UNIVERSITY OF LEEDS

Development of ClpXP as a tool for investigating the mechanical properties of biomolecules applied to polyglutamine repeat proteins

Christopher James Wilson

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

Astbury Centre for Structural Molecular Biology

November 2015

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

Acknowledgements

I would like to acknowledge the contributions made to this work by my supervisors David Brockwell and Sheena Radford and to thank them for their support and faith in me when thing have not been going well. They have provided excellent advice and instruction over the course of this project and have been very patient when I have been a less than model student.

I would like to thank my fiancée Clare Wishart, without whose love and support I would have given up many times over the course of this project. I love you and I hope that we can confuse people for many years to come by both being Dr C. Wilson.

I would like to thank Oliver Farrance and Toni Hoffman who have patiently answered all of my stupid questions about various aspects of forces and AFM.

I would also like to acknowledge all of the members of the Radford, Brockwell and Berry labs who have provided as appropriate: support, advice, helpful discussion, interesting and often inappropriate lunchtime conversation, friendship, and cake. I would particularly like to mention Lindsay McMorran for giving me someone to share my stress and frustration about thesis writing, and our fantastic laboratory manager Nasir Khan, who has again and again done everything in his power to make sure that things went smoothly, keeping me well supplied with not just equipment and reagents but also endless biscuits and even complete meals to take home.

Finally I would like to thank my parents, Keith and Wendy Wilson without whose moral and financial support undertaking this PhD would not have been possible.

Abstract

It has become increasingly apparent that mechanical force plays an important role in biology. Biophysical techniques such as optical tweezers and atomic force microscopy (AFM) have allowed the investigation of mechanical stability at the level of a single protein molecule. However, despite the increasing sensitivity of these techniques, it is still difficult to mimic precisely the geometry of extension, forces and loading rates applied *in vivo.*

Here we have developed a technique using a bacterial proteasome, ClpXP, which allowed the investigation of the mechanical stability of proteins at more biologically relevant forces and loading rates. It was demonstrated that various degradation signals can be used to target the proteins under investigation to ClpXP where they were unfolded, translocated and degraded. Several ClpX variants were investigated and an assay developed using a pseudohexameric ClpX variant that allowed robust degradation of several proteins.

This assay was used to investigate the properties of a protein containing a 30 polyglutamine repeat sequence. Previous studies of polyglutamine repeats, using AFM, have shown that it may have interesting mechanical properties: it has either extreme mechanical strength, or access to a conformation which has a high mechanical strength.

It was shown that this protein can be completely degraded using ClpXP without any intermediate product, and without reducing the degradation rate compared to a control protein without a polyglutamine repeat. This demonstrates that this assay can be used to investigate proteins whose mechanical properties are of interest and that the loading rates and application

of force applied by this assay differ enough from those of AFM that different and more biologically relevant results can be obtained.

Table of Contents

List of Abbreviations

AAA+ - Atpase Associated with cellular Activity

AFM – Atomic force microscope/microscopy

CD – Circular dichroism spectroscopy

Clp – Caesinolytic protease

DNA – deoxyribose nucleic acid

dNTP – deoxyribose nucleotide triphosphates

DRPLA - dentatorubropallidoluysian atrophy

DTT - dithiothreitol

E. coli – Eschericia coli

EDTA - Ethylenediaminetetraacetic acid

EM – Electron microscopy

HD - Huntington's disease

HTT^{NT} – N-terminal fragment of huntingtin

IDP – intrinsically disordered peptide

IPTG - isopropyl-β-thiogalactosidase

LB – Lysogeny Broth

MCS – multiple cloning site

MD – Molecular dynamics

MLCT – AFM cantilever model designation

MT – magnetic tweezers

MWCO – Molecular weight cut-off

NTD – N-terminal domain (of ClpX)

NTP – nucleotide triphosphate

OT – optical tweezers

PAGE – polyacrylamide gel electrophoresis

PBS – phosphate buffered saline

PCR – polymerase chain reaction

PolyQ – Polyglutamine

SBMA - spinal bulbar muscular atrophy

SCA –spinocerebrellar ataxia

SDS – sodium dodecyl sulphate

TEV – Tobacco Etch Virus

T_m – melting temperature

TS – Transition State

UPS – ubiquitin proteasome system

WLC – Worm like chain

X-gal - 5-bromo-4-chloro-3-indolyl-β-D-galactoside

List of Tables

List of Figures

1 Introduction

1.1 Forces in biology

We are beginning to understand that mechanical forces play an important role in biological function. For example, it is becoming apparent that in addition to chemical processes such as phosphorylation[1] and calcium-sensing[2] cellular events that are controlled by or produce mechanical force are also important. These occur on a wide range of scales, they can vary from subnanometer movements in the ear which we can sense and which allow us to hear[3], interactions such as the hydrolysis of ATP, which in ClpX produces a mechanical force with a step-size of 1nm [4], or ATP hydrolysis in skeletal muscle which leads to a force being applied with a stroke size of approximately 10nm [5] up to the whole organism level where many such events can combine to produce large macroscopic mechanical forces that can be applied for movement or the alteration of the organism's environment on a scale of metres[6]. Until recently it has been difficult or even impossible to study these forces on a microscopic scale and to understand how they are generated or sensed. However, since the first mechanical protein unfolding experiments were carried out in 1997 techniques have been refined and new techniques developed to allow us to study these events down to the single molecule level[7].

Force is known to make a vital contribution to processes such as cell motility[8–10], DNA replication[11] and protein unfolding[12]. In some forms of cell motility, blebs are formed in the cell membrane by the contraction of cortical acto-myosin, which detaches the membrane from the underlying cytoskeleton. This

bleb is expanded by the flow of cytoplasm into it and a mesh of actin and myosin II assembles in the bleb forming a contractile cortex attached to the cell membrane. This mesh then contracts forcing the cytosol back into the body of the cell and causing the cell to move[13]. DNA helicases are essential enzymes; they are molecular motors which hydrolyse nucleotide triphosphates in order to generate forces which they use to move along a DNA track. They are capable of generating enough force to not only move themselves but also to unwind the DNA while managing interactions between the DNA and other proteins such as transcription factors, histones and other repair and replication machinery[14]. Protein unfolding (which is the main focus of this thesis) is achieved by mechanical motor proteins of the AAA+ superfamily, one of which is ClpX which I will talk about in detail later in this chapter. In addition to these "obvious" uses of forces there are many other events in cells which require the use of mechanical force to unfold a protein, such as the translocation of proteins across membranes that is necessary during import into mitochondria[15] of proteins synthesised in the cytosol[16].

The relevance and importance of force in nature has been suspected for some time, but as we have only recently been able to measure the levels of force that can act at the level of a cell or even a single protein it has been difficult or even impossible to study. Using biophysical techniques such as atomic force microscopy (AFM) and optical tweezers (OT) it is possible to apply and measure forces that are applied to a single protein molecule[17] *in vitro*. Previous studies using these methods have shown interesting and counter-intuitive results such as the observation that protein complexes with very high affinity interactions, such as avidin/biotin[18] and colicin/immunity protein[19], are

easily dissociated by low levels (<18 pN) of mechanical force. Other complexes such as that between bacterial adhesins and mannose, (which decorates the cells of the urinary tract for example[20]) to which they bind have evolved a more complex reaction to force, whereby they operate as “catch bonds” which, under force, undergo an allosteric rearrangement into a conformation with higher affinity binding which allows them to remain attached under the forces they are exposed to[21,22].

1.2 Dynamic force spectroscopy

A small protein is able to fold reversibly because its final state is a thermodynamically stable state which has a lower free energy than other possible states[23]. As the number of amino acids in a polypeptide increases the number of possible conformations increases, soon reaching the point that the length of time required to randomly sample them all would be longer than the lifetime of the protein, the organism or even the universe itself. It is therefore apparent that proteins do not fold by randomly sampling all of the conformations available to them and have evolved to fold quickly into their native state. In its simplest form this means that we can imagine a system in which there is an unfolded polypeptide (U) which can move through some sort of transition state (TS) to reach it's folded state (F) where it will remain until conditions are changed in such a way that the energy barrier keeping it in its folded state is lowered sufficiently that the folded peptide can overcome it and become unfolded as illustrated in Figure 1.1. However, proteins are not static entities and the protein will transiently visit the unfolded state due to thermal fluctuations.

It is usual to study protein unfolding by the application of denaturants, which disrupt the protein-protein, protein-solvent and solvent-solvent interactions reducing the energy required to reach the transition state and unfold, or by increasing the temperature, which destabilises the folded state by reducing the entropic cost of unfolding. The application of a stretching mechanical force can be thought of as tilting the energy landscape lowering the energy of the transition state and making the unfolded state more energetically favourable, this is in contrast to an enzyme-catalysed reaction which would lower the

energy of the transition state without affecting the energy levels of the folded or unfolded states.

Importantly the ability to resist force-induced unfolding is governed by the protein's structure local to the points of application. The mechanical behaviour of a protein is thus distinct to the thermodynamic or kinetic properties measured when unfolded by global denaturants[24].

Several manipulation techniques exist which can give information on forces in biochemical systems such as protein unfolding, DNA elasticity, protein-protein interactions and molecular motors. These techniques can be used to either apply forces to these systems in order to investigate their properties, or to directly measure the forces applied by these systems[26]. Of the major techniques for force spectroscopy each has its optimal length and force scales. Optical tweezers (OT) and magnetic tweezers (MT) use a technique of tethering a molecule between two beads and applying forces to the beads in order to manipulate the molecule. OT use radiation pressure from a laser to manipulate the sample, it has excellent force resolution and time resolution, but the forces applied are relatively low (<300 pN). MT use magnetic fields to manipulate the sample, it can apply forces between 10-100 pN[27], the spatial resolution is excellent and can detect the unwinding of as little as 3 base pairs (~15nm) however a compromise must be reached between spatial and temporal resolution[28]. However, as in each assay only a single molecule is addressed, the experimental throughput is low, though this is improving. Particularly useful with MT is the ability to apply torque, or twisting forces, as well as linear forces[27]. While MT have the potential to be a powerful technique with applications in a variety of biophysical systems, they have been most

used for the study of nucleic acids and enzymes associated with them as they are particularly suited for this application. Atomic force microscopy (AFM) can apply a large range of forces (10-10,000pN) over large distances (1-10,000nm) but has a lower force resolution due to the thermal noise generated by the large cantilever (≈10pN) [29]. AFM was chosen as the technique to be used for this work as it is the most suitable for the system under investigation since it is capable of measuring forces in the same range as ClpX generates, around 20 pN [4,30], it has been used extensively for the study of proteins[17,24,31], and would allow direct comparisons with previous work on polyglutamine repeats[32,33].

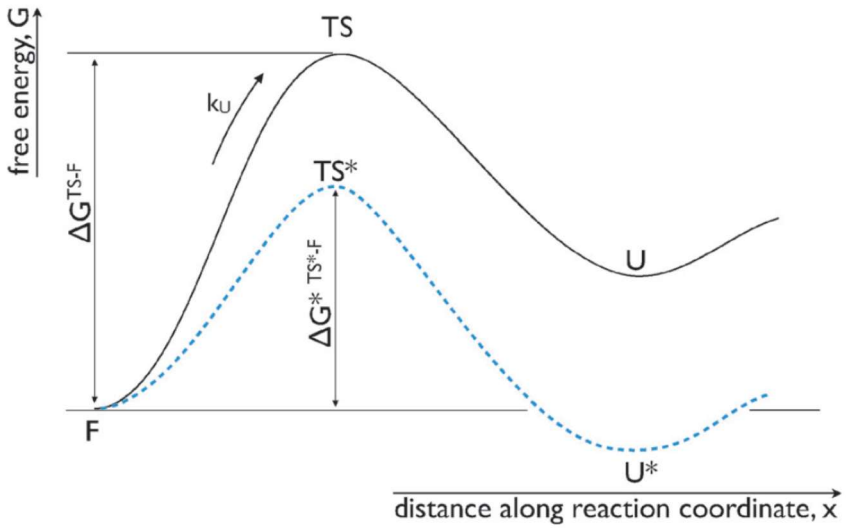


Figure 1.1: The effect of mechanical force on the free energy landscape of a polypeptide. The free energy landscape shows a simple two state model of polypeptide folding. Black line: the normal situation in which the polypeptide is "trapped" in its native folded state. Blue line: the situation after a mechanical force is applied which has tilted the energy landscape making the unfolded state more energetically favourable (U*) and reducing the free energy required to reach this unfolded state (ΔG^{*TS^*-F}) and increasing the unfolding rate k_U. Taken from Hoffman *et al*[30].

1.3 Atomic force microscopy

AFM uses a flexible cantilever onto the tip of which a laser can be focused. The laser is reflected by the tip, onto a photoelectric sensor. If the laser is spatially fixed, a small movement in the tip will cause an amplified movement in the reflected laser on the sensor (sometimes called an optical lever effect). This movement of the laser spot can be used to quantify the amount of movement at the tip. If the spring constant (measured in N/m) is known, the force applied at the tip can be calculated. When a force is applied to the tip its deflection is recorded by the sensor as described in Figure 1.2.

Experiments investigating the mechanical unfolding of proteins use multi-domain protein constructs, also known as a polyproteins, in order to be able to detect the events of interest and to separate these from noise in the system due to non-specific protein-tip and tip-substrate interactions. These polyproteins are proteins in which a single domain with a well characterised force response is repeated several times with a short linker sequence inserted between the repeated. Polyproteins do exist in nature but these contain a heterogeneous mixture of similar, but not identical, domains. As each domain in the polyprotein needs to be identical to provide a clearly identifiable force profile in AFM experiments it is necessary to produce bespoke polyproteins by protein engineering[24]. There are two approaches we can take when designing the polyprotein, we can construct a homopolymer of the domain we are interested in so that each event we observe will relate to that domain, or we can use one or more copies of our domain of interest inserted at alternating positions in a larger polyprotein made up of a domains which has already been well-characterised by AFM and so are easily separated from the events

relating to the domain of interest[24]. The approach taken will depend on several factors that include, ease of protein expression and the type of data to be acquired. Several proteins have been used for AFM experiments, probably the most used is the I27 domain from the I-band of the muscle protein Titin (Human cardiac titin I band module 27). It is an immunoglobulin like domain (Figure 1.3). It is ideal for force experiments due to its known structure and high stability to chemical, thermal and force denaturation[34].

Force mode AFM uses a sample immobilised onto a surface. The AFM tip is lowered towards the surface and when the cantilever contacts the surface it is deflected. When this deflection reaches a certain threshold the AFM tip is retracted away from the surface. If a molecule has attached to the tip it will begin to exert a force on the cantilever as it resists extension. As the tip continues to be retracted, the forces acting on the tip will change allowing a force profile of the event to be generated[26].

For protein mechanical unfolding experiments a gold surface is usually used to which a solution containing the protein of interest containing a single or two consecutive cysteines at one end of the protein can be attached by a strong covalent Au-S bond to the surface. Protein-tip attachment can vary dependent on the system being investigated, it is possible to tether a protein using functionalization of the tip[35], or by the non-specific interactions the silicon-nitride tip makes with the protein (as long as the interaction is stronger than the unfolding force to be measured). The concentration of the protein is varied to find a balance between an acceptable number of interactions and the number of interactions involving a single molecule[36].

Constant velocity experiments can be carried out, where the retraction of the tip occurs at a constant rate and the forces applied to the cantilever vary. This is the most common form of AFM mechanical unfolding due to its ease of use. It produces a saw-tooth force profile showing the unfolding of the multiple domains. From this profile it is possible to directly extract the forces at which unfolding events occur (Figure 1.4), and by fitting the length of polypeptide chain released upon unfolding to a worm like chain (WLC) model of polymer elasticity (Equation 1.1). The WLC model is a polymer physics model to explain the behaviour of semi-flexible polymers. In contrast to a freely-jointed chain model which models inflexible regions connected by joints which have free rotation, a WLC model describes a continuously flexible isotropic rod. This has been shown experimentally to be a better model for biological polymers at the levels of force we will be using to investigate them (>10pN)[37,38].

The WLC relates the contour length, (the theoretical length at maximum extension of an unfolded domain) to the entropic resting force generated:

$$\mathrm{F} = \frac{k_B \times \mathrm{T}}{L_p} \times \left[(0.25 \times (1 - \frac{\mathrm{x}}{L_c}))^{-2} - 0.25 + (\frac{\mathrm{x}}{L_c}) \right]$$

Equation 1.1: WLC model. Where F is force, T is the temperature, k_B is the Boltzman constant, L_c is the contour length, L_p is the persistance length (which quantifies the stiffness of the polymer) and x is the measured extension[37].

Fitting WLCs to the successive unfolding events thus allows the difference in length between folded and unfolded states (ΔLc) to be calculated. These results for unfolding distance and length are taken from many unfolding events and histograms generated which, by means of a Gaussian fit to the data, can be used to obtain the most probable value for ΔLc and for unfolding force. The expected ΔLc is easy to calculate if we know the number of amino acids in the

folded polypeptide and the structure is known. This value is compared to the measured value to satisfy ourselves that the events we are recording are the unfolding of the domains themselves and not spurious tip-protein artefacts. The more interesting value we can extract is the peak unfolding force (F_{UN}), which gives us a measure of the mechanical stability of the domain in question.

Constant force or force-clamp experiments can also be carried out. In this case a feedback system is employed to keep the force applied to the cantilever at a constant level. The protein unfolds when random thermal fluctuations allow the protein to traverse the barrier to unfolding (G^{*TS^*-F}, Figure 1.1). This reduces the force applied to the cantilever, and the piezo controller moves the tip away from the surface (shown as extension in Figure 1.5) until the force is restored, whereupon after a certain time thermal fluctuations will cause another domain to unfold. In this case, instead of the saw-tooth pattern seen in constant velocity experiments, a step pattern is seen by plotting extension against time (Figure 1.5). L_S is the increase in contour length at the applied force upon the unfolding of a single domain, which is distinct for each domain and the time taken for each unfolding event is recorded. If such traces are overlaid, a rate constant can be obtained by fitting directly to the data.

In terms of experimental design both constant force and constant velocity experiments are very similar. Both require the use of a polyprotein in order to both increase the number of events analysed and to ensure that each event is part of a single molecule interaction, both require the selection of a protein concentration that gives a balance between the rate of data acquisition and the likelihood that each event is a single molecule event and both are generally

done in a simple buffer with a pH and ionic strength close to “physiological” values such as phosphate-buffered saline[39].

In terms of the data obtained from each type of experiment, constant force experiments directly measure the extension of the protein over time. Many repetitions of this data collection are repeated and this allows the direct measurement of unfolding rate. Performing the experiment at various forces enables the calculation of $k_u^{(0F)}$ and x_u by fitting to Equation 1.2.

$$k_u^{(F)} = k_u^{(0F)} e^{\frac{F x_u}{k_B T}}$$

Equation 1.2: Where $k_u^{(F)}$ is the observed unfolding rate, $k_u^{(0F)}$ is the unfolding rate at zero applied force, F is the applied force, x_u is the distance between the native well and the unfolding transition state, k_B is Boltzmann’s constant and T is the temperature[39].

Constant velocity experiments allow direct measurement of unfolding force, they also allow direct measurement of the distance from the surface at which unfolding occurs but this is less useful than the contour length which is obtained by fitting to the raw data using Equation 1.1. Taking multiple measurements at various loading rates and plotting the most likely unfolding force against the logarithm of the loading rate then allows the calculation of $k_u^{(0F)}$ and x_u[39].

Data must be chosen carefully for analysis as not every event produces a perfect force profile. Sometimes more than one protein can become attached, resulting in a profile with more peaks than expected or unusual peaks resulting from tip-surface interactions or failure of a protein to cleanly detach from the tip. Ideally only “perfect” traces would be analysed, in reality few such traces exist and some analysis must be undertaken on shorter, less ideal traces to facilitate the acquisition of a statistically relevant data set in the available time.

It is important to choose the analysed traces in a consistent manner in order to minimize systematic bias[36].

There are several parameters we can obtain from mechanical unfolding experiments. As well as a simple measurement of unfolding force. It is possible to probe the energy barriers of unfolding. Plotting unfolding force against the log of loading rate allows us to calculate x_u, the distance between the unfolding barrier and the native energy well and $k_u^{(0F)}$, the unfolding rate constant in the absence of applied force, which is related to the height of the unfolding barrier. x_u and $k_u^{(0F)}$ remain the same while the same unfolding transition is being probed, but if mechanical unfolding covers a sufficiently wide range of forces, then deviations from these parameters can be detected giving us insight into "hidden" unfolding transition states[40].

These hidden transition states are not visible by other techniques such as chemical denaturation. They can be identified by examining a plot of unfolding force against the logarithm of loading rate. If a single barrier to unfolding is traversed then the plot produces a straight line. However, when one of these hidden transition states is present then more than one force regime (i.e. lines with different gradients) will be evident in the data. This is because upon the application of force, the "tilting" of the energy landscape results in a previously hidden barrier (close to the native well) becoming the new rate-limiting step[41]. One example of this is the dissociation under force of the colicin E9:Im9 complex which is understood to take place under the same theoretical framework. Figure 1.6 shows that there are two different force regimes in the data. At low, biologically accessible, loading rates the unbinding force is low, and has a small dependence on the loading rate. At higher loading rates the

unbinding forces are higher and have a stronger dependence on the loading rate. This indicates that the dissociation follows a three state mechanism involving a previously unknown “hidden” transition state[35].

A similar result is found when using AFM to investigate streptavidin-biotin binding, the data shows a similar result where two force regimes are evident in the data. Analysis of the data and comparison to a molecular dynamics simulation identified the different force regimes observed were due to the presence of a previously hidden transition state that was only visible due to the tilting of the energy landscape[41].

Although force spectroscopy using AFM is a powerful technique for the study of protein unfolding it is not always a good model for the events which take place in a cell. It applies forces only with a fixed directionality, since the polyprotein is anchored to the surface at a set point (the exposed cysteine residues) and the force will be applied at the N and C-termini of the protein. Figure 1.7 illustrates the situation with a hypothetical three β-stranded-protein. In Figure 1.7A the forces are applied to the termini of the protein, if the initial unfolding event is the separation of the blue and green strands it is necessary to rupture all of the interactions between the two strands simultaneously which may require a large amount of force and is not how the protein would be expected to be unfolded in the cell. Figure 1.7B shows a situation where the protein is being pulled by the C-terminus into a biological unfoldase; as the protein is pulled though the pore into the unfoldase the opposing forces can act more locally and the strands can be gradually peeled apart negating the need for a large force to separate the entire strand. In addition the forces

applied loading rates applied by AFM may not represent those applied in the cell.

It seems that in order to obtain a more complete understanding of how real forces are applied in a real system we may have to consider using a complementary technique. While this method may not be able to provide data at the resolution that is possible with AFM, it would better represent the way forces are applied *in vivo.* We can do this through use of an enzymic system which applies force in the cell.

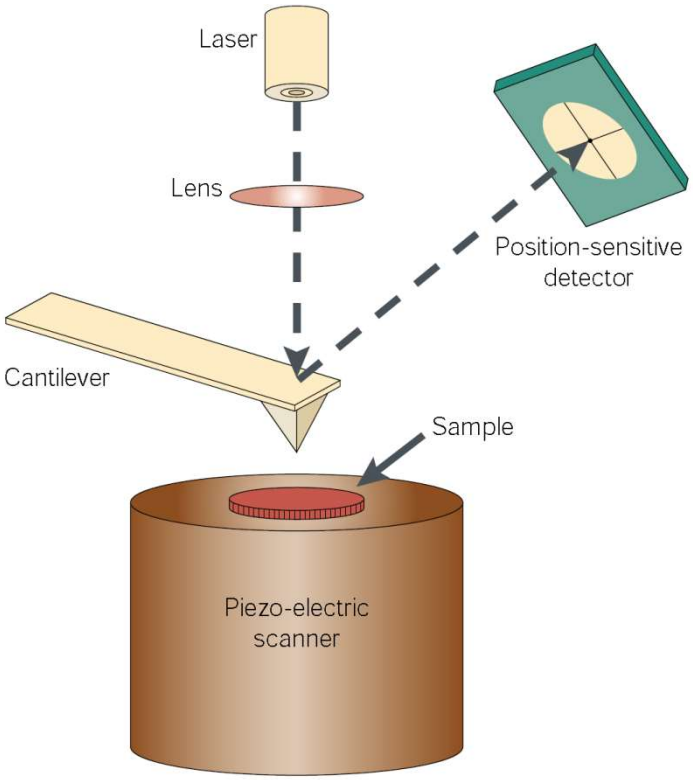


Figure 1.2: An illustration of the basic principle of atomic force microscopy. A laser is focussed onto the tip of the cantilever. Adjustments are made with the cantilever in an unloaded state so that the reflection of the laser is in the centre of a position-sensitive detector. Any deflection of the cantilever will cause the laser to move on the detector. Quantifying the distance moved allows the deflection of the cantilever to be calculated, if the spring constant of the cantilever is known this allows the force applied to the cantilever to be calculated. Adapted from Bustamante *et al*[25].

