the increased electrostatic repulsion between the two positively charged histidines at acidic pH.

A recent study by Tych et al. examined the impact of salt bridges on the mechanical properties of of three cold shock proteins; one from the mesophilic organism *Bacillus subtilis* (*Bs*CSP), one from the hyperthermophilic organism *Thermotoga maritima* (*Tm*CSP) and a mutated variation of *Bs*CSP, named the charge triple mutant (CTM)[29]. MD simulations revealed that *Tm*CSP contains a larger number of salt bridges than *Bs*CSP. In order to determine if a larger number salt bridges in CSP affect the mechanical properties of the a protein, a charge triple mutant of *Bs*CSP was produced by inserting an ionic cluster from *Tm*CSP into the *Bs*CSP structure. The CTM protein was found to be mechanically softer than *Bs*CSP, i.e. the distance between the folded state and the TS barrier, Δx_U, was larger. Δx_U is a measure of the deformation of the native state protein before unfolding. This increase in the mechanical softness for the CTM showed the protein closely mimicked the mechanical properties of *Tm*CSP.

1.4.7 Extracting the energy landscape of a protein from SMFS experiments

The use of SFMS experiments in combination with Monte Carlo techniques can be used to examine properties of the unfolding transition state of the protein under applied force[29,148,152,161]. As mechanical unfolding is a kinetic process, the rate at which force is applied to the protein affects its mechanical unfolding force. At a faster pulling velocity, the unfolding force of a protein will be higher as thermally activated unfolding will be less likely per unit time. By measuring the speed dependence of the unfolding force in combination with Monte Carlo techniques it is possible to extract the distance from the unfolded to the transitions state, Δx_U, (or as described by Tych et al. the mechanical softness) and the unfolding rate at zero force, k_U[29]. The parameters of the energy landscape are shown in figure 1.31.
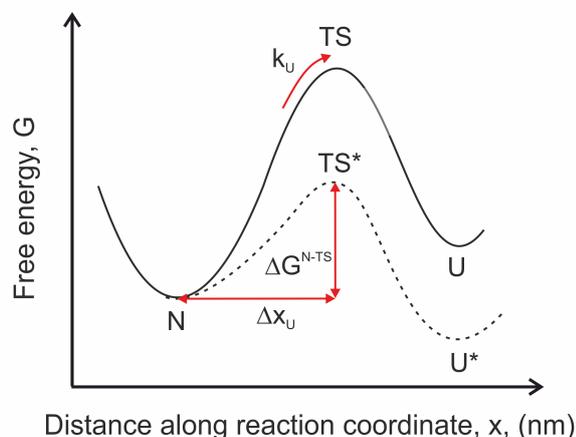
Figure 1.31: Energy landscape of a protein under applied force. The figure shows a schematic of a two-state protein free energy profile (black line) going from it native state, N, to its unfolded state, U, through the transition state, TS, at an unfolding rate, k_U. The dashed line shows the tiled energy landscape under application of force, lowering the energy barrier, ΔG^{N-TS}, to the unfolded state, U*. The distance from the unfolded to the transitions state is shown, Δx_U.

1.5 Discussion and aims

The aim of this project is to use protein engineering to produce a polyprotein containing a halophilic protein and to carry out SMFS experiments using the AFM to characterise the mechanical unfolding behaviour of a halophilic protein. The mechanical stability of a protein, which is determined in SMFS experiments, is derived from the kinetic description of the protein. It is not correlated with thermodynamic stability and hence an unfolding event only depends on the unfolding activation energy [147]. SMFS results will provide insight into the kinetic stability of a halophilic and mesophilic protein with increasing salt concentration, as well as the underlying energy landscape.

Since a unique feature of halophilic adaptation is an increase in amount of acidic residues, halophilic proteins allow the study of the contribution of electrostatic interactions to the mechanical stability of a protein. Only a limited number of studies are currently available that have investigated the role of the electrostatic contribution to mechanical stability, therefore the importance of electrostatic interactions is still under discussion.

Chapter 2

Materials and methods

This chapter gives details of the materials used and the experimental procedures carried out in this thesis including molecular biology, protein expression and purification, Single Molecule Force Spectroscopy (SMFS) experiments and biophysical characterisation. Other procedures specific to individual experiments are described in the appropriate results chapters (chapters 3,4, 5).

2.1 Materials

2.1.1 Chemicals

A list of chemicals used is found in table 2.1.

Item	Supplier
Bacterial growth medium component	
Tryptone, granulated	Fisher Scientific
Yeast extract, granulated	Fisher Scientific
Agar, granulated	Fisher Scientific
LB Broth Miller, granulated	VWR

N–Z–amine (casein hydrolysate)	Sigma
Ready mixed Auto–induction medium	Formedium
Ampicillin disodium salt	Formedium
Carbenicillin	Formedium
Chloroamphenicol	Formedium
IPTG, dioxane free	Formedium
Ammonium iron(II) sulphate hexahydrate	Sigma

Molecular biology

Deoxynucleotide solution mix	New England Biolabs (NEB)
Ethylenediaminetetraacetic acid	Acros organics
Acetic acid	Sigma
Bromophenol blue	Sigma
Xylene cyanol	Sigma
Sucrose	Fisher Scientific
Restriction enzymes	New England Biolabs (NEB)
Bovine serum albumin	New England Biolabs (NEB)
Antartic phosphatase	New England Biolabs (NEB)
Antartic phosphatase buffer	New England Biolabs (NEB)
T4 DNA ligase	New England Biolabs (NEB)
T4 DNA ligase buffer	New England Biolabs (NEB)

Protein electrophoresis

Tris–Tricine SDS buffer 10X (cathode)	Bio–Rad
Acrylamide 30% (w/v)	Severn Biotech
Presicion Plus Dual Xtra Protein Marker	Bio–Rad
Instant Blue Stain	Expedeon
TEMED	Sigma
Ammonium persulphate	Sigma
SDS	Fisher

Agarose gel electrophoresis

Agarose	Melford
DNA ladders	New England Biolabs (NEB)

Protein preparation

Bugbuster protein extraction reagent	Novagen
Urea	Sigma
Tris base	Melford
DTT	Melford
Sodium chloride	Fisher

Potassium chloride	Fisher
Imidazole	Fisher
DNaseI	Sigma
RNase A	Sigma
PMSF	Sigma
4–Aminobenzamidine dihydrochloride	Sigma
Protease inibitor cocktail tablets, EDTA free	ThermoFisher Scientific
Ni Sepharose Fast Flow	GE Healthcare Life Sciences
Ammonium bicarbonate	Sigma
AcTEV$^{\text{TM}}$ protease	ThermoFisher Scientific
1,11–bismaleimido–triethyleneglycol (BM(PEG)$_3$)	ThermoFisher Scientific
Miscellaneous	
β–mercaptoethanol	Sigma
Ethanol	Fisher
Hydrochloric acid	Fisher
Glycerol	Fisher
Magnesium sulphate	Sigma
Potassium phosphate monobasic	Sigma
Potassium phosphate dibasic	Sigma

Table 2.1: List of chemicals used and suppliers

2.1.2 Media

The types of media that were used for this study are listed in table 2.2. For agar plates, 15 gL^{-1} agar was added to the liquid medium prior to autoclaving.

2.1.3 Antibiotics

Antibiotics used in this study are listed in table 2.3

Each stock solution was made up using deionised water or ethanol, sterilised by passing through a 0.2 μm filter and stored in 1 mL aliquots at -20°C.

Medium	Ingredients	Amount
Luria–Bertani (LB) medium	Ready mixed LB–Broth	25 g
	Purite 18 MΩcm H_2O to	1 L
Terrific Broth (TB) medium	Tryptone	12 g
	Yeast extract	24 g
	Glycerol	4 mL
	Potassium phosphate monobasic	2.31 g
	Potassium phosphate dibasic	12.54 g
	Purite 18 MΩcm H_2O to	1 L
NZY Broth	Sodium chloride	5 g
	Magnesium sulphate	2 g
	Yeast extract	5 g
	NZ amine (casein hydrolysate)	10 g
	Purite 18 MΩcm H_2O to	1 L
Auto–Induction Medium (AIM)	Ready mixed Auto–induction medium	34.85 g
	Purite 18 MΩcm H_2O to	1 L

Table 2.2: Types of media used

Antibiotic	Stock solution	Final concentration used
Ampicillin	100 mg/mL in H_2O	100 μg/mL
Carbenicillin	100 mg/mL in H_2O	100 μg/mL
Chloroamphenicol	100 mg/mL in 100% ethanol	25 μg/mL

Table 2.3: Types of antibiotic used

2.1.4 *E. coli* strains

5 different cell strains were used for this project, shown in table 2.4.

Strain	Supplier	Genotype/Description
Cloning cells		
XL1–Blue	Stratagene	*recA1 endA1 gyrA96 thi–1 hsdR17 supE44 relA1 lac [FproAB lacI^{q}ZΔM15 Tn10 Tet^r)]* XL1–Blue cells are suitable for routine cloning purposes.
SURE2	Stratagene	*e14-(McrA-) Δ(mcrCB–hsdSMR–mrr)171 endA1 supE44 thi–1 gyrA96 relA1 lac recB recJ sbcC umuC::Tn5(Kan$'$) uvrC [FproAB lacI^{q}ZΔM15 Tn10 (Tet^r) Amy Cam^r]* SURE 2 cells are suitable for the cloning of unstable plasmids.
Expression cells		
JM83	Stratagene	*rpsL ara Δ(lac–proAB) Φ80dlacZΔM15* JM83 cells are suitable for transformation and protein expression.
BL21(gold)	Stratagene	*E. coli B F$^-$ ompT hsdS(r_B^- m_B^-) dcm+ Tet^r gal endA Hte* BL21 cells are suitable for transformation and protein expression.
BLR(DE3)PlysS	Stratagene	*E. coli B F$^-$ ompT hsdS(r_B^- m_B^-) dcm+ Tet^r gal λ (DE3) endA Hte [pLysS Cam^r]* BLR cells also help stabilise target plasmids that contain repeats in the sequence. DE3 cells contain T7 Polymerase, therefore they are used to express proteins with a T7 promoter. PLysS produce T7 Lysozyme in order to reduce the amount of protein expression before induction by IPTG.

Table 2.4: Types of *E. coli* cell strains used and their genotypes

2.1.5 Plasmids

All plasmids used in this project are shown in table 2.5.

Vector	Supplier	Properties
pCR2.1/	ThermoFisher Scientific	The pCR2.1 vector has a T7 promotor upstream of the gene sequence, which is specifically recognised by T7 RNA Polymerase (expressed in certain expression cells) and allows transcription of the gene to occur. The plasmid also includes an ampR gene and a kanR which are ampicillin and kanamycin resistance genes.
pET3a/	Stratagene	The pET3a vector has a T7 promotor upstream of the gene sequence. The vector also includes a T7 terminator to end transcription. An N–terminal T7 tag is included to facilitate detection of the target protein e.g. recognised by an antibody for western blotting. The plasmid also includes an ampR gene as mentioned above.
pET15b/	Stratagene	The pET15b plasmid codes an N–terminal His–tag sequence to facilitate purification of the target protein . The vector has a T7 promotor, a T7 terminator and an ampR gene as mention above. The vector also codes for a Lac repressor which reduces protein expression in the absense of IPTG
pMAL–c5x/	NEB	The pMAL–c5x vector includes a malE gene encoding maltose–binding protein (MBP) to express an MBP fusion protein. A tac promoter is coded in the vector, to bind *E. coli* RNA polymerase for transcription. The vector also codes for a Lac repressor. Factor Xa is encoded in the vector, after the malE gene, which allows cleavage of MBP from the target protein. The plasmid also includes an ampR gene.

Table 2.5: Types of protein expression vectors used and their properties

2.1.6 Oligonucleotides

All oligonucleotides were ordered from Eurofins MWG Operon.

2.1.7 Equipment

A list of equipment used is shown in table 2.6.

2.2 Molecular biology methods

2.2.1 Polymerase chain reaction (PCR)

Oligonucleotide primers were designed to modify and/or amplify required gene sequences. The estimated melting temperature of the primers (T_m) was calculated using equation 2.1 where n_{AT} corresponds to the number of Adenine and Thymine base pairs and n_{GC} corresponds to the number of Guanine and Cytosine base pairs.

$$T_m = (n_{AT} \times 2) + (n_{GC} \times 4) \tag{2.1}$$

The composition of a typical PCR experiment is shown in table 2.7. A negative control was also set up that did not include the template DNA. The temperature for a typical PCR reaction is shown in table 2.8 where x is 5°C below the calculated primer T_m.

2.2.2 Agarose gel electrophoresis

The agarose gel was prepared by adding 1-2% (w/v) agarose to Tris-acetate ethylenedi-aminetetraacetic acid (TAE) buffer (40 mM Tris-HCl, 20 mM acetic acid, 1 mM EDTA, pH 8.0) and heating in a microwave for approximately 2 minutes until the agarose was

Item	Supplier	Use
InGenius gel documentation system	Syngene	Imaging SDS–PAGE gels
vcx–130PB Vibra–cell^{TM} Tip sonicator	Sonics & Materials, Inc	Homogenising cell solutions and lysing cells
HAIVA cell disrupter	Constant Systems Ltd	Lysing cells
PowerDry PL 300 Lyophiliser	Heto	Lyophilising protein samples
3020 Bench pH meter	Jenway	Measuring the pH of a solution
PTC–100 PCR thermocycler	Bio–Rad	PCR reactions
Orbital Floor Incubator Shaker	Sanyo Gallenkamp	Large scale culture grows
Avanti J-26 XP high performance centrifuge	Beckman	Harvesting cells on a large scale
Eppendorf 5810R refrigerated bench-top centrifuge	Fisher	Harvesting cells on a small scale
ÄKTAprime protein purification system	GE Healthcare	Running columns to purify proteins
Ultrospec 2100 pro spectrometer	GE Healthcare	DNA, protein and cell concentration measurements
Chirascan^{TM} CD spectrometer	Applied Photophysics	Thermodynamic protein stability measurements

Table 2.6: List of biology equipment used and suppliers

Ingredient	Amount
Template DNA	100 ng
Forward primer	0.5 μM
Reverse primer	0.5 μM
Deoxynucleotide solution mix	200 μM
Magnesium sulphate	2 mM
Vent® DNA polymerase buffer (10x)	1x
Vent® DNA polymerase	1 U
Purite 18 MΩcm H$_2$O to	50 μL

Table 2.7: Composition of a typical PCR reaction

Step	Temperature	Time
Initial denaturation	98°C	5 minutes
Denaturation	98°C	30 seconds
Annealing	x°C	30 seconds
Elongation	72°C	1 minute per kb
Repeat last 3 steps x 25		
Final elongation	72°C	5 minutes

Table 2.8: Temperature cycling program for a typical PCR reaction

completely dissolved. The agarose solution was cooled to 50°C before addition of 5 μL 10 mg/ml ethidium bromide for a final concentration of 0.33 μg/ml. The agarose gel was then poured in a cast, which included a lane comb, and left to set at room temperature. The lane comb was removed from the agarose gel, which was then placed in a running tank. TAE was added onto the tank until the gel was completely submerged. DNA samples and size ladders were prepared by adding 6 x gel loading buffer (0.25% (w/v) bromophenol blue , 0.25% (w/v) xylene cyanol, 40% sucrose in H_2O) and were loaded into the wells. The running tank was connected to a power unit and run at 100 Volts for approximatly 1 hour. The resulting DNA was imaged using UV trans-illumination and DNA fragment sizes were estimated by comparison with 1kbp or 100bp DNA ladders.

2.2.3 Extraction of DNA from agarose gels

DNA was extracted from agarose gels using the QIAquick® gel extraction kit (Qiagen). DNA bands were cut out from the agarose gel using a scalpel and extracted using the protocol of the manufacturer. The DNA was eluted into a clean 1.5 mL eppendorf tube by carefully adding 50 μL Purite 18 MΩcm H_2O to the centre of the column followed by centrifugation at 1000 rpm for 60 s.

2.2.4 Digestion of DNA using restriction enzymes

Restriction digests were set up using the required restriction enzymes. The appropriate buffer for each reaction was chosen to allow at least 75% enzyme activity for each restriction enzyme. The composition of a typical restriction digest reaction is shown in table 2.9 below. In addition, three control reactions were also carried out simultaneously; one containing only restriction enzyme I, one containing only restriction enzyme II and one reaction not containing any restriction enzymes. The reactions were incubated at 37°C for 1 hour.

Ingredient	Amount
Plasmid DNA	3 μg
NEB buffer (10x)	1x
NEB Bovine Serum Albumin (BSA) (100x)	1x
Restriction enzyme I	10 U
Restriction enzyme II	10 U
Purite 18 MΩcm H$_2$O to	25 μL

Table 2.9: Composition of a typical restriction enzyme digestion reaction

2.2.5 Desphosphorylation of DNA

When digested plasmid DNA was used for subsequent ligation reactions, 5' phosphate groups were removed from the vector DNA to prevent re-ligation using the following ingredients listed in table 2.10. The reaction was incubated at 37°C for 15 minutes and subsequently heat inactivated at 65°C for 5 minutes.

Ingredient	Amount
Vector DNA	1-5 μg
Antarctic Phosphatase reaction buffer (10x)	1x
Antarctic Phosphatase	5 U

Table 2.10: Composition of a dephosphorylation reaction

2.2.6 Ligation of DNA

Ligation reactions were carried out using T4 DNA ligase. The composition of a ligase reaction is shown in table 2.11. 50 ng of vector DNA (X) was combined with a 3-fold molar excess of insert DNA (Y). A control reaction was also set up, which did not include any insert DNA. The reactions were incubated at 16°C overnight before cooling on ice.

2.2.7 Transformation of ligations into XL1-Blue supercompetent cells

E. coli Homemade XL1-Blue supercompetent cells were thawed on ice for 10 minutes. 2 μL β-mercaptoethanol was added to 100 μL of cells in a 14 mL round bottom falcon tube and incubated on ice for 10 minutes. 5 μL of ligated DNA solution was added to the

Ingredient	Amount
Vector DNA	X ng
Insert DNA	Y μg
T4 DNA ligase buffer	10x
T4 DNA ligase	1 U
Purite 18 MΩcm H$_2$O	20 μL

Table 2.11: Composition of a ligation reaction

cell suspension and a negative control was also set up containing no DNA; substituting sterile H$_2$O for the DNA ligation components. The solutions were further incubated on ice for 30 minutes and then heat shocked by submerging the tubes in a water bath at 42°C for 30 seconds. The cells were then returned to ice for 2 minutes before addition of 900 μL of sterile NZY medium and incubation at 37°C for 1 hour, with shaking at 200 rpm. After incubation, the cells were pelleted by centrifuging the sample at 1000 rpm for 10 minutes and discarding 800 μL of the medium. The pellet was resuspended in the remaining medium and the cells were spread onto LB agar plates containing 100 μg/mL ampicillin. Plates were incubated overnight at 37°C.

2.2.8   Small scale purification of DNA from *E.coli*

A single colony of cells, containing the plasmid to be purified, was used to inoculate 10 ml LB broth containing the appropriate antibiotic and incubated overnight at 37°C in a shaking incubator at 200 rpm. DNA was isolated from the culture using the QIAprep® Spin Miniprep Kit (Qiagen) by following the protocol of the manufacturer. The DNA was eluted by using Purite 18 MΩcm H$_2$O.

To determine the DNA concentration of the eluted sample, the A$_{260}$ was measured using an Amersham Biosciences Ultrospec 2100 Pro UV/Visible Spectrophotomer. An A$_{260}$ value of 1 is equal to 50 μg/mL of double stranded DNA.

2.2.9 Large scale purification of DNA from *E.coli*

A single colony, containing the plasmid to be purified, was used to inoculate 50 ml LB broth containing the appropriate antibiotic and incubated overnight at 37°C in a shaking incubator at 200 rpm. DNA was isolated from the culture using the QIAprep® Spin Midiprep Kit (Qiagen) by following the protocol of the manufacturer. The DNA was eluted by using Purite 18 MΩcmcm H_2O. The DNA concentration was calculated as outlined in section §2.2.8.

2.2.10 DNA sequencing

DNA was sent to Beckman Coulter Genomics for sequencing, which was used to confirm the success of cloning.

2.3 Preparation of 'His-tagged' hetero poly-proteins

2.3.1 Transformation of DNA into *E. coli* for protein expression

E. coli cells, of the required strain, were thawed on ice for 10 minutes. 5 μL of plasmid DNA was added to 50 μL of cells in an eppendorf tube and incubated on ice for 30 minutes. The cell suspension was then heat shocked by submerging the tubes in a water bath at 42°C for 45 seconds. The cells were then returned to ice for 2 minutes before addition of 450 μL of sterile LB medium and then incubated at 37°C for 1 hour, with shaking at 200 rpm. After incubation, 200 or 20 μL of cells were spread onto LB agar plates containing 100 μg/mL ampicillin. The agar plates were incubated overnight at 37°C.

2.3.2 Small scale over-expression of protein

A single cell colony of cells was picked and used to inoculate 10 mL LB medium containing the required antibiotic. This starter culture was incubated overnight at 37°C with shaking at 200 rpm. 2 mL of the starter culture was used to inoculate 100 mL of the appropriate medium, which also contained the required antibiotic in a 250 mL conical flask. The culture was incubated at 37°C with shaking at 200 rpm. Cell growth was monitored until an OD_{600} of 0.7 was reached and a 1 mL sample was taken. The culture was then induced by adding sterile IPTG to a final concentration of 1 mM. The cells were grown for up to 5 hours, 1 mL samples were taken every hour and OD_{600} was measured.

The cell samples were pelleted by centrifuging at 13000 rpm for 2 minutes in a bench-top centrifuge and removing the medium. The pellets were then resuspended in 2x reducing SDS loading buffer and analysed using SDS-PAGE (see section §2.6.1)

The remaining cells were harvested by centrifugation at 6000 rpm, 4°C for 20 minutes and the supernatant was discarded.

2.3.3 Small scale protein purification using Ni-NTA agarose beads

60 μL Ni Sepharose resin, suspended in a 1:1 slurry with 20% ethanol, was placed in a 1.5 mL eppendorf tube and centrifuged at 13000 rpm for 1 minute. After removing the supernatant, the beads were washed with 600 μL wash buffer (50 mM Tris-HCl, 500 mM NaCl, 20 mM Imidazole, pH 8.0). The solution was again centrifuged at 1500 rpm for 1 minute and the buffer was removed.

The harvested cells from the small scale expression (see section §2.3.2) were resuspended in 2.5 mL Bugbuster protein extraction reagent and incubated at room temperature for 20 minutes. The solution was centrifuged at 4000 rpm, 4°C for 20 minutes. The supernatant was applied to the beads and allowed to equilibrate for 30 minutes. After this binding period, the protein suspension was centrifuged at 1000 rpm to allow removal

of unbound protein. The beads were washed three times as described previously. The protein was eluted from the beads by adding 100 μL elution buffer (50 mM Tris-HCl, 500 mM NaCl, 250 mM Imidazole, pH 8.0) and the eluted protein solution was removed from the beads after centifuging. Finally, samples of the cell lysate, washes and elution were taken, diluted by a ratio of 1:1 in 2x reducing SDS loading buffer and analysed using SDS-PAGE (see section §2.6.1)

2.3.4   Large scale over-expression of protein

A single colony of cells was picked and used to innoculate 110 mL LB medium in a sterile 250 mL conical flask, also containing the required antibiotic. The culture was incubated overnight at 37°C with shaking at 200 rpm. 10 mL of starter culture was used to innoculate each of 10 x 2 L conical flasks filled with 1 L of the selected medium, also containing the required antibiotic. The expression procedure outlined in section §2.3.2 was followed. The cells were harvested by centrifugation using a Beckman Coulter Avanti J-26 XP centrifuge at 6000 rpm, 4°C for 15 minutes.

2.3.5   Large scale cell lysis

The cell pellet was resuspended in a lysis buffer (50 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole, 1 mM PMSF, 2 mM benzamidine, pH 8.0) and a small amount of DNase was added. The solution was stirred using a magnetic stirrer and incubated at 4°C until the lysate was no longer viscous. The solution was then passed through a glass homogeniser until a the solution was fully homogenised. The cells were then lysed using a cell disrupter. The insoluble extract was then removed by centrifuging the sample at 16000 rpm for 30 minutes at 4°C and the supernatant was collected.

2.3.6 Large scale protein purification using Ni-NTA affinity chromatography

The cell lysate was filtered through a 0.45 μm syringe filter and purified using an XK 50/20 column containing 30 mL of Ni Sepharose resin connected to an ÄKTA Prime chromatography system. The column was washed with one column volume of water and one column volume of lysis buffer (see section §2.3.5) at which point the recorded absorbance at 280 nm (A_{280}) was zeroed. The lysate was loaded onto the column and then the column was washed with lysis buffer until the A_{280} returned to zero. To elute the bound protein, an elution buffer (50 mM Tris-HCl, 500 mM NaCl, 250 mM imidazole, pH 8.0) was added to the column in a stepped process (25%, 50%, 75% and 100%), progressing to the next step each time the A_{280} reached a constant value. 2 mL fractions were collected during this process, which were analysed by SDS-PAGE (see section §2.6.1) to allow identification of the fractions that contained the desired protein. Elution fractions containing the desired protein were pooled together.

The pooled fractions were then diaysed using 3500 MWCO dialysis tubing, against 5 L 18 MΩcm H_2O containing 5 mM ammonium bicarbonate at 4°C. The dialysis solution was changed three times after incubation for 3 hours. The dialysed protein solution was then snap frozen in falcon tubes on dry ice and ethanol and lyophilsed.

2.3.7 Large scale protein purification using size exclusion chromatography

A Superdex 75 gel filtration column connected to an ÄKTA Prime chromatography system was washed with one column volume of water and one column volume of gel filtration buffer (25 mM Tris-HCl, 500 mM NaCl, pH 8.0). The lyophilised sample was resuspended in 9 mL gel filtration buffer and injected onto the column using 3 mL per run. Gel filtration buffer was run through the column at a rate of 3 mL/min while collecting 2 mL fractions. The collected fractions were analysed by SDS-PAGE (see section §2.6.1) to allow identification of the fractions that contained the desired

protein. Elution fractions containing the desired protein were pooled together. The pooled fractions were then diaysed as explained in section §2.3.6. The dialysed protein solution was then snap frozen in falcon tubes on dry ice and ethanol and lyophilsed.

2.3.8 Large scale protein purification using anion exchange chromatography

A 5 mL HiTrap Q HP anion exchange column connected to an ÄKTA Prime chromatography system was washed with one column volume of water and one column volume of anion exchange buffer (25 mM Tris-HCl, pH 8.0). Lysophilised protein sample was resuspended in anion exchange buffer. The protein was loaded onto the column and the column was washed with one column volume of anion exchange buffer. Protein was eluted from the column by adding anion exchange elution buffer (25 mM Tris-HCl, 1M NaCl, pH 8.0) in gradient from 0% to 100% over 20 column volumes, 2 mL fractions were collected during this process. The collected fractions were analysed by SDS-PAGE (see section §2.6.1) to allow identification of the fractions that contained the desired protein. Elution fractions containing the desired protein were pooled together. The pooled fractions were then diaysed as explained in section §2.3.6. The dialysed protein solution was then snap frozen in falcon tubes on dry ice and ethanol and lyophilsed.

2.4 Preparation of homo-polyproteins

2.4.1 Protein production

The required protein, bound to maltose binding protein (MBP) at the N-terminus separated by a Tobacco Etch Virus protease (TEVp) cleavage sequence, was produced following the methods outlined in sections §2.3.4, §2.3.5, §2.3.6 and §2.3.7.

2.4.2   Cleaving MBP fusion proteins using TEVp

The protein was resuspended in TEVp cleavage buffer (25 mM Tris-HCl, 500 mM NaCl, 14 mM β-mercaptoethanol, pH 8.0) to a concentration of 5 mg/mL. To produce a 1:3 molar ratio of TEVp to protein, 90 μL of 6 mg/ml TEVp was added per 1 mL of 0.5 mg/ml protein solution used. The solution was incubated at 25°C for 15 minutes. 30 μL samples were collected before and after this process which were analysed by SDS-PAGE (see section §2.6.1).

2.4.3   Protein purification using Ni-NTA agarose beads

625 μL Ni Sepharose resin, suspended in a 1:1 slurry with 20% ethanol, was placed in a 1.5 mL eppendorf tube and centrifuged at 13000 rpm for 1 minute. After removing the supernatant, the beads were washed with 625 μL wash buffer (50 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole, pH 8.0). The solution was again centrifuged at 13000 rpm for 1 minute and the buffer was removed.

The protein-TEVp solution was applied to the beads and allowed to equilibrate for 30 minutes. After this binding period, the protein suspension was centrifuged at 1000 rpm to allow removal of unbound protein, this fraction contained the desired protein (does not contain His-tag). The beads were washed three times as described previously. The supernatant from the first wash was pooled together with the unbound protein fraction. The TEVp and MBP proteins were eluted from the beads by addition of 100 μL elution buffer (50 mM Tris-HCl, 500 mM NaCl, 250 mM imidazole, pH 8.0) and the eluted protein solution was removed from the beads after centifuging. Finally, samples of the cell lysate, washes and elution were taken, diluted in 2x reducing SDS loading buffer and analysed using SDS-PAGE (see section §2.6.1)

2.4.4 Polymerisation of proteins using a crosslinker

The protein was resuspended in crosslinking buffer (50mM Tris-HCl, 500 mM NaCl, pH 7.4) to a concentration of 2 mg/mL. 10 mM 1,8-bis(maleimido)diethylene glycol (BM(PEG)$_3$) crosslinker was added to form a 1:1 ratio of crosslinker to protein. The solution was incubated at 25°C for 2 hours. The solution was then purified to remove smaller poly-proteins (see section §2.3.7).

2.5 Purification of insoluble proteins

2.5.1 Protein production

The required protein was expressed and cells were lysed following the methods outlined in sections §2.3.4 and §2.3.5, this time retaining the insoluble fraction of the lysate and discarding the supernatant. The solid faction was then resuspended in unfolding buffer (8 M Urea, 50 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole, 1 mM PMSF, 2 mM benzamidine, pH 8.0) and centrifuged at 16000 rpm for 30 minutes at 4°C, the supernatant was collected.

2.5.2 Refolding using Ni-NTA affinity chromatography

The cell lysate was filtered through a 0.45 μm syringe filter and purified using an XK 50/20 column containing 30 mL of Ni Sepharose resin connected to an ÄKTA Prime chromatography system. The column was washed with one column volume of water and one column volume of unfolding buffer (see section §2.5.1) at which point the recorded absorbance at 280 nm (A$_{280}$) was zeroed. The lysate was loaded onto the column and then the column was washed with lysis buffer until the A$_{280}$ returned to zero. A wash buffer (0.9 M arginine, 50 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole, 1 mM PMSF, 2 mM benzamidine, pH 8.0) was then added to the column in a 0-100% gradient over 5 column volumes to refold the protein. The protein was then eluted from the column

following elution steps described in §2.3.6.

2.6 General protein analysis methods

2.6.1 Sodium dodecyl sulphate polyacrylamide electrophoresis (SDS-PAGE)

SDS-PAGE was used to separate protein samples by molecular weight in order to determine protein yield and purity. A two-layered gel system was used, which consisted of a stacking gel and a resolving gel. The components of each gel are listed in table 2.12. Two glass plates were set up according to manufacturers instructions (Atto mini-gel kit), including a 1.5 mm spacer. Firstly the resolving gel was poured between the 2 glass plates, leaving a 2 cm gap at the top. The stacking gel was immediately poured on top of the resolving gel and a comb was inserted to create wells for sample loading. The gels were left for 1 hour to set.

Component	Resolving (x2)	Stacking (x2)
30% (w/v) acrylamide: 0.8% (w/v) bisacrylamide	5.1 mL	0.67 mL
3 M Tris-HCl, 0.3% (w/v) SDS, pH 8.45	5.00 mL	1.24 mL
H_2O	2.82 mL	3.05 mL
Glycerol	2 mL	-
10% (w/v) ammonium persulphate	150 μL	80 μL
Tetramethylethylenediamine (TEMED)	15 μL	8 μL

Table 2.12: Composition of a an SDS-PAGE gel

Samples, resuspended in 2x reducing SDS loading buffer (50 mM Tris-HCl, 100 mM DTT, 2% (w/v) SDS, 0.1% (w/v) bromophenol blue, 10% (v/v) glycerol, pH 6.8), were boiled at 100°C for 2 minutes before cooling on ice for 2 minutes. The samples were then centrifuged at 13000 rpm for 2 minutes. 15 μL of each sample was added to separate wells in the gel and 10 μL of a molecular weight marker was also added to one well to allow identification of the molecular weights of the sample gel bands.

Gel electrophoresis was carried out using cathode buffer (100 mM Tris-HCl, 100 mM

74

tricine, 0.1% (w/v) SDS, pH 8.3) in the inner reservoir of the gel tank and anode buffer (400 mM Tris-HCl, pH 8.8) in the outer reservoir. A current of 30 mA was applied across the gel until the samples entered the resolving gel. The current was then adjusted to 60 mA until the dye front reached the bottom of the gel. The gels were then removed from the glass plates and stained using InstantBlue stain. The gels were then photographed using a transilluminator with InGenius gel documentation system (Syngene).

2.6.2 Western blot

An SDS-PAGE gel was run using the technique described in section §2.6.1, using a pre-stained molecular weight marker and stained using Coomasie Blue stain. Protein bands from SDS-PAGE gels were transferred to Polyvinylidene Fluoride (PVDF) membranes using a Bio-Rad Trans-Blot semi-dry transfer cell. PDVF was activated by soaking in methanol and then in transfer buffer (25mM Tris-HCl, 192 mM glycine, 20% (m/v) methanol). 6 pieces of blotting paper were also soaked in transfer buffer. Residual buffer was removed, then the piece of PDVF was placed on top of 3 pieces of blotting paper. Next the SDS-PAGE gel was placed on top and finally 3 more pieces of blotting paper. The sandwich was pressed to remove any trapped air bubbles and placed on the bottom electrode. Once the top electrode was added on top, the transfer cell was set at 400 mA, 12 V and left for 1 hour. Once finished, the PVDF was removed from the cell and submerged in PBS buffer containing 10% dried skimmed milk and incubated overnight at 4°C on a rotary shaker.

The membrane was then transferred into a falcon tube with 5 mL PBS, 5% dried skimmed milk containing the 1st appropriate antibody and incubated for 1 hour at 4°C on a rotary shaker. The solution was discarded and PBS buffer containing 10% dried skimmed milk was added and placed on a rotary shaker for 5 minutes, after which the solution was discarded. This process was repeated 3 times.

The membrane was then transferred into a falcon tube with 5mL PBS, 5% dried skimmed milk solution containing the second appropriate antibody, using the same technique, the process was again repeated 3 times. 1 mL of each SuperSignal West Pico

Chemiluminescence Substrate from Thermo Scientific was then added according to manufacturer's instructions. The membrane was washed with dH_2O and sealed in a piece of clear plastic. The membrane was then exposed to an X-ray photographic film in a dark room.

2.6.3 Mass spectrometry

The molecular weight of the expressed protein was measured using Electrospray Ionisation Mass Spectrometry (ESI-MS). Samples were lyophilised and then ESI-MS experiments were carried out by the mass spectrometry facility managed by Dr James Ault. 0.05 mg of protein was unfolded in a 50% (v/v) acetonitrile, 0.1% (v/v) formic acid pH $\sim$2.5 in water.

2.7 Single Molecule Force Spectroscopy (SMFS)

SMFS experiments in this project were carried out using a custom built Atomic Force Microscope AFM instrument. This section will introduce the AFM and describe the process used to set up and calibrate an SMFS experiment. I will then conclude with an explanation of the data analysis procedure carried out for all experiments.

2.7.1 The Atomic Force Microscope (AFM)

An overview of the AFM can be seen in figure 2.1.
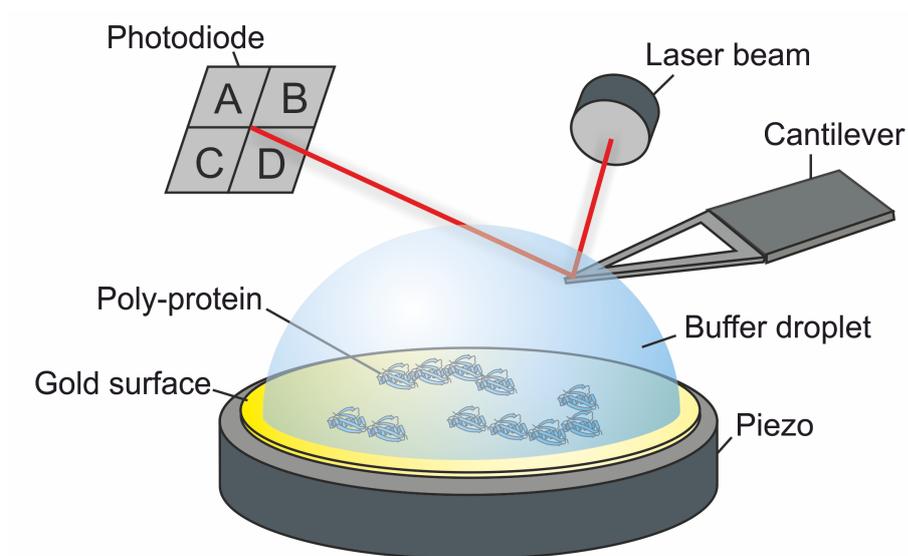
Figure 2.1: Overview of the Atomic Force Microscope.

2.7.1.1 Cantilevers

The measuring probe of the microscope consists of a cantilever with an integrated tip at its end that is able to interact with a sample. Here, probes were used which contained 6 silicon nitride cantilevers with lengths from 85–310 μm and thickness of 0.55 μm, each with a silicon nitride tip (obtained from Bruker, CA), shown in figure 2.2. The cantilevers used in this project have a small spring constant, ranging from 25 to 35 pN/nm, enabling a resolution in the force measurement in the order of piconewtons to be obtained.
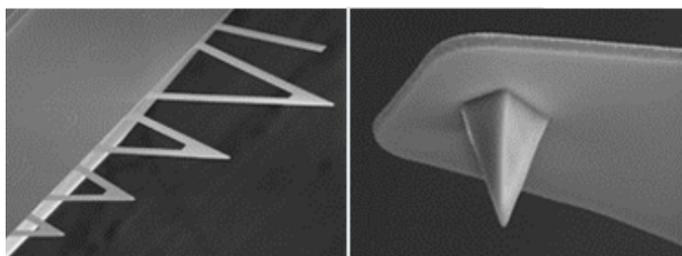
Figure 2.2: Overview of a cantilever. Image taken from Bruker AFM Probes[169].

2.7.1.2 The photodiode

The AFM measures small deflections of the cantilever so requires a sensitive detection method. Here, a laser beam is reflected off the cantilever onto a photodiode. Movements of the tip are measured by the any deviation of the laser light from the centre of the photodiode. Any deviation from the centre will cause a change in the amount of laser light detected in each quadrant of the photodiode, which can be used to determine the force exerted in the cantilever.

2.7.1.3 The piezo

To measure the unfolding force of a protein, the sample must be moved away from the cantilever with very accurate precision. This is done with the use of the piezoelectric effect, where certain crystals expand or contract in the direction in which a voltage is applied. The use of a piezo allows accurate control over the position to the surface on a nanometre length scale. The PicoCube$^{\text{TM}}$ XYZ Stage was obtained from Physik Instrumente Ltd, Bedford, which a 5 x 5 x 5 μm travel range with a $\sim$0.5 nm resolution.

2.7.1.4 Use of a gold surface

The polyprotein constructs are bound to the gold surface at one end. This is achieved by the addition of two cysteine residues at the C-terminus, which contain thiol groups. The thiol group binds to the gold surface (figure 2.3), tethering the poly-protein. Gold surfaces were produced by glueing a glass slide onto a gold tempered silicon wafers (ordered from Platypus technologies) using epoxy and incubating at 120° for 2 hours. A fresh slide was cleaved from the surface at the start of each experiment.

2.7.1.5 The fluid cell

The buffer droplet is contained in a fluid cell (not shown in figure 2.1), depicted in figure 2.4. The fluid cell holds the cantilever in position while containing the buffer in an o-ring
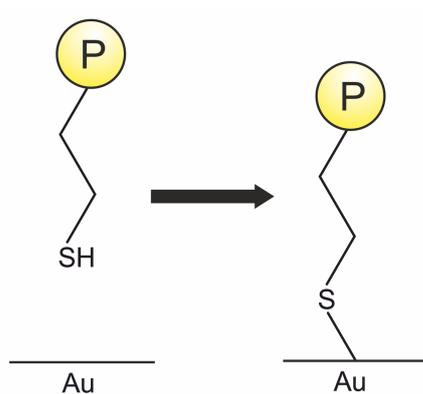
Figure 2.3: The reaction of a thiol group with a gold surface. Where P represents the polyprotein.

to prevent evaporation. MTFML fluid cells were obtained from Bruker, which hold one cantilever.
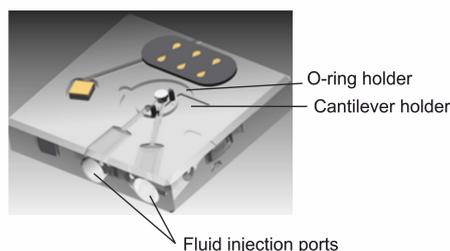
Figure 2.4: Overview of a fluid cell. Image taken from Bruker AFM Probes[169].

2.7.2   Calibration of the cantilever

2.7.2.1   Theory behind the calibration

A crucial part of the SMFS experiment is the calibration of the cantilever. In order to get the most accurate measurement a calibration is performed at the start of every experiment and is checked during each experiment. Here, the thermal noise method was used to calculate the spring constant of the cantilever using the equipartition theorem[170].

The equipartition theorem states that each degree of freedom has the same thermal energy in a thermodynamic system. Each degree of freedom, x, of the cantilever has

energy given by equation 2.2, where k_B is Boltzmann's constant and T is the absolute temperature.

$$E_x = \frac{1}{2}k_B T \tag{2.2}$$

When the tip is large distance from the gold surface, only thermal fluctuations of the cantilever will cause motion. This motion is limited to the vertical direction, z. The time average of the vibrations of the cantilever, caused by thermal fluctuations will, therefore, be directly related to the spring constant of the cantilever.

The spring constant of the cantilever, k_c, can then be calculated using the equation 2.3, where $\langle z^2 \rangle$ is the time-average square of the thermal fluctuation of the cantilever in the z direction.

$$\frac{1}{2}k_B T = \frac{1}{2}k_c \langle z^2 \rangle \tag{2.3}$$

2.7.2.2 The calibration method

Initially, the cantilever is submerged in the buffer droplet and the laser position is optimised. This is done by maximising the amount of light detected by the photodiode. The laser spot is also centred on the photodiode ready for calibration.

A power spectrum is then taken, shown in figure 2.5 and the area under the resonant peak, z_u is calculated. z_u estimates the voltage obtained from the time-average square of the thermal fluctuation.
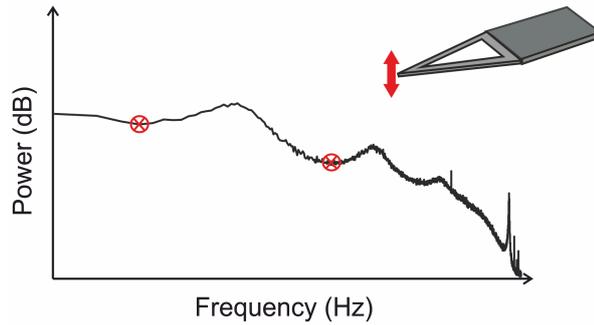
Figure 2.5: Power spectrum for the calibration of a cantilever in liquid at 23 °C. Measured by monitoring the deflections of the cantilever in equilibrium with the buffer far from the surface. The first peak shows the resonant frequency of the cantilever oscillation. The red markers define the edges of the resonance peak, placed in the two outside minima.

The gold surface is then moved up to the cantilever. A force-displacement curve is obtained by pushing at a force of 1500 pN, shown in figure 2.6. The gradient of the measured slope of displacement against detector voltage is used to to determine the conversion factor of the displacement of the cantilever in metres to the voltage changes of the detector, i.e. the sensitivity, S, given by equation 2.4[171].

$$S = \frac{\Delta V}{\Delta z} \tag{2.4}$$
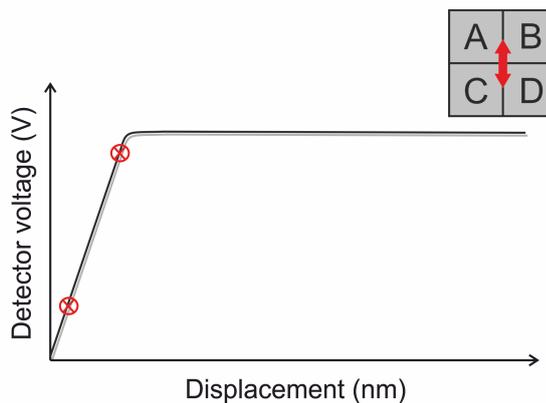
Figure 2.6: Force-displacement curve for the calibration of a cantilever in liquid. The slope of the deflection against voltage gives the sensitivity of the diode.

The value of z_u, previously calculated in volts, is then converted into metres, z, using the sensitivity measured by the force-displacement curve. The spring constant of the cantilever can then be calculated using equation 2.3[172].

2.7.3 A SMFS experiment

2.7.3.1 Setting up a SMFS experiment

Lyophilised protein samples were resuspended in the appropriate buffer to a final concentration of 0.5 g/mL and filtered through a 0.45 μm filter. 40 μL of the solution was then applied onto a freshly cleaved template-stripped gold surface. The sample was incubated at room temperature for 30-60 minutes to allow the sample to bind to the gold surface before gently removing the solution and adding 40 μL of buffer to reduce the amount of unbound protein.

2.7.3.2 Overview of a SMFS experiment

The gold surface, with bound polyprotein, was moved towards the silicon nitride tip at a constant velocity until a contact force of 1500 pN is achieved. At this point poly-protein may be adsorbed on the cantilever tip. The surface is then retracted away from from tip at a constant velocity. If a poly-protein does adsorb, a restoring force will be exerted by the poly-protein on the tip, causing the cantilever to bend. The movement of the cantilever is measured by the photodiode and is then used to calculate the applied force, shown in figure 2.7 A). The resulting AFM trace is shown in different colours.
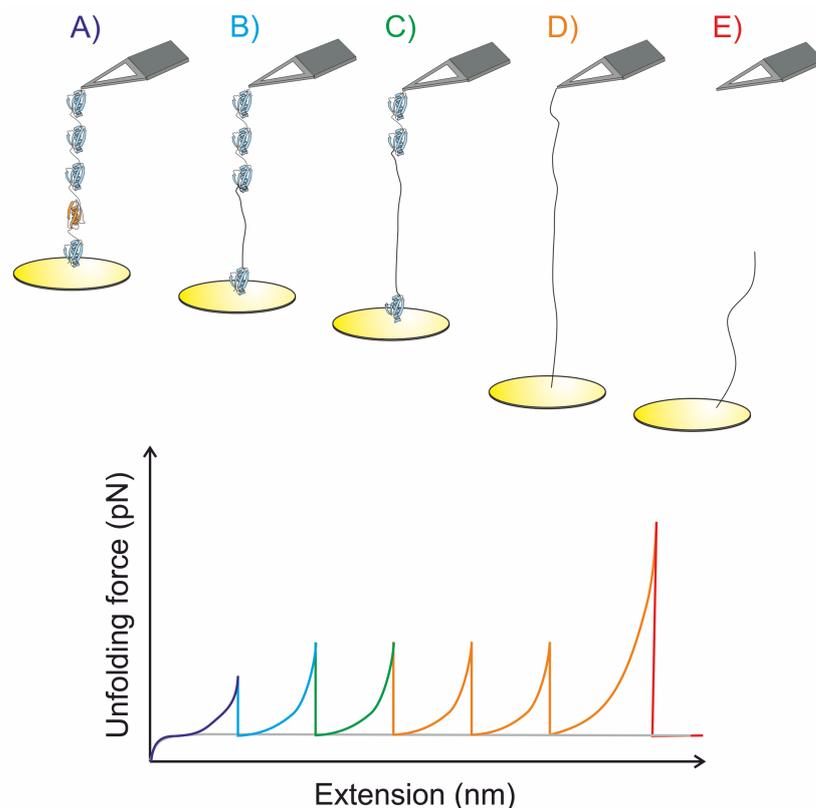
Figure 2.7: Schematic of a SMFS experiment using a poly-protein. A) represents the poly-protein construct adsorbed onto the tip, B) represents the mechanical unfolding of the first protein domain, C) represents the mechanical unfolding of the second protein domain, D) represents the mechanical unfolding of the rest of the protein domains and E) represents the detachment of the poly-protein from the tip. The resulting AFM trace is shown, where the grey line is the approach or the surface to the cantilever.

Once a particular force is reached, one of the proteins will unfold and the force exerted on the cantilever diminishes. The protein unravels as the gold substrate is moved further from the cantilever. When the exposed structure has fully unravelled, further movement will again be resisted by the restoring force of the poly-protein. This is depicted in figure 2.7 B) shown in blue.

Again, once a particular force is reached, one of the proteins will unfold and the process repeats, shown in figure 2.7 C) in green. This process will repeat for each of the domains in the poly-protein chain, 2.7 D) shown in orange. Once each domain in the poly-protein chain has fully unfolded, a typically larger force is needed to detach

the extended structure from the cantilever tip, shown in figure 2.7 E) shown in red. In this project, each SMFS experiment was carried out for up to 10 hours at retraction velocities of 400, 700, 1200 and 2000 nm/s. The traces collected were filtered according to the criteria stated for each experiment.

2.7.4 Analysing data collected from SMFS experiments

2.7.4.1 The Worm-Like Chain (WLC) model for polymer elasticity

The simplest way to describe a linear polymer is a random coil, a chain of monomer units that are able to freely rotate about each linkage. A real polymer is semi flexible, i.e. the rotation is restricted. The restriction may arise due to the chemical structure of each monomer or from the formation of a secondary structure.

When a polymer is extended at a constant velocity, it is elongated and work is needed to overcome the reduction of entropy. This creates a restoring force, $F(x)$, in the chain at a given extension, x. The WLC is an empirical model that describes the relationship between the extension of a polymer and the entropic restoring force generated, using equation 2.5[173]. Persistence length, p, is a measure of the stiffness of a chain. It can be described as the distance over which the correlations between sections of the chain are lost. If a chain is very flexible, correlations between sections of the chain will be lost quicker, therefore p is small. The p for a protein is about the length of one amino acid, estimated at 0.4 nm. L_c is the contour length, which is the full extension on the polymer, which is measured for each poly-protein trace.

$$F(x) = \frac{k_B T}{p} \left[\frac{1}{4} \left(1 - \frac{x}{L_c} \right)^{-2} - \frac{1}{4} + \frac{x}{L_c} \right] \tag{2.5}$$

A study by Marko et al. has shown that the WLC model describes the stretching of DNA very well[173,174]. The WLC model has also been used to describe poly-protein unfolding in number of publications[148,149,153,175,176].

For this project, it is fitted to each unfolding event in a force-extension trace, see figure 2.8. This was used as an initial check to see if the unfolding event behaved as expected for an unfolding protein, i.e. the shape of the curve fitted well to the unfolding event. As another check, the change in contour length upon unfolding, ΔL_c, for each unfolding event was then measured and compared to literature values where available.
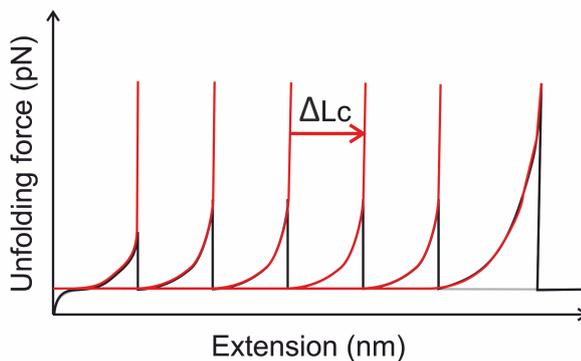
Figure 2.8: A force-extension trace fitted with the worm-like chain model for polymer elasticity. The worm-like chain fit is in red and the change in contour length upon unfolding is marked as ΔL_c

2.7.4.2 Analysis of the force-extension traces from SMFS experiments using chimera-polyproteins

The analysis procedure described here is for a chimera polyprotein (hetero-polyprotein), the structure of which is shown in figure 2.9. Each trace that fulfilled the acceptance criteria (described for each experiment) and could be fitted correctly with the worm-like chain was then analysed.

First, the peak to peak distance, p2p, and the unfolding force, F_{UN}, for each unfolding event was measured, shown in figure 2.9 A). The p2p and F_{UN} values that came from protein x were separated from the values that came from protein y and pooled for each pulling velocity (proteins x and y depicted in figure 2.9). The mean and median values for p2p and F_{UN} were also calculated for each pulling velocity.

The p2p and F_{UN} data values for both protein x and y were then plotted in a scatter
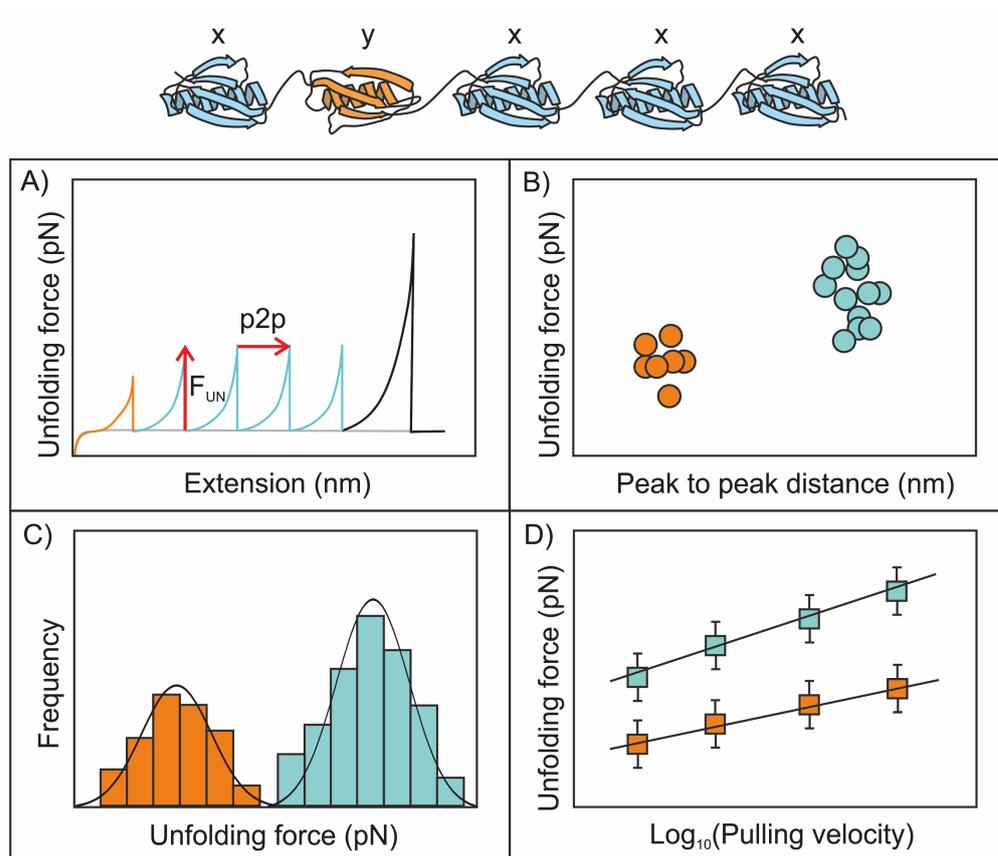
Figure 2.9: Overview of the analysis carried out for SMFS experiments. The poly-protein is shown containing 4 repeats of protein x in blue and one protein y in orange. A) represents a typical force-extension trace for this protein, showing the measurement of the unfolding force, F_{UN}, and peak to peak distance, p2p. The unfolding events for protein x are shown in blue and those for protein y are shown in orange. B) represents a scatter plot formed for one pulling velocity. C) represents the histograms plotted for the $_{UN}$ for both proteins, normalised for the number of unfolding events for each protein. D) represents the speed dependence diagram plotted, showing the linear fit for each protein and standard errors for each value.

diagram of the p2p vs. F_{UN}, an example for which is shown in figure 2.9 B). The scatter diagram is used to see if the values recorded formed a cluster for each protein and also to see if the cluster centred around literature values if available. Next, histograms were plotted with the p2p and F_{UN} data, depicted in figure 2.9 C). The values were normalised for the number of unfolding events for each protein. Each histogram was fitted with a gaussian distribution curve,the centre of the gaussian for each histogram was recorded. Finally the mean F_{UN} value for each pulling velocity was plotted in a speed dependence diagram, showing $\log_{10}$ (pulling velocity) vs. F_{UN}. An example of a speed dependence diagram is shown in figure 2.9 D). A linear fit is performed for each of the proteins. The properties of the distribution of the F_{UN} data, depicted by the histogram, and the slope of the linear from the speed dependence diagram can be used to provide more information about the mechanical unfolding energy landscape of a protein[147].

Chapter 3

Results 1: SMFS and polyprotein engineering experiments to examine the mechanical fingerprint of a halophilic protein

3.1 Introduction

This chapter will introduce and describe the experiments completed on the protein *Haloarcula marismortui* Ferredoxin (*Hm*Fd). The chapter will begin by explaining how the halophilic protein, (*Hm*Fd) was chosen. The organism, *Haloarcula marismortui*, and the protein, *Hm*Fd, will then be introduced, including what is known about the structure and stability of the protein. The experimental data collected, analysed and the conclusions drawn from SMFS experiments using the hetero-polyprotein, I27$_3$-*Hm*Fd-I27 will then be presented. Next, two techniques used in attempt to synthesise further halophilic polyprotein constructs will be described, including all the experimental steps carried out. The chapter will finish by concluding of all results shown.

3.1.1 Choosing a halophilic protein to study using SMFS experiments

3.1.1.1 Identifying a protein that will have high mechanical stability

A large number of proteins have now been mechanically unfolded using SMFS experiments[115,154,162,177–179] The force required to unfold a protein has been shown to range from tens to hundreds of piconewtons[133,148,180]. A collection of studies have been carried out to determine the structural features that control the stability of a protein under mechanical force (see section 1.4.4.1)[133–135]. These studies show that the mechanical unfolding force of a protein can be predicted by the secondary structure and topology. Proteins that are predominantly α–helical tend to need lower forces to overcome the energy barrier to unfolding, whereas proteins with a high β–sheet content tend to need high forces. This is thought to be due, in large part, to the shearing apart of two β strands by pulling apart from distal ends, which requires all non-covalent interactions between the strands to rupture at once[154]. Proteins that have a β-sandwich topology with directly hydrogen bonded β-strands, such as the Ig-like fold of I27, therefore have a high mechanical resistance[135,139].

The halophilic protein, HmFd comprises a 4-stranded β–sheet with direct hydrogen bonds, formed into a barrel-like structure, with 5 flanking α-helices[105]. It is, therefore, predicted to have significant mechanical stability. The structure of HmFd also includes an Iron-Sulphur (Fe-S) cluster bound to four cysteine residues, 63, 68 and 71, which are located in the loop region connecting helix α_3 and strand β_3 and 102, which is located after helix $\alpha_4$3, shown in figure 3.1. A review of the literature available for the halophilic protein, HmFd, can be found in section 1.3.2.4.

3.1.1.2 Mechanical unfolding of a structurally similar protein, Rubredoxin

Pyrococcus furiosus Rubredoxin (PfRd), a protein also containing an Fe–S cluster, has been studied using SMFS experiments[159,181–186]. These studies showed that the Fe–S bond could be ruptured using this technique and the mechanical force could be charac-
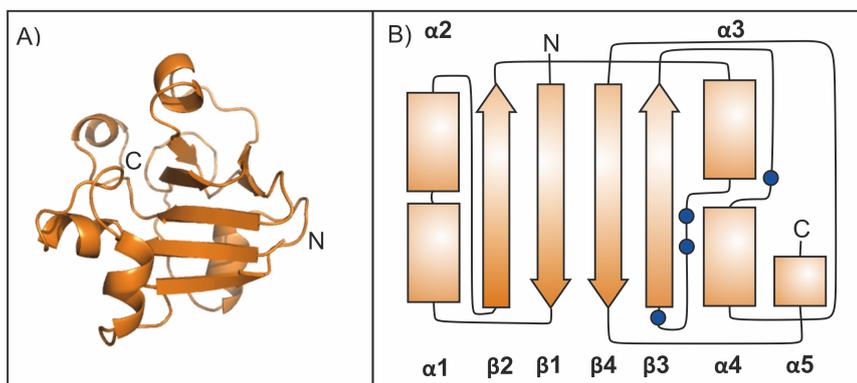
Figure 3.1: The structure of *Hm*Fd. A) shows the 3D topology of the protein. This figure was created using PDB file 1DOI and Pymol[90,105]. B) shows the secondary structure of the protein where the arrows represent β–sheets and the rectangle represents an α=-helix. The blue circles represent the cysteines which bind to the FeS cluster. The N–terminus is shown on β–strand, β1 and the C–terminus is on the α–helix, α5.

terised. SMFS experiments in this study were carried out in DTT to create a reducing environment to ensure Fe was in the $+2$ oxidation state[159].

The mechanical unfolding of *Pf*Rd (53 amino acids) showed firstly the secondary structure of the protein either side of the Fe-S cluster unravel without producing a resistive force on the cantilever tip (residues 1-4, 42-53)[187]. The strength of the Fe-S cluster then produces a resistive force which bends the cantilever[159]. The Fe-S cluster then ruptures, which allows the amino acid residues which are held by the cluster (residues 5-41) to unravel. The average measured peak to peak value, 12.6 $\pm$1.3 nm, reflected the distance between residues 5-41, estimated to be 12.1 - 13.9 nm.

The measured unfolding force was 211 $\pm$ 86 pN at a pulling speed of 400 nm/s, with a wide distribution range from 100 pN to 500 pN. The level of interaction between two atoms is determined by the mixing of their atomic orbitals, described by the covalency. An Fe-S bond has a high covalency due to the mixing of the Fe d-orbital and the S p-orbital. This unfolding force was surprisingly low since other highly covalent bonds have been found to be highly mechanically stable e.g. a disulphide bond which has a lower covalency than Fe-S has a mechanical strength >1 nN[188,189].

Further experiments with *Pf*Rd have been carried out to investigate the effect of covalency of the Fe-S cluster and the effect of the hydrogen bonds surrounding the Fe-S cluster on the mechanical stability of *Pf*Rd[182]. The direction of pulling force, the exact mechanism of the release of Fe and the reversible unfolding-refolding and the chemical reactivity of the Fe centre of rubredoxin have also been investigated[183–186].

Given the structural similarity of *Pf*Rd to *Hm*Fd, it was suggested that *Hm*Fd could be an ideal protein to study using SMFS.

3.2   Objectives

The objective of this chapter was to complete a series of SMFS experiments on a polyprotein construct containing *Hm*Fd to obtain experimental data and allow identification of the mechanical fingerprint of the protein.

3.3   Single molecule force spectroscopy on I27$_3$-*Hm*Fd-I27

In this section I will show data collected and analysed from three stages of SMFS experiments using I27$_3$-*Hm*Fd-I27:

i) I27$_3$-*Hm*Fd-I27 in 63 mM salt at a pulling speed of 400 nm/s (Ionic strength = 0.16 M)

ii) I27$_3$-*Hm*Fd-I27 in 500 mM and 2000 mM salt at a pulling speed of 400 nm/s (Ionic strength = 0.66 M)

iii) I27$_3$-*Hm*Fd-I27 in 63 mM salt at a pulling speed of 2000 nm/s (Ionic strength = 2.15 M)

3.3.1 Procedure for SMFS experiments using I27$_3$-HmFd-I27

The polyprotein I27$_3$-HmFd-I27 was provided by Dr. Toni Hoffman. Single molecule force spectroscopy (SMFS) experiments were performed using a custom built AFM. The spring constant of each silicon nitride cantilever, obtained from Veeco (Santa Barbara, CA), was calculated in buffer by applying the equipartition theorem (see section 2.7.2.1). Proteins were resuspended in the appropriate buffer at a concentration of 0.5 g/mL and applied onto a freshly cleaved template-stripped gold surface. The sample was incubated at room temperature for 30 minutes before mechanical unfolding experiments were carried out. Data was collected for timescales of up to 10 hours.

Figure 3.2: Schematic of the (I27-HmFd)$_3$-I27 construct. The construct has a His-tag, shown as a red triangle, at the N-terminal end and 2 cysteine residues, shown as green circles, at the C-terminal end.

3.3.2 SMFS experiments using I27$_3$-HmFd-I27 at 400 nm/s in in 63 mM sodium phosphate buffer

Example traces from these experiments are shown in figure 3.3. Figures A), B), C) and D) each show a force-extension trace with 3 I27 unfolding events. The traces show a range of events that may correspond to the unfolding of HmFd.

3.3.2.1 Analysis of the SMFS data for I27$_3$-HmFd-I27 at 400 nm/s in in 63 mM sodium phosphate buffer

Traces were accepted for analysis only if they contained 3 or more protein unfolding events, which corresponded to I27. Traces were not accepted if there were large surface interactions at the start of the trace, there were peaks with double tips or if there was not a clear detachment peak.
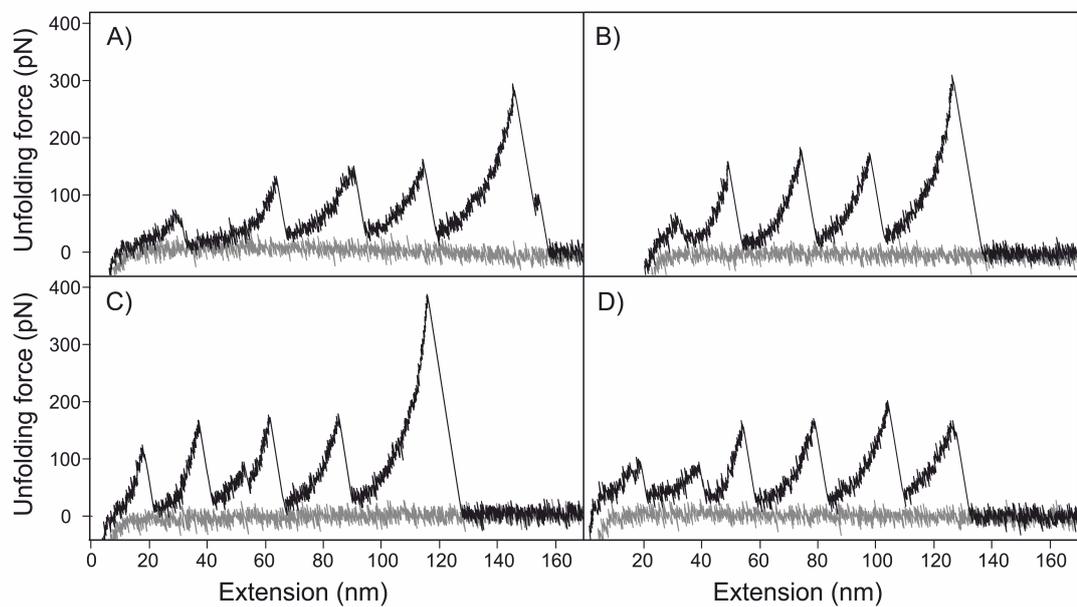
Figure 3.3: Example traces from SMFS experiments using I27₃-*Hm*Fd-I27 at 400 nm/s in 63 mM sodium phosphate buffer. Figures A), B), C) and D) each showing 3 I27 unfolding events and a range of events that could correspond to *Hm*Fd. The traces are taken from experiments carried out on 8 different days. The polyprotein was in 63 mM sodium phosphate, pulled at a constant velocity of 400 nm/s.

Traces were initially analysed by fitting the WLC model to confirm the number of I27 unfolding events, see section 2.7.4.1 for an explanation. For each protein unfolding event the peak unfolding force, F_{UN}, and the peak to peak distance, p2p, were recorded, overall values are recorded in tables 3.1 and 3.2. The mean F_{UN} obtained for I27 was 173 $\pm$ 2 pN and the p2p distance was 24.4 $\pm$ 0.1 pN. The values recorded were consistent with previous studies of I27 by SMFS by Hoffmann et al., of $F_{UN} = 168 \pm 7$ pN and p2p = 23.7 (error not quoted)[148].

Protein	No. events	Mean F_{UN} (pN)	Median F_{UN} (pN)	SD
I27	195	173	173	21
*Hm*Fd	88	114	104	43

Table 3.1: Unfolding force data and standard deviation for I27$_3$-*Hm*Fd-I27 in 63mM sodium phosphate buffer at 400 nm/s

Protein	No. I27 events	Mean p2p (nm)	Median p2p (nm)	SD
I27	195	24.4	24.3	1.3
*Hm*Fd	88	15.1	13.3	7.8

Table 3.2: Peak to peak distance data and standard deviation for I27$_3$-*Hm*Fd-I27 in 63mM sodium phosphate buffer at 400 nm/s

The values were plotted on a scatter diagram showing the unfolding force (F_{UN}) for each protein unfolding event against the measured peak to peak distance, p2p (figure 3.4). The diagrams show clear clusters of data points for I27 (grey), centred around the reference I27 data[148]. Other unfolding events, which are not expected to be I27, are also shown in the scatter plot (orange). There is no clear group of unfolding events which may correspond to *Hm*Fd. The wide spread of unfolding peaks may be due to non-specific interactions with the surface or impurities in the sample. They could also indicate a complex unfolding pathway for *Hm*Fd. Complex unfolding pathways have previously been reported for other proteins[160,190,191].

Histograms were then produced for the F_{UN} data and the p2p distances for both I27 (grey) and other events (orange), figure 3.5. Each histogram can be fitted with a Gaussian curve. The histograms for I27 show a clear Gaussian distribution for both the

p2p and F_{UN} data. There is a slight trend for the other data, i.e. a Gaussian can be fitted to the data, showing the results for a mechanical resistance for HmFd are promising but no definite unfolding force can be defined.

3.3.2.2 Refined analysis of the SMFS data for I27$_3$-HmFd-I27

In an attempt to extract some evidence of the unfolding of HmFd, the stringency of analysis was increased by only accepting traces that included four I27 unfolding events. If four I27 proteins have unfolded, the whole polyprotein construct must have been picked up from the surface and the HmFd protein must have been subjected to force, see figure 3.2 for the structure of the polyprotein. The F_{UN} and p2p distances of I27 for each experiment are recorded in tables 3.3 and 3.4. The mean F_{UN} obtained for I27 in 63 mM salt at 400 nm/s was 168 ± 5 pN and the p2p distance was 24.0 ± 0.3 pN, consistent with previous studies[148].

Salt conc.	Velocity	No. events	Mean F_{UN} (pN)	Median F_{UN} (pN)	SD
63 mM	400 nm/s	16	168	162	19

Table 3.3: Unfolding force data and the standard deviation for I27 in I27$_3$-HmFd-I27 using a more stringent analysis. The data shown only includes traces that have 4 I27 unfolding events.

Salt conc.	Velocity	No. events	Mean p2p (nm)	Median p2p (nm)	SD
63 mM	400 nm/s	16	24.0	24.2	1.1

Table 3.4: Peak to peak distance data and the standard deviation for I27 in I27$_3$-HmFd-I27 using a more stringent analysis. The data shown only includes traces that have 4 I27 unfolding events.