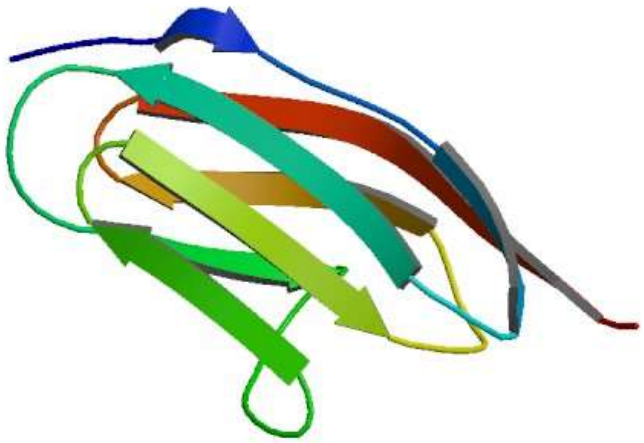


Figure 1.3: Titin, immunoglobulin repeat 27, NMR, minimized average structure. (I27) From PDB structure 1TIT[42]

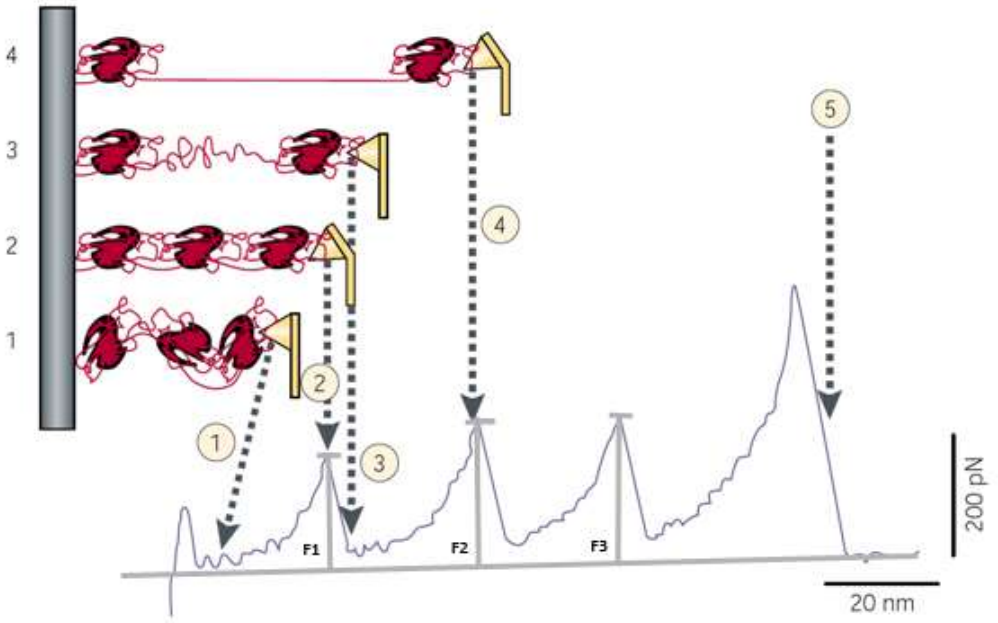


Figure 1.4: A force extension profile and diagram showing the interpretation of the profile for a constant velocity experiment. At point 1 the tip has adsorbed to the protein and is being withdrawn. As the tip is withdrawn the stretched protein exerts a force upon it until, at point 2, the force is sufficient to cause unfolding of one of the domains (F1) at point 2, at point 3 one domain of the protein has unfolded and the force acting on the cantilever is reduced, the process is then repeated as the protein is stretched until a second (point 4) and third unfolding event occurs (F2 and F3). The distance between points 2 and 4 gives the difference in length between the folded and unfolded states of the domain. At point 5 no domains remain to be unfolded and the force acting on the tip increases until the protein is pulled off the tip. Adapted from Bustamante *et al.*[26]

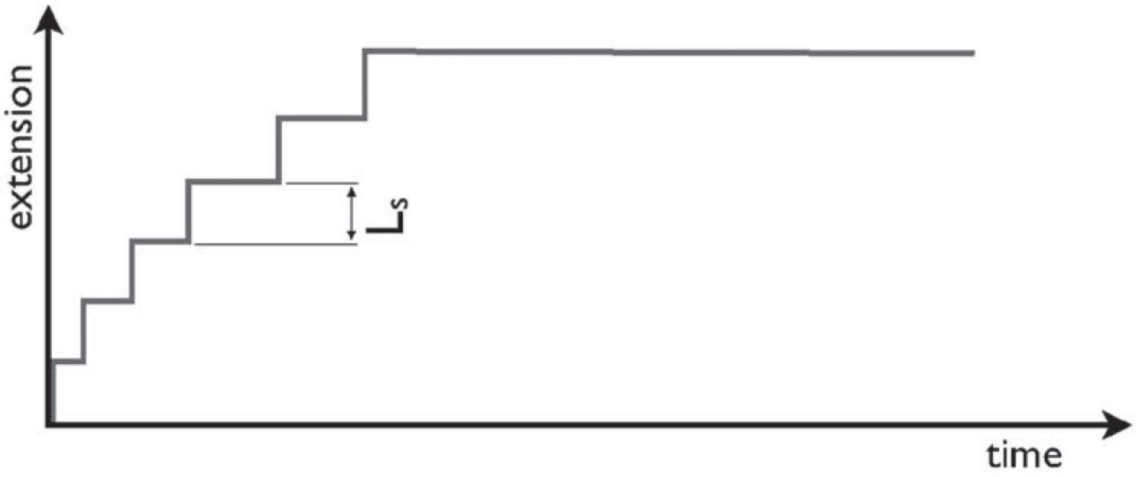


Figure 1.5: Plot of extension versus time for a polyprotein with 6 domains unfolded in a force-clamp experiment. From Dougan *et al*[32].

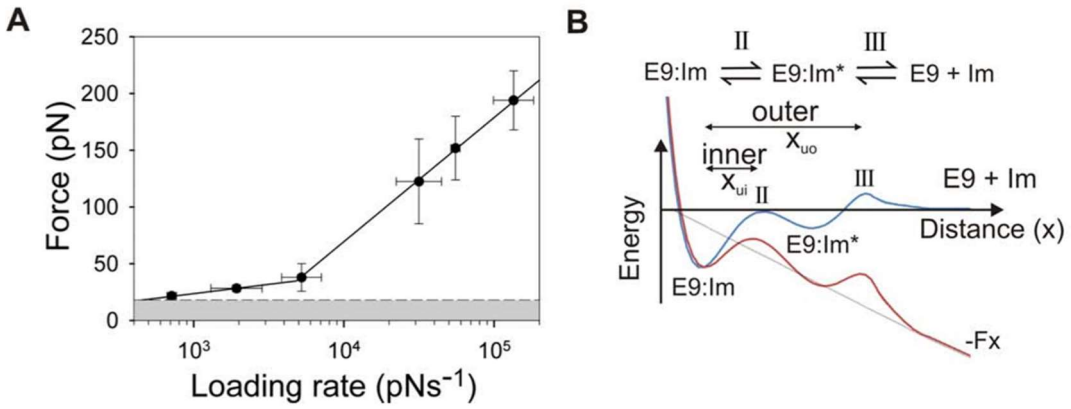


Figure 1.6: **A:** The dynamic force spectrum (a plot of the unbinding force versus the natural logarithm of the loading rate) of E9:Im9. Error bars are based on the standard deviation of measurements from triplicate datasets taken at each retraction velocity. Forces below the detection limit of the instrument and filtering software (18 pN, dashed line) are coloured grey. **B:** Postulated mechanism (top) and energy landscape (bottom) for the (un)binding of E9 from immunity proteins. The encounter complex (E9:Im*) is formed followed by the bound complex (E9:Im). Application of force causes the energy landscape to tilt (by –Fx) and the previously rate limiting step for unbinding (peak III, blue trace) is replaced by the previously hidden peak II on the red trace. Taken from Farrance *et al*[35].

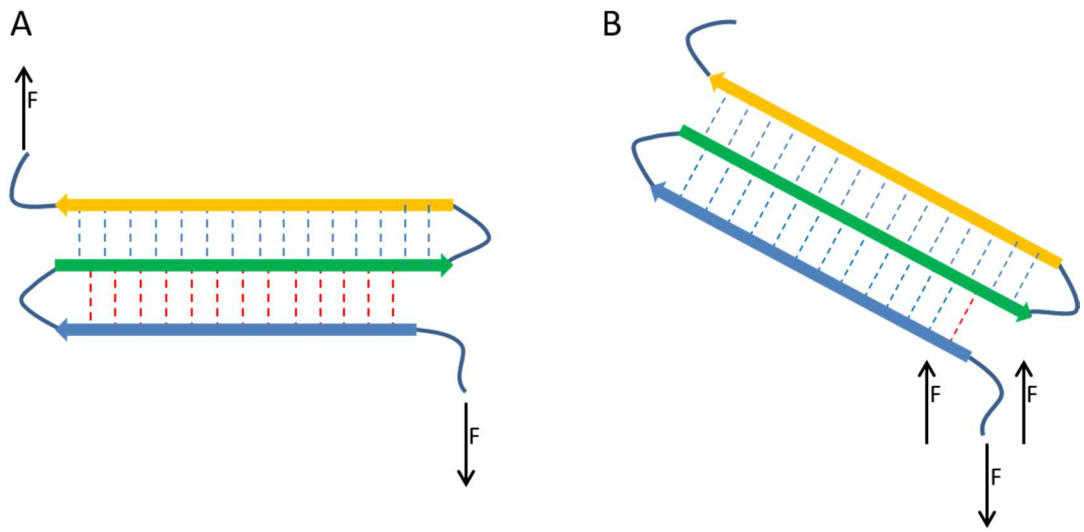


Figure 1.7: An illustration of force applied to a protein . **A:** by AFM. **B:** by a biological unfoldase. Black arrows indicate the locations at which forces are applied.

1.4 **ClpXP**

ClpXP is a member of the ATPase Associated with cellular Activity (AAA+) superfamily. These proteins contain a P-loop NTPase domain and have diverse functions such as molecular chaperones, helicases and ATPase subunits of proteases[43,44]. The ClpXP protease is a prokaryotic molecular machine found in several bacteria which has a similar structure and function to the eukaryotic 26S proteasome, with an ATPase subunit (ClpX) guarding the entrance to the chambered protease (ClpP)[45]. Its function is protein quality control and the degradation of a diverse array of unwanted proteins, it is responsible for degrading incomplete proteins on stalled ribosomes, degradation of secreted proteins which have failed to be secreted and can even function to degrade foreign proteins such as those from invading viruses[46,47].

ClpX is active as a hexameric ring with each of the 6 subunits containing an ATP binding site adjacent to the next subunit. ClpXP is made up of two stacked rings of heptameric ClpP oligomers to which a hexameric ring of ClpX is attached at one or both ends as shown in Figure 1.8.

ClpXP is a large complex; with the central ClpP protease and two ClpX rings, it is approximately 870kDa. ClpX is an unfoldase which recognizes a target protein by means of a peptide tag and unfolds it by means of pulling it through a small pore[49]. In the absence of ClpP, ClpX is able to remodel target proteins and their complexes, and to disaggregate and unfold aggregated proteins, allowing their refolding. It is more commonly associated with the chambered protease ClpP which degrades the protein as it is fed into the proteolysis

chamber by ClpX and releases it as short peptide fragments[50]. The action of ClpXP is summarized below (Figure 1.9).

It has been shown experimentally that ClpX is capable of translocating and degrading a large variety of substrates. For example, synthetic peptides with D-amino acids, altered peptide bond spacing, low complexity sequences consisting of polyglycine, polyproline sequences, sequences consisting of all positively or negatively charged residues, sequences with large aromatic side chains or up to 10 polyglutamine residues have been successfully translocated[51].

Unfolding and translocation by ClpX is powered by ATP hydrolysis, each power stroke is generated by the hydrolysis of a single ATP within one of the subunits, it is a robust system in that ATP hydrolysis does not proceed in a sequential manner and hydrolysis in any of the subunits in contact with the substrate peptide can drive translocation, preventing the stalling of the complex by any subunit being unable to hydrolyse ATP[53]. The hydrolysis of ATP causes a conformational change in the ClpX which is transmitted to the substrate as mechanical work by a tyrosine residue in the pore loop[54]. Once the ssrA tag and any unstructured regions have been translocated into the ClpXP pore any folded domains must be unfolded before further translocation can take place. Unfolding can be an energy intensive process. A previous study investigated the kinetics of ClpXP degradation of ssrA-tagged I27 and compared the amount of ATP used for unfolding to that required for translocation of the degraded substrate. Wild-type protein required up to 550 molecules of ATP for unfolding, whereas the same protein containing a destabilising V13P mutation required only 18 [55]. As ClpX has no way to store

energy from one unfolding attempt to another, each cycle of ATP hydrolysis represents a new attempt to unfold the protein. As each ATP hydrolysis event produces the same amount of force to unfold the protein, each unfolding attempt has the same probability of causing unfolding. Despite this, more thermodynamically stable proteins can still be unfolded by repeated application of this same force, suggesting that some change must occur in the substrate to allow it to unfold. If each unfolding attempt caused some slight change in the substrate's structure, perhaps some sort of partial unfolding, then the likelihood of unfolding with each ATP hydrolysis event would increase over time. As this is not the case, the most likely explanation is that the more stable substrate is unfolded only when an ATP hydrolysis event that lowers the barrier to unfolding coincides with a random thermal fluctuation that allows traversal of the lowered barrier[50,55,56].

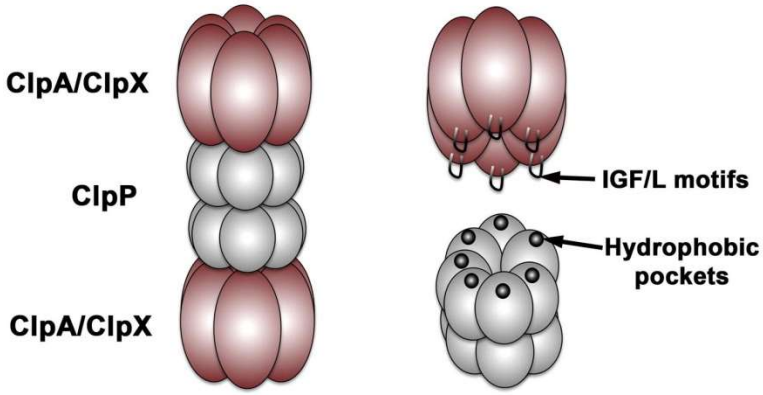


Figure 1.8: A diagrammatic representation of the ClpXP macromolecule showing the ClpP subunit in grey and the ClpX in brown. From Alexopoulos *et al*[55].

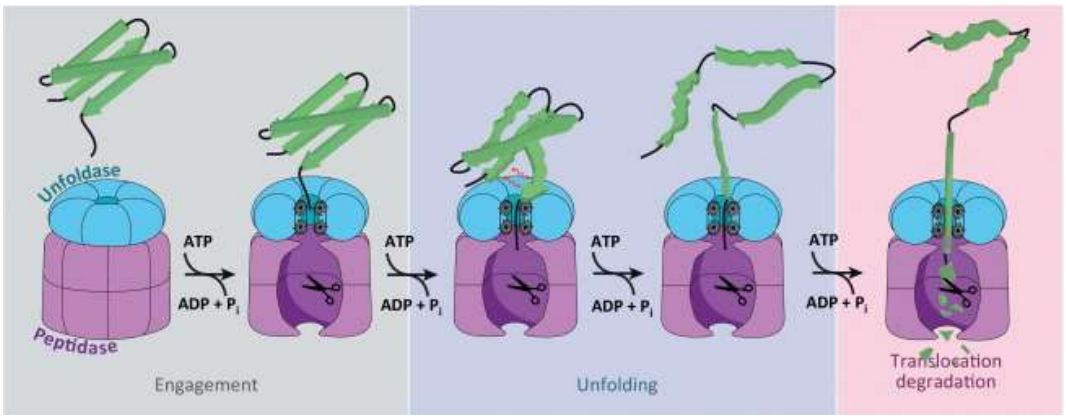


Figure 1.9: Cartoon model of substrate recognition and degradation by a AAA+ protease. The recognition step is mediated by binding of a peptide tag on the protein substrate to the AAA+ ATPase. The protein is unfolded, translocated into the compartmental peptidase and degraded. Peptide fragments are shown diffusing out of the peptidase, but active participation of the ATPase may be required for exit of large fragments. This figure is taken from Nyquist and Martin 2014[56].

1.4.1 ClpX

ClpX is a functional hexamer. It forms a ring structure in which each of its ATPase active sites is situated between two of the subunits. The central pore is lined with loops which, by conformational changes between the small and large domains of each monomer brought about by the hydrolysis of ATP, can apply a mechanical force to a peptide. The crystal structure of ClpX was solved to 2.8 Å resolution in 2003 by Kim and Kim[57] (Figure 1.10).

The wild type ClpX monomer is made up of three domains (Figure 1.11). There is an N-terminal domain (NTD) which is not necessary for degradation of all substrates as it folds independently of the other domains and is attached by a flexible linker. Each pair of adjacent NTDs form a dimer stabilised by the coordination of a zinc ion[58]. These are unnecessary for the degradation of all substrates and are instead thought to be involved in the recognition of some substrates and adaptor proteins and to contribute to the stability of the hexameric ring[59]. Deletion of the NTD (ΔNClpX) still allows hexamer formation, binding to ClpP and the degradation of certain substrates with near wild-type efficiency[60].

On formation of the hexamer each small subunit is adjacent to the large subunit of its neighbouring monomer. These interactions are stable and each small and large pair can be thought of as a single rigid body which is connected to its adjacent rigid body by a flexible hinge region which is also where the ATP binding and hydrolysis occurs[61]. The ring conformation is determined by the angles between these rigid-body units and the conformational changes in the hinge region[47]. Because of the stability of the ring, the arrangement of subunits relative to their topology interact and ability

to delete the NTD it is possible to construct ClpX variants in which the six monomers making up the hexamers are joined by a peptide linker, as pseudodimers, pseudotrimers and pseudohexamers to form a ring that is made up of three, two or a single covalently linked subunit[47] (Figure 1.12).

These constructs allow the structure and function of ClpX to be studied in new ways. The effects of mutations in a single subunit at a known position or several known positions can be studied without the uncertainty that mixing more than one variant of the monomer entails[61].

Previous crystal structures of ClpX have been obtained by determining the crystal structure of a single ClpX monomer and using the electron density obtained from cryo-EM to fit the monomers into a plausible hexameric structure[57]. More recently, the development of the covalently linked variants described above have been used to determine crystal structures of ClpX ring hexamers which showed that each rigid-body unit could exist in one of two different conformations: a “loadable” subunit orientated in a manner that allowed ATP binding and an “unloadable” conformation where rotation of the domains had removed the nucleotide binding site[62], resulting in an asymmetric structure in the ring as shown in Figure 1.10.

While it is still not known whether ClpX normally hydrolyses ATP in a sequential or a stochastic manner in relation to the level of co-ordination between the individual subunits of the hexamer during ATP hydrolysis, it is known that subunits have to be able to convert between the loadable and unloadable conformations for robust activity and that binding of ATP to one of the subunits causes allosteric changes in the ring that make binding of further ATP more favourable[61].

The pseudohexamer ($^{\Delta N}ClpX_6$) is also useful because it makes the process of using ClpX as a tool to study the mechanical unfolding of other proteins a simpler proposition, both because it can be tethered at a single anchor point for experiments using AFM or optical tweezers[30], and because it eliminates some of the issues that can arise when using wild type ClpX for *in vitro* experiments as shown in Chapter 3.

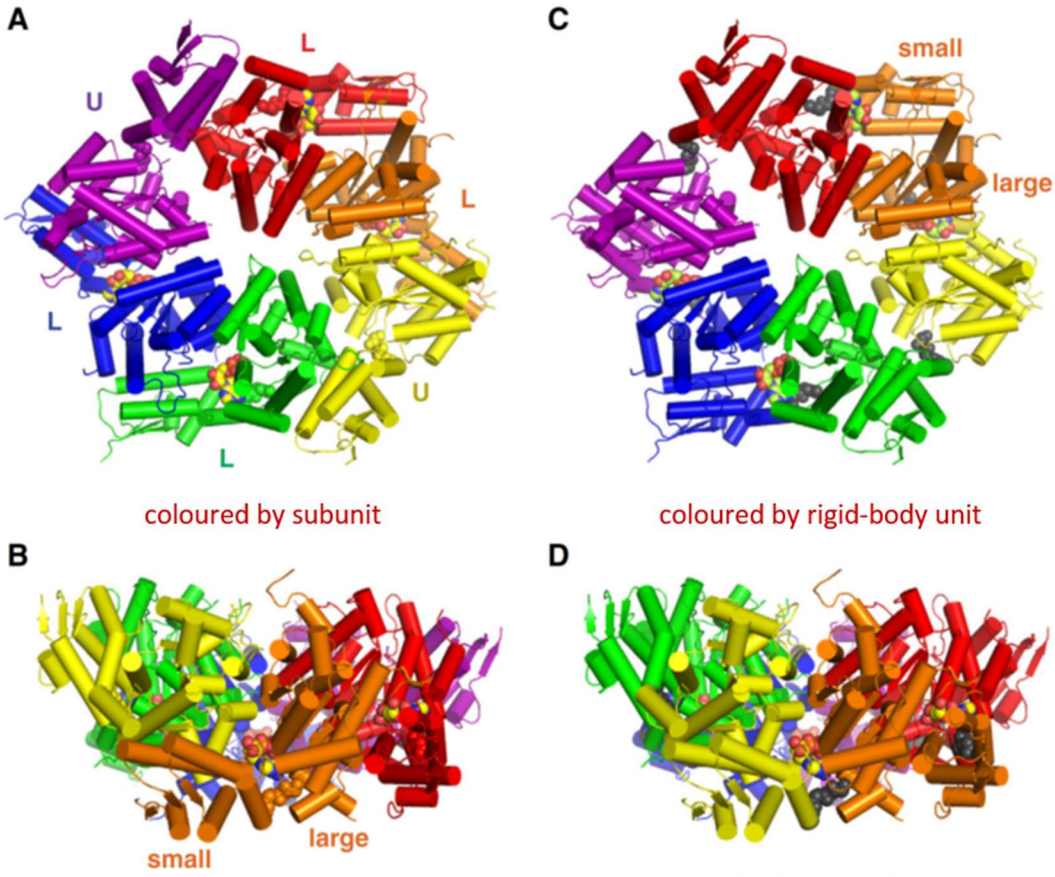


Figure 1.10: ClpX hexamer structures (3HWS). **A:** Face view (substrate side) showing loadable (L) and unloadable (U) subunits. **B:** Side View coloured by subunit. **C:** Face view (substrate side) coloured by rigid-body unit. **D:** Side view coloured by rigid-body unit. Adapted from Baker and Sauer[47] The atomic-resolution structure is for a pseudo-hexamer.

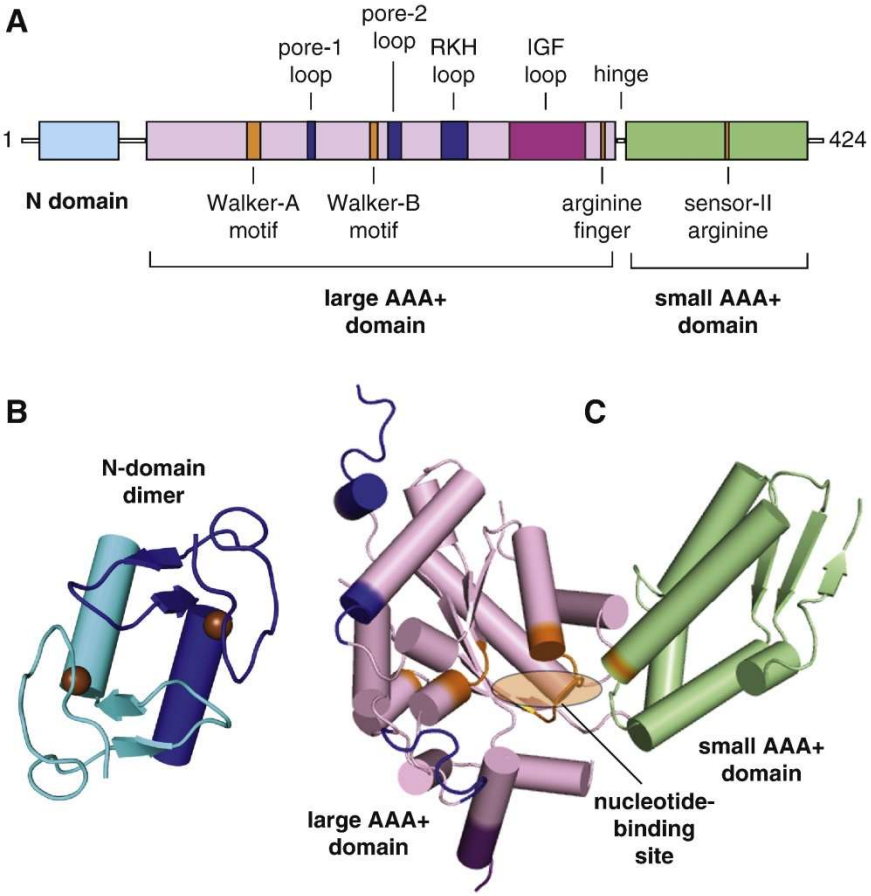


Figure 1.11: The domain structure of *E.coli* ClpX. **A:** Arrangement of the domains showing the functional motifs. **B:** Structure of the NTD dimer (1OVX) showing zinc ions (orange spheres). **C:** Structure of a single ClpX subunit, taken from a hexameric structure (3HWS) showing the nucleotide binding site. Colours represent: blue - ssrA tag binding region. Orange: ATP binding and hydrolysis. Purple: ClpP binding. Taken from Baker and Sauer[47].

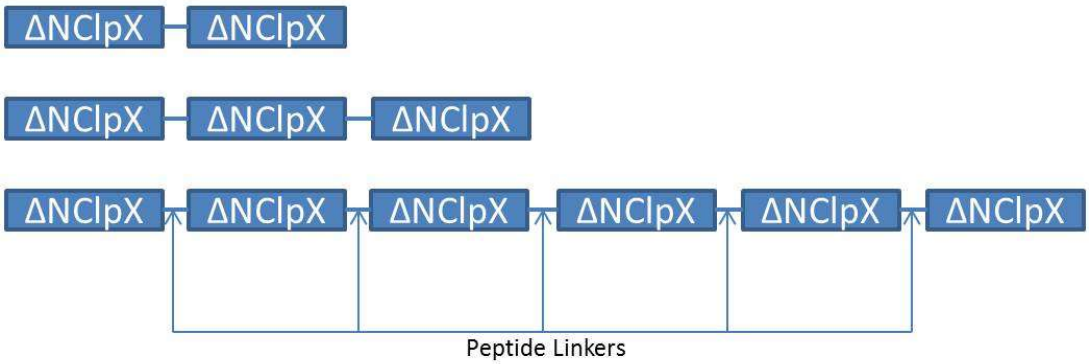


Figure 1.12: A schematic demonstrating **A:** pseudo-dimer. **B:** pseudo-trimer. **C:** pseudo-hexamer.