The refined data analysis values were plotted on a scatter diagram showing F_{UN} versus the p2p distance (figure 3.6 A)). The diagrams show clear clusters of data points for I27 (grey), centred around the reference I27 data[148]. Where only one unfolding event assigned to HmFd occurred in the same trace (in addition to the 4 I27 unfolding events), each data point was coloured orange. Where two unfolding events assigned to HmFd occurred in the same trace, each data point was coloured red. It is clear that not
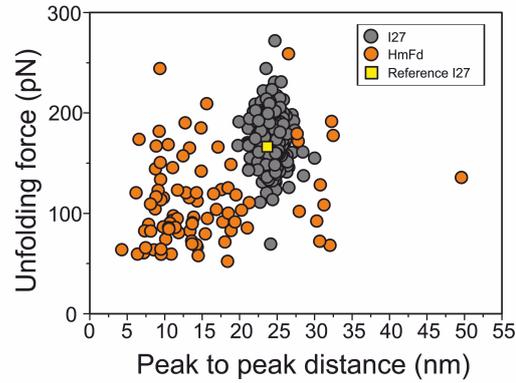
Figure 3.4: Scatter diagram of unfolding force vs. peak to peak distance of I27$_3$-*Hm*Fd-I27 taken from 8 SMFS experiments in 63 mM sodium phosphate, pulled at a constant velocity of 400 nm/s. The figure shows the data obtained for I27 in grey, clustered around the previously determined reference value for I27 in yellow[148]. The points which correspond to other peaks within the unfolding trace, interpreted as *Hm*Fd (orange) show no clear cluster.
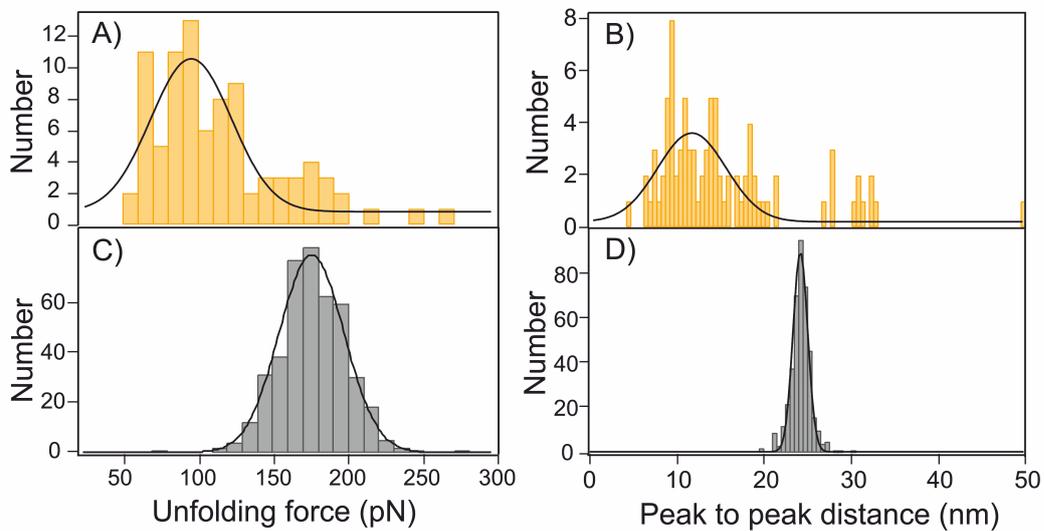
Figure 3.5: Data collected from SMFS experiments for I27$_3$-*Hm*Fd-I27 in 63 mM sodium phoshate at a pulling velocity of 400 nm/s. Figure A) shows the pooled data for the unfolding force of each *Hm*Fd unfolding event. Figure B) shows the pooled data for the peak to peak distance of each *Hm*Fd unfolding event. Figure C) shows the pooled data for the unfolding force of each I27 unfolding event. Figure D) shows the pooled data for the peak to peak distance of each I27 unfolding event.

enough data is available to clearly show the mechanical unfolding behaviour of *Hm*Fd.
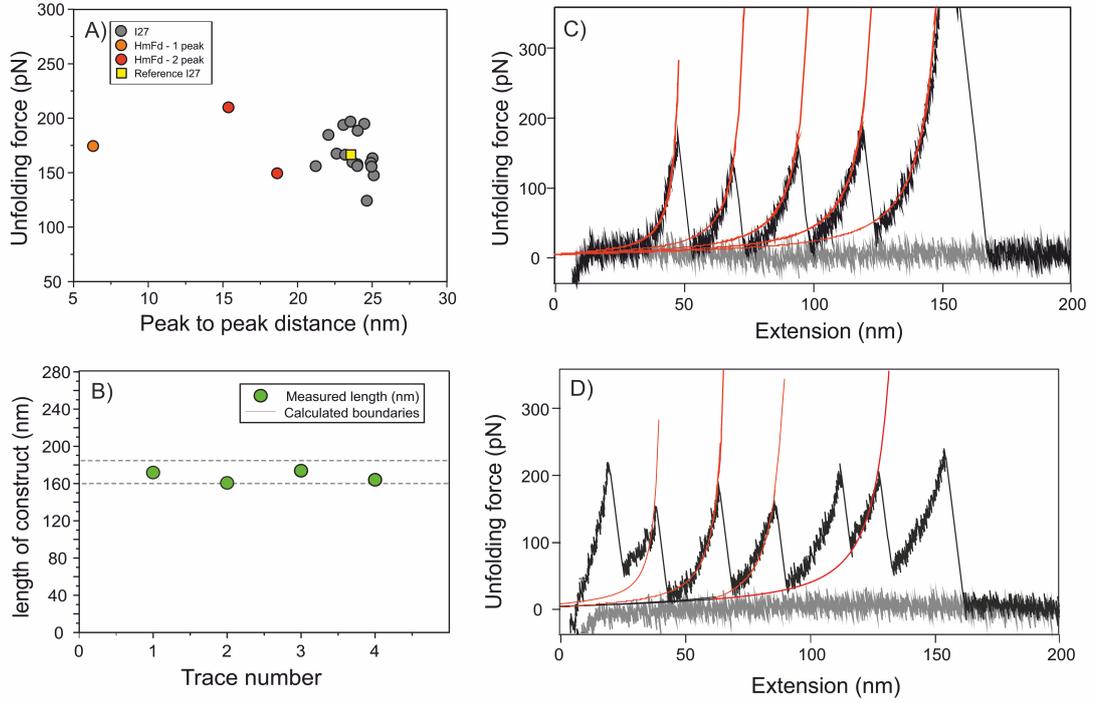
Figure 3.6: The refined data analysis of SMFS experiments using I27$_3$-*Hm*Fd-I27 at 400 nm/s. Figure A) shows the scatter diagram of the unfolding force vs. peak to peak distance for I27$_3$-*Hm*Fd-I27. The figure shows I27 in grey, clustered around the reference value for I27 in yellow[148]. Where only one unfolding event assigned to *Hm*Fd occurred in the same trace (in addition to the 4 I27 unfolding events), each data point was coloured orange. Where two unfolding events assigned to *Hm*Fd occurred in the same trace, each data point was coloured red. There is no clear agreement of *Hm*Fd (orange) or *Hm*Fd (red). Figure B) shows a scatter diagram comparing the lengths of the measured L_c of I27$_3$-*Hm*Fd-I27 in 63 mM salt to the calculated contour length L_c. All figures show the measured contour length in green. Figures C) and D) show example traces that contain 4 I27 unfolding events have an full extension length between 166.25 and 190.00 nm. The traces are taken from experiments carried out on 8 different days. The polyprotein was in 63 mM sodium phosphate, pulled at a constant velocity of 400 nms$_{-1}$. Trace D) has unfolding events that could correspond to *Hm*Fd. The worm like chain fits for the I27 unfolding events are shown in red. All figures show data collected across 8 experiments carried out in 63 mM sodium phosphate, pulled at a constant velocity of 400 nm/s.

The unfolded extension of the full construct before detachment was recorded for each trace that included 4 unfolding events corresponding to I27. If *Hm*Fd is unfolded, as expected, the recorded extension should correspond to the contour length of the polyprotein, L_c. Using knowledge of the number of amino acids in each protein, the length of an amino acid[132,192], and details of the linkers in the polyprotein construct,

L_c of the construct can be calculated as follows:

$$L_c = (\text{length of amino acid (nm)} \times \text{total no. of amino acids}) \quad (3.1)$$

$$\text{Lower estimate} = (0.35 \times 475) = 166.25 \text{ nm} \quad (3.2)$$

$$\text{Upper estimate} = (0.40 \times 475) = 190.00 \text{ nm} \quad (3.3)$$

The recorded extension of the unfolded polyprotein was plotted on a scatter diagram, the upper and lower estimates of L_c were shown as dashed lines (figure 3.6 B)). All four traces fall within the boundaries, suggesting the full polyprotein is unfolding in all of these traces. The traces with the correct extension length were then examined. Two example traces are shown in figure 3.6 C) and D). Trace D) has unfolding events that could correspond to HmFd but no correlation between the peak to peak distances or unfolding forces can be seen.

3.3.3 SMFS experiments using I27$_3$-HmFd-I27 in 500 mM and 2000 mM salt

Further SMFS experiments, using the same experimental procedure, were carried out in a range of salt concentrations to see if there was any effect on the results. Since halophilic proteins tend to be unstable at low salt concentrations, it is likely HmFd would follow the same trend. Increasing the salt concentration may then increase the stability of the protein, which could then show mechanical stability. The two extra buffers used for these experiments were:

i) 500 mM NaCl, 63 mM sodium phosphate, pH 7.4

ii) 2000 mM NaCl, 63 mM sodium phosphate, pH 7.4

Analysis was carried out only using traces that included four I27 unfolding events. F_{UN} and p2p data for I27 in each experiment are recorded in tables 3.5 and 3.6.

Interestingly, the unfolding force of I27 was found to increase slightly with the increasing salt concentration. The values were plotted on scatter diagrams, showing F_{UN}

Salt conc.	Velocity	No. events	Mean F_{UN} (pN)	Median F_{UN} (pN)	SD
500 mM	400 nm/s	51	184	184	31
2000 mM	400 nm/s	10	185	188	17

Table 3.5: Unfolding force data and the standard deviation for I27 in I27$_3$-HmFd-I27 in varying salt concentrations. The data shown only includes traces that have 4 I27 unfolding events.

Salt conc.	Velocity	No. events	Mean p2p (nm)	Median p2p (nm)	SD
500 mM	400 nm/s	51	23.9	24.1	1.3
2000 mM	400 nm/s	10	23.6	23.9	1.3

Table 3.6: Peak to peak distance data and the standard deviation for I27 in I27$_3$-HmFd-I27 in varying salt concentrations.

for each protein unfolding event against the measured p2p (figure 3.7 A) and B)). The diagrams show clear clusters of data points for I27 (grey), centred around the reference I27 data[148]. Where only one unfolding event assigned to HmFd occurred in the same trace (in addition to the 4 I27 unfolding events), each data point was coloured orange. Where two unfolding events assigned to HmFd occurred in the same trace, each data point was coloured red. There is no clear agreement of HmFd (orange) or HmFd (red) in either diagram.

The unfolded extension of the full construct before detachment was recorded for each trace that included 4 unfolding events corresponding to I27. The measured length of the unfolded polyprotein was plotted on a scatter diagram (for each condition), the upper and lower estimates of L_c were shown as dashed lines, shown in figures 3.7 C) and D). Both traces fall within the boundaries in figure C), suggesting the full polyprotein is unfolding in 2000 mM salt. Only 3 of 13 traces fall within the boundary limits in figure D), suggesting there are other factors affecting the traces. Theses factors could include impurities in the sample, interference of the polyprotein by other polyproteins or non-specific interactions. The traces with the correct extension length were then examined. One example trace in 2000 mM salt is shown in figure 3.7 E). The traces has two unfolding events that could correspond to HmFd but no correlation between the peak to peak distances or unfolding forces can be seen.
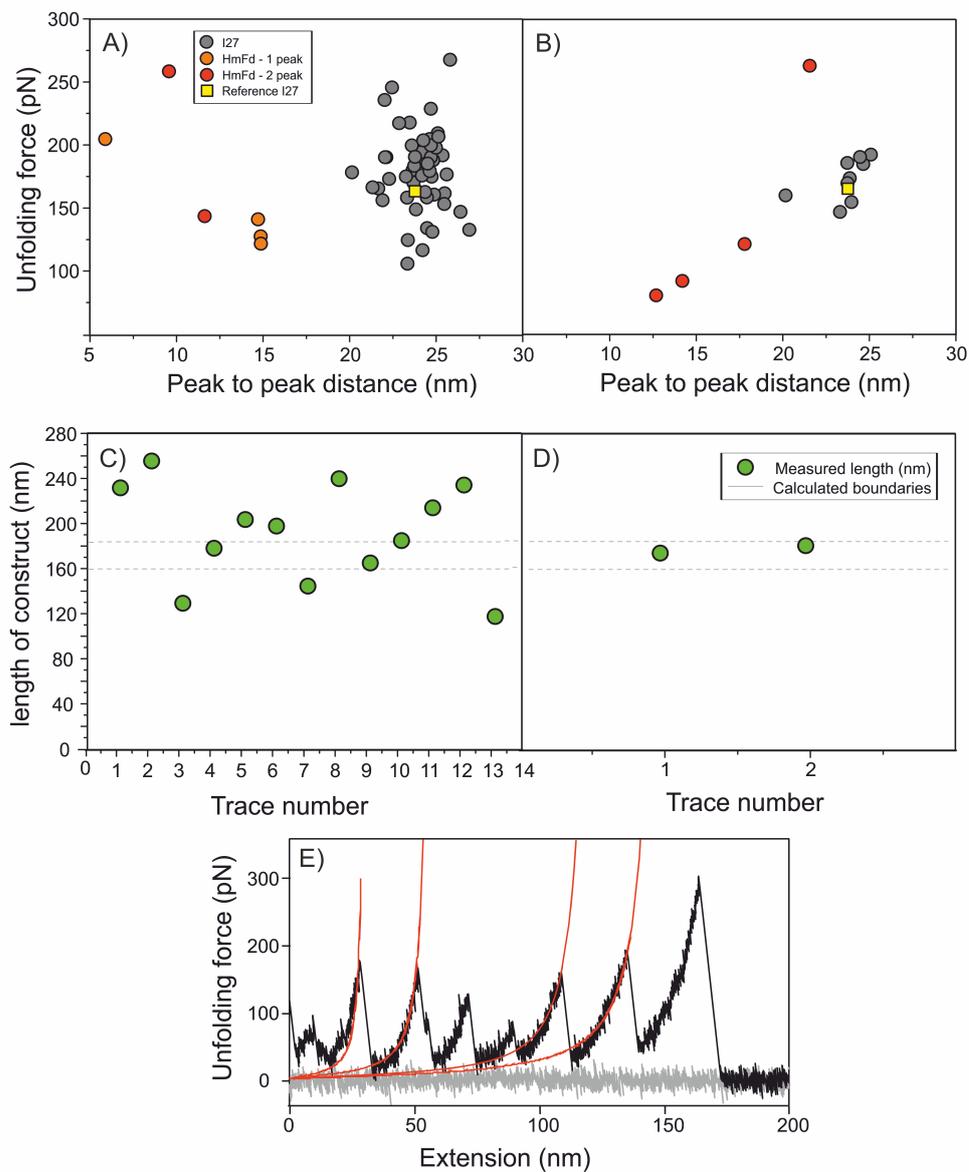
Figure 3.7: The data analysis of SMFS experiments using I27₃-*Hm*Fd-I27 in 500 mM and 2000 mM salt. Figures A) and B) show the scatter diagrams of the unfolding force vs. peak to peak distance for I27$_3$-*Hm*Fd-I27 in 500 mM and 2000 mM salt respectively. The figures both show I27 in grey, clustered around the reference value for I27 in yellow [148]. Where only one unfolding event assigned to *Hm*Fd occurred in the same trace (in addition to the 4 I27 unfolding events), each data point was coloured orange. Where two unfolding events assigned to *Hm*Fd occurred in the same trace, each data point was coloured red. Figures C) and D) show scatter diagrams comparing the measured L_c I27$_3$-*Hm*Fd-I27 in 500 mM and 2000 mM salt to the calculated contour length L_c. All figures show the measured contour length in green. Figure E) shows an example trace that contains 4 I27 unfolding events and has a full extension length between 166.25 and 190.00 nm. The trace has two unfolding events that could correspond to *Hm*Fd. The worm like chain fits for the I27 unfolding events are shown in red.

100

3.3.4 SMFS experiments using I27$_3$-HmFd-I27 at 2000 nm/s

Further SMFS experiments, using the same experimental procedure, were carried out at a different pulling velocity. At a faster pulling velocity, force is applied faster to the protein. This results in a higher unfolding force as the likelihood of unfolding due to thermal fluctuations is decreased. The rate of the increase in unfolding force with pulling velocity can be different for each protein. If I27 and HmFd have the same unfolding force at 400 nm/s, therefore masking the HmFd unfolding force, they may have different unfolding forces at 2000 nm/s[193]. Analysis was carried out only using traces that included four I27 unfolding events. The F$_{UN}$ and p2p data for I27 are recorded in tables 3.7 and 3.8.

Salt conc.	Velocity	No. events	Mean F$_{UN}$ (pN)	Median F$_{UN}$ (pN)	SD
63 mM	2000 nm/s	16	208	209	34

Table 3.7: Unfolding force data and the standard deviation for I27 in I27$_3$-HmFd-I27 at 2 pulling velocities. The data shown only includes traces that have 4 I27 unfolding events.

Salt conc.	Velocity	No. events	Mean p2p (nm)	Median p2p (nm)	SD
63 mM	2000 nm/s	16	23.6	23.7	1.5

Table 3.8: Peak to peak distance data and the standard deviation for I27 in I27$_3$-HmFd-I27 at 2 pulling velocities.

The values were plotted on a scatter diagram showing F$_{UN}$ against p2p (figure 3.8 A)). The diagram shows not enough data is available to clearly show the mechanical unfolding behaviour of HmFd, as described for the previous experiments.

The unfolded extension of the full construct before detachment was recorded for each trace that included 4 unfolding events corresponding to I27. The measured length of the unfolded polyprotein was plotted on a scatter diagram, the upper and lower estimates of L_c were shown as dashed lines, figure 3.8 B). None of the traces fall within the boundaries suggesting there are other factors affecting the traces. These factors could include impurities in the sample, interference of the polyprotein by other polyproteins or non-specific interactions.
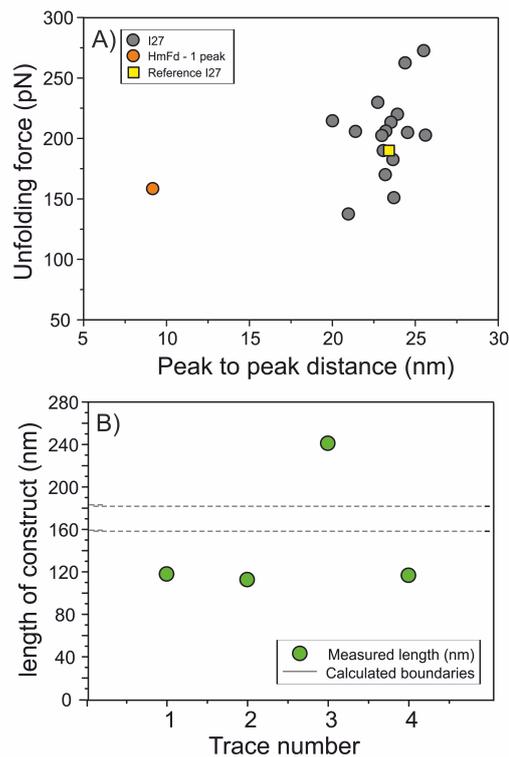
Figure 3.8: The data analysis of SMFS experiments using I27$_3$-*Hm*Fd-I27 at 2000 nm/s. Figure A) shows the scatter diagram of the unfolding force vs. peak to peak distance for I27$_3$-*Hm*Fd-I27 at 2000 nm/s. The figure shows I27 in grey, clustered around the reference value for I27 in yellow [148]. Unfolding events, interpreted as *Hm*Fd, are shown in orange. Figure B) shows a scatter diagram comparing lengths of the measured L_c of I27$_3$-*Hm*Fd-I27 at 2000 nm/s to the calculated L_c. The figure shows the measured contour length in green.

3.3.5 Summary

The results show it was possible to obtain the expected force-extension data for the reference protein, I27. It is likely that HmFd does have a mechanical resistance to force (figure 3.5). Full characterisation of the unfolding force and peak to peak distance was consequently unsuccessful for HmFd. The possible unfolding pathways of HmFd will be discussed further in section 6.1.

Conclusion: More repeats of HmFd are needed in the polyprotein construct in order generate a larger amount of data to carry out effective mechanical unfolding characterisation.

3.4 Production of a the hetero-polyprotein, (I27-HmFd)$_3$-I27

As it was not possible to identify the mechanical fingerprint of HmFd from the construct containing only one HmFd, a polyprotein containing more repeats of HmFd would make identification of the fingerprint easier by increasing the amount of data collected for HmFd in each AFM experiment. This would also remove HmFd from being close to the gold surface where nonspecific interactions can interfere with the protein unfolding. The design of the polyprotein chimera chosen is shown in figure 3.9 and includes 4 repeats of I27 and 3 repeats of HmFd. This chimera polyprotein design has been used to characterise other previously unstudied proteins[29,146,148]. This section will describe the experiments completed in the attempt to synthesise the polyprotein chimera containing I27 and HmFd. Further explanation of the techniques used can be found in chapter 2, including a basic description of each step and the protocols used. An overview of the synthesis method can be seen in figure 3.10.
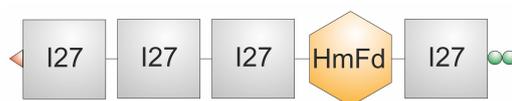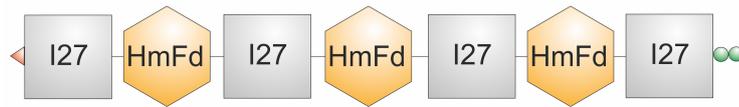
Figure 3.9: Schematic of the (I27-*Hm*Fd)₃-I27 construct. The construct has a His-tag, shown as a red triangle, at the N-terminal end and 2 cysteine residues, shown as green circles, at the C-terminal end.
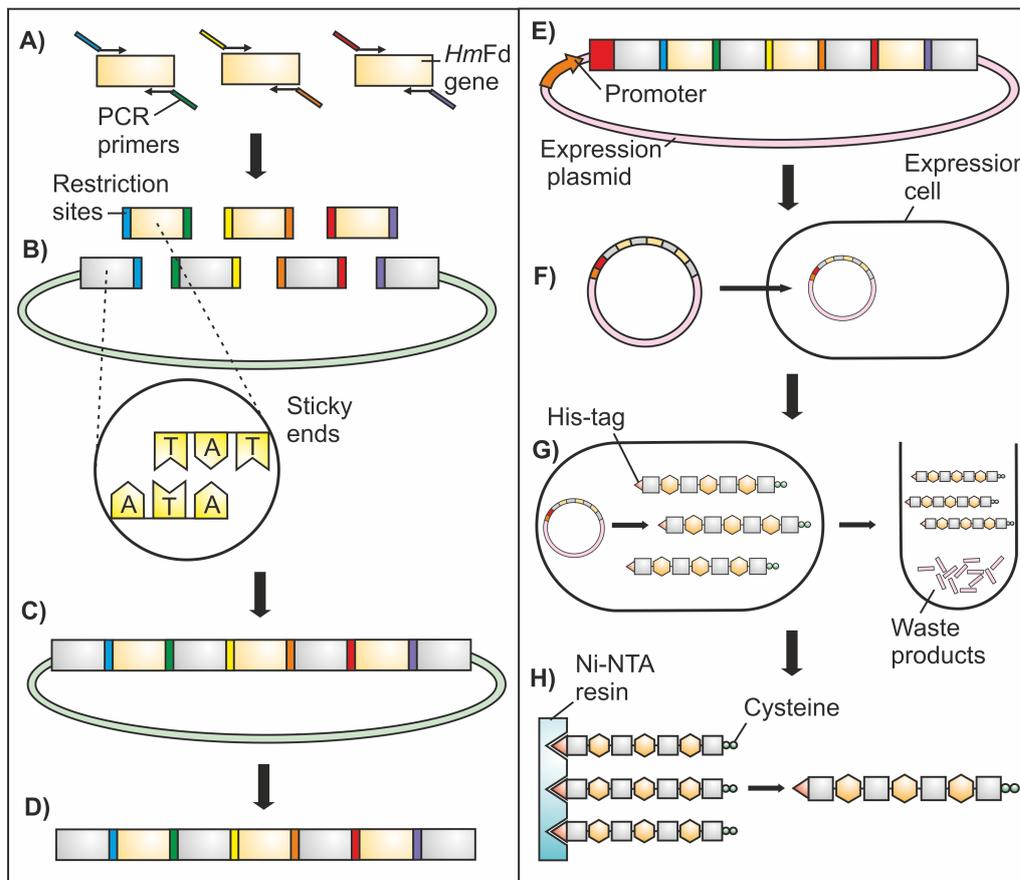
Figure 3.10: Schematic of the synthesis of (I27-*Hm*Fd)₃-I27. Where A) represents the PCR reactions to add the correct restriction sites to the *Hm*Fd inserts, B) represents digestion of the *Hm*Fd inserts and the plasmid to create sticky ends, C) represents the ligation of the each *Hm*Fd gene into the plasmid, D) represents the digestion of the whole polyprotein gene from the plasmid, E) represents the ligation of the whole polyprotein gene into an expression plasmid, F) represents the transformation of the plasmid into an expression *E. Coli* strain, G) represents the expression of the polyprotein, H) represents purification of the polyprotein. The restriction sites are shown in different colours, blue - SpeI, green - BssHII, yellow - ApaI, orange - SacI, red - AatII and purple - SacII.

3.4.1   Molecular biology for (I27-HmFd)$_3$-I27

The pET3a-(I27)$_7$ and pCR2.1-(I27)$_3$-HmFd-I27 plasmids were provided by Dr. David Brockwell. A 500 μL sample of the plasmid, pET3a-(I27)$_7$, was produced by first transforming 2 μL plasmid into XL1 Blue competent cells (§2.3.1), which were grown in colonies on agar plates. A colony was picked, grown in 50 mL LB medium, containing 100 μg/mL ampicillin, and DNA was purified using a QIAprep® Spin Midiprep Kit (§2.2.9).

The HmFd gene to be inserted into position 4 of the polyprotein gene was produced from the pCR2.1-(I27)$_3$-HmFd-I27 plasmid using the following procedure. The cassette was isolated by digesting the pCR2.1-(I27)$_3$-HmFd-I27 plasmid at position 4 using restriction enzymes ApaI and SacI using the protocol described in §2.2.4. The cassette was separated from the unwanted DNA by agarose gel electrophoresis, the procedure for which is outlined in §2.2.2, and subsequently extracted from the gel, as described in §2.2.3. The cassette was then ready to be inserted into the plasmid.
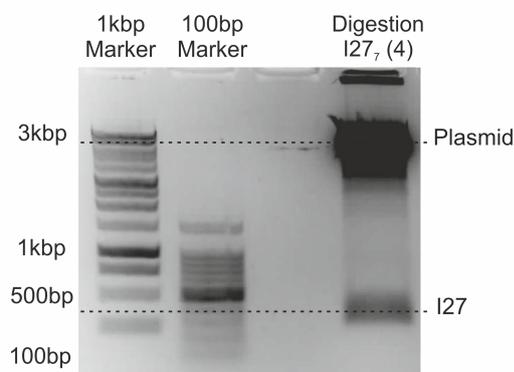
Figure 3.11: The plasmid pET3a-(I27)$_7$ was digested at position 4 using restriction enzymes ApaI and SacI. Agarose gel electrophoresis was used to purify and check the size of the cassette. The gel shows a correctly sized product for I27 - 279 bp

The (I27)$_7$ vector was digested using the same restriction enzymes as above and dephosphorylated to prevent self-ligation (§2.2.5). Agarose gel electrophoresis was used to isolate and check the size of DNA removed. The gel, figure 3.11, shows the excised

DNA to be about the correct size for I27, which is 279 bp. This shows one of the I27 genes was correctly removed from the pET3a-(I27)$_7$ plasmid.

Next, the HmFd insert for position 4 was ligated into the digested pET3a-(I27)$_7$ plasmid (see §2.2.6). A control ligation was also performed using all components in the reaction mixture, except for the insert HmFd DNA. The resulting plasmid pET3a-(I27)$_3$-HmFd-(I27)$_3$ was transformed into XL1 Blue Supercompetent cells, as well as the control reaction, and spread on agar plates. After incubating overnight, colonies were found on the plates with cells that had been transformed with the ligation reaction mixture. No colonies were found on plates with cells transformed with the negative control ligation mixture. 6 colonies were picked, grown in LB medium and the DNA was purified for each colony using a QIAprep® Spin Miniprep Kit (§2.2.8). To determine whether the vector now contained the desired insert, 1μg of plasmid from each colony was digested with the restriction enzymes ApaI and SacI. The results were analysed using agarose gel electrophoresis, shown in figure 3.12.
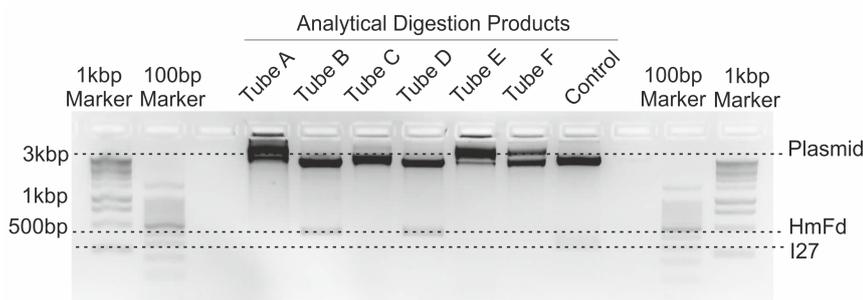
Figure 3.12: The plasmid pET3a-(I27)$_3$-HmFd-(I27)$_3$ was digested at position 4 using restriction enzymes ApaI and SacI. The gel shows a successful ligation product with the correctly sized band for HmFd - 420 bp in tubes B and D. Unsuccessful ligation products with the correctly sized band for I27 - 279 bp in tube C and the control containing no insert DNA.

Polymerase Chain Reaction (PCR) was used to introduce unique restriction enzyme sites to the 5$'$ and 3$'$ ends of the HmFd cassettes for positions 2 and 6. The restriction enzymes of choice were SpeI and BssHII for position 2 and AatII and SacII for position 6 (§2.2.1). The template plasmid used for PCR was the original plasmid (I27)$_3$-HmFd-I27. The agarose gel of the PCR products, used to separate the products from the rest of the

reaction mixture, is shown in figure 3.13.
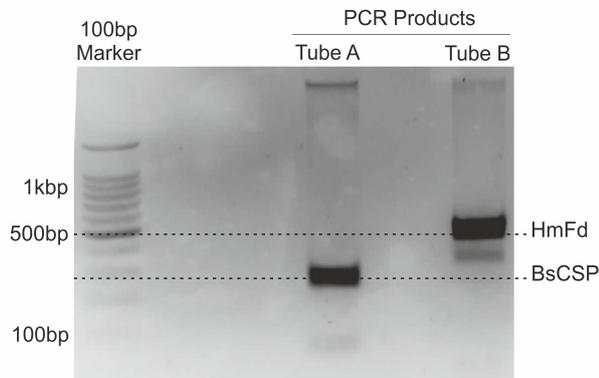
Figure 3.13: Agarose gel electrophoresis was used to purify and check the size of the PCR product. PCR was used to add short linker regions to each end of the *Hm*Fd insert, encoding the recognised sequence for restriction enzymes SpeI and BssHII (for position 2). The BsCSP gene in tube A is unrelated to this project.

The digestion, using the appropriate restriction enzymes, and ligation process was then repeated for position 6 to form $(I27)_3\text{-}(Hm\text{Fd-}I27)_2$, followed by 2 to form $(I27\text{-}Hm\text{Fd})_3\text{-}I27$. After each ligation process, the DNA was verified firstly using an analytical digest and then by sequencing. The agarose gel obtained for the analytical digestion of $(I27)_3\text{-}(Hm\text{Fd-}I27)_2$ in the 6th position is shown in figure 3.14. The plasmid pET3a-$(I27\text{-}Hm\text{Fd})_3\text{-}I27$ was then ready to be used for expression of the $(I27\text{-}Hm\text{Fd})_3\text{-}I27$ protein.
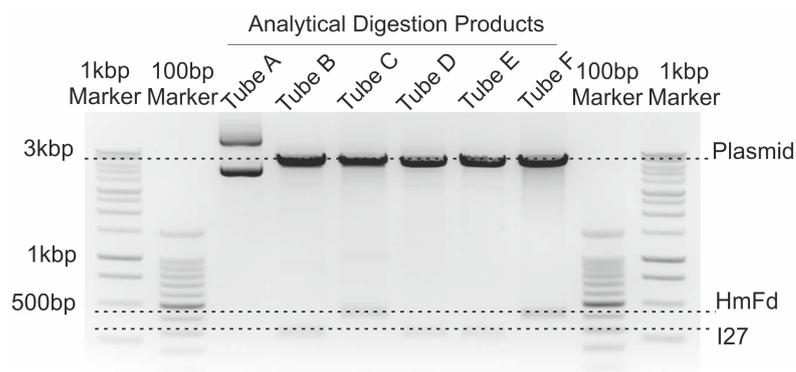
Figure 3.14: The plasmid pET3a-$(I27)_3\text{-}(Hm\text{Fd-}I27)_2$ was digested at position 6 using restriction enzymes SpeI and BssHII. The gel shows a successful ligation product with the correctly sized band for *Hm*Fd - 420 bp in tubes C and F. Unsuccessful ligation products with the correctly sized band for I27 - 279 bp in tubes B, C and E

3.4.2 Expression of the hetero-polyprotein, (I27-*Hm*Fd)$_3$-I27

The first step was to test for over-expression of the desired polyprotein on a small scale (§2.3.2). Once the final plasmid pET3a-(I27-*Hm*Fd)$_3$-I27 was confirmed by sequencing, it was transformed into *E.*coli BLR[DE3] cells (§2.3.1). 1 mL of an overnight culture of the cells was used to inoculate 100 mL LB medium and the culture was incubated at 37°C, 200 rpm. Protein expression was induced with addition of 1 mM IPTG at an optical density at 600 nm (OD$_{600}$) of 0.4. After further incubation for 3 hours, the cells were harvested by centrifugation and analysed using SDS-PAGE (§2.3.2). This polyprotein has an N-terminal hexa-histidine tag, therefore the cells were lysed and the resultant polyprotein was purified on Ni-NTA affinity resin (§2.3.3). The protein production and purification stages were analysed using SDS-PAGE at each stage, results are shown in figure 3.15.
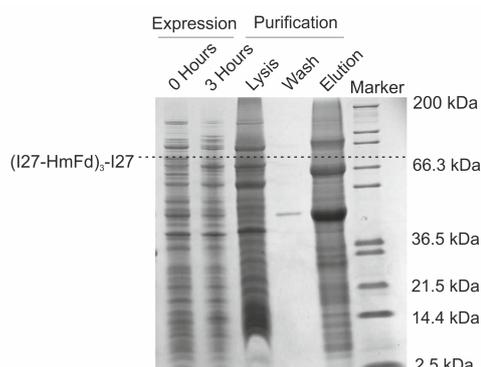
Figure 3.15: Small scale expression and purification of (I27-*Hm*Fd)$_3$-I27, analysed by SDS-PAGE with Coomassie blue staining. The gel shows no band with significant intensity at 87.7 kDa which would correspond to (I27-*Hm*Fd)$_3$-I27 in any of the expression or purification steps.

To further analyse the expression of (I27-*Hm*Fd)$_3$-I27, western blotting was carried out. While SDS-PAGE analysis with Coomassie blue staining shows all proteins in a solution, Western blotting uses antibodies to selectively recognise and only stain proteins with a specific amino acid sequence, e.g. a hexa-histidine tag (§2.6.2). The small scale expression and purification steps were repeated and analysed using SDS-PAGE with Coomassie blue staining and western blotting. The western blot, shown in figure 3.17

showed signs of protein degradation in all samples taken at different timepoints through $(I27\text{-}Hm\text{Fd})_3\text{-}I27$ expression. Each band must contain the histadine tag to be labelled with the antibody and, therefore, must come from the expressed polyprotein breaking down into smaller segments. Least degradation of the product can be seen at an expression time of 0.5 hours. The corresponding Coomassie stained gel is shown in figure 3.16. It was concluded that shorter expression times were more desirable, although the reduced amount of growth time for *E. coli* would produce less protein.

In order to try and increase the protein yield over a shorter timescale, the OD_{600} at which protein expression is induced was varied. By increasing the OD_{600} of induction, the amount of cells will be increased. The expression at induction OD_{600} values of 0.7, 1, 1.5 and 3 was analysed using SDS-PAGE and western blotting, shown in figures 3.18 and 3.19 respectively. Although the final OD_{600} after incubation for 0.5 hours was highest for the induction OD_{600} of 3, the intensity of the $(I27\text{-}Hm\text{Fd})_3\text{-}I27$ band was significantly reduced in the western blot.

To try and reduce the amount of degradation of the protein, the expression process was carried out with 5 x 100 mL samples at a temperature of 16 °C and incubated for 3 hours. The subsequent purification of one of the 100 mL samples was carried out at 4 °C and a protein inhibitor cocktail was used in all buffers. No evidence of the expression of $(I27\text{-}Hm\text{Fd})_3\text{-}I27$ can been using SDS-PAGE analysis, shown in figure 3.20. It was therefore concluded that the $(I27\text{-}Hm\text{Fd})_3\text{-}I27$ polyprotein construct could not be obtained using this method in sufficient quantities and another production route should be used.

The plasmid containing 2 repeats of $Hm\text{Fd}$, in the fourth and sixth positions, created during the initial molecular biology steps (§3.4.1), was then transformed into BLR[DE3] cells and 2 test cultures of 100 mL were grown. In one test culture, expression was induced with 1 mM IPTG at OD_{600} of 0.4 which was harvested after 3 hours. In the other test culture, expression was induced at OD_{600} of 0.7, which was harvested after 0.5 hours and the resulting polyprotein was purified. The protein production and purification stages were analysed using SDS-PAGE (data not shown). The analysis showed clear
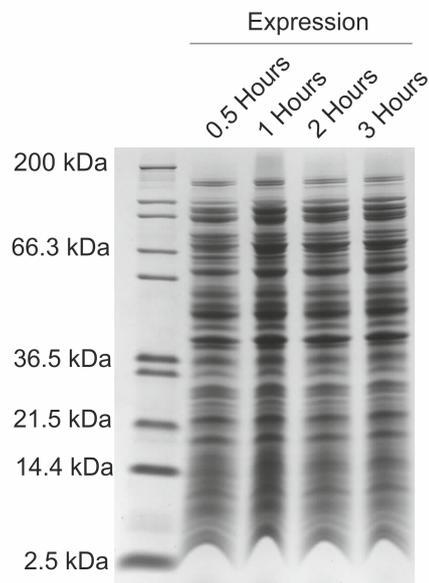
Figure 3.16: Small scale expression and purification of (I27-*Hm*Fd)₃-I27, analysed by SDS-PAGE with Coomassie blue staining. Again, the gel shows no band with significant intensity at 87.7 kDa which would correspond to (I27-*Hm*Fd)₃-I27 in any of the expression or purification steps.
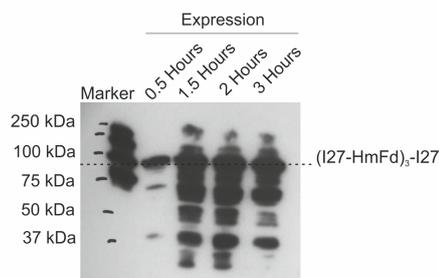
Figure 3.17: Small scale expression of (I27-*Hm*Fd)₃-I27, analysed by Western blotting. The gel shows a band with significant intensity at 87.7 kDa, which corresponds to (I27-*Hm*Fd)₃-I27 in all of the expression time points. The figure showed signs of protein degradation in all expression time points. Least degradation of the product can be seen at an expression time of 0.5 hours.
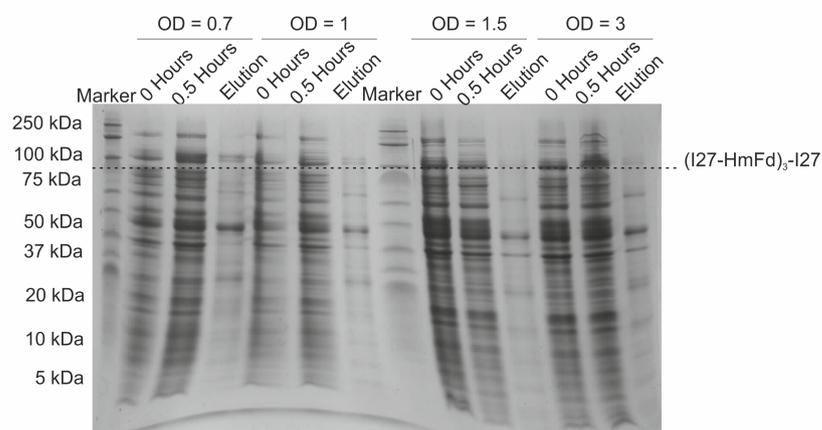
Figure 3.18: Four small scale expressions and purifications of (I27-*Hm*Fd)$_3$-I27 with stepwise increasing induction optical densities, analysed by SDS-PAGE with Coomassie blue staining. Again, the gel shows no band with significant intensity at 87.7 kDa which would correspond to (I27-*Hm*Fd)$_3$-I27 in any of the expression or purification steps.
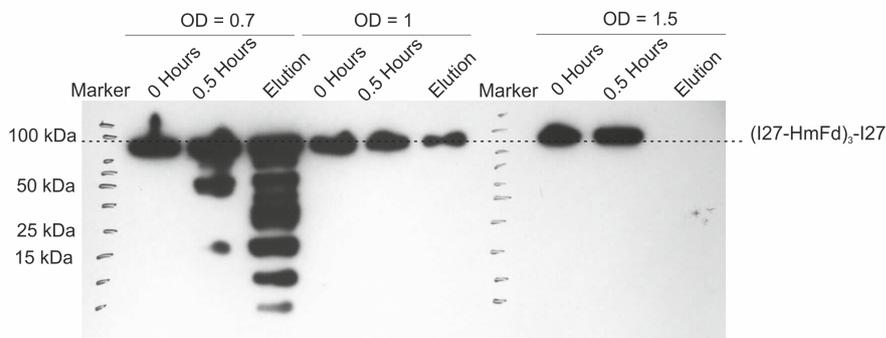
Figure 3.19: Four small scale expressions and purifications of (I27-*Hm*Fd)$_3$-I27 with stepwise increasing induction optical densities, analysed by Western blotting. The gel shows a band with significant intensity at 87.7 kDa, which corresponds to (I27-*Hm*Fd)$_3$-I27 in all of the expression expressions at induction OD$_{600}$s 0.7, 1 and 1.5 with the largest band at OD$_{600}$ of 0.7. The figure showed signs of protein degradation after 0.5 hours.
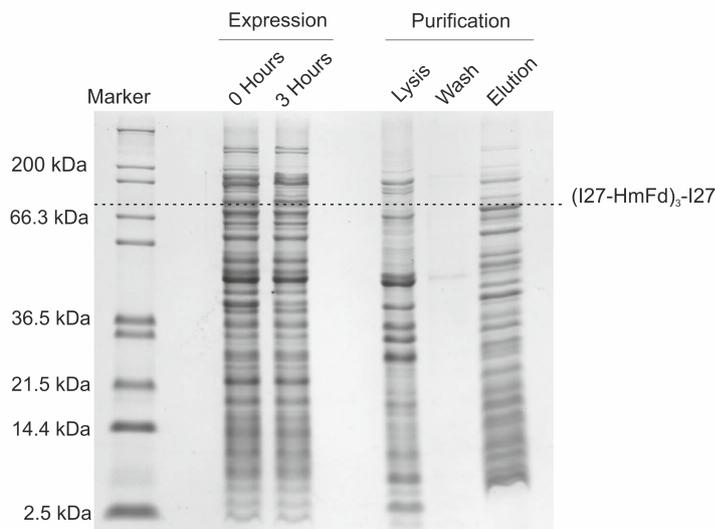
Figure 3.20: Small scale expression and purification of (I27-*Hm*Fd)$_3$-I27 at 16 °C, analysed by SDS-PAGE with Coomassie blue staining. The gel shows no band with significant intensity at 87.7 kDa which would correspond to (I27-*Hm*Fd)$_3$-I27 in any of the expression or purification steps.

impurities, which were again likely to be from degradation.

3.4.3 Summary

The results presented show attempts to synthesise a hetero-polyprotein construct containing four repeats of I27 protein and three repeats of *Hm*Fd. Construction of the hetero construct DNA was successful but the polyprotein construct could not be obtained using this method in sufficient quantities and another production route should be used.

Conclusion: An alternative technique is needed to produce a polyprotein containing more than one repeat of *Hm*Fd.

3.5 Production of a homo-polyprotein of *Hm*Fd using maltose binding protein

The results will be presented for the synthesis of a polyprotein of *Hm*Fd, shown in figure 3.21, using a fusion protein, Maltose binding protein (MBP)[194]. MBP facilitates the solubility, purification and proper folding of its fusion partners. A wide variety of polypeptides that are prone to aggregation have been recovered in soluble form as MBP fusion proteins[195,196]. After cleavage of *Hm*Fd from the fusion protein, cysteines, added to each end of *Hm*Fd allow for polymeristion via thiol coupling to a maleimide crosslinker, shown to be a successful method for the polymerisation of Rubredoxin[197]. An overview of the synthesis method can be seen in figure 3.22.

Figure 3.21: Schematic of the poly-*Hm*Fd construct. The construct is polymerised by N- and C-terminal cysteine residues, shown as green circles, at the C-terminal end.

3.5.1 Molecular biology for poly-*Hm*Fd

The *Hm*Fd insert was modified and amplified from the pET3a plasmid by PCR using Phusion®polymerase (NEB). Primers were designed to create an NdeI restriction site, a TEV protease recognition site and a cysteine at the 5 end of the *Hm*Fd insert and a cysteine upstream of a BamHI restriction site at the 3 end. The primers used are shown in table A.1. The primers and other PCR products were separated from the desired insert using agarose (1.5% w/v) gel electrophoresis, shown in figure 3.23.

The *Hm*Fd insert was digested using the restriction enzymes BamHI and NdeI. The vector, pMAL-c5x, was digested in an identical manner and was then 5 dephosphorylated. Agarose gel electrophoresis was used to separate the digested insert and the digested 5 dephosphorylated vector from other cleavage products. The gel is shown in figure 3.24.
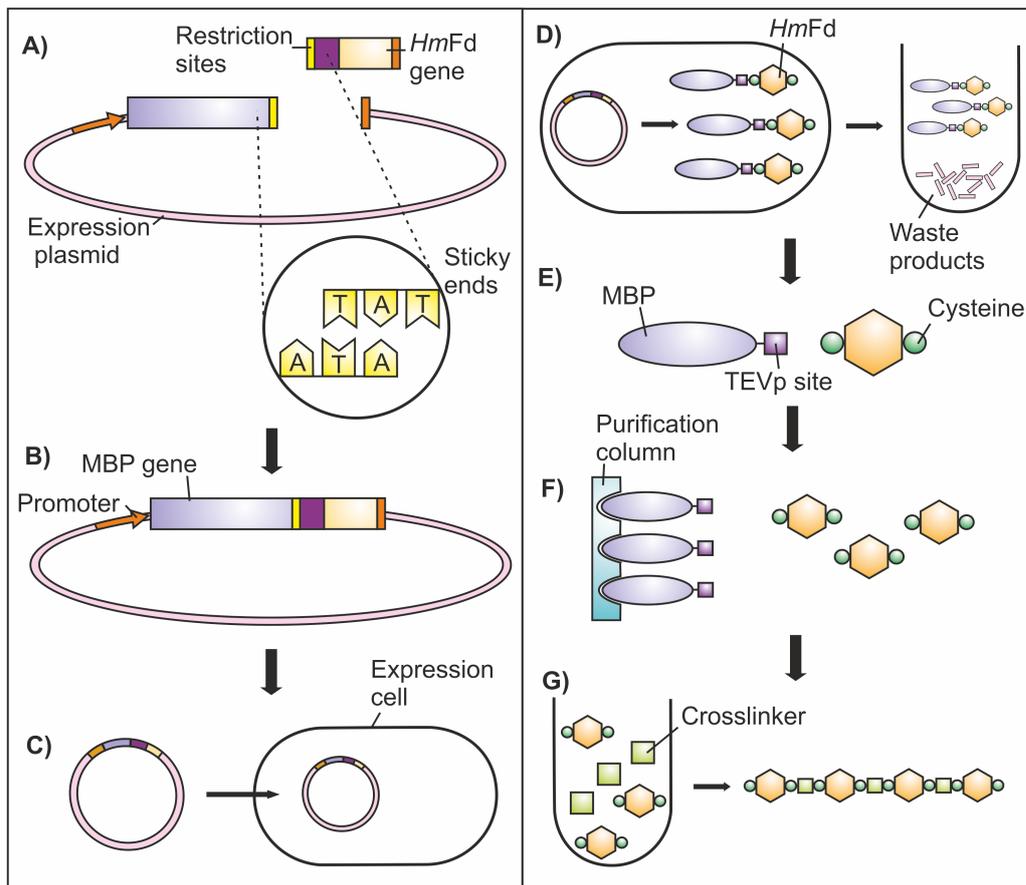
Figure 3.22: Schematic of the synthesis of poly-*Hm*Fd using MBP. Where A) represents the digestion of the *Hm*Fd insert and the plasmid to create sticky ends, B) represents the ligation of the *Hm*Fd gene into the plasmid, C) represents the transformation of the plasmid into an expression *E. Coli* strain, D) represents the expression of MBP-*Hm*Fd protein, E) represents cleavage of the *Hm*Fd monomer from MBP, F) reresents the purification of the *Hm*Fd monomer and G) represents the crosslinking of *Hm*Fd monomer to form a polymer. The restriction sites are shown in different colours, yellow - NdeI, orange - BamHI.
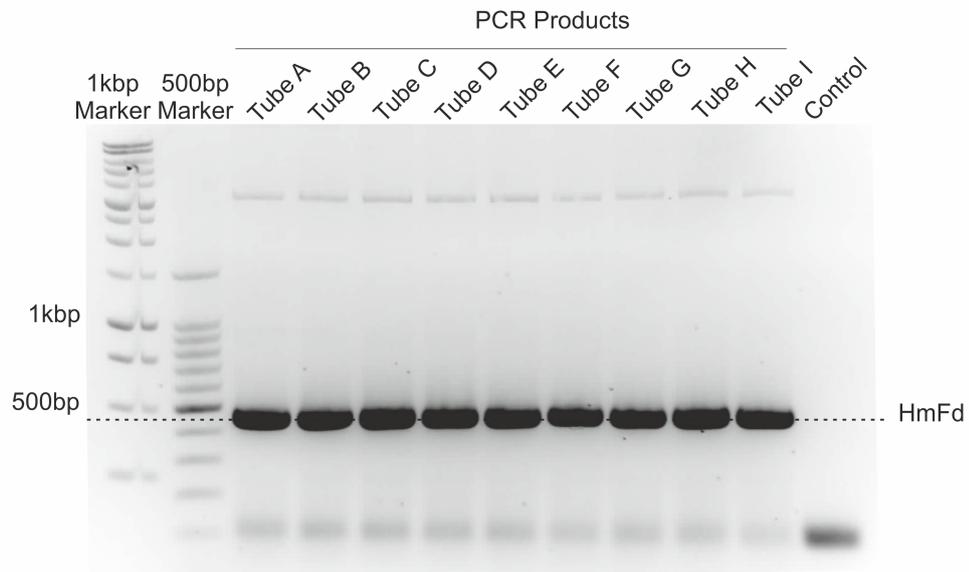
Figure 3.23: Agarose gel electrophoresis was used to clean the PCR product for *Hm*Fd. The gel shows the corect size band for *Hm*Fd - 420 bp in all products and no band corresponding to *Hm*Fd in the control lane (no template DNA).
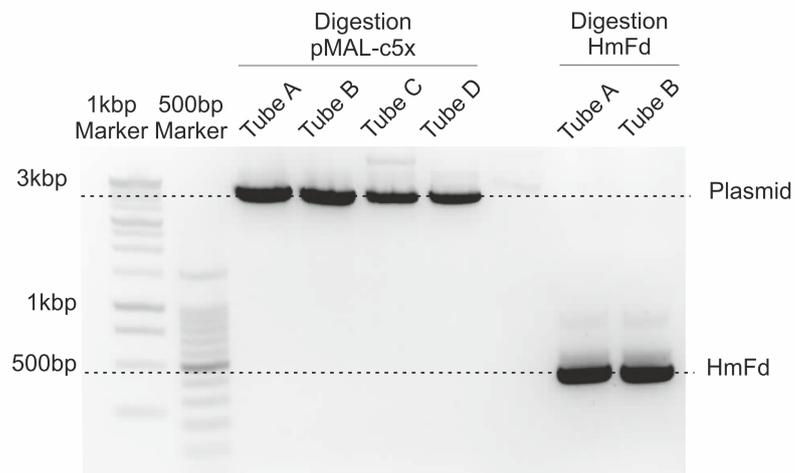
Figure 3.24: The plasmid pMAL-c5x and insert *Hm*Fd digest were digested using restriction enzymes NdeI and BamHI. Agarose gel electrophoresis was used to check the size of DNA. The gel shows a correctly sized product for the the plasmid - 5677 bp (tubes A-D) and *Hm*Fd - 420 bp.

A reaction was set up to ligate the digested *Hm*Fd insert into 5 dephosphorylated pMAL-c5x, using T4 quick ligase (Promega). In addition, a negative control containing no insert DNA was also assembled. 7 μL of these reactions were transformed into *E.coli* XL1-Blue supercompetent cells. DNA was purified using a QIAprep® Spin Miniprep Kit from nine single colonies picked from agar plates after transformation and used to innoculate 10 mL LB medium. Each sample was digested with NdeI and BamHI to check if the insertion of *Hm*Fd into pMAL-c5x had been successful. These digests were analysed by agarose gel electrophoresis, shown in figure 3.25. The ligation of *Hm*Fd into pMAL-c5x appeared to be successful for all plasmid samples, indicated by presence of a band at 417 bp. The success of cloning the pMAL-c5x-*Hm*Fd construct was subsequently confirmed by DNA sequencing.
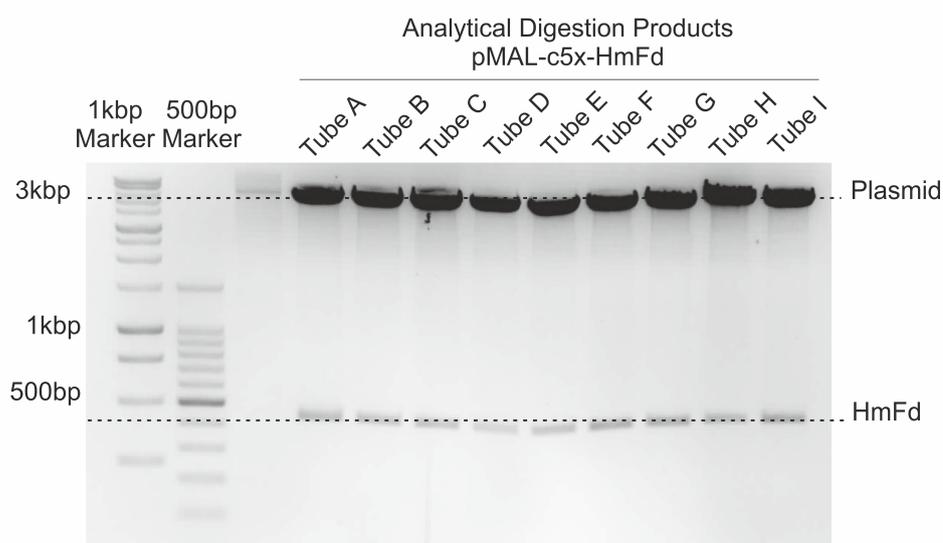
Figure 3.25: The plasmid pMAL-c5x-*Hm*Fd was digested using restriction enzymes NdeI and BamHI. The gel shows a successful ligation product with the correctly sized band for *Hm*Fd - 420 bp in all tubes

3.5.2 Expression of MBP-*Hm*Fd

A test expression of the pMAL-c5x-*Hm*Fd construct was carried out after transformation of the plasmid into *E.coli* BL21 expression cells. The trials were performed at a 100 mL

scale in LB medium, TB medium and M9 medium containing either no salt, 0.5 M NaCl, 0.2 mM $(NH_2)_2Fe(SO_4)_2$, or both 0.5 M NaCl and 0.2 mM $(NH_2)_2Fe(SO_4)_2$ in 250 mL conical flasks. Expression of the constructs was induced by IPTG at an OD_{600} of 0.4. Subsequent SDS-PAGE analysis (figures 3.26, 3.27, 3.28 and 3.29) showed a band corresponding to the expected size of MBP-HmFd (58.8 kDa) in the soluble lysate fraction after induction of MBP-HmFd in all tested media, which was not present in the lysate before induction.

The next step was to carry out a large scale expression with 10 x 1 L TB medium in 2.5L conical flasks. TB medium with 0.5M NaCl was chosen as the medium due to the large expression band at 3 hours shown in figure 3.27. The expression was unsuccessful as the cells did not grow fast enough, i.e. the OD_{600} did not reach 0.4 within 10 hours.

A test large scale expression was then set up using the following conditions:

- 1 L TB medium with 0.5 M NaCl, 2.5 L conical flask
- 1 L TB medium in a 2.5 L conical flask
- 0.5 L TB medium with 0.5 M NaCl, 2.5 L conical flask
- 0.5 L TB medium with 0.5 M NaCl, 2.5 L baffled flask

The only successful growth was that of 1 L TB medium in a 2.5 L conical flask without salt. It was concluded that the salt must affect the growth of the cells on a large scale.

Another test large scale expression was then set up using 1 L TB medium in a 2.5 L conical flask without salt with induction at an OD_{600} of 0.7. This time analysing the expression products at each hour using SDS-PAGE to find the most ideal time for harvesting the cells. The results showed expression was not successful as no band at 58.8 kDa, corresponiding to MBP-HmFd was present in any of the expression lanes (figure 3.30). BL21 cells were, therefore, shown to cease expression the protein, likely due to the batch of cells used rather than the general properties of BL21 cells.

The pMAL-c5x-HmFd construct plasmid and the original pMAL-c5x plasmid, were later transformed into JM83 expression cells. Expression trials were set up using 100 ml LB medium and expression was induced at an OD_{600} of 0.7, collecting samples at
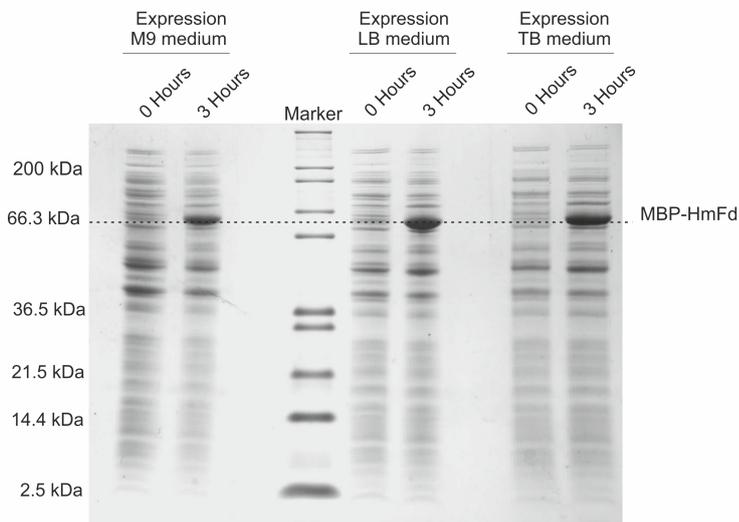
Figure 3.26: Three small scale expressions and purifications of MBP-*Hm*Fd in M9, LB and TB media, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 58.8 kDa which would correspond to MBP-*Hm*Fd after 3 hours of incubation after induction.
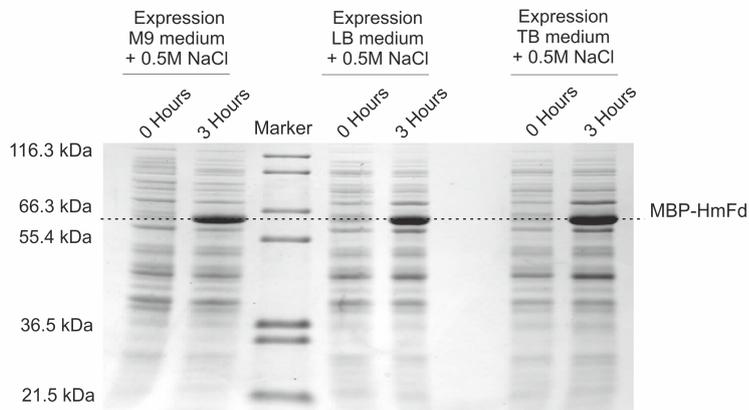
Figure 3.27: Three small scale expressions of MBP-*Hm*Fd in M9, LB and TB media with 0.5 NaCl, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 58.8 kDa which would correspond to MBP-*Hm*Fd after 3 hours of incubation after induction.
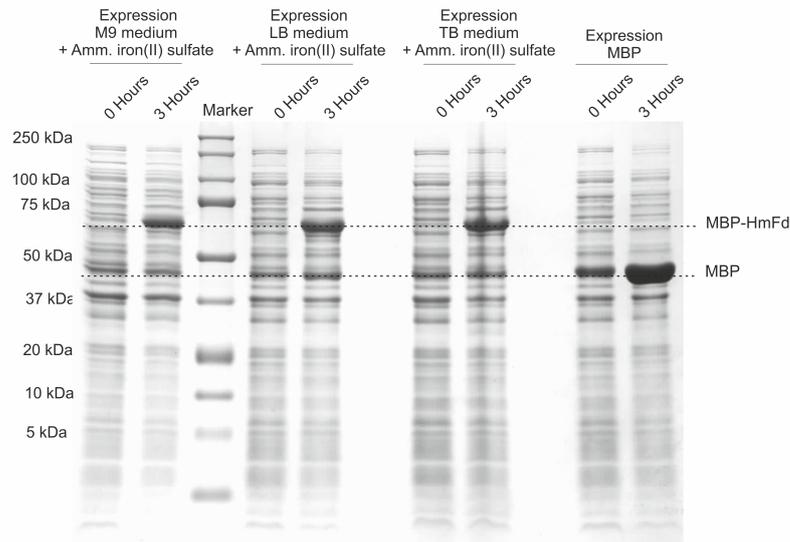
Figure 3.28: Three small scale expressions of MBP-*Hm*Fd in M9, LB and TB media with 0.2 mM Ammonium iron(II) sulfate, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band of significant intensity at 58.8 kDa which would correspond to MBP-*Hm*Fd after 3 hours of incubation after induction. This is shown next to a control experiment showing the expression of MBP from the pMAL-c5x plasmid.
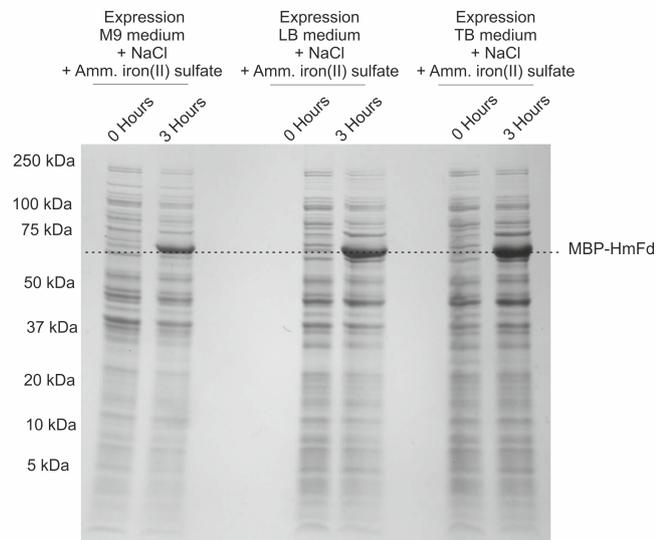
Figure 3.29: Three small scale expressions of MBP-*Hm*Fd in M9, LB and TB media with 0.5 NaCl and 0.2 mM Ammonium iron(II) sulfate, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 58.8 kDa which would correspond to MBP-*Hm*Fd after 3 hours of incubation after induction.
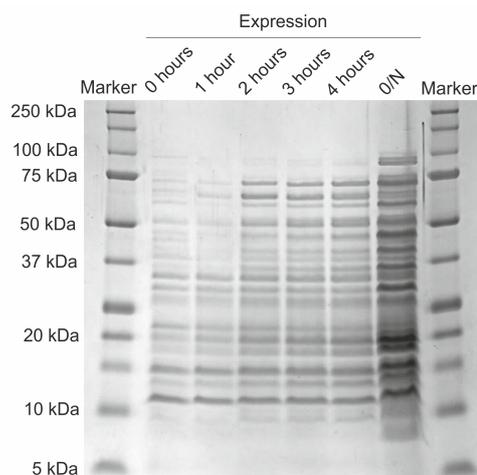
Figure 3.30: A 1 L test expression of MBP-*Hm*Fd in TB medium, monitored at each hour, analysed by SDS-PAGE with Coomassie blue staining. The gel shows no band at 58.8 kDa which would correspond to MBP-*Hm*Fd.

each hour of incubation afterwards. SDS-PAGE analysis of the expression is shown in figure 3.31. A clear band in the induced lysate can be seen between 50 and 75 kDa that is not present in the lysate before induction. This indicates successful expression of MBP-*Hm*Fd (58.8 kDa). The control expression shows successful production of MBP (43.8 kDa) indicated by a band between 50 and 37 kDa. SDS-PAGE analysis was then used to monitor the expression of MBP-*Hm*Fd at each hour after induction, shown in figure 3.32. The strongest MBP-*Hm*Fd band is after 2 hours of induction. This was chosen as the ideal time to harvest cells.

A 10 L expression of pMAL-c5x-*Hm*Fd in JM83 expression cells was set up using 10 x 1 L TB medium in 2.5 L conical flasks. Cultures were allowed to grow for 2 hours after induction before harvesting by centrifugation. SDS-PAGE analysis (figure 3.33) of the lysate revealed this expression was successful, showing the expected 58.8 kDa band in the lysate after expression for 2 hours.

The cell pellet was resuspended and cell lysis was carried out using a cell disrupter system. MBP-*Hm*Fd was purified from the soluble fraction by anionic exchange chromatography using a 5 mL Hi-Trap Q column, with the resulting elution profile shown in figure B.1 and SDS-PAGE of the fractions show in figure 3.34.

120
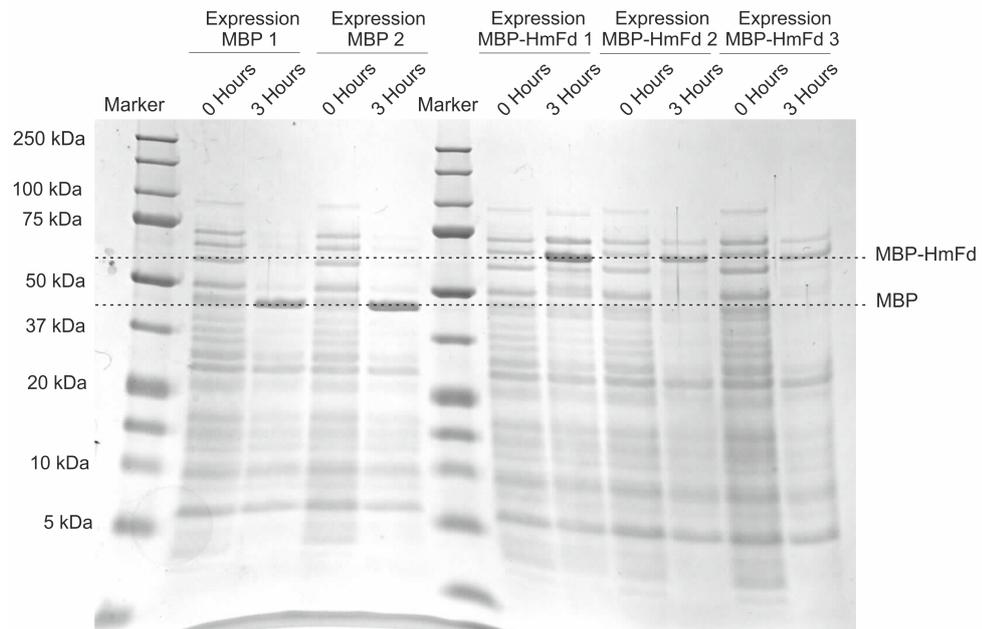
Figure 3.31: Small scale expressions of two samples of MBP and three samples of MBP-*Hm*Fd in JM83 expression cells in 100 mL LB medium, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 58.8 kDa which would correspond to MBP-*Hm*Fd after 3 hours for all three samples and a band at 43.8 kDa both samples corresponding to MBP.
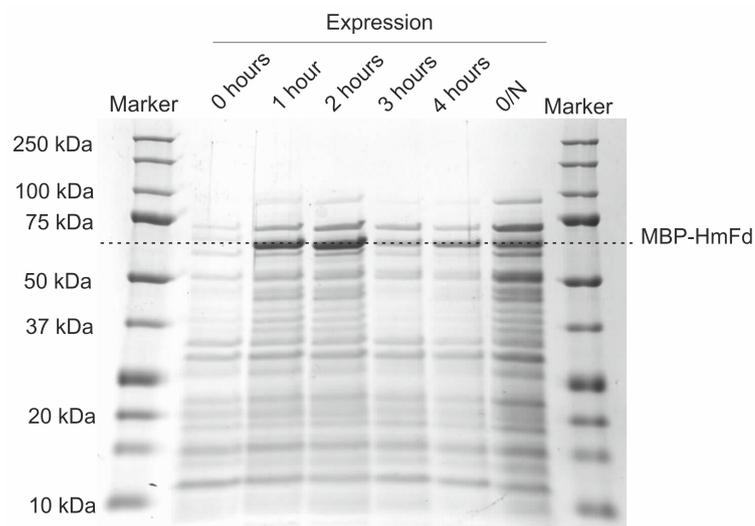
Figure 3.32: Small scale expression of MBP-*Hm*Fd in JM83 expression cells in 100 mL LB medium, monitored at each hour, analysed by SDS-PAGE with Coomassie blue staining. The MBP-*Hm*Fd band (58.8 kDa) is strongest after 2 hours of incubation after induction.
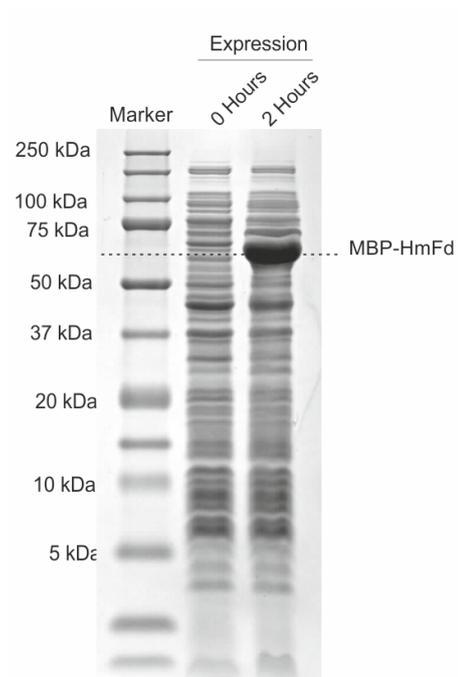
Figure 3.33: Large scale expressions MBP-*Hm*Fd in JM83 expression cells in LB medium, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 58.8 kDa which would correspond to MBP-*Hm*Fd after 2 hours of incubation after induction.
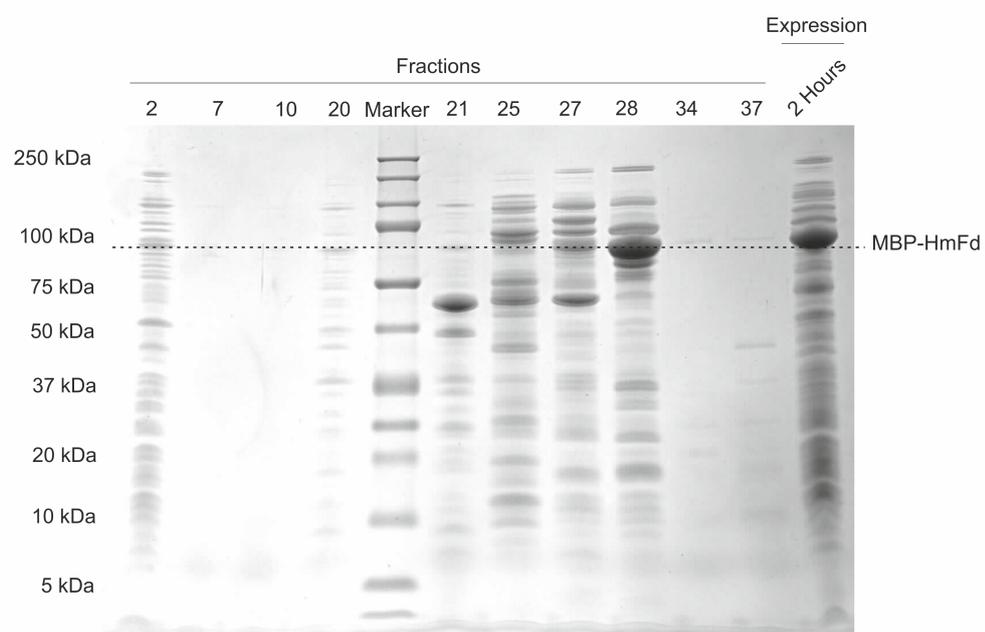
Figure 3.34: Purification of MBP-*Hm*Fd from the lysed cells after expression using anionic exchange chromatography, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 58.8 kDa which would correspond to MBP-*Hm*Fd in fraction 28.