1.4.2 Substrate targeting

ClpX can be considered a promiscuous enzyme as it has the ability to unfold and translocate a wide range of substrates as long as they have a short, unstructured, peptide sequence known as a degradation tag to which the ClpX can bind. This reflects its main function, protein quality control. The best studied of the degradation tags is the ssrA tag which is responsible for the targeting of polypeptides from stalled ribosomes to ClpX where they can be unfolded and translocated into ClpP for degradation releasing both the ribosome and the amino acids for re-use by the cell. As shown in Figure 1.13, SsrA is an RNA molecule which functions as both a transfer and messenger RNA. In this molecule, the tRNA-like domain is charged with an alanine which is added to the end of the peptide in the stalled ribosome by transpeptidation. The mRNA in the stalled ribosome is then released and a short open reading frame encoded by the SsrA is used to add a decapeptide degradation sequence (ANDENYALAA), known as an ssrA tag which targets the peptide for degradation by ClpXP[63]. Studies have shown that it is the final three amino acids (LAA) of the peptide and the negatively charged C-terminus that is important for recognition and binding to ClpX and that the other residues are involved in interaction with the SspB adaptor protein[55,64].

In addition to the degradation of truncated peptides from stalled ribosomes, ClpXP carries out several other functions including the regulation of cell stress proteins, such as RecN synthesised in response to DNA damage. Under this condition RecN is expressed at a sufficient level to allow an increase in its intracellular level despite them being targeted for constitutive degradation by ClpXP[66]. This means that they can carry out their functions while the stress is

on-going but are rapidly degraded once their synthesis is stopped. Degradation tags have also been discovered in large macromolecular complexes, such as ribosomes, which are buried within the structure when the complex is properly assembled but exposed if the complex is misfolded or damaged[64].

In addition to ssrA-tagged proteins, at least five other signal motifs located at both the N and C termini of peptides target proteins to ClpX[50,64]. The λO tag A Type 1 (polar-T/φ-φ-basic-φ) N-terminal motif[64] (TNTAKILNFGR) is used in this thesis. This tag originates from the enterobacteria phage λ where it is found at the N-terminus of the replication protein O. λO-tagged substrates are not well degraded by $ClpX^{WT}$ or $ClpX^{\Delta N}$. However, it is possible to introduce mutations onto the pore-loops of ClpX that will change its substrate preference from a ssrA-tagged substrate to a λO-tagged substrate by reducing the positive charge of the RKH loop making interactions with the positively charged N-terminus more favourable[67]. ClpX contains three different sets of pore loops (Figure 1.14). The RKH loops closest to the substrate side of the ring are involved in recognising the negatively charged carboxylate group. Pore 1 or GVYG loops and pore 2 loops have also been shown to be important for substrate recognition both by cross linking of the ssrA tag to both of these loops (showing an interaction can take place[68]) and by mutagenesis of the loops which alter the K_M and affinity for the ssrA tag[54,69,70]. It is also possible to take the human ClpX, which does not recognise ssrA-tagged substrates, and alter its specificity to recognise the ssrA tag by transplanting the *E.coli* loops into the human ClpX[68].

In addition to the pore-loops it has been shown that other elements of ClpX are important for substrate recognition, particularly for the less well studied degradation tags such as λO, which requires the NTD for recognition[71].

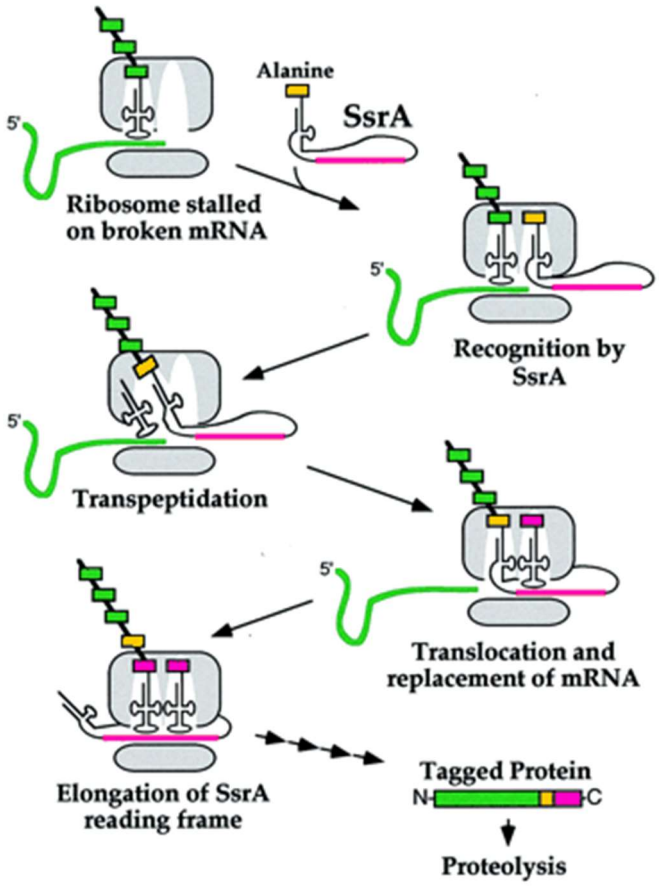


Figure 1.13: Upon stalling of the ribosome, SsrA is recruited to the A site of the ribosome. The nascent chain is transferred to the alanine-charged tRNA domain of the SsrA. The faulty mRNA is then exchanged for the SsrA open reading frame (magenta) by a message-switching event. Translation then continues until a stop codon is reached when the protein is released from the ribosome and can be degraded. Adapted from Karzai *et al*[71].

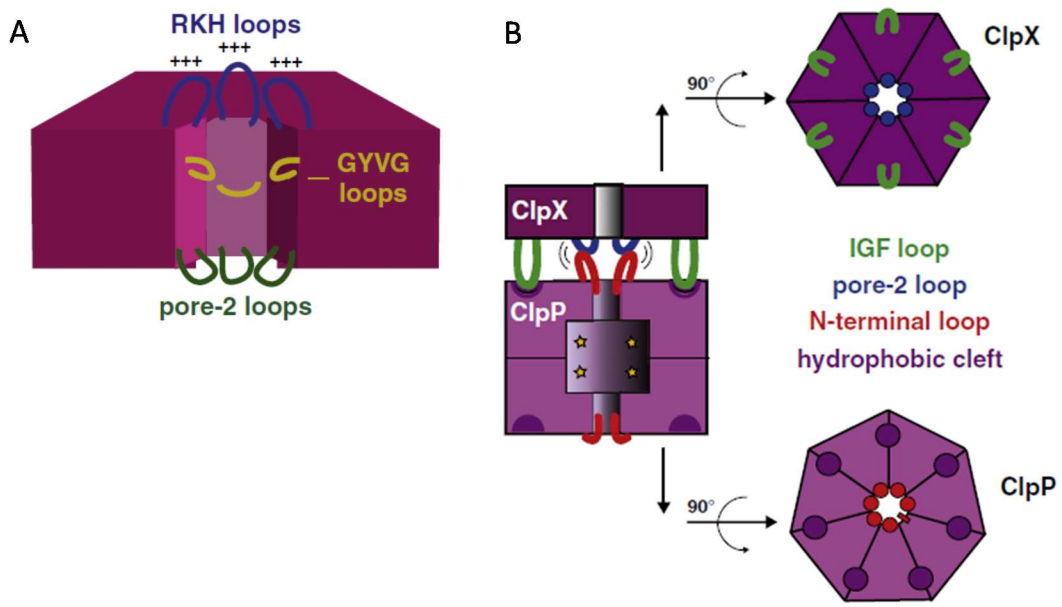


Figure 1.14: The pore loops of ClpX. **A:** Cartoon of the 3 types of ClpX pore loop involved in substrate recognition. **B:** Cartoon showing the assembled ClpXP complex. Adapted from Baker and Sauer[47].

1.4.2.1 Adaptor-mediated recognition

There are two adaptor proteins (RssB and SspB) known which can modulate the activity of ClpX by altering substrate delivery. RssB is a two component regulator of *E.coli* transcription initiation factor σ^s [64]. The better understood SspB binds to the NTD of ClpX and to the ssrA-tagged substrate, increasing the rate at which the substrate is degraded by preventing the substrate from diffusing away after an unsuccessful degradation attempt. It is not required for degradation to occur and does not affect the mechanism of degradation or change the substrate specificity but alters the overall rate of ClpXP mediated degradation. As it does not alter the mechanism it does not alter the kinetics of the individual steps in the unfolding and degradation of a substrate, but it does increase the overall rate of degradation in a bulk solution experiment by raising the effective substrate concentration[72]. It is interesting that SspB is capable of mediating the degradation of various substrates which have little sequence similarity and bind the adaptor differently[73].

1.4.2.2 Substrate translocation by ClpX

The exact mechanism of ClpX translocation is not known. However, as ATP binding and hydrolysis cause conformational changes in the ClpX ring it is likely that this conformational change drives translocation[47,61,74]. It has been demonstrated that, in addition to disrupting adaptor recognition, mutations in the GYVG loops can have an effect on ClpX translocation. In particular the Y153A mutation (which removes the tyrosine from the GYVG loop) completely prevents ClpX translocating its substrate[54]. Figure 1.15 shows a hypothesised situation in which movement of the loop tyrosine residue drives substrate

translocation. This GYVG loop is also thought to be responsible for preventing backtracking and loss of the substrate from the pore[54], although there is no conclusive evidence of this since mutations in this loop also affect the translocation of the substrate.

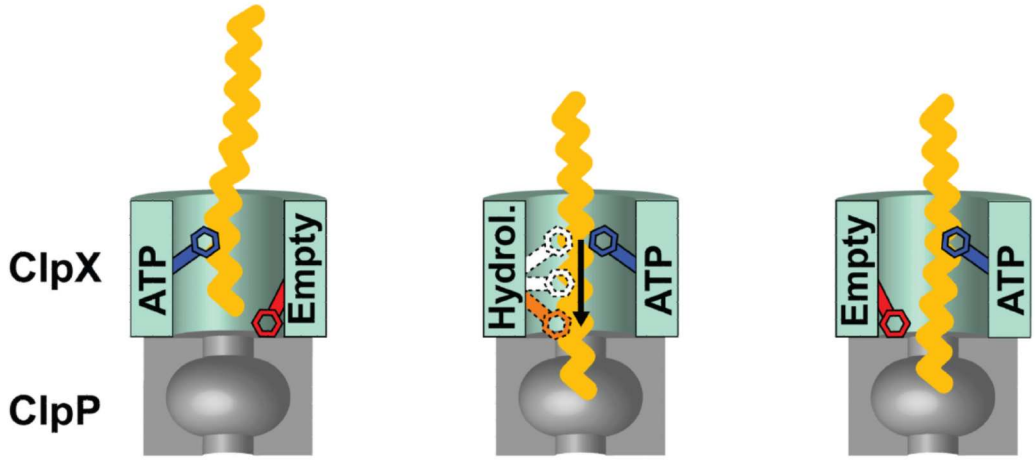


Figure 1.15: Cartoon representation of a possible mechanism of translocation by ClpX cycling between ATP bound (ATP), hydrolysing (Hydrol.), and empty states during substrate translocation. When ClpX is in the ATP-bound state the tyrosine residue is oriented towards the open side of the ClpX (blue) during ATP hydrolysis the tyrosine moves from being oriented towards the open side to being oriented towards the ClpP (white/orange) when there is no ATP bound the tyrosine is at the maximum orientation to the ClpP side of the ClpX (red). Adapted from Martin *et al*[54].

1.4.3 ClpP

ClpP is active as a 14-mer, its 14 subunits each have one serine protease active site and are arranged as two symmetric rings each of 7 subunits stacked face to face, producing a chambered peptidase with the active sites sequestered in a proteolytic chamber. Entry to the proteolytic chamber is through a narrow axial pore wide enough to accept only an unfolded peptide chain[75]. The crystal structure of ClpP was solved to 2.3 Å in 1997 by Wang et al[75]. The structure is shown in Figure 1.16 below.

In isolation ClpP does not degrade folded proteins, it is reliant on delivery of its substrates by an associated AAA+[50]. This lack of activity of free ClpP is important for cell viability: acyldepsipeptide antibiotics kill bacteria by allowing ClpP to degrade unfolded peptides without regulation[76]. This control of degradation is achieved by the first 21 residues of each ClpP subunit which form a stem-loop structure effectively blocking the channel by which proteins enter the degradation chamber unless it is interacting with ClpX. The pore into the proteolytic site of ClpP has a diameter of only 9 Å at its narrowest point. Binding of the aforementioned acyldepsipeptide antibiotics or an AAA+ ring to ClpP can allow entry of larger substrates allowing ClpP to degrade unfolded peptides including those containing disulphide bonds[77,78].

There is also a much reduced activity by single ClpP subunits and single ring heptamers which in theory have exposed active sites. The cleavage of peptides by ClpP does not seem to be sequence specific although cleavage after non-polar residues is preferred[79,80].

The inner chamber of ClpP is approximately spherical and had a diameter of around 50 Å, the 14 active sites it contains have an equivalent concentration of about 350 mM which allows it to degrade even sub-optimal substrates very efficiently. The sites are located such that each one is about 25Å from three other sites, a distance spanned by around eight residues in an unfolded substrate allowing multiple binding and cleavage events to occur simultaneously. This helps prevent stalling of the ClpXP as the ClpP can degrade a peptide chain much faster than ClpX can translocate it[47].

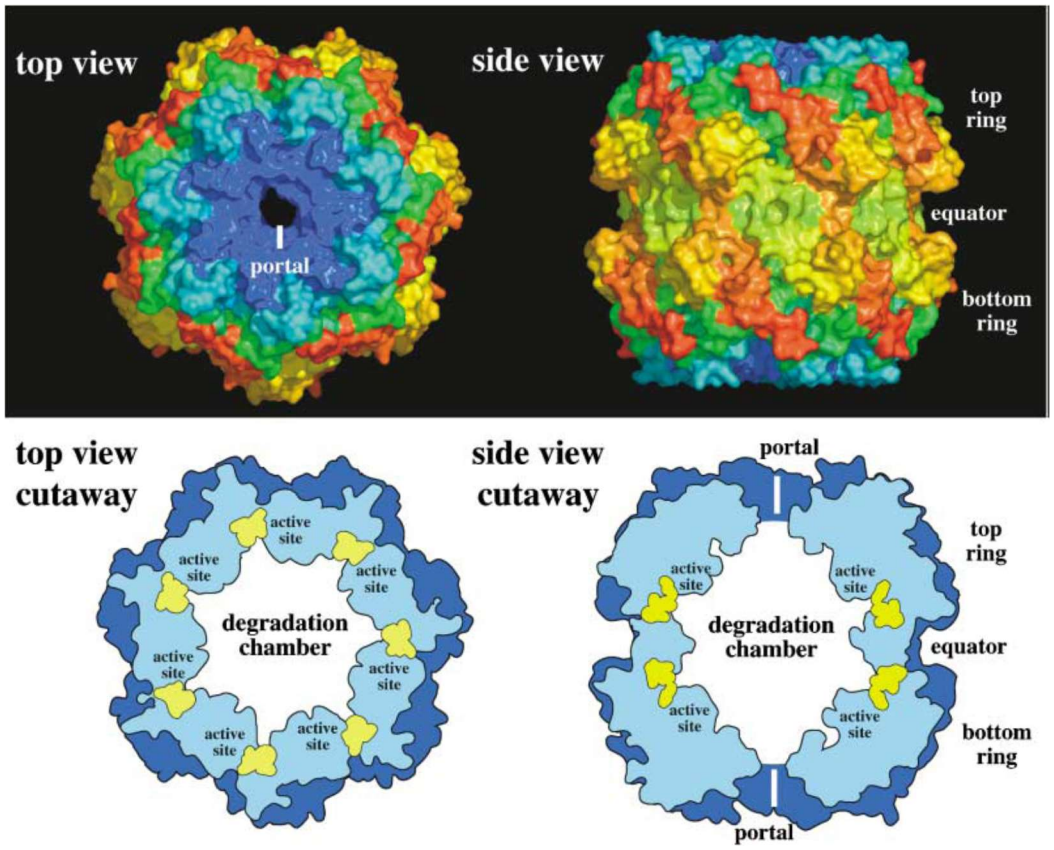


Figure 1.16: Structural views of the ClpP peptidase. Top: each of the 14 identical subunits is shown, the colouring is from red (C terminus) to blue (N terminus) for each subunit. Bottom: cutaway views showing the location of the protease active site residues within the chamber. Taken from Sauer *et al*[50].

1.4.4 ClpX and ClpP interactions

EM images have shown that ClpX can bind to either one or both sides of a ClpP 14-mer (Figure 1.17). The axial pores of ClpX align with the pores in ClpP providing a route for substrate translocation into the proteolytic chamber. Interestingly even in ClpXP complexes with two ClpX caps, only one of them will actively translocate a substrate at any given time, indicating some communication must occur through the ClpP[81,82].

As ClpX is a hexamer and each ClpP ring a heptamer there is inevitably a symmetry mismatch between the two different components. As no high resolution structures of complete ClpXP exist (only EM reconstructions) it is not known exactly how this mismatch is resolved, as can be seen in Figure 1.17 the resolution of the available EM images is too low to resolve this mismatch. It is known from mutagenesis studies that the pore-2 loops of ClpX and the N-terminal stem-loop residues of ClpP are involved, as is the binding of ATP but it is not known exactly how [69,83].

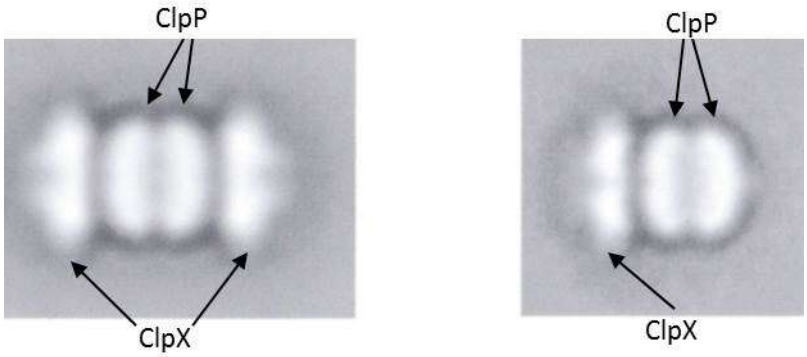


Figure 1.17: Cryo EM image showing the assembled ClpXP complex. Left :ClpXP with two ClpX rings, right: ClpXP with one ClpX ring. Adapted from Ortega *et al*[82].

1.4.5 Forces applied by ClpX

Two studies have been undertaken using optical tweezers to manipulate beads to which ClpX and a substrate have been tethered (Figure 1.18) in order to investigate the forces applied by ClpX[4,29]. While both studies used optical tweezers the precise experimental design differed. While both used one polystyrene bead to which ClpX had been tethered, and a second bead to which the substrate protein was attached at the N-terminus by a DNA linker, the substrate proteins were different. Maillard *et al*. used a fusion protein with an ssrA tag and either one or two copies of green fluorescent protein, Aubin-Tam *et al.* used an ssrA tag fused to 8 filamin A domains. The findings of these studies showed good agreement regarding the main detail of ClpX unfolding and translocation detailed in Table 1.1. They were successful in determining the force applied by the ClpX and to measure the translocation rates which show good agreement with earlier bulk solution studies[84].

The force applied by ClpX was measured up to 20 pN with a translation rate (approximately analogous to a loading rate) of 4.5 nm s^{-1} [4,30] which means that both the applied force and loading rate are at the lower end of what can be achieved by biophysical methods such as AFM or optical tweezers. Both sets of results supported the expected power stroke unfolding mechanism where ATP hydrolysis supplies the energy for a conformational change that provides a method of doing work on the substrate protein rather than some Brownian ratchet mechanism where the ClpX just provides a mechanism by which random Brownian motions of the substrate can only move in one direction. The studies also supported the hypothesis that ClpX will repeatedly pull on the

substrate and that unfolding will occur when this pulling coincides with random thermal fluctuations in the substrate protein. In this way substrates can be unfolded with a wide range of thermodynamic and mechanical properties.

The main disagreement between the two studies was the different level of unfolding intermediate seen in the unfolding events, but as each was using a different substrate protein this in unsurprising.

Table 1.1: Summary of the results of Maillard et al. and Aubin-Tam et al. regarding the mechanical unfolding of a substrate by ClpX investagated by optical tweezers.

Finding	**Maillard *et al*[4].**	**Aubin-Tam *et al*[29].**
20 pN stall force	Yes	Yes
Slower translocation at higher foces	Yes	Yes
1 nm minimum translocation step size	Yes	Yes
Pauses before unfolding	Yes	Yes
Unfolding intermediates detected	70% of events	15% of events
Peak work applied by ClpX	5 *k*T	5 *k*T

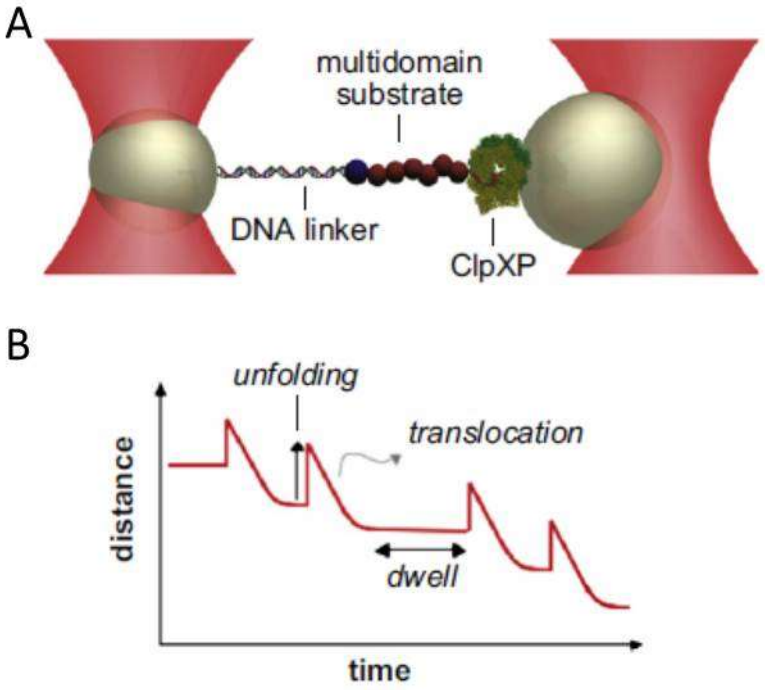


Figure 1.18: A: experimental setup of an optical tweezer experiment for measuring the force applied by ClpXP. Two polystyrene beads are held in an optical double trap with a passive force-clamp geometry. The multidomain substrate is attached to one bead by a DNA linker, ClpXP is tethered to the other bead and contact between the two beads is maintained by the contacts between the substrate and ClpXP. B: an idealised trace showing the typical data obtained from such an experiment. Increases in the distance (between the beads) represents an unfolding event, decreases in the distance represent substrate translocation and the dwell is the time required for unfolding of the next domain. Adapted from Aubin-Tam *et al*[30].

1.4.6 ClpXP as a tool to study mechanical forces

The degradation of substrates by ClpXP has several steps:

1. Productive binding of substrate by ClpX
2. Mechanical unfolding of the substrate by ClpX
3. Translocation of the substrate by ClpX
4. Degradation of the substrate by ClpP

It has been shown that mechanical unfolding of the substrate by ClpX is the rate-limiting step in a degradation reaction and that the rate of degradation depends on the mechanical stability of the substrate[49]. While it is not possible with the system I have used to directly measure the unfolding of a substrate, subsequent degradation of the substrate by ClpP provides an all-in-one linked assay as we can assume that as unfolding is the rate-limiting step then the rate of unfolding is proportional to the rate of degradation.

Consequently care must be taken when using the data from degradation experiments to infer the mechanical unfolding rate of a substrate. What we can measure using the assay developed here is the degradation rate of the protein and while these should be linked (as described above) it is important to realise that we are not measuring unfolding directly.

Most studies involving ClpXP have concentrated on understanding the mechanism of ClpXP itself. My project involves the use of ClpXP as a model system to study how different proteins react to the mechanical forces which ClpX can apply. The lack of substrate specificity (except for the requirement for a degradation tag) and the ability of ClpXP to degrade a wide variety of substrates make it an ideal model system for this study which investigates the

mechanical strength of proteins containing long stretches of polyglutamine repeats.

1.5 Polyglutamine Repeats

Polyglutamine (polyQ) expansion in proteins is caused by the genetic expansion of a CAG nucleotide repeat. It causes several hereditary, progressive neurodegenerative diseases. Since the identification and sequencing of the Huntington's disease (HD) gene in 1993[85] much research has been carried out to determine the mechanisms of these diseases. Briefly, polyQ disease proteins contain a polyQ region coded for by repeated CAG codons which become expanded in one of the genes of diseased individuals. While it is possible to inherit a gene with an already expanded CAG region, expansion can also occur during meiosis and mitosis. While the mechanisms of CAG expansion are understood, the pathogenic consequences of polyglutamine expansion are not, and neither is the structure of polyglutamine regions of proteins. The polyglutamine repeat disorders share several similarities in their pathogenicity and are thought to be caused by similar mechanisms, involving a toxic gain-of-function in the disease protein. Current therapeutic strategies have concentrated on relief of the symptoms of these disorders rather than treating the underlying condition. Gene therapy has gained popularity as a possible treatment; if the underlying genetic defect can be corrected or expression of the disease protein reduced, the condition can be treated without having to fully understand the mechanisms by which the toxic effect is produced. More traditional therapies such as small molecule inhibitors of toxicity or aggregation do not yet exist, but further study of polyglutamine structure and aggregation may allow their development. Recent studies have started to shed some light on the structure of polyglutamine repeats (see section 1.5.6) which may in time lead to the

effective development of more traditional therapies for polyglutamine expansion disorders. This project will attempt to elucidate more details about the mechanical properties of polyglutamine repeats using a traditional biochemical method of enzyme assays, exploiting the ability of ClpXP to apply forces to degrade proteins, and the biophysical technique of dynamic force spectroscopy, using an AFM to directly apply and measure the force required for the unfolding of polyglutamine and to investigate its unfolding resistance.