After pooling of the fractions containing MBP-*Hm*Fd, the protein was further purified by anionic exchange chromatography using a Source 15Q column. The elution profile from this column is shown in figure B.2 and SDS-PAGE analysis of each fraction is shown in figure 3.35. The fractions containing MBP-*Hm*Fd identified by SDS-PAGE were pooled, dialysed into H_2O and lyophilised.
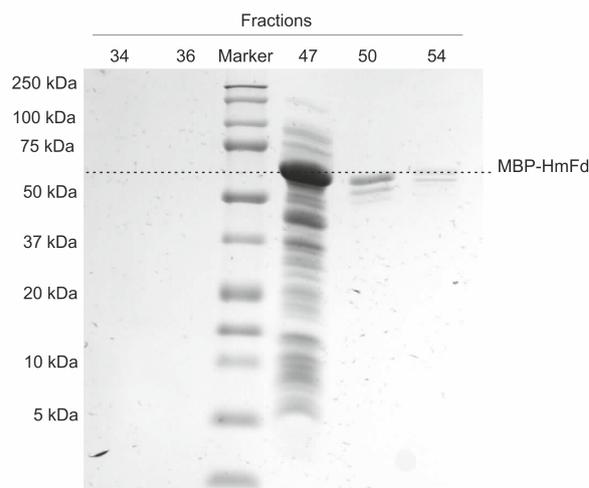
Figure 3.35: Further purification of MBP-*Hm*Fd using anionic exchange chromatography with a SOURCE^{TM} 15Q column, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 58.8 kDa which would correspond to MBP-*Hm*Fd in fraction 47 and smaller amountes in fractions 50, 54.

3.5.3   Cleaving *Hm*Fd from MBP

0.5 mg of MBP-*Hm*Fd was resuspended in buffer containing 50 mM Tris.HCl, 0.5 M NaCl and 2 mM DTT. TEV protease was added to two separate reactions with final molar ratios of MBP-*Hm*Fd to TEV protease of 1:1 and 10:1. Both reactions were incubated at room temperature. Samples of each reaction were taken at 0 hrs, 15 mins, 30 mins, 1 hr and 3 hrs to find the optimal conditions for maximum cleavage with minimal non-specific proteolysis (§2.4.2). The samples were analysed by SDS-PAGE, showing a band decreasing in intensity over time at the expected weight for MBP-*Hm*Fd (58.8 kDa) and a band increasing in intensity over time at the expected weight for MBP (43.8 kDa) in figure 3.36. This implies *Hm*Fd is successfully being cleaved from the the MBP-*Hm*Fd

construct by TEV protease. There does not, however, appear to be a band increasing in intensity at the expected weight for *Hm*Fd (15.0 kDa).
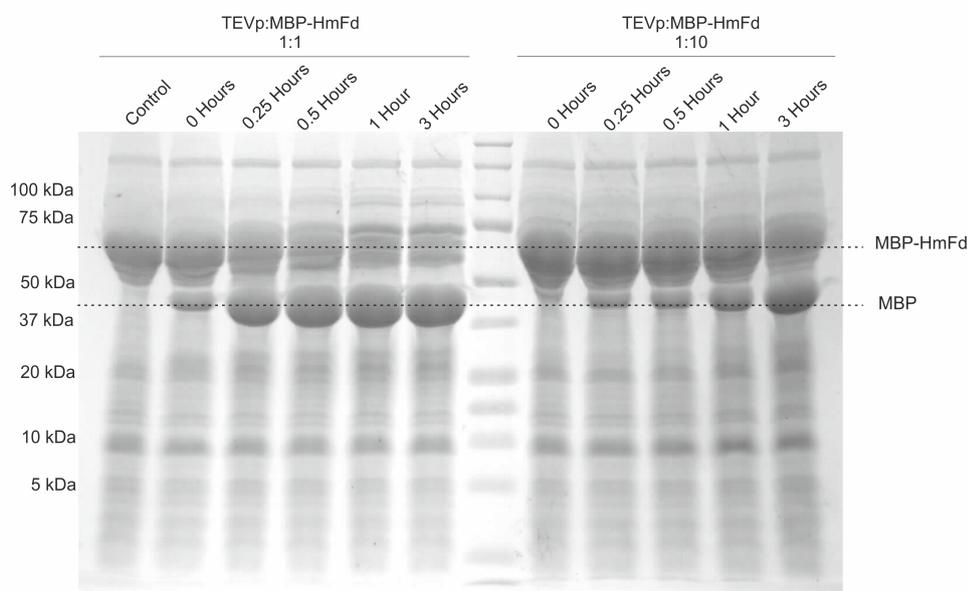
Figure 3.36: Trial cleavage of *Hm*Fd from MBP using TEV protease at two different ratios of protease:protein, analysed by SDS-PAGE with Coomassie blue staining. The gel shows the MBP-*Hm*Fd band at 58.8 kDa reducing in intensity from 0 hours to 3 hours and the MBP band at 43.8 kDa increasing from 0 to 3 hours. There is no band at 15.0 kDa of significant intensity that increased with time, which would correspond to the production of cleaved *Hm*Fd.

To facilitate identification of any cleaved *Hm*Fd product, if present, MBP-*Hm*Fd was further purified by size exclusion chromatography with a Superdex 75 30/300 column prior to TEV cleavage. The SDS-PAGE analysis of the fractions obtained is shown in figure 3.37. Fractions 24-28 were pooled together, dialysed into H_2O and lyophilised.

Another trial digestion was then set up on the purified protein with a final molar ratio of MBP-*Hm*Fd to TEV protease of 3:1. Again, samples were taken at 0 hrs, 15 mins, 30 mins, 1 hr and 3 hrs. The SDS-PAGE analysis, figure 3.38, showed a band decreasing in intensity over time for MBP-*Hm*Fd (58.8 kDa) and a band for MBP (43.8 kDa) but again no band increasing in intensity at the expected weight for *Hm*Fd (15.0 kDa).

A third trial cleavage was then attempted at a 4 x higher concentration of MBP-
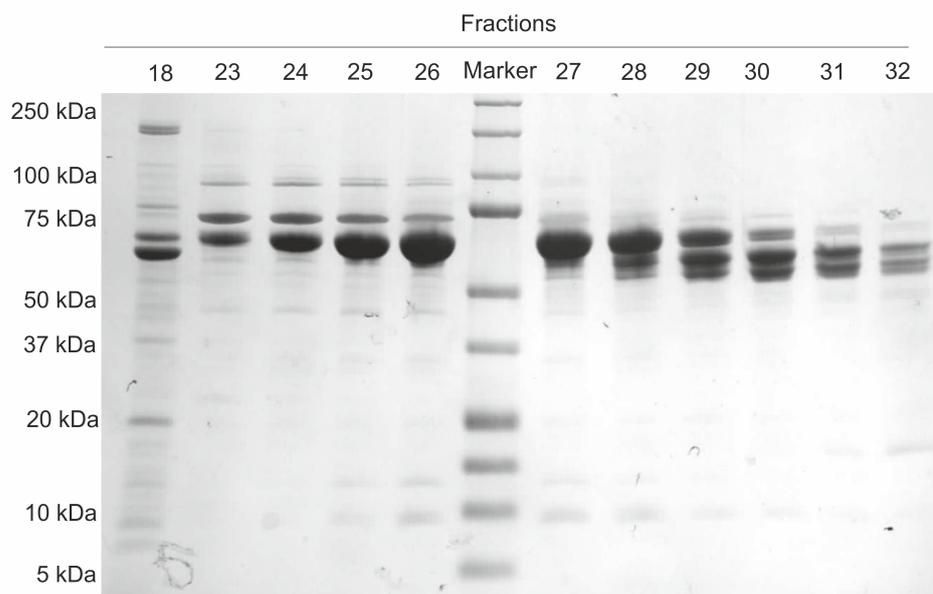
Figure 3.37: Further purification of MBP-*Hm*Fd using size exclusion chromatography, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 58.8 kDa which would correspond to MBP-*Hm*Fd in fractions 24-28, these fractions were pooled together for further use.

*Hm*Fd with the same reaction conditions. A control reaction was also set up which contained only TEVp and no MBP-*Hm*Fd. A band with very low intensity which could correspond to *Hm*Fd can be found in the SDS-PAGE analysis when run with and without DTT in the loading buffer (see figure 3.39 and 3.40) but the intensity is too low to be of use. The analysis was also carried out without DTT to make sure it was not affecting the cleaved *Hm*Fd, since DTT can co-ordinate to metal atoms. The control reaction shows TEVp reducing over time. Precipitated material produced in the reaction was also analysed but found no evidence of cleaved *Hm*Fd (figure 3.41).

A final TEVp cleavage attempt was carried out by performing the cleavage and then purifying the reaction product using a SUP75 30/300 size exclusion column. The SDS-PAGE analysis of the fractions is shown in figure 3.42. No evidence of cleaved HmFd can be seen. Eluted fractions were then analysed using a UV/Vis spectrometer to record full wavelength absorbance scans. *Hm*Fd has a distinctive wavelength absorption at about 420 nm due to the presence of the iron-sulfur cluster[102]. Unfortunately, this distinctive
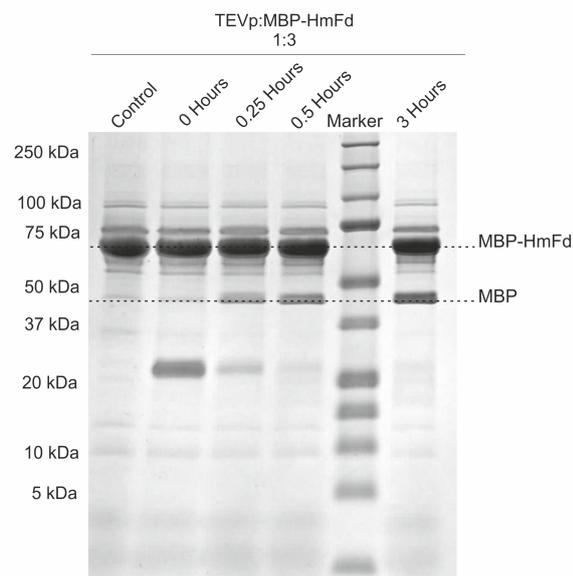
Figure 3.38: Trial cleavage of *Hm*Fd from MBP using TEV protease after futher purification of MBP-*Hm*Fd using size exclusion chromatography, analysed by SDS-PAGE with Coomassie blue staining. The gel shows the MBP-*Hm*Fd band at 58.8 kDa reducing from 0 hours to 3 hours and the MBP band at 43.8 kDa increasing from 0 to 3 hours. It is also possible to see the TEVp band at 27 kDa reducing over time. There is no band with significant intensity at 15.0 kDa increasing in intensity, which would correspond to the production of *Hm*Fd

127
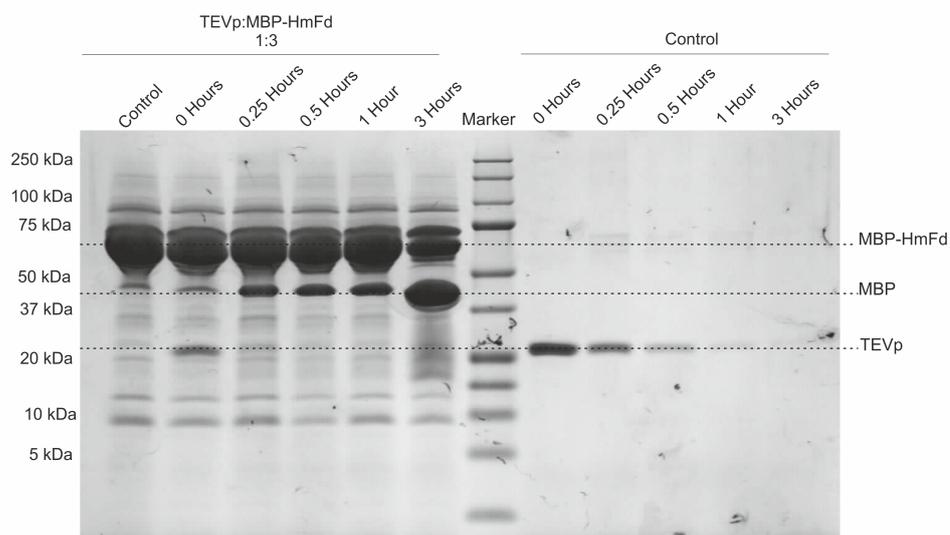
Figure 3.39: Repeat trial cleavage of _Hm_Fd from MBP using TEV protease after purification of MBP-_Hm_Fd using size exclusion chromatography, concentrated using centrifugation and analysed by SDS-PAGE with Coomassie blue staining. Again, The gel shows the MBP-_Hm_Fd band at 58.8 kDa reducing in intensity from 0 hours to 3 hours and the MBP band at 43.8 kDa increasing in intensity from 0 to 3 hours. It is also possible to see the TEVp band at 27 kDa reducing in intensity over time. There is a band with very low intensity which could correspond to _Hm_Fd (15.0 kDa) after cleavage for 3 hours but is too low to be of use.
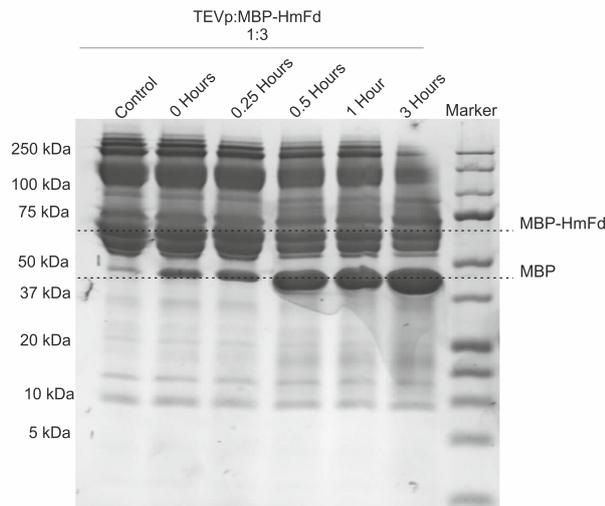
Figure 3.40: Repeat trial cleavage of *Hm*Fd from MBP using TEV protease after purification of MBP-*Hm*Fd using size exclusion chromatography, analysed by SDS-PAGE with Coomassie blue staining without the use of DTT in the loading buffer. The gel shows the MBP-*Hm*Fd band at 58.8 kDa reducing from 0 hours to 3 hours, the MBP band at 43.8 kDa increasing from 0 to 3 hours and the TEVp band at 27 kDa reducing over time. There is a band with very low intensity which could correspond to *Hm*Fd but is too low to be of use.

absorption could not be seen in the analysed fractions (data not shown).

3.5.4 Summary

The results presented show the work carried out in attempt to synthesise a homo–polyprotein construct containing only *Hm*Fd. The results show that the construction of this polyprotein was also unsuccessful using the techniques shown.

Conclusion: Construction of a polyprotein containing *Hm*Fd was unsuccessful using the techniques shown, therefore, an alternative protein must be used.
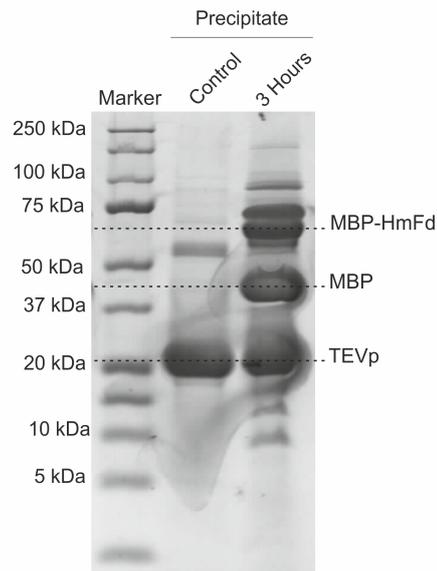
Figure 3.41: Precipitate formed in the repeat trial cleavage of *Hm*Fd from MBP using TEV protease after purification of MBP-*Hm*Fd using size exclusion chromatography, analysed by SDS-PAGE with Coomassie blue staining. The gel shows the MBP-*Hm*Fd band at 58.8 kDa, the MBP band at 43.8 kDa and the TEVp band at 27 kDa but no band with significant intensity at 15.0 kDa, corresponding to the production of *Hm*Fd.
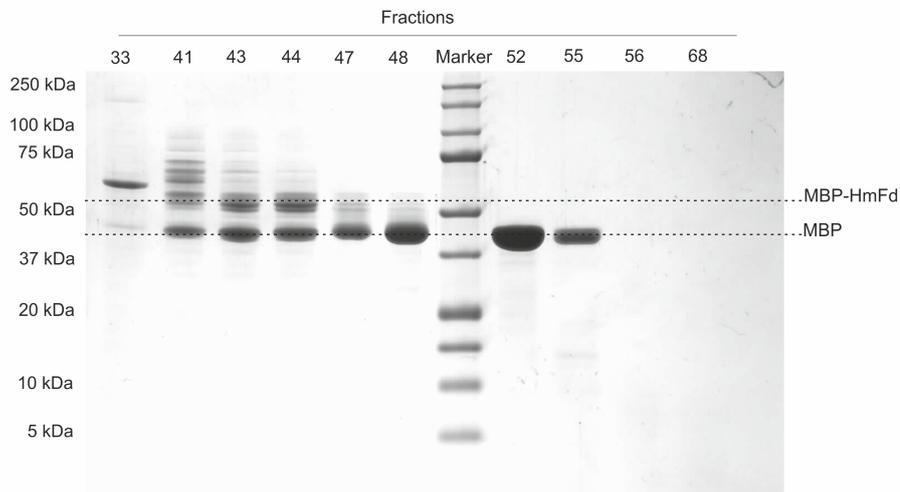
Figure 3.42: Repeat trial cleavage of *Hm*Fd from MBP using TEV protease immediatly purified using size exclusion chromatography, analysed by SDS-PAGE with Coomassie blue staining. The gel shows the MBP-*Hm*Fd band at 58.8 kDa in fractions 41-55 and the MBP band at 43.8 kDa in fractions 33-44 but no evidence of a band at 15.0 kDa, corresponding to the production of *Hm*Fd.

3.6   Chapter summary

This chapter initially presented the work that was carried out on a hetero-polyprotein construct I27$_3$-HmFd-I27 using SMFS. Although force-extension data was successfully collected, the full characterisation of the unfolding force and peak to peak distance was unsuccessful for HmFd. A further discussion of the SMFS analysis can be found in section 6.1. It was concluded that more repeats of HmFd were needed in a polyprotein construct in order generate a larger amount of data to carry out effective mechanical unfolding characterisation.

This chapter also describes two techniques used in attempt to synthesise further polyprotein constructs, a hetero-polyprotein, (I27-HmFd)$_3$-I27, and a homo-polyprotein, poly-HmFd using the fusion protein, Maltose-biding protein (MBP). The production of both of these polyprotein constructs was shown to be unsuccessful using the techniques shown. A discussion of the possible reasons for this is given in section 6.1.

Chapter 4

Results 2: The effect of salt on the mechanical stability of an obligate halophilic and mesophilic hetero-polyproteins

4.1 Introduction

This chapter will introduce and describe the experiments completed on the protein *Peptostreptococcus magnus* Protein L (ProtL). The chapter will begin by introducing the protein, Protein L. Studies carried out on the effects of salt on the protein and mutations to increase the salt stability of the protein will then be introduced. All experimental steps to synthesise two polyproteins will be described, (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27, where ProtL Kx6E is protein L that has had six lysine residues mutated to glutamic acid residues. The experimental data collected, analysed and concluded from SMFS experiments using the polyproteins, (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in two salt concentrations will then be presented. The re-analysis of the SMFS experimental data by separating data sets will then be presented

for both polyproteins. Finally the investigation into the effects of the unfolding history on the SMFS results for the two polyproteins will be presented.

4.1.1 Halophilic protein model: Protein L

The B1 domain of Protein L from *Streptococcus magnus* (ProtL) is a small protein that has a simple topology (figure 4.1). It is one domain of a large multi-domain virulence factor composed of a tandem array of highly homologous domains, and has previously mechanically characterised using AFM by Brockwell et al[145,146].
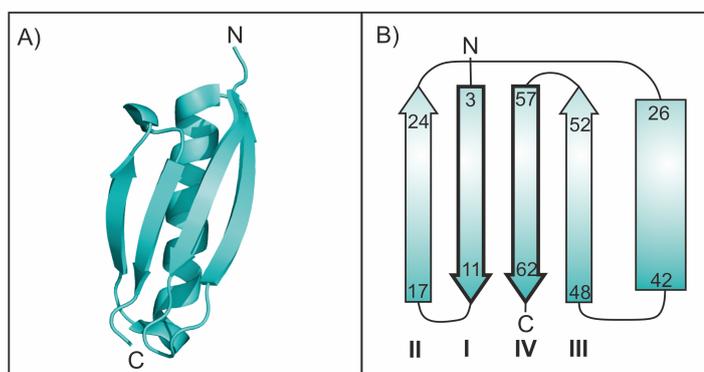
Figure 4.1: The structure of Protein L. A) shows the 3D topology of the protein. This figure was created using PDB file 1HZ6 and Pymol[7,90]. B) shows the secondary structure of the protein where the arrows represent β-strands and the rectangle represents an α-helix. The two bold outlined β–strands show the mechanical clamp of the protein.

The effect of salt concentration on the thermodynamic stability of wild type Protein L has been studied by Tadeo et al. (previously described in section 1.3.3)[111]. It was shown that ProtL WT has fully reversible thermal and chemical unfolding, monitored by circular dichroism and fluorescence spectroscopy. The stability of Prot L, represented by the free energy change of unfolding ($\Delta G^{\text{U-F}}$), was unaffected by the presence of 3.2 M KCl/NaCl. A systematic mutation study was performed, including the generation of charge preservation variants, size preserving variants and variants that changed both size and charge. $\Delta G^{\text{U-F}}_{3.2 \text{ M salt}}$ was determined for each variation of ProtL by equilibrium denaturation experiments.

Previous comparative studies of mesophilic and halophilic proteins from the same family have shown that halophilic proteins tend to have a reduced number of lysine residues[75]. Tadeo reported that changing K for R, Q, S, or E in ProtL leads to salt stabilization for all replacements, independent of their charge, shown in figure 1.22. It was suggested that the low prevalence of lysine residues in a halophilic protein can be explained by a destabilising effect of long lysine side chain at high salt concentration. The NMR structure of ProtL Kx6E was determined (PDB code 2KAC). ProtL Kx6E was found to be an obligate halophile (i.e. unfolded at low salt concentrations).

4.1.1.1   The chosen halophilic protein model: Protein L Kx6E

In order to test the mechanical stability of a salt stabilised protein compared to that of the mesophilic homologue, a wild type ProtL (ProtL WT) and a Protein L with six lysine residues in the ProtL sequence mutated to glutamic acids (ProtL Kx6E)(figure 4.2) was chosen to be expressed and purified. Poly-proteins that include these proteins were then characterised using force extension experiments. One of the glutamic acid residues in Protein L is position 61, which is located in β-strand IV, the mechanical clamp region of Prot L, shown in figure 4.2. Two other glutamic acids reside in the secondary structure, positions 23 and 28, of the protein, while the remaining three are within the loop regions, 41, 42, 54.

4.1.2   Previous SMFS studies carried out on *Peptostreptococcus magnus* Protein L (ProtL)

Protein L is a small protein that found in the bacterial cell wall of *Peptostreptococcus magnus*. It binds specifically to immunoglobulins (Ig) of a number of animal species[198]. Protein L has been first studied using SMFS in 1995 by Brockwell at al[145,146,199]. It is a small protein with a simple topology that was predicted to have mechanical stability due to its hydrogen-bonded parallel terminal β-strands (I and IV), see figure 4.1. The mechanical unfolding of the protein was shown to be highly reproducible in a two state transition with a well defined unfolding pathway. Molecular dynamics simulations of
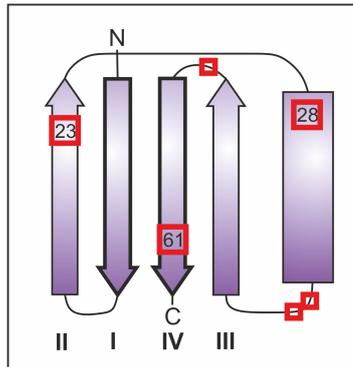
Figure 4.2: The positions of the ProtL Kx6E mutations. The figure shows the secondary structure of the protein, where the arrows represent β–strands and the rectangle represents an α–helix. The two bold outlined β–strands show the mechanical clamp of the protein. The numbers given are the positions of mutations within the secondary structure. All positions of mutations are outlined in red, the remaining mutations (41, 42, 54) are all positioned in the loop regions of the protein.

Protein L mechanical unfolding were used to investigate the process on an atomistic scale[146]. Firstly, the β-stand at the N-terminus (I) reorients to the direction of the applied force, which disrupts the bonds between β-stand I and the hydrophobic core (see figure 4.3 A)). Since this only causes a very small extension, the force is quickly applied to the rest of the protein. The full protein then unfolds after the bonds (hydrogen bonds and hydrophobic contacts) are sheared between the N- and C-terminus β-stands (I and IV) as a single step.

To gain greater insight into the role of side chains on the mechanical properties of protein L, a contact maps was constructed (figure 4.3 B)). The amino acid residues that form β–strands I and IV of Protein L have a total of 65 contacts. The two β–strands only form 22 long-range contacts. Ubiquitin is a protein that is structurally very similar to protein L but has a significantly higher mechanical stability (∼70 pN higher) at the same pulling velocities. It was calculated to have a similar total number of contacts but significantly more long-range contacts, 38. It was suggested that a protein with a larger number of long-range contacts that entangle each unit may need a higher force to unfold the protein[146].

The effect of side-chain packing on the mechanical strength of protein L was also
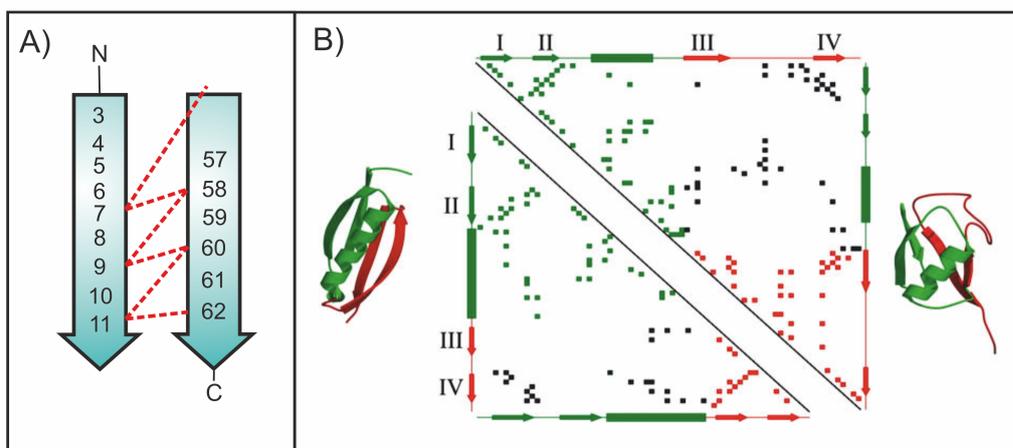
Figure 4.3: The mechanical clamp region of Protein L. Figure A) shows the two β-stands (I and IV) that form the mechanical clamp for Protein L. The numbers of the corresponding residues have been included and interstrand hydrogen bonds, calculated using DSSP, are shown as red dashed lines[146,200]. Figure B) shows a contact map of protein L (bottom left) and ubiquitin (top right). Side-chain contacts (nearest distance between atoms of two residues <5 Å, calculated by CSU software) made by pairs of amino acids within structural unit 1 (β-hairpin I and the helix - green) or within structural unit 2 (β-hairpin 2 - red)[201]. Contacts made between these structural units are shown in black. β-strands are shown as arrows and α-helices are shown as rectangles alongside each contact map. The two structural units are colored green (unit 1) and red (unit 2) in each protein and are also shown superimposed onto the three-dimensional structure of protein L(left) and ubiquitin (right). Figure adapted from [146].

examined using SMFS experiments[145]. The mechanical stability of 5 different Protein L mutants, designed to change the number of interactions of β-stands (I and IV) with the hydrophobic region, was measured. One variant, I60V, showed a significant decrease in the mechanical stability ($\sim$36 pN) compared to the wild type. Two other variants also showed significant results; L10F, a variant that increases core packing, increased the mechanical strength of Protein L by $\sim$13 pN compared to ProtL WE. I60F, which increases both core and long-range contacts, increased the mechanical strength $\sim$72 pN. This study highlighted the importance of hydrophobic contacts on the mechanical properties of Protein L.

4.2 Objectives

The objective of this chapter was, firstly, to produce two polyprotein constructs, (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27. Secondly, to complete a series of SMFS experiments on a polyprotein construct containing ProtL WT and the halophilic model protein, ProtL Kx6E to obtain experimental data to:

i) allow identification of the mechanical fingerprint of ProtL WT and ProtL Kx6E

ii) characterise the speed dependence of the unfolding force of both ProtL WT and ProtL Kx6E

iii) investigate the effect of salt on the mechanical stability of both ProtL WT and ProtL Kx6E.

4.3 Production of the hetero-polyprotein, (I27-ProtL WT)$_3$-I27

This section will describe the experiments completed to synthesise a polyprotein chimera containing I27 and ProtL WT. The design of the polyprotein chimera chosen is shown in figure 4.4 and includes 4 repeats of I27 and 3 repeats of ProtL WT. Further explanation of the techniques used can be found in chapter 2, including a basic description of each

step and the protocols used.

Figure 4.4: Schematic of the (I27-ProtL WT)$_3$-I27 construct. The construct has a His-tag, shown as a red triangle, at the N-terminal end and 2 cysteine residues, shown as green circles, at the C-terminal end.

4.3.1 Expression of (I27-ProtL WT)$_3$-I27

The pET3d-(I27-ProtL WT)$_3$-I27 was provided by Dr David Brockwell. pET3d-(I27-ProtL WT)$_3$-I27 plasmid was transformed into competent *E. coli* strain BL21 (DE3) PLysS (§2.3.1). 10 mL of an overnight culture of the cells was used to inoculate 10 x 1 L LB medium in 2.5 L conical flasks. The cultures were incubated at 37°C, 200 rpm. Expression of the construct was induced by IPTG at an OD$_{600}$ of 0.7. Cultures were allowed to grow for a further 3 hours before harvesting by centrifugation (§2.3.4). The cell pellet was resuspended in an appropriate buffer and cell lysis was performed using a cell disrupter system. (I27-ProtL WT)$_3$-I27 was purified from the soluble fraction by Ni affinity chromatography, using a Ni Sepharose column and the His-tagged protein was eluted by competition with imidazole. Fractions containing (I27-ProtL WT)$_3$-I27 were pooled, dialysed into H$_2$O and lyophilised. SDS-PAGE analysis of the fractions is shown in figure 4.5.

The polyprotein was further purified by size exclusion chromatography using a Superdex 75 10/300 GL column. SDS-PAGE analysis (figure 4.6) of the resulting fractions show a band with significant intensity at 65.1 kDa corresponding to (I27-ProtL WT)$_3$-I27 in fractions 36-42. The lanes also contain other bands, showing further purification of the polyprotein is needed. The fractions containing (I27-ProtL WT)$_3$-I27 were again pooled, dialysed into H$_2$O and lyophilised.

To attempt to remove the extra unwanted proteins in the sample, purification by Ni affinity chromatography was repeated using a 1 mL His-Trap column to reduce binding
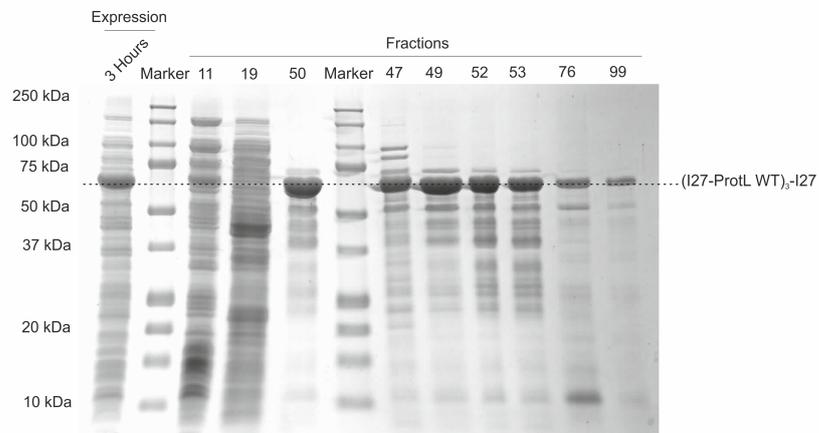
Figure 4.5: Large scale purification of (I27-ProtL WT)$_3$-I27 using Ni-NTA chromatography, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 65.1 kDa which would correspond to (I27-ProtL WT)$_3$-I27 in fractions 49-53.

Figure 4.6: Purification of (I27-ProtL WT)$_3$-I27 using size exclusion chromatography, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 65.1 kDa which would correspond to (I27-ProtL WT)$_3$-I27 in fractions 36-42.

of unwanted proteins. The polyprotein was eluted by increasing the concentration of imidazole in a 0-40% ratio over 5 column volumes. The SDS-PAGE of the fractions shown in figure 4.7. The number of protein bands in each lane is significantly reduced, showing the purified polyprotein, although a few unwanted bands can still be seen. The fractions containing (I27-ProtL WT)$_3$-I27 were again pooled, dialysed into H$_2$O and lyophilised.
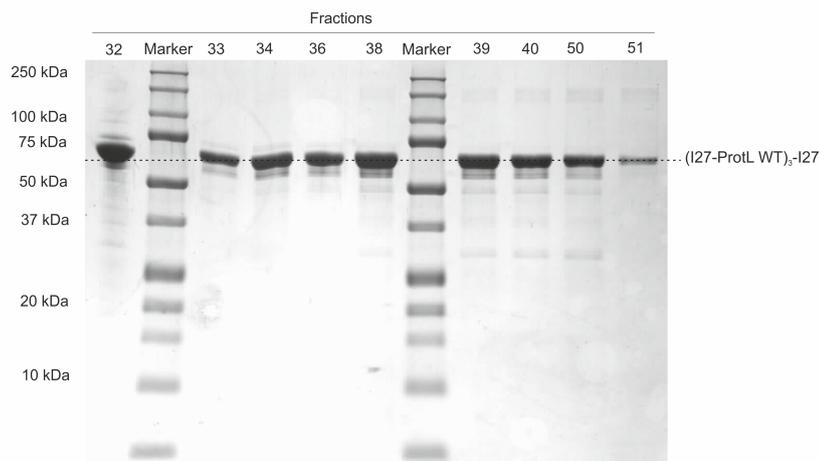
Figure 4.7: Further purification of (I27-ProtL WT)$_3$-I27 using Ni affinity chromatography, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band of significant intensity at 65.1 kDa which would correspond to (I27-ProtL WT)$_3$-I27 in all fractions tested.

To try and remove the final contaminants, the polyprotein was purified by anionic exchange chromatography using a 5 mL Hi-Trap Q column (§2.3.8). The SDS-PAGE analysis of the fractions is shown in figure 4.8. A number of bands that do not correspond to the polyprotein can still be seen. Since the polyproteim could not be fully purified after a number of purification steps, the process of purification needed to be optimised.

In order to optimise the first step of purification of the polyprotein, Ni affinity chromatography, three components were varied:

i) The concentration of imidazole in the wash buffer. An increased imidazole concentration is desirable to stop unwanted proteins from binding to the column. If the concentration of imidazole is too high, the amount of his-tagged polyprotein bound to the column will be reduced.
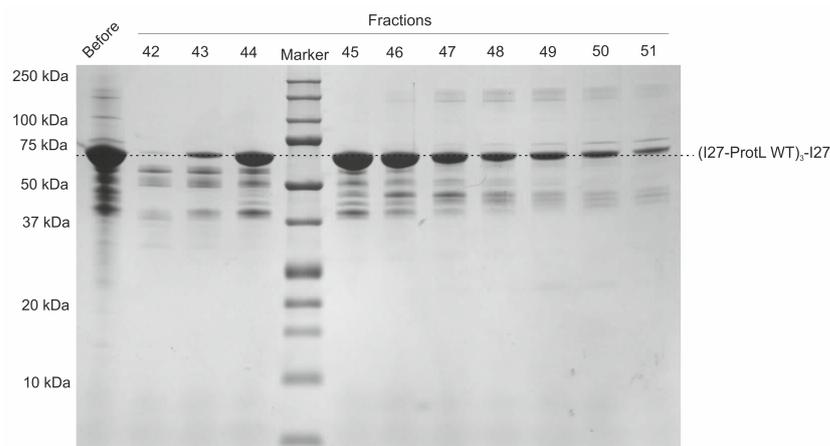
Figure 4.8: Purification of (I27-ProtL WT)$_3$-I27 using ionic exchange chromatography, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band of significant intensity at 65.1 kDa which would correspond to (I27-ProtL WT)$_3$-I27 in fractions 44-51.

ii) The salt concentration in all buffers. The addition of salt in a buffer helps screen ionic interactions and helps prevent non-specific binding of unwanted proteins to the column and also prevent aggregation of proteins.

iii) Inclusion of glycerol in the buffers can prevent the aggregation of proteins.

Firstly, 4 x 1 mg samples of polyprotein were resuspended in buffers containing 10 mM, 20 mM, 30 mM and 40 mM imidazole. (I27-ProtL WT)$_3$-I27 was purified using Ni-NTA affinity resin (§2.3.3). The resin was also washed with buffers containing the appropriate concentration of imidazole for each purification. SDS-PAGE analysis of one fraction for each sample after elution with 250 mM Imidazole is shown in figure 4.9. 20 mM imidazole was chosen as the optimal concentration in the resuspending buffer because the fractions corresponding to 30 and 40 mM imidazole show a reduced intensity in the band corresponding to (I27-ProtL WT)$_3$-I27.

Next, 3 x 1 mg samples of polyprotein were resuspended in wash buffers containing 0.5 M, 1 M and 2 M NaCl all with no glycerol. 3 x 1 mg samples of polyprotein were also resuspended in wash buffers containing 0.5 M, 1 M and 2 M NaCl all with 25% glycerol. (I27-ProtL WT)$_3$-I27 was then purified by Ni affinity chromatography for each sample, using Ni-NTA affinity resin and the His-tagged protein was eluted by competition with
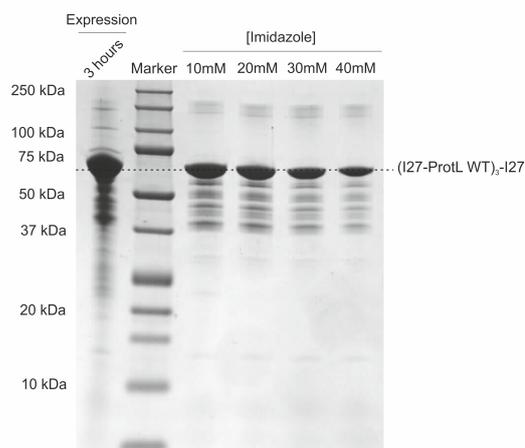
Figure 4.9: The 1st optimisation of the purification of (I27-ProtL WT)$_3$-I27 using Ni-NTA chromatography, analysed by SDS-PAGE with Coomassie blue staining. The gel shows one lane for each concentration of imidazole. The intensity of each band is reduced with the increasing concentration of imidazole.

imidazole. SDS-PAGE analysis of one fraction for each sample is shown in figure 4.10. The gel shows that neither the salt concentration or the glycerol concentration affect the purification of the polyprotein.

4.3.2 Repeat expression of (I27-ProtL WT)$_3$-I27 after optimisation of the purification technique

Expression of the polyprotein was then repeated using 10 x 1 L LB medium in 2.5 L conical flasks. The cultures were incubated at 37°C, 200 rpm. Expression of the construct was induced by IPTG at an OD$_{600}$ of 0.7. Cultures were allowed to grow for a further 3 hours before harvesting by centrifugation.

The cell pellet was resuspended in an appropriate buffer containing 20 mM imidazole. Protease inhibitor cocktail tablets were added to remove any proteases that might be present. Cell lysis was performed using a cell disrupter system. (I27-ProtL WT)$_3$-I27 was purified from the soluble fraction using Ni-NTA affinity resin and the His-tagged protein was eluted by competition with imidazole. The amount of resin used to purify the polyprotein was optimized for its binding capacity. Fractions containing (I27-ProtL
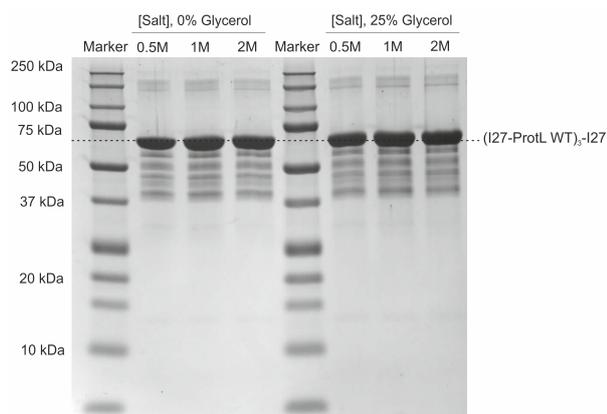
Figure 4.10: The 2nd optimisation of the purification of (I27-ProtL WT)$_3$-I27 using Ni-NTA chromatography, analysed by SDS-PAGE with Coomassie blue staining. The gel shows one lane for each sample. The gel shows that neither the salt concentration or the glycerol concentration affect the purification of the polyprotein.

WT)$_3$-I27 were pooled, dialysed into H$_2$O and lyophilised.

Polyprotein was further purified by size exclusion chromatography using a Superdex 75 column 10/300 GL. SDS-PAGE analysis of the resulting fractions, shown in figure 4.11, shows a band with with significant intensity at 65.1 kDa corresponding to (I27-ProtL WT)$_3$-I27 in fractions 35-37 with no other unwanted protein bands. The elution profile is shown in figure C.1. These fractions were pooled, dialysed into H$_2$O and lyophilised. 5 mg of (I27-ProtL WT)$_3$-I27 was obtained, ready to be used for SMFS analysis.

4.3.3 Summary

The results presented show the work carried out to optimise the purification of a hetero-polyprotein construct (I27-ProtL WT)$_3$-I27. Production of the hetero construct was successful using this technique and produced enough sample for SMFS experiments.

Conclusion: This method of production of the hetero-polyprotein construct (I27-ProtL WT)$_3$-I27 was successful.
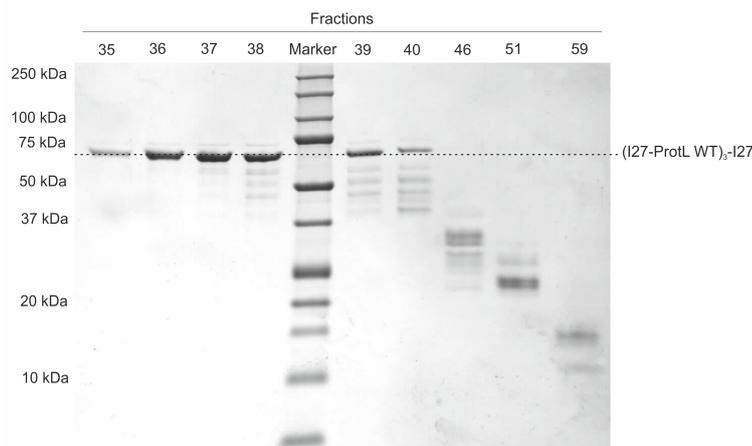
Figure 4.11: Purification of re-expressed (I27-ProtL WT)$_3$-I27 using size exclusion chromatography, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 65.1 kDa which would correspond to (I27-ProtL WT)$_3$-I27 in fractions fractions 35-37 with no other unwanted protein bands.

4.4 Production of a the hetero-polyprotein, (I27-ProtL Kx6E)$_3$-I27

This section will describe the experiments completed to synthesise a polyprotein chimera containing I27 and ProtL Kx6E. The design of the polyprotein chimera chosen is shown in figure 4.12 and includes 4 repeats of I27 and 3 repeats of ProtL Kx6E.

Figure 4.12: Structure of the (I27-ProtL Kx6E)$_3$-I27 construct. The construct has a His-tag, shown as a red triangle, at the N-terminal end and 2 cysteine residues, shown as green circles, at the C-terminal end.

4.4.1 Expression of (I27-ProtL Kx6E)$_3$-I27

The pET3d-(I27-ProtL Kx6E)$_3$-I27 was provided by Dr David Brockwell. pET3d-(I27-ProtL Kx6E)$_3$-I27 plasmid was transformed into competent *E. coli* strain BL21 (DE3) PLysS. 10 mL of an overnight culture of the cells was used to inoculate 10 x 1 L LB

medium in 2.5 L conical flasks. The cultures were incubated at 37°C, 200 rpm. Expression of the construct was induced by IPTG at an OD_{600} of 0.7. Cultures were allowed to grow for a further 3 hours before harvesting by centrifugation. The cell pellet was resuspended in an appropriate buffer and the cells were lysed using a cell disrupter system. (I27-ProtL Kx6E)$_3$-I27 was purified from the soluble fraction by Ni affinity chromatography using a Ni Sepharose column. Fractions containing (I27-ProtL Kx6E)$_3$-I27 were pooled, dialysed into H_2O and lyophilised, SDS-PAGE analysis of the fractions is shown in figure 4.13. The gel shows a band at 65.1 kDa which would correspond to (I27-ProtL Kx6E)$_3$-I27 in fractions 39 and 40. The band does not have a large intensity, indicating only a small amount of (I27-ProtL Kx6E)$_3$-I27 was expressed.
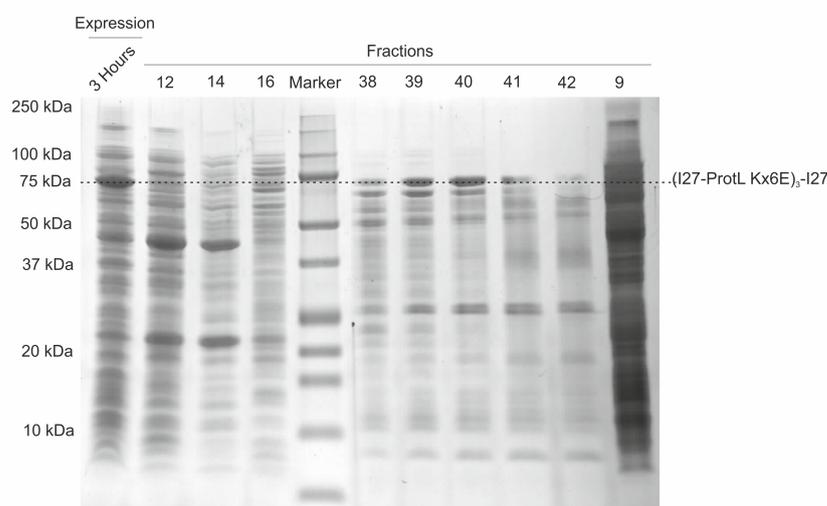
Figure 4.13: Large scale purification of (I27-ProtL Kx6E)$_3$-I27 using Ni-NTA chromatography, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 65.1 kDa which would correspond to (I27-ProtL Kx6E)$_3$-I27 in fractions 39 and 40 and in the expression band.

(I27-ProtL Kx6E)$_3$-I27 was then purified by size exclusion chromatography using a Superdex 75 column. SDS-PAGE analysis of the resulting fractions, figure 4.14, shows a band with significant intensity at 65.1 kDa corresponding to (I27-ProtL Kx6E)$_3$-I27 in fractions 33-41. The lanes also contain other bands, showing further purification of the polyprotein is needed. Size exclusion chromatography separates proteins by size, larger proteins should be eluted in the later fractions compared to smaller fractions. Figure 4.14 shows proteins with a variety of sizes in each lane, suggesting the protein was degrading

after purification.
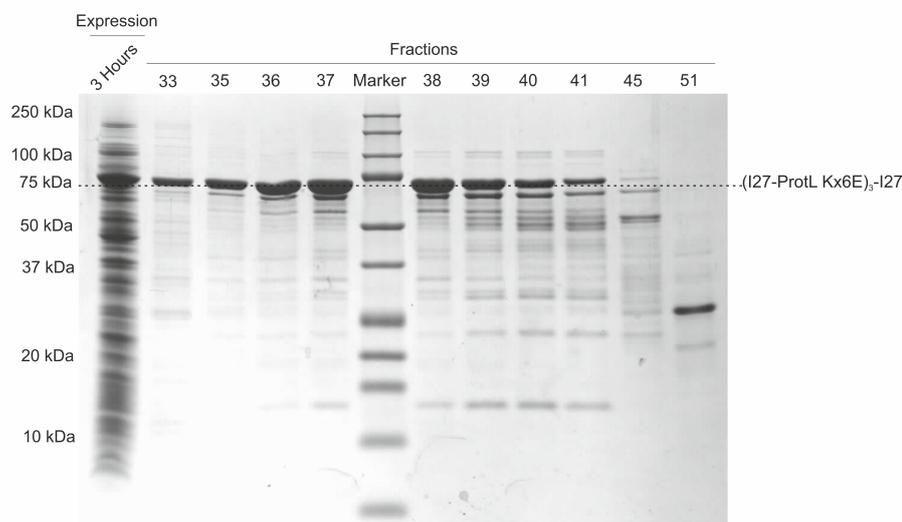
Figure 4.14: Purification of (I27-ProtL Kx6E)$_3$-I27 using size exclusion chromatography, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 65.1 kDa which would correspond to (I27-ProtL Kx6E)$_3$-I27 in fractions 33-41.

4.4.2 Repeat expression of (I27-ProtL Kx6E)$_3$-I27 to reduce degradation of the polyprotein

Expression of the polyprotein was then repeated using 10 x 1 L LB medium in 2.5 L conical flasks, using the same process stated above. The cell pellet was resuspended in an appropriate buffer, containing 20 mM imidazole. Protease inhibitor cocktail tablets were added to inactivate any proteases that might be present. Cell lysis was performed using a cell disrupter system. (I27-ProtL Kx6E)$_3$-I27 was purified from the soluble fraction using Ni-NTA affinity resin and the His-tagged protein was eluted by competition with imidazole. The amount of resin used the purify the polyprotein was optimized for its binding capacity. Fractions containing (I27-ProtL Kx6E)$_3$-I27 were pooled, dialysed into H$_2$O and lyophilised.

The re-expressed polyprotein was further purified by size exclusion chromatography using a Superdex 75 column. The elution profile is shown in figure C.2. SDS-PAGE

analysis of the resulting fractions, shown in figure 4.15, shows a band with with significant intensity at 65.1 kDa corresponding to (I27-ProtL Kx6E)$_3$-I27 in fractions 35-37. A single contaminant is present at $\sim$60 kDa. It was deduced that, given the relative intensity of this, the fractions were sufficiently pure for SMFS analysis. These fractions were pooled, dialysed into H$_2$O and lyophilised. Using this technique, only 1 mg (I27-ProtL Kx6E)$_3$-I27 was produced, which was not enough for analysis by SMFS.
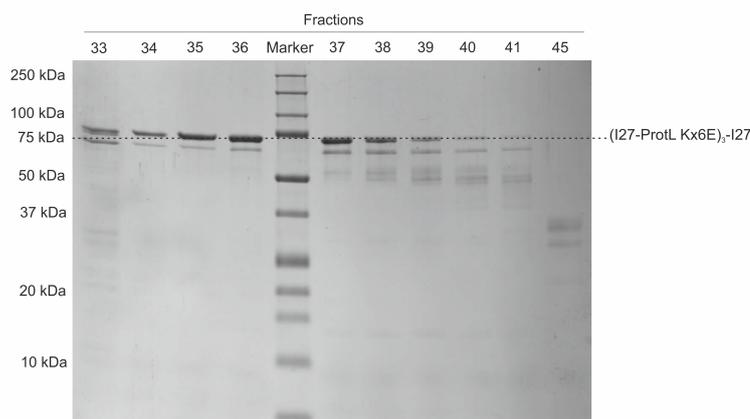
Figure 4.15: Purification of re-expressed (I27-ProtL Kx6E)$_3$-I27 using size exclusion chromatography, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 65.1 kDa which would correspond to (I27-ProtL Kx6E)$_3$-I27 in fractions fractions 35-37.

4.4.3 Repeat expression of (I27-ProtL Kx6E)$_3$-I27 to improve the yield of the polyprotein

To try and produce a larger amount of protein, expression trials were set up using LB, TB and auto-induction media. Expression of the polyprotein was carried out using 1 L of each medium in 3 x 2.5 L conical flasks. The cultures were incubated at 37°C, 200 rpm. Expression of the construct was induced by IPTG at an OD$_{600}$ of 0.7. Samples were taken at each hour after induction (up to 6 hours) with a final sample being taken after incubation overnight. The SDS-PAGE analysis for each sample is shown in figures 4.16, 4.17 and 4.18. The largest expression of polyprotein was found to be in auto-induction medium with overnight incubation. The amount of cells at the optimal protein expression time was also largest in auto-induction medium (OD$_{600}$ = 5.51).
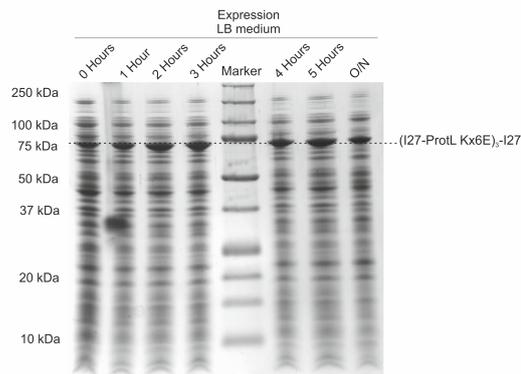
Figure 4.16: Test expression of (I27-ProtL Kx6E)$_3$-I27 in 1 L LB medium, monitored at each hour, analysed by SDS-PAGE with Coomassie blue staining. The (I27-ProtL Kx6E)$_3$-I27 (65.1 kDa) has the highest intensity 5 hours after induction.
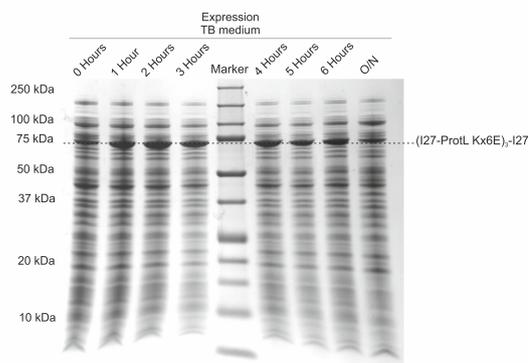
Figure 4.17: Test expression of (I27-ProtL Kx6E)$_3$-I27 in 1 L TB medium, monitored at each hour, analysed by SDS-PAGE with Coomassie blue staining. The (I27-ProtL Kx6E)$_3$-I27 (65.1 kDa) has the highest intensity 2 hours after induction.
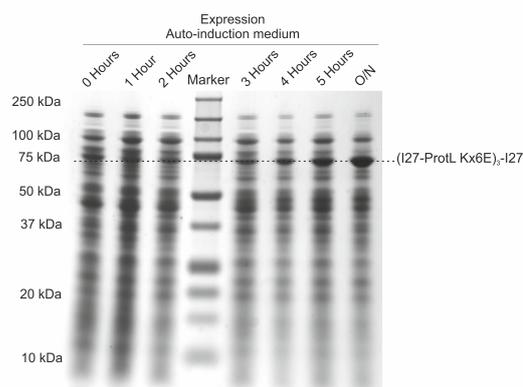
Figure 4.18: Test expression of (I27-ProtL Kx6E)$_3$-I27 in 1 L auto-induction medium, monitored at each hour, analysed by SDS-PAGE with Coomassie blue staining. The (I27-ProtL Kx6E)$_3$-I27 (65.1 kDa) has the highest intensity after incubation overnight.

A large scale expression of (I27-ProtL Kx6E)$_3$-I27 was then set up in 10 x 1 L Auto-induction medium in 2.5 L conical flasks. The cultures were incubated at 37°C, 200 rpm and were allowed to grow overnight before harvesting by centrifugation. The cell pellet was resuspended in an appropriate buffer and the cells were lysed using a cell disrupter system. (I27-ProtL Kx6E)$_3$-I27 was purified from the soluble fraction by Ni affinity chromatography using a Ni Sepharose column. Fractions containing (I27-ProtL Kx6E)$_3$-I27 were pooled, dialysed into H$_2$O and lyophilised, SDS-PAGE analysis of the fractions is shown in figure 4.19. The insoluble fraction was also analysed using SDS-PAGE since the yield of polyprotein in the soluble fraction was small. Evidence of (I27-ProtL Kx6E)$_3$-I27 was found in the insoluble fraction, shown by a band at 65.1 kDa in figure 4.19.
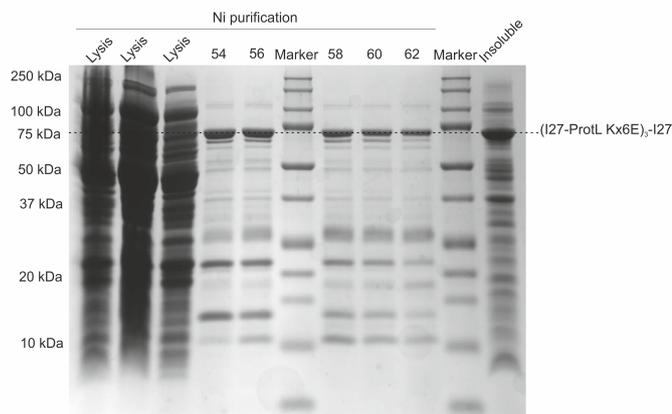
Figure 4.19: Purification of (I27-ProtL Kx6E)$_3$-I27 using Ni-NTA chromatography, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 65.1 kDa which would correspond to (I27-ProtL Kx6E)$_3$-I27 in fractions 54-62. The insoluble sample also shows a band with significant intensity at 65.1 kDa.

Half of the insoluble fraction was then resuspended in 8 M urea buffer. The other half was resuspended in a 6 M guanidinium chloride buffer. Both samples were purified in a denatured form using by Ni affinity chromatography using a 1 mL His-trap Ni Sepharose column The elution profile for the sample in 6 M guanidinium chloride is shown in figure C.3. Fractions containing (I27-ProtL Kx6E)$_3$-I27 were pooled. SDS-PAGE analysis of the fractions is shown in figures 4.20 and 4.21. The resulting polyprotein from this purification was unfolded.
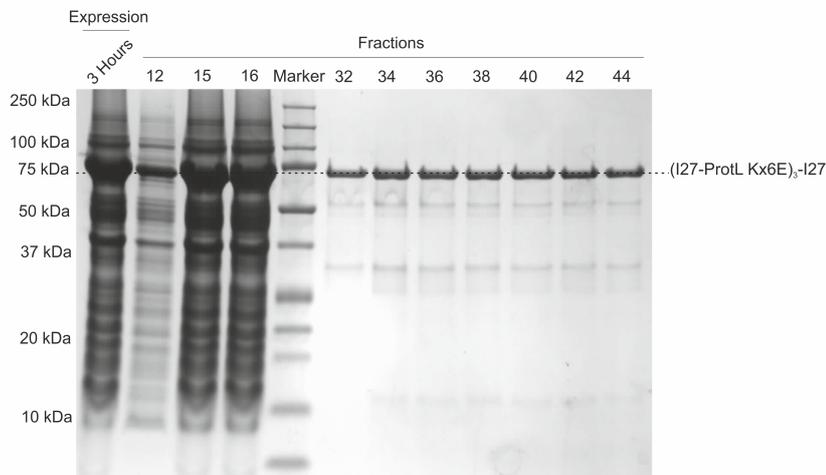
Figure 4.20: Large scale purification of (I27-ProtL Kx6E)$_3$-I27 from the insoluble fraction using Ni-NTA chromatography in 8 M urea, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 65.1 kDa which would correspond to (I27-ProtL Kx6E)$_3$-I27 in all fractions, including initial flow through fractions (12-16), which were re-purified on the column afterwards.
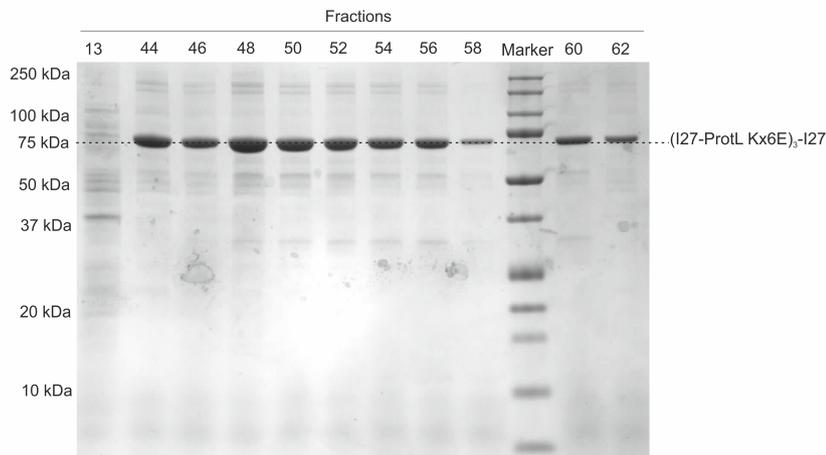
Figure 4.21: Large scale purification of (I27-ProtL Kx6E)$_3$-I27 from the insoluble fraction using Ni-NTA chromatography in 6 M GdmCl, analysed by SDS-PAGE with Coomassie blue staining. The gel shows a band with significant intensity at 65.1 kDa which would correspond to (I27-ProtL Kx6E)$_3$-I27 in fractions 44-56.

Further purification of the polyprotein was carried out with the sample unfolded in urea since the SDS-PAGE analysis showed fewer unwanted protein bands. To refold the polyprotein, (I27-ProtL Kx6E)$_3$-I27, purified using 8 M urea, was bound to a Ni Sepharose column and then refolded washing the column with a 0.9 M arginine buffer. The His-tagged polyprotein was eluted by competition with imidazole.

Given the success of this method, a large scale expression of (I27-ProtL Kx6E)$_3$-I27 was repeated using the same procedure. The relative amounts of soluble/insoluble protein were found to be highly variable. Given the relative amount of soluble protein purified, it was decided that this method should be optimised.

4.4.4 Repeat expression of (I27-ProtL Kx6E)$_3$-I27 to reduce degradation of the polyprotein at an increased yield

In order to reduce the amount of degradation of the polyprotein, the polyprotein was re-expressed in cells cultivated in 20 L x auto-induction medium. The length of time of the purification process was reduced by carrying out using Ni-NTA affinity resin. The amount of resin used to purify the polyprotein was optimized for its binding capacity after elution with imidazole. The solution was concentrated and injected onto a Superdex 75 10/300 GL column, to remove the dialysis and lyophilisation steps, and purified by size exclusion chromatography. The elution profile is shown in figure C.4. Each step of the purification process was carried out at 4 °C. SDS-PAGE analysis of the full procedure is shown in figure 4.22. A band with significant intensity at 65.1 kDa corresponding to (I27-ProtL WT)$_3$-I27 is observed in fractions 36-38 with minimal contamination of unwanted protein bands. These fractions were pooled, dialysed into H$_2$O and lyophilised. 4 mg of (I27-ProtL Kx6E)$_3$-I27 was obtained, ready to be used for SMFS analysis.

4.4.5 Summary

The results presented show the work carried out to synthesise a hetero-polyprotein construct (I27-ProtL Kx6E)$_3$-I27. Production of the hetero construct was successful using

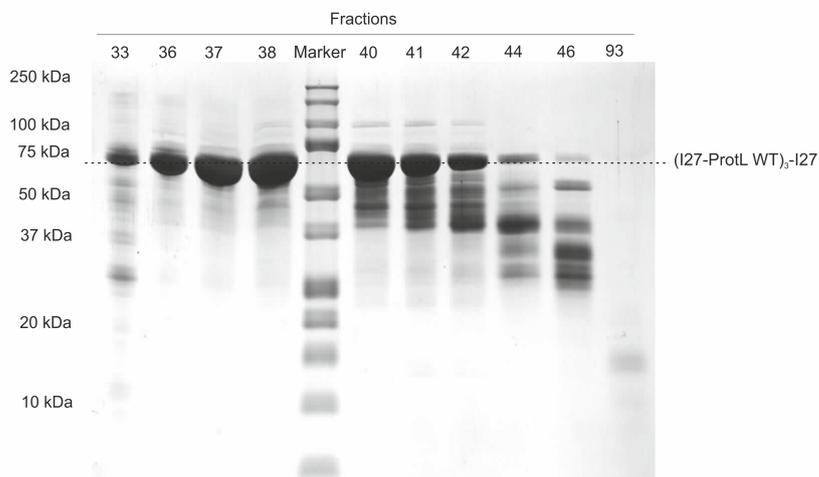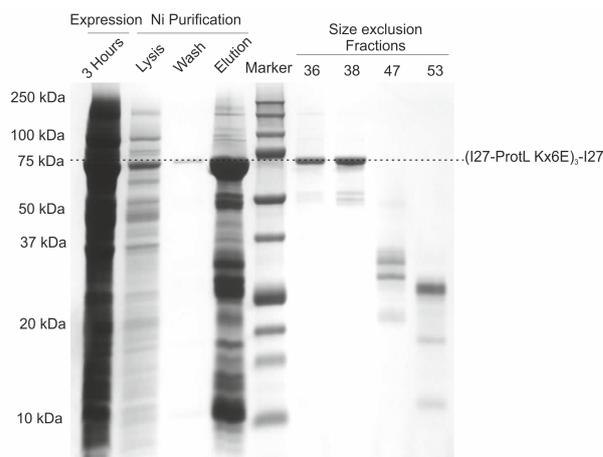**Figure 4.22: Purification of (I27-ProtL WT)$_3$-I27 in one day at 4 °C, analysed by SDS-PAGE with Coomassie blue staining.** The gel shows a band with significant intensity at 65.1 kDa which would correspond to (I27-ProtL Kx6E)$_3$-I27 in fractions 36-38.

this technique and produced enough for SMFS analysis after optimisation of the method used.

Conclusion: This method of production of the hetero-polyprotein construct (I27-ProtL Kx6E)$_3$-I27 was successful.

4.5 Single molecule force spectroscopy on (I27-ProtL WT)$_3$-I27 in 0.5 M KCl

This section will show data collected from SMFS experiments using (I27-ProtL WT)$_3$-I27 in 0.5 M KCl, 5 mM MOPS, pH 7.4. The analysis carried out on this data will also be presented.

4.5.1 Procedure for SMFS experiments using (I27-ProtL WT)$_3$-I27 in 0.5 M KCl

Proteins were resuspended in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer at a concentration of 0.5 μg/mL and applied onto a freshly cleaved template-stripped gold surface. The

sample was incubated at room temperature for 30 minutes before mechanical unfolding experiments were carried out. Example traces from these experiments are shown in figure 4.23. The traces show the poly-protein does have mechanical resistance as expected and the rising edge of each unfolding event could be fitted with the Worm-Like Chain (WLC) model ($\S$2.7.4.1).

4.5.2 Analysis of the SMFS data for (I27-ProtL WT)$_3$-I27 in 0.5 M KCl

Traces were accepted for analysis if they contained at least 2 I27 unfolding events. Traces were not accepted if there were large surface interactions at the start of the trace, there were splitting of peaks or if they lacked a clear detachment peak. For each protein unfolding event the peak unfolding force, F_{UN}, and the peak to peak distance, p2p, were recorded. The mean of all the F_{UN}, median F_{UN} and p2p distance values collected at each pulling velocity were calculated and recorded in tables 4.1, C.1 and C.2 respectively.

The values recorded for I27 were consistent with previous studies of I27 by SMFS by Hoffmann et al. (F_{UN} = 168 $\pm$ 7 pN, 194 $\pm$ 4 pN, p2p = 23.7 (error not quoted) at 400 and 2000 nm/s respectively[148]. The values recorded for ProtL WT were consistent with previous studies of ProtL by SMFS by Brockwell et al. (F_{UN} = 136 $\pm$ 1 pN, 152 $\pm$ 5 pN, p2p = 16.5 $\pm$ 0.1, 16.9 $\pm$ 0.2 at 400 nm/s and 700 nm/s respectively[146].

Pulling velocity (nm/s)	Events	Mean F_{UN} (pN)	SD	Events	Mean F_{UN} (pN)	SD
	ProtL			I27		
400	96	138	23	110	178	23
700	114	145	23	134	179	24
1200	111	160	30	119	196	31
2000	119	169	25	127	203	28

Table 4.1: Mean unfolding force data and standard deviation for (I27-ProtL WT)$_3$-I27 in 0.5M KCl, 5mM MOPS, pH 7.4.

The next step was to plot the data for each pulling velocity in scatter diagrams, showing the F_{UN} and p2p, shown in figure C.5. The diagrams show clear clusters of data
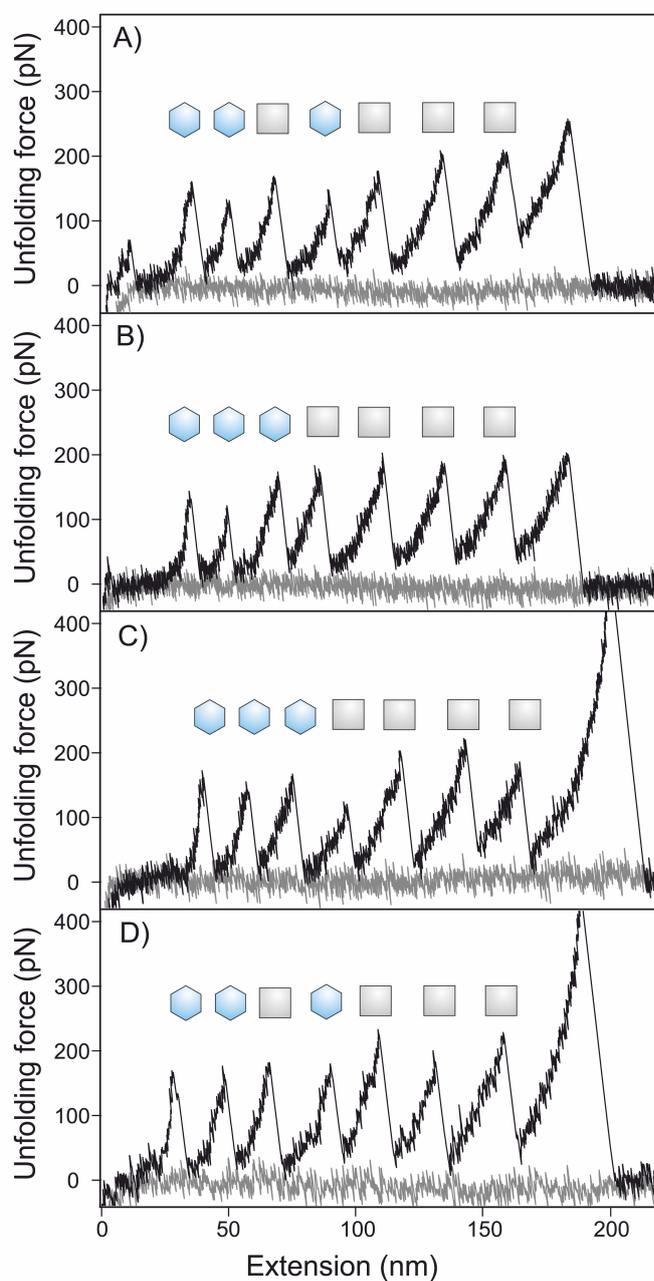
Figure 4.23: Example traces from SMFS experiments using (I27-ProtL WT)$_3$-I27 in 0.5 M KCl, 5 mM MOPS, pH 7.4. Figures A), B), C) and D) all show 4 I27 unfolding events and 3 ProtL unfolding events at 400, 700, 1200 and 2000 nm/s respectively. Blue hexagons represent ProtL WT unfolding events and light grey squares represent I27 unfolding events.

points for each experiment centred around the reference Protein L and I27 data[146,148]. Histograms were then produced for the unfolding force data and the peak to peak distances for each pulling velocity, shown in figure 4.24. Each histogram was fitted with a gaussian curve.

4.5.3 Summary

The results presented show the data from SMFS experiments carried out using the hetero-polyprotein construct (I27-ProtL WT)$_3$-I27. The mechanical fingerprint of the polyprotein was successfully characterised and the unfolding force and peak to peak distance values were calculated. These were shown to agree with published data for both I27 and ProtL, showing the SMFS experiments were successful.

Conclusion: The hetero-polyprotein construct (I27-ProtL WT)$_3$-I27 can successfully be characterised using SMFS.

4.6 Further single molecule force spectroscopy experiments on (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27

This section will show data collected from SMFS experiments using:

i) (I27-ProtL Kx6E)$_3$-I27 in 0.5 M KCl, 5 mM MOPS, pH 7.4

ii) (I27-ProtL WT)$_3$-I27 in 2 M KCl, 5 mM MOPS, pH 7.4

iii) (I27-ProtL Kx6E)$_3$-I27 in 2 M KCl, 5 mM MOPS, pH 7.4.

The analysis carried on this data will also be presented.

Single molecule force spectroscopy (SMFS) experiments were carried out using the same procedure described in section 4.5.1. Although in 2 M KCl, the sample was incubated at room temperature for 60 minutes before mechanical unfolding experiments
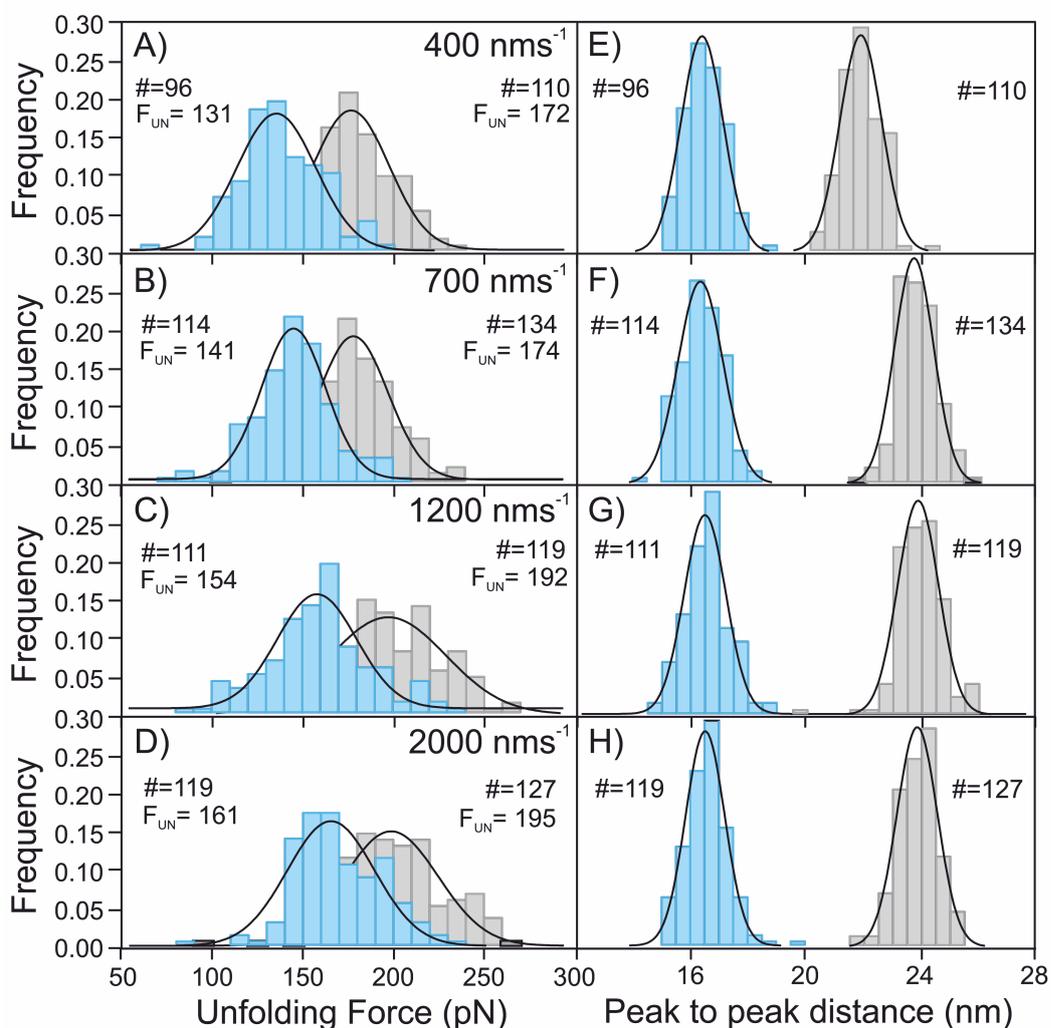
Figure 4.24: Histograms of the unfolding force data and peak to peak distance data for poly-ProtL WT at 400, 700, 1200 and 2000 nm/s in 0.5 M KCl. All figures show ProtL WT in blue and I27 in light grey. Figures A), B), C) and D) show the peak to peak distance data and figures E), F), G) and H) show the unfolding force data from each pulling velocity respectively. All histograms are fitted with a gaussian curve. The poly-protein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL WT and I27.

156

were carried out. The longer incubation time was to improve the binding on cysteine residues to the gold surface. Example traces from these experiments are shown in figures 4.25, 4.26 and 4.27.

4.6.1 Analysis of the SMFS data for (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27

The analysis procedure described in section 4.5.2 was followed. The mean F_{UN}, median F_{UN} and p2p distance values collected at each pulling velocity were calculated and recorded in tables 4.2, C.3 and C.4 for (I27-ProtL Kx6E)$_3$-I27 in 0.5 M KCl, tables 4.3, C.5 and C.5 for (I27-ProtL WT)$_3$-I27 in 2 M KCl and tables 4.4, C.7 and C.8 for (I27-ProtL Kx6E)$_3$-I27 in 2 M KCl.

Scatter diagrams are shown in figures C.6 for (I27-ProtL Kx6E)$_3$-I27 in 0.5 M KCl, C.8 for (I27-ProtL WT)$_3$-I27 in 2 M KCl and C.10 for (I27-ProtL Kx6E)$_3$-I27 in 2 M KCl. Histograms are shown in figures C.7 for (I27-ProtL Kx6E)$_3$-I27 in 0.5 M KCl, C.9 for (I27-ProtL WT)$_3$-I27 in 2 M KCl and C.11 for (I27-ProtL Kx6E)$_3$-I27 in 2 M KCl.

4.6.2 Summary

The mechanical fingerprint of each polyprotein was successfully characterised and the unfolding force and peak to peak distance values were calculated.

Conclusion: The hetero-polyprotein constructs, (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 can both successfully be characterised using SMFS in 0.5 M and 2 M KCl.
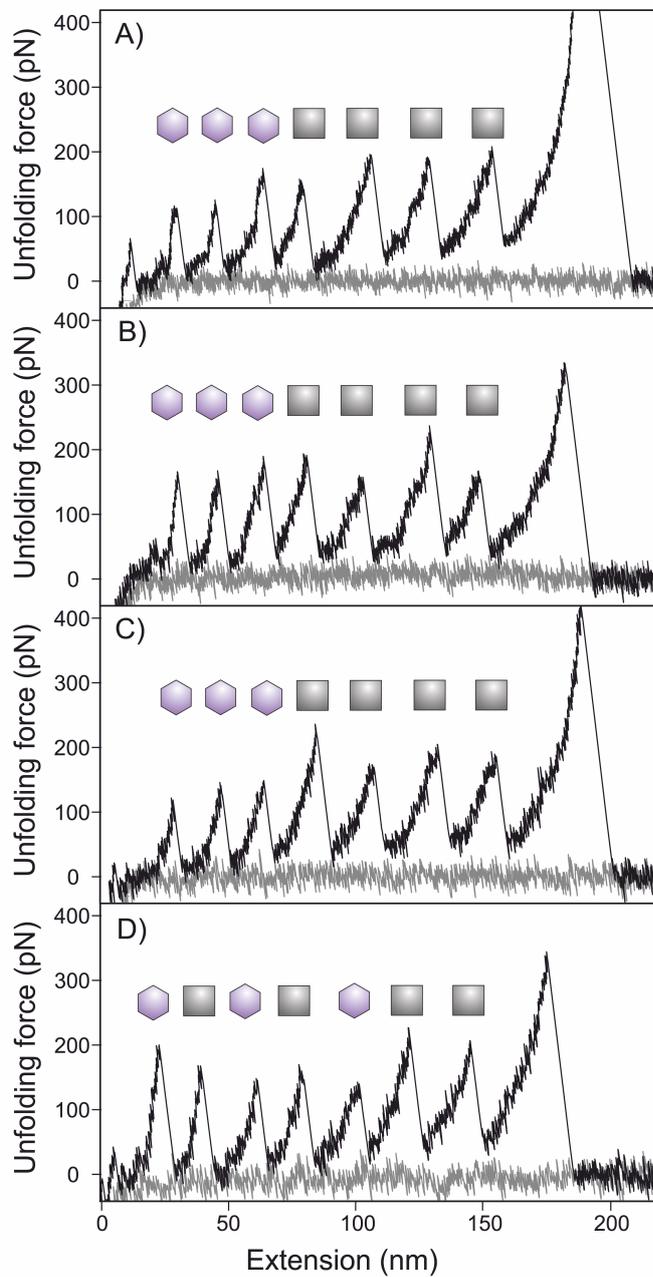
Figure 4.25: Example traces from SMFS experiments using (I27-ProtL Kx6E)$_3$-I27 in in 0.5 M KCl, 5 mM MOPS, pH 7.4. Figures A), B), C) and D) all show 4 I27 unfolding events and 3 ProtL unfolding events at 400, 700, 1200 and 2000 nm/s respectively. Purple hexagons represent ProtL Kx6E unfolding events and grey squares represent I27 unfolding events.
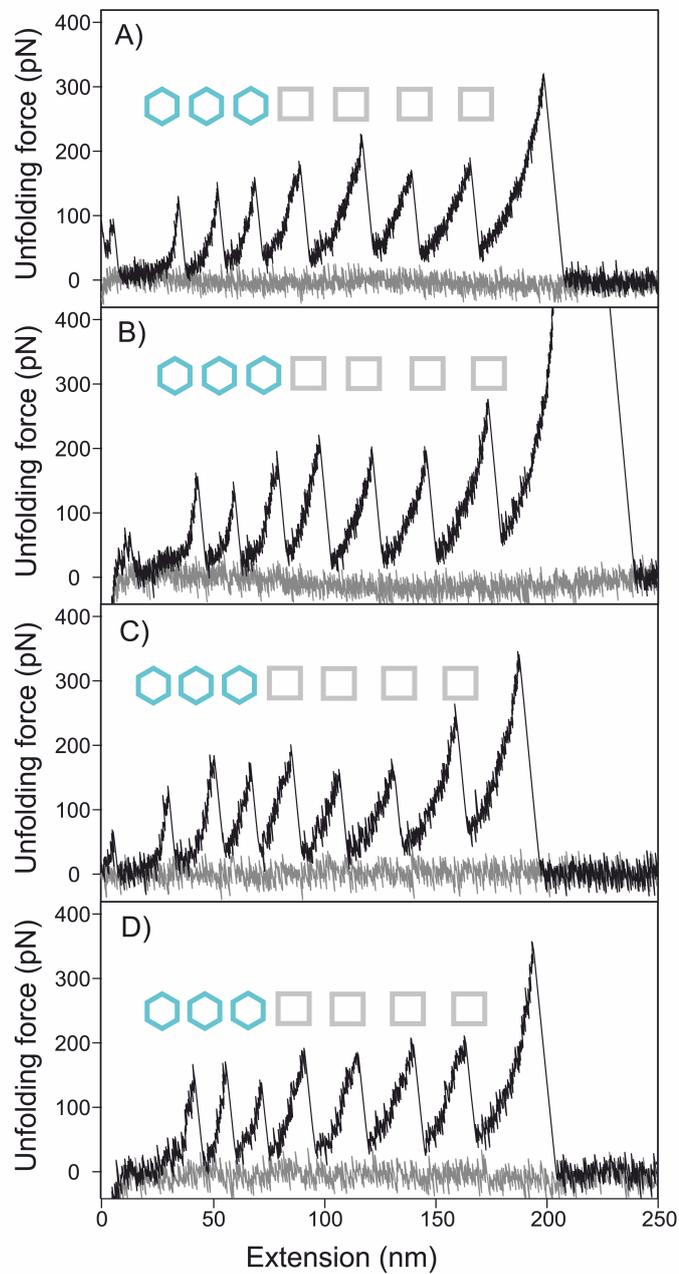
Figure 4.26: Example traces from SMFS experiments using (I27-ProtL WT)$_3$-I27 in 2 M KCl, 5 mM MOPS, pH 7.4. Figures A), B), C) and D) all show 4 I27 unfolding events and 3 ProtL unfolding events at 400, 700, 1200 and 2000 nm/s respectively. Blue hexagons represent ProtL WT unfolding events and light grey squares represent I27 unfolding events.
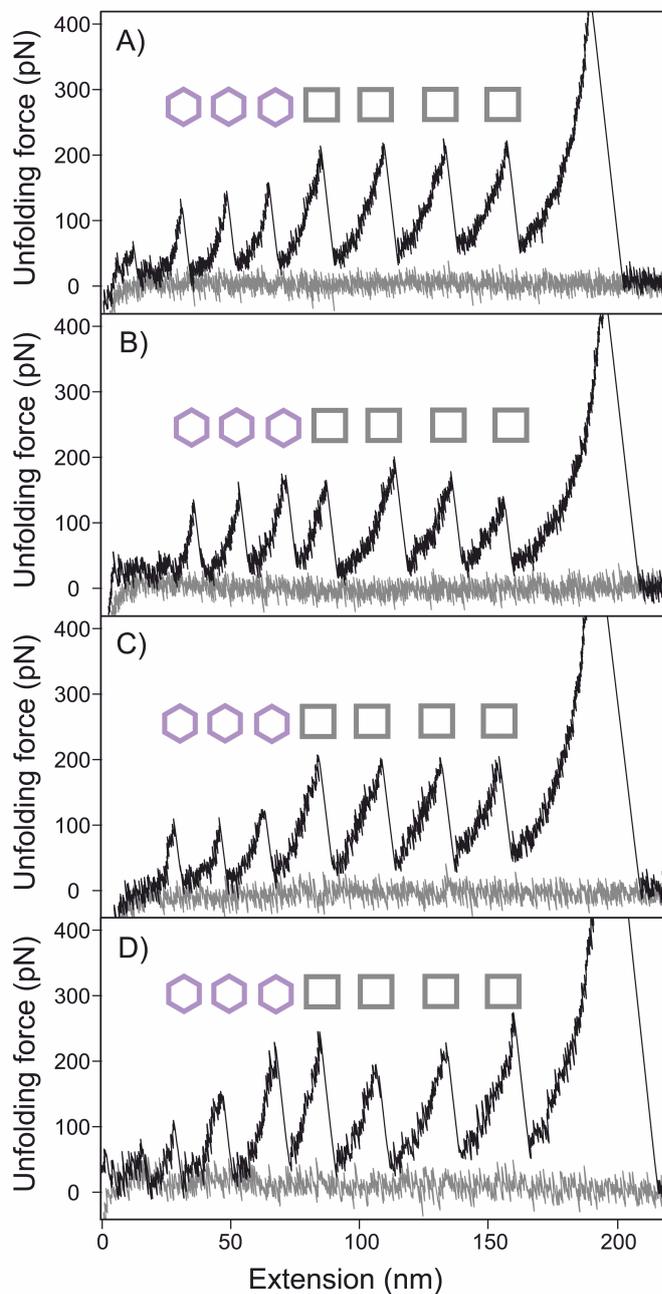
Figure 4.27: Example traces from SMFS experiments using (I27-ProtL Kx6E)$_3$-I27 in in 2 M KCl, 5 mM MOPS, pH 7.4. Figures A), B), C) and D) all show 4 I27 unfolding events and 3 ProtL unfolding events at 400, 700, 1200 and 2000 nm/s respectively. Purple hexagons represent ProtL Kx6E unfolding events and grey squares represent I27 unfolding events.

Pulling velocity (nm/s)	Events	Mean F_{UN} (pN)	SD	Events	Mean F_{UN} (pN)	SD
	ProtL			I27		
400	74	124	22	93	186	25
700	56	133	23	71	188	27
1200	75	135	25	106	195	33
2000	104	154	26	132	210	30

Table 4.2: Mean unfolding force data and standard deviation for (I27-ProtL Kx6E)$_3$-I27 in 0.5M KCl, 5mM MOPS, pH 7.4.

Pulling velocity (nm/s)	Events	Mean F_{UN} (pN)	SD	Events	Mean F_{UN} (pN)	SD
	ProtL			I27		
400	71	155	23	81	197	24
700	50	159	24	62	213	30
1200	82	170	32	92	220	29
2000	68	172	31	86	214	34

Table 4.3: Mean unfolding force data and standard deviation for (I27-ProtL WT)$_3$-I27 in 2M KCl, 5mM MOPS, pH 7.4.

Pulling velocity (nm/s)	Events	Mean F_{UN} (pN)	SD	Events	Mean F_{UN} (pN)	SD
	ProtL			I27		
400	8	137	14	11	199	16
700	49	132	22	67	196	33
1200	53	142	28	65	199	27
2000	23	154	24	30	224	34

Table 4.4: Mean unfolding force data and standard deviation for (I27-ProtL Kx6E)$_3$-I27 in 2M KCl, 5mM MOPS, pH 7.4.

4.7 Comparing the data analysis of the SMFS data for all 4 systems

In order to illuminate the changes in unfolding force values between the different constructs and salt concentrations used, speed dependence diagrams were plotted for all 4 systems. This section will include 4 speed dependence diagrams to highlight differences between each set of data. The differences will be further discussed in sections 6.2 and 6.3. For each speed dependence diagram, the mean F_{UN} value for each pulling velocity was plotted against the $\log_{10}$(pulling velocity). A linear fit is performed for each of the proteins. The speed dependence diagrams for the following data groups are shown in figure 4.28:

A) (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 0.5 M KCl

B) (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 2 M KCl

C) I27 in 0.5 M and 2 M KCl

D) ProtL in 0.5 M and 2 M KCl

Figure 4.28 A) and C) highlights the difference in unfolding force between ProtL WT and ProtL Kx6E, which is consistent at all pulling velocities. Figure 4.28 B) and C) shows that the difference in unfolding force between ProtL WT and ProtL Kx6E is consistent in a higher salt concentration. The reduction of the unfolding force for the halophilic protein is likely to be due to a reduced number of interactions across the two sub-units of protein L that are sheared apart in the unfolding pathway of Protein L[145]. A discussion of the reduction in the unfolding force of ProtL Kx6E compared to ProtL WT can be found in section 6.2.

Figure 4.28 A) and D) show that the I27 unfolding force values are similar for both constructs but the I27 from (I27-ProtL WT)$_3$-I27 is slightly lower than for I27 in (I27-ProtL Kx6E)$_3$-I27 in 0.5 M KCl. Figure 4.28 B) and D) shows the linear fit for I27 from (I27-ProtL WT)$_3$-I27 is slightly higher than for I27 in (I27-ProtL Kx6E)$_3$-I27 in 2 M KCl. The difference in unfolding force for each construct could be due to an unfolding history effect, which is examined in section 4.10.
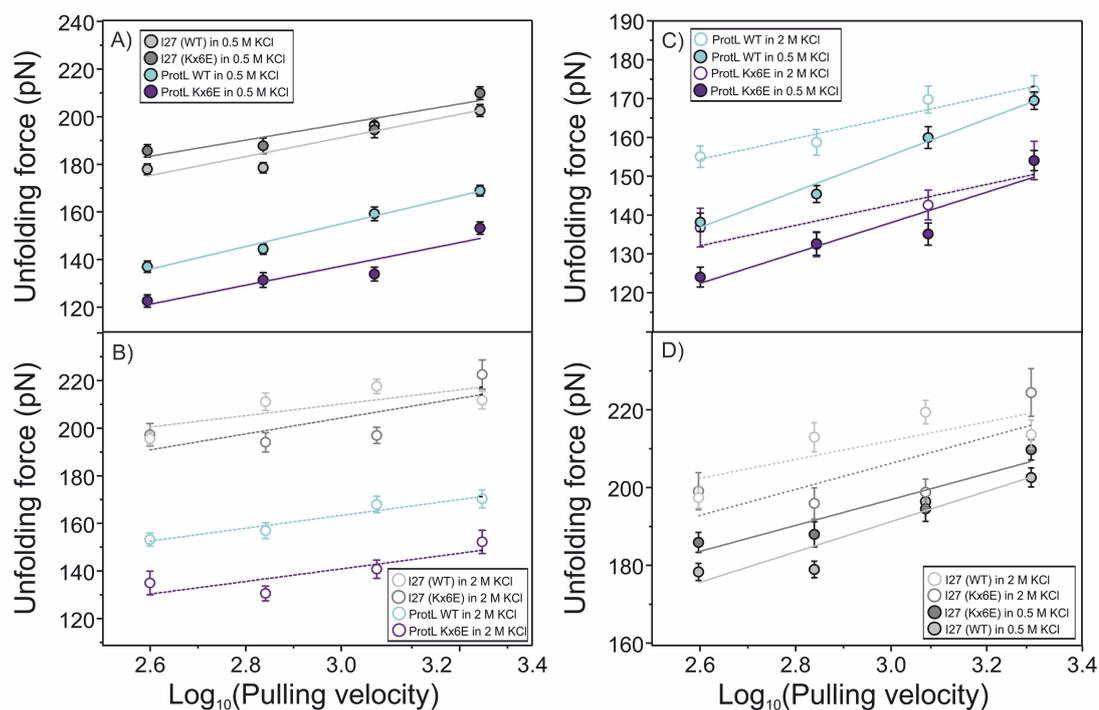
Figure 4.28: Speed dependence diagram of the unfolding force vs. $\log_{10}$(pulling velocity) for all 4 systems. A) shows speed dependence of (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 0.5 M KCl. B) shows the speed dependence of (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 2 M KCl. C) shows the speed dependence of I27 in 0.5 M and 2 M KCl. D) shows the speed dependence of ProtL in 0.5 M and 2 M KCl. The linear fit for each set of data is also plotted as full lines in 0.5 M KCl and dashed lines in 2 M KCl.

163

Figure 4.28 B) shows that all data points collected in 2 M KCl do not lie on, or close to, the linear fit. This suggests that these data may be unreliable. The linear fits for the the data collected at 2 M KCl should, therefore be interpreted with caution. Although each data point does not sit on the linear fit, the data points does seem to follow the same trend so hypotheses may be formed. A diagram of the comparison all of the data in one diagram can be found in figure C.12.

Figure 4.29 shows the percentage increase in the unfolding force for each protein with increased salt concentration, calculated using equation 4.1. This figure highlights the consistent increase in unfolding force of I27 from both constructs and ProtL WT, i.e. all 3 mesophilic proteins, from 0.5 to 2 M salt, with percentage increase averages of 8.6% for ProtL WT, 13.1% for I27 (WT) and 6.8% for I27 (Kx6E). ProtL Kx6E does not show such a consistent increase in force from 0.5 M to 2 M KCl with a percentage increase average of 1.6%. It should be noted that the value quoted for ProtL Kx6E at 400 nm/s in 2 M does not fit the same trend as it shows a large increase in unfolding force from 0.5 M KCl. This value comes from only 8 data points and is, therefore, likely to be inaccurate. It was, therefore, omitted from the calculation of the average percentage increase.

$$\text{Percentage change} = \frac{F_{UN}^{2M} - F_{UN}^{0.5M}}{F_{UN}^{0.5M}} \tag{4.1}$$

The increase in force in 2 M KCl for the mesophilic proteins is likely due to an increase in the hydrophobic effect acting on the protein in the higher salt concentration (see section 1.2.4). The increased hydrophobic effect may not affect ProtL Kx6E as this protein has an increased number of negatively charged residues, which are able to remain flexible. This hypothesis is further discussed in section 6.3.

4.7.1 Summary

The results presented in this section show the speed dependence diagrams formed from SMFS experiments carried out using the constructs (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 0.5 M and 2 M KCl. Although it is possible to form hypotheses from the
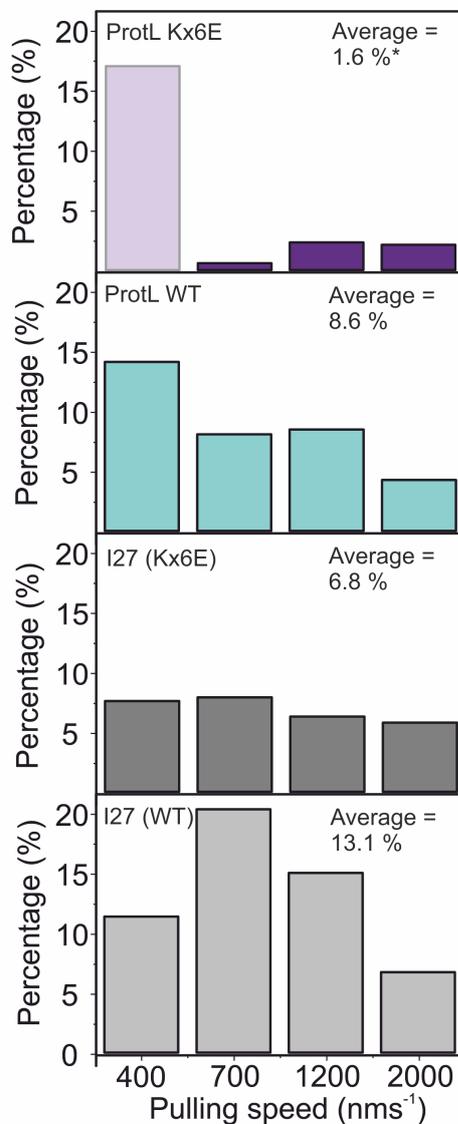
Figure 4.29: Percentage change in the unfolding force of each protein between 0.5 and 2 M KCl. The median unfolding force of each protein in 2 M KCl is shown as a percentage change from its unfolding force in 0.5 M KCl. Where ProtL Kx6E is represented in purple, ProtL WT is represented in blue, I27 from the construct, (I27-ProtL Kx6E)$_3$-I27 is represented in grey and I27 from the construct, (I27-ProtL WT)$_3$-I27, is represented in light grey. *the data for ProtL Kx6E at 400 nm/s comes from only 8 data points and is likely to be inaccurate. It was, therefore, omitted from the calculation of the average percentage increase.

presented speed dependence diagrams, the data collected in 2 M KCl does not form an accurate speed dependence plot as expected, i.e. all data points do not lie on the linear fit.

The linear fit of the speed dependence is very important because it allows access to information about the unfolding energy landscape of the protein. Monte Carlo (MC) simulations use this information and the distribution of forces generated to extract the distance from the unfolded to the transitions state, Δx_U, and the unfolding rate at zero force, k_U[145,146].

Conclusion: Data analysis refinement is needed to increase the accuracy of the data for MC simulations.

4.8 Refining the data analysis of the SMFS data

This section will introduce the refinement analysis procedure carried out on the SMFS data for (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 0.5 M and 2 M KCl. There are a number of elements that may affect the results of an SMFS experiment, the temperature of the room, the calibration of the cantilever, build up of protein on the cantilever, movement of the cantilever in its holder during an experiment and evaporation of the solvent. For example, if one experiment is not calibrated correctly, it will skew the results for the whole set if data it is included in. In each experiment a new cantilever is used to identify any errors in the calibration of the cantilever and reduce these errors as much as possible. Each cantilever comes from the same batch. In order to reduce the effect of these elements on the data, the data was refined by producing histograms for each experiment carried out, i.e. for each day, rather than pooling all of the data. The comparison of these histograms was used to decide whether a data set should be accepted. A data set was accepted if:

- the data set contained 5 or more ProtL unfolding events in total
- one or both of the unfolding force histograms for I27 and ProtL showed a gaussian distribution.

Table 4.5 shows the total number of experiments carried out (across all speeds) for each protein and salt concentration, and the number of accepted data sets for each. It is interesting to note the number of accepted data sets in 2 M KCl is significantly less than for those in 0.5 M KCl. This is likely due to the salt concentration affecting the binding of the polyprotein to the cantilever tip and gold surface, reducing the amount of data collected in one experiment. This will be discussed further in section 6.3.

Protein construct	KCl conc.	No. experiments	No. data sets accepted
(I27-ProtL WT)$_3$-I27	0.5 M	16	12
(I27-ProtL WT)$_3$-I27	2 M	29	6
(I27-ProtL Kx6E)$_3$-I27*	0.5 M	36	12
(I27-ProtL Kx6E)$_3$-I27	2 M	38	5

Table 4.5: Number of accepted data sets vs. number attempted for each system, *includes first experimental optimisation

The mean and the median of the F_{UN} and the p2p distance values were calculated for each separate data set for all systems and recorded in tables 4.6, C.9, C.10, 4.7, C.11, C.12, 4.8, C.13, C.14, 4.9, C.15 and C.16.

The next step was to plot the data for each pulling speed in scatter diagrams, showing F_{UN} and p2p. The scatter diagrams for (I27-ProtL WT)$_3$-I27 in 0.5 M KCl are shown in figures C.13, C.14, C.15 and C.16 for 400 nm/s, 700 nm/s, 1200 nm/s and 2000 nm/s respectively. . The scatter diagrams for (I27-ProtL Kx6E)$_3$-I27 in 0.5 M KCl are shown in figures C.17, C.18, C.19 and C.20. The scatter diagrams for (I27-ProtL WT)$_3$-I27 in 2 M KCl are shown in figures C.21, C.22, C.23 and C.24. The scatter diagrams for (I27-ProtL Kx6E)$_3$-I27 in 2 M KCl are shown in figures C.25, C.26 and C.27. The scatter diagrams all show clear clusters of data points for each experiment centred around the reference Protein L and I27 data where applicable[146,148]. Histograms were then produced for the F_{UN} and p2p distances for each data set. The histograms for the accepted data sets for (I27-ProtL WT)$_3$-I27 in 0.5 M KCl are shown in figures C.28, C.29, C.30 and C.31. The histograms for each accepted data set for (I27-ProtL Kx6E)$_3$-I27 in 0.5 M KCl are shown in figures C.32, C.33, C.34 and C.35. The histograms for the accepted data sets for (I27-ProtL WT)$_3$-I27 in 2 M KCl are shown in figures C.36, C.37, C.38 and C.39. The histograms for the accepted data sets for (I27-ProtL Kx6E)$_3$-I27 in 2 M

Pulling velocity (nm/s)	Events	Mean F_{UN} (pN)	SD	events	Mean F_{UN} (pN)	SD
	ProtL			I27		
400	36	136	22	42	169	21
400	31	142	20	35	184	20
400	29	137	26	33	183	25
		138	**2**		**179**	**5**
700	49	140	24	55	174	21
700	29	152	25	43	184	27
700	36	147	19	36	181	24
		146	**4**		**180**	**3**
1200	19	154	30	24	193	29
1200	43	162	33	47	195	31
1200	49	160	26	48	200	31
		159	**3**		**196**	**2**
2000	36	168	24	39	201	34
2000	39	165	25	43	200	24
2000	44	175	23	45	207	24
		169	**3**		**203**	**2**

Table 4.6: Mean unfolding force data and standard deviation for (I27-ProtL WT)$_3$-I27 in 0.5M KCl, 5mM MOPS, pH 7.4. The numbers quoted in bold are the mean and standard error of the 3 unfolding force values for each pulling velocity.

Pulling velocity (nm/s)	Events	Mean F_{UN} (pN)	SD	events	Mean F_{UN} (pN)	SD
	ProtL			I27		
400	39	132	21	46	187	21
400	28	131	25	32	185	24
400	35	120	20	39	184	29
		128	**4**		**185**	**1**
700	21	134	26	25	187	32
700	20	123	17	27	188	34
700	15	144	19	29	186	20
		133	**6**		**188**	**1**
1200	19	138	25	27	209	25
1200	31	131	23	42	195	40
1200	25	137	25	37	184	25
		135	**2**		**196**	**7**
2000	29	156	29	40	211	39
2000	44	149	25	54	206	23
2000	31	159	24	38	215	28
		155	**3**		**210**	**3**

Table 4.7: Mean unfolding force data and standard deviation for (I27-ProtL Kx6E)$_3$-I27 in 0.5M KCl, 5mM MOPS, pH 7.4. The numbers quoted in bold are the mean and standard error of the 3 unfolding force values for each pulling velocity

Pulling velocity (nm/s)	Events	Mean F_{UN} (pN)	SD	events	Mean F_{UN} (pN)	SD
	ProtL			I27		
400	32	157	26	41	197	22
400	29	152	23	36	196	30
		154	**2**		**197**	**1**
700	17	155	18	20	209	38
700	30	157	25	35	216	31
		156	**1**		**212**	**3**
1200	32	165	33	36	226	22
		165	**2**		**226**	**4**
2000	27	163	26	40	218	34
		163	**5**		**218**	**5**

Table 4.8: Mean unfolding force data and standard deviation for (I27-ProtL WT)$_3$-I27 in 2 M KCl, 5mM MOPS, pH 7.4. The numbers quoted in bold are the mean and standard error of the 3 unfolding force values for each pulling velocity

Pulling velocity (nm/s)	Events	Mean F_{UN} (pN)	SD	events	Mean F_{UN} (pN)	SD
	ProtL			I27		
700	34	130	23	43	186	27
700	16	141	16	23	211	28
		135	**4**		**200**	**10**
1200	25	142	26	32	200	34
		142	**7**		**200**	**5**
2000	17	154	32	23	222	34
2000	10	153	15	15	235	30
		153	**1**		**236**	**6**

Table 4.9: Mean unfolding force data and standard deviation for (I27-ProtL Kx6E)$_3$-I27 in 2 M KCl, 5mM MOPS, pH 7.4. The numbers quoted in bold are the mean and standard error of the 3 unfolding force values for each pulling velocity

KCl are shown in figures C.40, C.41 and C.42 for 700 nm/s, 1200 nm/s and 2000 nm/s respectively. Each histogram was fit to a Gaussian curve.

4.8.1 Summary

The results presented show the refined data analysis from SMFS experiments carried out using the hetero-polyprotein constructs (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 0.5 M and 2 M KCl. In order to increase the accuracy of the data, data sets were separated into separate experiments and either accepted or rejected based on set criteria. Although three data sets were accepted for each speed for both polyproteins in 0.5 M KCl, only a few data sets were accepted for both polyproteins in 2 M, shown in table 4.5. This highlights the difficulties in obtaining SMFS data in a high salt concentration. This will be discussed further in section 6.3.

Conclusion: A refined data analysis was carried out for both (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 0.5 M and 2 M KCl. A significant reduction in the amount of accepted data was shown for both polyproteins in 2 M KCl.

4.9 Comparing the refined data analysis of the SMFS data for (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 0.5 M and 2 M KCl

In order to illuminate the changes in unfolding force values between the different constructs and salt concentrations used, speed dependence diagrams were plotted for the refined analysis of all 4 systems. The F_{UN} value for each pulling velocity was plotted against the $\log_{10}$ (pulling velocity). A linear fit is performed for each of the proteins. The speed dependence diagrams for the refined analysis of the following data groups are shown in figure 4.30:

A) (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 0.5 M KCl

B) (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 2 M KCl

C) I27 in 0.5 M and 2 M KCl

D) ProtL in 0.5 M and 2 M KCl
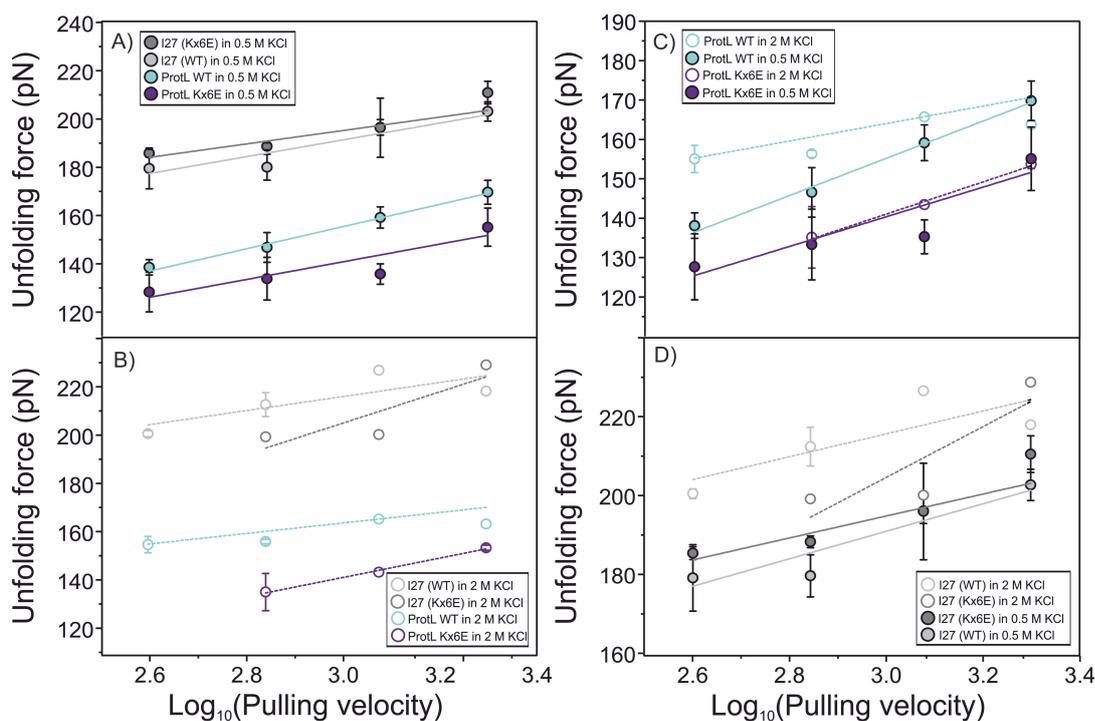
Figure 4.30: **Speed dependence diagram of the unfolding force vs. log$_{10}$(pulling velocity) for the refined analysis of all 4 systems.** A) shows speed dependence of (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 0.5 M KCl. B) shows the speed dependence of (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 2 M KCl. C) shows the speed dependence of I27 in 0.5 M and 2 M KCl. D) shows the speed dependence of ProtL in 0.5 M and 2 M KCl. The linear fit for each set of data is also plotted as full lines in 0.5 M KCl and dashed lines in 2 M KCl.

Figures 4.30 A) - D) all show the same trends as the first analysis (figure 4.28), showing the refined data analysis does not affect the interpretation of the behaviour of each polyprotein. The data for ProtL Kx6E is shown to be more accurate with refined data analysis, i.e. the linear fitting agrees with all three data points. This figure highlights that with the refined speed dependence, the linear fit of ProtL Kx6E in 2 M KCl now follows the same fit for ProtL Kx6E in 0.5 M KCl. Confirming the hypothesis that the salt concentration does not affect the unfolding force of the halophilic ProtL

172

Kx6E (see section 4.7).

The accuracy of the data for figure 4.30 A) does not improve with the refined analysis. This is not surprising as this data was described as accurate previously. The refined data analysis does not improve the accuracy of the I27 data for both constructs in 2 M KCl, shown in figures 4.30 B) and D). None of the data points sit on the linear fit. This is possibly due to the lack of data in 2 M KCl, highlighted in table 4.5. A diagram of the comparison all of the data in one diagram can be found in figure C.43.

4.9.1 Summary

The results presented in this section show the speed dependence diagrams formed from the refined data anlysis of SMFS experiments carried out using the constructs (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 0.5 M and 2 M KCl.

It is clear the reduced amount of data in 2 M KCl severely affects the accuracy of the data. The refined data analysis does show promise for ProtL Kx6E in 2 M KCl. The linear fit of ProtL Kx6E data in 2 M KCl now follows the same fit for ProtL Kx6E in 0.5 M KCl, confirming the hypothesis that the salt concentration does not affect the unfolding force of the halophilic ProtL Kx6E.

Conclusion: An increased number of unfolding events per day is needed to help improve the accuracy speed dependence plot and, therefore, form solid conclusions.

4.10 Analysing the unfolding history of the SMFS data for (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 0.5 M and 2 M KCl

Many previous SMFS studies assumed that in a hetero-polyprotein, the domain with the fastest unfolding rate must unfold first at a given force[147]. Also it was assumed that all

unfolding forces are equivalent in a homo-polyprotein, the only differences originating from thermal fluctuations of the polyprotein [132,202]. The unfolding force of a protein has been shown to depend on two competing effects, with every domain that unfolds, the probability of an extra domain unfolding is reduced [152]. As a consequence the unfolding force of later events is larger than earlier events. The other effect comes from that fact that later unfolding events have a lower loading rate, i.e. the force is applied to the domain in a longer time-scale, causing a reduction in the unfolding force of later events. The use of a polyprotein for determining the unfolding force for a single protein is justified by the dependence of the unfolding force on the unfolding history is less than the thermal fluctuations of the cantilever [203].

In order to check the effect of the unfolding history on the measured values of the unfolding force, all traces from were named either type A or type B:

- Type A - All traces where all ProtL unfolding events occur before the I27 unfolding events
- Type B - All traces where the unfolding events are in a different order, e.g. any traces that have one or more I27 unfolding events occurring before a ProtL unfolding event.

The traces were separated for each pulling velocity and the mean unfolding force for each trace type was calculated.

4.10.1 The unfolding history of the SMFS data for all 4 systems

The calculated mean unfolding force for each trace type for is shown in the tables 4.10, 4.11, 4.12 and 4.13.

The results show that (I27-ProtL WT)$_3$-I27 has a large amount of type B traces as well as type A in both 0.5 M and 2 M KCl (tables 4.10 and 4.12). ProtL WT has an unfolding force of 138 pN at 400 nm/s, which sometimes unfolds after the I27 (shown in table 4.1. (I27-ProtL Kx6E)$_3$-I27 for both 0.5 M and 2 M KCl, on the other hand, does not have a large amount of type B, they are mostly type A (tables 4.11 and 4.13). ProtL

Pulling velocity (nm/s)	Events	F_{UN} (pN)	Events	F_{UN} (pN)	Events	F_{UN} (pN)	Events	F_{UN} (pN)
	ProtL				I27			
	Type A		Type B		Type A		Type B	
400	67	136	29	142	73	181	37	172
700	73	146	47	142	76	183	58	174
1200	58	156	53	164	60	202	54	190
2000	52	171	51	170	49	208	55	196

Table 4.10: Unfolding history analysis for (I27-ProtL WT)$_3$-I27 in 0.5 M KCl where type A. are the traces that unfold all ProtL first and all I27 second and type B. are all traces that unfold in a different order. The table shows the number of unfolding events and the mean unfolding force, F_{UN} for each system.

175

Pulling velocity (nm/s)	ProtL				I27			
	Type A		Type B		Type A		Type B	
	Events	F$_{\text{UN}}$ (pN)	Events	F$_{\text{UN}}$ (pN)	Events	F$_{\text{UN}}$ (pN)	Events	F$_{\text{UN}}$ (pN)
400	54	125	9	125	71	187	10	175
700	35	133	3	150	47	192	2	185
1200	52	130	0	-	68	192	0	-
2000	68	156	12	156	84	213	15	193

Table 4.11: Unfolding history analysis for (I27-ProtL Kx6E)$_3$-I27 in 0.5 M KCl where type A. are the traces that unfold all ProtL first and all I27 second and type B. are all traces that unfold in a different order. The table shows the number of unfolding events and the mean unfolding force, F$_{\text{UN}}$ for each system.

176

Pulling velocity (nm/s)	ProtL				I27			
	Type A		Type B		Type A		Type B	
	Events	F_{UN} (pN)	Events	F_{UN} (pN)	Events	F_{UN} (pN)	Events	F_{UN} (pN)
400	50	154	13	153	52	201	19	194
700	46	159	4	159	56	215	6	198
1200	x	x	x	x	62	222	30	213
2000	44	162	23	171	47	224	32	203

Table 4.12: Unfolding history analysis for (I27-ProtL WT)$_3$-I27 in 2 M KCl where type A. are the traces that unfold all ProtL first and all I27 second and type B. are all traces that unfold in a different order. The table shows the number of unfolding events and the mean unfolding force, F_{UN} for each system.

Pulling velocity (nm/s)	ProtL				I27			
	Type A		Type B		Type A		Type B	
	Events	F$_{\text{UN}}$ (pN)	Events	F$_{\text{UN}}$ (pN)	Events	F$_{\text{UN}}$ (pN)	Events	F$_{\text{UN}}$ (pN)
400	8	137	0	-	11	199	0	-
700	45	133	4	126	62	197	5	184
1200	48	142	5	146	59	199	6	200
2000	22	153	1	178	28	226	2	200

Table 4.13: Unfolding history analysis for (I27-ProtL Kx6E)$_3$-I27 in 2 M KCl where type A. are the traces that unfold all ProtL first and all I27 second and type B. are all traces that unfold in a different order. The table shows the number of unfolding events and the mean unfolding force, F$_{\text{UN}}$ for each system.

Kx6E has a smaller unfolding force (124 pN at 400 nm/s, shown in table 4.2) compared to ProtL WT, which is less likely to unfold after the I27.

The unfolding forces of ProtL for type A compared to type B traces are very similar for all pulling velocities for both ProtL WT and ProtL Kx6E (for all tables). This suggests the unfolding history does not have an effect on the mean unfolding force of Protein L. The unfolding force of I27 from type A is higher in force than for type B for all pulling velocities. This suggests that the unfolding history does have an affect on the mean unfolding force of I27. Although since there are less type B traces for (I27-ProtL Kx6E)$_3$-I27 than for (I27-ProtL WT)$_3$-I27, the unfolding history should have less of an effect on I27 from (I27-ProtL Kx6E)$_3$-I27 than I27 from (I27-ProtL WT)$_3$-I27.

4.10.2 Summary

The results presented in this section show the unfolding history analysis of SMFS experiments carried out using the constructs (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 0.5 M and 2 M KCl. The results suggest the unfolding history does not have an effect on the mean unfolding force of ProtL, i.e. the unfolding history does not account for the difference in force between ProtL WT and ProtL Kx6E. The results also suggest that the unfolding history does have an affect on the mean unfolding force of I27, it is expected that the unfolding force I27 from the wild type construct will have a lower unfolding force than I27 from the halophilic construct. This trend is shown to be consistent for the speed dependence diagram of both constructs in 0.5 M KCl, figure 4.28 A). It is also consistent for the speed dependence diagram of both constructs in 2 M KCl, figure 4.28 B) at 400 and 2000 nm/s but not at 700 and 1200 nm/s. This could be due to the inaccuracy of the data in 2 M KCl or from additional effects from being in a construct with a highly negatively charged protein. This will be discussed further in section 6.3.

Conclusion: the unfolding force of I27 from the wild type construct has a lower unfolding force than I27 from the halophilic construct in 0.5 M KCl, which may be due to the effect of unfolding history.

4.11 Chapter summary

This chapter first presented the work that was carried out on to synthesise two hetero-polyprotein constructs

 i) (I27-ProtL WT)$_3$-I27
 ii) (I27-ProtL Kx6E)$_3$-I27.

Production of both constructs was shown to be successful.

This chapter also presents the experimental data collected and analysed from SMFS experiments using these polyprotein constructs in two salt concentrations. A further discussion of the results can be found in sections 6.2 and 6.3. Initial analysis showed some inaccuracy of the data collected in 2 M KCl and refinement of the data analysis procedure was needed. Re-analysis of the SMFS experimental data by separating data sets was then be presented for both polyproteins. A significant reduction in the amount of accepted data was shown for both polyproteins in 2 M KCl and it was concluded that an increased number of unfolding events per day was needed. Finally an investigation of the unfolding history effects on the SMFS results for the two polyproteins was presented. The results showed that the unfolding history effect causes the I27 from the wild type construct to have a lower unfolding force than I27 from the halophilic construct. This trend was shown to be consistent for the speed dependence diagram of both constructs in 0.5 M KCl.

Chapter 5

Results 3: Synthesis and characterisation of two homo-polyproteins to optimise the collection of SMFS data

5.1 Introduction

This chapter will describe all experiments carried out to synthesise and mechanically unfold two homo-polyproteins containing wild type Protein L and a halophilic model protein, Protein L Ex6D. First, the studies by Tadeo at al., which were described in section 1.3.3, will be be revisited. The rationale for using the halophilic model protein, ProtL Ex6D, as a model halophilic system for the homo-polyprotein studies will then be explained[111]. Next the synthesis techniques in the literature which have previously been used to produce homo-polyproteins will be reviewed. The results of my studies will then be presented, which includes:

   i) Experimental results will be presented for the synthesis of a homo-polyprotein

of Protein L (poly-ProtL WT) that includes two mutations to provide cysteine residues at either end of the protein with the use of a crosslinker.

ii) The data collected from force spectroscopy experiments carried out on the poly-ProtL WT sample will then be presented and discussed.

iii) All the experimental steps carried out to synthesise a homo-polyprotein construct using ProtL Ex6D (poly-ProtL Ex6D), using the same technique used for poly-ProtL WT will be presented.

iv) All the experimental steps carried out to synthesise poly-ProtL Ex6D using second technique, which utilises maltose binding protein (MBP) will be presented.

v) Finally, the data collected from force spectroscopy experiments carried out on the poly-ProtL Ex6D sample will then be presented and discussed.

5.1.1 Properties of ProtL Ex6D

A study was carried out by Tadeo at al. on the mechanism of salt stabilisation of halophilic proteins using a systematic investigation of mutated variants of ProtL, this study was described in more detail in section §1.3.3. A range of mutations were investigated, including charge preservation mutations, size preserving mutations and mutations changing both size and charge[111]. $\Delta G^{\text{U-F}}_{\text{3.2M salt}}$ was determined for each mutation of ProtL by equilibrium denaturation experiments and was plotted against the number of substituted residues. The results for ProtL glutamic acid to aspartic acid (E to D) mutations is shown in figure 5.1 as black circles.
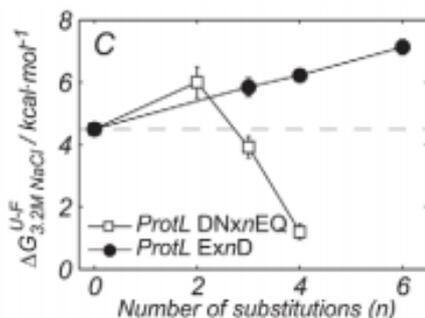
Figure 5.1: The free energy of ProtL at 3.2 M NaCl) versus the number of substituted residues of glutamic acid for aspartic acid. Figure taken from [111].

182

Halophilic proteins have been previously reported to have a large increase in the number of acidic residues, mainly aspartic acids, although the reason for this increase is not yet fully understood[75]. The study by Tadeo et al. showed that a salt dependant stabilising effect was generated for ProtL by reducing one or more amino acids in the protein sequence by one methylene group (e.g. E and Q to D and N) while conserving charge[111]. The opposite mutation, increasing one or more amino acids by one methylene group (D and N to E and Q), was found to have reverse effect by destabilised Prot L in 3.2 M NaCl, shown in figure 5.1 as white squares.

It was found that charge variation has a smaller effect on salt stabilisation compared to size variation. ProtL Ex6D was, therefore, chosen as a halophilic model protein because of the increased salt stabilisation to see if this had a noticeable effect on the mechanical stability of the protein.

5.1.2 Synthesis techniques of homo polyproteins

Homo-polyproteins may be formed by crosslinking monomer proteins, this provides a faster method of producing polyproteins than a recombinant method[197]. The disadvantage of this method is that it provides less control over the precise orientation or sequence of domains (if a hetero-polyprotein is being produced) and length of polyproteins produced.

The crosslinking of proteins using cysteines was originally carried out using solid state synthesis by formation of a disulphide bridge between the side chains[202]. The technique was later shown to be successful and regularly used for proteins in solution[144,204,205]. Polymerisation via this method allows precise control over the binding position along the protein chain and, therefore, over the position of force application[204].

The formation of disulphide bonds between protein monomers is a slow process and the use of disulphide bonds limits the study of the mechanical properties of a protein in different redox conditions[186]. Zheng et al. developed a method of polymerisation of a protein using a crosslinker[197]. This procedure provided a much less time-consuming

method of polyprotein formation while still allowing study under different redox conditions. This method was thus chosen for the synthesis of the polyproteins in this chapter.

The crosslinker chosen was 1,11-bismaleimidotriethyleneglycol (BM(PEG)$_3$), shown in figure 5.2. It contains two maleimide groups, which react with sulphydryl groups to form a thioether linkage, depicted in figure 5.3. The reaction is not reversible. It has three polyethylene glycol (PEG) units that provide a spacer arm of 17.8 Å.
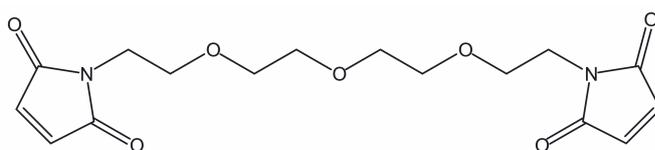
Figure 5.2: Structure of 1,11-bismaleimidotriethyleneglycol (BM(PEG)$_3$.
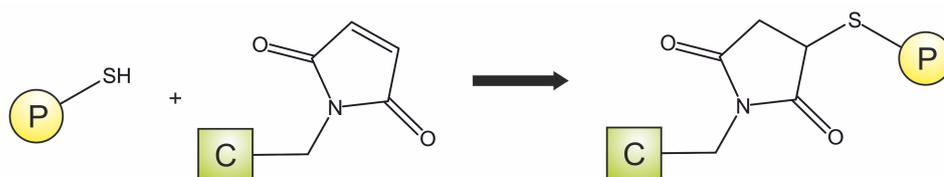
Figure 5.3: Schematic of the reaction of a maleimide with a thiol. Where P represents the protein and C represents the crosslinker.

5.2 Objectives

The objective of this chapter was firstly to produce a homo-polyprotein of ProtL WT to determine if the planned method of polyprotein production would yield a construct suitable for analysis by AFM. The second objective was to produce a homo polyprotein of the halophilic protein L variant. The third objective was to carry out an AFM experiment using the halophilic model polyprotein and identify a mechanical fingerprint. The final objective was to analyse the SMFS data produced and determine whether this method of polyprotein production is suitable for increasing the amount of data produced in SMFS experiments.

5.3 Experimental Results

5.3.1 Production of a homo-polyprotein of wild type Protein L

This section will describe the experiments completed to synthesise a homo-polyprotein of Protein L, shown in figure 5.4.
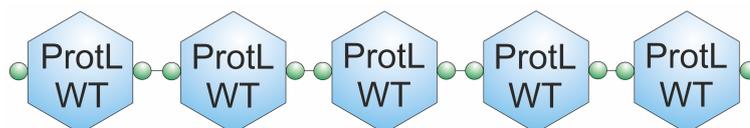
Figure 5.4: Structure of the poly-ProtL WT construct. The construct is polymerised by N- and C-terminal cysteine residues, shown as green circles.

The cysteine residues, located at each terminus of the protein monomer, facilitate the production of a polyprotein via maleimide-thiol coupling[206]. Further explanation of the techniques used can be found in chapter 2, including a basic description of what each step is and the protocols used to carry the technique out. A schematic of the overall process used to produce the polyprotein is shown in figure 5.5. Steps A) and B) in the figure were carried out by Dr David Brockwell.

5.3.1.1 Expression of ProtL WT

pET15b-ProtL WT plasmid was transformed into competent *E. coli* BL21 (DE3) pLysS cells using the protocol described in §2.3.1. A 9 L growth of ProtL WT was set up using 9 x 1 L LB medium in 2.5 L conical flasks. Expression of the construct was induced by IPTG at a final concentrations of 1 mM at an OD_{600} of 0.7. Cultures were allowed to grow for a further 3 hours before harvesting by centrifugation. SDS-PAGE analysis showed the expression of ProtL WT; a band at the expected sizes of ProtL WT of 8.1 kDa was found in the lysate 3 hours after induction. The cell pellet was resuspended in an appropriate buffer and cell lysis was performed using a cell disrupter system. ProtL WT was purified from the soluble fraction by Ni affinity chromatography (see §2.3.6).
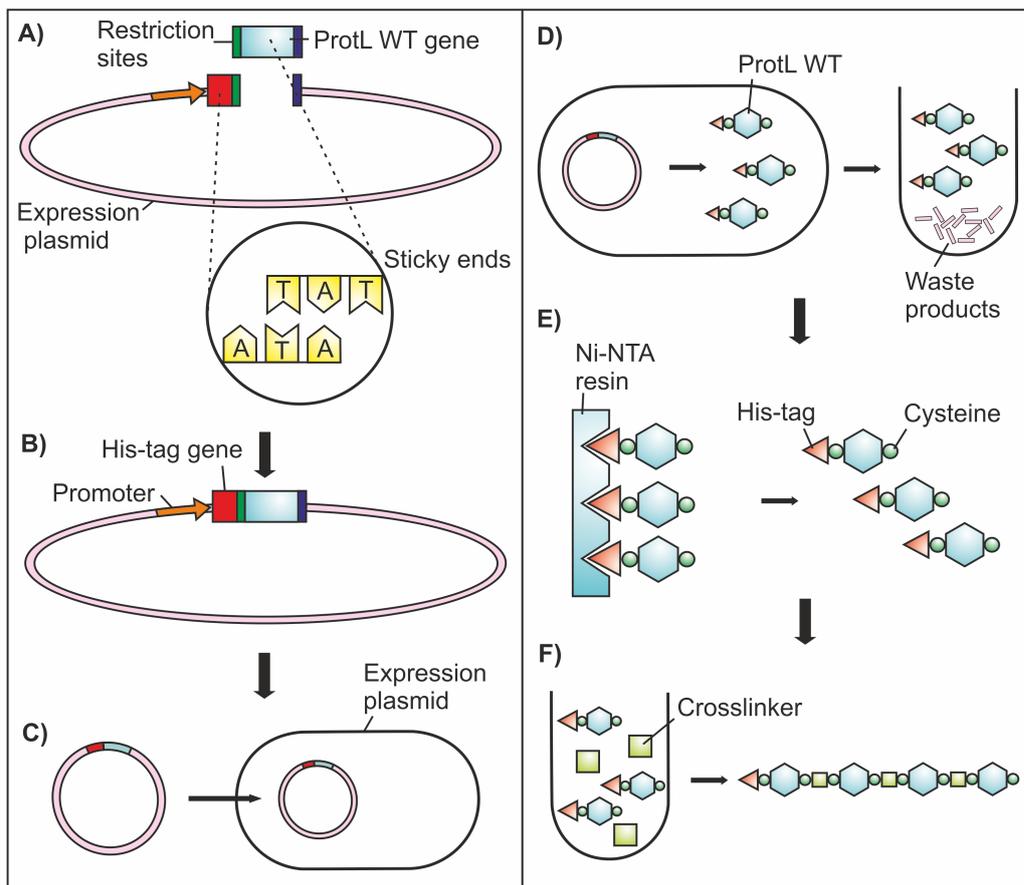
Figure 5.5: Schematic of the synthesis of poly-ProtL WT. Where A) represents the digestion of the ProtL insert and the plasmid to create sticky ends, B) represents the ligation of the ProtL WT gene into the plasmid, C) represents the transformation of the plasmid into an expression *E. coli* strain, D) represents the expression of ProtL WT protein, E) represents purification of the ProtL WT monomer and F) represents the crosslinking of ProtL monomer to form a polymer.

SDS-PAGE analysis of the fractions is shown in figure 5.6. Fractions containing ProtL WT were pooled, dialysed into H_2O and lyophilised.
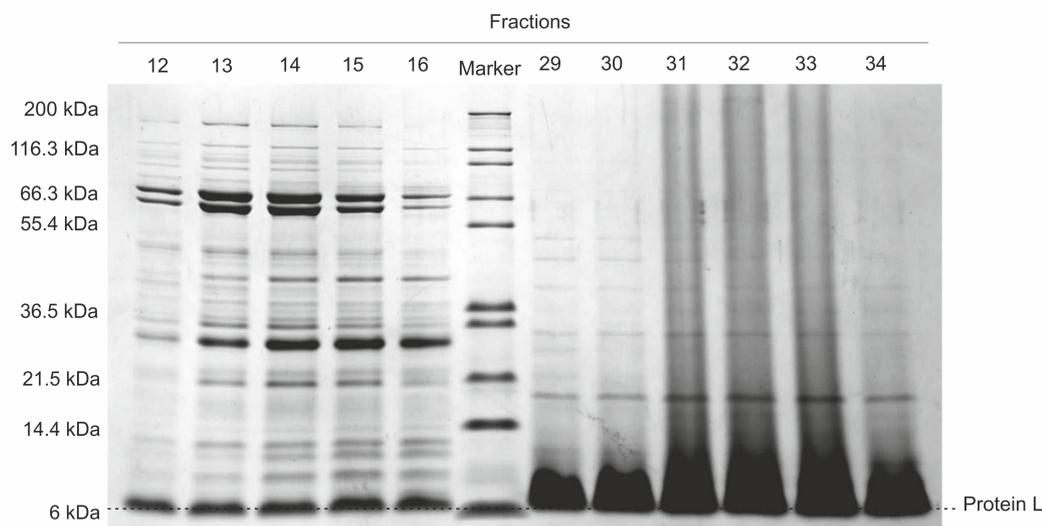
Figure 5.6: Large scale purification of ProtL WT using Ni-NTA chromatography, analysed by SDS-PAGE with Coomasie blue staining. The gel shows a band of significant intensity above 6 kDa, which would correspond to ProtL WT in all Ni fractions. ProtL WT was shown to run slightly lower than expected.

The protein was further purified by size exclusion chromatography using a Superdex 75 HiLoad 26/600 column. SDS-PAGE analysis (figure 5.7) of the resulting fractions shows a band with significant intensity above 6 kDa, with no other protein bands, indicating resulting ProtL WT samples contain no other proteins. Fractions containing ProtL WT were again pooled, dialysed into H_2O and lyophilised.

5.3.1.2 Crosslinking of ProtL WT

A sample of ProtL WT was resuspended in buffer containing 50 mM Tris, pH 7.4 to a concentration of 2 mg/ml and a volume of 5 mL. 1,11-bis(maleimido)triethylene glycol (BM(PEG)$_3$) was added to three separate reactions with final molar ratios of ProtL WT to BM(PEG)$_3$ of 1:0.8, 1:1 and 0.8:1. All 3 reactions were incubated at room temperature. Samples of each reaction were quenched in 50 mM cysteine at separate time-points of 2 hrs, 4 hrs and overnight to find optimal conditions. The samples were

187
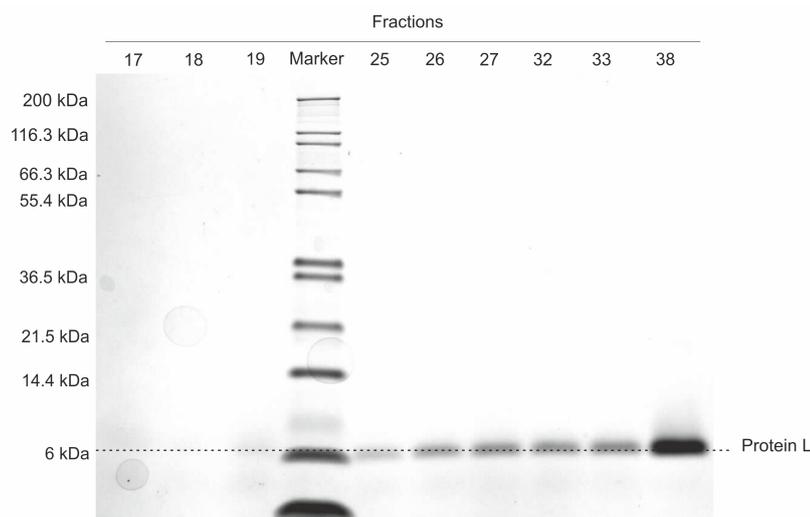
Figure 5.7: Purification of ProtL WT using size exclusion chromatography, analysed by SDS-PAGE with Coomasie blue staining. The gel shows a band with significant intensity above 6 kDa, which would correspond to ProtL WT in fractions 25-38.

analysed using SDS-PAGE, shown in figure 5.9. A wide range of polymeric species was formed. The molecular weight of a specific polymer of ProtL WT can be calculated using equation 5.1, where $MW_{polymer}$ is the molecular weight of the polymer, $MW_{monomer}$ is the molecular weight of the monomer (ProtL WT including His–tag = 8.1 kDa), MW_{RS} is the molecular weight of the repeating section (ProtL WT not including the His–tag = 6.9 kDa), n is the number of protein monomers in the polymer and MW_{CL} is the molecular weight of the cross linker. A representation of this calculation is shown in figure 5.8.
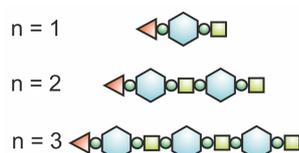
Figure 5.8: Representation of the estimation of crosslinking length. Where the His–tag is represented by a red triangle, the cysteines are represented by dark green circles, ProtL WT is represented by blue hexagons and the crosslinker is represented by green squares.

$$MW_{\text{polymer}} = MW_{\text{monomer}} + (MW_{\text{RS}})_{n-1} + (MW_{\text{CL}})_{n+1} \qquad (5.1)$$
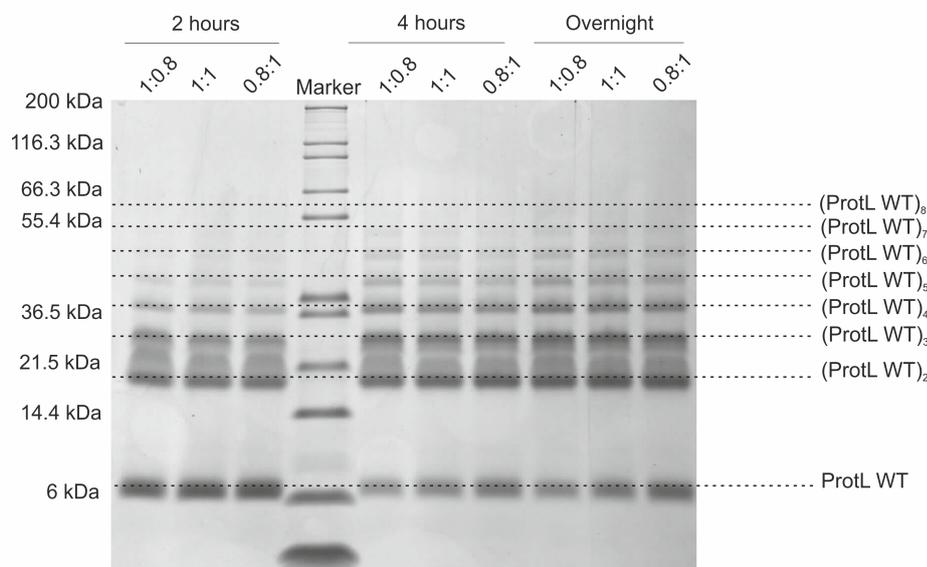
Figure 5.9: The crosslinking of ProtL WT using BM(PEG)$_3$ using 3 different ratios at 2 hours, 4 hours and overnight, analysed by SDS-PAGE with Coomasie blue staining. The gel shows a band of significant intensity above 6 kDa which would correspond to the monomer ProtL WT, 16.0 kDa - (ProtL WT)$_2$, 23.3 kDa - (ProtL WT)$_3$, 30.5 kDa - (ProtL WT)$_4$, 37.8 kDa - (ProtL WT)$_5$, 45.0 kDa - (ProtL WT)$_6$, 52.3 kDa - (ProtL WT)$_7$ and 59.5 kDa - (ProtL WT)$_8$.

The polymerised ProtL WT solution was dialysed into H$_2$O and lyophilised. The polyprotein was purified by size exclusion chromatography using a Superdex 75 column to remove monomer, dimer and trimer polyprotein constructs. The SDS-PAGE analysis of the fractions is shown in figure 5.10.