1.5.1 Polyglutamine function

Analysis of the human genome suggests polyQ-containing proteins are enriched in transcriptional co-activator and transcription factor binding proteins[86]. This leads to two major biological functions for polyQ containing genes; DNA dependent transcriptional regulation and neurogenesis. Interestingly, most eukaryotic proteins containing homo-peptide repeats are involved in nucleic acid interactions in some way; transcription, translation or direct binding to nucleic acids or chromatin[87].

1.5.2 Polyglutamine diseases

Polyglutamine expansion diseases, also known as CAG repeat diseases are hereditary neurodegenerative diseases caused by the expansion of a CAG repeat in the gene for a cellular protein leading to the expansion of a polyQ tract in the protein which causes a toxic gain-of-function[88]. To date there are nine known polyQ expansion diseases, Huntington's disease (HD), dentatorubropallidoluysian atrophy (DRPLA), spinal bulbar muscular atrophy (SBMA) and six types of spinocerebrellar ataxia (SCA1, 2, 3, 6, 7, and 17)[85] (see Table 1.2). The diseases, of which HD and SCA3 are the most common,

share some features indicating a common toxicity related to the expanded polyQ. They are autosomal dominant diseases and the toxic protein is normally expressed at the same level and alongside the normal length protein. Most of the proteins involved are expressed systematically but the toxic effects are seen only in neuronal cells, perhaps due to increased expression in these cells[89]. The accumulation of these proteins in neuronal cells seems to lead to their dysfunction and death. The diseases are progressive in nature, both in the symptoms presented and the accumulation of the protein. The specific pathogenic mechanism of polyQ diseases is complex and not fully understood and several models have been devised to explain it. None, however, have been conclusively proven, but an inverse correlation between repeat length and the age of onset of clinical symptoms has been observed, and the risk of suffering from a polyglutamine disease is increased in individuals with a long uninterrupted polyglutamine tract[89].

1.5.3 Length

As shown in Table 1.2, repeat lengths in both normal and disease proteins can vary considerably. Polyglutamine repeats in disease proteins are generally longer than ~35 Q; however in SCA6 it can be as low as 21 or, in SCA3, as high as 62. This indicates that length of the polyQ repeat alone cannot be responsible for the disease, as polyQ lengths that would be considered normal in one protein can cause disease in another, so the protein context of the polyQ repeat must be important[90].

Despite disease proteins being expressed throughout an individual's life, polyQ diseases are typically considered to have an adult onset, although increased length of polyQ repeat is associated with both earlier age of onset

and disease severity[92] thus childhood onset is possible yet uncommon. For example, onset of HD in juveniles (defined as before age 20) occurs in approximately 10% of cases, the majority of these having a repeat length of 60 or greater[93]. While infantile onset (before 1 year) occurs in ~1% of cases with typical repeat lengths of 80-90[94] and repeat lengths of up to 250 have been recorded[95].

Table 1.2: List of polyglutamine expansion diseases, with protein names and functions, showing the variation of polyQ length in normal and disease proteins[85,95].

Disease	*Protein*	*Protein Function*	*polyQ length*	
			Normal	**Disease**
SBMA	Androgen receptor	Testosterone-activated steroid receptor	6-36	38-62
DRPLA	Atrophin-1	Possible transcriptional repressor	3-38	49-88
HD	Huntingtin	Possible scaffolding protein linked to diverse cellular pathways	6-35	36-121
SCA1	Ataxin-1	Transcriptional repressor involved in transcription regulation, cell specification and synaptic activity	6-39	41-83
SCA2	Ataxin-2	Component of RNA processing and translational regulation pathways	14-32	34-77
SCA3	Ataxin-3	Deubiquitinating enzyme involved in protein quality control	12-40	62-86
SCA6	P/Q-type calcium channel subunit α1A	Voltage-dependent P/Q-type calcium-channel subunit	4-18	21-30
SCA7	Ataxin-7	Component of histone acetyltransferase complex and transcriptional regulation pathways	7-18	38-200
SCA17	TATA-binding protein	Component of core transcriptional complex TFIID	25-43	45-63

1.5.4 Pathogenicity

Studies have concluded that the expansion that occurs in polyglutamine disorders is responsible for a toxic gain-of-function rather than loss of the normal protein function[89]. However, the specific pathogenic mechanisms of each disease seem to be unrelated and it is unlikely that a common pathway will emerge[92]. The most likely candidate for a shared mechanism was the ability of expanded polyglutamine peptides to form aggregates and inclusions, since inclusions containing the relevant protein are found in the regions of the brain subject to neurodegeneration[96]. However it has been questioned whether these aggregates and inclusions are directly pathogenic[86,97]. It seems that the polyQ expansion causes a pathogenic mechanism related to a change in conformation of the normal protein resulting in a change in function[86,97]. This idea is supported by reports that overexpression of molecular chaperones can reduce polyQ toxicity in model systems[92].

There are many proposed models to explain polyQ toxicity and it is unlikely that any one will be sufficient to fully explain the diverse observations. Instead several of the proposed mechanisms are likely acting together in each of the diseases[91]. Table 1.3 summarises several of the proposed mechanisms.

Table 1.3: Proposed mechanisms of polyglutamine toxicity[86–88].

Proposed mechanisms of polyglutamine toxicity
Protein misfolding causing a toxic gain-of-function
Adverse protein interactions between the disease protein and the proteome
Formation of toxic oligomers
Transcriptional dysregulation
Mitochondrial misfunction impairing bioenergetics and causing oxidative stress
Impaired axonal transport
Abnormal neuronal signalling and excitotoxicity
RNA toxicity
Ubiquitin-proteosome system impairment

1.5.5 Therapy

Although current treatments for polyglutamine diseases in clinical use can treat only the symptoms, the hereditary nature of the diseases and the ease of identifying at-risk individuals through genetic screening make it a good target for therapy. As most current therapies concentrate on symptomatic relief[88] there is therefore a need to develop our understanding of the diseases at a molecular level in order to design effective treatments.

One therapeutic target may be the gene itself, turning off, or down-regulating expression of the mutant gene. RNA interference using siRNA would, in theory, provide a cure[98,99], however this approach is not without its challenges and a clinical application may be some years away.

If expression of the protein cannot be prevented another strategy is to enhance its clearance. Improving degradation by either autophagy or the ubiquitin proteasome system (UPS) may be effective, although as discussed later the UPS itself may be impaired by polyQ. A final strategy is to prevent polyQ forming toxic structures; however without a greater understanding of the structure of polyQ and its aggregation this will not be possible.

1.5.6 Polyglutamine structure

Polyglutamine peptides are examples of intrinsically disordered proteins (IDP), sequences that under non-denaturing conditions do not form an ordered tertiary structure. Polyglutamine is a typical IDP in that it has a low sequence complexity and a lack of hydrophobic residues[100]. It has been shown by several different methods that polyglutamine peptides form a random coil conformation in water. Circular dichroism spectroscopy (CD) indicates a high

degree of random coil conformation over several polyQ lengths, indicating no significant structural change that could account for the difference in toxicity of the different length polyQ tracts[100–102]. Nuclear magnetic resonance (NMR) also shows that polyglutamine adopts a random coil conformation[103].

Structural information about polyQ at an atomic level is very difficult to obtain: until recently the database of protein structures contained no structures for polyQ repeats greater than 3 [104]. Only one crystal structure at atomic resolution exists for a repeat greater than Q_{10} and this does not give a single definitive structure. This crystal structure, obtained by Kim *et al*[105], is of the first exon of huntingtin which containined 17 glutamine repeats located between the 17 amino acid N-terminal region and a polyproline rich region. Data was collected from 30 crystals, and a partial structure for the polyQ region was achieved for 7 of them. The N terminal region was shown to be α-helical, with the α-helix continuing into the polyQ region for between 1 and 12 glutamines before becoming a random coil. The 4 glutamine residues immediately before the polyproline region adopted an extended loop conformation[105].

Water is a poor solvent for polyglutamine peptides longer than 16Q[106] and they form a compact collapsed globular structure in water stabilized by extensive intra-chain hydrogen bonding[100,107]. Indeed, polyQ longer than 16Q does not remain in solution and quickly forms aggregates, but when in the context of a soluble protein or a peptide with sufficient charged residues, it can maintain solubility. The poor solubility of polyQ is counterintuitive given the hydrophilic nature of glutamine and the high solubility of small amides in water. Some change must occur between the small amides and the polyamide and it

seems that the polyglutamine backbone is more favourably solvated by interacting with itself than with the solvent water causing a collapse of the chain to minimize its exposure to the solvent[100,107]. This is thought to be due to the increased probability of the formation of intermolecular hydrogen bonds, which can be backbone-backbone, side chain-side chain or side chain-backbone, as the peptide length increases. Once several intermolecular hydrogen bonds have formed the peptide becomes trapped in a collapsed conformation and further interchain hydrogen bonds can form[106]. The disorder of the structure may lie in the fact that there are many potential hydrogen bond donors and acceptors in a polyglutamine sequence. Exploration of a polyQ sequence by use of molecular dynamics simulations (MD) shows there are a large number of hydrogen bonds able to form because they can form almost anywhere along the chain, in any of the accessible conformations. For example, a Q_{15} peptide in a compact or semi-compact globule can form an average of 4-5 sidechain-backbone hydrogen bonds[108]. To date, it has not been possible to obtain a definitive structure for expanded polyQ regions in context as a part of a native protein. A study using a polyQ/glutathione S-transferase (GST) fusion protein in order to look at the structure of polyQ tracts in a more native-like environment by both CD and NMR indicated a random coil conformation[104]. Most studies of polyQ structure, if they have considered the structure of the individual polyQ tract, have concentrated on how its structure facilitates its aggregation.

1.5.7 Polyglutamine aggregation

From both *in vitro* studies of peptides and *in vivo* observations it has been observed that polyQ can aggregate to form amyloid-like fibrils. Given that the

majority of polyglutamine expansion disorders are characterised by the formation of both cytoplasmic and intranuclear inclusions within neurons it seems clear that this aggregation is important to the pathogenesis of polyQ diseases. However, it is unclear whether these aggregates are directly responsible for toxicity, merely represent the end point of a toxic mechanism involving an aggregation intermediate, or even have a protective role[85]. It seems reasonable to assume that the proposed conformational change brought about by an increase in polyQ length which allows aggregation to occur has other effects which may cause toxicity.

It has been known for some time that aggregated polyQ has a β-sheet structure, which may be stabilized by a polar zipper of hydrogen bonds formed between the main chain and side chain amides; other suggested structures include parallel β-sheets and β-hairpins[109–111] these structures are summarised in Figure 1.19.

Aggregation appears to follow a nucleation-dependent mechanism. When fibril growth is observed *in vitro* there is a lag phase, representing the formation of a nucleus, followed by a rapid increase in aggregation; the lag phase can be eliminated by the seeding of a solution with pre-formed aggregate.

The length of the lag phase and the critical concentration for aggregation (the concentration at which it becomes more thermodynamically favourable for the peptide to form aggregate) was found to be inversely correlated with the length of the polyQ[112]. This likely explains the correlation of polyQ length with severity of disease and the inverse correlation with age of onset. Interestingly it has been suggested that in addition to aggregation of expanded polyQ

peptides, normal length polyQ peptides also play a role in enhancing aggregation kinetics[113].

Since polyQ is inherently disordered without a stable structure, we can assume that in solution it will sample several of the conformations available to it. One of these may be relatively extended and have some β-sheet character. This conformation, however briefly occupied, may form the nucleus for subsequent aggregation[101]. It may be that as polyQ length increases, the likelihood that it will occupy this conformation, or that part of it may occupy this conformation increases, explaining the increased aggregation of longer polyQ sequences. It has also been shown that the context of the polyQ expansion strongly affects the rate of onset of disease[90].

1.5.8 Protein context

The physico-chemical environment of the polyQ residues will be strongly affected by the adjacent through-space sequence of the protein in which it resides. Thus, in addition to the conformation and repeat-length, the characteristics of adjacent domains, posttranslational modifications, protein interactions, cellular location and normal function of the protein may all be important in the mechanisms of polyQ diseases and aggregation[90]. As shown in Figure 1.20 it may be that the flanking regions themselves form an initial aggregate that causes or accelerates polyQ aggregation by bringing polyQ regions into close proximity.

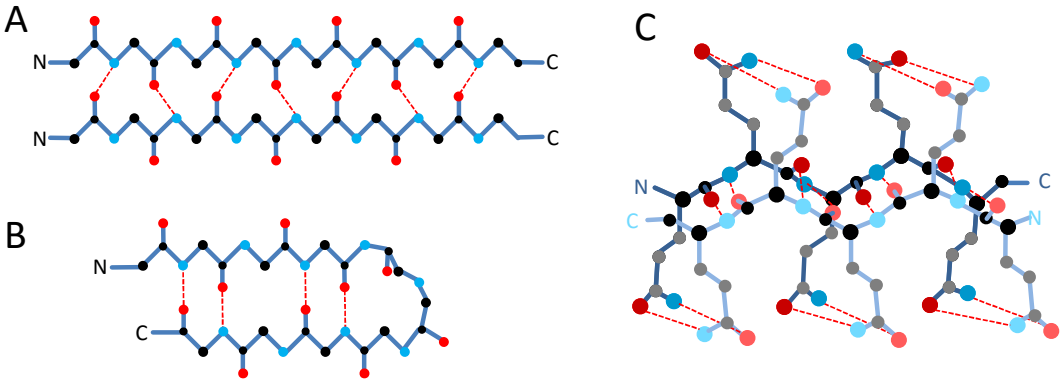


Figure 1.19: A summary of the possible structures for aggregated polyglutamine. **A** and **B** show only the backbone and backbone hydrogen bonding. They represent parallel β-sheets and β-hairpins respectively. **C:** shows a polyglutamine polar zipper formed by the hydrogen bonding between both the backbone and the side chains. Key: red: oxygen, blue: nitrogen, black: backbone carbon, grey: side-chain carbon.

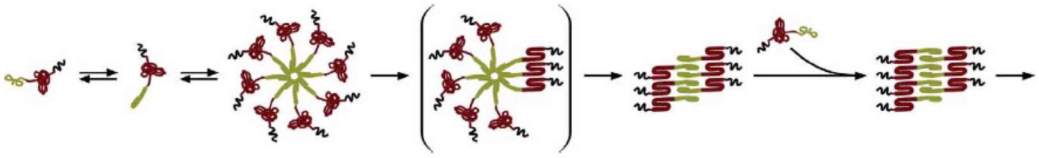


Figure 1.20: Mechanism of aggregation for huntingtin exon 1 (HTT^{NT}) mediated aggregation of polyQ. The HTT^{NT} domain (green) unfolds in a polyQ repeat length–dependent fashion and, once unfolded, self-aggregates without a nucleation barrier to form oligomers with cores comprising HTT^{NT} and not polyQ (red). The next identified aggregates involve both HTT^{NT} and polyQ in amyloid-like structure; the proline rich domain (black) is not incorporated into the core. This drawing is schematic and is not meant to imply any details of aggregate structure, except that final aggregates are rich in β-sheet, are fibrillar and involve both HTT^{NT} and polyQ. Taken from Thakur *et al*[114].

Table 1.4 shows the polyQ region of the known polyQ disease proteins together with the amino-acids adjacent to their primary sequence. It is immediately notable from Table 1.4 that not all of the “polyglutamine regions” contain exclusively polyglutamine. For example, ataxin-3 contains a lysine and Ataxin-1 contains a central histidine-glutamine-histidine sequence. This does not alter the random coil conformation of the polyglutamine tract, but it does have an effect on the aggregation kinetics as aggregation is slowed by reducing the nucleation equilibrium constant, K_{n^*}, which is calculated by fitting to Equation 1.3, and is approximately analogous to K_a[113,115]. This is thought to be due to the structure of His containing aggregates being more restricted due to the positioning of His residues in reverse turns[115].

$$\Delta = \frac{1}{2} K_{n^*} k_+^2 C^{n^*+2} t^2$$

Equation 1.3: where Δ is the slope of a log-log plot of the initial aggregation kinetics versus concentration, K_{n^*} is the nucleation equilibrium constant, k_+ is the second order elongation rate constant (calculated seperatley from the molar concentration of viable growth points in the seed fibrils), n* is the critical nucleus (n* = slope -2), C is concentration and t is time[115].

The presence of the H-Q-H motif may explain the longer polyQ region required for a pathological polyQ repeat in SCA1 [116]. Another notable feature is that many of the polyQ regions are flanked by homopeptide repeats or regions rich in a single amino acid; in particular, polyproline or proline rich regions, but polyalanine, polylysine and polyhistidine are all represented. It has been found that the addition of a polyproline sequence of 5-6 amino acids C-terminal to the polyQ region can reduce the rate of aggregation[117]. It is proposed that this occurs due to the structural constraints of polyproline, it favours a P_{II} helix that may influence the conformation of residues N-terminal to the polyproline, imposing a degree of order on the polyQ region[118,119]. Studies of exon 1 of

huntingtin (consisting of the 17 N-terminal amino acids (HTT^{NT}), the polyQ repeat and the polyproline region) have shown that as well as changing the aggregation kinetics, sequence context can also alter the aggregation mechanism. Instead of a simple nucleation-aggregation mechanism, the presence of HTT^{NT} causes the formation of an aggregation intermediate made up of aggregated unfolded HTT^{NT} which forms without having to overcome a nucleation barrier, followed by aggregation of the attached polyQ, forming the nucleus for further polyQ aggregation (see Figure 1.20)[114]. Studies of polyQ aggregation have mostly used polyQ peptides, whereas in the disease the focus has been on full length proteins. However, proteasomal degradation of polyQ containing proteins may release shorter polyQ peptides into the cell[114,120–122], making studies of polyQ peptides more biologically relevant than it first appears.

Table 1.4: The polyglutamine regions (in red) and surrounding amino acids of the known polyglutamine expansion disorders.

Protein	*UniProt Identifier*	*PolyQ region sequence*	*Disease*
Huntingtin	P42858	MATLEKLMKA FESLKSFQQQ QQQQQQQQQQ QQQQQQQQQQ PPPPPPPPPP PQLPQPPPQA	Huntington's Disease
Ataxin-1	P54253	LANMGSLSQT PGHKAEQQQQ QQQQQQQQHQ HQQQQQQQQQ QQQQQHLSRA PGLITPGSPP	SCA1
Ataxin-2	Q99700	ACEPVYGPLT MSLKPQQQQQ QQQQQQQQQQ QQQQQQQQPP PAAANVRKPG	SCA2
Ataxin-3	P54252	LRKRREAYFE KQQQKQQQQQ QQQQQQQQQQ QQQQQQQRDL SGQSSHPCER	SCA3
Ataxin-7	O15265	GEPRRAAAAA GGAAAAAARQ QQQQQQQQQP PPPQPQRQQH PPPPPRRTRP	SCA7
TATA-box-binding protein	P20226	ILEEQQRQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQAVAAA AVQQSTSQQA	SCA17
Voltage-dependent P/Q-type calcium channel subunit alpha-1A	O00555	SPVIRKAGGS GPPQQQQQQQ QQQQAVARPG RAATSGPRRY	SCA6
Androgen Receptor	P10275	GASLLLLQQQ QQQQQQQQQQ QQQQQQQQET SPRQQQQQQG	SBMA
Atrophin-1	P54259	TAHPPVSTHH HHHQQQQQQQ QQQQQQQHHG NSGPPPPGAF	DRPLA

1.5.9 Polyglutamine degradation

Two main methods exist for eukaryotic cells to remove misfolded and toxic proteins: the autophagy-lysosome pathways and the ubiquitin proteasome system[123]. In autophagy, cytoplasmic proteins are bound within autophagic vacuoles, autophagosomes, which fuse with lysosomes to form autophagolysosomes, within which lysosomal enzymes degrade the enclosed proteins[123]. This pathway may be activated in cells containing aggregated polyQ if the ubiquitin-proteasome system (UPS) is unable to degrade the misfolded protein[123], and evidence exists that its up-regulation may reduce clinical symptoms of polyQ diseases[124].

UPS degradation is carried out by the 26S proteasome which is an energy dependent multicatalytic protease which recognizes proteins tagged with ubiquitin as targets for degradation[125]. However, despite having 3 different types of catalytic sites (chymotrypsin-like, trypsin-like and caspase-like, which cleave peptides after hydrophobic, basic and acidic residues respectively) they are unable to hydrolyse polyQ tracts after the initial glutamine and must release the polyQ region as a polyQ peptide[126]. Not only is the proteasome unable to degrade polyQ effectively, it can also become irreversibly associated with it, leading to its sequestration in polyQ aggregates[121]. Several studies have found different results when examining impairment of the UPS by polyQ. There appears to be no global impairment of the UPS[127], but several different assays have detected some degree of impairment[123]. Depleted levels of free proteasome are not detected in cells containing aggregates[128], which may indicate that the proteasome is still present, but its function is impaired by the presence of polyQ. Recent research has shown that induction of mutant

huntingtin expression in a mouse model does transiently affect UPS function. When the mutant huntingtin is initially expressed, UPS function is reduced. On the formation of aggregated inclusion bodies, however, UPS function was restored[129]. This may indicate that monomeric mutant huntingtin or an oligomeric form free in the cytoplasm has the ability to block or otherwise disrupt the proteasome, but on aggregate formation the protein is no longer accessible to the UPS and so function is restored.

The mechanical stability of monomeric polyglutamine may be a factor in its inability to be properly degraded by the proteasome[107] and may explain some of the observed impairment before aggregation. In a recent study, it was found that the polyQ domains of an engineered protein were extremely resistant to mechanical force. Several lengths of polyQ from 15-75 were investigated by dynamic force spectroscopy using a force-clamp AFM and found to resist unfolding at an applied force of 180 pN. By comparison, the mechanically strong I27 domain of the muscle protein titin was unfolded at a force between 150-200 pN. Further experiments found a polyQ chain of 50 repeats was resistant to a mechanical force of up to 800 pN, demonstrating extreme mechanical stability; a surprisingly high figure, but not inconsistent with molecular dynamics simulations[32]. A particularly surprising result was that the mechanical resistance of the polyQ was length-independent, as both the physiological symptoms and aggregation propensity are known to increase with polyQ length. This may explain some of the impairment of the UPS observed, since the UPS uses mechanical force to unfold its substrate[125], if the UPS is unable to unfold the polyQ it may block the proteasome preventing its function.

1.5.10 Other repeat expansion disorders

Polyglutamine is not the only repeat expansion disorder, nine genes containing an alanine tract expansion are also known. Similar to polyglutamine diseases these can cause misfolding and aggregation, although the disease pathology appears more likely to be due to an accumulation of cytotoxic aggregates or a loss of protein function rather than the more complex and less well understood mechanism of polyQ diseases[130]. It is interesting that other repeat expansion disorders cause such symptoms as it is understood that polyQ's aggregation properties are related to the propensity of its side chains to hydrogen bond to each other rather than to the surrounding solvent, with polyalanine this is of course not possible.

At least another 11 disorders associated with repeat expansion are known, however these operate by conceptually simpler mechanisms or as yet unknown mechanisms. Expansions in non-coding regions can cause loss of protein function as in fragile X syndrome where a non-coding CGG codon is expanded causing transcriptional silencing and loss of the protein product, fragile-x mental retardation protein. This protein binds RNA and its silencing leads to severe mental retardation. Expansions can also occur in coding but unexpressed regions such as in dystpophia myotonica. Here the CAG repeat is part of the 3' untranslated region of the mRNA, so it is transcribed as RNA but the sequence comes after the stop codon. Consequently while it is not translated into a protein it can cause an alteration of RNA function. In this case RNA containing the expansion binds aberrantly to RNA-binding proteins that control splicing, this leads to abnormal splicing and dysregulation of protein

function. Others have an as yet unknown pathogenic mechanism though most are thought to disrupt RNA function[97,130].

1.6 Forces and polyglutamine

Given the putative involvement of an energy-dependent unfoldase in its cytotoxic mechanism, the idea of using force to study polyglutamine is not new. There are two previous studies in this area which are very relevant to my work. Dougan *et al*[32]. looked at the collapsed, mechanically rigid conformations adopted by homopolypeptide chains by force-clamp AFM. Hervás *et al*[33]. studied common features at the start of the neurodegeneration cascade by investigating the mechanical properties of a range of protein sequences involved in amyloid-like pathology. I will briefly review both here and then attempt to reconcile their differing conclusions.

1.6.1 Single homopolypeptide chains collapse into mechanically rigid conformations

This study used a single-molecule force-clamp technique where a chimeric polyprotein containing several I27 domains combined with one or more polyQ segments was stretched at a constant force of 180 pN in order to investigate the force response of the polyQ. They used the length fingerprint of the I27, which extends by 24 ±0.4nm on unfolding, to ensure that they were pulling on a single polyprotein. They looked at the initial length ($L_{Initial}$) of the polyprotein before any of the domains unfolded, this was in good agreement with an expected initial length calculated using a worm-like chain model (L_{WLC}) when the length of linkers (short amino acid sequences between and at the ends of the construct) and an expected size for a globular polyQ region were accounted for. In order to demonstrate that they were able to see a difference in $L_{initial}$ if the protein contained a domain which unfolded before the constant

force of 180 pN was reached, which would allow them to determine whether the polyglutamine region unfolded at this level of force, they used a protein containing the titin PEVK region which has a random-coil conformation but little mechanical strength. When the $L_{initial}$ of a protein construct containing two of these PEVK regions was compared to the same construct without these domains the expected increase in $L_{initial}$ was observed.

In the experiments with polyQ they saw the three unfolding events they would expect to observe from the unfolding of I27 but they did not see any polyQ unfolding events and the $L_{Initial}$ observed corresponded to the polyQ being in a globular conformation. They used polyQ regions of lengths between 15 and 75 glutamine and saw no difference in the $L_{Initial}$ or any unfolding events that indicated any unfolding of the polyQ. (Figure 1.21).

They also used a force-ramp technique where the force applied to the polyprotein was increased in a linear manner up to 800 pN. This also failed to detect any disruption of the polyQ region and both the initial and final lengths were in agreement with what was expected from the extension of the linkers and I27 domains. Interestingly similar results were also observed for polyalanine.

In order to explain these unexpected results they used molecular dynamics simulations to explore what might be happening in the polyQ. They took a polyQ region, unfolded it by heating to 800 K then allowed it to anneal at a temperature of 300 K before applying a mechanical unfolding force to it. They used this approach to probe 10,000 collapsed structures. These conformations were found to unfold *in silico* over a large range of mechanical forces but some were very stable and required up to 1,500 pN to unfold.

They theorised that this may be due to the extensive network of intra-chain hydrogen bonding found in these polyQ structures. It has been found in the unfolding of β-sheet rich proteins that those which require the rupture of a large number of hydrogen bonds tend to have more mechanical strength[40]. Dougan *et al*. thus hypothesised that the unfolding of polyQ might require the simultaneous rupture of a large number of hydrogen bonds and that this may explain the extreme mechanical strength seen in the polyQ. Another explanation could be that their polyglutamine containing peptides were not mono-disperse and some intermolecular interaction between the polyglutamine domains of different proteins was lending them an appearance of an increase in intrinsic mechanical strength. Such an observation could be seen if polyQ was in a fibre or pre-fibrillar aggregated state and I27 was in a soluble form decorating the fibre.

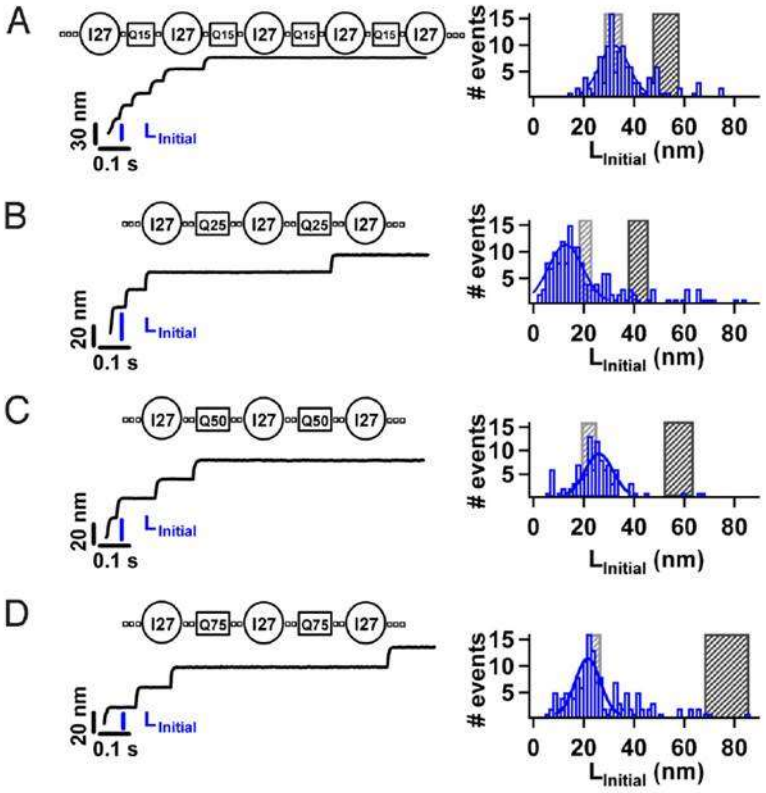


Figure 1.21: Probing the mechanical properties of homopolypeptide chains. Chimeric proteins were constructed comprising I27 and polyglutamine chains of different length, **A:** Q15, **B:** Q25, **C:** Q50, **D:** Q75. A constant force of 180 pN was applied, resulting in a series of step increases in length. Full mechanical extension of a complete construct was identified by the presence I27 unfolding steps (24 nm). Initial extension ($L_{Initial}$) were measured for each trajectory that satisfied this criterion. Histograms of $L_{Initial}$ are shown for each of the constructs. A Gaussian fit to the histograms (solid line) gave an average $L_{Initial}$ for each construct For all polyproteins, the measured $L_{Initial}$ was significantly shorter than that expected for full extension of the construct if the polyQ had unfolded (black shaded area). Instead $L_{Initial}$ was in close agreement with the expected length extension of only the folded I27 proteins and linkers (grey shaded area). From Dougan *et al*[32].

1.6.2 Common features at the start of the neurodegeneration cascade

This study used single molecule force spectroscopy by constant velocity AFM to investigate the mechanical resistance of the monomeric form of several amyloidogenic peptides. To maintain these aggregation-prone peptides in a soluble form they devised a novel method called a carrier-guest strategy (illustrated in Figure 1.22). This involves placing the peptide of interest (guest) within a loop of a ubiquitin domain (carrier). They claim that this allows a definitive identification of the force events corresponding to the unfolding of the peptide of interest as it is isolated from the application of force until the carrier ubiquitin is unfolded. They claim that this also eliminates any issues with non-specific tip or surface interaction by the guest as it unfolds far from the surface and guarantees that the analysed data comes from the stretching of a single guest protein.

They used this system to analyse the force response of a variety of polyglutamine lengths (Table 1.5). They saw only a few polyQ regions which were very mechanically stable, although the number seen did show a positive correlation with polyQ length. The reason for their results not agreeing with the results from Dougan *et al.* are unclear, they say that their results are not directly comparable with the results by Dougan *et al.* because they have a clear marker to indicate which are the unfolding events that should be analysed and have better controls for inter-molecular interactions.

It is also possible that the results are different because they used a different technique (constant velocity rather than constant force) with different loading rates and the isolation of the polyQ by the carrier means that when the carrier unfolds the polyQ experiences a sudden loading of force rather than a more gradual loading that would be experienced if the polyQ were a normal part of the chain.

Whatever the reason for these differences it is clear that polyQ reacts to the application of force in an interesting manner and that some mechanically resistant conformers are present, these could provide an explanation for polyQ appearing to block the proteasome. Even if the proteasome was only blocked by a relatively rare mechanically resistant polyQ (as suggested by Hervas *et al.* and the simulation of Dougan *et al.*), if it were to remain blocked it would prevent further degradation by that proteasome until the blockage was cleared. It is clear from both of these results that polyQ reacts to force in a remarkable manner which is worthy of further study. It will be particularly interesting to investigate how it reacts to force applied at only one end as illustrated in Figure 1.7B and at lower loading rates. For this reason it was chosen to investigate it by using the bacterial proteasome ClpXP.

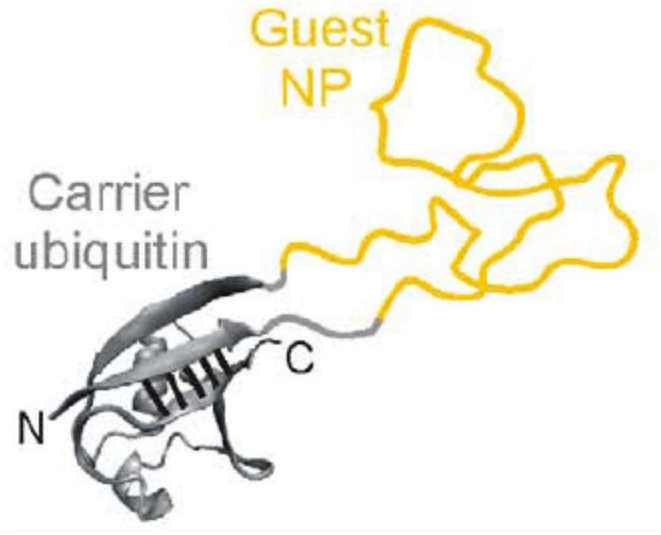


Figure 1.22: An illustration of the carrier-guest strategy. A guest neurotoxic protein (yellow) was placed in the loop of a ubiquitin carrier (grey) Adapted from Hervas *et al*[33].

Table 1.5: The results of SMFS experiments performed with guest proteins consisting of several different lengths of polyQ. n is the number of events analysed.

	Unfolding force of polyQ			
Q Length	n	≤20 pN (%)	20-400 pN (%)	≥400 pN (%)
19	111	100	0	0
35	100	95	5	1
62	107	92.5	7.5	2.8

1.7 Other proteins used in this work

While conducting this thesis research two other proteins were studied, which were used here as either a protein to calibrate an assay or as a substrate for ClpXP degradation.

1.7.1 Beta-2-microglobulin

Beta-2-microglobulin (β2m) is the non-covalently bound light chain of the major histocompatibility complex class 1 where it plays an important role in the assembly of the complex for antigen presentation. β2m contains 99 amino acids and adopts a classical β-sandwich fold with seven anti-parallel β-strands stabilised by a disulphide bridge. It has been well studied due to its aggregation being involved in dialysis related amyloidosis[131]. It was used in this study as an invariant standard to allow normalisation of densitometry data due to its availability in the laboratory.

1.7.2 Colicin E9

E. coli produce endonuclease colicins as a defence mechanism against competition by other bacteria. The endonuclease domain of the colicin E9 (herein called E9) is noted for its tight binding to its cognate immunity protein, Im9. The structure of the E9 endonuclease domain is shown in Figure 1.23. It was used in this study to provide a test-substrate for N-terminal degradation by ClpXP.

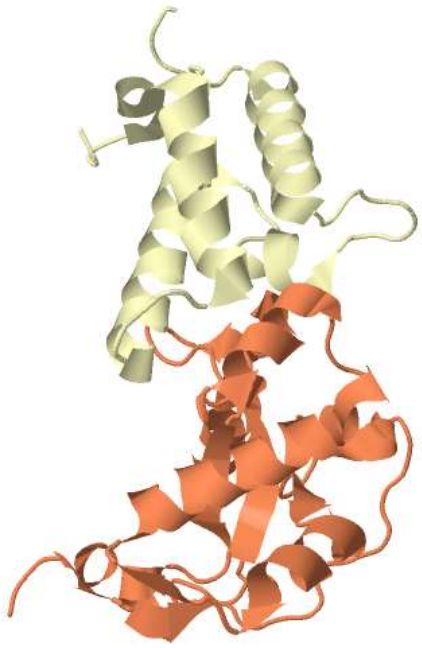


Figure 1.23: Cartoon rendering of colicin E9 endonuclease domain (orange) in complex with its cognate immunity protein Im9 (yellow). From PDB ID 1EMV[132]

1.7.3 Arc repressor

The *E. coli* arc repressor (Arc) was chosen as a substrate to test N-terminal degradation of a substrate by ClpXP because the Sauer group have used it successfully for this purpose previously[67]. It is dimeric in solution and its structure is shown in Figure 1.24.

Figure 1.24: Cartoon depiction of the solution structure of Arc from 1ARQ. Determined by NMR.[133]. Each monomer is shown in a different colour.

1.8 Aims and Objectives

1.8.1 Aims

The aims of this project are to develop ClpXP for use as a tool by which the mechanical properties of proteins can be investigated in a more biologically relevant manner than is available by the use of biophysical methods. This will then be used to investigate the mechanical properties of a polyglutamine repeat containing protein and to compare the results to those obtained by AFM.

1.8.2 Objectives

The initial objectives of this project are (i) to synthesise the necessary components of a ClpXP proteasome and find a suitable test protein; (ii) to develop an assay that can be used to investigate the mechanical properties of a substrate protein and (iii) to further develop this assay to allow investigation of different substrate proteins and conditions such as alternative unfolding/degradation directions and a non-reducing environments.

Once these have been achieved a polyglutamine-repeat-containing protein will be synthesised, its suitability assessed by circular dichroism, fluorescence emission spectroscopy and analytical ultracentrifugation. It will then be subjected to the assay developed and investigated by dynamic force spectroscopy using AFM and the results compared.

2 Materials and Methods

2.1 Reagents and Materials

2.1.1 General

Dithiothreitol (DTT) was purchased from Melford Laboratories, UK. Ethylenediaminetetraacetic acid disodium salt (EDTA) was purchased from Fisher Scientific, UK.

All other chemicals were purchased from Sigma-Aldrich, UK and were of analytical grade.

All solutions were made up with 18 MΩ.cm resistivity de-ionised water.

2.1.2 Molecular biology

Water used for molecular biology was sterilised by autoclaving for 20 minutes at 120 °C. Isopropyl-β-thiogalactosidase (IPTG), and carbenicillin disodium salt, were purchased from Melford Laboratories, UK. Ampicillin sodium salt was purchased from Formedium, UK. LB bouillon was purchased from Merck, Germany. Restriction enzymes, Quick ligase, T4 phosphonucleotide kinase, Vent polymerase and Antarctic phosphatase were purchased from New England Biolabs (UK) Ltd. pGEM-T Vector Systems, 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) and T4 DNA ligase were purchased from Promega Corporation, USA. QIAprep Spin Miniprep Kit, HiSpeed Plasmid Midi Kit and QIAquick Gel Extraction Kit were purchased from Qiagen, UK. Oligonucleotides were purchased and depending on their size purified by HPSF or HPLC+ from MWG biotech, UK. DNA sequencing was carried out in house, by DBS genomics, UK or by Beckman Coulter Genomics (UK).

2.1.3 Protein purification and analysis

Solutions used during protein purification were vacuum filtered using 0.45 µm Durapore® membrane filters purchased from Millipore, UK.

Ammonium persulphate and 30% *(w/v)* acrylamide:0.8% *(w/v)* bis-acrylamide were purchased from Severn Biotech Ltd, UK. 10× Tris-Tricine SDS electrophoresis buffer was purchased from Fermentas, UK. Mark 12 unstained standard molecular mass markers were purchased from Invitrogen, USA. Instant Blue was purchased from Expedion Protein Solutions, UK. Ni Sepharose high performance affinity chromatography resin, Source 15Q, Resource Q, Superdex 75 and Superdex 300 columns were purchased from GE Healthcare, Sweden. Spectra/Por (3,500 Da MWCO) dialysis tubing was purchased from Spectrum Medical Industries, USA.

2.1.4 Atomic force microscopy

Silicon nitride contact probes, model MLCT, were purchased from Veeco Instruments Ltd, UK. Phosphate buffered saline (Dulbecco A) (PBS) tablets were purchased from Oxoid Ltd., UK. Silicon wafers covered with a 1000 Å surface layer of evaporated gold were purchased from Platypus Technologies, USA. Microscope slides were purchased from Menzel GmbH, Germany. Epo-Tek 377 was purchased from Epoxy Technology, Inc. USA. Viva Spin 20 ultrafiltration spin columns (3,500 Da MCWO) were purchased from Sartorius Stedim Biotech GmbH, Germany.

2.2 Molecular biology

Escherichia coli strains used are described in Table 2.1.

2.2.1 Growth media

Bacterial strains were typically cultured in lysogeny broth (LB) prepared by dissolving 25 g per litre of LB bouillon in distilled water and sterilised by autoclaving for 20 minutes at 120 °C and 121 psi. Media were allowed to cool before the addition of the appropriate filter-sterilised antibiotic.

Agar plates were prepared from media as above with the addition of 15 g bacto-agar per litre of medium prior to autoclaving. After cooling, filter-sterilised antibiotics were added and the solution thoroughly mixed before being poured into 100 mm diameter petri dishes, 25 ml per dish. Plates were allowed to set before being stored at 4 °C.

Blue-white screening plates were prepared in the same way with the addition of 80 µg/ml X-gal and 0.5 mM IPTG prior to pouring.

Antibiotics were used at the following final concentrations: ampicillin 100 µg/ml, carbenicillin 100 µg/ml both dissolved in sterile water and chloramphenicol 25 µg/ml dissolved in 100% ethanol and filter-sterilised through a 0.22 µm filter.

Table 2.1: *E. coli* strains used; including their source, genotype and reason for their use in this project.

Strain	***Source***	***Genotype***	***Purpose***
SURE2	Stratagene	endA1 glnV44 thi-1 gyrA96 relA1 lac recB recJ sbcC umuC::Tn5 uvrC e14- Δ(mcrCB-hsdSMR-mrr)171 F'[$proAB^{+}$ $lacI^{q}$ lacZΔM1 5 Tn10 Amy Cm^{R}]	Cloning of DNA containing repeated sequences (I27 concatamers, ΔNClpX6)
BLR(DE3) pLysS	Novagen	F^{-} ompT gal dcm lon $hsdS_B(r_B^{-} m_B^{-})$ λ(DE3) pLysS(cm^R) $recA^{-}$	Expression of proteins containing repeated sequences (I27 concatamers, $^{\Delta N}ClpX_6$)
BL21(DE3)	Laboratory Stocks	F^{-} ompT gal dcm lon $hsdS_B(r_B^{-} m_B^{-})$ λ(DE3)	ClpX expression
BL21(DE3) pLysS	Laboratory Stocks	F^{-} ompT gal dcm lon $hsdS_B(r_B^{-} m_B^{-})$ λ(DE3) pLysS(cm^R)	ClpP expression

2.2.2 Agarose gel electrophoresis

1.5-3 g agarose was added to 100 ml of TAE buffer (200 mM Tris, 0.114% *(v/v)* glacial acetic and 1 mM EDTA) in a conical flask and the solution heated until the agarose was dissolved. The solution was allowed to cool until it could be handled comfortably before addition of 1 mg/ml ethidium bromide to a final concentration of 1 µg/ml. The gel was then poured into a mould containing a comb with the appropriate number of lanes. For most purposes 1.5% *(w/v)* agarose gels were used apart from separation of ligated polyQ fragments, where 3% *(w/v)* agarose gels were used.

Gels were then completely covered with TAE buffer in a gel tank making sure no air was trapped between the tank and the casting unit. DNA samples (containing 100 ng-1 µg DNA) and markers were added to 6×gel loading buffer (0.25% (*w/v*) bromophenol blue, 0.25% (*w/v*) xylene cyanol, 40% (*w/v*) sucrose).

Gels were electrophoresed at 10 volts per centimetre for around 1 h until the bottom marker reached no further than half-way through the gel. The DNA was then visualised by UV trans-illumination (Syngene, GVM20). If the ethidium bromide counter-migrated further than the lowest expected DNA band then the gel was re-stained for one hour in TAE buffer containing 1 µg/ml ethidium bromide.

2.2.3 PCR mutagenesis

All PCR mutagenesis of I27 was performed on single I27 cassettes in pGEM derived plasmids before ligation into a concatamer. Mutagenic primers were designed with the reverse complement of the strand to which they were to be

annealed with a length sufficient to provide an annealing temperature between 55-65 °C (using the formula of $T_m = 4(C+G) + 2(A+T)$ °C) terminating in a 3' C or G. 5' to this they contained the required mutation followed by 9 random bases 5' to the annealing sequence to facilitate cutting by restriction enzymes.

PCR mixture consisted of:

- 50-100 ng dsDNA template
- 50 pmol of each primer
- dNTPs at a final concentration 750 µM dNTP
- Thermo pol buffer at 1×final concentration
- $MgSO_4$ to a final concentration of 2, 4 or 6 mM
- 0.5 µl (1 unit) Vent DNA polymerase
- 1 µl dimethyl sulfoxide

The solution was made up to 100 µl final volume with chilled (on ice) distilled water.

Temperature cycling was performed in a PTC-100 programmable thermal controller (MJ Research Inc.). The thermal cycle was:

1. 95 °C 30 seconds
2. 95 °C 30 seconds
3. *n* °C 60 seconds
4. 72 °C 60 seconds per kb of expected product
5. Repeat steps 2-4 24 times
6. 72 °C 5 minutes

n = 5 °C below primer T_m.

The reaction mixture was stored until use at 4 °C or frozen at -20 °C until required.

See Appendix – Section 7.2 for a full list of primers used.

2.2.4 Site directed mutagenesis

Small additions, deletions and mutations were carried out using a QuikChange Site-Directed Mutagenesis Kit (Agilent Technology, USA) according to the manufacturer's instructions with the exception that DpnI digests were carried out overnight rather than for 1 hour as instructed. Primers were designed using QuikChange primer design software available at: https://www.genomics.agilent.com/primerDesignProgram.jsp.

2.2.5 Plasmid DNA preparation

Small scale plasmid preparation (to obtain approximately 3 µg in 28 µl buffer) was carried out using a QIAprep Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions with the following modification. DNA was eluted with 30 µl elution buffer; the eluate was then added back into the column and eluted again.

Large scale plasmid preparation (to obtain approximately 200 µg in 1 ml buffer) was carried out by use of a HiSpeed Plasmid Midi Kit (Qiagen) according to manufacturer's instructions. Eluted DNA was then concentrated by use of a QIAquick Gel Extraction Kit (Qiagen) with the omission of the steps involved in dissolving the agarose gel.

2.2.6 Preparation of competent cells

A 10 ml overnight culture of desired strain was grown in LB medium from a single colony with the appropriate antibiotics. This was inoculated (1:20) into 100 mL LB and grown at 37 °C until the OD_{600} was 0.40–0.45.

Cells were harvested in sterile 50 ml Falcon™ tubes (4000 rpm, 4 °C for 10min, taking care that the Beckman Coulter JS 5.3 rotor was pre-chilled) then the pellet gently resuspended in 10mL of sterile, pre-chilled 100 mM $CaCl_2$. They were then incubated on ice for 10 min and centrifuged as before. The pellet was resuspended in 2 mL of 100 mM $CaCl_2$, 30% (*v/v*) glycerol then divided into 100 μL aliquots in micro-centrifuge tubes resting on dry ice and allowed to freeze. The aliquots were then stored at -80 °C until use.

2.2.7 Transformation

Transformations of commercial competent cells with plasmid DNA were carried out according to the manufacturer's instructions.

Transformation of laboratory stocks of competent cells was by addition of 2 µl DNA at a concentration of 1ng/µl to 50 µl competent cells freshly thawed on ice, in a sterile 14 ml polypopyline transformation tube (Starlab, UK). The cells were then incubated on ice for 30 minutes with occasional gentle mixing. The cells were heat shocked in a water bath at 42 °C for 30 seconds before being incubated on ice for a further 5 minutes. 500 µl sterile LB was then added to the transformation tube and incubated at 37 °C in an orbital incubator at 200 rpm. Cells were then spread onto two selective agar plates, (100 µl on one plate and 450 µl on a second plate) and the agar was allowed to dry before inversion and incubation overnight at 37 °C.

2.2.8 Restriction digests

Restriction digests were performed using NEB restriction enzymes according to the manufacturer's instructions. The optimal buffer composition for double digests were calculated according to the NEB double digest finder.[134] Typical digests contained 1µg of DNA.

2.2.9 Ligation of cassettes into vector

Vector plasmids were digested with the appropriate restriction enzymes. Phosphate groups were removed from 5' overhangs by use of Antarctic Phosphatase according to the manufacturer's instructions. After inactivation of the phosphatase (5 minutes at 70 °C) the DNA was then separated by electrophoresis on a 1.5% *(w/v)* agarose gel and the correct band was excised with a scalpel and the DNA purified using a QIAquick Gel Extraction Kit according to the manufacturer's instructions.

Inserts were prepared in the same way as the vector but the dephosphorylation step was omitted.

Ligation was carried out by using Quick Ligase according to the manufacturer's instructions (50 ng template DNA with a 3 fold molar excess of insert) and the ligation mixture (5 ng in a 2 µl volume) transformed into the appropriate competent cells.

2.2.10 Polyglutamine preparation

PolyQ tracts were prepared by the use of terminator and elongator DNA fragments as described in section 4.3.1. Each fragment was made from two complimentary oligomers with 5' and 3' overhangs that were complementary to upstream or downstream cassettes (Appendix – Section 7.2.1).

The oligomers were phosphorylated by use of T4 polynucleotide kinase according to the manufacturer's instructions. Each pair of oligomers required to make the duplex terminators and elongators were mixed together in a 1:1 ratio and 3 M NaCl was added to a final concentration of 100 mM. The terminators and elongators were then annealed by heating to 95 °C then cooling to 25 °C at 1 °C per minute.

Ligation was performed by mixing terminators and elongators in a 1:10 ratio to a total mass of 2 µg DNA. 10 µl of 10× T4 ligase buffer and 5 µl T4 DNA ligase was added and the solution made up to 100 µl with distilled water. The mixture was incubated at 15 °C for 3 hours then 12 °C for 15 hours. A control was prepared in the same way but the T4 DNA ligase was replaced with distilled water.

Small unligated DNA fragments and enzymes were removed from the mixture by use of a QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's instructions.

2.2.11 Blunt-ended ligation

The purified polyQ ligation mixture was blunt-end ligated into a pGEM-T plasmid by use of a pGEM-T Vector system. Briefly, the mixture was "A-tailed" and ligated into the plasmid as per the manufacturer's instructions then transformed into *E. coli* SURE 2 super-competent cells. The transformed cells were plated onto blue-white colour screening plates (Section 2.2.1) and grown overnight.

White colonies containing the insert were selected and inoculated into 5 ml LB with 100 µg/ml ampicillin using a sterile wire loop and grown overnight at 37 °C

with shaking at 200 rpm. The presence of an insert was confirmed by a restriction digest and the DNA was sequenced (Sections 2.2.5 and 2.2.8).