To carry out the crosslinking on a large scale, 5 mg ProtL WT was resuspended in buffer 50 mM Tris, pH 7.4 to a concentration of 2 mg/ml. BM(PEG)$_3$ was added with a final molar ratios of ProtL WT to BM(PEG)$_3$ of 1:1 . The reaction was incubated at room temperature and quenched after 6 hours.

The resultant product was purified by size exclusion chromatography using a Superdex 75 10/300 column to remove smaller constructs. SDS-PAGE analysis showed the
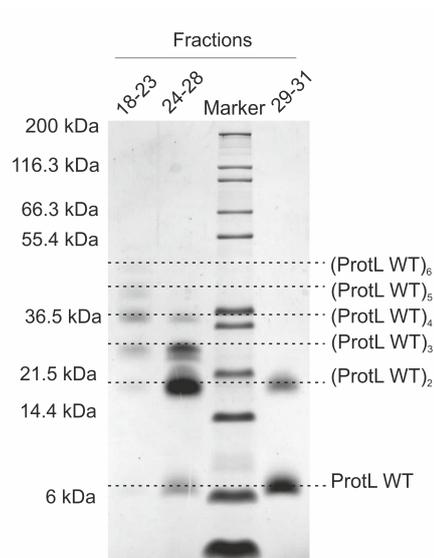
Figure 5.10: Purification of crosslinked ProtL WT using size exclusion chromatography, analysed by SDS-PAGE with Coomasie blue staining. The gel shows a band with significant intensity above 6 kDa which would correspond to the monomer ProtL WT in fractions 29-31, at 16.0 kDa - $(ProtL\ WT)_2$ and 23.2 kDa - $(ProtL\ WT)_3$ in fractions 24-28, 30.5 kDa - $(ProtL\ WT)_4$, 37.8 kDa - $(ProtL\ WT)_5$ and 45.0 kDa - $(ProtL\ WT)_6$ in fractions 18-23.

separation of the resultant poly-ProtL WT by size, figure 5.11.

5.3.1.3   Summary

The results presented show the work carried out to synthesise a homo polyprotein construct of ProtL WT. Construction of the homo construct was successful using this technique.

Conclusion: Production of the polyprotein of ProtL WT using a crosslinker between cysteine residues was shown to be a successful approach.

5.3.2   Single Molecule Force Spectroscopy on poly-ProtL WT

In this section I will show data collected from SMFS experiments using poly-ProtL WT. I will also present the analysis carried on this data.
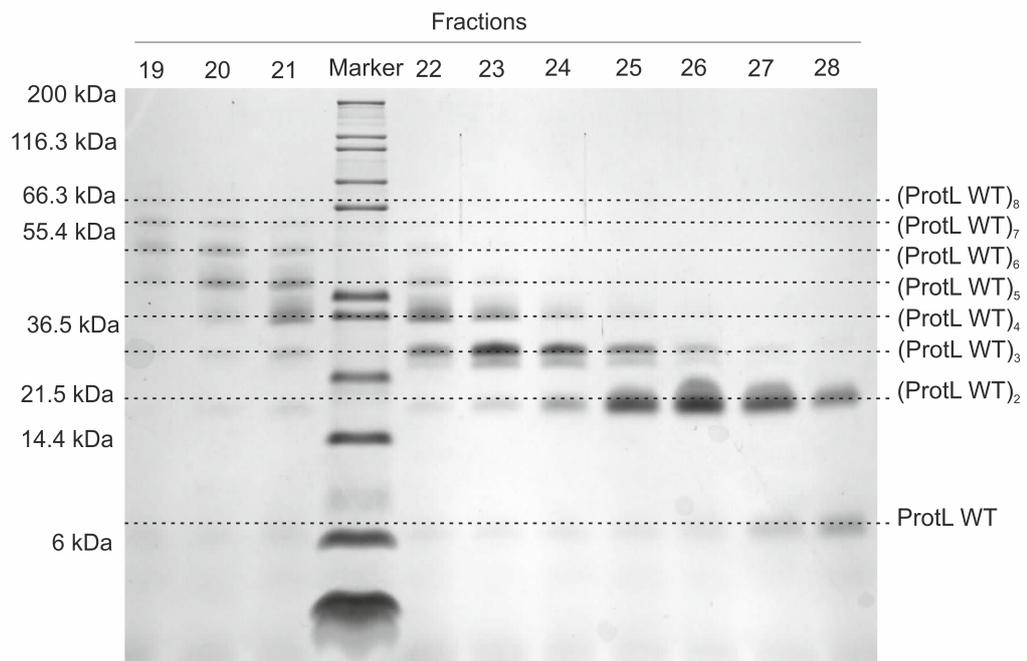
Figure 5.11: Large scale purification of crosslinked ProtL WT using size exclusion chromatography, analysed by SDS-PAGE with Coomasie blue staining. The gel shows a band with significant intensity above 6 kDa which would correspond to the monomer ProtL WT, 16.0 kDa - (ProtL WT)$_2$, 23.3 kDa - (ProtL WT)$_3$, 30.5 kDa - (ProtL WT)$_4$, 37.8 kDa - (ProtL WT)$_5$, 45.0 kDa - (ProtL WT)$_6$, 52.3 kDa - (ProtL WT)$_7$ and 59.5 kDa - (ProtL WT)$_8$.

5.3.2.1 Procedure for SMFS experiments using poly-ProtL WT

In order to validate this synthesis method for the production of poly-ProtL WT for SMFS analysis, three SMFS experiments were carried out. The results of these experiments were then compared with previous results obtained for SMFS experiments on ProtL$_5$ synthesised using recombinant expression[146].

Single molecule force spectroscopy experiments were performed using a custom built AFM. The spring constant of each silicon nitride cantilever, obtained from Veeco (Santa Barbara, CA), was calculated in buffer by applying the equipartition theorem (see section 2.7.2.1). Proteins were resuspended in 63 mM sodium phosphate buffer at a concentration of 0.5 μg/mL and applied onto a freshly cleaved template-stripped gold surface. The sample was incubated at room temperature for 30 minutes before mechanical unfolding experiments were carried out at a pulling speed of 400 nm/s.

Example traces from these experiments are shown in figure 5.12. The traces show the polyprotein does have mechanical resistance as expected and leading edge of each sawtooth could be fitted with the Worm-Like Chain (WLC) model (§2.7.4.1).

5.3.2.2 Analysis of the SMFS data for poly-ProtL WT

Traces were accepted for analysis only if they contained 2 or more protein unfolding events. The traces were fitted with the WLC model (see section 2.7.4.1 for further details) to confirm the number of ProtL WT unfolding events.

The unfolding forces (F_{UN}) and peak to peak distances (p2p) of ProtL WT for each experiment are recorded in tables 5.1 and 5.2. The average F_{UN} obtained for ProtL WT was 133 ± 2 pN and the p2p distance was 16.4 ± 0.1 pN. The values recorded were consistent with previous studies of ProtL WT (data collected using a pentameric ProtL WT construct) by SMFS by Brockwell et al. at this pulling velocity, of $F_{UN} = 136 \pm 1$ pN and p2p $= 16.5 \pm 0.1$ nm[146].

The next step was to plot the data in a scatter diagram showing the F_{UN} and p2p for
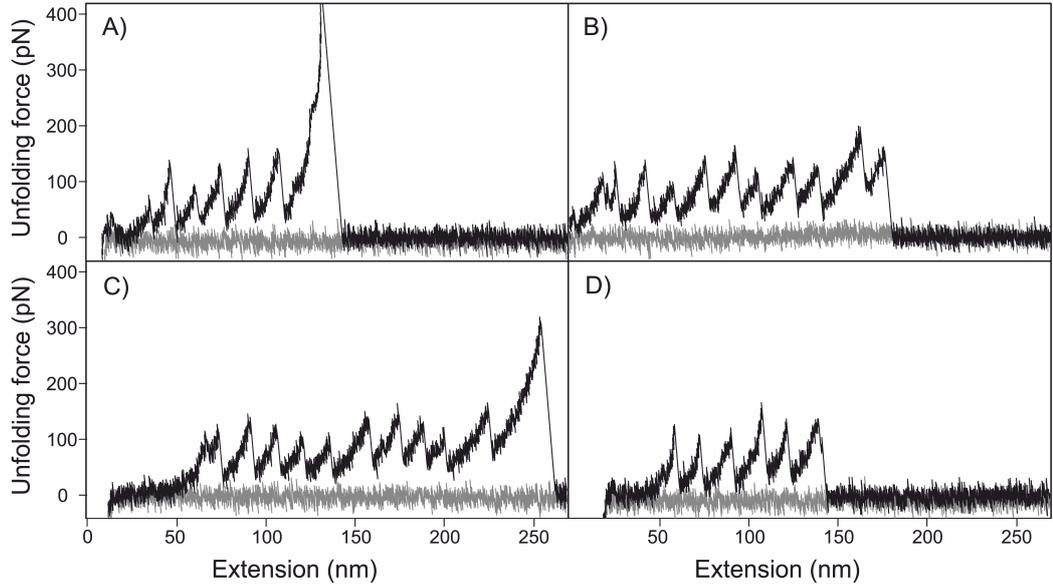
Figure 5.12: Example traces from SMFS experiments using poly-ProtL WT at 400 nm/s in PBS buffer. Figures A), B), C) and D) showing 6, 9, 10 and 5 protein unfolding events respectively. The traces are taken from 3 data sets from experiments carried out on 3 different days.

Protein	No. events	Mean F_{UN} (pN)	Median F_{UN} (pN)	SD
ProtL WT	36	137	135	31
ProtL WT	25	132	131	24
ProtL WT	40	130	132	23
		133	**133**	**2**

Table 5.1: Unfolding force data and standard deviation for poly-ProtL WT at 400 nm/s in PBS buffer for three separate experiments. The numbers quoted in bold are the mean and standard error of the 3 unfolding force values.

Protein	No. events	Mean p2p (nm)	Median p2p	SD
ProtL WT	36	16.2	16.2	0.9
ProtL WT	25	16.6	16.6	1.0
ProtL WT	40	16.3	16.2	1.0
		16.4	**16.4**	**0.1**

Table 5.2: Peak to peak distance data and standard deviation for poly-ProtL WT at 400 nm/s in PBS buffer for three separate experiments. The numbers quoted in bold are the mean and standard error of the 3 peak to peak values.

each experiment, shown in figure 5.13. The diagrams show clear clusters of data points for each experiment centred around the previously published Protein L data[146].

Histograms were then produced for the F_{UN} data and the p2p distances for each experiment, figure 5.14. Each histogram can be fitted with a Gaussian curve.
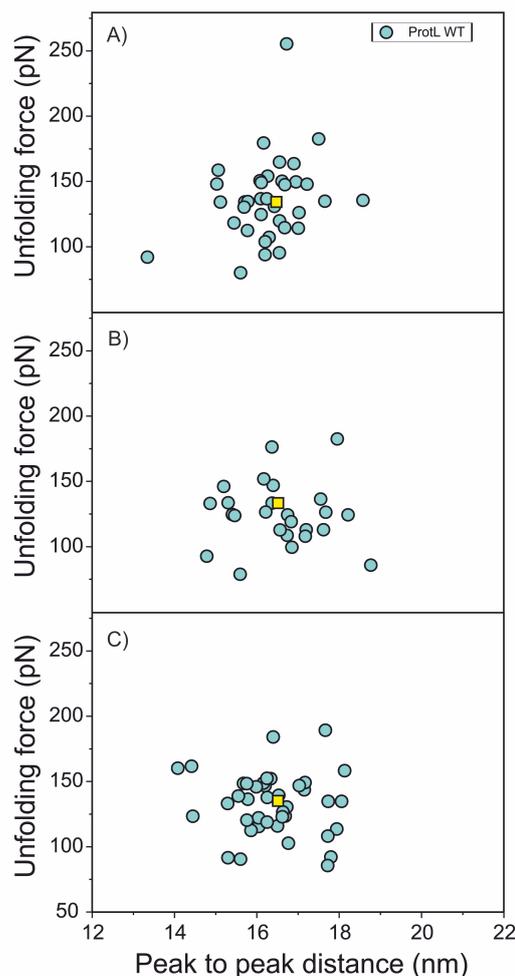
Figure 5.13: Scatter diagrams of the unfolding force versus peak to peak distance for poly-ProtL WT. All figures show ProtL WT in blue. Figures A), B) and C) show the data from 3 data sets from experiments carried out on 3 different days. The polyprotein was in PBS buffer, pulled at a constant velocity of 400 nm/s. The ProtL WT data show a cluster in each of the 3 diagrams. A reference value of unfolding force and peak to peak distance, taken from Brockwell et al. for data accumulated at an identical speed, is shown as a yellow square, error bars are negligable[146].
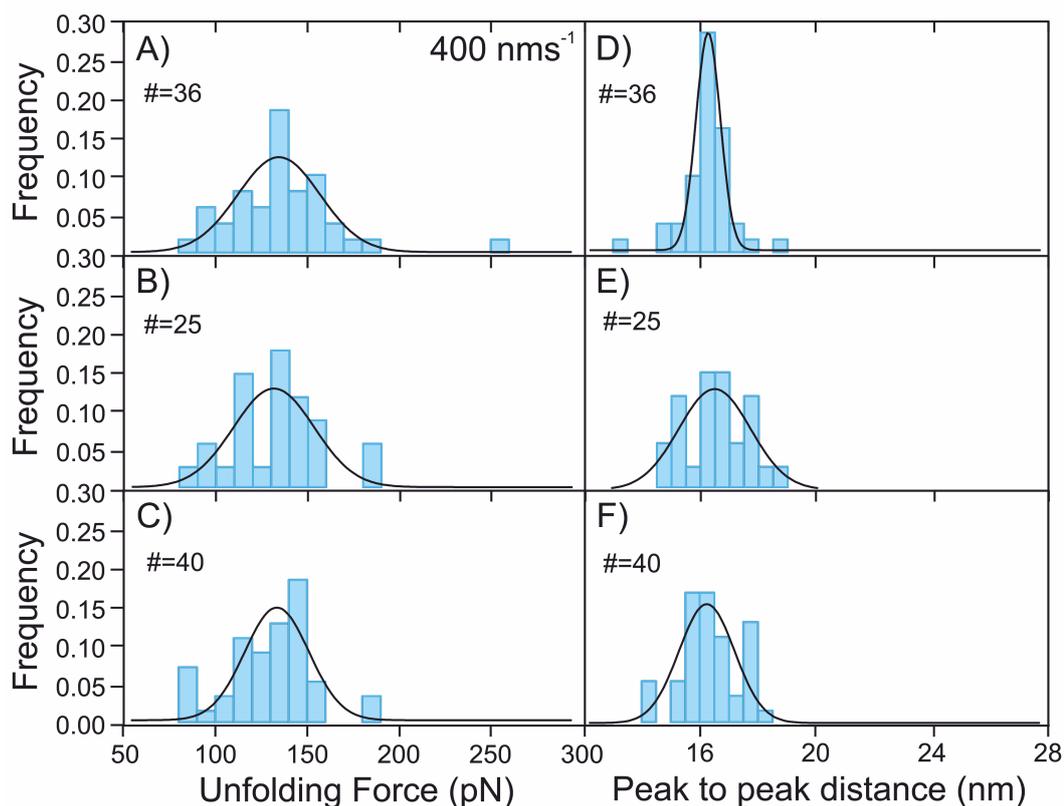
Figure 5.14: Histograms of the unfolding force data and peak to peak distance data for poly-ProtL WT. All figures show ProtL WT in blue. Figures A), B) and C) show the peak to peak distance data and figures D), E) and F) show the unfolding force data. All histograms are fitted with a Gaussian curve. The polyprotein was in PBS buffer, pulled at a constant velocity of 400 nm/s. The number of unfolding events are displayed on each histogram for ProtL WT.

5.3.2.3 Summary

All three experiments carried out had a large number of unfolding events for ProtL WT. If time was taken to optimise the experiments for this sample, i.e. perform with a range of concentrations and with a range of incubation times, it is likely the number of unfolding events in each experiment could be increased. The use of this a homo-polyprotein construct in SMFS experiments was, therefore, shown to be promising as an alternative poly protein construct that could provide large amounts of data for SMFS experiments.

Conclusion: This method of production of the polyprotein of ProtL WT by crosslinking provides a viable alternative to the standard recombinant method.

5.3.3 Production of a homo-polyprotein of the halophilic model protein, Protein L Ex6D

In this section I will describe the first set of experiments I carried out to synthesise a homo polyprotein of ProtL Ex6D, shown in figure 5.15. These experiments were carried out using the same technique as shown in §5.3.1, the schematic for which is shown in figure 5.5.
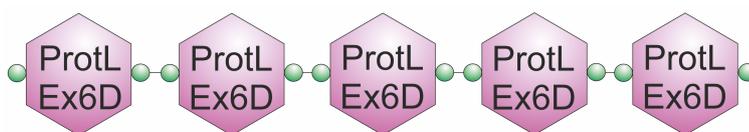
Figure 5.15: Structure of the poly-ProtL Ex6D construct. The construct is polymerised by N- and C-terminal cysteine residues, shown as green circles.

5.3.3.1 Molecular biology for ProtL Ex6D

pEX-A-ProtL Ex6D plasmid was ordered from Eurofins. pET-15b-ProtL WT was provided by Dr. David Brockwell.

To produce a larger, 500 μL, sample of the ordered plasmid, 2 μL was transformed into XL1 blue competent cells (§2.3.1). One colony was picked, the cells were grown in 50 mL LB medium and the DNA was purified using a QIAprep® Spin Midiprep Kit (§2.2.9).

The ProtL Ex6D gene was digested from the pEX-A plasmid using restriction enzymes NcoI and MluI §2.2.4. The excised cassette was separated from the plasmid DNA by agarose gel electrophoresis §2.2.2, and extracted from the gel, as described in §2.2.3. The pET-15b vector was digested using the same restriction enzymes as above and dephosphorylated (§2.2.5). Agarose gel electrophoresis was used to clean and check the size of DNA removed. The gel, figure 5.16, shows the digested sections with controls.
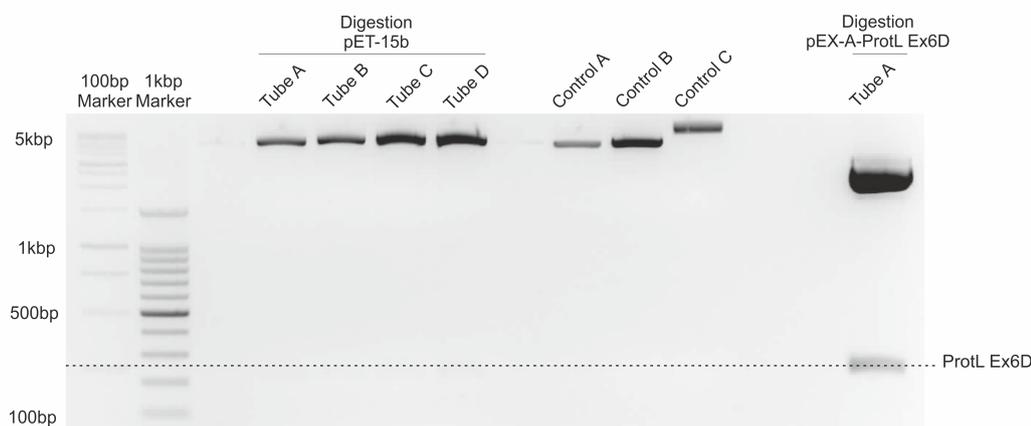
Figure 5.16: The plasmids pET15b and pEX-A-ProtL Ex6D were digested using restriction enzymes NcoI and MluI. Agarose gel electrophoresis was used to clean and check the size of DNA removed. The gel shows a correctly sized product for the removal of ProtL Ex6D from pEX-A - 201 bp.

Next, the ProtL Ex6D insert was ligated into the digested pET-15b plasmid (see §2.2.6). A control ligation was also performed using all components in the reaction mixture, except for the insert DNA. The resulting plasmid, pET-15b-ProtL Ex6D was then transformed into XL1 blue Supercompetent cells, as well as the control reaction, and spread on agar plates. After incubating overnight, colonies were found on the plates with cells that had been transformed with the ligation reaction mixture and none for the negative control ligation mixture. A few colonies were picked, the cells were grown in LB medium and the DNA was purified for each colony using a QIAprep® Spin Miniprep

Kit (§2.2.8). To determine whether the vector now contained the desired insert, $1\mu g$ of plasmid from each colony was digested with the restriction enzymes NcoI and MluI. The results were analysed using agarose gel electrophoresis, shown in figure 5.17. Plasmid DNA containing the insert, positively identified from the analytical digestion was sent off for sequencing.
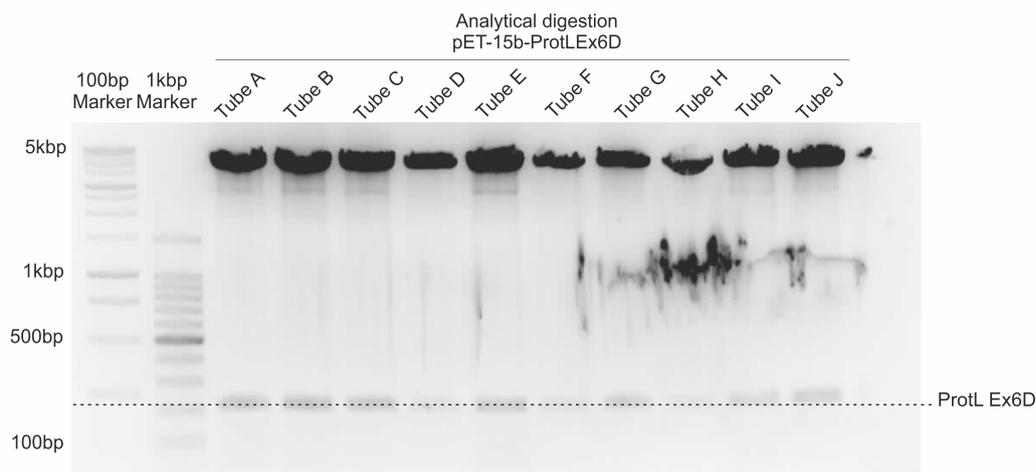
Figure 5.17: The plasmid pET15b-ProtL Ex6D was digested using restriction enzymes NcoI and MluI. The gel shows a successful ligation product with the correctly sized band for ProtL Ex6D - 201 bp in all tubes tested.

5.3.3.2 Expression of ProtL Ex6D

A test expression of the pET15b-ProtL Ex6D construct was carried out after transformation of the plasmid into *E.coli* BL21 expression cells. The trial was performed at a 100 mL scale in LB medium in 250 mL conical flasks. A sample of pET15b-ProtL WT was also expressed as a control. Expression of the constructs was induced by 1 mM IPTG at an OD_{600} of 0.7 and the cells were harvested after 3 hours. The cells were lysed and the resultant protein was purified on Ni-NTA affinity resin §2.3.3. The protein production and purification stages were analysed using SDS-PAGE at each stage, results are shown in figure 5.18. ProtL WT was again shown to run slightly lower than expected. The gel shows a band at 5 kDa in the expression and elution lanes for ProtL WT but there is no evidence for the expression of ProtL Ex6D.
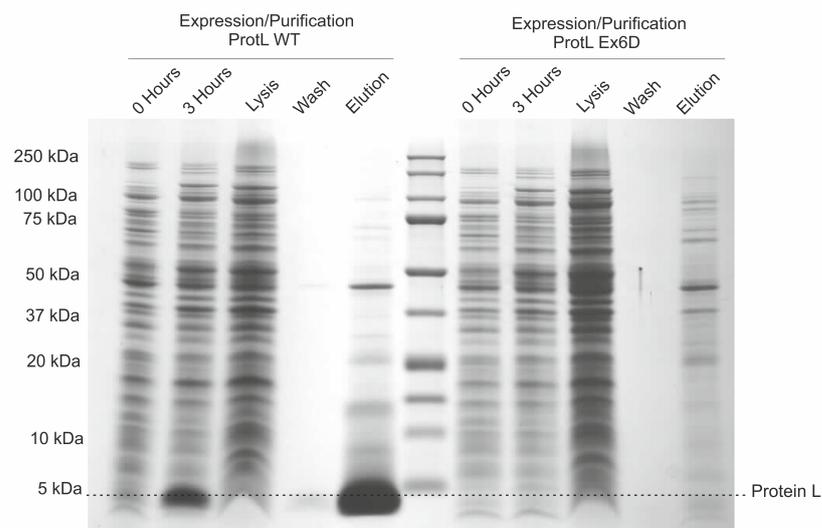
Figure 5.18: Small scale expression and purification of ProtL Ex6D, analysed by SDS-PAGE with Coomasie blue staining. The gel shows a band of significant intensity, corresponding to Protein L, in the expression and purification steps for ProtL WT. No such band, however, is present in the expression and purification steps for ProtL Ex6D.

5.3.3.3   Summary

The results presented show the work carried out to synthesise a homo polyprotein construct of ProtL Ex6D. Construction of the DNA for this construct was successful but the protein did not show any sign of expression.

Conclusion: An alternative technique is needed for of production of Poly-ProtL Ex6D.

5.3.4   Production of a homo polyprotein of ProtL Ex6D using a Maltose Binding Protein fusion

In this section, the results will be presented for the synthesis of a homo polyprotein of ProtL Ex6D by fusing the protein L frame with Maltose binding protein (MBP) (§3.5). An overview of the synthesis method can be seen in figure 5.19. Note that the MBP encoding plasmid (pMAL-c5x) was modified by Dr David Brockwell, before use with Protein L Ex6D, to contain a His tag N-terminal to the MBP protein, in order to aid

purification.
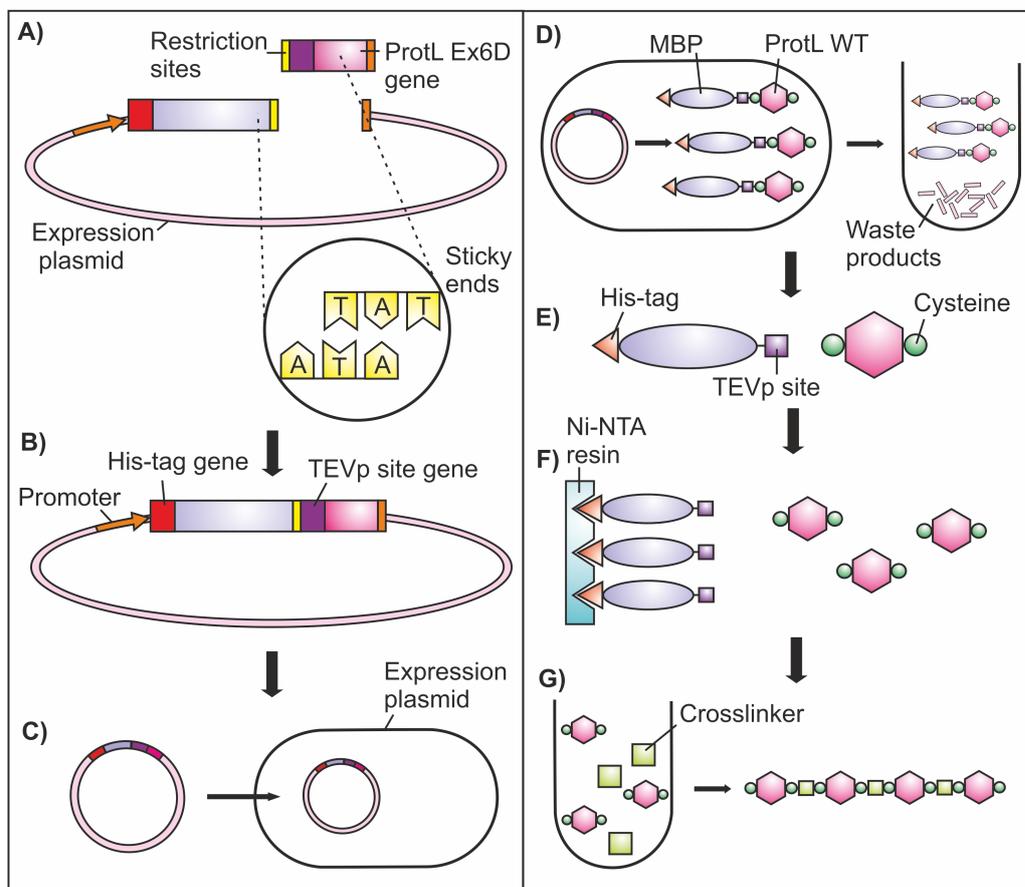
Figure 5.19: Schematic of the synthesis of poly-ProtL Ex6D using MBP. Where A) represents the digestion of the ProtL Ex6D insert and the plasmid to create sticky ends, B) represents the ligation of the ProtL WT gene into the plasmid, C) represents the transformation of the plasmid into an expression *E. Coli* strain, D) represents the expression of MBP-ProtL Ex6D protein, E) represents cleavage of the ProtL Ex6D monomer from MBP, F) represents the purification of the ProtL Ex6D monomer and G) represents the crosslinking of ProtL Ex6D monomer to form a polymer.

5.3.4.1 Molecular biology for MBP-ProtL Ex6D

The modified pMAL-c5x-HmFx plasmid from section §3.5 and the ProtL Ex6D gene from section §5.3.3 were used to carry out these experiments.

Silent mutations were carried out in order to remove BamHI and EcoRI restriction sites from within the ProtL Ex6D sequence. This was done using a Quikchange muta-

genesis kit. The products were transformed into XL1 blue cells and after incubation a colony was picked, grown in 10 mL LB medium and DNA was purified using a QIAprep® Spin Miniprep Kit. The success of the mutation of the DNA was verified by sequencing.

The mutated ProtL Ex6D insert was modified and amplified from the pET15b plasmid by PCR (§2.2.1). Primers were designed to create BamHI and EcoRI restriction sites on the N- and and C-terminal ends of the Prot Ex6D sequence respectively, to create a TEVp recognition sequence and to remove the His-tag. The agarose gel of the PCR products, used to separate the products from the rest of the reaction mixture, is shown in figure 5.20.
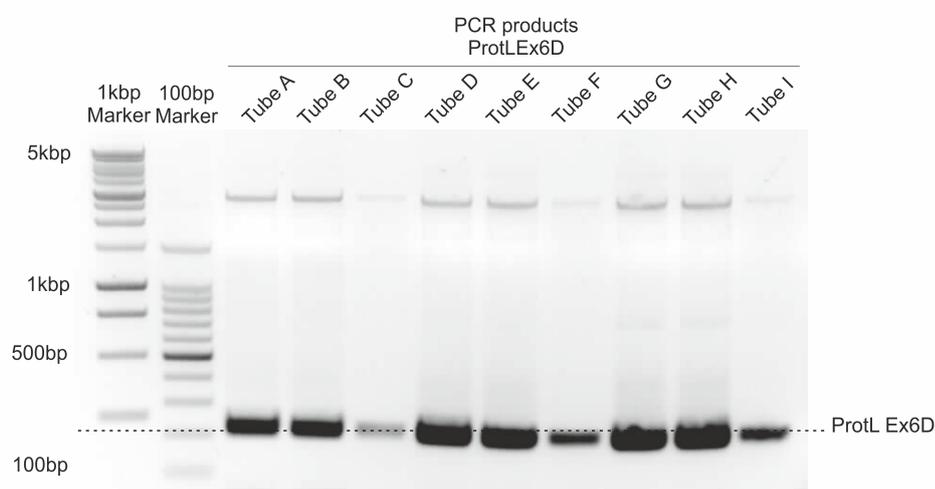
Figure 5.20: Agarose gel electrophoresis was used to clean and check the size of the PCR product for ProtL Ex6D. Short linker regions were added to each end of the ProtL Ex6D insert, encoding the recognised sequence for restriction enzymes BamHI and EcoRI. A band can be seen in all tubes at 231 bp, corresponding to the PCR product of ProtL Ex6D.

To produce a 500 μL sample of the pMAL-c5X plasmid (including His-tag), 2 μL of DNA was transformed into XL1 blue competent cells and grown on agar plates. One colony was picked, the cells were grown in 50 mL LB medium and the DNA was purified using a QIAprep® Spin Midiprep Kit.

The purified PCR product (ProtL Ex6D) was digested using restriction enzymes BamHI and EcoRI. The insert was separated from the the plasmid DNA by agarose

gel electrophoresis and extracted from the gel, shown in figure 5.21. The pMAL-c5x vector was digested using the same restriction enzymes as above and dephosphorylated. Agarose gel electrophoresis was used to clean and check the size of DNA removed. The gel, figure 5.21, shows the digested sections with controls.
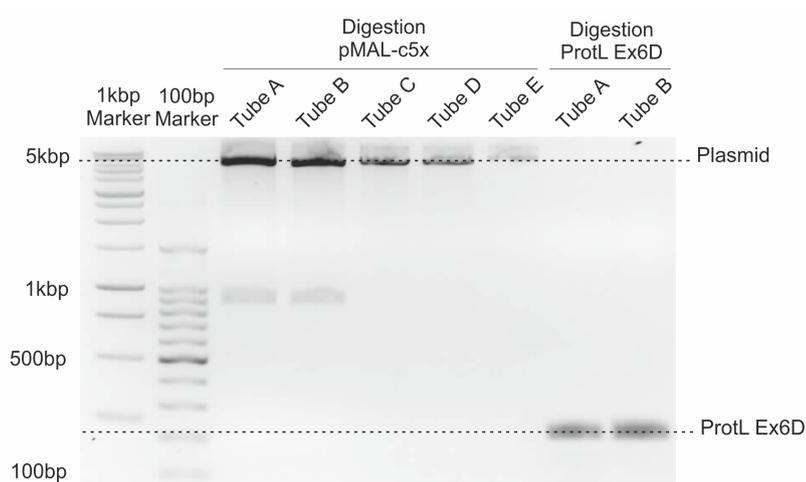
Figure 5.21: The plasmid pMAL-c5x and insert ProtL Ex6D were digested at using restriction enzymes BamHI and EcoRI. Agarose gel electrophoresis was used to clean and check the size of DNA removed. The gel shows a correctly sized product for ProtL Ex6D - 211 bp.

The ProtL Ex6D insert was then ligated into the digested pMAL-c5x plasmid along with a control ligation containing all components in the reaction mixture, except for the insert DNA. The resulting plasmid, pMAL-ProtL Ex6D was then transformed into SURE2 Supercompetent cells, as well as the control reaction, and spread on agar plates. After incubating overnight, colonies were found on the plates with cells with the ligation reaction mixture and none for the negative control ligation mixture. A few colonies were picked, the cells were grown in LB medium and the DNA was purified for each colony using a QIAprep® Spin Miniprep Kit.

An analytical digestion was then carried out using restriction enzymes BamHI and EcoRI with 1μg of plasmid from each colony. The results were analysed using agarose gel electrophoresis, shown in figure 5.22. Plasmid DNA that contained the insert identified from the analytical digestion was sent off for sequencing.
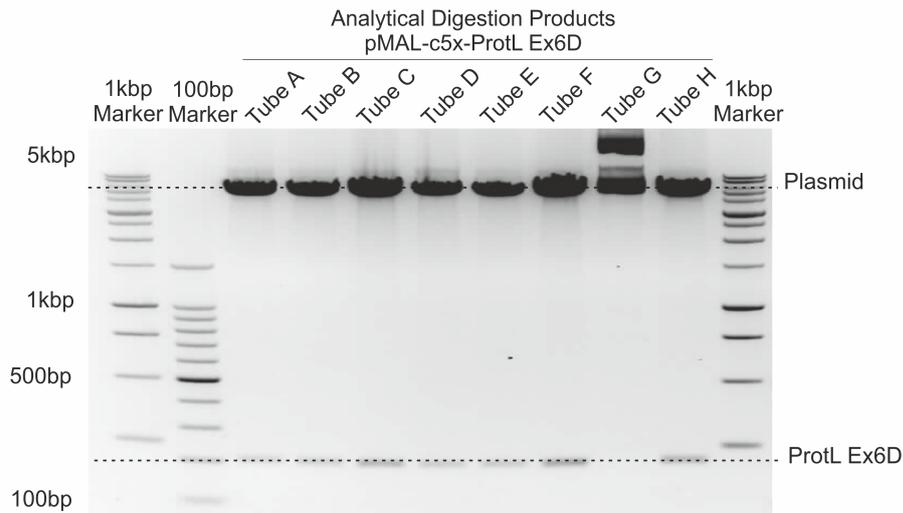
Figure 5.22: The plasmid pMAL-c5x-ProtL Ex6D was digested using restriction enzymes BamHI and EcoRI. The gel shows a successful ligation product with the correctly sized band for ProtL Ex6D - 211 bp in tubes all lanes except G.

5.3.4.2 Expression of MBP-ProtL Ex6D

A test expression of the pMAL-ProtL Ex6D construct was carried out after transformation of the plasmid into *E.coli* JM83 expression cells. The trial was performed at a 100 mL scale in LB medium in 250 mL conical flasks. Expression of the construct was induced by IPTG at an OD_{600} of 0.7 and the cell growth and protein expression was monitered by taking samples at 0, 1, 2, 3, 4 hours after expression and overnight. The resultant protein was purified on Ni-NTA affinity resin. The protein production and purification stages were analysed using SDS-PAGE, shown in figure 5.23. The gel shows a band corresponding to the expected size of MBP-ProtL Ex6D (50.9 kDa) in the expression lanes (except for 0 hours as expected) and elution lane.

A large scale, 10 L growth of MBP-ProtL Ex6D was then set up using 10 x 1 L LB medium in 2.5 L conical flasks. Protein expression was induced by IPTG at an OD_{600} of 0.7. Cultures were allowed to grow for a further 4 hours before harvesting by centrifugation.

The cell pellet was resuspended in lysis buffer and the cells were lysed using a cell
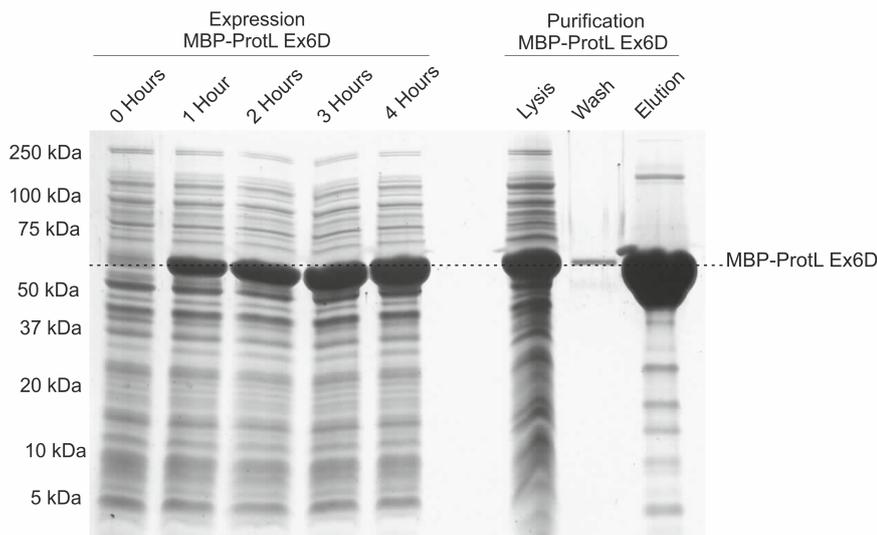
Figure 5.23: A small scale expression and purification of MBP-ProtL Ex6D, analysed by SDS-PAGE with Coomasie blue staining. The gel shows a band of significant intensity at 50.9 kDa which would correspond to MBP-ProtL Ex6D in the expression and purification steps.

disrupter system. MBP-ProtL Ex6D was then purified from the soluble fraction by Ni affinity chromatography, using a Ni Sepharose column. The soluble lysate was passed through the column so the His-tagged MBP-ProtL Ex6D would bind. After washing away non-specifically bound protein, the protein was eluted by competition with imidazole in the elution buffer. SDS-PAGE analysis of the fractions is shown in figure 5.24. Fractions shown to contain MBP-ProtL Ex6D were pooled, dialysed into H_2O and lyophilised.

5.3.4.3   Cleaving ProtL Ex6D from MBP

Two test cleavage reactions were set up; 2 mg of MBP-ProtL Ex6D was resuspended in buffer containing 50 mM Tris.HCl and 2 mM DTT and another 2 mg of MBP-ProtL Ex6D was resuspended in buffer containing 50 mM Tris.HCl, 0.5 M NaCl and 2 mM DTT. Commercial AcTEVp was added to the two separate reactions with final molar ratios of MBP-HmFd to TEV protease of 1:3. Both reactions were incubated at room temperature and samples of each reaction were taken at separate time-points of 0, 1, 2, 4 and 6 hrs to find optimal conditions. The samples were analysed by SDS-PAGE, and this
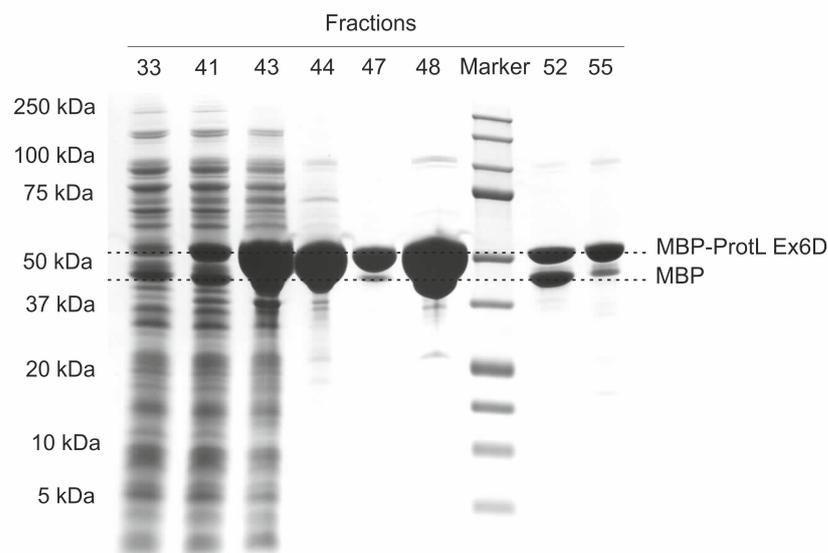
Figure 5.24: Large scale purification of MBP-ProtL Ex6D using Ni-NTA chromatography, analysed by SDS-PAGE with Coomasie blue staining. The gel shows a band with significant intensity at 50.9 kDa which would correspond to ProtL Ex6D in fractions 43-48, these fractions were pooled.

showed a band decreasing in intensity over time at the expected weight for MBP-ProtL Ex6D (50.9 kDa) and a band increasing in intensity over time at the expected weight for MBP (43.8 kDa) in figure 5.25. This implies MBP is successfully being cleaved from the the MBP-HmFd construct by TEV protease. There is also a band, increasing in intensity, at the expected weight for ProtL Ex6D (7.2 kDa), showing ProtL Ex6D was successfully cleaved from MBP. It should be noted that the weight of ProtL Ex6D is less here since the protein no longer contains a His-tag.

A large scale cleavage experiment was then set up using 10 mg MBP-ProtL Ex6D resuspended in the buffer containing 0.5 M NaCl. Expressed and purified TEVp (§2.4.2) was then added to the reaction with a final molar ratio of MBP-HmFd to TEV protease of 1:3. The reaction was incubated at room temperature and samples were taken at 15 mins, 30 mins, 1 hour, and 3 hours. The samples were analysed by SDS-PAGE, figure 5.26, showing a band corresponding to MBP-ProtL Ex6D (50.9 kDa) only at 0 hours, a band at the expected weight for MBP (43.8 kDa), a band at the expected weight for TEVp (27 kDa) and a band, increasing in intensity, at the expected weight for ProtL
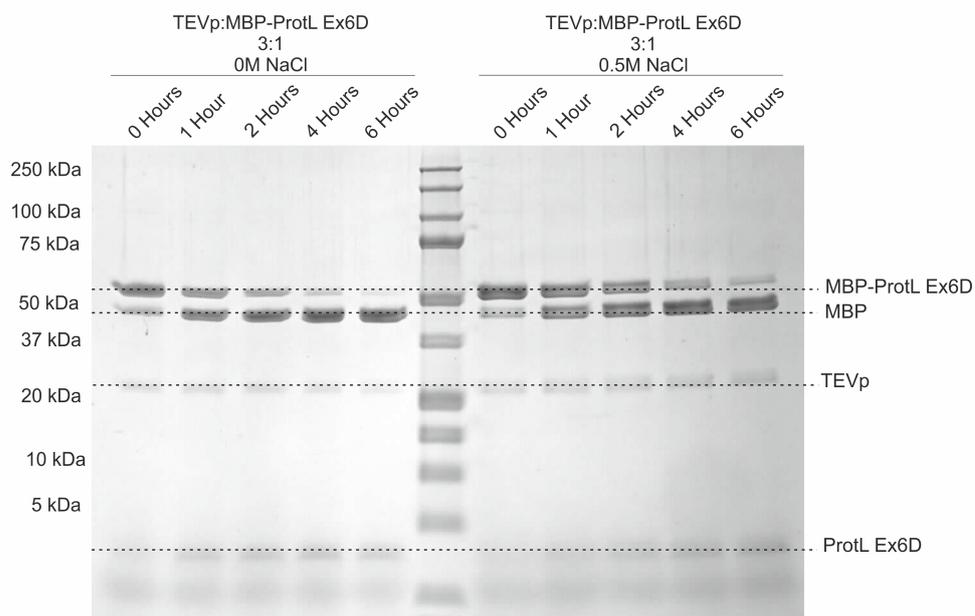
Figure 5.25: Trial cleavage of ProtL Ex6D from MBP using TEV protease at two different salt concentrations, analysed by SDS-PAGE with Coomasie blue staining. The gel shows the MBP-ProtL Ex6D band at 50.9 kDa reducing in intensity from 0 hours to 6 hours and the MBP band at 43.8 kDa increasing from 0 to 6 hours. There is a band at 7.2 kDa that increases in intensity over time, which corresponds to the production of ProtL Ex6D.

Ex6D (7.2 kDa), showing ProtL Ex6D was successfully cleaved from MBP using the purified TEVp.
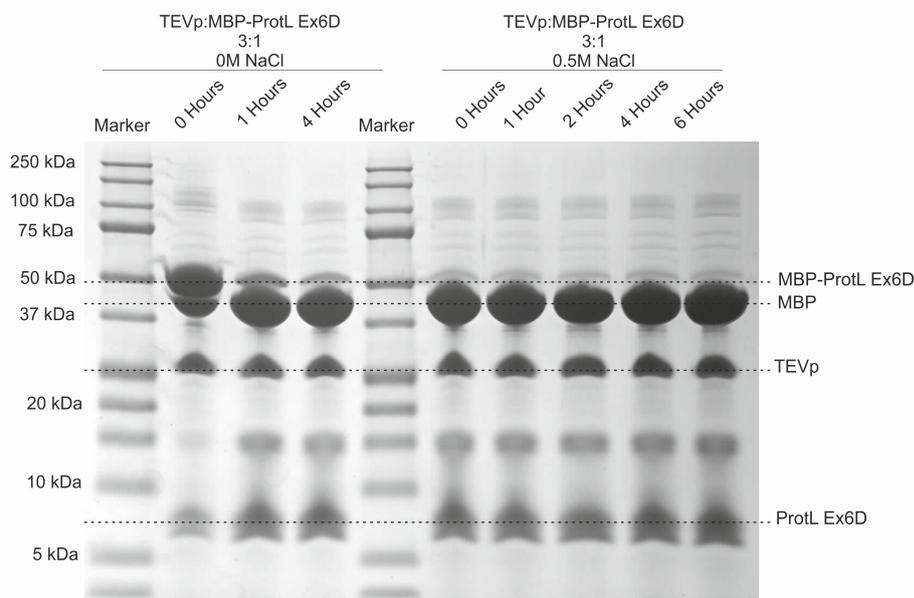
Figure 5.26: Trial cleavage of ProtL Ex6D from MBP using a purified sample of TEV protease at a ratio of protease to protein 1:3, analysed by SDS-PAGE with Coomasie blue staining. The gel shows the MBP-ProtL Ex6D band at 50.9 kDa at 0 hours and the MBP band at 43.8 kDa increasing in intensity over time. There is a band with significant intensity at 7.2 kDa increasing in size, which corresponds to the production of ProtL Ex6D.

ProtL Ex6D was then purified using using Ni-NTA affinity resin. Both MBP and TEVp contained His-tags, therefore would bind to the resin, leaving only ProtL Ex6D in the flow through. Analysis by SDS-PAGE (Figure 5.27) shows the flow through that contains ProtL Ex6D (7.2 kDa) with a small contamination of MBP (43.8 kDa) and TEVp (27 kDa) that have not bound to the resin. The elution clearly shows only MBP and TEVp.

ProtL Ex6D was then further purified by size exclusion chromatography using a Superdex 75 column to remove the remaining MBP and TEVp. SDS-PAGE analysis showed the pure ProtL Ex6D, figure 5.28. Fractions 64-75 were pooled, dialsed into H_2O and lyophilised.
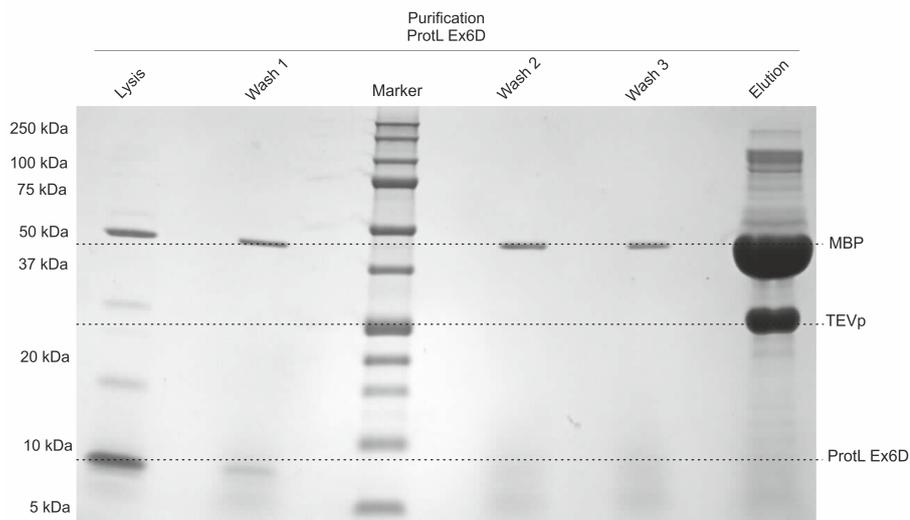
Figure 5.27: Trial cleavage of ProtL Ex6D from MBP purified using Ni-NTA chromatography, analysed by SDS-PAGE with Coomasie blue staining. The gel shows a band at 43.8 kDa corresponding to MBP in the flow through, wash steps and in the elution. There is a band at 7.2 kDa corresponding to ProtL Ex6D in the flow through of the column.

5.3.4.4 Crosslinking of ProtL Ex6D

A 0.7 mg of ProtL Ex6D was resuspended in buffer containing 50 mM Tris, pH 7.4 to a concentration of 2 mg/ml. 1,11-bis(maleimido)triethylene glycol (BM(PEG)$_3$) was added to a final molar ratio of ProtL Ex6D to BM(PEG)$_3$ of 1:1. The reaction was incubated at room temperature. After 3 hours the resultant product was purified by size exclusion chromatography using a Superdex 75 column to remove smaller unpolymerised monomeric or low molecular weight polymeric ProtL Ex6D. SDS-PAGE analysis showed the separation of the resultant poly-ProtL Ex6D by size (figure 5.29). The elution profile is shown in figure D.3. A wide range of polymeric species were formed, bands corresponding to ProtL Ex6D monomer (7.2 kDa), dimer (15.1 kDa) and trimer (22.4 kDa), up to at least an octomer are visible. The molecular weight of each polymer can be calculated, using equation 5.1. The before sample also shows signs of crosslinking, which is likely due to disulphide bonding that has not been reduced by the DTT in the loading buffer. Fractions 15-19 were pooled, dialsed into H$_2$O and lyophilised in 0.1 mg aliquots.

208
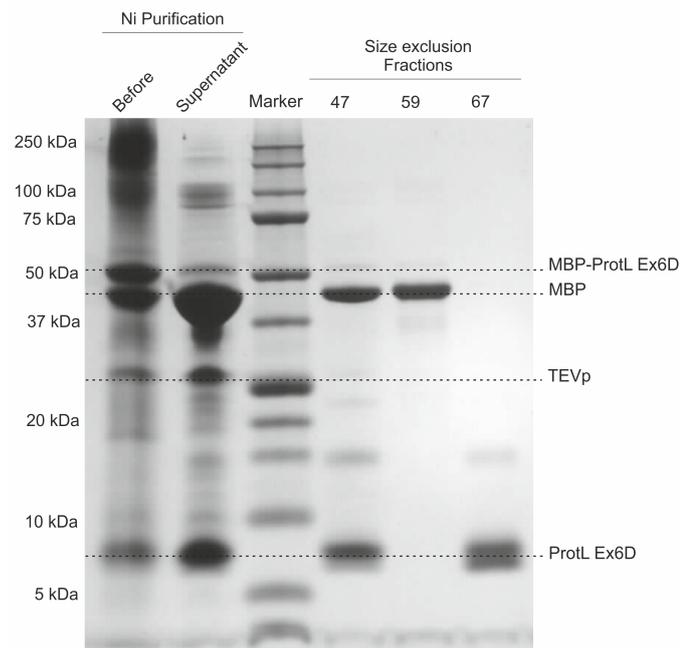
Figure 5.28: Trial cleavage of ProtL Ex6D from MBP purified using Ni-NTA chromatography and size exclusion chromatography, analysed by SDS-PAGE with Coomasie blue staining. The gel shows a band at 43.8 kDa corresponding to MBP in the supernantant and fractions 47-59. There is a band at 7.2 kDa corresponding to ProtL Ex6D in the supernatant and fraction 67.
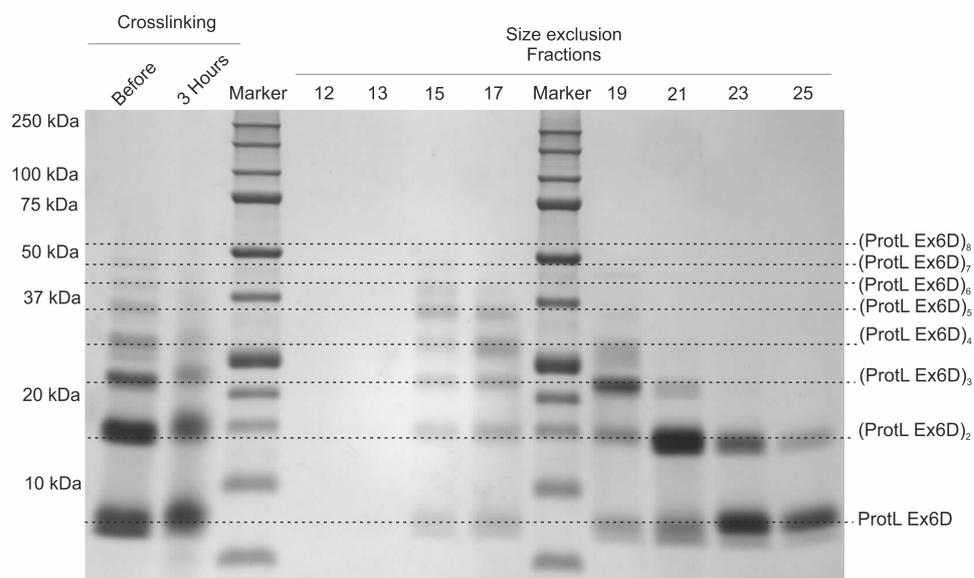
Figure 5.29: The crosslinking of ProtL Ex6D using BM(PEG)$_3$ and purification using size exclusion chromatography, analysed by SDS-PAGE with Coomasie blue staining. The gel shows a band of significant intensity at 7.2 kDa which would correspond to the monomer ProtL Ex6D, 15.1 kDa - (ProtL Ex6D)$_2$, 22.4 kDa - (ProtL Ex6D)$_3$, 29.6 kDa - (ProtL Ex6D)$_4$, 36.9 kDa - (ProtL Ex6D)$_5$, 44.1 kDa - (ProtL Ex6D)$_6$, 51.4 kDa - (ProtL Ex6D)$_7$ and 58.6 kDa - (ProtL Ex6D)$_8$.

5.3.4.5 Summary

The results presented show the work carried out to synthesise a homo polyprotein construct of ProtL Ex6D using the fusion protein, Maltose binding protein. Construction of the DNA for this construct was successful and the protein was successfully expressed, cleaved and crosslinked to form a homo polyprotein.

Conclusion: This method of production of the polyprotein of ProtL Ex6D was successful.

5.3.5 Single Molecule Force Spectroscopy on poly-ProtL Ex6D

In this section I will describe the SMFS experiments carried out using poly-ProtL Ex6D and the analysis of the results. I will finish by discussing the usefulness of the construct for experimentation using SMFS.

5.3.5.1 Procedure for SMFS experiments using poly-ProtL Ex6D

The next step was to see how useful poly-ProtL Ex6D was for SMFS experiments. 0.1 mg of poly-ProtL Ex6D was resuspended in 0.6 mL buffer (0.5 M KCl, 5 mM MOPS, pH 7.4 buffer). The solution was incubated on a freshly cleaved gold surface for 30 minutes. A mechanical unfolding experiment was carried out at a pulling speed of 1200 nm/s. Figure 5.30 shows example traces from the experiment.

5.3.5.2 Analysis of the SMFS data for poly-ProtL Ex6D

Firstly, only traces that contained at least 2 protein unfolding events were accepted. The traces were fitted with the WLC model (see section 2.7.4.1 for further details) to confirm that bona fide ProtL Ex6D unfolding events were being observed. For each ProtL Ex6D unfolding event, the F_{UN} and p2p were recorded (see tables 5.3 and 5.4 respectively).
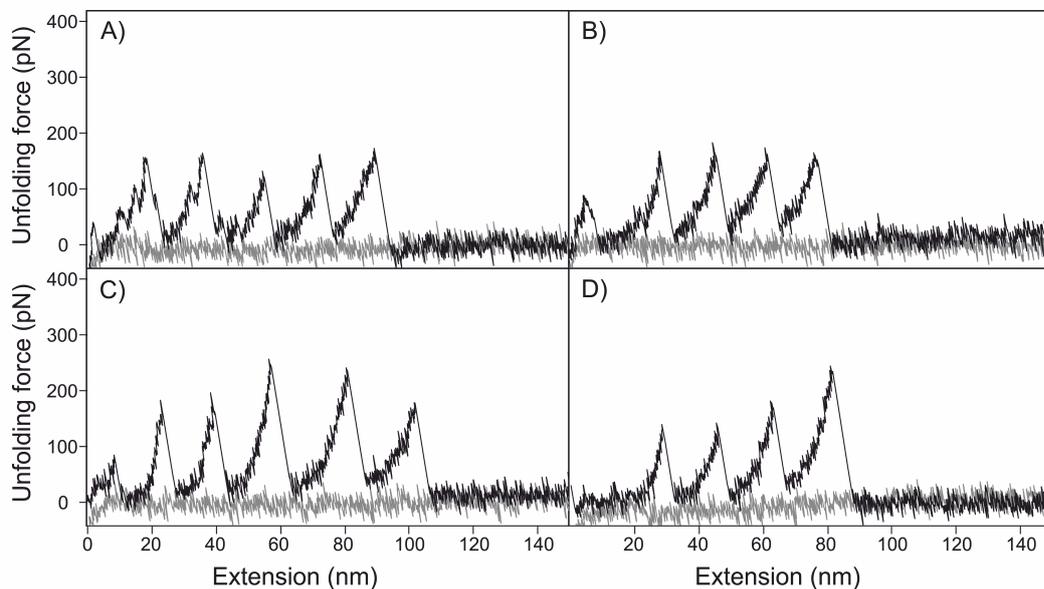
Figure 5.30: Example traces from the SMFS experiment using poly-ProtL Ex6D at 1200 nm/s in 0.5 M KCl buffer. Figures A), B), C) and D) showing 4, 3, 5 and 3 protein unfolding events respectively. The polyprotein was in 0.5 M KCl buffer, pulled at a constant velocity of 1200 nm/s.

Protein	No. events	Mean F_{UN} (pN)	Median F_{UN} (pN)	SD
ProtL Ex6D	38	158	160	23

Table 5.3: Unfolding force data and standard deviation for poly-ProtL Ex6D at 1200 nm/s in 0.5 M KCl buffer.

Protein	No. events	Mean p2p (nm)	Median p2p (nm)	SD
ProtL Ex6D	38	16.8	16.7	0.8

Table 5.4: Peak to peak distance data and standard deviation for poly-ProtL Ex6D at 1200 nm/s in 0.5 M KCl buffer.

Next, the values were plotted in a scatter diagram showing the F_{UN} for each protein unfolding event against the measured p2p distance (figure 5.31). The plot shows an average p2p for ProtL Ex6D of 16.8 ± 0.1 nm and an average F_{UN} of 158 ± 4 pN. The p2p distance is slightly higher than the value recorded for ProtL WT in the poly-ProtL WT construct in PBS buffer (16.4 ± 0.1 nm).
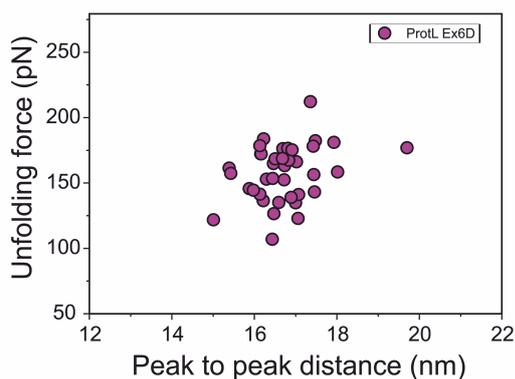
Figure 5.31: Scatter diagram of the unfolding force vs. peak to peak distance for poly-ProtL Ex6D in 0.5M KCl at 1200 nms[1]. The figure shows ProtL Ex6D in pink. The ProtL Ex6D points show a cluster in the diagram.

The clear cluster in the scatter plot, figure 5.31 validates the use of this method of synthesising the poly-ProtL Ex6D construct for SMFS experiments. The histograms for the F_{UN} and p2p data, figure 5.32 show a Gaussian distribution.
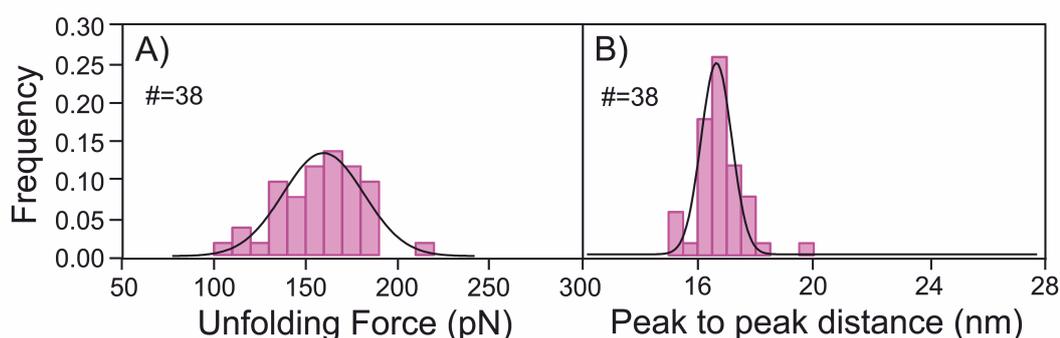
Figure 5.32: Histogram of the unfolding forces for poly-ProtL Ex6D. The figure shows ProtL Ex6D in pink. Figure A) shows the peak to peak distance data and figure B) shows the unfolding force data. Both histograms are fitted with a Gaussian curve. The number of unfolding events are displayed on each histogram for ProtL Ex6D.

5.3.5.3 Summary

The above results show the data collected and analysed from one SMFS experiment on poly-ProtL Ex6D at 1200 nm/s in 0.5 M KCl. For this data set, there were 38 unfolding events. This is larger than most data sets (the number of unfolding events in one day) recorded for ProtL Kx6E in chapter 4. As above, if time was taken to optimise the experiments for this sample it is likely the number of unfolding events in each experiment could be increased. Further discussion of the results can be sound in section 6.2. The use of this a homo-poly protein construct in SMFS experiments was, therefore, shown to be promising as an alternative poly protein construct for a halophilic model protein that could provide large amounts of data for SMFS experiments.

Conclusion: The polyprotein of ProtL Ex6D, formed by crosslinking, provides an alternative construct to use in SMFS experiments, which could provide more data on a shorter timescale.

5.4   Chapter summary

This chapter presented the work that was carried out on to synthesise two homo-polyprotein constructs:

i) poly-ProtL WT
ii) poly ProtL Ex6D

Production of both constructs was shown to be successful.

The experimental results for SMFS experiments for both homo-polyprotein constructs was presented. The results showed that the method of production of the polyproteins by crosslinking provides a viable alternative to the standard recombinant method, which could provide more data on a shorter timescale. Further discussion of the SMFS results for poly-ProtL Ex6D can be found in section 6.2.

Chapter 6

Discussion and future work

The aim of this project was to use protein engineering to produce polyprotein constructs containing a halophilic protein and to carry out Single Molecule Force Spectroscopy (SMFS) experiments using the Atomic Force Microscope (AFM) to characterise the mechanical unfolding behaviour of a halophilic protein.

This chapter presents a discussion of all of the results presented in this thesis. First, the data collected for SMFS experiments with the hetero-polyprotein construct I27$_3$-HmFd-I27 is discussed. Possible unfolding pathways for HmFd are suggested and compared to the data. This is followed with a discussion about the attempts to express and purify further poly-protein constructs containing HmFd and hypotheses as to why the expression may not have been successful. The data collected for SMFS experiments with three hetero-polyprotein constructs, (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 and poly-ProtL Ex6D, in 0.5 M KCl is discussed. Reasons for differences between the measured unfolding forces for the three constructs are explained. Next, the data collected for SMFS experiments with two hetero-polyprotein constructs, in 0.5 M is compared to data collected in 2 M KCl. An explanation for the differences found between the two constructs is presented. Finally, possible future directions of this project are outlined.

6.1 The mechanical fingerprint of *Hm*Fd

Chapter 3 presents the work that was carried out on a hetero-polyprotein construct, I27$_3$-*Hm*Fd-I27 using SMFS. Although force-extension data was successfully collected, the unambiguous characterisation of the unfolding force, F$_{UN}$, and peak to peak distance, p2p, was consequently unsuccessful for *Hm*Fd.

The histograms plotted for *Hm*Fd, figure 3.5, reveal it is possible to see a slight trend in the measured unfolding force, i.e. a gaussian can be fitted to the data. It is, therefore, likely that *Hm*Fd does have mechanical resistance but the unfolding behaviour could not be characterised using this analysis technique due to insufficient data.

Refinement of the data analysis was carried out by only accepting traces where the full construct had unfolded. The p2p values of the *Hm*Fd candidate unfolding events are shown in table 6.1. Examples of the corresponding traces are shown in figures 3.6 and 3.7.

Salt conc.	Ionic strength	Velocity	*Hm*Fd event number	p2p distance (nm)
63 mM	0.17 M	400 nm/s	1	15.6
63 mM	0.17 M	400 nm/s	2	6.5
63 mM	0.17 M	400 nm/s	1	18.8
500 mM	0.66 M	400 nm/s	1	14.9
500 mM	0.66 M	400 nm/s	1	6.0
2000 mM	2.15 M	400 nm/s	1	17.8
2000 mM	2.15 M	400 nm/s	2	12.7
2000 mM	2.15 M	400 nm/s	1	14.8
2000 mM	2.15 M	400 nm/s	2	21.5

Table 6.1: Peak to peak distance data for HmFd in I27$_3$-*Hm*Fd-I27 for only full length traces. The event number corresponds the order of *Hm*Fd unfolding events where traces have more than one. Example traces for these unfolding events are shown in figures 3.6 and 3.7.

Given the measured p2p values, it is interesting to consider the possible unfolding pathway of *Hm*Fd. To do this, the structure of *Hm*Fd was simplified by only drawing the β–strands (figure 6.1 A). It is assumed that the β–sheet structure of a protein that is most important for mechanical stability (see section 1.4.4.1). The number of amino

acids and, therefore, the total extension on the protein was calculated for each mechanical unfolding scenario illustrated in figure 6.1.
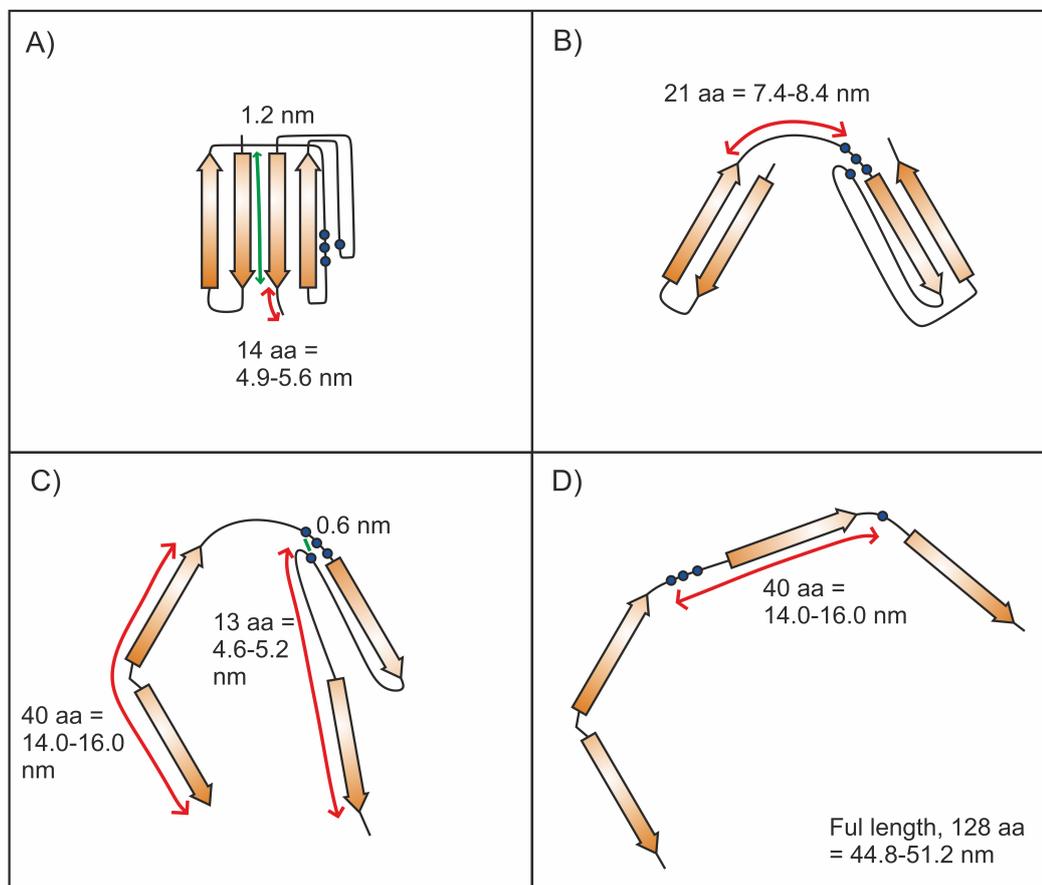
Figure 6.1: Estimating the possible unfolding pathway of *Hm*Fd. Figures A)–D) show all possible regions that may have mechanical resistance and the extension of the chain upon unfolding each region. Each amino acid is estimated to have a length of 0.35-0.4 nm. Figure A) shows the length of chain that will readily unfold, which is comprised of 14 amino acids. Figure B) shows the shearing of β–strands 1 and 4. Figure C) shows the peeling of β-strands 1 and 2 and the peeling of β–stands 3 and 4. Figure D) shows the rupture of the Fe–S cluster. Red arrows represent the number of amino acids exposed upon unfolding, green arrows represent initial separation before unfolding. In this schematic the α-helices are not shown because they are assumed to not show mechanical resistance.

It is clear from figure 6.1 that *Hm*Fd is likely to have a complicated unfolding pathway that may result in more than one mechanical clamp motif. Figure 6.1 A) shows the length of protein that will readily unfold, which is composed of 14 amino acids making up an α-helix, this is expected to unravel as soon as force is applied to the protein and will not

have any mechanical resistance (see section 1.4.4). Figure 6.1 B) shows the separation of β-strands 1 and 4. As the force is acting on opposite ends on the β-strands, the hydrogen bonds between the β-strands will need to be ruptured simultaneously. Observation of the mechanical resistance of this separation event is highly likely (i.e. it is likely to be greater than the thermal noise, ~ 15 pN). Figure C) shows the separation of β-strands 1 and 2 and the separation of β-stands 3 and 4. For both of these events, the force is acting on the same ends of the β–strands, the hydrogen bonds can be broken one by one. Mechanical resistance of these two separation events is likely to be very low (see section 1.4.4.2). Figure D) shows the rupture of the Fe-S cluster.

A plausible unfolding pathway for HmFd might be:

i) The 14 amino acids (4.9-5.6 nm) at the C-terminus forming α-helix 5 will unravel as soon as force is applied to the protein with no mechanical resistance.

ii) The force is quickly applied to the β–strands 1 and 4. The hydrogen bonds of β–strands 1 and 4 are sheared, creating an unfolding force. β–strands 1 and 2 and β–stands 3 and 4 readily unfold. The extension for this event is 74 amino acids (26.3–29.6 nm), the separation before extension is 1.2 nm, therefore the extension upon shearing of β–strands 1 and 4 is estimated as 25.2–28.4 nm.

iii) The force is then applied to the Fe-S cluster. A further extension of 40 amino acids (14.0-16.0 nm) after rupture of the Fe-S bond (see section 3.1.1.2). The separation before extension is 0.6 nm, therefore, the extension upon rupture of the Fe–S bond is estimated as 13.4–15.4 nm.

If the Fe-S cluster is not present in the protein, the final 40 amino acids will extend during step ii) giving an estimated extension 38.7–44.4 nm.

The values of the extension measured for the SFMS data, shown in table 6.1, do not show a significant correlation with the estimated extension values in the steps above. A few values agree with the extension estimated for the rupture of the Fe-S cluster (13.4–15.4 nm) but it is not possible to correctly identify the unfolding pathway based on the limited SMFS data collected.

It is possible that HmFd was expressed without the Fe-S cluster. Marg et al. recorded expression of the apo-protein of a mutant variant of ferredoxin from *Halobacterium salinarum* (HsFd) when expressed in *E. coli*[101]. If the Fe-S cluster is not present in the protein, it is possible that the cysteine residues, no longer bound, are free to interact with the gold surface. If the cysteines do bind to the gold surface, it would then be very unlikely to see unfolding traces with the full construct unfolding. The polyprotein construct is bound to the gold surface at the C-terminal end by two cysteine residues (see figure 3.2). If HmFd is bound to the gold surface, it is likely only the 3 I27 domains and the first 75 amino acids of HmFd will unfold. Only a small number of full unfolding traces were recorded (i.e. containing 4 unfolding events).

It is therefore clear that the mechanical fingerprint cannot be characterised using the SMFS data presented here. More copies of HmFd are needed in the polyprotein construct in order to generate a larger amount of data to carry out effective mechanical unfolding characterisation. Chapter 3 also described two techniques used in an attempt to synthesise polyprotein constructs containing multiple copies of HmFd. However the production of a polyprotein containing HmFd was shown to be unsuccessful using the techniques presented.

It has been reported that the excess of negatively charged amino acids, typically found in halophilic proteins, can create problems for expression in a host system that is a different species, such as *E. coli*. Halophilic proteins typically misfold and aggregate in conditions of low ionic strength, so the purification of halophilic enzymes from *E. coli* has been shown to involve the recovery of insoluble protein from inclusion bodies[207]. Further to this, archaeal proteins that are expressed in bacterial hosts do not include posttranslational modifications, such as acetylation or ubiquitination[208]. HmFd has been found to be acetylated in vivo, the amino acid sequences contains a unique acetylated lysine close to its C-termini, therefore causing further complications for the expression of HmFd in *E. coli*[209].

6.2 Modifying the mechanical stability of protein L using specific mutations

Chapters 4 and 5 presents the work that was carried out on mesophilic and mutated variants of Protein L. Figure 6.2 shows the differences in the amino acid structures for the mutations carried out on Protein L.
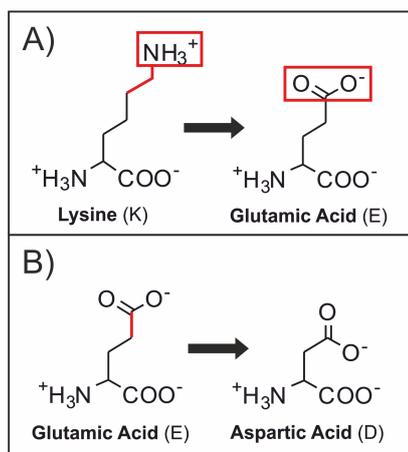
Figure 6.2: The amino acid structures of the two mutations carried out on Protein L. The figure shows the difference in the amino acid R-groups for A) lysine to glutamic acid (Kx6E) and B) glutamic acid to aspartic acid (Ex6D). The changed portions of the amino acid are shown in red.

ProtL Kx6E has six lysine residues mutated to six glutamic acid residues. Here the length and the charge of the R-group varies, shown in figure 6.2 A). ProtL Ex6D has six glutamic acid residues mutated to six aspartic acid residues. Here only the length of the R-group varies, shown in figure 6.2 B).

The importance of the mechanical clamp region on the mechanical stability of a protein has been outlined in section 1.4.4. The mechanical clamp for ProtL has been shown to be the N- and C-terminal β-stands (I and IV). After the hydrogen bonds between these strands are sheared, the protein unfolds in a single step. This process is dicussed further in section 4.1.2. The N- and C-terminal β-stands (I and IV) of ProtL WT are shown in figure 6.3 A), along with the intervening hydrogen bonds. Both of the

ProtL variants (Kx6E and Ex6D) have one of their six substitutions occuring within the mechanical clamp region, the exact position of the mutation is shown in figure 6.3 B) for ProtL Kx6E and figure 6.3 C) for ProtL Ex6D.
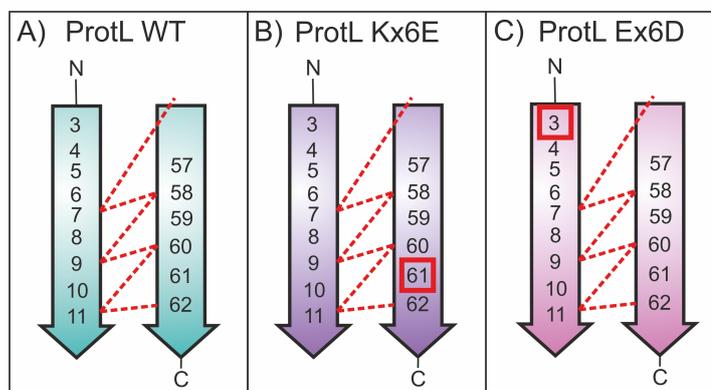
Figure 6.3: The mechanical clamp region of Protein L for WT, Kx6E and Ex6D. The figures show the two β-stands (I and IV) that form the mechanical clamp for Protein L. The numbers of the corresponding residues have been included and interstrand hydrogen bonds, calculated using DSSP, are shown as red dashed lines [146,200]. Figure A) shows the mechanical clamp of ProtL WT, figure B) shows the position of the mutated amino acid of ProtL Kx6E within the mechanical clamp region and C) shows the position of the mutated amino acid of Ex6D within the mechanical clamp region. All other mutations occur outside of the mechanical clamp region for both ProtL Kx6E and ProtL Ex6D.

The mutation in the mechanical clamp region of ProtL Kx6E is located adjacent to one of the hydrogen bonding residues. It is, therefore, highly likely that the change in length and the charge of the R-group will have an effect on the strength of the hydrogen bonds. Figure 4.28 A) and C) highlights the difference in the measured unfolding force between ProtL WT and ProtL Kx6E from the SFMS analysis. This difference is consistent at all pulling velocities. The reduction of the unfolding force from ProtL WT to ProtL Kx6E protein is likely due to disruption of the contacts, which could be hydrogen bonds, hydrophobic contacts etc., involved in the mechanical clamp.

The mutation in the mechanical clamp region of ProtL Ex6D is located at top of β–strand I, a few residues away from hydrogen bonding residues. It is, therefore, unlikely that the change in length the R-group will have an effect on the strength of the hydrogen bonds. Only one SMFS experiment was carried out for poly-ProtL Ex6D so it is not possible to give an exact description of how this variation affects the mechanical

properties of the unfolding force of Protein L. The mean unfolding force for ProtL Ex6D was measured at 158 ± 4 pN. This is very similar to the mean unfolding force measured for ProtL WT from the hetero-polyprotein construct, (I27-ProtL WT)$_3$-I27) at 158 ± 4 pN at 1200 nm/s (value taken form the first analysis, shown in table 4.1). The similarity of the unfolding forces of ProtL WT and ProtL Ex6D protein suggests that this mutation does not disrupt the contacts involved in the mechanical clamp.

The peak to peak distance reflects the number of amino acids unfolded during the structural transitions. A change in this value between Protein L variants may indicate a change in the unfolding pathway of the protein. The values for the extension of the protein upon unfolding (p2p distance) are all very similar for each variant of Protein L:

i) ProtL WT from the hetero–polyprotein construct, (I27-ProtL WT)$_3$-I27) = 16.4-16.6 nm (the variation between the different speeds shown in table C.2).

ii) ProtL WT from the homo–polyprotein construct (poly-ProtL WT) = 16.4 nm (shown in table 5.2).

iii) ProtL Kx6E from the hetero–polyprotein construct, (I27-ProtL Kx6E)$_3$-I27) = 16.2-16.4 nm (the variation between the different speeds shown in table C.4).

iv) ProtL Ex6D from the homo–polyprotein construct (poly-ProtL Ex6D) = 16.8 nm (shown in table 5.4).

Mutating a protein has been shown previously to have significant effects on its mechanical properties [29,145,155,168]. A study by Sadler et al. also measured the mechanical stability of ProtL variants. Although no mutations occur in exactly the same position relative to this study, a few were adjacent.

The mutations for ProtL Kx6E occurred at residue numbers 23, 28, 41, 54 and 61, and the mutations for ProtL Ex6D occurred at residue numbers 2, 3, 21, 27, 32 and 46. Sadler et al. carried out a single mutation at position 22, which is located in β–strand II. This was shown to have no effect on the mechanical stability of Protein L. Another mutation carried out by Sadler et al. was in position 60, which is located in β–strand IV and is one of the important hydrogen bonding residues in the mechanical clamp. When mutated from an isoleucine to a valine, the mechanical stability of Protein L was significantly

reduced (36 pN at 447 nm/s) relative to the wild type. When mutated instead to a phenylalanine residue, the mechanical stability of Protein L was significantly increased (72 pN at 447 nm/s). The study showed the only mutation that caused a significant change to the unfolding force of the protein was at this location, position 60. It was previously reported that the mechanical strength of Protein L may be modulated by the number of hydrophobic contacts across the two sub-units that are sheared apart[146]. Residue I60 has 10 contacting side chains in the core, 6 of these contacting side chains occur between the sub-units, as shown by the two units depicted in red and green on figure 4.3 B).

K61 for protein L has 5 contacting side chains in the core, 3 of these contacting side chains occur between the sub-units. When mutated to glutamic acid (K61E), the number of contacting side chains is reduced to 3 with only 1 occurring between sub-units. This mutation has, therefore, reduced the number of hydrophobic contacts across the two sub-units, resulting in a lower unfolding force relative to the wild type. The contact map for ProtL Kx6E is shown in figure 6.4. Only this one mutation (K61E) of the 6 mutations (K23E, K28E, K41E, K42E, K54E and K61E) is shown to change the number of contacting side chains occur between the sub-units.

Only one of the positions of mutations of ProtL Ex6D (E46D) has a contacting side chain that occurs between the sub-units of protein L. These mutations are therefore unlikely to modify the mechanical properties of protein L. This agrees with the SMFS data collected.

This study supports the claims by Sadler et al. that modulating the hydrophobic contacts across mechanical interfaces may be a generic method to tune the mechanical strength of a protein[145].
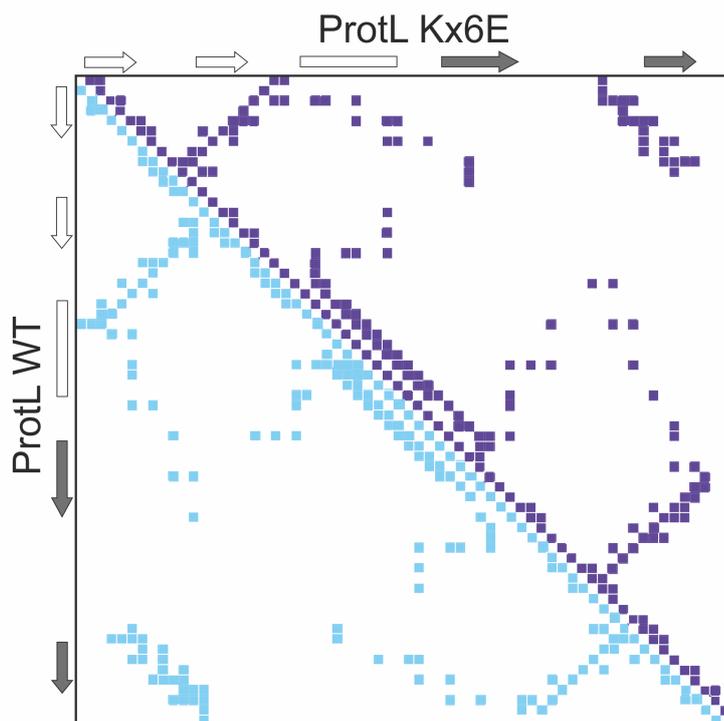
Figure 6.4: Contact maps of ProtL WT and ProtL Kx6E. The contact maps of ProtL WT (bottom left) and ProtL Kx6E (top right). Side-chain contacts (nearest distance between atoms of two residues <5 Å, calculated by CSU software) made by pairs of amino acids within structural unit 1 (β I + II and the helix - white) or within structural unit 2 (β III + IV - grey)[201]. Contacts made between the structural units of ProtL WT are shown in blue and contacts made between the structural units of ProtL Kx6E are shown in purple. β-strands are shown as arrows and α-helices are shown as rectangles alongside each contact map. The two structural units are coloured white (unit 1) and grey (unit 2) in each protein and are also shown superimposed onto the three-dimensional structure of ProtL WT(left) and ProtL Kx6E (right).

6.3 How does an increase in salt concentration affect protein mechanical properties?

Chapter 4 presents the work carried out on ProtL WT and ProtL Kx6E. The chapter presents the SMFS data collected for two hetero-polyprotein constructs, (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27, at four pulling velocities and two KCl concentrations (0.5 M and 2 M KCl).

Previous studies that have examined the effect salt concentration for mechanical protein unfolding

A few studies have studied the effect salt concentration for mechanical unfolding or other SMFS studies[210–214]. One study by Leake et al. used optical tweezers to measure the change in persistence length of the protein (0.4-2.7 nm) by increasing the ionic strength (from 15 to 300 mM) of its environment[210]. Another study, carried out by Zhang et al. examined the salt dependent binding of single–stranded DNA to *Bacillus subtilis* DNA-Binding Protein, using a NaCl concentrations of 0.1 M and 0.5 M, reporting a higher unbinding force in 0.5 M NaCl[213]. A very interesting study by Pirzer et al. used SMFS experiments to probe the influence of a number of salts from the Hofmiester series on the adhesion of recombinant spider silk proteins onto various solid substrates[212]. A spider silk protein was covalently bound at one end to an AFM tip and was brought into contact with a substrate and the desorption force was measured. The study showed a small rise in the desorption forces with the kosmotropicity of the anions (at a concentration of 1 M salt) from 80 ± 2 pN for NaClO$_4$ to 87 ± 2 pN for NaF. By contrast, the kosmotropicity of the cations was found not to have any effect on the desorption force. The study also observed that the force measured had a linear dependence on the concentration of three salts (NaCl, NaH$_2$PO$_4$ and NaI) from 0 M to 4 M salt. NaI was found to reduce desorption force with increasing salt concentration and NaH$_2$PO$_4$ and NaI were both found to increase desorption force with increasing salt concentration. Finally a study by Gronau at al. used steered molecular dynamics to simulate the response of a spider silk protein domain under mechanical force[214]. They

reported a lower unfolding force for the protein dimer in 0.5 M NaCl than 0 M NaCl (266 and 383 pN respectively) due to the breaking of the salt bridges at the interface between to two monomers in the higher salt concentration. None of these studies measure the mechanical strength of a protein in a salt concentration of 2 M.

What effect does an increased salt concentration have on the mechanical stability of a mesophilic and a halophilic protein?

This section will discuss the speed dependence diagrams from the first set of data analysis carried out, which was presented in section 4.7. The comparison of all of the mean unfolding forces for these data is shown in figure 4.28. Although the data collected in 2 M KCl does not form an accurate speed dependence plot as expected, i.e. all data points do not lie on the linear fit, it was possible to form hypotheses based on the presented speed dependence diagrams.

In the present study, ProtL Kx6E was shown to be mechanically weaker than ProtL WT in 0.5 M KCl. This trend was shown to remain at 2 M KCl (figure 4.28 B) and C)). Figure 4.29 depicts the change in the unfolding force from 0.5 M to 2 M KCl for each protein as a percentage of the unfolding force of the protein in 0.5 M KCl. The figure highlights the consistent increase in unfolding force of I27 from both constructs and ProtL WT, i.e. all 3 mesophilic proteins, from 0.5 to 2 M salt. The average increases were calculated as 8.6% for ProtL WT, 13.1% for I27 (WT) and 6.8% for I27 (Kx6E). ProtL Kx6E does not show a consistent increase in force from 0.5 M to 2 M KCl with a average increase of 1.6%.

Section 1.2.4 discusses the effect of salt on a mesophilic protein including three models. These will be discussed in the context of this new data on protein L.

i) The electrostatic screening of surface charge-charge interactions. This theory is only valid in low concentrations of salt, i.e. lower than about 100 mM. The effect of 2 M KCl on the proteins, therefore, cannot be explained using electrostatic screening alone (see section 1.2.4.1).

ii) The increase in the unfolding force in 2 M KCl, compared to 0.5 M KCl for the

mesophilic proteins may be due to an ion specific binding effect. In order to check this effect, the SMFS experiments would need to be carried out in a range of salts.

iii) The increase in the unfolding force in 2 M KCl, compared to 0.5 M KCl for the mesophilic proteins may be due to the high concentrations of hydrated salt ions being excluded from the protein surface, leading to an increase in the hydrophobic interactions within the protein[65–67]. Increased hydrophobic interactions cause an increase in the stability of a protein.

Changes in hydrophobic effects on the mechanical strength of a protein have previously been studied by SMFS. Two studies carried out by Sadler et al. and Bu et al. studied the effects of mutations on the mechanical properties of a protein[145,215]. Both of these studies reported significant effects on the measured unfolding force due to mutations within the hydrophobic core. One study by Dougan et al[151]. examined the influence of the solvent on protein mechanical stability. They reported that hydrophobic interactions are enhanced in D_2O, which caused an increase in mechanical strength of the protein. The results of this study provided the first single-molecule level measurement of the influence of D_2O on the hydrophobic effect during protein folding.

Why is it important for the stability of the obligate halophilic protein not to increase in high salt?

SMFS in this project is used to measure the stability of a protein. However, nature must balance protein stability with conformational flexibility to obtain optimal protein function. The flexibility of a protein is closely linked to its function[216]. For a protein to interact with another molecule, it must have the ability to change its conformation. Any perturbation to the flexibility of protein may, therefore, interfere with its function.

Evidence of the reduced flexibility of a protein in salt has been shown previously by Majumdar et al. who studied the effects of salts from the Hofmeister series on the local flexibility of an IgG1 monoclonal antibody, using hydrogen/deuterium exchange with mass spectrometry[217]. They showed that the addition of 0.5 M NaCl caused only a few significant changes in the local flexibility, with an overall trend of decreased flexibility.

The protein in 0.5 M Na$_2$SO$_4$ and 0.5 M NaSCN both exhibited an overall trend of increased flexibility. For a protein to function in high salt it must, therefore, remain flexible by overcoming the increased hydrophobic effect acting on the protein.

Negatively charged residues, especially aspartic acid have been shown to be characteristic for flexible but ordered proteins[218]. Soluble halophilic enzymes have highly negative surface charge densities, therefore, their flexibility may be achieved by repulsion between those charges[85]. Since ProtL Kx6E has an increased number of negatively charged residues compared to ProtL WT, it may remain flexible in the high salt concentration, despite the increased hydrophobic interactions. This may explain why the mechanical stability of the protein did not increase in 2 M KCl. Evidence of this effect is shown in figure 4.29, where ProtL Kx6E does not show such a consistent increase in force from 0.5 M to 2 M KCl. Further evidence of this effect is shown for the overlapping linear fits for ProtL Kx6E in 0.5 M and 2 M KCl in the refined data analysis speed dependence diagrams, figure 4.30.

Why does the change in thermodynamic stability of ProtL WT and ProtL Kx6E in an increased salt concentration differ from the change in the measured mechanical stability?

The thermodynamic stability of ProtL WT, measured by Tadeo et al., was shown to be independent of salt concentration[111]. The thermodynamic stability of ProtL Kx6E was shown to increase with the concentration of NaCl (see section 4.1.1.1 for further details). It has been shown previously that the thermodynamic stability does not correlate with the mechanical stability of a protein[139,219]. This is because the thermodynamic stability is be defined as the difference in free energy between its folded and unfolded states, whereas the mechanical stability describes the kinetic stability of localised areas of the protein, i.e. the mechanical clamp[135]. It should also be mentioned that the buffer Tadeo et al. used to measure the thermodynamic stability of the protein (20 mM phosphate buffer at pH 6.0) is different to that used here to measure the mechanical stability (5 mM MOPS, pH 7.4)

6.3.1 Difficulties of SMFS experiments in a high concentration of salt

SMFS experiments carried out in 2 M KCl were found to be particularly difficult. The number of protein unfolding events recorded for one day was significantly reduced (see table 4.5). There are a number of factors that could be affecting the number of unfolding events observed in a high salt environment over a certain time period. The salt could interfere with the binding of the cysteine residues to the gold surface as the chemistry of the sulphur-gold bonding is not fully understood[220]. In order to try and accommodate for this possibility, the protein solution in 2 M KCl was incubated on the gold surface for an hour instead of 30 minutes for the protein solution in 0.5 M KCl. Non-specific interactions occur when the AFM tip comes into contact with the protein construct. A high salt concentration might affect the strength of the interaction. If this interaction is reduced, the protein will have less chance of binding to the tip and, therefore, a lower hit rate will be observed.

6.4 Future directions

6.4.1 Production of a polyprotein containing *Hm*Fd

A future attempt at the expression of a polyprotein construct containing more repeats of *Hm*Fd could be carried out using an expression system specifically designed for the conditional overexpression of halophilic proteins in halophilic organisms[221–223]. A series of plasmid vectors and host strains for this purpose were developed by Allers et al. in *Haloferax volcanii*[221]. This organism was introduced in section 1.3.2.2. The plasmids utilise a tryptophan–inducible promoter that promotes rapid and strong induction of protein expression to increase protein yield.

Since the expression of I27$_3$-*Hm*Fd-I27 was previously shown to be successful, the inclusion of only one *Hm*Fd domain in a polyprotein with four I27 domains must be preferable. The polyprotein could be redesigned to move the HmFd domain further away from the surface. This would decrease the chance of 'non–specific' interactions

interfering with the HmFd unfolding pathway. The chimera polyprotein, I27$_2$-*Hm*Fd-I27$_2$, would be an ideal candidate alternative since only three I27 domains would need to unfold to be sure the *Hm*Fd domain had been subjected to force. This would make a *Hm*Fd statistically more likely.

6.4.2 Further characterisation of ProtL Kx6E compared to ProtL WT

To further investigate the effect that mutating six lysine residues to glutamic acid residues has on the stability of ProtL, the thermodynamic stability of the wild type and the ProtL Kx6E variant could be measured. Although the study carried out by Tadeo et al., which introduced ProtL Kx6E, does examine the thermodynamic stability, this is carried out in a different buffer with NaCl instead of KCl. Since potassium is found in higher concentrations than sodium in the cytoplasm of halophilic organisms, it is important to characterise the protein in KCl[224]. These experiments should be performed using Circular Dichroism (CD) and fluorescence spectroscopy.

CD spectroscopy is a form of light absorption spectroscopy that measures the interactions of chiral molecules with circularly polarised light. Ellipticity is the unit of CD and is recorded as a function of the wavelength of the light going through the sample. The absorption at particular wavelengths can give information about the secondary structure of a protein[225]. CD spectroscopy can be used to follow the unfolding of a protein as a function of temperature. As temperature is increased, the protein can undergo conformational changes and the secondary structure may change. These changes can, therefore, be measured using CD and the melting temperature, T_m, of the protein can be calculated. Fluorescence spectroscopy measures the intensity of photons emitted from a sample after it has absorbed photons. Fluorescence from the amino acid tryptophan has been shown to be sensitive to the polarity of its local environment[226]. The native state of a protein can be altered by increasing the temperature, as the protein unfolds, the tryptophan, previously buried in the hydrophobic core, can be exposed to the solvent. Changes in intrinsic protein fluorescence can be used to monitor the protein unfolding. Again, the T_m of the protein can be calculated. CD and fluorescence spec-

troscopy are more high-throughput techniques so the thermodynamic stability of ProtL WT and ProtL Kx6E could be measured by equilibrium denaturation, using urea or guanidinium chloride, in a wider range of salt concentrations (or wider range of salts) for a wider picture of how both proteins are affected by salt.

Monte Carlo (MC) simulations can be used to determine the underlying energy landscape parameters: Δx_U, the distance from the unfolded to the transition state and k_U, the unfolding rate at zero force. This enables the height of the TS to be determined[148,152,227]. The simulations are used to produce histograms of F_{UN} for each protein domain (e.g. I27 and ProtL WT for (I27-ProtL WT)$_3$-I27). For each pulling velocity, the values of the unfolding rate constant at zero force, k_U, and the distance between the folded state and the transition state, Δx_U, are optimised iteratively until the values of mean force and standard deviation of the peak unfolding force histograms agree with the values obtained for the experimental data. The dependence of the mechanical force of a protein on the pulling velocity should be described by a linear fit. The simulations use this linear fit as well as the average spring constant of the cantilever, temperature, persistence length, contour length, the linker length (the amino acids between proteins in a polyprotein), the length of a single folded and unfolded protein domain and the distribution of forces generated. Since the F_{UN} values measured for (I27-ProtL WT)$_3$-I27) and (I27-ProtL Kx6E)$_3$-I27) in 0.5 M KCl do show a linear fit to the pulling velocity, MC simulations could be used to extract values of k_U and Δx_U and, therefore, learn more about the energy landscape of the model halophilic protein, ProtL Kx6E.

The effect of the mutations on the structural properties of the unfolding transition state of protein L compared to that of the wild type can be investigated using molecular dynamics (MD) simulations. Mechanical unfolding of polyproteins can be modelled using MD simulations. This technique has been used to gain atomistic detail of the unfolding process of a number of proteins[29,146,156,204].

A larger mutation study could be carried out on Protein L to further examine the effect of mutations on the mechanical stability of Protein L. The contact map, figure 6.4, suggests that the mutation at position 61 is the cause of the lower unfolding force of ProtL

Kx6E compared to ProtL WT. This hypothesis could be further examined by varying the number of mutations of K to E carried out. It would be interesting to determine if a mutation at position 61 alone resulted in the observed reduction of the mechanical stability of the protein. If this is not the case, it would be interesting to quantify the additional contribution of each sequential mutation to the reduced mechanical stability, going from ProtL KxE to ProtL Kx6E.

SMFS experiments could be carried out on variants of the mutation position 61 in Protein L. The mutation of lysine to glutamic acid varies both the length and charge of the amino acid side chain. If the charge remained constant but the length of the side chain was varied, would the unfolding force of Protein L change?

6.4.3 Examining the behaviour of a halophilic protein L in high salt concentrations

A homo-polyprotein of a halophilic model protein, poly-ProtL Ex6D, formed by crosslinking, was shown as a viable alternative construct to use in SMFS experiments in chapter 5. It was suggested that the construct could provide more data on a shorter timescale and, therefore, increase the accuracy of data collected by providing enough data to use a refined data analysis technique. In order to accurately characterise the mechanical unfolding behaviour of ProtL Kx6E in higher salt concentrations, a homo-polyprotein of ProtL Kx6E would, therefore, need to be produced. Once the mechanical unfolding data has been collected and analysed using the refined analysis technique. If accurate mechanical unfolding data is collected, Monte Carlo (MC) simulations could then be used to determine the parameters controlling the energy landscape of the protein in an increased salt concentration, as described above.

Modelling the data using MD simulations could again be used to gain atomistic detail of the mechanical unfolding process of a protein in an increased salt concentration. These simulations provide a method to experimentally map the structure of the transition state. Obtaining atomistic details of the mechanical unfolding process of a halophilic protein in a high and low concentration of salt could lead to a greater understanding of the effect

of salt on the mechanical stability of a protein and the associated underlying molecular mechanisms.

This study could be developed to examine the effect of the Hofmeister salts on the mechanical stability of the protein. If the hypothesis that the increase in the concentration of KCl causes an increase in the hydrophobic interactions within the protein and, therefore, the mechanical stability of the protein, is correct, the properties of the salt should affect the outcome. For example, would an increase in concentration of a chaotropic salt cause a decrease in the mechanical stability of a mesophilic protein? Also would an increase in an extremely kosmotropic salt cause a larger increase in the stability of a mesophilic protein?

This project has shown that SMFS is a useful tool to examine the effect of salt on the stability of a halophilic protein. Further study using SMFS experiments in combination with MD simulations has the potential to gain atomistic information of the mechanical unfolding behaviour of a halophilic protein in high and low salt concentrations.

Appendix A

Supplementary information for materials and methods

Primer name	Nucleic acid sequence
HmFd_F	caccatcaccatatggaaaacctgtactttcagggctgttctggctctccgaccgtggaatatctgaact
HmFd_R	gtgatggtgggatccttaacagccagaaatcacgcggttctgcagataat
PL_WT_a75g	atggatccacacaaactgcagagttcaaaggaacatttgaaaaag
PL_WT_a75g_anti	ctttttcaaatgttcctttgaactctgcagtttgtgtggatccat
PL_WT_a57t	taacctaatctttgcaaatggttccacacaaactgcagaattc
PL_WT_a57t_anti	gaattctgcactttgtgtggaaccatttgcaaagattaggtta
PL_Ex6D_a57t	taacctaatctttgcaaatggttccacacaaactgcagatttc
PL_Ex6D_a57t_anti	gaaatctgcagtttgtgtggaaccatttgcaaagattaggtta
PL_WT_F	catcaccatggatccggttgtggtgaagaagtaacaat
PL_WT_R	atggtgatggaattcctatccacaaaatttaatattttttgtat
PL_Ex6D_F	catcaccatggatccggttgtggtgatgatgtaacaat

Table A.1: All primers where red letters are non–complementary, underlined letters are restriction sites, blue letters is a TEV site and green letters are linker regions (containing cysteine).

Protein	Amino acid sequence
HmFd	P T V E Y L N Y E V V D D N G W D M Y D D D V F G E A S D M D L D D E D Y G S L E V N E G E Y I L E A A E A Q G Y D W P F S C R A G A C A N C A A I V L E G D I D M D M Q Q I L S D E E V E D K N V R L T C I G S P D A D E V K I V Y N A K H L D Y L Q N R V I
ProtL WT (Tadeo/hetero)	M E E V T I K A N L I F A N G S T Q T A E F K G T F E K A T S E A Y A Y A D T L K K D N G E Y T V D V A D K G Y T L N I K F A G
ProtL WT (homo)	M G C E V T I K A N L I F A N G S T Q T A E F K G T F E K A T S E A Y A Y A D T L K K D N G E W T V D V A D K G Y T L N I K F C G
ProtL Kx6E (hetero)	M E E V T I K A N L I F A N G S T Q T A E F E G T F E E A T S E A Y A Y A D T L E E D N G K W T V D V A D E G Y T L N I E F A G
ProtL Ex6D (homo)	M G C G D D V T I K A N L I F A N G S T Q T A D F K G T F D K A T S D A Y A Y A D T L K K D N G D W T V D V A D K G Y T L N I K F C G

Table A.2: Amino acid sequences of proteins used. Sequences do not include the His–tag

Appendix B

Supplementary information for Results 1
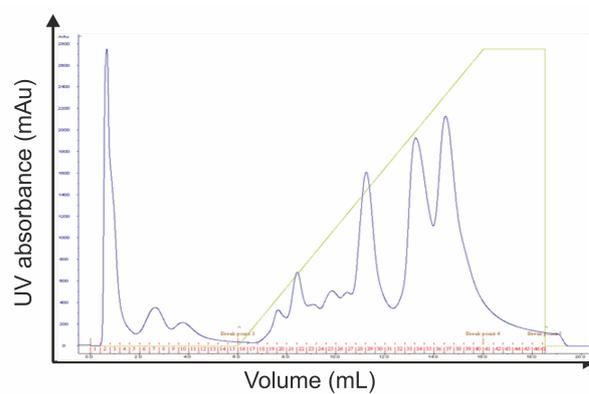
Figure B.1: Elution profile of the purification of MBP-HmFd on a Hi-Trap Q column. The figure shows the absorbance in blue, the concentration of buffer 2 in green and fraction number in red.
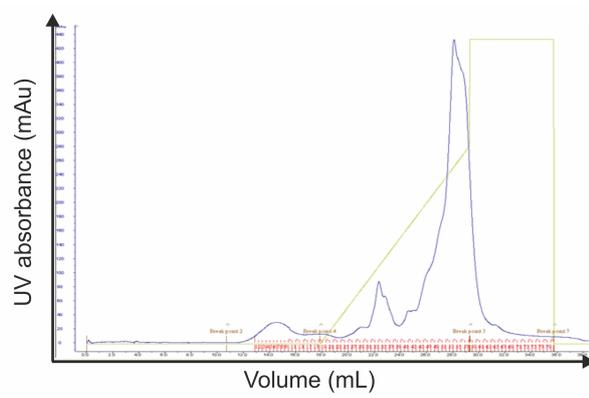
Figure B.2: Elution profile of the purification of MBP-HmFd on a SOURCE™ 15Q column. The figure shows the absorbance in blue, the concentration of buffer 2 in green and fraction number in red.

Appendix C

Supplementary information for Results 2
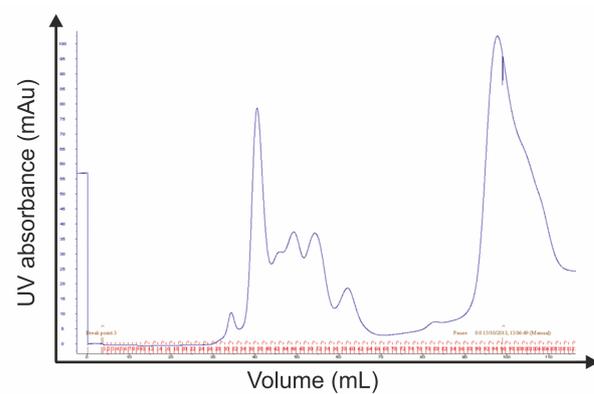
Figure C.1: Elution profile of the purification of (I27-ProtL WT)$_3$-I27 on a size exclusion column. The figure shows the absorbance in blue and fraction number in red.
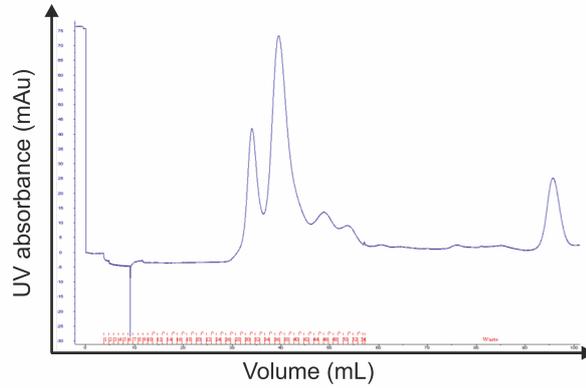
Figure C.2: Elution profile of the purification of (I27-ProtL Kx6E)$_3$-I27 on a size exclusion column. The figure shows the absorbance in blue and fraction number in red.
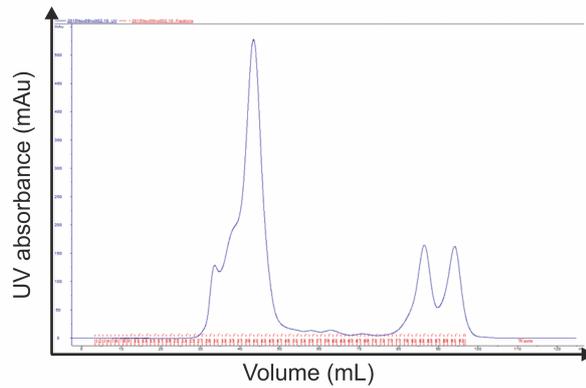
Figure C.3: Elution profile of the purification of (I27-ProtL Kx6E)$_3$-I27 using Ni-affinity chromatography in 6 M GdmCl. The figure shows the absorbance in blue and fraction number in red.

Pulling velocity (nms^{-1})	Events	Mean F$_{UN}$ (pN)	SD	Events	Mean F$_{UN}$ (pN)	SD
	ProtL			I27		
400	96	136	23	110	178	23
700	114	146	23	134	179	24
1200	111	160	30	119	195	31
2000	119	165	25	127	201	28

Table C.1: Median unfolding force data and standard deviation in (I27-ProtL WT)$_3$-I27 in 0.5M KCl, 5mM MOPS, pH 7.4.
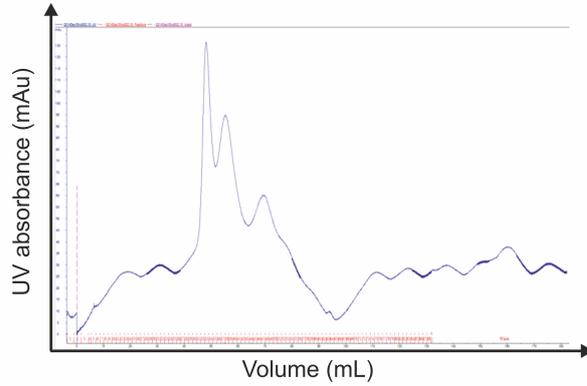
Figure C.4: Elution profile of the purification of (I27-ProtL Kx6E)$_3$-I27 on a size exclusion column at 4 °C. The figure shows the absorbance in blue and fraction number in red.

Pulling velocity (nms^{-1})	Events	Mean p2p (nm)	SD	Events	Mean p2p (nm)	SD
	ProtL			I27		
400	96	16.5	0.7	110	23.8	0.7
700	114	16.4	0.7	134	23.8	0.7
1200	111	16.6	0.8	119	23.9	0.8
2000	119	16.6	0.8	127	23.8	0.7

Table C.2: Mean peak to peak distance data and standard deviation in (I27-ProtL WT)$_3$-I27 in 0.5M KCl, 5mM MOPS, pH 7.4.

Pulling velocity (nms^{-1})	Events	Mean F_{UN} (pN)	SD	Events	Mean F_{UN} (pN)	SD
	ProtL			I27		
400	74	122	22	93	187	25
700	56	135	23	71	186	27
1200	75	139	25	106	189	33
2000	104	151	26	132	212	30

Table C.3: Median unfolding force data and standard deviation for (I27-ProtL Kx6E)$_3$-I27 in 0.5M KCl, 5mM MOPS, pH 7.4.
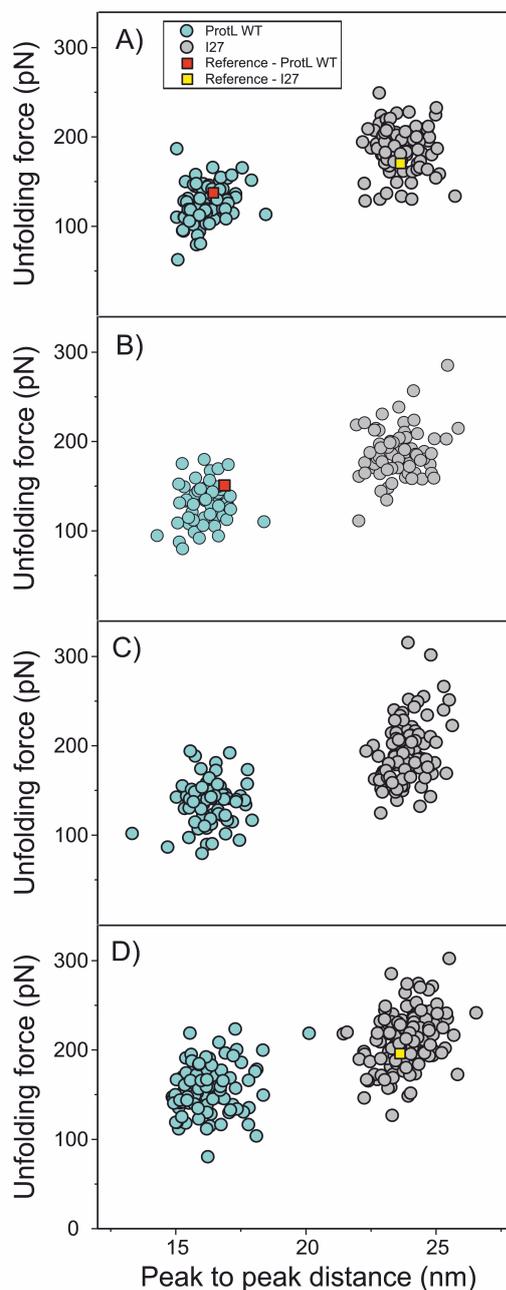
Figure C.5: Scatter diagrams of the unfolding force vs. peak to peak distance for (I27-ProtL WT)$_3$-I27 at 4 pulling velocities in 0.5 M KCl. All figures show ProtL WT in blue and I27 in light grey. Figures A), B), C) and D) show the data from 400, 700, 1200 and 2000 nms^{-1} respectively. The poly-protein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer. The ProtL WT and I27 points show a cluster around their respective reference points. A reference value of unfolding force and peak to peak distance is shown as a yellow square for ProtL WT and I27, error bars are negligible.
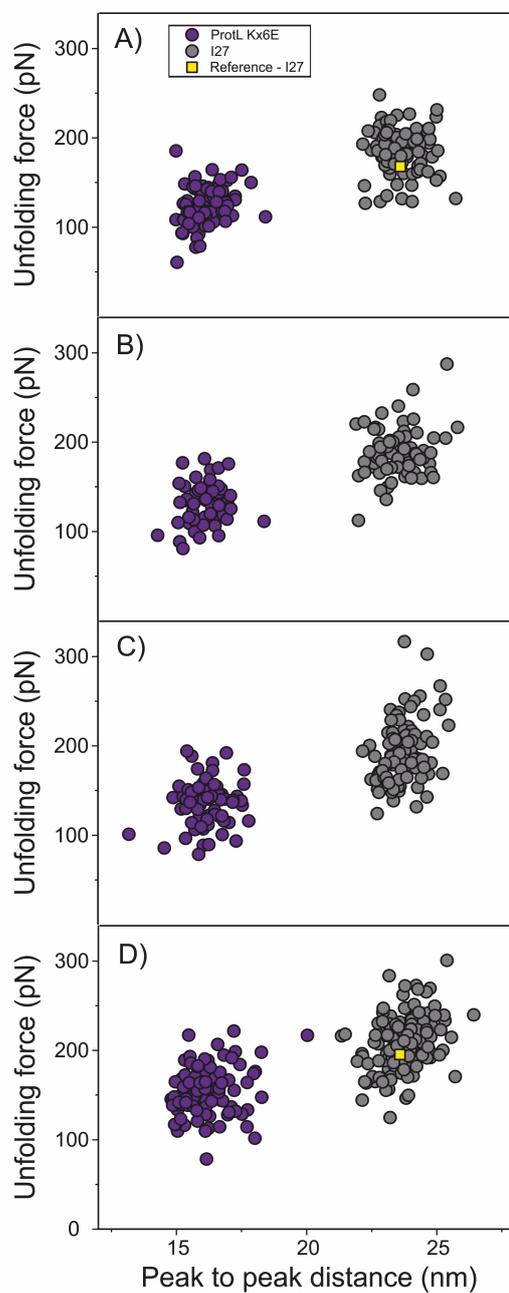
Figure C.6: Scatter diagrams of the unfolding force vs. peak to peak distance for (I27-ProtL Kx6E)$_3$-I27 at 4 pulling velocities in 0.5 M KCl. All figures show ProtL Kx6E in purple and I27 in grey. Figures A), B), C) and D) show the data from 400, 700, 1200 and 2000 nms^{-1} respectively. The poly-protein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer.
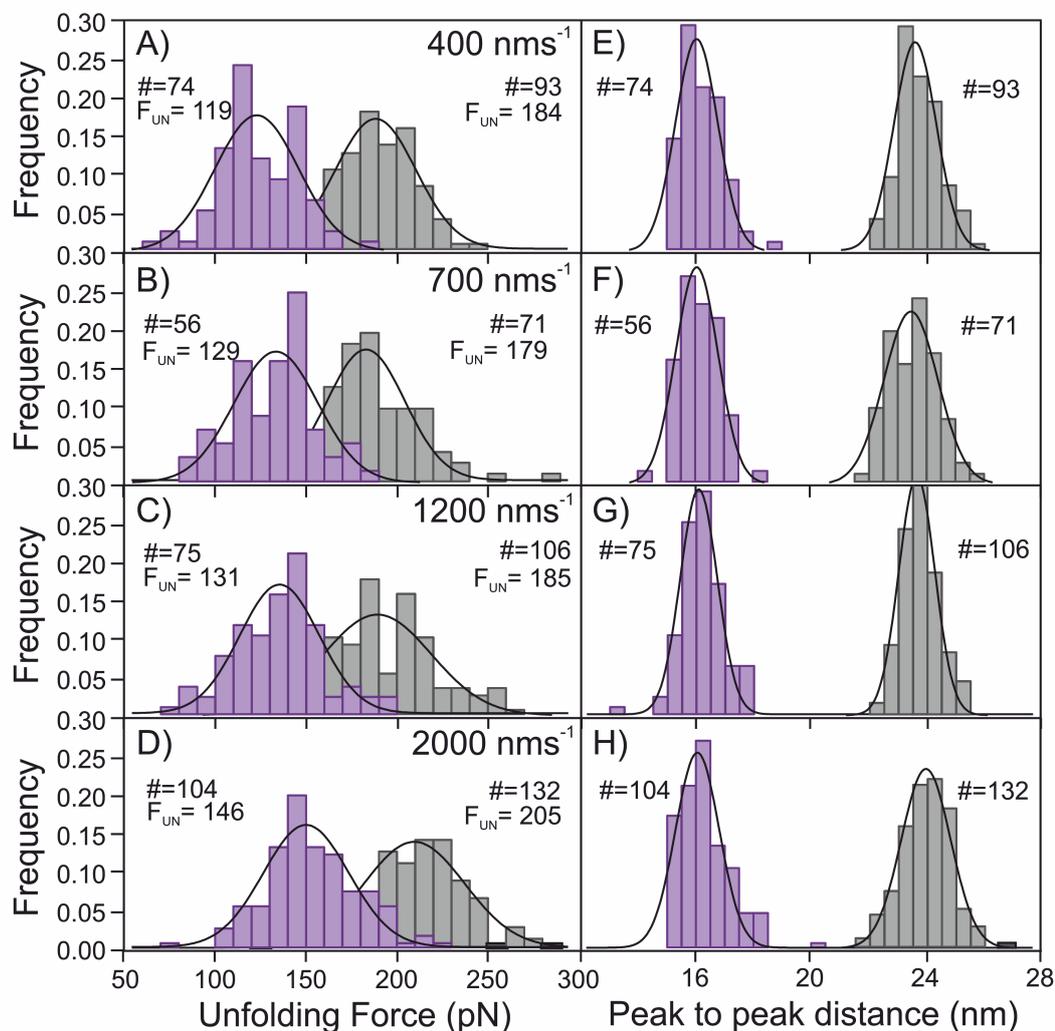
Figure C.7: Histograms of the unfolding force data and peak to peak distance data for poly-ProtL Kx6E at 400, 700, 1200 and 2000 nms⁻¹ in 0.5 M KCl. All figures show ProtL Kx6E in purple and I27 in grey. Figures A), B), C) and D) show the peak to peak distance data and figures E), F), G) and H) show the unfolding force data from each pulling velocity respectively. All histograms are fitted with a gaussian curve. The poly-protein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL Kx6E and I27.

Pulling velocity (nms^{-1})	Events	Mean p2p (nm)	SD	Events	Mean p2p (nm)	SD
	ProtL			I27		
400	96	16.2	0.7	110	23.7	0.7
700	114	16.2	0.7	134	23.5	0.9
1200	111	16.2	0.8	119	23.7	0.7
2000	119	16.4	0.9	127	24.0	0.9

Table C.4: Mean peak to peak distance data and standard deviation in (I27-ProtL Kx6E)$_3$-I27 in 0.5M KCl, 5mM MOPS, pH 7.4.

Pulling velocity (nms^{-1})	Events	Mean F$_{UN}$ (pN)	SD	Events	Mean F$_{UN}$ (pN)	SD
	ProtL			I27		
400	71	155	23	81	198	24
700	50	158	24	62	214	30
1200	82	173	32	92	224	29
2000	68	172	31	86	214	34

Table C.5: Median unfolding force data and standard deviation for (I27-ProtL WT)$_3$-I27 in 2M KCl, 5mM MOPS, pH 7.4.

Pulling velocity (nms^{-1})	Events	Mean p2p (nm)	SD	Events	Mean p2p (nm)	SD
	ProtL			I27		
400	96	16.7	0.7	110	23.9	0.7
700	114	16.8	0.7	134	24.1	0.7
1200	111	16.4	0.7	119	23.9	0.8
2000	119	16.7	0.8	127	24.0	0.8

Table C.6: Mean peak to peak distance data and standard deviation in (I27-ProtL WT)$_3$-I27 in 2 M KCl, 5mM MOPS, pH 7.4.

Pulling velocity (nms^{-1})	Events	Mean F$_{UN}$ (pN)	SD	Events	Mean F$_{UN}$ (pN)	SD
	ProtL			I27		
400	8	143	14	11	201	16
700	49	136	22	67	201	33
1200	53	142	28	65	201	27
2000	23	154	24	30	224	34

Table C.7: Median unfolding force data and standard deviation for (I27-ProtL Kx6E)$_3$-I27 in 2M KCl, 5mM MOPS, pH 7.4.
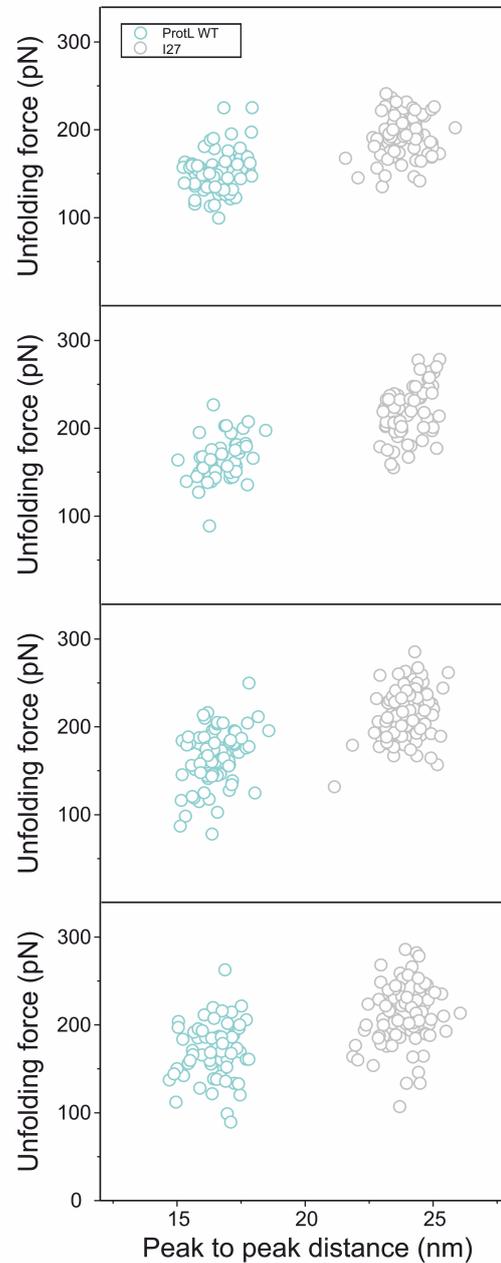
Figure C.8: Scatter diagrams of the unfolding force vs. peak to peak distance for (I27-ProtL WT)$_3$-I27 at 4 pulling velocities in 2 M KCl. All figures show ProtL WT outlined in blue and I27 outlined in light grey. Figures A), B), C) and D) show the data from 400, 700, 1200 and 2000 nms^{-1} respectively. The poly-protein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer.
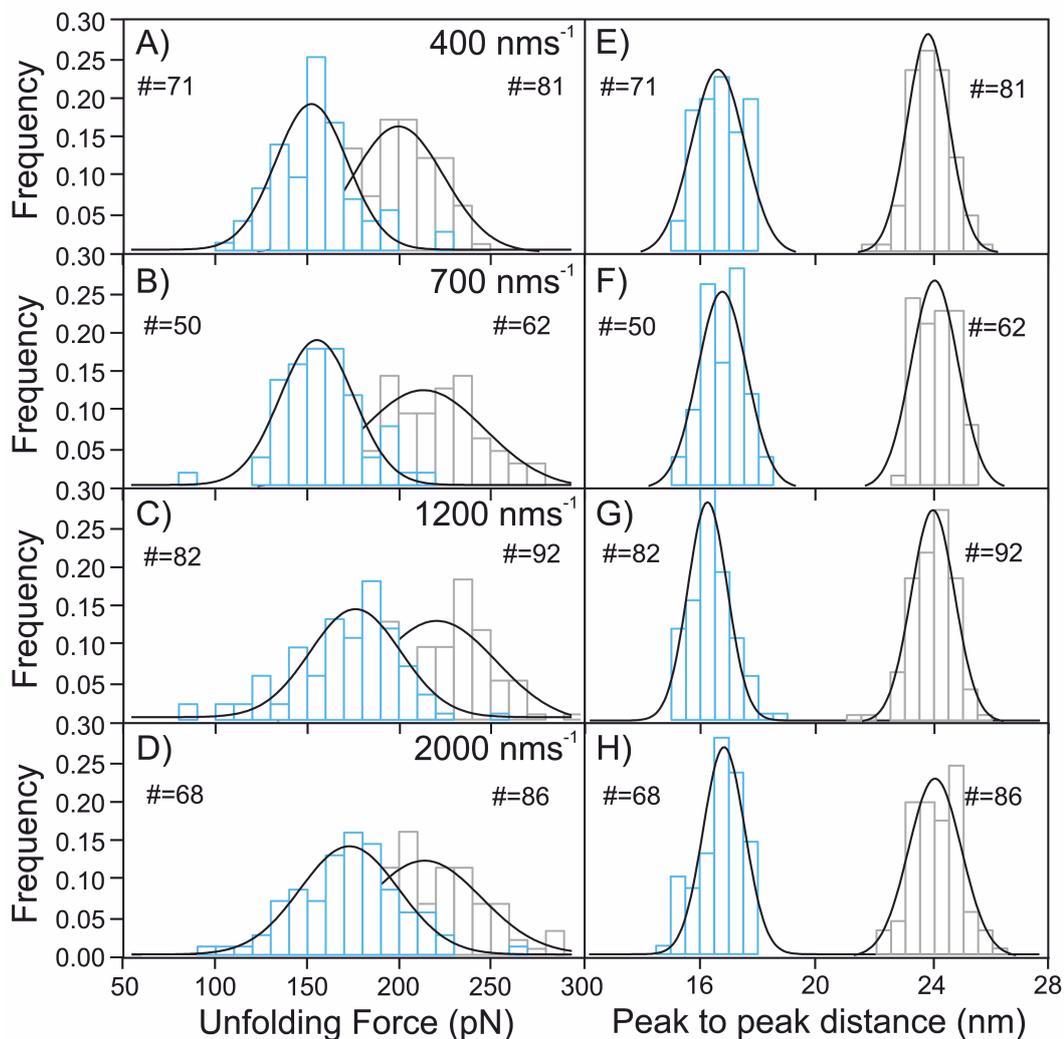
Figure C.9: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL WT)$_3$-I27 at 400, 700, 1200 and 2000 nms^{-1} in 2 M KCl. All figures show ProtL WT outlined in blue and I27 outlined in light grey. Figures A), B), C) and D) show the peak to peak distance data and figures E), F), G) and H) show the unfolding force data from each pulling velocity respectively. All histograms are fitted with a gaussian curve. The poly-protein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL WT and I27.
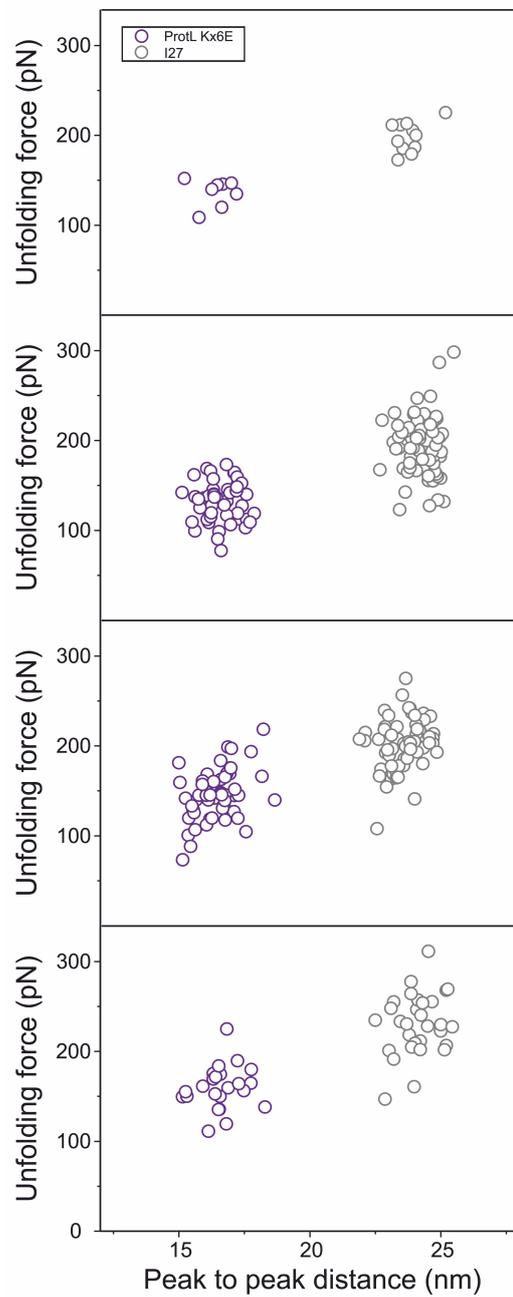
Figure C.10: Scatter diagrams of the unfolding force vs. peak to peak distance for (I27-ProtL Kx6E)$_3$-I27 at 400, 700, 1200 and 2000 nms^{-1} in 2 M KCl. All figures show ProtL Kx6E outlined in purple and I27 outlined in grey. Figures A), B), C) and D) show the data from each pulling velocity respectively. The poly-protein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer.
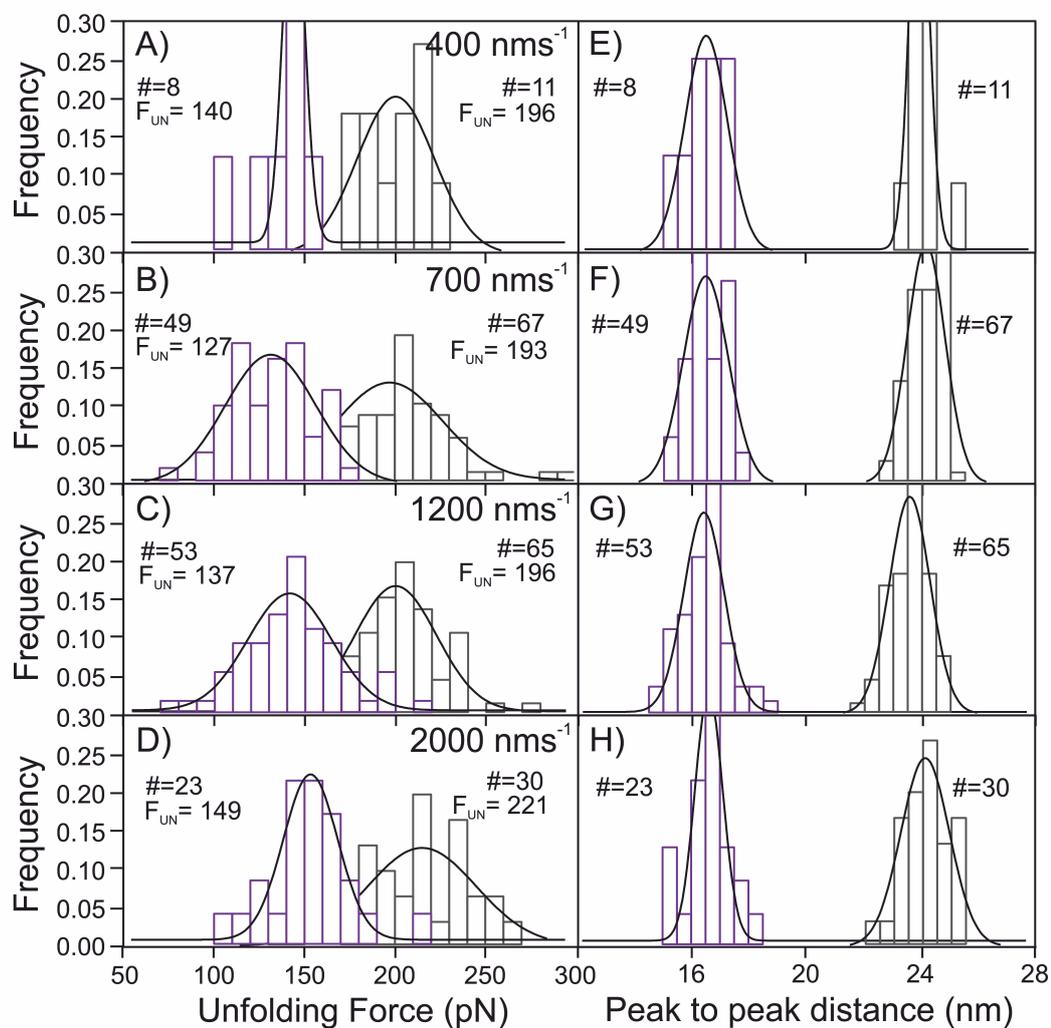
Figure C.11: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL Kx6E)$_3$-I27 at 400, 700, 1200 and 2000 nms^{-1} in 2 M KCl. All figures show ProtL Kx6E outlined in purple and I27 outlined in grey. Figures A), B), C) and D) show the peak to peak distance data and figures E), F), G) and H) show the unfolding force data from each pulling velocity respectively All histograms are fitted with a gaussian curve. The polyprotein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL Kx6E and I27.

Pulling velocity (nms^{-1})	Events	Mean p2p (nm)	SD	Events	Mean p2p (nm)	SD
		ProtL			I27	
400	96	16.4	0.6	110	24.0	0.5
700	114	16.5	0.7	134	24.1	0.6
1200	111	16.4	0.8	119	23.5	0.7
2000	119	16.6	0.8	127	24.1	0.8

Table C.8: Mean peak to peak distance data and standard deviation in (I27-ProtL Kx6E)$_3$-I27 in 2 M KCl, 5mM MOPS, pH 7.4.
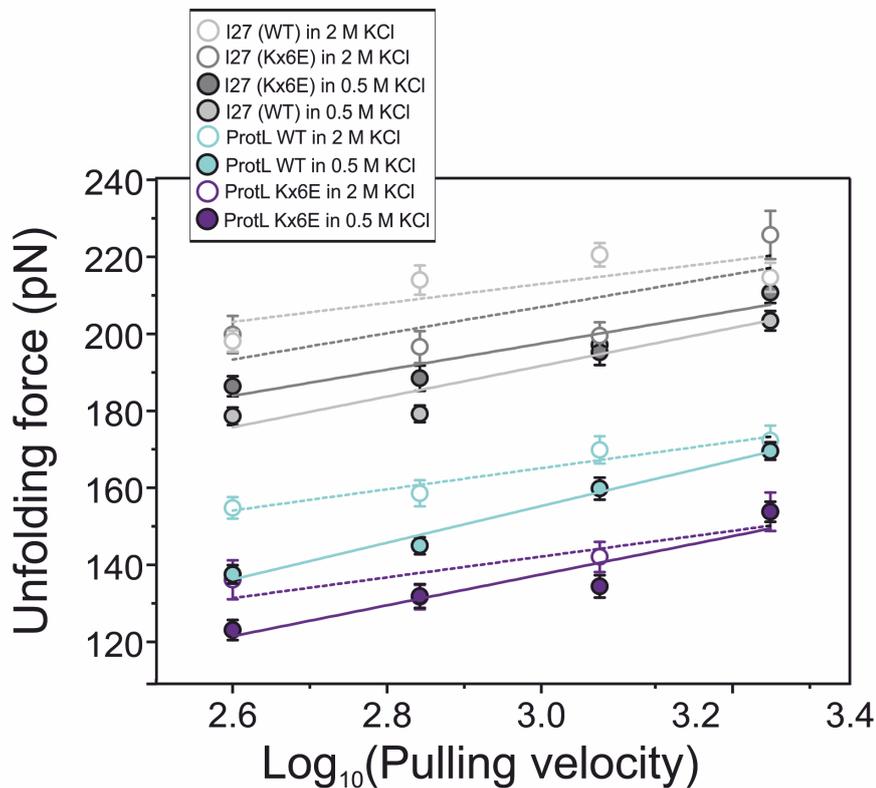
Figure C.12: Speed dependence diagram of the unfolding force vs. $\log_{10}$(pulling velocity) for all 4 systems in one. The figure shows the speed dependence of (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 0.5 M and 2 M KCl. The linear fit for each set of data is also plotted as full lines in 0.5 M KCl and dashed lines in 2 M KCl.

Pulling velocity (nms^{-1})	Events	Median F$_{UN}$ (pN)	SD	events	Median F$_{UN}$ (pN)	SD
	ProtL			I27		
400	36	135	22	42	171	21
400	31	136	20	35	182	20
400	29	135	26	33	183	25
		135	**1**		**179**	**4**
700	49	142	24	55	176	21
700	29	155	25	43	188	27
700	36	147	19	36	178	24
		148	**4**		**180**	**4**
1200	19	154	30	24	195	29
1200	43	162	33	47	197	31
1200	49	160	26	48	193	31
		159	**3**		**195**	**1**
2000	36	169	24	39	197	34
2000	39	161	25	43	197	24
2000	44	176	23	45	205	24
		169	**4**		**200**	**3**

Table C.9: Median unfolding force data and standard deviation for (I27-ProtL WT)$_3$-I27 in 0.5M KCl, 5mM MOPS, pH 7.4. The numbers quoted in bold are the mean and standard error of the 3 unfolding force values for each pulling velocity

Pulling velocity (nms^{-1})	Events	Mean p2p (nm)	SD	events	Mean p2p (nm)	SD
	ProtL			I27		
400	36	16.6	0.6	42	23.9	0.7
400	31	16.6	0.8	35	23.6	0.7
400	29	16.1	0.6	33	23.7	0.6
		16.5	**0.1**		**23.8**	**0.1**
700	49	16.2	0.7	55	23.8	0.6
700	29	16.6	0.7	43	23.8	0.7
700	36	16.6	0.7	36	23.8	0.7
		16.4	**0.1**		**23.8**	**0**
1200	19	16.7	0.8	24	24.1	1.1
1200	43	16.6	0.8	47	23.8	0.6
1200	49	16.6	0.7	48	23.9	0.7
		16.6	**0**		**23.9**	**0.1**
2000	36	16.5	0.8	39	24.0	0.7
2000	39	16.6	0.7	43	23.8	0.8
2000	44	16.7	0.8	45	23.7	0.7
		16.6	**0.1**		**23.8**	**0.1**

Table C.10: Mean peak to peak distance data and standard deviation for (I27-ProtL WT)$_3$-I27 in 0.5M KCl, 5mM MOPS, pH 7.4. The numbers quoted in bold are the mean and standard error of the 3 peak to peak distance values for each pulling velocity

Pulling velocity (nms^{-1})	Events	Median F$_{UN}$ (pN)	SD	events	Median F$_{UN}$ (pN)	SD
	ProtL			I27		
400	39	135	21	46	189	21
400	28	134	25	32	182	24
400	35	121	20	39	184	29
		130	**5**		**185**	**2**
700	21	138	26	25	183	32
700	20	126	17	27	188	34
700	15	144	19	19	186	20
		136	**6**		**186**	**2**
1200	19	143	25	27	210	25
1200	31	135	23	42	188	40
1200	25	139	25	37	181	25
		139	**2**		**193**	**9**
2000	29	161	29	40	210	39
2000	44	145	25	54	209	23
2000	31	155	24	38	215	28
		154	**3**		**212**	**2**

Table C.11: Median unfolding force data and standard deviation for (I27-ProtL Kx6E)$_3$-I27 in 0.5M KCl, 5mM MOPS, pH 7.4. The numbers quoted in bold are the mean and standard error of the 3 unfolding force values for each pulling velocity

Pulling velocity (nms^{-1})	Events	Mean p2p (nm)	SD	events	Mean p2p (nm)	SD
	ProtL			I27		
400	39	16.4	0.7	46	23.6	0.7
400	28	16.1	0.8	32	23.8	0.8
400	35	16.2	0.6	39	23.5	0.6
		16.2	**0.1**		**23.7**	**0.1**
700	21	15.9	0.6	25	22.9	0.7
700	20	16.3	0.7	27	23.8	0.8
700	15	16.3	0.5	19	23.8	0.9
		16.2	**0.2**		**23.5**	**0.2**
1200	19	16.3	0.7	27	23.9	0.5
1200	31	15.9	0.5	42	23.5	0.7
1200	25	16.4	0.9	37	23.9	0.7
		16.2	**0.1**		**23.8**	**0.1**
2000	29	16.8	0.9	40	23.8	1.0
2000	44	16.2	0.9	54	24.1	0.8
2000	31	16.2	0.9	38	23.9	0.8
		16.4	**0.2**		**24.0**	**0**

Table C.12: Mean peak to peak distance data and standard deviation for (I27-ProtL Kx6E)$_3$-I27 in 0.5M KCl, 5mM MOPS, pH 7.4. The numbers quoted in bold are the mean and standard error of the 3 peak to peak distance values for each pulling velocity

Pulling velocity (nms^{-1})	Events	Median F$_{UN}$ (pN)	SD	events	Median F$_{UN}$ (pN)	SD
400	32	158	26	41	202	22
400	29	149	23	36	194	30
		154	**4**		**198**	**3**
700	17	151	18	20	206	38
700	30	156	25	35	223	31
		157	**2**		**213**	**7**
1200	32	177	33	36	231	22
		177	**6**		**231**	**4**
2000	27	163	26	40	222	34
		163	**5**		**222**	**5**

Table C.13: Median unfolding force data and standard deviation for (I27-ProtL WT)$_3$-I27 in 2 M KCl, 5mM MOPS, pH 7.4. The numbers quoted in bold are the mean and standard error of the unfolding force values for each pulling velocity

Pulling velocity (nms^{-1})	Events	Mean p2p (pN)	SD	events	Mean p2p (pN)	SD
400	32	16.9	0.7	41	24.1	0.7
400	29	16.4	0.7	36	23.7	1.7
		16.6	**0.2**		**23.9**	**0.1**
700	17	17.0	0.6	20	24.0	0.7
700	30	16.6	0.7	35	24.0	0.6
		16.8	**0.2**		**24.0**	**0**
1200	32	16.3	0.7	36	23.8	0.6
		16.3	**0.1**		**23.8**	**0.1**
2000	27	16.6	0.8	40	24.1	0.7
		16.6	**0.2**		**24.1**	**0.1**

Table C.14: Mean peak to peak distance data and standard deviation for (I27-ProtL WT)$_3$-I27 in 2 M KCl, 5mM MOPS, pH 7.4. The numbers quoted in bold are the mean and standard error of the unfolding force values for each pulling velocity

Pulling velocity (nms^{-1})	Events	Median F$_{UN}$ (pN)	SD	events	Median F$_{UN}$ (pN)	SD
	ProtL				I27	
700	34	131	23	43	188	27
700	16	142	16	23	212	28
		137	**4**		**200**	**10**
1200	25	142	26	32	203	34
		142	**7**		**203**	**5**
2000	17	151	32	23	226	34
2000	10	155	15	15	247	30
		153	**2**		**236**	**8**

Table C.15: Median unfolding force data and standard deviation for (I27-ProtL Kx6E)$_3$-I27 in 2 M KCl, 5mM MOPS, pH 7.4. The numbers quoted in bold are the mean and standard error of the unfolding force values for each pulling velocity

Pulling velocity (nms^{-1})	Events	Mean p2p (nm)	SD	events	Mean p2p (nm)	SD
	ProtL			I27		
700	34	16.7	0.7	43	24.3	0.5
700	16	16.2	0.6	23	23.8	0.7
		16.5	**0.2**		**24.1**	**0.2**
1200	25	16.3	0.9	32	23.3	0.7
		16.3	**0.2**		**23.3**	**0.1**
2000	17	16.7	0.6	23	24.2	0.8
2000	10	16.4	0.9	15	24.0	0.7
		16.6	**0.2**		**24.1**	**0.1**

Table C.16: Mean peak to peak distance data and standard deviation for (I27-ProtL Kx6E)$_3$-I27 in 2 M KCl, 5mM MOPS, pH 7.4. The numbers quoted in bold are the mean and standard error of the unfolding force values for each pulling velocity

Figure C.13: Scatter diagrams of the unfolding force vs. peak to peak distance for (I27-ProtL WT)$_3$-I27 with separated data sets in 0.5 M KCl at a pulling velocity of 400 nms^{-1}. All figures show ProtL WT in blue and I27 in light grey. Figures A), B) and C) show the data from each data set. The polyprotein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer.