2.2.12 λO-Arc

DNA encoding λO-Arc was created synthetically using a commercial supplier (MWG Eurofins). To design the sequence to be ordered the *E. coli* arc repressor protein sequence [*Escherichia coli* P0299483.3] was obtained from the NCBI protein database adding the protein sequence for the λO degradation tag to the N-terminus and a hexahistidine tag to the C-terminus. The sequence was then reverse translated to optimise *E. coli* codon usage using the reverse translation tool at http://www.bioinformatics.org/sms2/rev_trans.html adding restriction sites to the obtained DNA sequence 5' XhoI and 3' NdeI. The sequence was then ordered form MWG Eurofins and a restriction digest and ligation used (Sections 2.2.8 and 2.2.9) to transfer the gene into a pET23a plasmid for expression.

2.2.13 λO-E9

A plasmid containing a variant of the colicin domain of E9 (TEV-(T1-16)-E9) was obtained (a gift from C. Kleanthous, University of York). PCR (Section 2.2.3) was used to add the λO sequence (MTNTAKILNFGR) to the 5' end to produce λO-(T1-16)-TEV-E9. this is referred to as pET23λO-(T1-16)-TEV-E9. This was then ligated into a pET-derived plasmid for transformation and expression (Section 2.2.9), this is referred to as pET23λO-(T1-16)-TEV-E9..

2.3 Plasmid source summary

Table 2.2: Plasmids used in this thesis and their sources

Plasmid	Protein	Source	Notes
pET23His$_6$-ClpX	ClpX with N terminal His tag	Gift of J.Pullen (Brockwell Group)	
pET23ClpXΔRKH	ClpX with N terminal His tag and R228A, K229A, and H230A mutations	New	See Appendix: Figure 7.1 for plasmid map
pET23ClpXΔRKH-TC	ClpX with a thrombin cleavable N terminal His tag and R228A, K229A, and H230A mutations	New	See Appendix: Figure 7.2 for plasmid map
pAYC$^{\Delta N}$ClpX$_6$-His$_6$	ClpX psuedohexamer with deleted N-terminal domain	Gift of R.T.Sauer group[135]	
pET23ClpP	ClpP	Gift of J. Pullen (Brockwell Group)	
pET23-His$_6$ClpP	ClpP with N-terminal His tag	Gift of J. Pullen (Brockwell Group)	
pET23λOARC-His6	Arc repressor with N terminal λO sequence and C-terminal His Tag	New	See Appendix: Figure 7.3 for plasmid map
pET23λO-(T1-16)-E9	λO-(T1-16)-E9 colicin domain	New	See Appendix: Figure 7.4 for plasmid map
pET23His6-I27-ssrA	I27$_3$-ssrA concatamer with N-terminal His tag	New	See Appendix: Figure 7.5 for plasmid map
pET23His6-Q30-ssrA	I27-Q30-(I27)$_2$-ssrA concatamer with N-terminal His tag	New	See Appendix: Figure 7.6 for plasmid map

2.4 Protein preparation

2.4.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Gel electrophoresis was routinely used to monitor over-expression and purification of proteins as well as protein quantification in degradation assays. Tris-tricine buffered SDS-PAGE gels were used for this purpose, using a two-layered system consisting of a stacking gel and a resolving gel.

Two glass plates (Atto mini-gel kit) separated by a 1.5 mm spacer were assembled, according to the manufacturer's recommendations.

The resolving gel mixture (Table 2.3) was made and rapidly poured to within 2 cm of the top of the glass plates. The stacking gel mixture was then mixed and poured on top of the resolving gel and a comb inserted to create sample wells. Anode buffer (200 mM Tris-HCl pH 8.9) and cathode buffer (1×Tris-tricine SDS electrophoresis buffer) were added to the gel tank. The samples were diluted two-fold in 2× loading buffer (50 mM Tris-HCl pH 6.8, 100 mM DTT, 2% (*w/v*) SDS, 0.1% (*w/v*) bromophenol blue, 10% (*v/v*) glycerol) and boiled for a minimum of 2 minutes before loading onto the gel. 15 µl Invitrogen Mark 12 unstained standard molecular mass markers were loaded into one lane to aid identification of protein bands. The gels were electrophoresed at a constant current of ~30 mA until the sample entered the resolving gel when the current was then adjusted to ~60 mA. Gels were rinsed with distilled water then stained in Instant Blue (Expedeon)for one hour before being rinsed again. The gel was visualised using a light box and an image of the gel acquired using Gene Snap (Synoptics Ltd).

Table 2.3: Components of Tris-tricine buffered SDS-PAGE gels. The volumes indicated are sufficient for casting two 8 cm × 10 cm mini gels using a 1.5 mm spacer.

Solution component	**Resolving gel** Volume to add (ml)	**Stacking gel** Volume to add (ml)
30% (*w/v*) Acrylamide:0.8% (*w/v*) *bis*-acrylamide	7.5[1]	0.83
H_2O	0.44	3.72
3 M Tris, 0.3% *(w/v)* SDS pH 8.45	5	1.55
Glycerol	2	-
10% (*w/v*) ammonium persulphate	0.1	0.2
N,N,N',N'-tetramethyl-ethylenediamine (TEMED)	0.01	0.01

[1] For $^{\Delta N}ClpX_6$ 3.75 ml acrylamide and 4.19 ml H_2O were used.

2.5 Over-expression

2.5.1 Expression trial

After transformation *E. coli* with of a plasmid harbouring the appropriate gene each of four colonies were used to inoculate 5 ml of LB growth medium containing appropriate selection antibiotics. The culture was shaken overnight at 200 rpm and 37 °C. 500 µl of these cultures were mixed with 500 µl sterile 30% glycerol *(v/v)* and frozen at -20 °C. 1 ml of each overnight culture was used to inoculate 50 ml of LB medium under the same antibiotic selection. These flasks were shaken at 200 rpm and 37 °C until an OD_{600} of ~0.6 was attained. Protein expression was induced for three of the four cultures with filter sterilised IPTG at a final concentration of 1 mM for 4 h. 1 ml of culture was then removed and centrifuged for 1 min at 16,300 RCF using a bench-top centrifuge (Spectrafuge 24D Labnet international Inc.). The supernatant was discarded and the cell pellets were stored at -20 °C until required. For $^{\Delta N}ClpX_6$ the same procedure was followed except that incubation was at 18 °C.

2.5.2 Sample analysis by SDS-PAGE

E. coli cell pellets were resuspended in 500 µl of 15 mM Tris-HCl pH 7.0 buffer containing lysozyme (1 mg/ml), RNAse (0.1 mg/ml) and DNAse (0.1 mg/ml and left at room temperature for 20 minutes. A 50 µl sample was removed and added to 50 µl of 2× loading buffer (50 mM sodium phosphate pH 6.8, 100 mM DTT, 2% *(w/v)* SDS, 0.1% *(w/v)* bromophenol blue, 10 % *(v/v)* glycerol). The remaining sample was then centrifuged for 5 minutes, 16,300 RCF in a bench-top centrifuge. The supernatant was then mixed with an equal volume of 2× loading buffer and the pellet was re-suspended in 450 µl of 15mM Tris-HCl pH

7.0 and mixed with an equal volume of 2× loading buffer. The samples were boiled for 5 minutes. 15µl aliquots were loaded on to an SDS-PAGE gel.

2.5.3 Large scale over-expression

A starter culture of 200 ml LB with appropriate selection antibiotics was inoculated by a sterile wire loop from the glycerol stock of the best expressing transformant taken during the expression trial. The culture was incubated overnight at 37 °C shaking at 200 rpm. 10 × 1 l LB with appropriate selection antibiotics was then was inoculated with 20 ml of the starter. This was incubated (37 °C, 200 rpm) until an $OD_{600} \approx 0.6$ was reached, when expression was induced by addition of IPTG to a final concentration of 1 mm. After 4 hours, cells were harvested in a Heraeus Contrifuge at 15,000 rpm (Rotor 8575) at 4 °C. Cell pellets were frozen at – 20 °C until required. The procedure for $^{\Delta N}ClpX_6$ was the same except that cells were incubated at 18 °C.

2.5.4 Auto-induction

Stock solutions (Tables 2.4 - 2.7) were autoclaved separately at 120 °C and 121 psi for 20 min. Trace elements (Table 2.7) were filter-sterilised through a 0.22 µm filter. 2ZYM – 1×LAC expression medium was prepared by addition of 50× LAC, 20× NPSC and trace elements to a 2.5 l sterile baffled conical flask containing 2 ZY as described in Table 2.8.

Cultures were grown in for 24 hours after inoculation with 10 ml from an overnight starter culture grown in LB medium. Cells were then harvested as described in Section 2.5.3.

Table 2.4: Components of 2 ZY solution (per 465 ml)

Yeast extract	5 g
Bactotryptone	10 g

Table 2.5: Components of 50× LAC solution (per 50 ml)

Glycerol	12.5 g
Glucose	1.25 g
α-Lactose	5.00 g

Table 2.6: Components of 20 × NPSC solution (per 50 ml)

NH_4Cl 2.675 g	1 M
Na_2SO_4 1.61 g	0.1 M
KH_2PO_4 3.4 g	0.5 M
Na_2HPO_4 3.55 g	0.5 M

Table 2.7: Components of Trace elements 1000× (per 1 l)

$FeSO_4.7H_2O$	1.0 g
$ZnSO_4.7H_2O$	8.8 g
$CuSO_4.7H_2O$	0.4 g
$MnSO_4.4H_2O$	0.15 g
$Na_2B_4O_7.10H_2O$	0.1 g
$(NH_4)_6Mo_7O_{24}.4H_2O$	0.05 g

Table 2.8: Components of 2ZYM – 1× LAC media.

2ZY solution	465 ml
1M $MgSO_4$	1 ml
50 x LAC	10 ml
20 x NPSC	25 ml
Trace Elements	500 µl

2.6 Protein extraction and purification

2.6.1 General procedures

Before purification, cell pellets were defrosted and then re-suspended in 10 ml per gram of cell pellet of the buffer to be used in the first stage of purification. Cells were fragmented using a cell disrupter (Constant Cell Disruption Systems) and the debris removed by centrifugation at 12,000 rpm, 4 °C, 30 minutes (rotor JLA 16.250, Beckman Coulter). The cleared lysate was subsequently decanted and filtered through a 0.2 µm syringe filter.

Estimated extinction coefficients for all proteins were determined using the online protein parameter calculator at http://web.expasy.org/protparam/.

2.6.2 Lyophilisation

Before lyophilisation the A_{280} of the protein solution was measured using a spectrophotometer (Ultraspec 2100 pro, Amersham Biosciences). Protein concentration was calculated using Equation 2.1:

$$concentration \quad \frac{absorbance}{extinction\ coefficient \times path\ length}$$

Equation 2.1: The Beer-Lambert equation

The volume of protein required in each aliquot was determined by using Equation 2.2.

$$volume = \frac{amount\ of\ protein\ required}{concentration}$$

Equation 2.2: Used to determine the volume of liquid required for a given amount of protein when the concentration is known.

The protein was divided into aliquots (typically 10 × 0.5 or 1 mg with the remainder of 5 or 10 mg) and placed in either a 1.5 ml micro-centrifuge tube or a 50 ml Falcon™ tube. The protein solution was then frozen by immersing the sealed container in a bath of ethanol containing dry ice. Once the solution was frozen the lid was replaced with one which had been pierced to allow evaporation and the protein placed in a lyophiliser (PowderDry PL3000, Heto). Once all the liquid was removed the pierced lid was replaced with the original lid and the protein stored at -20 °C until required.

2.6.3 ClpP purification

ClpP over-expression was achieved from a pET23 derived plasmid conferring ampicillin resistance transformed into *E. coli* BL21 [DE3].

All steps were carried out on ice using chilled buffers. The cell pellet of a large scale culture of *E. coli* BL21 [DE3] pET23aClpP (Section 2.5.3) was lysed in Ni Sepharose wash buffer (50 mM Tris-HCl pH 7.6. 500 mM NaCl, 10% Glycerol *(v/v)*, 20 mM imidazole) and the lysate clarified and filtered as described in Section 2.6.1. The filtered lysate was made up to 100 ml with Ni Sepharose wash buffer.

Beads of Ni Sepharose High Performance (GE Healthcare, Sweden) were re-suspended by gentle shaking and 0.5 ml was added to each of 4 50 ml Falcon™ tubes. The beads were equilibrated by adding 25 ml Ni Sepharose wash buffer, mixing thoroughly (roller mixer SRT6, Stuart) and centrifuged for 5 minutes, 4,000 rpm, 4 °C (rotor Beckman Coulter JS 5.3). The supernatant was discarded and the procedure repeated. The filtered lysate was then added

to the beads and incubated for 20 minutes. 4 wash steps were carried out in Ni Sepharose wash buffer by sequential re-suspension and centrifugation steps as described above. Three further washes were carried out at increasing imidazole concentrations (100, 150 and 200 mM). Two elutions were carried out by resuspending the beads in 50 ml elution buffer (as wash buffer but with 500 mM imidazole) and centrifuging as before. The supernatant was retained and the two elutions were pooled. Samples were retained from each step to determine purity and the eluted sample was dialysed (3,500 MWCO tubing) into ClpXP storage buffer (25 mM HEPES pH 7.6, 15% *(v/v)* glycerol, 200 mM KCl and 2 mM EDTA) with stirring and two buffer changes allowing a minimum of three hours between buffer changes. The dialysed sample was then concentrated using Vivaspin 20 ultrafiltration spin columns to a concentration between 5-20 μM of $ClpP_{14}$ (tetradecameric ClpP has an estimated ε_{280} = 109,200 M^{-1} cm^{-1}). ClpP was aliquoted into 50 μl samples, snap frozen in liquid nitrogen and stored at -80 °C until required.

2.6.4 ClpX purification

ClpX over-expression was achieved from a pET23 derived plasmid conferring ampicillin resistance transformed into *E. coli* BL21 [DE3] pLysS, except $^{\Delta N}ClpX_6$, which is in a pACYC derivative plasmid which confers chloramphenicol resistance (J. Davis, personal communication) and is transformed into *E. coli* BLR [DE3].

Purification of all ClpX variants was the same as for hexahistidine tagged ClpP (Section 2.6.3) except that the intermediate wash steps at 100, 150 and 200 mM imidazole were omitted. Additionally, before concentration the sample

was subjected to size exclusion chromatography using a Sephacryl S300 HR 16/60 column (G.E. Healthcare, Sweden,120 ml bed volume) attached to an ÄKTA prime chromatography system. After equilibration of the column with 2 column volumes of ClpXP storage buffer, 5 ml volumes of the sample were injected into the loading loop using a syringe and after injection onto the column the sample was eluted with 350 ml ClpXP storage buffer at a flow rate of 0.5 ml/minute. Fractions showing increased A_{280} values were collected and analysed by SDS-PAGE and fractions containing ClpX were pooled and concentrated by Vivaspin 20 ultrafiltration spin columns to a concentration between 5-20 µM ($ClpX_6$). $ClpX_6$ variants and $^{\Delta N}ClpX_6$ have an estimated ε_{280} = 82,700 M^{-1} cm^{-1}). ClpX was aliquoted into 50 µl volumes, snap frozen in liquid nitrogen and stored at -80 °C until required.

2.6.5 Thrombin cleavage

Thrombin cleavage was carried out using a Thrombin Clean-Cleave Kit (Sigma). 100 µl thrombin resin was equilibrated with ClpXP storage buffer by first pelleting the thrombin beads by centrifugation at 4,000 rpm for 5 minutes at 4 °C (rotor Beckman Coulter JS 5.3). After removing the supernatant the resin was re-suspended by adding 10× its volume of ClpXP storage buffer. This procedure was then repeated. The beads were then re-suspended with an equal volume of ClpXP storage buffer and ClpX with a thrombin-cleavable His tag was then added and incubated for 4 hours at room temperature with shaking. A 15 µl aliquot was taken every hour for analysis by SDS-PAGE. The beads were chilled to 4 °C and incubated for a further 18 hours. Once the time taken for complete cleavage of the thrombin tag was optimised, the procedure was scaled up to use 1ml of thrombin beads with a larger amount of ClpX (1 ml

of ~10μM $ClpX_6$). The cleaved ClpX was recovered by washing the beads twice with ClpXP storage buffer. This solution then had imidazole added to a concentration of 20 mM and was added to Ni Sepharose beads equilibrated with the same solution. The Ni Sepharose beads were then centrifuged as above and the supernatant containing the ClpX was recovered and stored as described above.

Both the Thrombin and Ni Sepharose beads were washed and regenerated as per the manufacturer's instructions for re-use.

2.6.6 λO-Arc purification

λO-Arc over-expression is achieved from a pET23 derived plasmid conferring ampicillin resistance transformed into *E. coli* BL21 [DE3] pLysS.

λO-Arc was expressed and purified in the same way as ClpX (Sections 2.5.3 and 2.6.4, respectively) except that the buffers used for Ni Sepharose purification were 25 mM Tris, 500 mM NaCl, 20 mM imidazole at pH 7.6 for binding and the elution buffer contained 500 mM imidazole. Gel filtration was carried out using 25 mM Tris, 500 mM NaCl, at pH 7.6 using a HiLoad 26/60 Superdex 75 prep grade column (120 ml bed volume).

2.6.7 λO-(T1-16)-TEV-E9 purification

λO-(T1-16)-TEV-E9 over-expression is achieved from a pET23 derived plasmid conferring ampicillin resistance transformed into *E. coli* BL21 [DE3] pLysS.

λO-(T1-16)-TEV-E9 is co-expressed with a hexa-histidine tagged variant of E9's cognate immunity protein (Im9) to which it binds very tightly[19]

(K_d = 10^{-16} M). Both proteins are encoded on the same pET derived plasmid, conferring ampicillin resistance, from the same T7 promoter.

As the E9:Im9 complex is extremely avid, when Im9 is purified by affinity chromatography, λO-(T1-16)-TEV-E9 will be co-purified with it. The complex is left bound to the Ni Sepharose and the proteins can then be separated by denaturing them with guanidine hydrochloride. This denatures and dissociates the λO-(T1-16)-TEV-E9:Im9 complex, eluting the λO-(T1-16)-TEV-E9 while leaving the Im9 bound to the resin. λO-(T1-16)-TEV-E9 was purified by nickel affinity using Ni Sepharose beads in an XK26 column (with a bed volume of 35 ml) on an ÄKTA prime purification system.

Three buffers were used for the affinity chromatography:

Binding buffer: 25 mM Tris, 200 mM NaCl, 20 mM imidazole pH 8.0. Elution buffer: binding buffer with the addition of 6 M guanidine hydrochloride. Wash buffer: binding buffer with the addition of 500 mM imidazole. The cells were lysed after being re-suspended in binding buffer (Section 2.6.3) and the cleared lysate was loaded on to the Ni Sepharose column pre-equilibrated with binding buffer. The column was then washed with binding buffer until no changes in the A_{280} of the eluant were detected. The λO-(T1-16)-TEV-E9 was dissociated from the His-tagged Im9 by addition of denaturing elution buffer and eluted; collecting fractions that contained λO-(T1-16)-TEV-E9. The Im9 was eluted from the column with wash buffer and discarded.

The fractions containing λO-(T1-16)-TEV-E9 were dialysed overnight into gel filtration buffer (25 mM Tris, 100 mM NaCl, pH 8.0) before further purification

using a Superdex S75 column (G.E. Healthcare, Sweden, 120 ml bed volume). Fractions containing λO-(T1-16)-TEV-E9 were then dialysed (3,500 MWCO dialysis tubing) into distilled water with stirring and two buffer changes, allowing a minimum of three hours between buffer changes. The protein was then lyophilised as described in Section 2.6.2.

2.6.8 $I27_3$-ssrA and I27-Q30-$(I27)_2$-ssrA purification

I27 concatamer over-expression is achieved from a pET23 derived plasmid conferring ampicillin resistance transformed into *E. coli* BL21 [DE3] pLysS.

Cells were re-suspended in lysis buffer (25 mM Tris-HCl pH8.0, 2 mM DTT), and lysed as described in Section 2.6.1. The cleared lysate was made up to 100 ml with lysis buffer and an ammonium sulphate precipitation procedure was performed. Briefly, the lysate was placed in a beaker on ice and powdered ammonium sulphate was added stepwise with constant stirring. The mass of ammonium sulphate to be added at each step was determined from Table 20.1 of Burgess, 2009[136].

Ammonium sulphate was added to 20% saturation (106 g/l), then insoluble material removed by centrifugation at 15,000 rpm, 0 °C for 10 minutes (rotor JLA 25.50 Beckman Coulter). The insoluble pellet was re-dissolved in lysis buffer and a sample taken for analysis by SDS-PAGE, the volume of the supernatant was measured and further ammonium sulphate added in intervals of 10% saturation until 80% saturation was achieved, with centrifugation at each stage as described. After analysis by SDS-PAGE, samples containing significant amounts of the concatamer were pooled and dialysed against distilled water before lyophilisation (Sections 2.6.6 and 2.6.2 respectively).

Lyophilised protein was then re-dissolved in buffer A (25 mM Tris-HCl pH 8.0, 2 mM DTT, 1 mM EDTA). A 5ml Q anion exchange column on an ÄKTA prime purification system was equilibrated with two column volumes of buffer A before addition of 2 ml protein, *via* the injection loop and washing with a further two column volumes of buffer A. The concentration of buffer B (as buffer A with the addition of 1 M NaCl) was adjusted to 30% and the column washed with a further two column volumes. A gradient of 30-70% buffer B was then applied over ten column volumes before 100% buffer B was used for a further two column volumes. The fractions showing an increase in A_{280} were analysed by SDS-PAGE and fractions containing the protein were pooled, dialysed against distilled water and lyophilised as described above.

Lyophilised protein was re-dissolved in gel filtration buffer (25 mM Tris-HCl pH 8.0, 2 mM DTT, 1 mM EDTA and 200 mM NaCl). A Superdex 75 prep grade HiLoad 26/60 gel filtration column, 318 ml bed volume (G.E. Healthcare, Sweden) attached to an ÄKTA prime purification system was equilibrated with 1.5 column volumes of gel filtration buffer before addition of 5 ml protein via the injection loop and elution with gel filtration buffer. The fractions showing an increase in A_{280} were analysed by SDS-PAGE and those containing the pure concatamer were dialysed against distilled water and lyophilised as described above.

I27 constructs with a hexahistidine affinity tag were purified in the same way except that the ammonium sulphate precipitation was replaced with nickel affinity purification as described for ClpX (section 2.6.4).

2.7 **Protein source summary**

Table 2.9: The proteins used in this thesis, their source and abbreviation used.

Protein	Abbreviation	Source
ClpX with N terminal His tag	ClpX	New protocol
ClpX with N terminal His tag and R228A, K229A, and H230A mutations	$ClpX^{\Delta RKH}$	New Protocol
ClpX with a thrombin cleavable N terminal His tag and R228A, K229A, and H230A mutations	Thrombin cleaved $ClpX^{\Delta RKH}$	New Protocol
ClpX psuedohexamer with deleted N-terminal domain	$^{\Delta N}ClpX_6$	New Protocol
ClpP	ClpP	New Protocol
ClpP with N-terminal His tag	His_6-ClpP	New Protocol
Im9 with N-terminal His tag and C-terminal ssrA tag	Im9-ssrA	Gift (J.Pullen, Brockwell group, Leeds)
Creatine phosphokinase		Purchased (Sigma)
Beta-2-microglobulin	β2M	Gift (G.N.Khan, Radford group, Leeds)
Myoglobin		Purchased (Sigma)
Arc repressor with N terminal λO sequence and C-terminal His tag	λO-Arc	New Protocol
λO-(T1-16)-E9 colicin domain-His_6	E9	Modified existing Protocol provided by C. Kleanthus Group
His_6-$I27_3$-ssrA	$(I27)_3$	New Protocol
His_6-I27-Q30-I27-ssrA	Q30	New Protocol
His_6-$(I27)_5$	$(I27)_5$	Gift (O.Farrance, Brockwell group, Leeds)

2.8 **Degradation assay**

The degradation assay was performed at 30 °C in a reaction mixture containing 200 nM $ClpX_6$, 600 nM $ClpP_{14}$, 5 mM ATP, 16 mM creatine phosphate, 0.032 mg/ml creatine phosphokinase, 25 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 200 mM KCl, 10% glycerol *(v/v)* and 1 mM DTT. A typical reaction volume was 200 µl.

Immediately before the degradation assay, aliquots of ClpX, ClpP, ATP and 10× reaction buffer (containing creatine phosphate, creatine phosphokinase, Tris-HCl pH 7.6, $MgCl_2$, KCl, and DTT) were removed from the freezer, thawed on ice, added to glycerol and mixed. A control was carried out in which the ATP was replaced with distilled water.

Lyophilised substrate protein was dissolved in distilled water and centrifuged in a bench-top centrifuge for 5 minutes at 16,300 RCF to remove insoluble aggregates. The reaction mixture and substrate were then pre-incubated separately at 30 °C for 5 minutes. The substrate protein was added (concentrations varied but were typically 10 mM) and a sample taken immediately (time = 0). ClpXP-catalysed degradation in this and subsequent samples were quenched immediately by addition to an equal volume of quenching/loading buffer (50 mM Tris-HCl pH 6.8, 100 mM DTT, 2% *(w/v)* SDS, 0.1% *(w/v)* bromophenol blue, 30 mM EDTA) and mixed thoroughly. Further samples were taken at various time points throughout the experiment. The samples were then analysed by SDS-PAGE running each sample in duplicate to reduce the impact of pipetting errors during analysis. The rate of degradation was quantified by densitometry as described in Section 2.8.1.

2.8.1 Densitometry

Degradation assays were quantified through the use of densitometry. Images of the SDS-PAGE gels were taken using an IN-Genious SynGene Bio Imaging (Synoptics Ltd.) gel documentation system with the following settings: shutter speed 40 ms, focal length 25 mm, aperture 12, with the focus set half way between 5 m and ∞.