Figure C.14: Scatter diagrams of the unfolding force vs. peak to peak distance for (I27-ProtL WT)$_3$-I27 with separated data sets in 0.5 M KCl at a pulling velocity of 400 nms^{-1}. All figures show ProtL WT in blue and I27 in light grey. Figures A), B) and C) show the data from each data set. The polyprotein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer.

Figure C.15: Scatter diagrams of the unfolding force vs. peak to peak distance for (I27-ProtL WT)$_3$-I27 with separated data sets in 0.5 M KCl at a pulling velocity of 1200 nms^{-1}. All figures show ProtL WT in blue and I27 in light grey. Figures A), B) and C) show the data from each data set. The polyprotein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer.
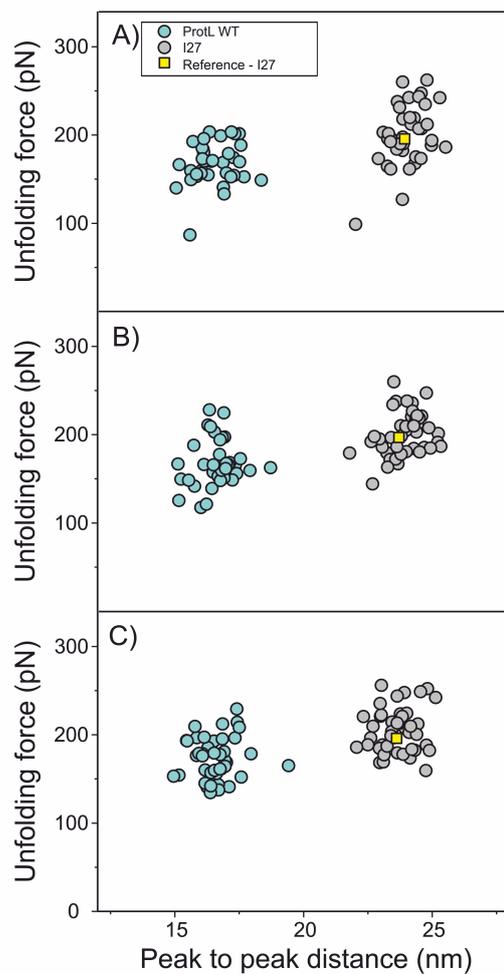
Figure C.16: Scatter diagrams of the unfolding force vs. peak to peak distance for (I27-ProtL WT)$_3$-I27 with separated data sets in 0.5 M KCl at a pulling velocity of 2000 nms^{-1}. All figures show ProtL WT in blue and I27 in light grey. Figures A), B) and C) show the data from each data set. The polyprotein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer.
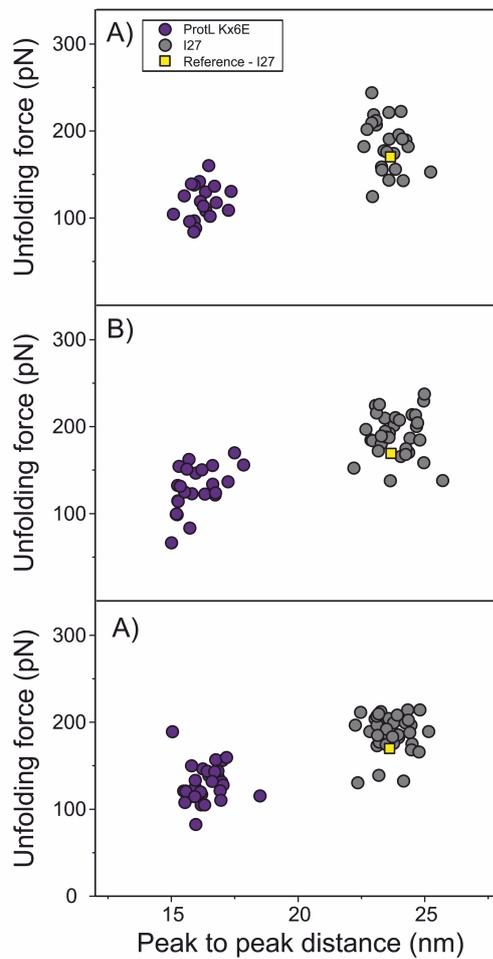
Figure C.17: Scatter diagrams of the unfolding force vs. peak to peak distance for (I27-ProtL Kx6E)$_3$-I27 with separated data sets in 0.5 M KCl at a pulling velocity of 400 nms^{-1}. All figures show ProtL Kx6E in purple and I27 in grey. Figures A), B) and C) show the data from each data set. The polyprotein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer.
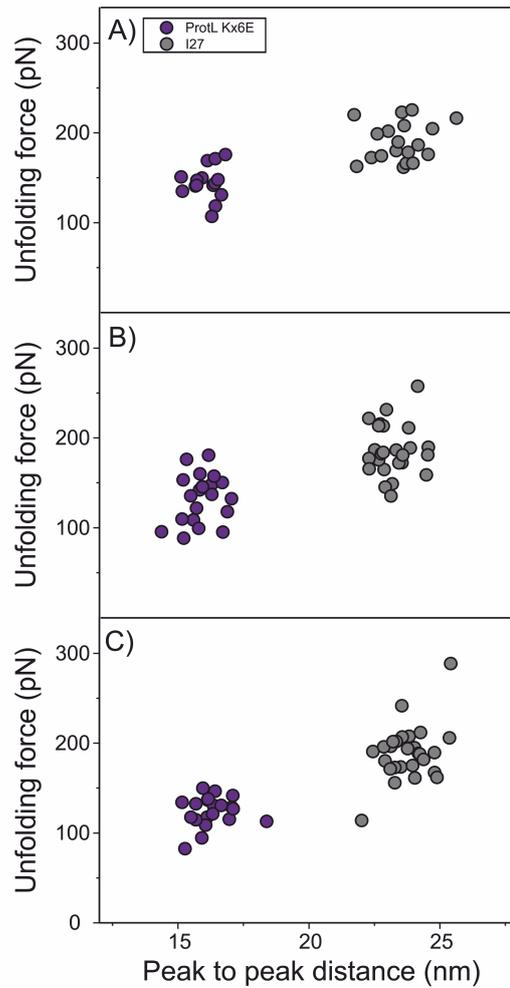
Figure C.18: Scatter diagrams of the unfolding force vs. peak to peak distance for (I27-ProtL Kx6E)$_3$-I27 with separated data sets in 0.5 M KCl at a pulling velocity of 700 nms^{-1}. All figures show ProtL Kx6E in purple and I27 in grey. Figures A), B) and C) show the data from each data set. The polyprotein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer.
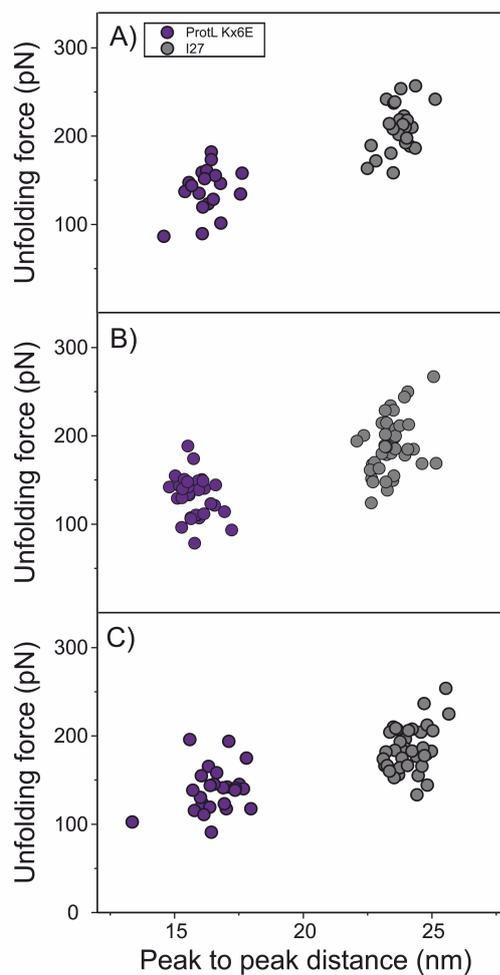
Figure C.19: Scatter diagrams of the unfolding force vs. peak to peak distance for (I27-ProtL Kx6E)$_3$-I27 with separated data sets in 0.5 M KCl at a pulling velocity of 1200 nms^{-1}. All figures show ProtL Kx6E in purple and I27 in grey. Figures A), B) and C) show the data from each data set. The polyprotein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer.
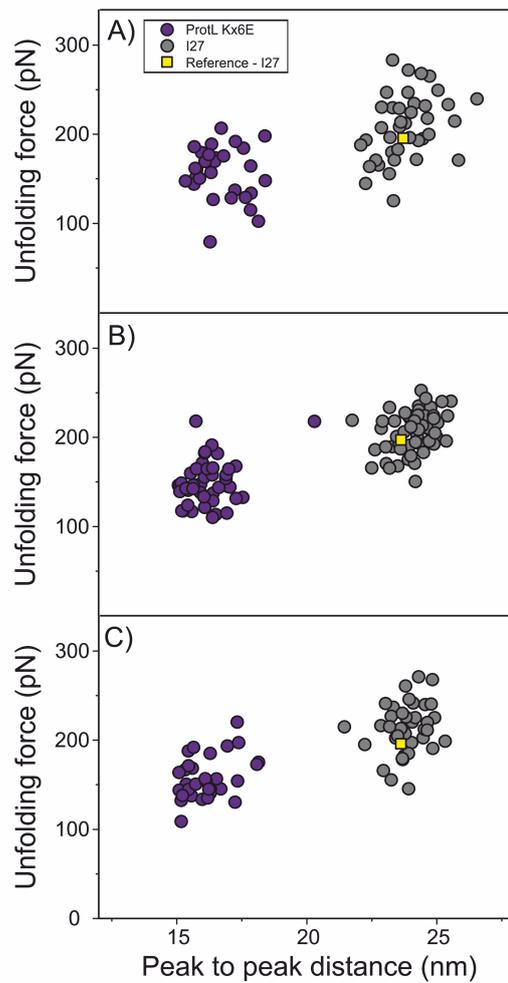
Figure C.20: Scatter diagrams of the unfolding force vs. peak to peak distance for (I27-ProtL Kx6E)$_3$-I27 with separated data sets in 0.5 M KCl at a pulling velocity of 2000 nms^{-1}. All figures show ProtL Kx6E in purple and I27 in grey. Figures A), B) and C) show the data from each data set. The polyprotein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer.
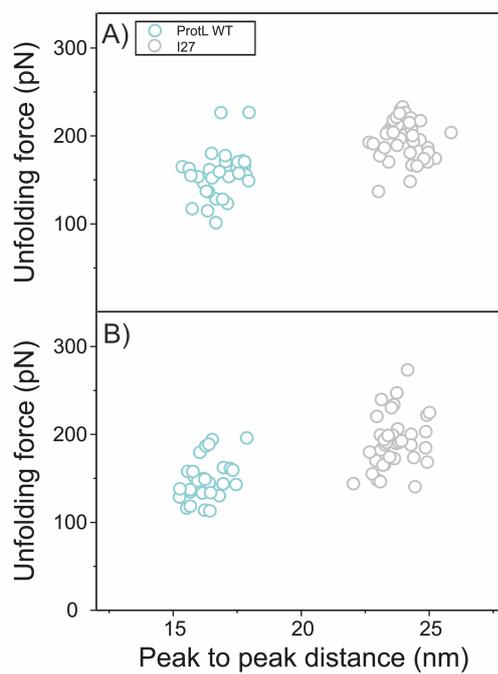
Figure C.21: Scatter diagrams of the unfolding force vs. peak to peak distance for (I27-ProtL WT)$_3$-I27 with separated data sets in 2 M KCl at a pulling velocity of 400 nms^{-1}. All figures show ProtL WT outlined in blue and I27 outlined in light grey. Figures A) and B) show the data from each data set. The polyprotein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer.
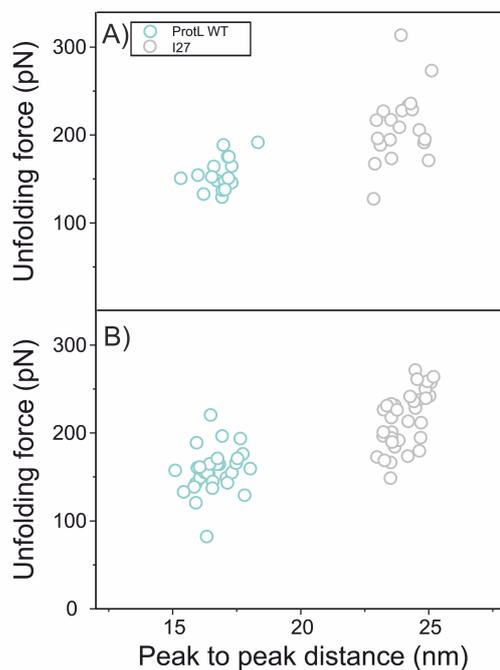
Figure C.22: Scatter diagrams of the unfolding force vs. peak to peak distance for (I27-ProtL WT)$_3$-I27 with separated data sets in 2 M KCl at a pulling velocity of 700 nms^{-1}. All figures show ProtL WT outlined in blue and I27 outlined in light grey. Figures A) and B) show the data from each data set. The polyprotein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer.
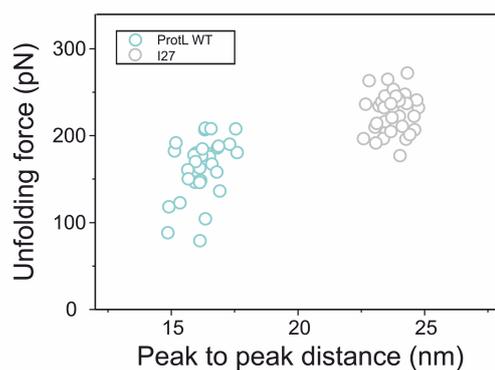
Figure C.23: Scatter diagram of the unfolding force vs. peak to peak distance for (I27-ProtL WT)$_3$-I27 with separated data sets in 2 M KCl at a pulling velocity of 1200 nms^{-1}. All figures show ProtL WT outlined in blue and I27 outlined in light grey. The figure shows the data from one data set. The polyprotein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer.
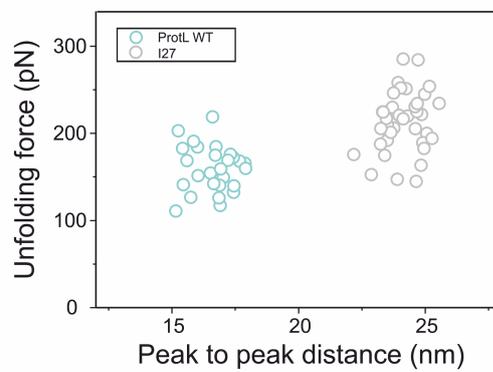
Figure C.24: Scatter diagram of the unfolding force vs. peak to peak distance for (I27-ProtL WT)$_3$-I27 with separated data sets in 2 M KCl at a pulling velocity of 2000 nms^{-1}. All figures show ProtL WT outlined in blue and I27 outlined in light grey. The figure shows the data from one data set. The polyprotein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer.
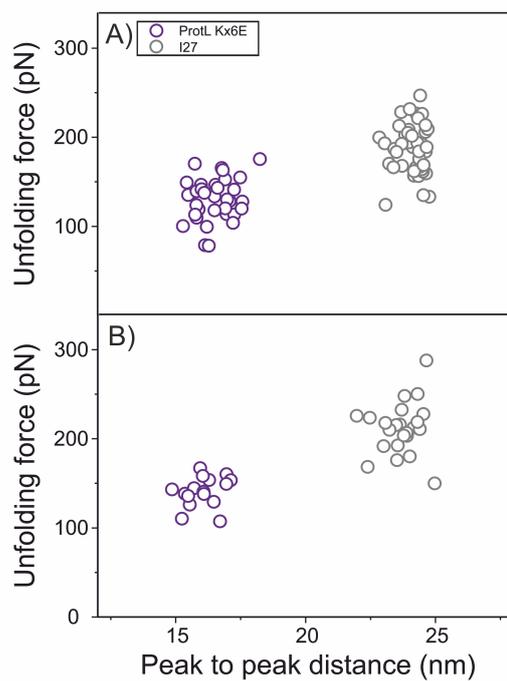
Figure C.25: Scatter diagrams of the unfolding force vs. peak to peak distance for (I27-ProtL Kx6E)$_3$-I27 with separated data sets in 2 M KCl at a pulling velocity of 700 nms^{-1}. All figures show ProtL WT outlined in purple and I27 outlined in grey. Figures A) and B) show the data from each data set. The polyprotein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer.
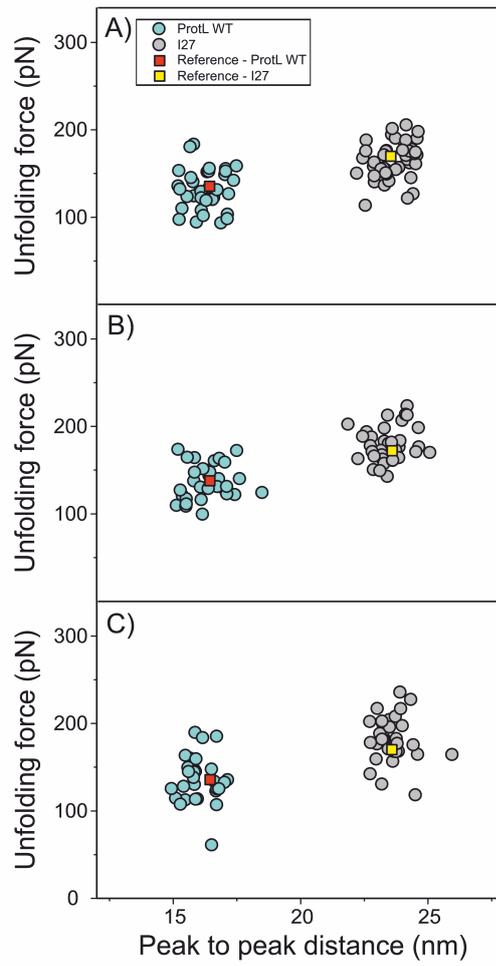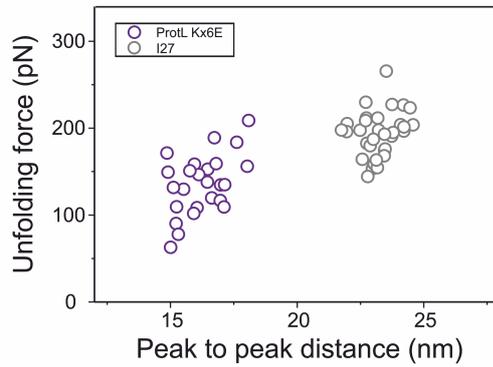
Figure C.26: Scatter diagram of the unfolding force vs. peak to peak distance for (I27-ProtL Kx6E)$_3$-I27 with separated data sets in 2 M KCl at a pulling velocity of 1200 nms^{-1}. All figures show ProtL Kx6E outlined in purple and I27 outlined in grey. The figure shows the data from one data set. The polyprotein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer.
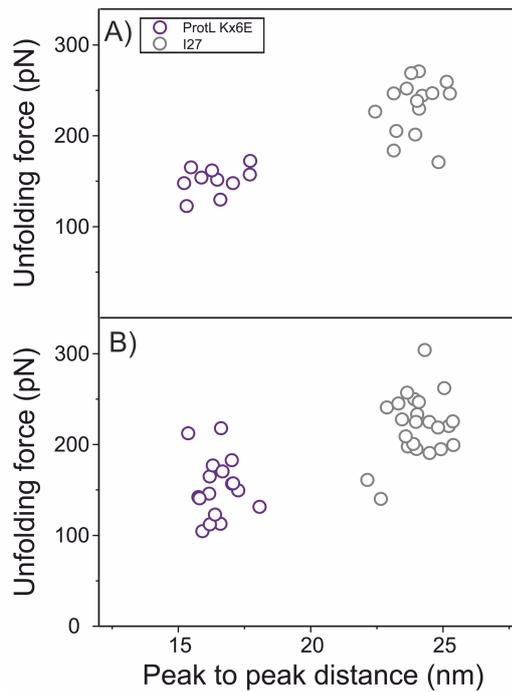
Figure C.27: Scatter diagrams of the unfolding force vs. peak to peak distance for (I27-ProtL Kx6E)$_3$-I27 with separated data sets in 2 M KCl at a pulling velocity of 2000 nms^{-1}. All figures show ProtL WT outlined in purple and I27 outlined in grey. Figures A) and B) show the data from each data set. The polyprotein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer.
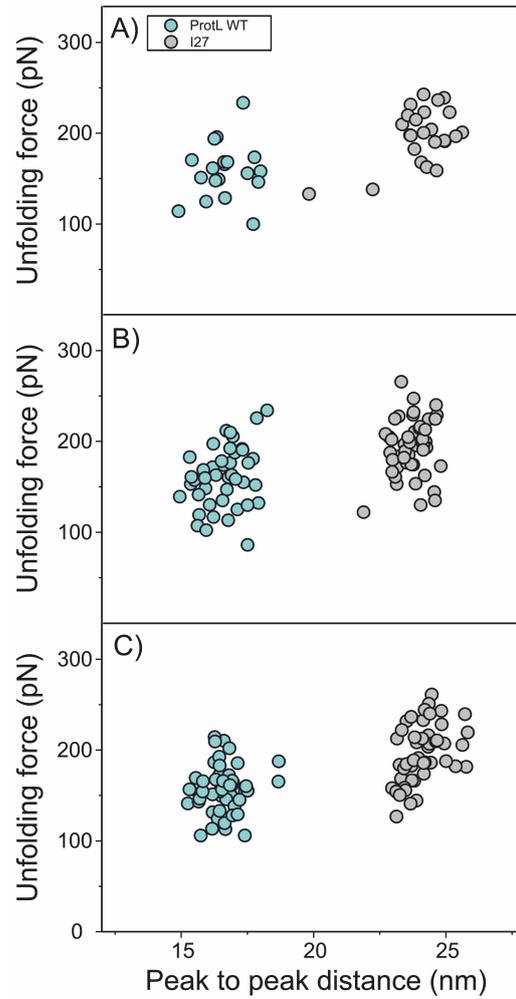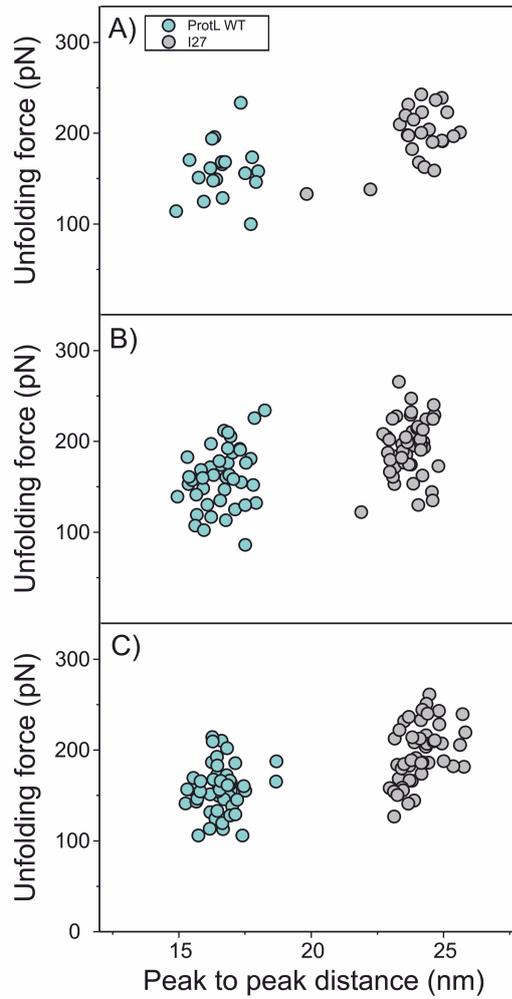
Figure C.28: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL WT)$_3$-I27 with separated data sets at 400 nms^{-1} in 0.5 M KCl. All figures show ProtL WT in blue and I27 in light grey. Figures A), B) and C) show the peak to peak distance data and figures D), E) and F) show the unfolding force data from each data set respectively. All histograms are fitted with a gaussian curve. The polyprotein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL WT and I27.

Figure C.29: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL WT)$_3$-I27 with separated data sets at 700 nms^{-1} in 0.5 M KCl. All figures show ProtL WT in blue and I27 in light grey. Figures A), B) and C) show the peak to peak distance data and figures D), E) and F) show the unfolding force data from each data set respectively. All histograms are fitted with a gaussian curve. The polyprotein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL WT and I27.

Figure C.30: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL WT)$_3$-I27 with separated data sets at 1200 nms^{-1} in 0.5 M KCl. All figures show ProtL WT in blue and I27 in light grey. Figures A), B) and C) show the peak to peak distance data and figures D), E) and F) show the unfolding force data from each data set respectively. All histograms are fitted with a gaussian curve. The polyprotein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL WT and I27.

Figure C.31: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL WT)$_3$-I27 with separated data sets at 2000 nms^{-1} in 0.5 M KCl. All figures show ProtL WT in blue and I27 in light grey. Figures A), B) and C) show the peak to peak distance data and figures D), E) and F) show the unfolding force data from each data set respectively. All histograms are fitted with a gaussian curve. The polyprotein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL WT and I27.
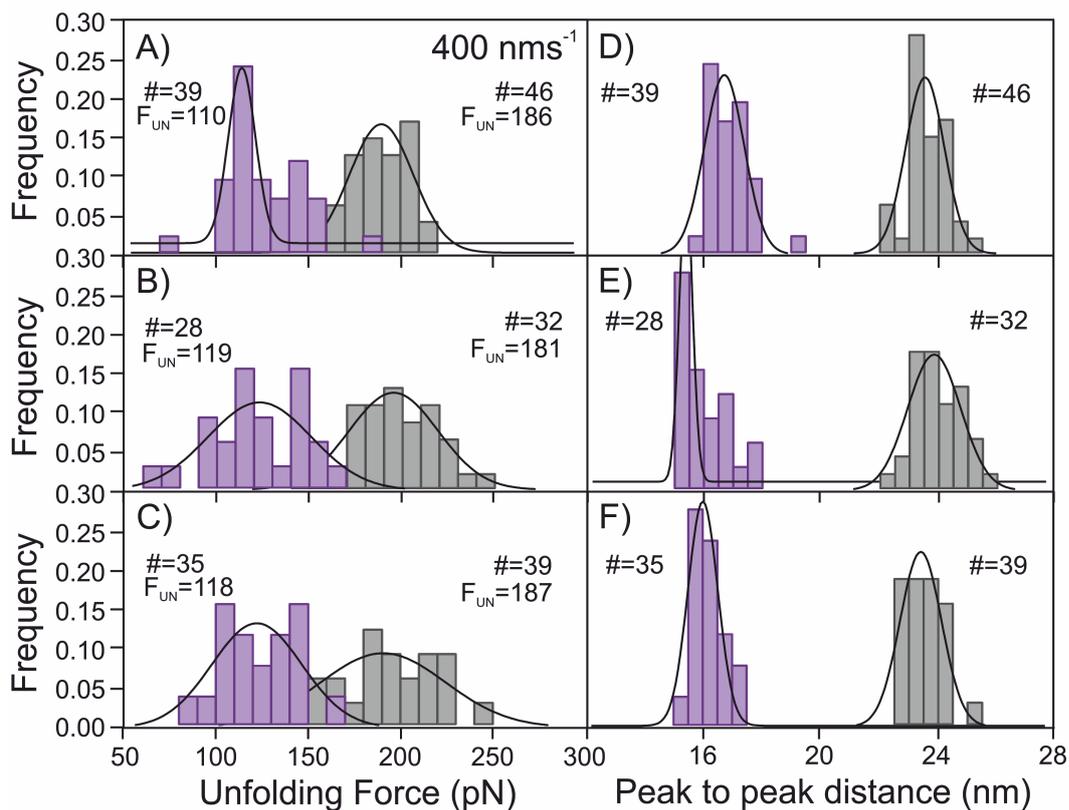
Figure C.32: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL Kx6E)$_3$-I27 with separated data sets at 400 nms^{-1} in 0.5 M KCl. All figures show ProtL Kx6E in purple and I27 in grey. Figures A), B) and C) show the peak to peak distance data and figures D), E) and F) show the unfolding force data from each data set respectively. All histograms are fitted with a gaussian curve. The polyprotein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL Kx6E and I27.
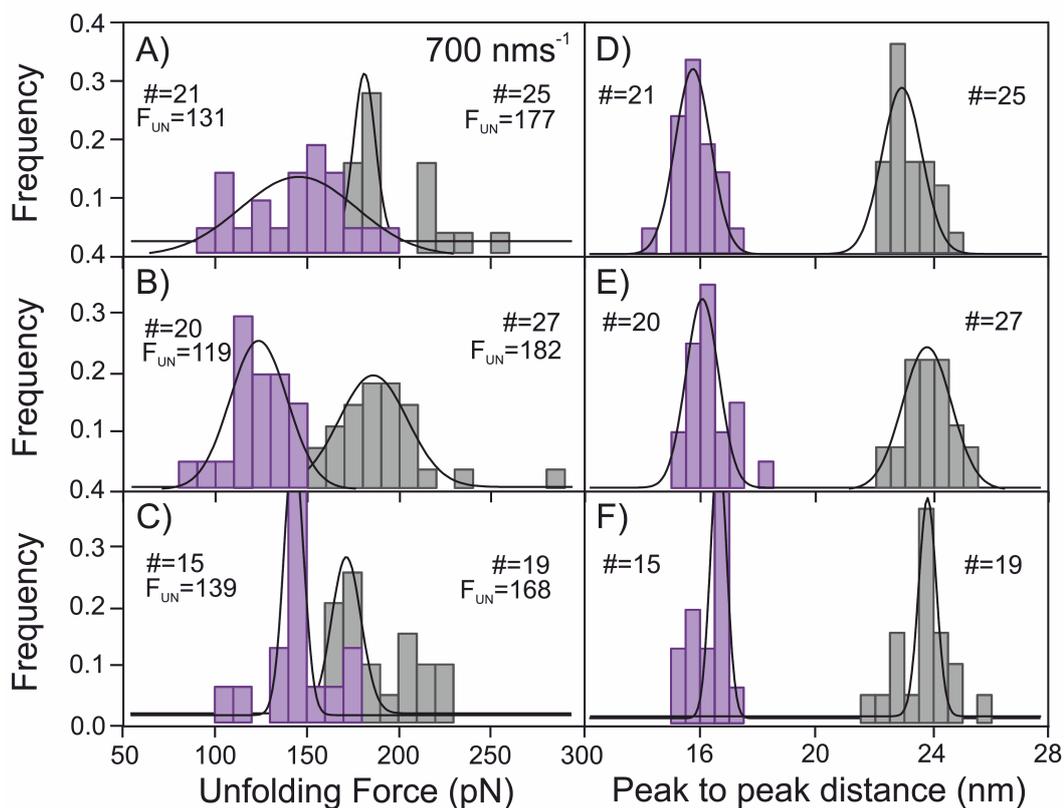
Figure C.33: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL Kx6E)$_3$-I27 with separated data sets at 700 nms^{-1} in 0.5 M KCl. All figures show ProtL Kx6E in purple and I27 in grey. Figures A), B) and C) show the peak to peak distance data and figures D), E) and F) show the unfolding force data from each data set respectively. All histograms are fitted with a gaussian curve. The polyprotein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL Kx6E and I27.
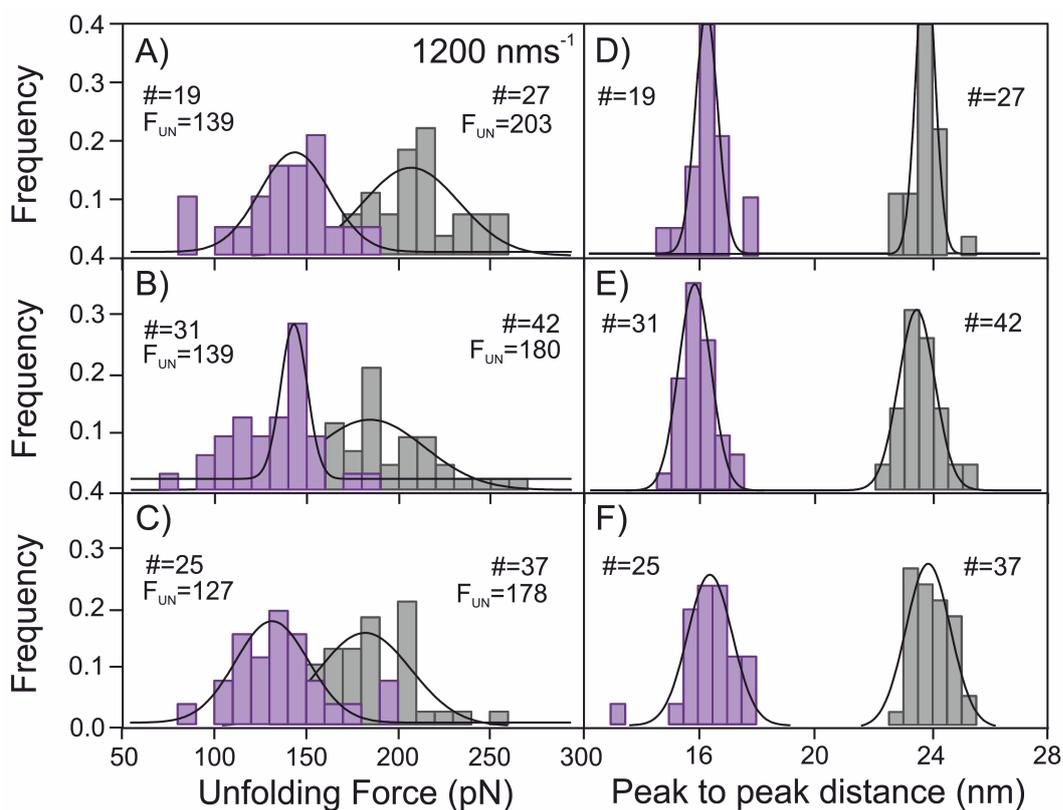
Figure C.34: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL Kx6E)$_3$-I27 with separated data sets at 1200 nms^{-1} in 0.5 M KCl. All figures show ProtL Kx6E in purple and I27 in grey. Figures A), B) and C) show the peak to peak distance data and figures D), E) and F) show the unfolding force data from each data set respectively. All histograms are fitted with a gaussian curve. The polyprotein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL Kx6E and I27.
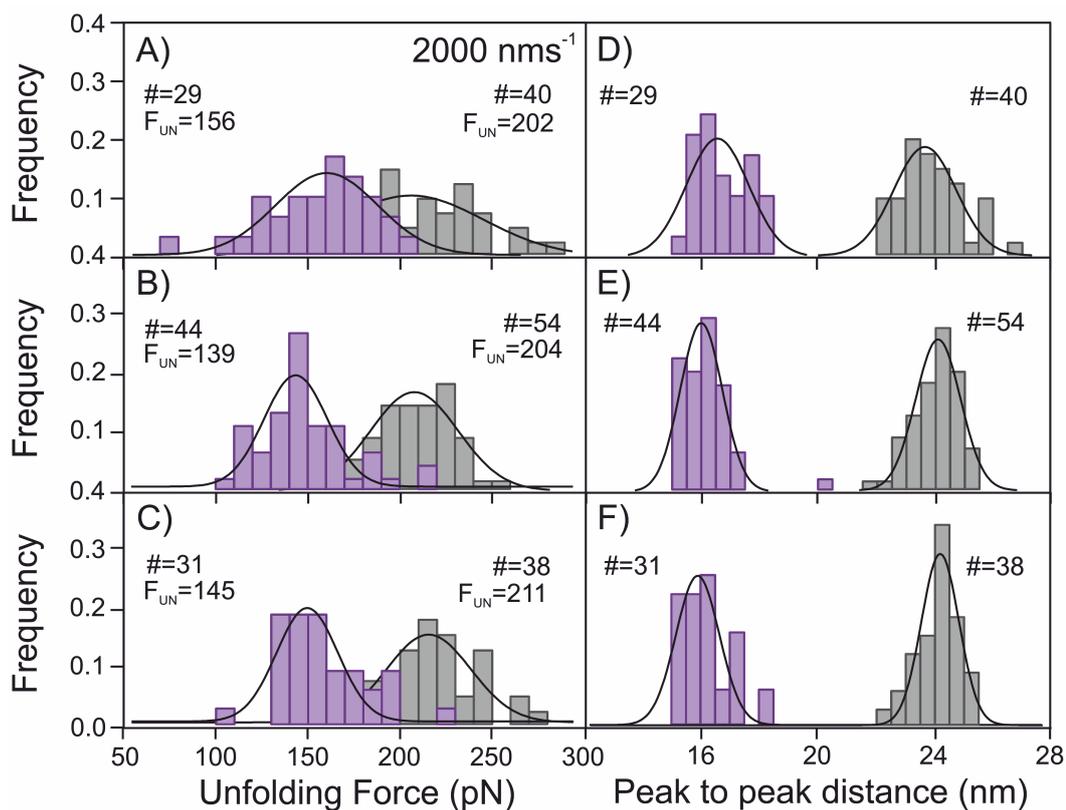
Figure C.35: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL Kx6E)$_3$-I27 with separated data sets at 2000 nms^{-1} in 0.5 M KCl. All figures show ProtL Kx6E in purple and I27 in grey. Figures A), B) and C) show the peak to peak distance data and figures D), E) and F) show the unfolding force data from each data set respectively. All histograms are fitted with a gaussian curve. The polyprotein was in 0.5 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL Kx6E and I27.
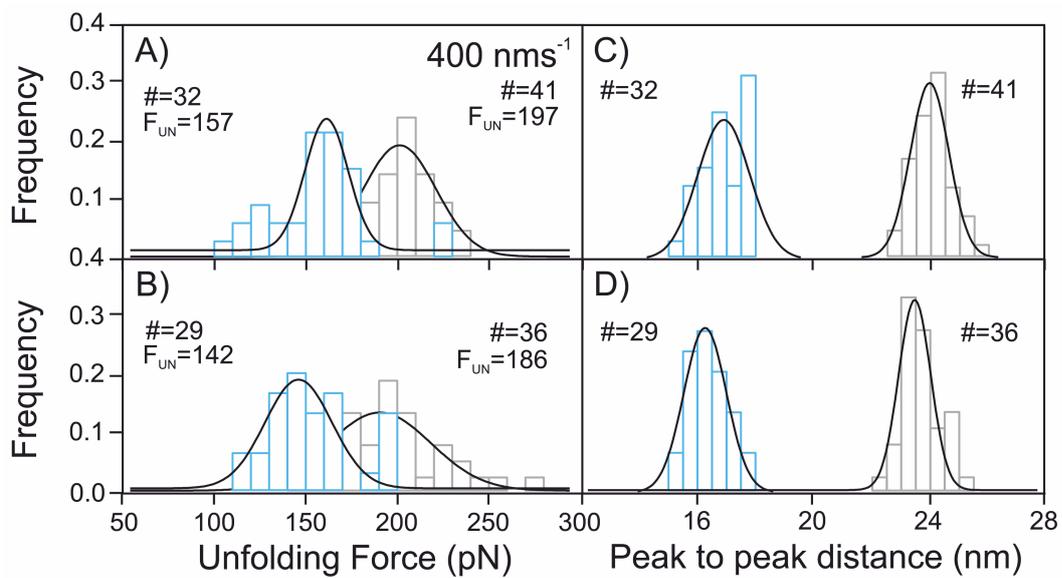
Figure C.36: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL WT)$_3$-I27 with separated data sets at 400 nms^{-1} in 2 M KCl. All figures show ProtL WT outlined in blue and I27 outlined in light grey. Figures A), B) and C) show the peak to peak distance data and figures D), E) and F) show the unfolding force data from each data set respectively. All histograms are fitted with a gaussian curve. The polyprotein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL WT and I27.
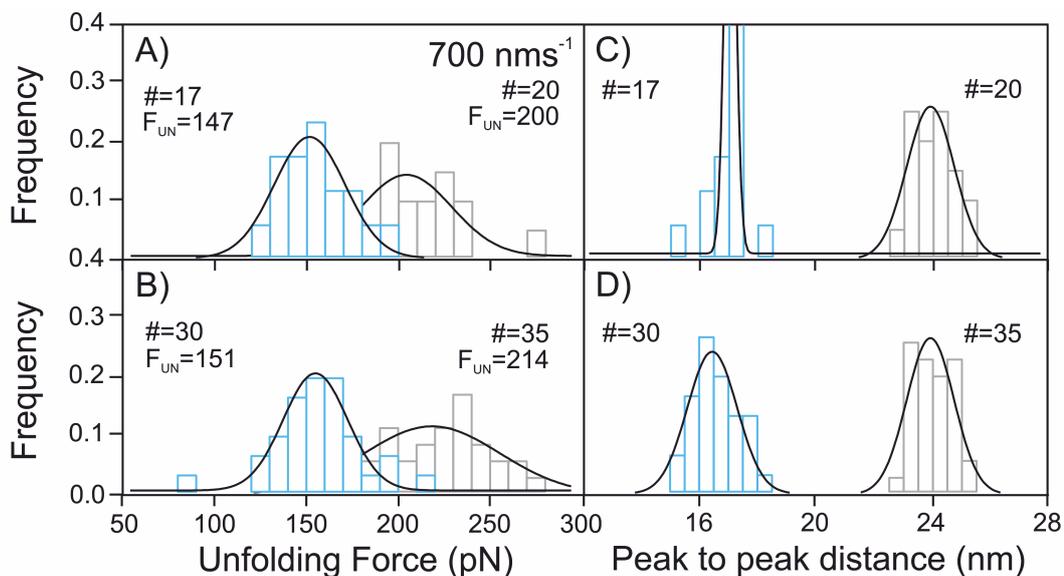
Figure C.37: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL WT)$_3$-I27 with separated data sets at 700 nms^{-1} in 2 M KCl. All figures show ProtL WT outlined in blue and I27 outlined in light grey. Figures A), B) and C) show the peak to peak distance data and figures D), E) and F) show the unfolding force data from each data set respectively. All histograms are fitted with a gaussian curve. The polyprotein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL WT and I27.
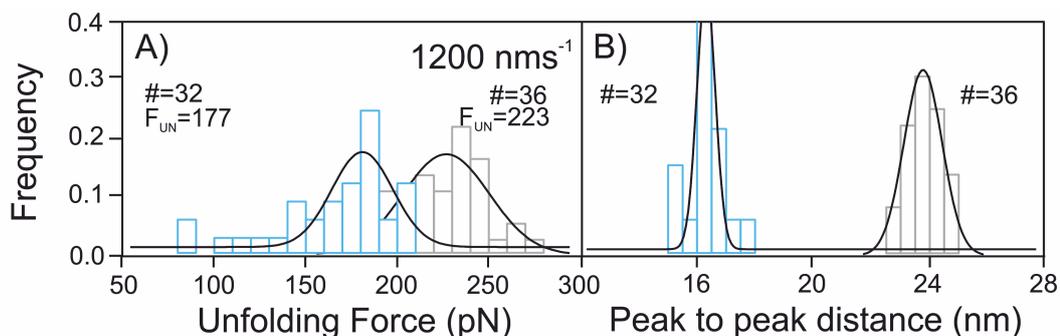
Figure C.38: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL WT)$_3$-I27 with separated data sets at 1200 nms^{-1} in 2 M KCl. All figures show ProtL WT outlined in blue and I27 outlined in light grey. Figures A), B) and C) show the peak to peak distance data and figures D), E) and F) show the unfolding force data from each data set respectively. All histograms are fitted with a gaussian curve. The polyprotein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL WT and I27.
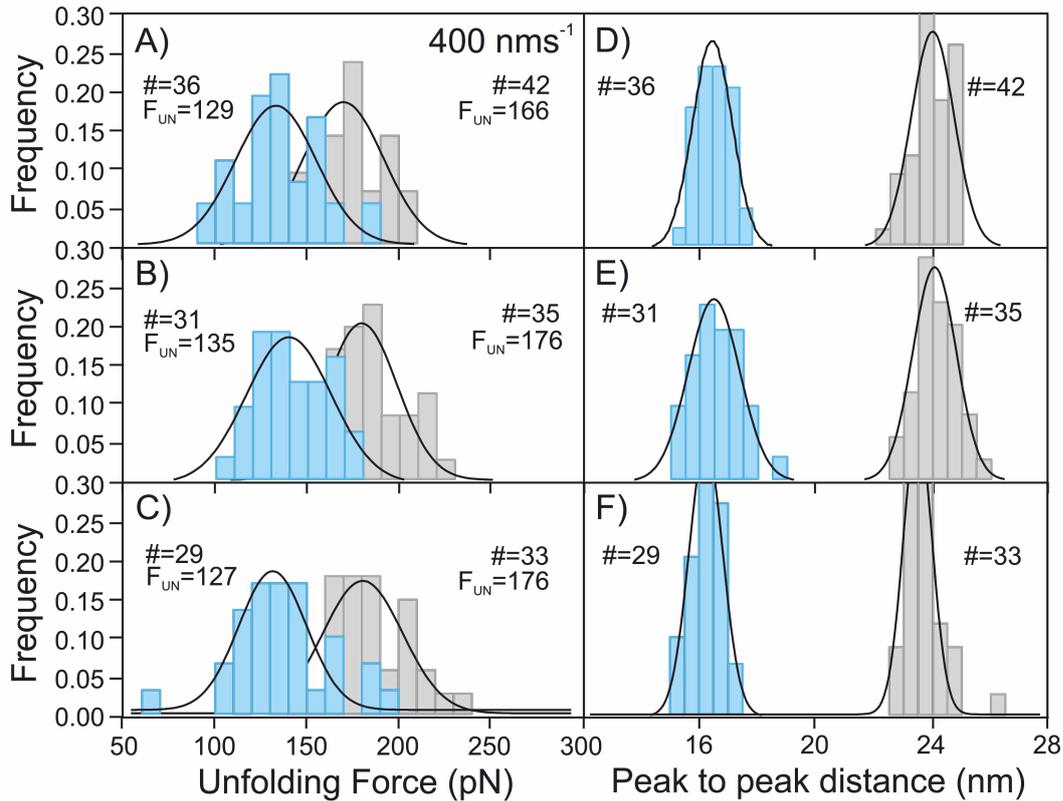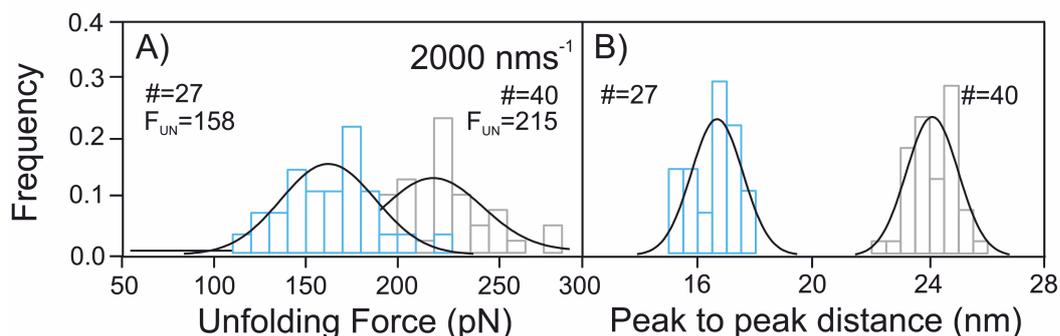
Figure C.39: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL WT)$_3$-I27 with separated data sets at 2000 nms^{-1} in 2 M KCl. All figures show ProtL WT outlined in blue and I27 outlined in light grey. Figures A), B) and C) show the peak to peak distance data and figures D), E) and F) show the unfolding force data from each data set respectively. All histograms are fitted with a gaussian curve. The polyprotein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL WT and I27.
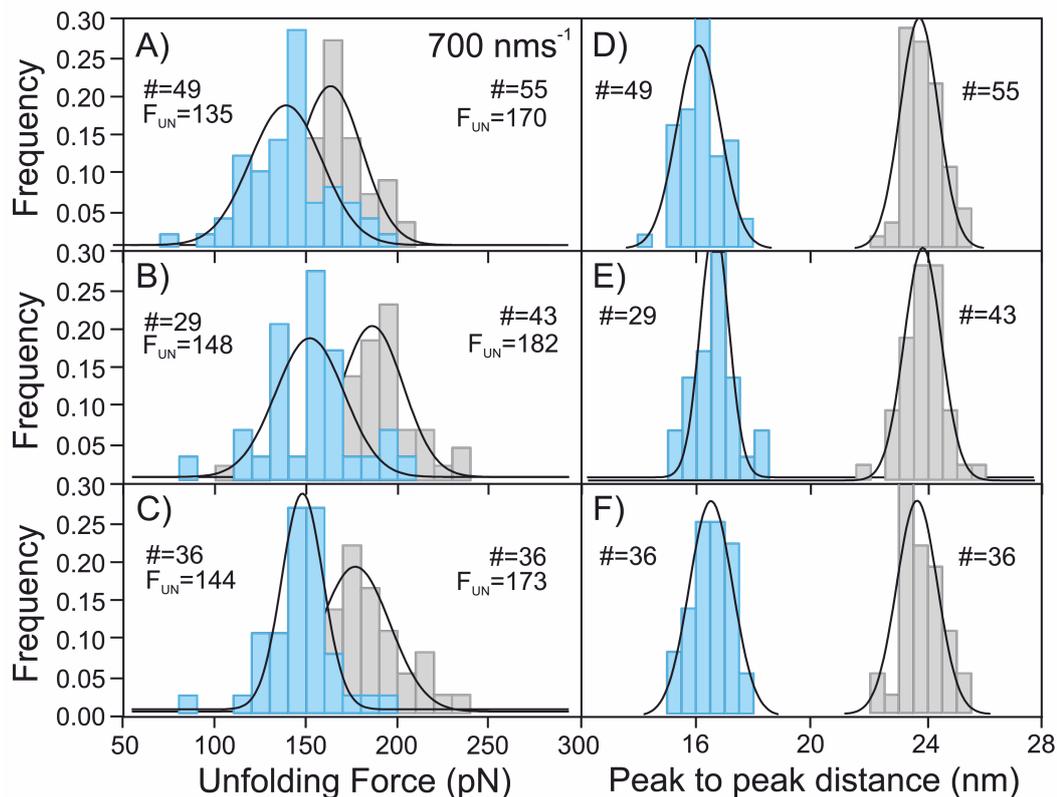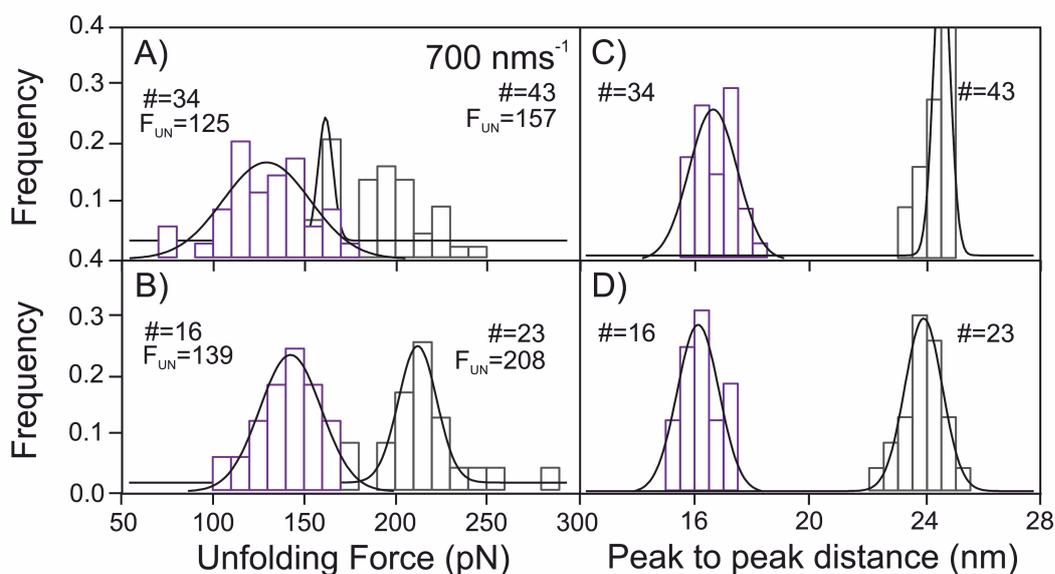

Figure C.40: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL Kx6E)$_3$-I27 with separated data sets at 700 nms^{-1} in 2 M KCl. All figures show ProtL Kx6E outlined in purple and I27 outlined in grey. Figures A), B) and C) show the peak to peak distance data and figures D), E) and F) show the unfolding force data from each data set respectively. All histograms are fitted with a gaussian curve. The polyprotein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL Kx6E and I27.
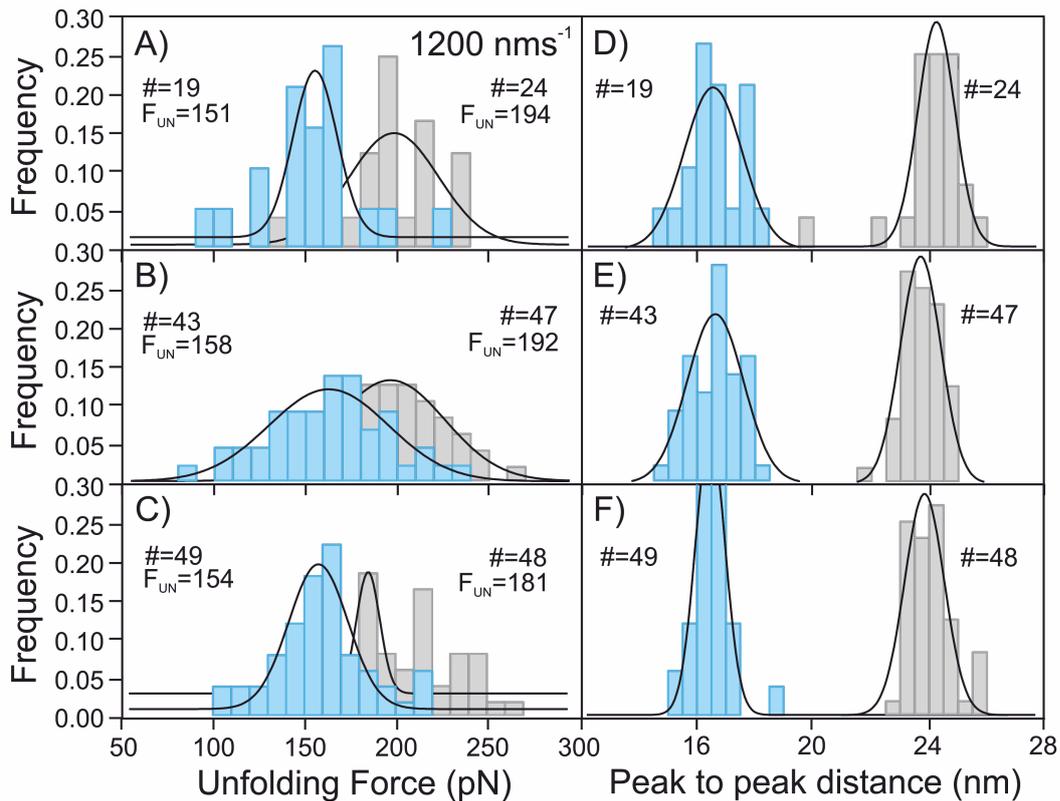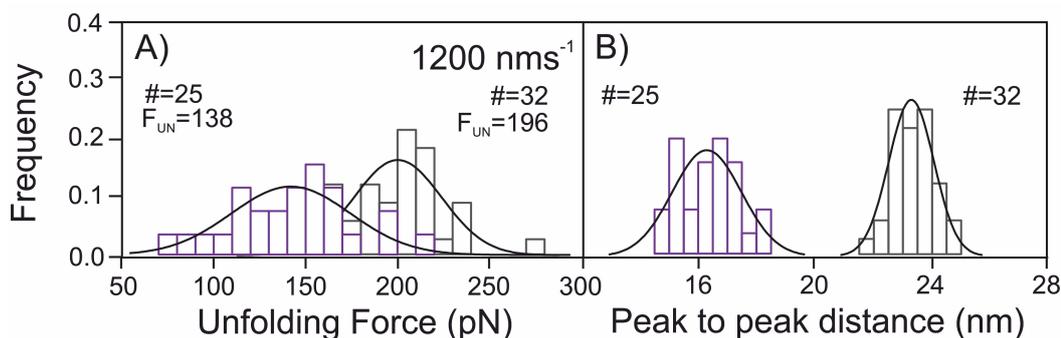
Figure C.41: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL Kx6E)$_3$-I27 with separated data sets at 1200 nms^{-1} in 2 M KCl. All figures show ProtL Kx6E outlined in purple and I27 outliined in grey. Figures A), B) and C) show the peak to peak distance data and figures D), E) and F) show the unfolding force data from each data set respectively. All histograms are fitted with a gaussian curve. The polyprotein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL Kx6E and I27.
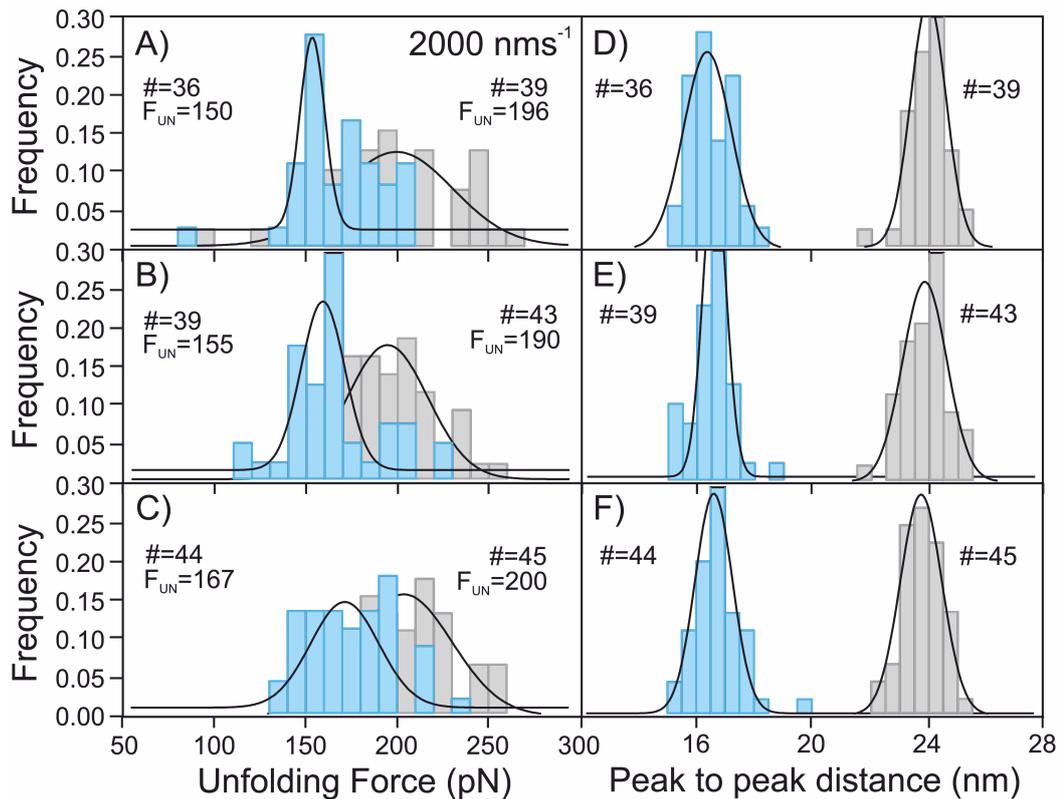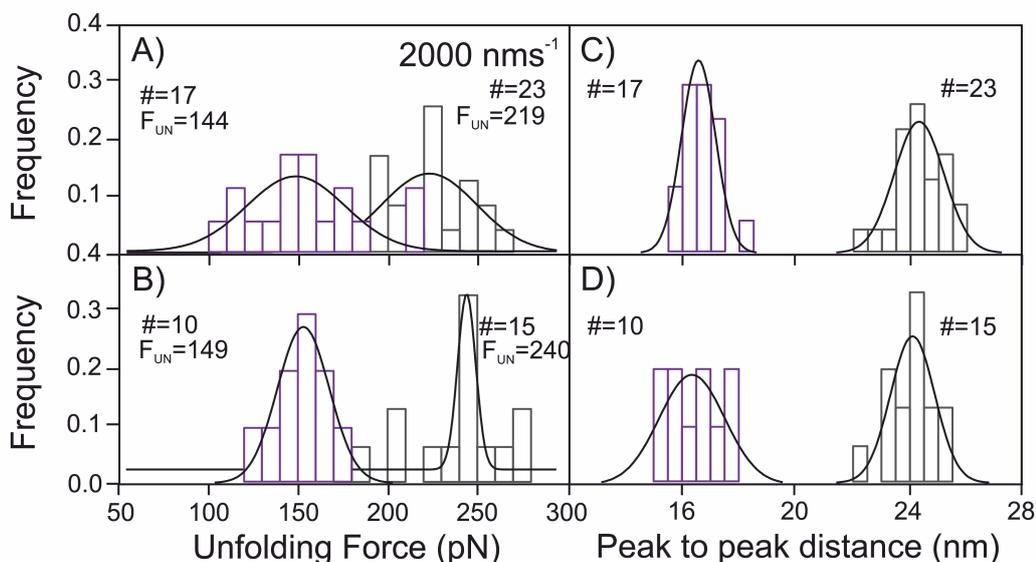

Figure C.42: Histograms of the unfolding force data and peak to peak distance data for (I27-ProtL Kx6E)$_3$-I27 with separated data sets at 2000 nms^{-1} in 2 M KCl. All figures show ProtL Kx6E outlined in purple and I27 outlined in grey. Figures A), B) and C) show the peak to peak distance data and figures D), E) and F) show the unfolding force data from each data set respectively. All histograms are fitted with a gaussian curve. The polyprotein was in 2 M KCl, 5 mM MOPS, pH 7.4 buffer. The number of unfolding events and modal unfolding force of the gaussian fit are displayed on each histogram for ProtL Kx6E and I27.
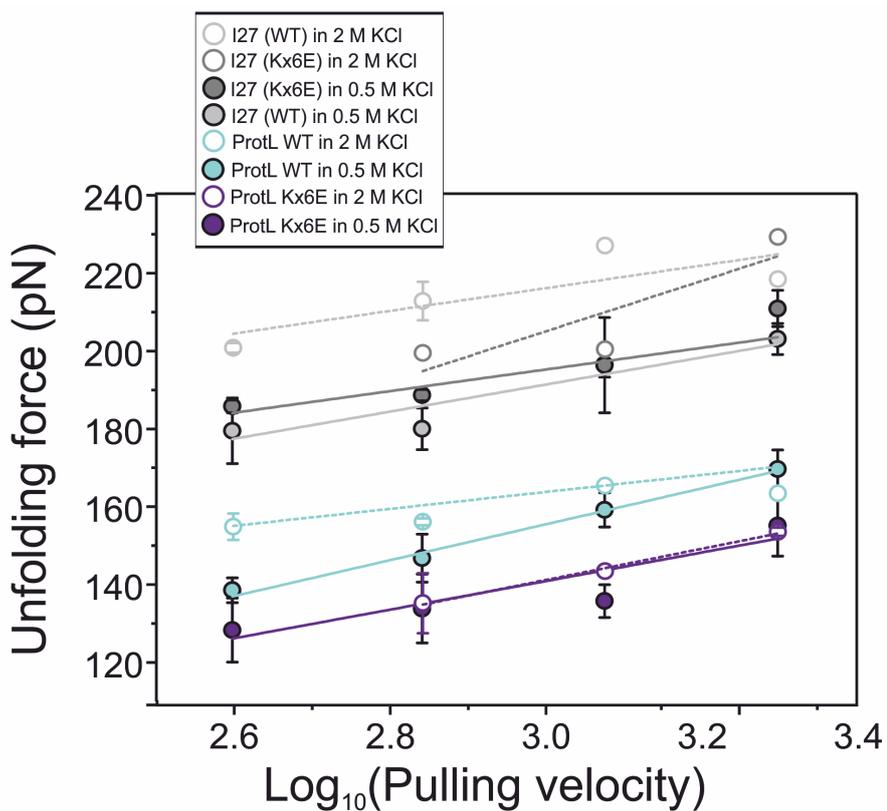
Figure C.43: Speed dependence diagram of the unfolding force vs. $\log_{10}$(pulling velocity) for the refined analysis of all 4 systems in one. The figure shows the speed dependence of (I27-ProtL WT)$_3$-I27 and (I27-ProtL Kx6E)$_3$-I27 in 0.5 M and 2 M KCl. The linear fit for each set of data is also plotted as full lines in 0.5 M KCl and dashed lines in 2 M KCl.

Appendix D

Supplementary information for Results 3

The data from the 3 experiments was then pooled together to see if there was an effect in the final values. The unfolding force and the peak to peak distances are shown in tablesD.1 and D.2 respectively.

Protein	No. events	Mean F_{UN} (pN)	Median F_{UN} (pN)	SD
ProtL WT	101	133	133	27

Table D.1: Pooled unfolding force data and standard deviation for poly-ProtL WT at 400 nms^{-1} in PBS buffer.

Protein	No. events	Mean p2p (nm)	Median p2p (nm)	SD
ProtL WT	101	16.4	16.3	1.0

Table D.2: Pooled peak to peak distance data and standard deviation for poly-ProtL WT at 400 nms^{-1} in PBS buffer.

The data was again plotted in a scatter diagram showing the unfolding force for each protein unfolding event against the measured peak to peak distance for each experiment(figure D.1.

A histogram was then produced for the unfolding force data and the peak to peak distances for the pooled data. Figure D.2 shows both histograms can be fitted correctly with a gaussian curve.
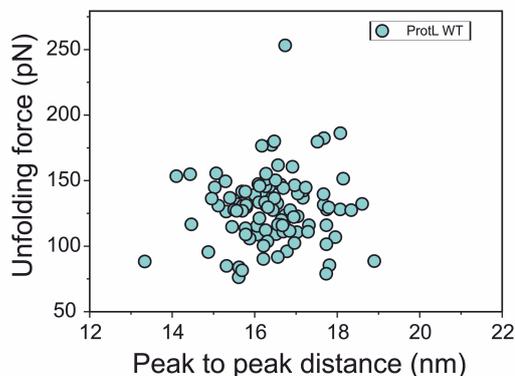
Figure D.1: Scatter diagram of the pooled unfolding force vs. peak to peak distance data for poly-ProtL WT The figure, where ProtL WT is blue, shows a clear cluster for the combined ProtL WT data.
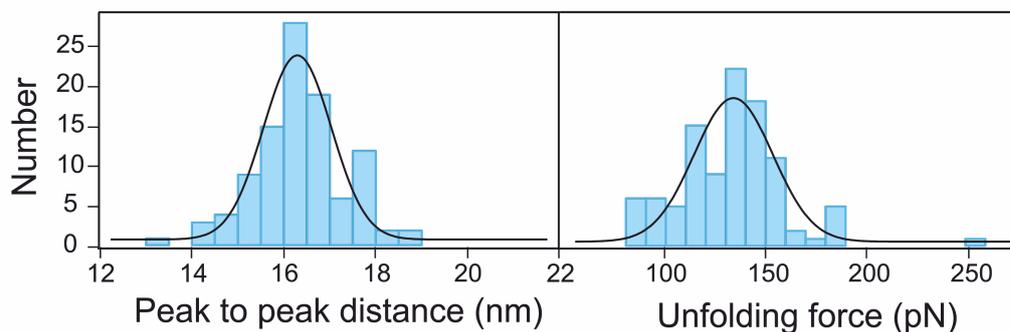
Figure D.2: Histograms of the pooled unfolding force data and peak to peak distance data for poly-ProtL WT. Both figures show ProtL WT in blue. Both histograms are fitted with a gaussian curve. The poly-protein was in 63 mM sodium phosphate, pulled at a rate of 400 nms$_{-1}$.
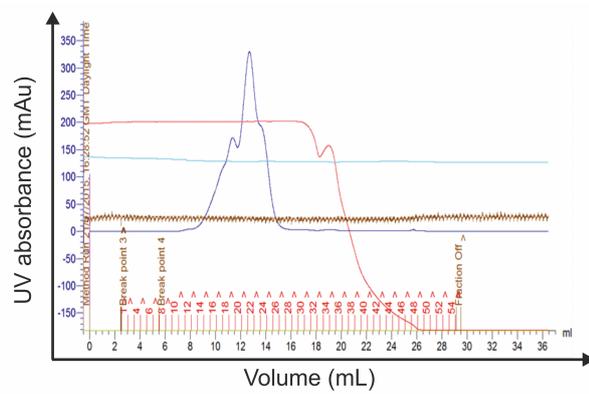
Figure D.3: Elution profile of the purification of poly-ProtL Ex6D on a size exclusion column. The figure shows the absorbance in blue and fraction number in red.