Gene Tools software (Synoptics Ltd.) was used to analyse the gels. The band for undigested substrate protein was selected in each lane along with a band of constant intensity (usually ClpP). The software then converted these bands to a line graph showing each band as a peak on the graph. Care was taken to ensure that the area of the peak selected was applied consistently across the gel. Figure 2.1 illustrates how this was achieved. Both the reference band and band of interest were selected in each lane. The software was allowed to pick the bands itself to ensure consistency and they were only adjusted in cases where it had clearly made an error in picking the bands. Examples of errors encountered are: identifying two proximal bands as a single band (Figure 2.1B) and picking only half of a band (Figure 2.1C). Bands which had been badly selected by the software were manually adjusted to be consistent with the picking of the other bands. The picked bands were then analysed to obtain the total peak volume which accounts for the density and area of the band. As the reference band contained the same amount of protein in each sample and should vary only due to loading errors it was possible to obtain a corrected value for the peak of interest in each lane (V_C) by taking the apparent peak area (V_A) and using the ratio of the peak area of the invariant band in the first

lane (generally time = 0) (I_0) and the peak area of the invariant band for that lane (I_T) using Equation 2.3.

$$V_C = V_A \times \frac{I_0}{I_T}$$

Equation 2.3: Used to calculate the corrected values for densitometry data

In the example in Figure 2.1 V_A is the value obtained for the band of interest in lane 2, I_0 is the value obtained for the reference band in lane 1 and I_T is the value obtained for the reference band in lane 2. This normalises the apparent peak area of the band of interest in lane 2 with respect to the band of interest in lane 1, correcting for any loading errors or differential staining of the gel.

This was applied to each band of interest to obtain corrected values which could then be analysed.

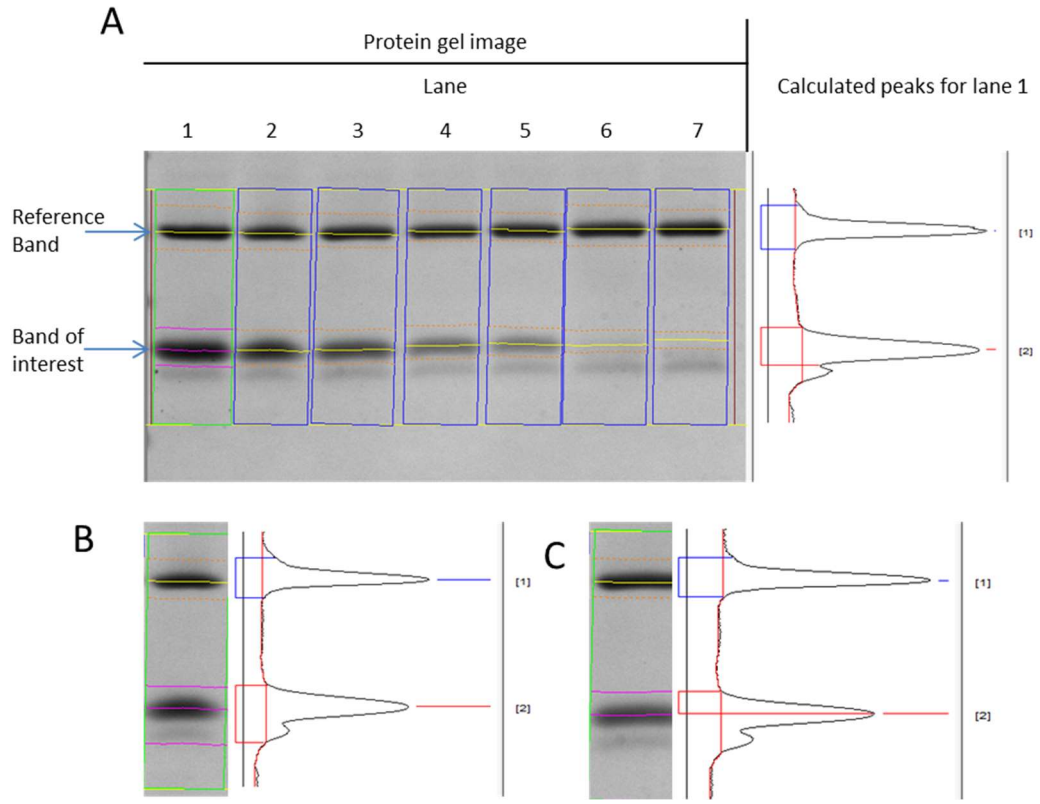


Figure 2.1: Screen capture of a portion of the Gene Tools software demonstrating densitometry analysis of the bands. **A**: typical example of a gel with well picked bands. **B**: an example of a badly picked band, the lower band covers too large an area, **C**: an example of a badly picked band, the lower band covers too small an area.

2.9 Force spectroscopy

2.9.1 Atomic force microscopy slide preparation

Glass microscope slides (76 × 26 mm) were cut into 10 × 10 mm squares using a glass cutter. The glass squares were stuck onto a gold coated (100 Å thickness) silicon wafer (Platypus Technologies) with a small drop of Epo-Tek 377 2 part epoxy resin, taking care not to trap any air bubbles and leaving a minimum 0.5 mm border between the glass squares (Figure 2.2A). The silicon wafer was then heated to 120 °C for one hour to allow the glue to cure. Immediately before use, a glass square was removed from the wafer with a scalpel and attached to a larger glass microscope slide using vacuum grease (Figure 2.2B).

2.9.2 Sample preparation

The lyophilised protein prepared as described in Section 2.6.8 was dissolved in Phosphate Buffered Saline (Dulbecco A), to a concentration of 1 mg/ml then centrifuged in a bench-top centrifuge (16,300 RCF for 10 minutes) to remove any insoluble residue before being added to the gold microscope slide.

2.9.3 Atomic force microscopy

AFM was carried out on an MFP-3D-SA (Asylum Research) atomic force microscope using silicon nitride MLCT contact probes. Typically cantilever D was used with a nominal spring constant of 30 pN/nm.

50 µl of solution containing ~1 mg/ml protein was placed on a freshly template-stripped gold covered glass square. The tip was lowered until the meniscus of the protein solution covered the tip. The laser was then focussed onto the

cantilever and adjusted to give a good signal indicating the reflected beam was centred on the photodetector. The deflection sensitivity of the cantilever was calibrated by engaging the tip and running a single approach-retract cycle then setting the virtual deflection line and the inverse optical lever sensitivity. The spring constant was calibrated using the thermal method[137]; the tip was retracted so it was no longer in contact with the surface, but was within the protein solution and the MFP-3D software thermal tune function was used to obtain a resonance curve for the cantilever. The thermal tune function continually samples the resonance curve and calculates an average over time. This was allowed to run until the resonance curve shown was no longer changing. (1-2 minutes) obtaining a curve as shown in Figure 2.3. Fitting to the first resonance curve provided the spring constant for the cantilever. The tip was then re-engaged and data collection started. A pull distance of 1 μm was used and a balance between mostly single molecule events and a reasonable data collection rate was achieved by altering the contact force and surface dwell time. Approximately one valid event per ten approach-retract cycles with a total of 1,000 cycles was typically achieved.

2.9.4 Data analysis

The data was analysed using the MFP-3D software; a worm like chain[37] fitting tool is used to determine the contour length of the chain at each unfolding event using Equation 1.1. F, T, k_B and x are either constants or are known from the data and L_p is assumed to be 0.4 nm. The difference in L_c between successive events is used to calculate the unfolding length, which is related to the polypeptide length, and the unfolding force can be determined from the

heights of the peaks. The expected unfolding length of I27 was not calculated but a value of 28.1 nm from the literature was used.[138]

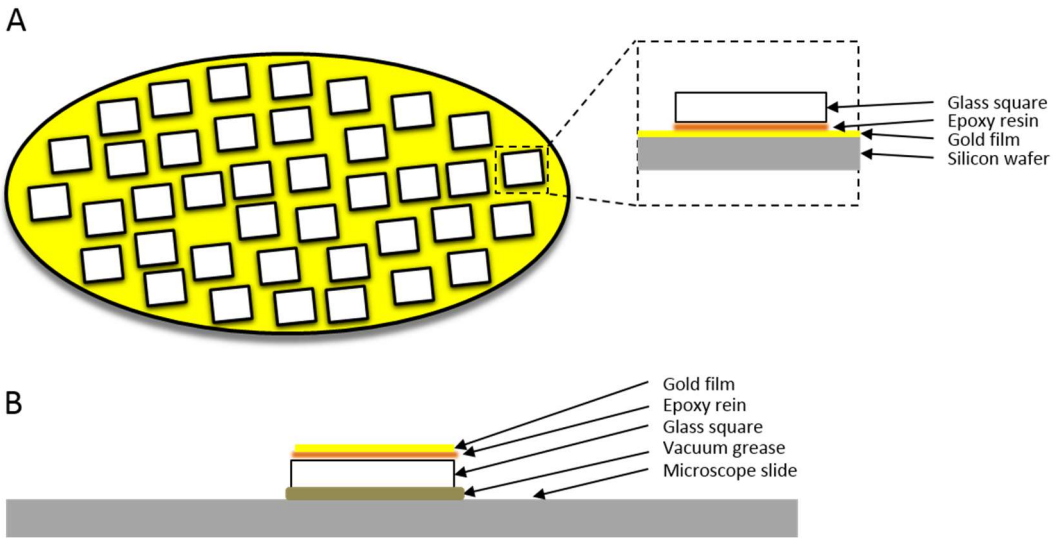


Figure 2.2: Cartoon demonstrating the method of producing gold coated AFM slides. A: the glass squares are attached to the gold coated wafer with epoxy resin. B: The squares are removed from the wafer bringing the freshly exposed gold which was previously attached to the wafer with them, these squares are then attached to a microscope slide with vacuum grease.

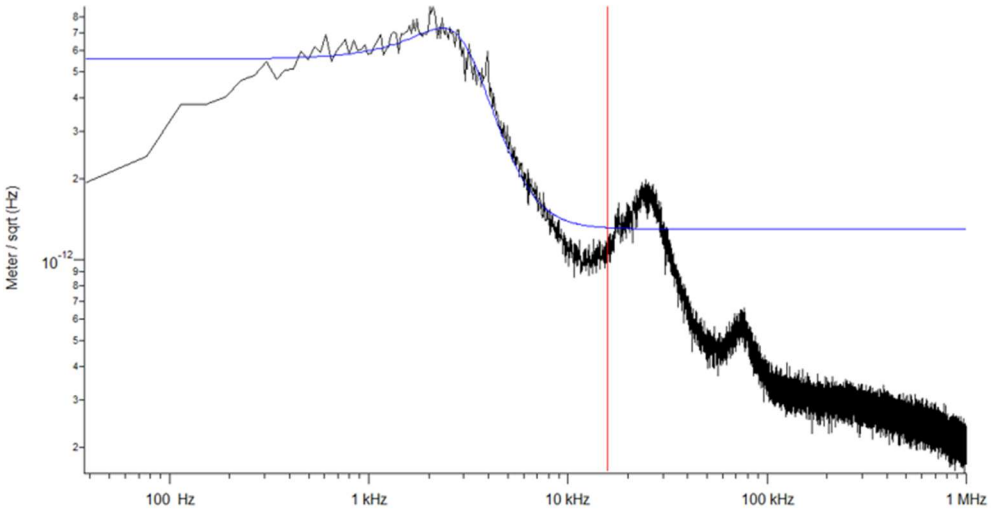


Figure 2.3: Screen capture of thermal tune function showing an acceptable resonance curve (black) and the fit to the 1st peak (blue).

2.10 Biophysical characterisation

2.10.1 Sedimentation velocity analytical ultracentrifugation

A 0.32 mL sample was centrifuged in a 1.2 cm path-length 2-sector epon centrepiece cell built with sapphire windows using an 8-place An50 Ti analytical rotor running in an Optima XL-I analytical ultracentrifuge (Beckman Instruments, Inc., Palo Alto, California 94304) at 48,000 rpm and at a temperature of 20.0˚C. Changes in solute concentration were detected by Interference optics and absorbance scans at 280nm. A total of 100 scans per cell were collected for ~10 hours.

Sedimentation analysis was performed with the program Sedfit[139]. Buffer densities and viscosities were calculated using the program Sednterp[140].

All sedimentation AUC was carried out by Amy M. Barker.

2.10.2 Fluorescence emission spectra

Fluorescence emission spectra were measured using a Photon Technology International Fluorimeter (Ford, West Sussex, UK). For spectra of native and denatured proteins, each protein was dissolved in buffer (10 mM Tris-SO_4, 1 mM EDTA, 42 mM K_2SO_4, 0.5 mM TCEP pH 7.6) containing 0 M or 8 M urea (protein concentration ~ 5 μM) and incubated at 30 °C for one hour. Excitation slit widths were set to 2 nm, emission slit widths were adjusted to give a good signal. Each spectrum was recorded from 290 nm to 500 nm in 1 nm increments, using an excitation wavelength of 280 nm. Spectra of all denatured states were assumed to have the same maximum intensity. The spectra of each native protein were normalised to the intensity of their

respective denatured state, allowing direct comparison of the fluorescence intensity between proteins.

2.10.3 Circular dichroism spectroscopy

Circular dichroism spectroscopy (CD) was performed using a J-715 spectrophotometer (JASCO Corp.). Measurements were taken between 190 and 260 nm in a 1 mm path-length cuvette. The recording parameters were: 1 nm bandwidth, 1 nm resolution, 20 nm/min scan speed, 8 second response time and the average of 3 scans was taken. Spectra were measured at 30 °C.

Observed ellipticity, after subtraction of the buffer only sample was converted to mean residue ellipticity by use of Equation 2.4.

$$[\theta]_{molar,\lambda} = \frac{100 \times \theta_{\lambda}}{m \times d}$$

Equation 2.4: For calculation of mean residue molar ellipticity Where: $[\theta]_{molar,\lambda}$ is mean residue molar ellipticity, θ_{λ} is the observed ellipticity (degrees), d is the path length (cm) and m is the mean residue molar concentration.

3 Production and testing of the ClpXP system

3.1 Introduction

As described in section 1, ClpXP is a bacterial proteasome capable of degrading diverse protein substrates if suitably tagged. Most previous work reported for ClpXP has focussed on ClpXP itself; determining its structure, function and mechanism of action. Great progress had been made in this field and both ClpXP structure and function are now well understood[12,47]. Although the precise mechanism of action and processivity are still unknown, its general mechanism has been elucidated[12,47]. In summary: In the presence of ATP, ClpX forms a hexameric ring. These Hexamers can then bind to one or both ends of a cylinder composed of two heptameric rings formed from ClpP monomer; the pore loops of ClpX are capable of recognising and binding a degradation tag; ATP hydrolysis occurs causing a conformational change in ClpX which drives translocation of the substrate through the central pore of ClpX. This applies force onto the substrate accelerating its unfolding rate; the unfolded peptide is then moved into the central degradation chamber of ClpP where it is degraded in a non-sequence-specific manner by the serine protease sites contained in the degradation chamber, finally the degradation products (short peptides) are then released by the ClpP.

For most stable globular proteins the rate-determining step of this mechanism is the unfolding rate of the substrate peptide. For active degradation this can be assumed to be dependent on the mechanical stability of the substrate (see introduction). Following the rate of degradation of a folded protein by ClpXP by analysing the disappearance of the substrate band on an SDS-PAGE gel

will allow investigation of the mechanical resistance of a degradation tagged substrate peptide at loading rates that are inaccessible by AFM but found in *in vivo* protein degradation and other processes.

In this chapter I will describe my attempts to develop ClpXP as a biological applicator of mechanical force which can be used to investigate the mechanical strength of a range of tagged substrate peptides.

3.2 **Aims**

The aims of this chapter are (i) to describe the expression and purification of a selection of ClpX variants, ClpP and suitable substrate proteins, (ii) develop a robust degradation assay and (iii) investigate a variety of conditions under which degradation can be observed.

3.3 ClpXP production

All of the ClpX and ClpP variants used were based on the published *E. coli* ClpX/P sequence (see Appendix – Section 7.1) and inserted into plasmids containing the T7 promoter from the T7 bacteriophage (pET23, pET14 and pAYC). Overexpression was subsequently achieved by transformation of *E. coli* [DE3] which are lysogenic for λ-DE3, and contain the T7 bacteriophage gene I, encoding T7 RNA polymerase under the control of the *lac* UV5 promoter allowing inducible expression by the addition of IPTG[141].

3.3.1 ClpX

An array of ClpX variants was used in this work. These are summarised in Figure 3.1 and described in detail in Appendix – Section 7.1.1.

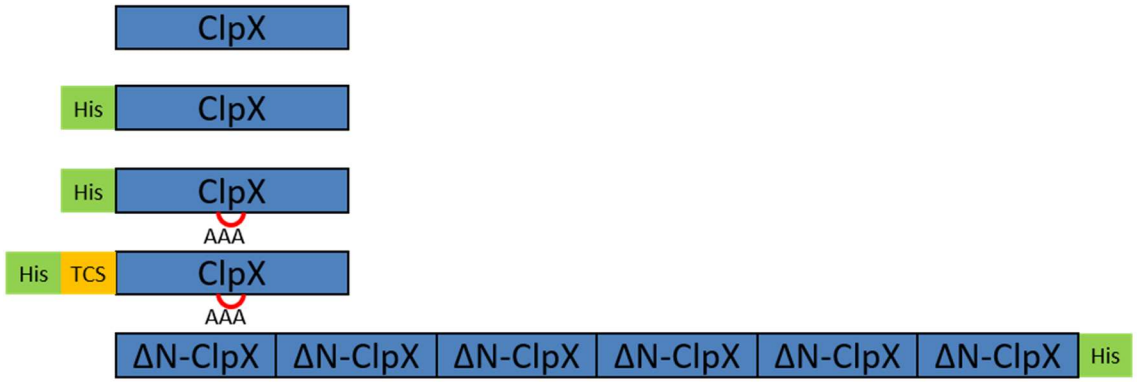
3.3.1.1 His_6-ClpX

The first variant produced was His_6-ClpX with an N-terminal hexahistidine affinity tag to facilitate purification. This had been previously studied in the Brockwell group and the plasmid (pET23His_6-ClpX) was available in the laboratory.

pET23His_6-ClpX was transformed into *E. coli* BL21 [DE3] pLysS (Section 2.2.7) and a test expression was performed as described in Section 2.5.1. Figure 3.2 shows the results from the test expressions from four single colonies. A band with significant density was present in both the soluble and insoluble fraction at the expected mass for His_6-ClpX. It was decided not to attempt to solubilise the insoluble protein as sufficient was present in the soluble fraction for further purification. This obviates the requirement to find

suitable conditions (if any) that would allow the refolding of denatured His_6-ClpX to an active form.

A large scale expression was then carried out using 10 l of LB (Section 2.5.3) and the cells lysed (Section 2.6.1) and a batch purification using nickel affinity followed by size-exclusion chromatography (Section 2.6.4) was carried out. The final product was concentrated to 1 μM and snap frozen as shown in Figure 3.3.



Name	Size /A.A.	Calculated Mass /Da
$ClpX^{WT}$	424	46356
His_6-ClpX	431	47310
His_6-$ClpX^{\Delta RKH}$	431	47102
Thrombin cleaved His_6-$ClpX^{\Delta RKH}$	444 (427 cleaved)	48311 (46429 cleaved)
$^{\Delta N}ClpX_6$	2305	247192

Figure 3.1: schematic representation of ClpX variants used in this work. His - hexahistidine tag, TCS - thrombin cleavage site, ClpX - ClpX sequence, ΔN-ClpX - ClpX with N-terminal domain removed.

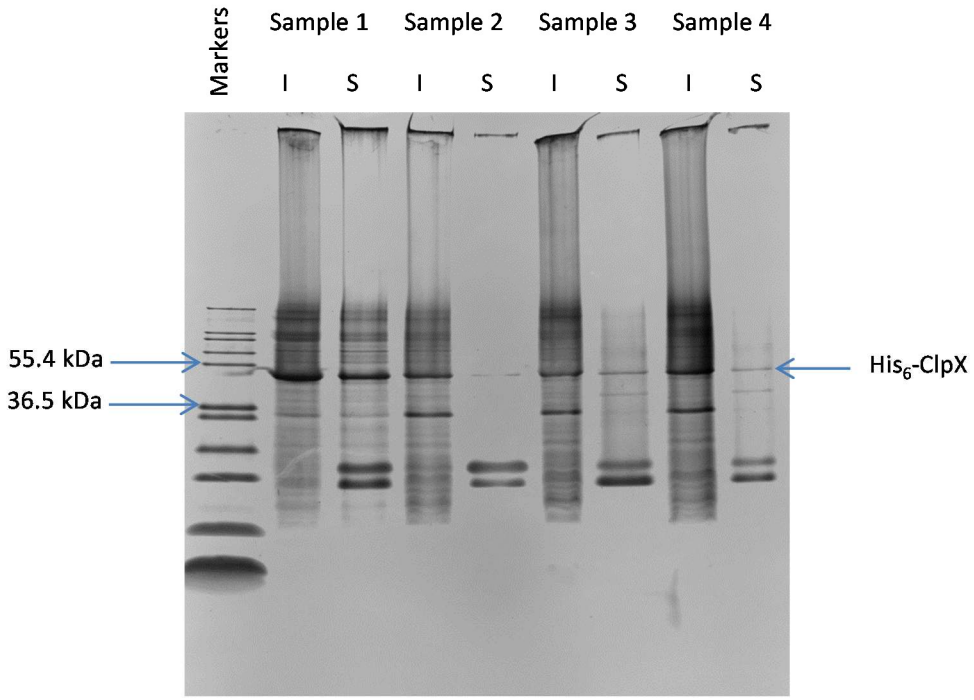


Figure 3.2: SDS-PAGE gel of His_6-ClpX test expression. After cell lysis the soluble and non-soluble proteins were separated by centrifugation then the supernatant and resuspended pellet were analysed by SDS-PAGE. Four test expressions were analysed, labelled Sample 1-4. Fractions are labelled as follows. **I:** re-suspended pellet, **S:** supernatant. The migration distance expected for His_6-ClpX is shown on the right.

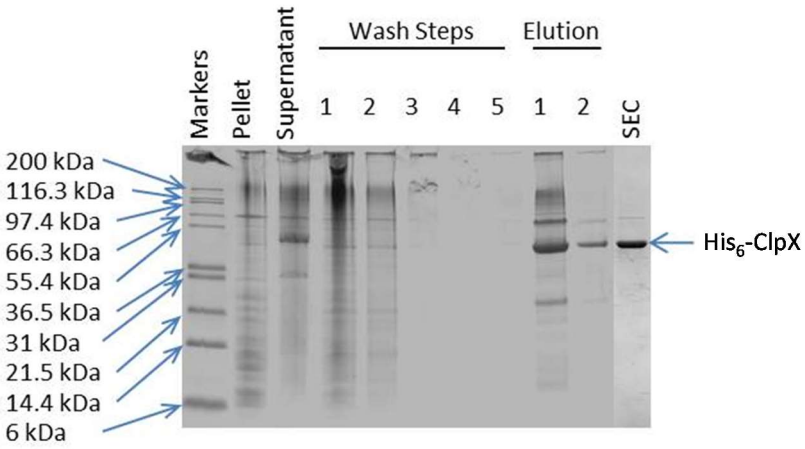


Figure 3.3: SDS-PAGE gel showing the batch purification of His_6-ClpX. The lanes marked pellet and supernatant show the fractions obtained after cell lysis and centrifugation, the resuspended pellet and the supernatant containing the soluble proteins respectively. The lanes marked wash steps show the supernatant obtained after washing the Ni sepharose beads with Ni sepharose wash buffer then centrifuging the sample to pellet the beads, numbers refer to repeated washes. The lane marked elution shows the supernatant obtained after washing the beads with elution buffer (wash buffer with the addition of 500 mM imidazole). Numbers refer to repeated washes. The lane marked SEC shows the final His_6-ClpX protein obtained after size-exclusion chromatography and concentration.

3.3.1.2 $^{\Delta N}ClpX_6$

In the absence of ATP the ClpX homohexamer is unstable and prone to dissociation and inactivation. The Sauer group (MIT) have overcome this limitation by constructing a pseudohexameric single chain ClpX variant with a C-terminal $(His)_6$ tag (Section 1.4.1). This group had worked extensively and successfully with this variant.[30,54,61,135].The plasmid containing $^{\Delta N}ClpX_6$, a pAYC derivative (pAYC$^{\Delta N}ClpX_6$) was, therefore obtained from the Sauer group and transformed into BLR (DE3) pLysS cells and a test expression performed (Figure 3.4). While a band was observed which was consistent with that expected for $^{\Delta N}ClpX_6$ ($^{\Delta N}ClpX_6$ was too large to enter the resolving gel with a molecular weight of 247 kDa), it was present only in the insoluble fraction.

A further expression trial was undertaken, where the culture was incubated at 22 °C upon addition of IPTG. It was rationalised that such a large protein was vulnerable to misfolding and aggregation, and expression at lower temperature has been shown to minimise these effects due to the temperature dependence of the hydrophobic interactions (which determine the rate of protein aggregation), the increase in expression and activity of *E. coli* chaperones, and the reduced induction of heat shock proteases[142]. The lower temperature also slows the rate of RNA translation and so protein production allowing the protein more time to fold successfully as it leaves the ribosome. Lower temperature expression did indeed improve the soluble expression of $^{\Delta N}ClpX_6$ and the results of a 22 °C expression trial are shown in Figure 3.5. Although a large amount was also present in the insoluble fraction it was felt that enough was present in the soluble fraction for it to be worthwhile continuing with a large-scale overexpression and purification.