Bibliography

[1] L.J. Rothschild and R.L. Mancinelli. Life in extreme environments. *Nature*, 409: 1092–1101, 2001. doi: 10.1038/35059215. URL `http://www.nature.com/nature/journal/v409/n6823/pdf/4091092a0.pdf`.

[2] A. Nissenbaum. The Microbiology and Biogeochemistry of the Dead Sea. *Microb. Ecol.*, 2:139–161, 1975. doi: 10.1007/BF02010435. URL `http://link.springer.com/article/10.1007/BF02010435`.

[3] General structural characteristics of proteins. In H.S. Stoker, editor, *Organic and Biological Chemistry*, page 371. Brooks/Cole, Cengage Learning, California, seventh edition, 2015.

[4] J.C. Kendrew, G. Bodo, H.M. Dintzis, R.G. Parrish, H. Wyckoff, and D.C. Phillips. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610):662–666, 1958. doi: 10.1038/181662a0. URL `http://www.nature.com/nature/journal/v181/n4610/pdf/181662a0.pdf`.

[5] M.F. Perutz, M.G. Rossmann, A.F. Cullis, H. Muirhead, G. Will, and A.C. North. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. *Nature*, 185(4711):416–422, 1960. doi: 10.1038/185416a0. URL `http://www.nature.com/nature/journal/v185/n4711/pdf/185416a0.pdf`.

[6] 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):535–42, 1977. doi: 10.1016/j.jmb.2014.12.021. URL `http://www.ncbi.nlm.nih.gov/pubmed/875032`.

[7] J.W. O'Neill, D.E. Kim, D. Baker, and K.Y.J. Zhang. Structures of the B1 domain of protein L from *Peptostreptococcus magnus* with a tyrosine to tryptophan substitution. *Acta Crystallographica Section D*, 57(4):480–487, 2001. doi: 10.1107/S0907444901000373. URL `http://dx.doi.org/10.1107/S0907444901000373`.

[8] 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. URL `http://www.sciencedirect.com/science/article/pii/S0969212696000366`.

[9] C.B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181 (96):223–230, 1973. doi: 10.1126/science.181.4096.223. URL `http://web.iitd.ac.in/~amittal/1973_Anfinsen_Science.pdf`.

[10] K.A. Dill. Dominant Forces in Protein Folding. *Biochemistry*, 29(31):7133–7155, 1990. doi: 10.1021/bi00483a001. URL `http://pubs.acs.org/doi/pdf/10.1021/bi00483a001`.

[11] C.N. Pace and B.A. Shirley. Forces contributing to the conformational stability of proteins. *FASEB J.*, 10:75–83, 1996. URL `http://0-www.fasebj.org.wam.leeds.ac.uk/content/10/1/75.full.pdf+html`.

[12] M. Cohen, V. Potapov, and G. Schreiber. Four Distances between Pairs of Amino Acids Provide a Precise Description of their Interaction. *PLoS Comput. Biol.*, 5(8), 2009. doi: 10.1371/journal.pcbi.1000470. URL `http://dx.plos.org/10.1371/journal.pcbi.1000470`.

[13] J. Israelachvili and R. Pashley. The hydrophobic interaction is long range, decaying exponentially with distance. *Nature*, 300(5890):341–342, 1982. doi: 10.1038/300341a0. URL `http://www.nature.com/nature/journal/v300/n5890/pdf/300341a0.pdf`.

[14] S.K. Burley and G.A. Petsko. Weakly Polar Interactions In Proteins. *Adv. Protein Chem.*, 39:125–189, 1988. doi: 10.1016/S0065-3233(08)60376-9. URL `http://www.sciencedirect.com/science/article/pii/S0065323308603769`.

[15] D.F. Stickle, L.G. Presta, K.A. Dill, and G. Rose. Hydrogen bonding in globular proteins. *J. Mol. Biol.*, 226:1143–1159, 1992. doi: 10.1016/0022-2836(92)91058-W. URL `http://www.sciencedirect.com/science/article/pii/002228369291058W`.

[16] Berg J.M., Tymoczko J.L., and Stryer L. *Biochemistry*. W H Freeman, New York, fifth edition, 2002. URL `http://www.ncbi.nlm.nih.gov/books/NBK21154/`.

[17] J.N. Israelachvili. *Intermolecular and Surface Forces*. Academic Press, London, 2010. doi: 10.1002/bbpc.19860901226. URL `http://dx.doi.org/10.1002/bbpc.19860901226`.

[18] J. Seras-Franzoso, R. Affentranger, M. Ferrer-Navarro, X. Daura, A. Villaverde, and E. García-Fruitósa. Disulfide bond formation and activation of Escherichia coli β-galactosidase under oxidizing conditions. *Appl. Environ. Microbiol.*, 78:2376–2385, 2012. doi: 10.1128/AEM.06923-11. URL `http://aem.asm.org/content/78/7/2376.full.pdf+html`.

[19] S. Basu, D. Bhattacharyya, and R. Banerjee. Self-complementarity within proteins: bridging the gap between binding and folding. *Biophys. J.*, 102(11):2605–14, 2012. doi: 10.1016/j.bpj.2012.04.029. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3368132&tool=pmcentrez&rendertype=abstract`.

[20] A.B.P Lever. Lewis Structures and the Octet Rule. *J. Chem. Educ.*, 49(12):819–21, 1972. doi: 10.1021/ed049p819. URL `http://pubs.acs.org/doi/pdf/10.1021/ed049p819`.

[21] P.W. Atkins and P.W. J. de Paula. *Atkins' Physical Chemistry*. Oxford Univeristy Press, Oxford, eighth edition, 2006.

[22] J. Clayden, N. Greeves, Warren S., and P. Wothers. *Organic Chemisrty*. Oxford Universuty Press, Oxford, 2001.

[23] G.I. Makhatadze, V.V. Loladze, D.N. Ermolenko, X. Chen, and S.T. Thomas. Contribution of Surface Salt Bridges to Protein Stability: Guidelines for Protein Engineering. *J. Mol. Biol.*, 327(5):1135–1148, 2003. doi: 10.1016/S0022-2836(03)00233-X. URL `http://linkinghub.elsevier.com/retrieve/pii/S002228360300233X`.

[24] D. Perl, U. Mueller, U. Heinemann, and F.X. Schmid. Two exposed amino acid residues confer thermostability on a cold shock protein. *Nature Struct. Biol.*, 7 (5):380–383, 2000. doi: 10.1038/75151. URL `http://www.nature.com/nsmb/journal/v7/n5/pdf/nsb0500_380.pdf`.

[25] S. Kumar, C. Tsai, and R. Nussinov. Factors enhancing protein thermostability. *Protein Eng.*, 13(3):179–191, 2000. doi: 10.1093/protein/13.3.179. URL `http://peds.oxfordjournals.org/content/13/3/179.full.pdf+html`.

[26] S. Lindman, W. Xue, O. Szczepankiewicz, M.C. Bauer, H. Nilsson, and S. Linse. Salting the charged surface: pH and salt dependence of protein G B1 stability. *Biophys. J.*, 90(8):2911–2921, 2006. doi: 10.1529/biophysj.105.071050. URL `http://www.cell.com/biophysj/abstract/S0006-3495(06)72471-7`.

[27] D. Bordo and P. Argos. The Role of Side-chain Hydrogen Bonds in the Formation and Stabilization of Secondary Structure in Soluble Proteins. *J. Mol. Biol.*, 243(3): 504–519, 1994. doi: 10.1006/jmbi.1994.1676. URL `http://www.sciencedirect.com/science/article/pii/S0022283684716767`.

[28] S. Kumar, C.J. Tsai, B. Ma, and R. Nussinov. Contribution of salt bridges toward protein thermostability. *J. Biomol. Struct. Dyn.*, 17(Special Issue: Proceedings of the Eleventh Conversation):79–85, 2000. doi: 10.1080/07391102.2000.10506606. URL `http://www.tandfonline.com/doi/pdf/10.1080/07391102.2000.10506606`.

[29] K.M. Tych, M.R. Batchelor, T. Hoffmann, C.M. Wilson, E. Paci, D.J. Brockwell, and L. Dougan. Tuning protein mechanics through an ionic cluster graft from an extremophilic protein. *Soft Matter*, page to appear, 2016. doi: 10.1039/c1sm06330h. URL `http://pubs.rsc.org/en/content/articlepdf/2014/SM/C5SM02938D?page=search`.

[30] R.L Baldwin. The new view of hydrophobic free energy. *FEBS Lett.*, 587(8):1062–6, 2013. doi: 10.1016/j.febslet.2013.01.006. URL `http://www.febsletters.org/article/S0014579313000203/fulltext`.

[31] W. Kauzmann. Advances in Protein Chemistry. In C.B. Anfinsen, Anson M.L., K. Bailey, and Edsall J.T., editors, *Advances in Protein Chemistry*, volume 14, pages 1–63. Elsevier, 1959. doi: 10.1016/S0065-3233(08)60608-7. URL `http://www.sciencedirect.com/science/article/pii/S0065323308606087`.

[32] M.L. Anson. Protein Denaturation and the Properties of Protein Groups. *Adv. Protein Chem.*, 2:361–386, 1945. doi: 10.1016/S0065-3233(08)60629-4. URL `http://www.sciencedirect.com/science/article/pii/S0065323308606294`.

[33] D. Baker and D.A. Agard. Kinetics versus thermodynamics in protein folding. *Biochemistry*, 33(24):7505–7509, 1994. doi: 10.1021/bi00190a002. URL `http://pubs.acs.org/doi/pdf/10.1021/bi00190a002`.

[34] Andrew D Robertson and Kenneth P Murphy. Protein Structure and the Energetics of Protein Stability. *Chem. Rev.*, 97(5):1251–1268, 1997. doi: 10.1021/cr960383c. URL `http://dx.doi.org/10.1021/cr960383c`.

[35] A.V. Finkelstein and O.B. Ptitsyn. Lecture 5. In *Protein Physics*, pages 43–55. Elsevier, 2002. doi: 10.1016/B978-012256781-0/50007-X. URL `http://www.sciencedirect.com/science/article/pii/B978012256781050007X`.

[36] P.E. Leopold, M. Montal, and J.N. Onuchic. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA*, 89(18): 8721–8725, 1992. doi: 10.1073/pnas.89.18.8721. URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC49992/$\delimiter"026E30F$nhttp://www.ncbi.nlm.nih.gov.gate1.inist.fr/pmc/articles/PMC49992/pdf/pnas01092-0322.pdf`.

[37] K.A. Dill and H.S. Chan. From Levinthal to pathways to funnels. *Nature Struct. Biol.*, 4(1):10–19, 1997. doi: 10.1038/nsb0197-10. URL `http://faculty.uml.edu/vbarsegov/teaching/bioinformatics/papers/dill.pdf`.

[38] M. Manning and W. Colón. Structural basis of protein kinetic stability: Resistance to sodium dodecyl sulfate suggests a central role for rigidity and a bias toward β-sheet structure. *Biochemistry*, 43(35):11248–11254, 2004. doi: 10.1021/bi0491898. URL `http://pubs.acs.org/doi/pdf/10.1021/bi0491898`.

[39] M. Machius, N. Declerck, R. Huber, and G. Wiegand. Kinetic Stabilization of Bacillus licheniformis - Amylase through Introduction of Hydrophobic Residues at the Surface. *J. Biol. Chem.*, 278(13):11546–11553, 2003. doi: 10.1074/jbc.M212618200. URL `http://www.jbc.org/cgi/doi/10.1074/jbc.M212618200`.

[40] J. Mansfeld, G. Vrient, B.W. Dijkstra, O.R. Veltman, B. Van den Burg, G. Venema, R. Ulbrich-Hofmann, and V.G.H. Eijsink. Extreme Stabilization of a Thermolysin-like Protease by an Engineered Disulfide Bond. *J. Biol. Chem.*, 272(17):11152–11156, 1997. doi: 10.1074/jbc.272.17.11152. URL `http://www.jbc.org/content/272/17/11152`.

[41] S. Solis-Mendiola, L.H. Gutierrez-Gonzalez, A. Arroyo-Reyna, J. Padilla-Zuniga, A. Rojo-Dominguez, and A. Hernandez-Arana. pH dependence of the activation parameters for chymopapain unfolding: influence of ion pairs on the kinetic stability of proteins. *Biochim. Biophys. Acta*, 1388(2):363–372, 1998. doi: S0167-4838(98)00195-2[pii]. URL `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9858767`.

[42] I. Pozdnyakova, J. Guidry, and P. Wittung-Stafshede. Copper stabilizes azurin by decreasing the unfolding rate. *Arch. Biochem. Biophys.*, 390(1):146–148, 2001. doi: 10.1006/abbi.2000.2369. URL `http://www.sciencedirect.com/science/article/pii/S0003986100923695`.

[43] *IUPAC Nomenclature of Inorganic Chemistry*. Blackwell Scientific Publications, Oxford, third edition, 1990. URL `http://goldbook.iupac.org/S05447.html`.

[44] K.P.C. Vollhardt and N.E. Schore. *Organic Chemistry Structure and Function*. W. H. Freeman, New York, 2007.

[45] C.E. Housecroft and A.G. Sharpe. *Inorganic Chemistry*. Pearson Education Ltd., Madrid, second edition, 2005.

[46] R. Leberman and A.K. Soper. Effect of high salt concentrations on water structure. *Nature*, 378(6555):364–366, 1995. doi: 10.1038/378364a0. URL `http://www.nature.com/nature/journal/v378/n6555/pdf/378364a0.pdf`.

[47] B. Hribar, N.T. Southall, V. Vlachy, and K.A. Dill. How ions affect the structure of water. *J. Am. Chem. Soc.*, 124(11):12302–12311, 2002. doi: 10.1021/ja026014h. URL `http://pubs.acs.org/doi/pdf/10.1021/ja026014h`.

[48] S. Park and M.D. Fayer. Hydrogen bond dynamics in aqueous NaBr solutions. *Proc. Nat. Acad. Sci. USA*, 104(43):16731–16738, 2007. doi: 10.1073/pnas.0707824104. URL `http://www.pnas.org/content/104/43/16731.full.pdf?sid=e904d9c6-0d5f-4306-b8ed-a5d61155230a`.

[49] C.D. Cappa, J.D. Smith, B.M. Messer, R.C. Cohen, and R.J. Saykally. Effects of cations on the hydrogen bond network of liquid water: new results from X-ray absorption spectroscopy of liquid microjets. *J. Phys. Chem. B*, 110(11):5301–9, 2006. doi: 10.1021/jp054699c. URL `http://www.ncbi.nlm.nih.gov/pubmed/16539461`.

[50] A.W. Omta, M.F. Kropman, S. Woutersen, and H.J. Bakker. Negligible effect of ions on the hydrogen-bond structure in liquid water. *Science*, 301(5631):347–349, 2003. doi: 10.1126/science.1084801. URL `http://www.sciencemag.org/content/301/5631/347.full.pdf`.

[51] J.M. Andric, G/V. Janjic, D.B. Ninkovic, and S.D. Zaric. The influence of water molecule coordination to a metal ion on water hydrogen bonds. *Phys. Chem. Chem. Phys.*, 14(31):10896–10898, 2012. doi: 10.1039/c2ce25621e. URL `http://dx.doi.org/10.1039/C2CP41125C$\delimiter"026E30F$nhttp://pubs.rsc.org/en/content/articlepdf/2012/cp/c2cp41125c`.

[52] D. Paschek and R. Ludwig. Specific ion effects on water structure and dynamics beyond the first hydration shell. *Angew. Chem. Int. Ed.*, 50(2):352–353, 2011. doi: 10.1002/anie.201004501. URL `http://onlinelibrary.wiley.com/doi/10.1002/anie.201004501/pdf`.

[53] K.J. Tielrooij, N. Garcia-Araez, M. Bonn, and H.J. Bakker. Cooperativity in Ion Hydration. *Science*, 328(5981):1006–1009, 2010. doi: 10.1126/science.1183512. URL `http://www.ncbi.nlm.nih.gov/pubmed/20489020`.

[54] R. Mancinelli, A. Botti, F. Bruni, M.A. Ricci, and A.K. Soper. Perturbation of water structure due to monovalent ions in solution. *Phys. Chem. Chem. Phys.*, 9(23):2959, 2007. doi: 10.1039/b701855j. URL `http://xlink.rsc.org/?DOI=b701855j`.

[55] F. Hofmeister. Zur Lehre von der Wirkung der Salze. *Arch. Exp. Pathol. Pharmakol.*, 24:247–260, 1887. doi: 10.1007/BF01918191. URL `http://link.springer.com/article/10.1007/BF01918191`.

[56] P. Jungwirth and P.S. Cremer. Beyond Hofmeister. *Nature Chem.*, 6(4):261–263, 2014. doi: 10.1038/nchem.1899. URL `http://www.nature.com/doifinder/10.1038/nchem.1899`.

[57] F. Hofmeister. Zur Lehre von der Wirkung der Salze. *Arch. Exp. Pathol. Pharmakol.*, 25:1–30, 1888. doi: 10.1007/BF01838161. URL `http://link.springer.com/article/10.1007/BF01838161`.

[58] A.S. Thomas and A.H. Elcock. Molecular dynamics simulations of hydrophobic associations in aqueous salt solutions indicate a connection between water hydrogen bonding and the hofmeister effect. *J. Am. Chem. Soc.*, 129(13):14887–14898, 2007. doi: 10.1021/ja073097z. URL `http://pubs.acs.org/doi/pdf/10.1021/ja073097z`.

[59] R. Perez-Jimenez, R. Godoy-Ruiz, B. Ibarra-Molero, and J.M. Sanchez-Ruiz. The efficiency of different salts to screen charge interactions in proteins: a Hofmeister effect? *Biophys. J.*, 86(4):2414–2429, 2004. doi: 10.1016/S0006-3495(04)74298-8. URL `http://dx.doi.org/10.1016/S0006-3495(04)74298-8`.

[60] Y. Zhang and P. Cremer. Interactions between macromolecules and ions: the Hofmeister series. *Curr. Opin. Chem. Biol.*, 10(6):658–663, 2006. doi: 10.1016/j.cbpa.2006.09.020. URL `http://linkinghub.elsevier.com/retrieve/pii/S1367593106001517`.

[61] R. Kumar and A.G. Mauk. Atypical effects of salts on the stability and iron release kinetics of human transferrin. *J. Phys. Chem. B*, 113(36):12400–9, 2009. doi: 10.1021/jp903257c. URL `http://www.ncbi.nlm.nih.gov/pubmed/19685917`.

[62] D.F. Parsons, M. Bostrom, P.L. Nostro, and B.W. Ninham. Hofmeister effects: interplay of hydration, nonelectrostatic potentials, and ion size. *Phys. Chem. Chem. Phys.*, 13:12352–12367, 2011. doi: 10.1039/c1cp20538b. URL `http://pubs.rsc.org/en/content/articlepdf/2011/cp/c1cp20538b`.

[63] S. Kumar, D. Sharma, R. Kumar, and R. Kumar. Electrostatic effects control the stability and iron release kinetics of ovotransferrin. *J. Biol. Inorg. Chem.*, 19(6): 1009–1024, 2014. doi: 10.1007/s00775-014-1145-2. URL `http://link.springer.com/10.1007/s00775-014-1145-2`.

[64] B. García-Moreno E. Estimating Binding Constants for Site-Specific Interactions between Monovalent Ions and Proteins. *Methods Enzymol.*, 240:645–667, 1994. doi: 10.1016/S0076-6879(94)40067-9. URL `http://www.sciencedirect.com/science/article/pii/S0076687994400679`.

[65] R.L Baldwin. -How Hofmeister Interactions Affect Protein Stability *. 71:2056–2063, 1996. doi: 10.1016/S0006-3495(96)79404-3. URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1233672/`.

[66] E.J. van Asselt and B.W. Dijkstra. Binding of calcium in the EF-hand of Escherichia coli lytic transglycosylase Slt35 is important for stability. *FEBS letters*, 458(3):429–35, 1999. doi: 10.1016/S0014-5793(99)01198-9. URL `http://www.ncbi.nlm.nih.gov/pubmed/10570954`.

[67] L.M. Pegram and M.T. Record. Thermodynamic origin of hofmeister ion effects. *J. Phys. Chem. B*, 112(31):9428–9436, 2008. doi: 10.1021/jp800816a. URL `http://pubs.acs.org/doi/pdf/10.1021/jp800816a`.

[68] K.B. Rembert, J. Paterová, J. Heyda, C. Hilty, P. Jungwirth, and P.S. Cremer. Molecular mechanisms of ion-specific effects on proteins. *J. Am. Chem. Soc.*, 134(24):10039–10046, 2012. doi: 10.1021/ja301297g. URL `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=22687192$\delimiter"026E30F$npapers://6ef7d602-1904-44f2-bd27-ffa0776e3fdb/Paper/p298`.

[69] G. Ortega, T. Diercks, and O. Millet. Halophilic Protein Adaptation Results from Synergistic Residue-Ion Interactions in the Folded and Unfolded States. *Chemistry & Biology*, pages 1597–1607, 2015. doi: 10.1016/j.chembiol.2015.10.010. URL `http://linkinghub.elsevier.com/retrieve/pii/S1074552115004147`.

[70] D.J. Kushner. Life in High Salt and Solute Concentrations. In *Microbial Life in Extreme Environments*, pages 317–368. Academic Press, London, 1978. doi: DOI:10.1002/jobm.19790190621.

[71] M.B.V. Roberts. *Biology: A Functional Approach*. Biology Series. Nelson Thornes, Cheltenham, 1986.

[72] D.D. Martin, R.A. Ciulla, and M.F. Roberts. Osmoadaptation in archaea. *Applied and environmental microbiology*, 65(5):1815–1825, 1999. URL `http://aem.asm.org/content/65/5/1815.short`.

[73] P.H. Yancey, M.E. Clark, S.C. Hand, R.D. Bowlus, and G.N. Somero. Living with water stress: evolution of osmolyte systems. *Science*, 217(4566):1214–1222, 1982. doi: 10.1126/science.7112124. URL `http://www.sciencemag.org/content/217/4566/1214.full.pdf`.

[74] J.H. Christian and J.A. Waltho. Solute concentrations within cells of halophilic and non-halophilic bacteria. *Biochim. Biophys. Acta*, 65:506–508, 1962. doi: 10.1016/0006-3002(62)90453-5. URL `http://www.sciencedirect.com/science/article/pii/0006300262904535`.

[75] J.K. Lanyi. Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol. Rev.*, 38(3):272–90, 1974. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=413857&tool=pmcentrez&rendertype=abstract`.

[76] H. Eisenberg, M. Mevarech, and G. Zaccai. Advances in Protein Chemistry Volume 43. In *Advances in Protein Chemistry*, volume 43 of *Advances in Protein Chemistry*, pages 1–62. Elsevier, Amsterdam, 1992. doi: 10.1016/S0065-3233(08)60553-7. URL `http://www.sciencedirect.com/science/article/pii/S0065323308605537`.

[77] P. Von Hippel and T. Schleich. The Effects of Neutal Salts on the Stuctural and Conformational Stability of Macromolecules in Solution. In *Structure and stability of biological macromolecules*, pages 417–575. Dekker, New York, 1969. doi: 10.1016/0032-3861(69)90115-3.

[78] G. Zaccai, E. Wachtel, and H. Eisenberg. Solution structure of halophilic malate dehydrogenase from small-angle neutron and X-ray scattering and ultracentrifugation. *J. Mol. Biol.*, 190(1):97–106, 1986. doi: 10.1016/0022-2836(86)90078-1. URL `http://www.sciencedirect.com/science/article/pii/0022283686900781`.

[79] G. Zaccai, F. Cendrin, Y. Haik, N. Borochov, and H. Eisenberg. Stabilization of halophilic malate dehydrogenase. *J. Mol. Biol.*, 208(3):491–500, 1989. doi: 10.1016/0022-2836(89)90512-3. URL `http://www.sciencedirect.com/science/article/pii/0022283689905123`.

[80] F. Bonnete, C. Ebel, G. Zaccai, and H. Eisenberg. Biophysical study of halophilic malate dehydrogenase in solution: revised subunit structure and solvent interactions of native and recombinant enzyme. *J. Chem. Soc., Faraday Trans.*, 89:

2659–2666, 1993. doi: 10.1039/FT9938902659. URL `http://dx.doi.org/10.1039/FT9938902659`.

[81] O. Dym, M. Mevarech, and J.L. Sussman. Structural features that stabilize halophilic malate dehydrogenase from an archaebacterium. *Science*, 267 (5202):1344–1346, 1995. doi: 10.1126/science.267.5202.1344. URL `http://www.sciencemag.org/content/267/5202/1344.abstract`.

[82] D. Madern, C. Pfister, and G. Zaccai. Mutation at a single acidic amino acid enhances the halophilic behaviour of malate dehydrogenase from Haloarcula marismortui in physiological salts. *Eur. J. Biochem.*, 230:1088–1095, 1995. doi: 10.1111/j.1432-1033.1995.tb20659.x. URL `http://onlinelibrary.wiley.com/doi/10.1111/j.1432-1033.1995.1088g.x/pdf`.

[83] D. Madern and G. Zaccai. Stabilisation of halophilic malate dehydrogenase from Haloarcula marismortui by divalent cations: effects of temperature, water isotope, cofactor and pH. *Eur. J. Biochem.*, 249(2):607–611, 1997. doi: 10.1111/j.1432-1033.1997.00607.x. URL `http://onlinelibrary.wiley.com/doi/10.1111/j.1432-1033.1997.00607.x/pdf`.

[84] D. Madern, C. Ebel, M. Mevarech, S.B. Richard, C. Pfister, and G. Zaccai. Insights into the molecular relationships between malate and lactate dehydrogenases: Structural and biochemical properties of monomeric and dimeric intermediates of a mutant of tetrameric L-[LDH-like] malate dehydrogenase from the halophilic archaeon Halo. *Biochemistry*, 39(5):1001–1010, 2000. doi: 10.1021/bi9910023. URL `http://pubs.acs.org/doi/pdf/10.1021/bi9910023`.

[85] M. Mevarech, F. Frolow, and L.M. Gloss. Halophilic enzymes: proteins with a grain of salt. *Biophys. Chem..*, 86(2):155–164, 2000. doi: 0.1016/S0301-4622(00)00126-5. URL `http://www.ncbi.nlm.nih.gov/pubmed/11026680`.

[86] S.B. Richard, D. Madern, E. Garcin, and G. Zaccai. Halophilic adaptation: Novel solvent protein interactions observed in the 2.9 and 2.6 ÅResolution structures of the wild type and a mutant of malate dehydrogenase from Haloarcula marismortui. *Biochemistry*, 39(5):992–1000, 2000. doi: 10.1021/bi991001a. URL `http://pubs.acs.org/doi/pdf/10.1021/bi991001a`.

[87] M. Ginzburg, L. Sachs, and B.Z. Ginzburg. Ion metabolism in a Halobacterium. I. Influence of age of culture on intracellular concentrations. *J. Gen. Physiol.*, 55(2):187–207, 1970. doi: 10.1085/jgp.55.2.187. URL `http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC2202994&blobtype=pdf`.

[88] A. Oren, M. Ginzburg, B.Z. Ginzburg, L.I. Hochstein, and B.E. Volcan. Halophilic Bacterium from the Dead Sea. *Int. J. Syst. Bacteriol.*, 40:209–210, 1990. doi: 10.1099/00207713-40-2-209. URL `http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-40-2-209`.

[89] N.S. Baliga, R. Bonneau, M.T. Facciotti, M. Pan, G. Glusman, E.W. Deutsch, P. Shannon, Y. Chiu, R.S. Weng, R.R. Gan, P. Hung, S.V. Date, E. Marcotte, L. Hood, and W.V. Ng. Genome sequence of *Haloarcula marismortui*: A halophilic archaeon from the Dead Sea. *Genome Res.*, pages 2221–2234, 2004. doi: 10.1101/gr.2700304.Halobacterium. URL `http://genome.cshlp.org/content/14/11/2221.full.pdf+html`.

[90] Schrödinger, L.L.C. Pymol the pymol molecular graphics system, version 1.8. 2015.

[91] H. Eisenberg and E.J. Wachtel. Structural Studies of Halophilic Proteins, Ribosomes, and Organelles of Bacteria Adapted to Extreme Salt Concentrations. *Annu. Rev. Biophys. Biomol. Struct.*, 16:69–92, 1987. doi: 10.1146/annurev.bb.16.060187. 000441. URL `http://www.annualreviews.org/doi/pdf/10.1146/annurev.bb.16.060187.000441`.

[92] J. Birktoft, G. Rhodes, and L. Banaszak. Refined crystal structure of cytoplasmic malate dehydrogenase at 2.5-A resolution. *Biochemistry*, 28(14):6065–6081, 1989. doi: 10.1021/bi00440a051. URL `http://pubs.acs.org/doi/pdf/10.1021/bi00440a051`.

[93] G. Böhm and R. Jaenicke. A structure-based model for the halophilic adaptation of dihydrofolate reductase from Halobacterium volcanii. *Protein Eng.*, 7(2):213–220, 1994. doi: 10.1093/protein/7.2.213. URL `http://peds.oxfordjournals.org/content/7/2/213.long`.

[94] U. Pieper, G. Kapadia, M. Mevarech, and O. Herzberg. Structural features of halophilicity derived from the crystal structure of dihydrofolate reductase from the Dead Sea halophilic archaeon, Haloferax volcanii. *Structure*, 6(1):75–88, 1998. doi: 10.1016/S0969-2126(98)00009-4. URL `http://www.sciencedirect.com/science/article/pii/S0969212698000094$\delimiter"026E30F$nhttp://www.ncbi.nlm.nih.gov/pubmed/9493269$\delimiter"026E30F$nhttp://linkinghub.elsevier.com/retrieve/pii/S0969212698000094`.

[95] D.B. Wright, D.D. Banks, J.R. Lohman, J.L. Hilsenbeck, and L.M. Gloss. *J. Mol. Biol.* doi: 10.1016/S0022-2836(02)00916-6.

[96] M.F. Mullakhanbhai and H. Larsen. Halobacterium volcanii spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Archives of microbiology*, 104(3): 207–14, 1975. doi: 10.1007/BF00447326. URL `http://www.ncbi.nlm.nih.gov/pubmed/1190944`.

[97] A.L. Hartman, C. Norais, J.H. Badger, S. Delmas, S. Haldenby, R. Madupu, J. Robinson, H. Khouri, Q. Ren, T.M. Lowe, J. Maupin-Furlow, M. Pohlschroder, C. Daniels, F. Pfeiffer, T. Allers, and J.A. Eisen. The Complete Genome Sequence of Haloferax volcanii DS2, a Model Archaeon. *PLoS One*, 5(3):e9605, 2010. doi: 10.1371/journal.pone.0009605. URL `http://dx.plos.org/10.1371/journal.pone.0009605`.

[98] C. Bystroff, S.J. Oatley, and J. Kraut. Crystal structures of escherichia coli dihydrofolate reductase: the nadp+ holoenzyme and the folate .cntdot. nadp+ ternary complex. substrate binding and a model for the transition state. *Biochemistry*, 29(13):3263–3277, 1990. doi: 10.1021/bi00465a018. URL `http://dx.doi.org/10.1021/bi00465a018`.

[99] J. Soppa. From genomes to function: haloarchaea as model organisms. *Microbiology*, 152:585–90, 2006. doi: 10.1099/mic.0.28504-0. URL `http://www.ncbi.nlm.nih.gov/pubmed/16514139`.

[100] F. Pfeiffer, S.C. Schuster, A. Broicher, M. Falb, P. Palm, K. Rodewald, A. Ruepp, J. Soppa, J. Tittor, and D. Oesterhelt. Evolution in the laboratory: the genome of Halobacterium salinarum strain R1 compared to that of strain NRC-1. *Genomics*,

91(4):335–46, 2008. doi: 10.1016/j.ygeno.2008.01.001. URL `http://www.ncbi.nlm.nih.gov/pubmed/18313895`.

[101] B.L. Marg, K. Schweimer, H. Sticht, and D. Oesterhelt. A two-α-helix extra domain mediates the halophilic character of a plant-type ferredoxin from Halophilic Archaea. *Biochemistry*, 44(1):29–39, 2005. doi: 10.1021/bi0485169. URL `http://pubs.acs.org/doi/pdf/10.1021/bi0485169`.

[102] A.K. Bandyopadhyay and H.M. Sonawat. Salt dependent stability and unfolding of [Fe2-S2] ferredoxin of Halobacterium salinarum: spectroscopic investigations. *Biophys. J.*, 79(1):501–10, 2000. doi: 10.1016/S0006-3495(00)76312-0. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1300954&tool=pmcentrez&rendertype=abstract`.

[103] A.K. Bandyopadhyay, G. Krishnamoorthy, and H.M. Sonawat. doi: 10.1021/bi001614j.

[104] A.K. Bandyopadhyay, G. Krishnamoorthy, L.C. Padhy, and H.M. Sonawat. Kinetics of salt-dependent unfolding of [2Fe-2S] ferredoxin of Halobacterium salinarum. *Extremophiles*, 11(4):615–25, 2007. doi: 10.1007/s00792-007-0075-0. URL `http://www.ncbi.nlm.nih.gov/pubmed/17406782`.

[105] F. Frolow, M. Harel, J.L. Sussman, M. Mevarech, and M. Shoham. Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin. *Nat. Struc. Biol.*, 3(5):452–458, 1996. doi: 10.1038/nsb0596-452. URL `http://www.nature.com/nsmb/journal/v3/n5/pdf/nsb0596-452.pdf`.

[106] V.M. Coghlan and L.E. Vickery. Expression of human ferredoxin and assembly of the [2Fe-2S] center in Escherichia coli. *Proc. Natl. Acad. Sci. USA*, 86(3):835–839, 1989. doi: 10.1073/pnas.86.3.835. URL `http://www.pnas.org/content/86/3/835.full.pdf`.

[107] H.M. Holden, B.L. Jacobson, J.K. Hurley, G. Tollin, B.H. Oh, L. Skjeldal, Y.K. Chae, H. Cheng, B. Xia, and J.L. Markley. Structure-function studies of [2Fe-2S] ferredoxins. *J. Bioenerg. Biomembr.*, 26(1):67–88, 1994. doi: 10.1007/BF00763220. URL `http://link.springer.com/article/10.1007%2FBF00763220`.

[108] M. M. Werber and M. Mevarech. Purification and characterization of a highly acidic 2Fe-ferredoxin from Halobacterium of the Dead Sea. *Arch. Biochem. Biophys.*, 187(2):447–456, 1978. doi: 10.1016/0003-9861(78)90056-5. URL `http://www.sciencedirect.com/science/article/pii/0003986178900565`.

[109] A.H. Elcock and J.A. McCammon. Electrostatic contributions to the stability of halophilic proteins. *J. Mol. Biol.*, 280(4):731–748, 1998. doi: 10.1006/jmbi.1998.1904. URL `http://www.sciencedirect.com/science/article/pii/S0022283698919040`.

[110] R. Chang. *Physical Chemistry for the Chemical and Biological Sciences*. University Science Books, California, 2000. doi: 10.1021/ed078p594.

[111] X. Tadeo, B. López-Méndez, T. Trigueros, A. Laín, D. Castaño, and O. Millet. Structural Basis for the Aminoacid Composition of Proteins from Halophilic Archea. *PLoS Biol.*, 7(12):e1000257, 2009. doi: 10.1371/journal.pbio.1000257. URL `http://dx.plos.org/10.1371/journal.pbio.1000257`.

[112] J. Qvist, G. Ortega, X. Tadeo, O. Millet, and B. Halle. Hydration dynamics of a halophilic protein in folded and unfolded states. *J. Phys. Chem. B*, 116(10): 3436–3444, 2012. doi: 10.1021/jp3000569. URL `http://pubs.acs.org/doi/pdf/10.1021/jp3000569`.

[113] A. Siglioccolo, A. Paiardini, M. Piscitelli, and S. Pascarella. Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface. *BMC Struct. Biol.*, 11(1):50, 2011. doi: 10.1186/1472-6807-11-50. URL `http://www.biomedcentral.com/1472-6807/11/50`.

[114] G. Ortega, A. Laín, X. Tadeo, D. López-Méndez, B.and Castaño, and O. Millet. Halophilic enzyme activation induced by salts. *Sci. Rep.*, 1:6, 2011. doi: 10.1038/srep00006. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3216494&tool=pmcentrez&rendertype=abstract`.

[115] T. Hoffmann and L. Dougan. Single molecule force spectroscopy using polyproteins. *Chem. Soc. Rev.*, 41(14):4781–4796, 2012. doi: 10.1039/C2CS35033E. URL `http://pubs.rsc.org/en/content/articlehtml/2012/cs/c2cs35033e`.

[116] K. Wagner, D.U. Mick, and P. Rehling. Protein transport machineries for precursor translocation across the inner mitochondrial membrane. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1793(1), 2009. doi: 10.1016/j.bbamcr.2008.05.026. URL `http://dx.doi.org/10.1016/j.bbamcr.2008.05.026`.

[117] R.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. URL `http://dx.doi.org/10.1016/j.cell.2011.04.010`.

[118] J. Alegre-Cebollada, P. Kosuri, and J.M. Fernández. Protease power strokes force proteins to unfold. *Cell*, 145:339–340, 2011. doi: 10.1016/j.cell.2011.04.007. URL `http://www.cell.com/abstract/S0092-8674(11)00424-7`.

[119] S. Huang, K.S. Ratliff, and A. Matouschek. Protein unfolding by the mitochondrial membrane potential. *Nature Structl. Biol.*, 9(4):301–7, 2002. doi: 10.1038/nsb772. URL `http://www.ncbi.nlm.nih.gov/pubmed/11887183`.

[120] K. Shariff, S. Ghosal, and A. Matouschek. The force exerted by the membrane potential during protein import into the mitochondrial matrix. *Biophys. J.*, 86(6):3647–52, 2004. doi: 10.1529/biophysj.104.040865. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1304266&tool=pmcentrez&rendertype=abstract`.

[121] M.A. Schwartz. The force is with us. *Science*, 323(5914), 2009. doi: 10.1126/science.1169414. URL `https://www.sciencemag.org/content/323/5914/588`.

[122] J. Christof, M. Gebhardt, and M. Rief. Force Signaling in Biology. *Science*, 324 (5932), 2009. doi: 10.1126/science.1175874. URL `https://www.sciencemag.org/content/324/5932/1278`.

[123] E.M. Puchner, A. Alexandrovich, A.L. Kho, U. Hensen, L.V. Schafer, B. Brandmeier, F. Grater, H. Grubmuller, H.E. Gaub, and M. Gautel. Mechanoenzymatics of titin kinase. *Proc. Natl. Acad. Sci.*, 105(36):13385–13390, 2008. doi: 10.1073/pnas.0805034105. URL `http://www.pubmedcentral.nih.`

gov/articlerender.fcgi?artid=3001449&tool=pmcentrez&rendertype=
abstracthttp://www.pnas.org/cgi/doi/10.1073/pnas.0805034105.

[124] M.L. Smith, D. Gourdon, William C. Little, K.E. Kubow, R.A. Eguiluz, S. Luna-Morris, and V. Vogel. Force-Induced Unfolding of Fibronectin in the Extracellular Matrix of Living Cells. *PLoS Biology*, 5(10):e268, 2007. doi: 10.1371/journal.pbio.0050268. URL http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371%2Fjournal.pbio.0050268.

[125] K.E. Kubow, R. Vukmirovic, L. Zhe, E. Klotzsch, M.L. Smith, D. Gourdon, S. Luna, and V. Vogel. Mechanical forces regulate the interactions of fibronectin and collagen I in extracellular matrix. *Nat. Commun.*, 6:8026, 2015. doi: 10.1038/ncomms9026. URL http://www.nature.com/doifinder/10.1038/ncomms9026.

[126] 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. URL http://dx.doi.org/10.1016/j.ceb.2010.12.001.

[127] G. Binnig and C.F. Quate. Atomic Force Microscope. *Phys. Rev. Lett.*, 56(9):930–933, 1986. doi: 10.1103/PhysRevLett.56.930. URL http://journals.aps.org/prl/pdf/10.1103/PhysRevLett.56.930.

[128] D. Fotiadis, S.A. Scheuring S Müller, A. Engel, and D.J. Müller. Imaging and manipulations of biological structures with the AFM. *Micron*, 33(4):385–397, 2002. doi: 10.1016/S0968-4328(01)00026-9. URL http://www.sciencedirect.com/science/article/pii/S0968432801000269.

[129] A. Alessandrini and P. Facci. AFM: a versatile tool in biophysics. *Meas. Sci. Technol.*, 16(6):R65–R92, 2005. doi: 10.1088/0957-0233/16/6/R01. URL http://jilawww.colorado.edu/perkinsgroup/(MCDB-4550)_Alessandrini_et_al_2005_AFM_review.pdf.

[130] K. Mitsui, M. Hara, and A. Ikai. Mechanical unfolding of α_2-macroglobulin atomic force microscope. *FEBS Lett.*, 385:29–33, 1996. doi: 10.1016/0014-5793(96)00319-5. URL http://www.sciencedirect.com/science/article/pii/0014579396003195.

[131] Rief M., M. Gautel, F. Oesterhelt, J.M. Fernandez, and Gaub H.E. Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM. *Science*, 276(5315), 1997. doi: 10.1126/science.276.5315.1109. URL http://www.sciencemag.org/content/276/5315/1109.long.

[132] 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. USA*, 96(7):3694–3699, 1999. doi: 10.1073/pnas.96.7.3694. URL http://www.pnas.org/content/96/7/3694.full.pdf.

[133] N. Crampton and D.J. Brockwell. Unravelling the design principles for single protein mechanical strength. *Curr. Opin. Struct. Biol.*, 20(4):508–517, 2010. doi: 10.1016/j.sbi.2010.05.005. URL http://linkinghub.elsevier.com/retrieve/pii/S0959440X10000758.

[134] W. Lu, S.S. Negi, A.F. Oberhauser, and W. Braun. Engineering proteins with enhanced mechanical stability by force-specific sequence motifs. *Proteins*, 80(5):1308–15, 2012. doi: 10.1002/prot.24027. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3323705&tool=pmcentrez&rendertype=abstract`.

[135] 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.*, 15:15767–80, 2013. doi: 10.1039/c3cp52142g. URL `http://www.ncbi.nlm.nih.gov/pubmed/23989058`.

[136] 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):662–71, 1998. doi: 10.1016/S0006-3495(98)77556-3. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1299741&tool=pmcentrez&rendertype=abstract`.

[137] A. Galera-Prat, A. Gómez-Sicilia, A.F. Oberhauser, M. Cieplak, and M. Carrión-Vázquez. Understanding biology by stretching proteins: recent progress. *Curr. Opin. Struct. Biol.*, 20(1):63–69, 2010. doi: 10.1016/j.sbi.2010.01.003. URL `http://www.sciencedirect.com/science/article/pii/S0959440X10000060`.

[138] M. Sikora and M. Cieplak. Mechanical stability of multidomain proteins and novel mechanical clamps. *Proteins: Struct., Funct., Bioinf.*, 79(6):1786–99, 2011. doi: 10.1002/prot.23001. URL `http://www.ncbi.nlm.nih.gov/pubmed/21465555`.

[139] 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. URL `http://www.sciencedirect.com/science/article/pii/S0079610700000171`.

[140] 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. URL `http://www.sciencedirect.com/science/article/pii/S0022283698924664`.

[141] N. Bhasin. Chemistry on a Single Protein, Vascular Cell Adhesion Molecule-1, during Forced Unfolding. *J. Biol. Chem.*, 279(44):45865–45874, 2004. doi: 10.1074/jbc.M404103200. URL `http://www.jbc.org/cgi/doi/10.1074/jbc.M404103200`.

[142] R.B. Best, E. Paci, G. Hummer, and O.K. Dudko. Pulling direction as a reaction coordinate for the mechanical unfolding of single molecules. *J. Phys. Chem. B*, 112(19):5968–5976, 2008. doi: 10.1021/jp075955j. URL `http://pubs.acs.org/doi/pdf/10.1021/jp075955j`.

[143] 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. URL `http://www.nature.com/nsmb/journal/v10/n9/pdf/nsb968.pdf`.

[144] H. Dietz, F. Berkemeier, M. Bertz, and M. Rief. Anisotropic deformation response of single protein molecules. *Proc. Natl. Acad. Sci. USA*, 103(34):12724–8, 2006. doi: 10.1073/pnas.0602995103. URL `http://www.pnas.org/content/103/34/12724.abstract`.

[145] 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), 2009. doi: 10.1016/j.jmb.2009.08.015. URL http://dx.doi.org/10.1016/j.jmb.2009.08.015.

[146] 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), 2005. doi: 10.1529/biophysj.105.061465. URL http://dx.doi.org/10.1529/biophysj.105.061465.

[147] H. Li, A.F. Oberhauser, S.B. Fowler, J. Clarke, and J.M. Fernandez. Atomic force microscopy reveals the mechanical design of a modular protein. *Proc. Natl. Acad. Sci. USA*, 97(12):6527–31, 2000. doi: 10.1073/pnas.120048697. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=18646&tool=pmcentrez&rendertype=abstract.

[148] 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. URL http://www.ncbi.nlm.nih.gov/pubmed/23293964.

[149] A.F. Oberhauser, P.K. Hansma, M. Carrion-Vazquez, and J.M. Fernandez. doi: 10.1073/pnas.98.2.468.

[150] 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):342–50, 2012. doi: 10.1016/j.bpj.2011.12.007. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3260691&tool=pmcentrez&rendertype=abstract.

[151] 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):2836–47, 2008. doi: 10.1002/cphc.200800572. URL http://www.ncbi.nlm.nih.gov/pubmed/19058277.

[152] R.C. Zinober, D.J. Brockwell, G.S. Beddard, A.W. Blake, P.D. Olmsted, S.E. Radford, and D.A. Smith. doi: 10.1110/ps.0224602.

[153] M. Carrion-Vazquez, P.E. Marszalek, A.F. Oberhauser, and J.M. Fernandez. Atomic force microscopy captures length phenotypes in single proteins. *Proc. Natl. Acad. Sci. USA*, 96:11288–11292, 1999. doi: 10.1073/pnas.96.20.11288. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC18026/pdf/pq011288.pdf.

[154] M. Rief, M. Gautel, F. Oesterhelt, J.M. Fernandez, and H.E. Gaub. Reversible Unfolding of Individual Titin Immunoglobulin Domains by Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM. *Science*, 1109(1997), 2013. doi: 10.1126/science.276.5315.1109. URL http://www.sciencemag.org/content/276/5315/1109.

[155] H. Li, M. Carrion-vazquez, A.F. Oberhauser, P.E. Marszalek, and J.M. Fernandez. letters Point mutations alter the mechanical stability of immunoglobulin modules. *Nat. Struc. Biol.*, 7(12):5–8, 2000. doi: 10.1038/81964. URL http://www.nature.com/nsmb/journal/v7/n12/pdf/nsb1200_1117.pdf.

[156] R.B. Best, S.B. Fowler, J.L. 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. URL `http://linkinghub.elsevier.com/retrieve/pii/S0022283603006181$\ delimiter"026E30F$nhttp://www.ncbi.nlm.nih.gov/entrez/query.fcgi? cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12850153`.

[157] M. Sandal, F. Valle, I. Tessari, S. Mammi, E. Bergantino, F. Musiani, M. Brucale, L. Bubacco, and B. Samorì. Conformational equilibria in monomeric α-synuclein at the single-molecule level. *PLoS Biol.*, 6(1):0099–0108, 2008. doi: 10.1371/journal. pbio.0060006. URL `http://www.plosbiology.org/article/fetchObject. action?uri=info:doi/10.1371/journal.pbio.0060006&representation=PDF`.

[158] V. Aggarwal, S.R. Kulothungan, M.M. Balamurali, S.R. Saranya, R. Varadarajan, and S.R.K. Ainavarapu. Ligand-modulated Parallel Mechanical Unfolding Pathways of Maltose-binding Proteins. *J. Biol. Chem.*, 286(32):28056–28065, 2011. doi: 10.1074/jbc.M111.249045. URL `http://www.jbc.org/lookup/doi/ 10.1074/jbc.M111.249045`.

[159] P. Zheng and H. Li. Highly covalent ferric-thiolate bonds exhibit surprisingly low mechanical stability. *J. Am. Chem. Soc.*, 133(17):6791–6798, 2011. doi: 10.1021/ ja200715h. URL `http://pubs.acs.org/doi/abs/10.1021/ja200715h`.

[160] H.C. Kotamarthi, R. Sharma, S. Narayan, S. Ray, and S.R.K. Ainavarapu. Multiple unfolding pathways of leucine binding protein (LBP) probed by single-molecule force spectroscopy (SMFS). *J. Amer. Chem. Soc.*, 135(39):14768–14774, 2013. doi: 10.1021/ja406238q. URL `http://pubs.acs.org/doi/pdf/10.1021/ja406238q`.

[161] T. Hoffmann, K.M. Tych, T. Crosskey, B. Schiffrin, D.J. Brockwell, and L. Dougan. Rapid and robust polyprotein production facilitates single-molecule mechanical characterization of β-barrel assembly machinery polypeptide transport associated domains. *ACS Nano*, 9(9):8811–8821, 2015. doi: 10.1021/acsnano.5b01962. URL `http://pubs.acs.org/doi/10.1021/acsnano.5b01962`.

[162] 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. URL `http://www.sciencedirect.com/science/article/pii/ S0959440X07000073`.

[163] R. Schweitzer-Stenner. *Protein and Peptide Folding, Misfolding, and Non-Folding*. Wiley series in protein and pepride science. Wiley, New Jersey, 2012. doi: 10. 1002/9781118183373. URL `http://onlinelibrary.wiley.com/book/10.1002/ 9781118183373`.

[164] G. Hummer, S. Garde, A.E. García, and L.R. Pratt. New perspectives on hydrophobic effects. *Chem. Phys.*, 258(2-3):349–370, 2000. doi: 10. 1016/S0301-0104(00)00115-4. URL `http://www.sciencedirect.com/science/ article/pii/S0301010400001154`.

[165] C. Chyan, F. Lin, H. Peng, J. Yuan, C. Chang, S. Lin, and G. Yang. Reversible mechanical unfolding of single ubiquitin molecules. *Biophysical journal*, 87(6):3995–4006, 2004. doi: 10.1529/biophysj.104.042754.

URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1304909&tool=pmcentrez&rendertype=abstract.

[166] D. Craig, M. Gao, K. Schulten, and V. Vogel. Tuning the Mechanical Stability of Fibronectin Type III Modules through Sequence Variations. *Structure*, 12(1): 21–30, 2004. doi: 10.1016/j.str.2003.11.024. URL http://linkinghub.elsevier.com/retrieve/pii/S0969212603002946.

[167] S.P. Ng and J. Clarke. Experiments Suggest that Simulations May Overestimate Electrostatic Contributions to the Mechanical Stability of a Fibronectin Type III Domain. *J. Mol. Biol.*, 371(4):851–854, 2007. doi: 10.1016/j.jmb.2007.06.015. URL http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=7276DE0A5D2690E091E92D9BE2835D4C?doi=10.1.1.279.8174&rep=rep1&type=pdf.

[168] P. Zheng, Y. Cao, T. Bu, S.K. Straus, and H. Li. Single molecule force spectroscopy reveals that electrostatic interactions affect the mechanical stability of proteins. *Biophys. J.*, 100(6):1534–41, 2011. doi: 10.1016/j.bpj.2011.01.062. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3059573&tool=pmcentrez&rendertype=abstract.

[169] Bruker afm probes. URL http://www.brukerafmprobes.com/p-3444-mlct.aspx. Accessed: 2016-01-10.

[170] E.L. Florin, M. Rief, H. Lehmann, M. Ludwig, C. Dornmair, V.T. Moy, and H.E. Gaub. Sensing specific molecular interactions with the atomic force microscope. *Biosen. Bioelectron.*, 10(9-10):895–901, 1995. doi: 10.1016/0956-5663(95)99227-C. URL http://chaos.ph.utexas.edu/wp-uploads/2011/03/18.-1995-Biosensors-Bioelectronics.pdf.

[171] S.M. Cook, K.M. Lang, K.M. Chynoweth, M. Wigton, R.W. Simmonds, and T.E. Schäffer. Practical implementation of dynamic methods for measuring atomic force microscope cantilever spring constants. *Nanotech.*, 17:2135–2145, 2006. doi: 10.1088/0957-4484/17/9/010.

[172] R. Levy and M. Maaloum. Measuring the spring constant of atomic force microscope cantilevers : thermal fluctuations and other methods. *Nanotech.*, 13: 33–37, 2002. doi: 10.1088/0957-4484/13/1/307. URL http://iopscience.iop.org/article/10.1088/0957-4484/13/1/307/pdf.

[173] J.F. Marko, E.D. Siggia, and S. Smith. Entropic Elasticity of X-Phage DNA Explicit and Implicit Learning and Maps of Cortical Motor Output. *Science*, 265(5178):1599–1600, 1994. doi: 10.1126/science.8079175. URL http://www.sciencemag.org/content/265/5178/1599.full.pdf.

[174] J.F. Marko and E.D. Siggia. Stretching DNA. *Macromolecules*, 28 (26):8759–8770, 1995. doi: 10.1021/ma00130a008. URL http://pubs.acs.org/doi/abs/10.1021/ma00130a008$\delimiter"026E30F$npapers2://publication/doi/10.1021/ma00130a008.

[175] 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. URL http://dx.doi.org/10.1016/S0006-3495(01)75881-X.

[176] Y. Cao and H. Li. Polyprotein of GB1 is an ideal artificial elastomeric protein. *Nat. Mater.*, 6(2):109–114, 2007. doi: 10.1038/nmat1825. URL `http://www.nature.com/nmat/journal/v6/n2/pdf/nmat1825.pdf`.

[177] R. Rounsevell, J.R. Forman, and J. Clarke. Atomic force microscopy: Mechanical unfolding of proteins. *Methods*, 34, 2004. doi: 10.1016/j.ymeth.2004.03.007. URL `http://www.sciencedirect.com/science/article/pii/S1046202304000556`.

[178] G. Žoldák and M. Rief. Force as a single molecule probe of multidimensional protein energy landscapes. *Curr. Opin. Struct. Biol.*, 23(1):48–57, 2013. doi: 10.1016/j.sbi.2012.11.007. URL `http://linkinghub.elsevier.com/retrieve/pii/S0959440X12001911`.

[179] A. Galer-Prat, A. Gomez-Sicilia, A.F. Oberhauser, and M. Carrion-Vazquez. Understanding biology by stretching proteins: recent progress. *Curr. Opin. Struct. Biol.*, 20(1):63–69, 2010. doi: 10.1016/j.sbi.2010.01.003. URL `http://www.ncbi.nlm.nih.gov/pubmed/20138503`.

[180] T. Bornschlögl and M. Rief. Single-molecule dynamics of mechanical coiled-coil unzipping. *Langmuir*, 24(4):1338–1342, 2008. doi: 10.1021/la7023567. URL `http://pubs.acs.org/doi/pdf/10.1021/la7023567`.

[181] 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):1467–73, 2011. doi: 10.1016/j.bpj.2011.08.021. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3177063&tool=pmcentrez&rendertype=abstract`.

[182] P. Zheng, S.I.J. Takayama, A.G. Mauk, and H. Li. Hydrogen bond strength modulates the mechanical strength of ferric-thiolate bonds in rubredoxin. *J. Amer. Chem. Soc.*, 134(9):4124, 2012. doi: 10.1021/ja2078812. URL `http://pubs.acs.org/doi/pdf/10.1021/ja2078812`.

[183] P. Zheng, C. Chou, Y. Guo, Y. Wang, and H. Li. Single Molecule Force Spectroscopy Reveals the Molecular Mechanical Anisotropy of the FeS 4 Metal Center in Rubredoxin. *J. Amer. Chem. Soc.*, 134:17783–17792, 2013. doi: dx.doi.org/10.1021/ja406695g. URL `http://pubs.acs.org/doi/pdf/10.1021/ja406695g`.

[184] P. Zheng, S.I.J. Takayama, A.G. Mauk, and H. Li. Single molecule force spectroscopy reveals that iron is released from the active site of rubredoxin by a stochastic mechanism. *J. Amer. Chem. Soc.*, 135(21):7992–8000, 2013. doi: 10.1021/ja402150q. URL `http://pubs.acs.org/doi/pdf/10.1021/ja402150q`.

[185] P. Zheng, Y. Wang, and H. Li. Reversible Unfolding-Refolding of Rubredoxin: A Single-Molecule Force Spectroscopy Study. *Angew. Chem. Int. Ed.*, 53(51):14060–14063, 2014. doi: 10.1002/anie.201408105. URL `http://doi.wiley.com/10.1002/anie.201408105`.

[186] P. Zheng, G.M. Arantes, M.J. Field, and H. Li. Force-induced chemical reactions on the metal centre in a single metalloprotein molecule. *Nat. Commun.*, 6:7569, 2015. doi: 10.1038/ncomms8569. URL `http://www.nature.com/doifinder/10.1038/ncomms8569`.

[187] M.W. Day, B.T. Hsu, L. Joshua-Tor, J.B. Park, Z.H. Zhou, M.W. Adams, and D.C. Rees. X-ray crystal structures of the oxidized and reduced forms of the rubredoxin from the marine hyperthermophilic archaebacterium Pyrococcus furiosus. *Protein Sci.*, 1:1494–1507, 1992. doi: 10.1002/pro.5560011111. URL `http://onlinelibrary.wiley.com/doi/10.1002/pro.5560011111/epdf`.

[188] M. Grandbois, M. Beyer, M. Rief, H. Clausen-Schaumann, and H.E. Gaub. How Strong is a Covalent Bond? *Science*, 283(5408):1727–1730, 1999. doi: 10.1126/science.283.5408.1727. URL `http://www.jstor.org/stable/2897510`.

[189] M.A. Lantz. Quantitative Measurement of Short-Range Chemical Bonding Forces. *Science*, 291(5513):2580–2583, 2001. doi: 10.1126/science.1057824. URL `http://www.sciencemag.org/cgi/doi/10.1126/science.1057824`.

[190] M. Bertz and M. Rief. Mechanical unfoldons as building blocks of maltose-binding protein. *J. Mol. Biol.*, 378(2):447–58, 2008. doi: 10.1016/j.jmb.2008.02.025. URL `http://www.sciencedirect.com/science/article/pii/S0022283608002106`.

[191] 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. USA*, 105(6):1885–1890, 2008. doi: 10.1073/pnas.0706775105. URL `http://www.pnas.org/content/105/6/1885.full.pdf`.

[192] D.N. Greene, T. Garcia, R.B. Sutton, K.M. Gernert, G.M. Benian, and A.F. Oberhauser. Single-molecule force spectroscopy reveals a stepwise unfolding of Caenorhabditis elegans giant protein kinase domains. *Biophys. J.*, 95(3):1360–1370, 2008. doi: 10.1529/biophysj.108.130237. URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2479574/`.

[193] 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. URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1184350/pdf/biophysj00037-0057.pdf`.

[194] R.B. Kapust and D.S. Waugh. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.*, 8(8):1668–1674, 1999. doi: 10.1110/ps.8.8.1668. URL `http://onlinelibrary.wiley.com/doi/10.1110/ps.8.8.1668/epdf`.

[195] M.R. Dyson, S.P. Shadbolt, K.J. Vincent, R.L. Perera, and J. McCafferty. Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. *BMC Biotechnol*, 4:32, 2004. doi: 10.1186/1472-6750-4-32. URL `http://www.ncbi.nlm.nih.gov/pubmed/15598350`.

[196] I. Kataeva, J. Chang, H. Xu, C. Luan, J. Zhou, V.N. Uversky, D. Lin, P. Horanyi, Z.J. Liu, L.G. Ljungdahl, J. Rose, M. Luo, and B. Wang. Improving Solubility of Shewanella oneidensis MR-1 and Clostridium thermocellum JW-20 Proteins Expressed into Esherichia coli. *J. Proteome Res.*, 4:1942–1951, 2005. doi: 10.1021/pr050108j. URL `http://pubs.acs.org/doi/pdf/10.1021/pr050108j`.

[197] P. Zheng, Y. Cao, and H. Li. Facile method of constructing polyproteins for single-molecule force spectroscopy studies. *Langmuir*, 27(10):5713–5718, 2011. doi: 10.1021/la200915d. URL `http://www.ncbi.nlm.nih.gov/pubmed/21486060`.

[198] 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. URL `http://www.sciencedirect.com/science/article/pii/S0022283685703646`.

[199] N. Crampton, K. Alzahrani, G.S. Beddard, S.D. Connell, and D.J. Brockwell. Mechanically unfolding protein L using a laser-feedback-controlled cantilever. *Biophys. J.*, 100(7):1800–9, 2011. doi: 10.1016/j.bpj.2011.02.021. URL `http://www.sciencedirect.com/science/article/pii/S0006349511002384`.

[200] W. Kabsch and C. Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen bonded and geometrical features. *Biopolymers*, 22:2577–2637, 1983. doi: 10.1002/bip.360221211. URL `http://onlinelibrary.wiley.com/doi/10.1002/bip.360221211/pdf`.

[201] V. Sobolev, A. Sorokine, J. Prilusky, E.E. Abola, and M. Edelman. Automated analysis of interatomic contacts in proteins. *Bioinformatics*, 15(4):327–332, 1999. doi: 10.1093/bioinformatics/15.4.327. URL `http://bioinformatics.oxfordjournals.org/content/15/4/327.full.pdf+html`.

[202] G. Yang, C. Cecconi, W.A. Baase, I.R. Vetter, W.A. Breyer, J.A. Haack, B.W. Matthews, F.W. Dahlquist, and C. Bustamante. Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme. *Proc. Natl. Acad. Sci. USA*, 97 (1):139–144, 2000. doi: 10.1073/pnas.97.1.139. URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC26629/pdf/pq000139.pdf`.

[203] S.P. Ng, L.G. Randles, and J. Clarke. Single molecu;e studiesof protein unfolding using AFM. In Y. Bai and R. Nussinov, editors, *Protein folding protocols*, pages 139–168. Humana Press, New Jersey, 2006. doi: 10.1385/1597451894. URL `http://www.springer.com/gp/book/9781588296221`.

[204] M. Carrion-Vazquez, H. 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. URL `https://groups.chem.ubc.ca/hongbin/pdf/LHB-nsb-ubiquitin-reprint.pdf`.

[205] J.M. Zhao, H. Lee, R.A. Nome, S. Majid, N.F. Scherer, and W.D. Hoff. Single-molecule detection of structural changes during Per-Arnt-Sim (PAS) domain activation. *Proc. Natl. Acad. Sci. USA*, 103(31):11561–6, 2006. doi: 10.1073/pnas.0601567103. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1544209&tool=pmcentrez&rendertype=abstract`.

[206] Greg T. Hermanson. Bioconjugate Techniques. In *Bioconjugate Techniques*, pages 276–335. Elsevier, Amsterdam, 2008. doi: 10.1016/B978-0-12-370501-3.00005-9. URL `http://www.sciencedirect.com/science/article/pii/B9780123705013000059`.

[207] O. Blecher, S. Goldman, and M. Mevarech. High expression in Escherichia coli of the gene coding for dihydrofolate reductase of the extremely halophilic archaebacterium Haloferax volcanii. Reconstitution of the active enzyme and mutation studies. *Eur. J. Biochem.*, 216(1):199–203, 1993. doi: 10.1111/j.1432-1033.1993.tb18133. URL `http://www.ncbi.nlm.nih.gov/pubmed/8365406`.

[208] N. Altman-Price and M. Mevarech. Genetic evidence for the importance of protein acetylation and protein deacetylation in the halophilic archaeon Haloferax volcanii. *J. Bacteriol.*, 191(5):1610–1617, 2009. doi: 10.1128/JB.01252-08. URL `http://jb.asm.org/content/191/5/1610.full.pdf+html`.

[209] T. Hase, S. Wakabayashi, H. Matsubara, M. Mevarech, and M.M. Werber. Amino acid sequence of 2Fe-2S ferredoxin from extreme halophile, Halobacterium of the Dead Sea. *Biochim. Biophys. Acta*, 623:139–145, 1980. doi: 10.1016/0005-2795(80)90016-1. URL `http://www.ncbi.nlm.nih.gov/pubmed/7378468`.

[210] M.C. Leake, D. Wilson, M. Gautel, and R.M. Simmons. The elasticity of single titin molecules using a two-bead optical tweezers assay. *Biophys. J.*, 87(2):1112–1135, 2004. doi: 10.1529/biophysj.103.033571. URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1304451/pdf/1112.pdf`.

[211] K. Pant, R.L. Karpel, I. Rouzina, and M.C. Williams. Salt dependent binding of T4 gene 32 protein to single and double-stranded DNA: single molecule force spectroscopy measurements. *J. Mol. Biol.*, 349(2):317–30, 2005. doi: 10.1016/j.jmb.2005.03.065. URL `http://www.ncbi.nlm.nih.gov/pubmed/15890198`.

[212] T. Pirzer, M. Geisler, T. Scheibel, and T. Hugel. Single molecule force measurements delineate salt, pH and surface effects on biopolymer adhesion. *Phys. Biol.*, 6:025004, 2009. doi: 10.1088/1478-3975/6/2/025004. URL `http://iopscience.iop.org/article/10.1088/1478-3975/6/2/025004/pdf`.

[213] W. Zhang, X.J. Lu, W.K. Zhang, and J.C. Shen. EMSA and Single-Molecule Force Spectroscopy Study of Interactions between Bacillus subtilis Single-Stranded DNA-Binding Protein and Single-Stranded DNA. *Langmuir*, 27(24):15008–15015, 2011. doi: Doi10.1021/La203752y. URL `http://pubs.acs.org/doi/pdf/10.1021/la203752y`.

[214] G. Gronau, Z. Qin, and M.J. Buehler. Effect of sodium chloride on the structure and stability of spider silk's N-terminal protein domain. *Biomater. Sci.*, 1(3):276–284, 2013. doi: 10.1039/C2BM00140C. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3700428&tool=pmcentrez&rendertype=abstract`.

[215] T. Bu, H.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):12319–25, 2012. doi: 10.1021/la301940g. URL `http://www.ncbi.nlm.nih.gov/pubmed/22823458`.

[216] T. Kaare, J.G. Olsen, and B.B. Kragelund. Protein stability, flexibility and function. *Biochim. Biophys. Acta, Proteins Proteomics*, 1814:969–976, 2011. doi: 10.1016/j.bbapap.2010.11.005. URL `http://dx.doi.org/10.1016/j.bbapap.2010.11.005`.

[217] R. Majumdar, P. Manikwar, J.M. Hickey, H.S. Samra, H.A. Sathish, S.M. Bishop, C.R. Middaugh, D.B. Volkin, and D.D. Weis. Effects of salts from the Hofmeister series on the conformational stability, aggregation propensity, and local flexibility of an IgG1 monoclonal antibody. *Biochemistry*, 52:3376–3389, 2013. doi: 10.1021/bi400232p. URL `http://pubs.acs.org/doi/pdf/10.1021/bi400232p`.

[218] P. Radivojac, Z. Obradovic, D.K. Smith, G. Zhu, S. Vucetic, C.J. Brown, J.D. Lawson, and A.K. Dunker. Protein flexibility and intrinsic disorder. *Protein Sci.*, 13(1):71–80, 2004. doi: 10.1110/ps.03128904.structures. URL `http://www.proteinscience.org/cgi/content/abstract/13/1/71`.

[219] S.P. Ng, K.S. Billings, T. Ohashi, M.D. Allen, R.B. Best, L.G. Randles, H.P. Erickson, and J. Clarke. Designing an extracellular matrix protein with enhanced mechanical stability. *Proc. Natl. Acad. Sci. USA*, 104(23):9633–7, 2007. doi: 10.1073/pnas.0609901104. URL `http://www.pnas.org/content/104/23/9633.full`.

[220] E. Pensa, E. Cortés, G. Corthey, P. Carro, C. Vericat, M.H. Fonticelli, G. Benítez, A.A. Rubert, and R.C. Salvarezza. The chemistry of the sulfur-gold interface: In search of a unified model. *Acc. Chem. Res.*, 45(8):1183–1192, 2012. doi: 10.1021/ar200260p. URL `http://pubs.acs.org/doi/pdf/10.1021/ar200260p`.

[221] T. Allers, S. Barak, S. Liddell, K. Wardell, and M. Mevarech. Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in Haloferax volcanii. *Appl. Environ. Microbiol.*, 76(6):1759–1769, 2010. doi: 10.1128/AEM. 02670-09. URL `http://aem.asm.org/content/76/6/1759.full.pdf+html`.

[222] D. Kixmüller and J.C. Greie. Construction and characterization of a gradually inducible expression vector for Halobacterium salinarum, based on the kdp promoter. *Appl. Environ. Microbiol.*, 78:2100–2105, 2012. doi: 10.1128/AEM.07155-11. URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3302624/pdf/zam2100.pdf`.

[223] J. Lv, S. Wang, C. Zeng, Y. Huang, and X. Chen. Construction of a shuttle expression vector with a promoter functioning in both halophilic Archaea and Bacteria. *FEMS Microbiol. Lett.*, 349(1):9–15, 2013. doi: 10.1111/1574-6968. 12278. URL `http://www.ncbi.nlm.nih.gov/pubmed/24106795`.

[224] M. Ginzburg, L. Sachs, and B.Z. Ginzburg. Ion Metabolism in a Halobacterium. *J. Gen. Physioly*, 55(4):187–207, 1970. doi: 10.1007/BF01870826. URL `http://jgp.rupress.org/content/55/2/187.long`.

[225] N.J. Greenfield. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protoc.*, 1(6):2876–2890, 2007. doi: 10.1038/nprot.2006.202. Using. URL `http://www.nature.com/nprot/journal/v1/n6/abs/nprot.2006.202.html`.

[226] Protein fluorescence. In J.R. Lakowicz, editor, *Principles of Fluorescence Spectroscopy*, pages 529–575. Springer, 2006. doi: 10.1007/978-0-387-46312-4_16. URL `http://dx.doi.org/10.1007/978-0-387-46312-4_16`.

[227] 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*, 91:012710, 2015. doi: 10.1103/PhysRevE.91.012710. URL `http://link.aps.org/doi/10.1103/PhysRevE.91.012710`.