A large-scale over-expression using 10 l of LB broth with chloramphenicol cultured at 22 °C after induction was performed and the cells harvested and lysed as described in Section 2.6.1 and purified using Ni sepharose followed by size-exclusion chromatography carried out as described in Section 2.6.4. This initial purification protocol resulted in a protein of low purity and the method was refined by using several wash steps with a variety of different imidazole concentrations to try and minimise non-specific binding of contaminant proteins in the first purification step. As can be seen from Figure 3.6 it was not possible to find the ideal conditions that would wash off the contaminants while leaving the $^{\Delta N}ClpX_6$ bound to the resin. The eluate from the 100 and 500 mM imidazole washes was thus pooled and concentrated (Vivaspin 20, 30,000 MWCO (Sartorius Stedim)) to a volume of 5 ml and further purified by size-exclusion chromatography (Sephacryl S-300 HR, using ClpXP storage buffer as the elution buffer) (Figure 3.7). Fractions 15-17 were pooled and concentrated to 1 µM, divided into 20 µl aliquots then snap-frozen in liquid nitrogen before storage at -80 °C. It is noticeable in Figure 3.7 that a second protein band is visible at approximately 66 kDa, this band was present in all $^{\Delta N}ClpX_6$ purifications carried out and could not be eliminated entirely. Since the amount of contaminant present was low, it did not appear to affect the activity of the purified $^{\Delta N}ClpX_6$ and due to the sensitivity of ClpX to prolonged purification it was decided that the $^{\Delta N}ClpX_6$ produced was pure enough for use in further experiments. If it were necessary to quantify the exact concentration of $^{\Delta N}ClpX_6$ then densitometry could be used to estimate the proportion of this contaminant present.

While it is the ClpX component of the ClpXP complex that acts as a mechanical unfoldase which is able to unfold substrate proteins and dissociate

aggregates in isolation, it is not possible to use ClpX independently in an ensemble assay as any unfolded protein would be able to refold after translocation through the ClpX ring. Monitoring this by SDS-PAGE (which is how the degradation was to be followed) would thus lead to no change in signal. It was therefore necessary to express and purify ClpP to use in conjunction with the ClpX. This renders the mechanical unfolding step of the substrate irreversible and allows the rate of unfolding (rate limiting step) to be measured by rate of disappearance of full-length protein by SDS-PAGE coupled with densitometry.

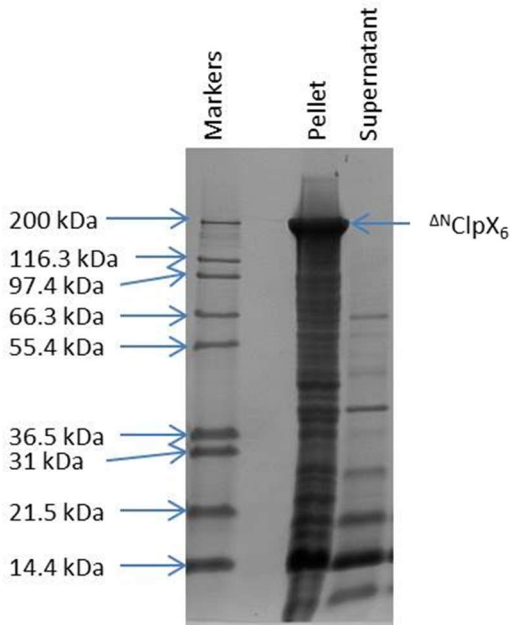


Figure 3.4: SDS-PAGE gel of $^{\Delta N}ClpX_6$ test expression after cell lysis. The cell lysate was centrifuged to separate the soluble and non-soluble proteins then the supernatant and re-suspended pellet were analysed by SDS-PAGE. The lane labelled Pellet shows the re-suspended pellet, the lane labelled Supernatant shows the supernatant containing the soluble proteins.

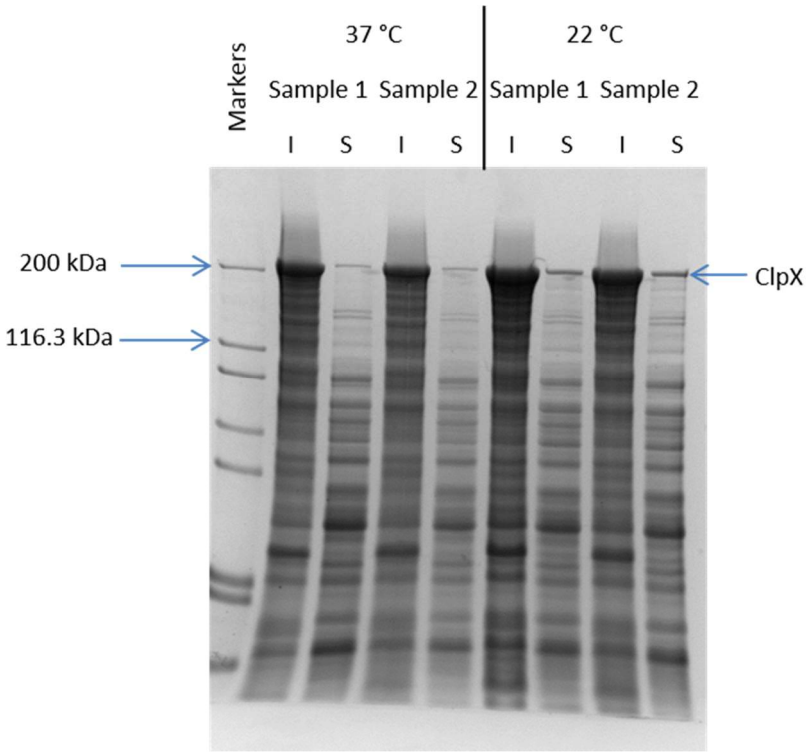


Figure 3.5: SDS-PAGE gel of $^{\Delta N}ClpX_6$ test expression at two different temperatures after cell lysis. The cell lysate was centrifuged to separate the soluble and non-soluble proteins then the supernatant and re-suspended pellet were analysed by SDS-PAGE. Four test expressions were analysed, two at 22 °C and two at 37 °C. Fractions are labelled as follows: I - re-suspended pellet, S - supernatant. The migration distance expected for $^{\Delta N}ClpX_6$ is shown on the right, this is at the interface of the stacking and resolving gels as the expressed protein is too large to enter the resolving gel.

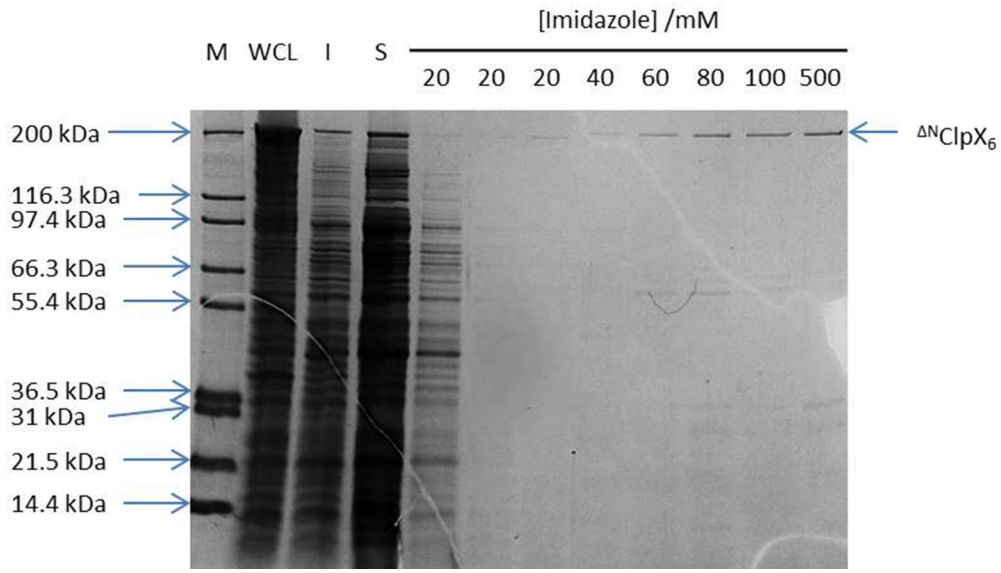


Figure 3.6: SDS-PAGE showing the initial purification step of $^{\Delta N}ClpX_6$. The lane marked M shows the molecular weight markers. The lanes marked WCL, I and S show: the whole cell lysate before centrifufation (WCL), and after centrifugation the re-suspended pellet containing the insoluble proteins (I) and the supernatant containing the soluble proteins (S), which was used for further purification. The lanes showing imidazole concentrations are the supernatant from sequential wash-steps with the stated imidazole concentration after centrifugation to pellet the Ni sepharose beads. The migration distance expected for $^{\Delta N}ClpX_6$ is shown on the right, this is at the interface of the stacking and resolving gels as the expressed protein is too large to enter the resolving gel.

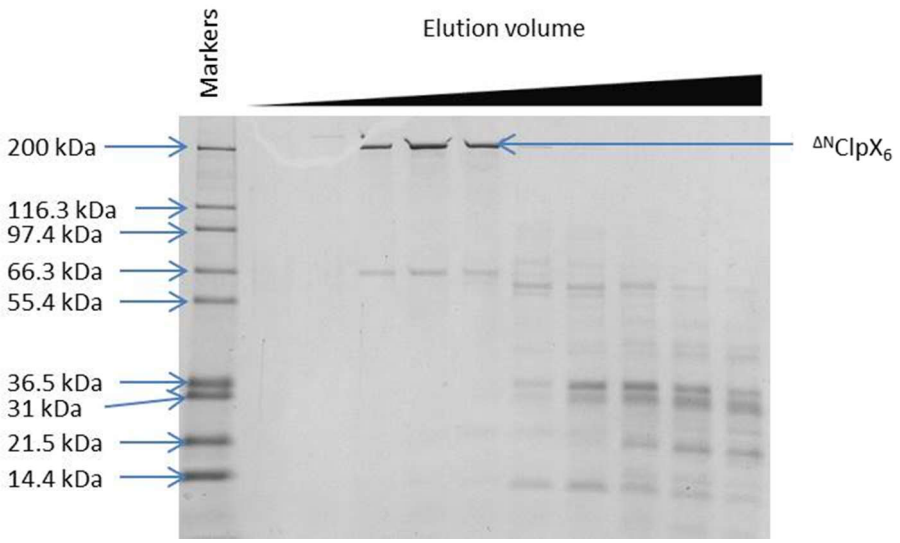


Figure 3.7: SDS-PAGE showing fractions which showed an increase in absorbance at 280 nm eluted from size-exclusion column during the purification of $^{\Delta N}ClpX_6$. The migration distance expected for $^{\Delta N}ClpX_6$ is shown on the right, this is at the interface of the stacking and resolving gels as the expressed protein is too large to enter the resolving gel.

3.3.2 ClpP

The ClpP used throughout this study is the *E. coli* ClpP (Appendix 7.1.2). It is inserted into a pET14 derived vector conferring ampicillin resistance. pET14ClpP was transformed into *E. coli* Bl21 [DE3] (Section 2.6.3). An initial test expression was performed as shown in Figure 3.8. ClpP was found to express well and was found mostly in the soluble fraction, there was another prominent band present in the soluble fractions at approximately 14 kDa, no attempt was made to identify this as it would be expected to be removed by purification but it is likely to be lysozyme (Molecular mass = 14,307 Da) used in the cell lysis procedure. A large scale overexpression using 10 l of LB medium was performed and the resulting cells harvested and lysed before purification as described in Section 2.6.3.

The initial attempt at purification using ammonium sulphate precipitation and anion exchange chromatography steps were unsuccessful and resulted in a poorly purified protein as shown in Figure 3.9. As this was unsuccessful a new strategy was devised. A ClpP variant with a C-terminal hexahistidine affinity tag had been constructed previously in the Brockwell group (D. Brockwell, personal communication) but had not been used as it was unknown whether the tag would interfere with ClpP complex formation. As $(His)_6$ purification greatly facilitates production of high purity proteins it was decided to purify ClpP$(His)_6$ as described in section 2.6.3, because if this was successful it would greatly reduce the amount of work required for ClpP production during the course of this project. A transformation and expression trial was carried out successfully as described above and a small scale test purification was carried out in order to optimise the purification strategy. A successful scheme

was devised using several different washes with different imidazole concentrations (Figure 3.10). After purification ClpP was dialysed into ClpXP storage buffer and snap frozen as described for $^{\Delta N}ClpX_6$.

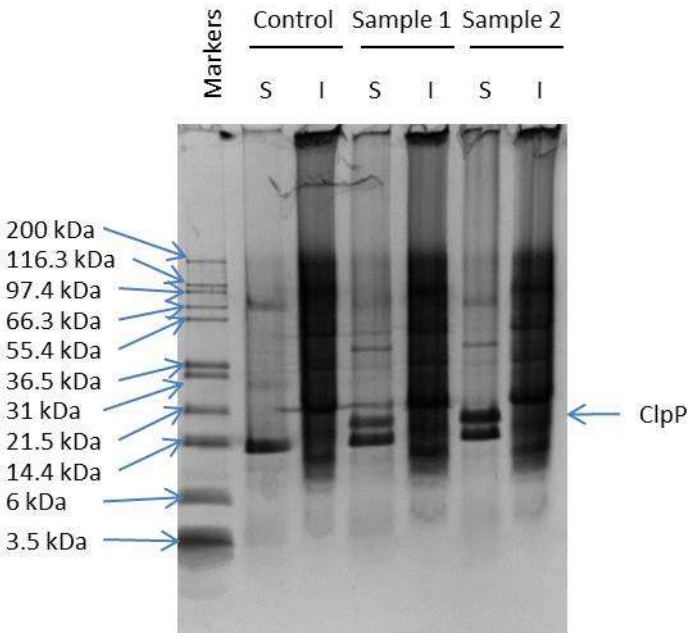


Figure 3.8: SDS-PAGE of lysed and centrifuged samples from a ClpP expression trial. The lanes marked Control show an uninduced control, Samples 1 and 2 have been induced with 1 mM IPTG and grown for a further 4 hours. I and S refer to the re-suspended pellet containing the insoluble proteins and the supernatant containing the soluble proteins respectively. The migration distance expected for ClpP is shown on the right.

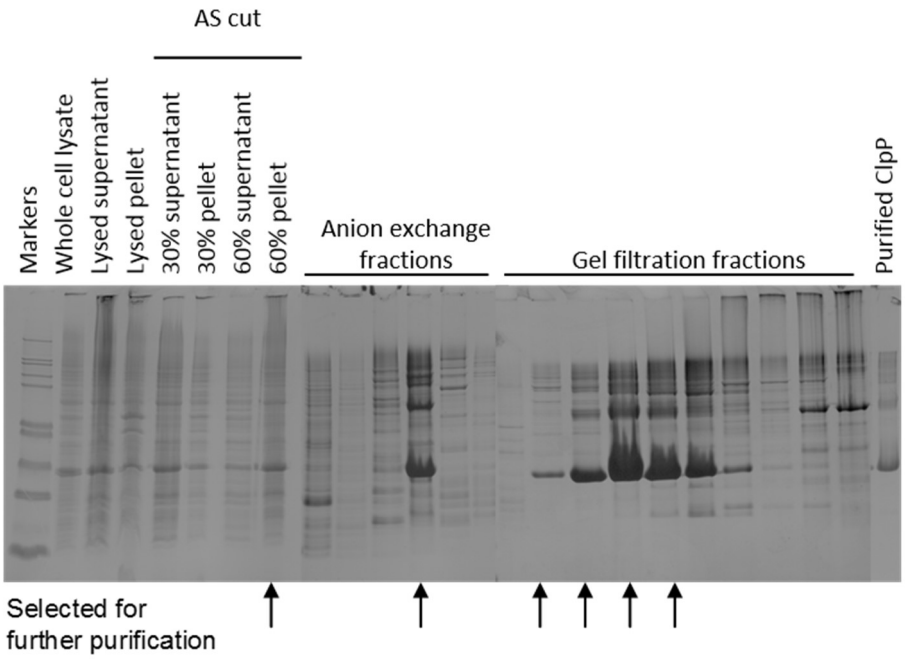


Figure 3.9: Composite of SDS-PAGE gels showing the results of the ClpP purification protocol. Whole cell lysate, Lysed supernatant and Lysed pellet show the results of the initial cell lysis and centrifugation. The section labelled AS cut refers to the ammonium sulphate precipitation step of the purification. Ammonium sulphate was added to a saturation of 30% and 60%, separating the resulting suspension at each stage by centrifugation. Shown are the supernatant containing the soluble proteins and the pellet containing precipitated proteins (after it had been redissolved). The fraction which was insoluble at 60% ammonium sulphate was selected for further purification by anion exchange (Anion exchange fractions). The fractions indicated by arrows were selected for further purification by size exclusion chromatography (Gel filtration fractions). The fractions indicated by the arrows were pooled, concentrated and frozen. A sample of this is shown labelled Purified ClpP.

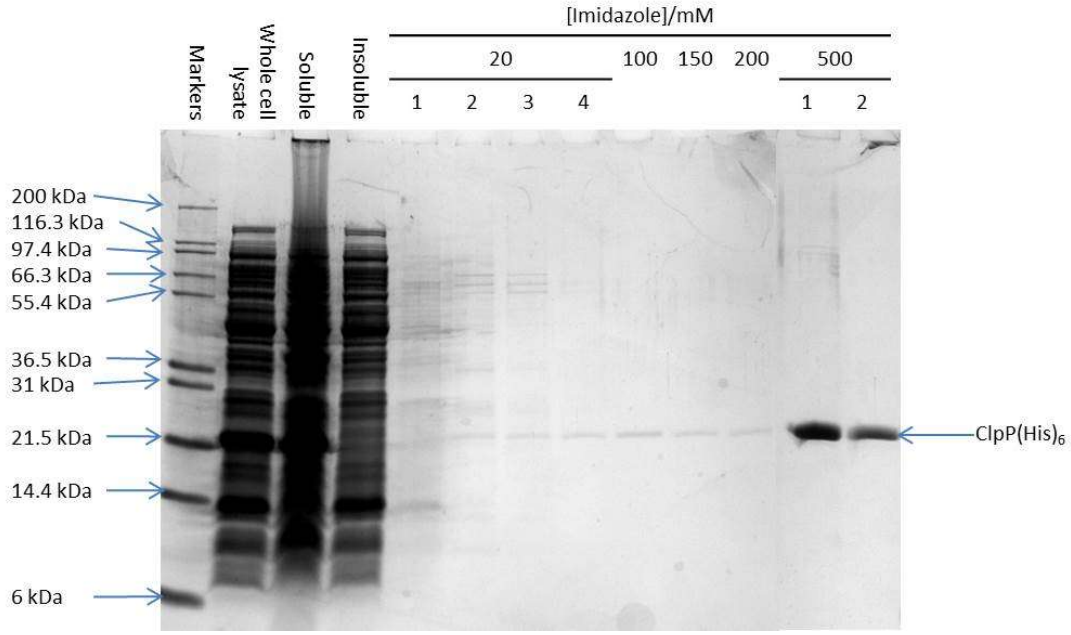


Figure 3.10: SDS-PAGE showing purification of ClpP. Nickel resin was washed 4 times with a buffer containing 20 mM imidazole, a single wash at each of 100, 150 and 200 mM and eluted with 2 washes at 500 mM.

3.4 Degradation assays

In addition to determining its approximate mass by SDS-PAGE, an excellent test to verify the identity of the proteins isolated in sections 3.3.1.2 and 3.3.2 is the ability of these proteins to assemble and carry out ATP dependent degradation of tagged substrates. The substrate chosen for this was Im9-ssrA. This is the *E. coli* immunity protein 9; produced as a defence against the bacterial E colicin, E9 which is produced as part of the SOS stress response, with the ssrA degradation tag added to the C-terminus. It has been used previously in the Brockwell group as a substrate for ClpXP and a large stock of lyophilised ssrA-tagged Im9 produced by James Pullen was available for use and has so far proven to be sufficient for all of the experiments for which it has been required.

3.4.1 Initial degradation

The reaction mixture for the degradation assay contains: ClpXP to unfold and degrade the substrate protein; Tris-HCl (pH 7.6), KCl, glycerol and DTT to provide the conditions in which ClpXP is stable and active; $MgCl_2$ to provide the Mg^{2+} ions necessary for the hydrolysis of ATP; creatine phosphokinase and creatine phosphate to act as an ATP regeneration system; and ATP to provide the energy source for the ClpX and allow complex formation. ATP was omitted in the control reactions. The reaction mixture and substrate was pre-warmed to the reaction temperature (30 °C), the substrate added and samples of the mixture are taken at the appropriate time-points for analysis, the results are shown in Figure 3.11.

The initial reaction conditions were based on previously successful experiments performed in the Brockwell group (J. Pullen, personal communication). Optimisation concentrated on changing the concentrations of ClpX and ClpP in the reaction in order to have the reaction take place in a time frame which made a good balance between the time taken for the experiments and allowed a reasonable number of samples to be taken for analysis. It was attempted to keep reaction times to under 1 hour while taking around 8 samples. An example of typical reaction components is given in Table 3.1. The concentration of most components remained the same, the ones that might typically be varied was the ClpX and ClpP concentrations in order to alter the rate of the reaction. The fraction of ClpXP which was active was not quantified.

While it is the ClpX component of the ClpXP complex that acts as a mechanical unfoldase which is able to unfold substrate proteins and dissociate aggregates in isolation, it is not possible to use ClpX independently in an ensemble assay as any unfolded protein would be able to refold after translocation through the ClpX ring. Monitoring this by SDS-PAGE (which is how the degradation was to be followed) would thus lead to no change in signal. It was therefore necessary to express and purify ClpP to use in conjunction with the ClpX. This renders the mechanical unfolding step of the substrate irreversible and allows the rate of unfolding (rate limiting step) to be measured by rate of disappearance of full-length protein by SDS-PAGE coupled with densitometry.

Table 3.1: Typical concentrations of components used in ClpXP mediated degradation reactions. ClpX and ClpP concentrations sometimes varied from this, if so it is stated in the experimental description. ATP was omitted from control reactions. pH was adjusted to 8 by the addition of HCl.

Reaction Component	*Concentration*
Tris base	25 mM
KCl	200 mM
Glycerol	10% v/v
DTT	2 mM
$MgCl_2$	10 mM
Creatine phosphokinase	0.032 mg/ml
Creatine phosphate	16 mM
ATP	5 mM
ClpX (pseudohexamer or homohexamer)	200 mM
ClpP (homo-14mer)	600 mM

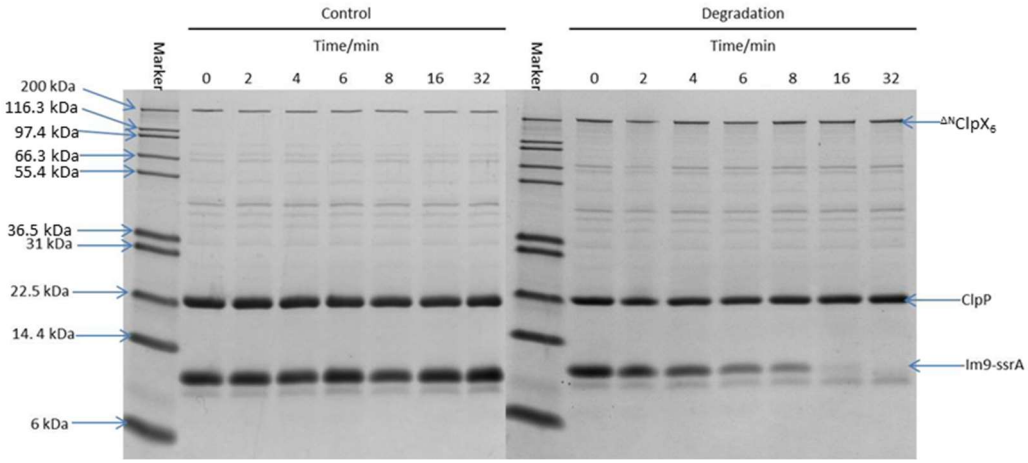


Figure 3.11: Initial test degradation showing that $^{\Delta N}ClpX_6P$ is active against an ssrA-tagged substrate.

3.4.2 Assay validation

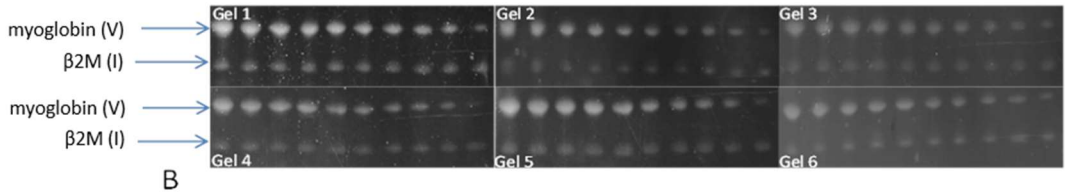
In order to determine whether it was possible to use SDS-PAGE gels in conjunction with densitometry to follow the progress of a degradation reaction accurately it was necessary to conduct an experiment where a known amount of protein was loaded onto a gel then densitometry used to quantify the resultant bands. Two proteins were chosen that were readily available: Beta-2-microglobulin (prepared by Nasir Khan) (β2m) and myoglobin from equine heart (Sigma).

β2m has a molecular weight of 11,860 Da and myoglobin 17,600 Da. The β2m was used as an invariant standard to correct for loading errors and 1 μg was loaded in each lane, myoglobin was used as the variable protein and was loaded in amounts varying from 0.25 μg to 3 μg which spanned the range that would be used in the degradation assays. To quantify the error of such an approach, six gels were loaded and run in identical conditions. The gels were stained by two methods: SYPRO red protein gel stain (Lonza, USA) (Figure 3.12A), which is a fluorescent stain visible under UV light (λ_{EX} = 300 nm, λ_{EM} = 550/630 nm)) whose manufacturer claims it provides a superior staining with a detection limit to Coomassie of 1ng/band[143] and has a linear range between ng/band to μg/band allowing accurate quantification over a greater range of protein quantity. The second method was using a Coomassie-based stain (sensitivity = 8ng/band)[143] (Instant Blue, Expedion) (Figure 3.12B). The intensity of each band was determined and normalised to the reference protein β2M as described in Section 2.8.1. The results are shown in Figures 3.13 and 3.14.

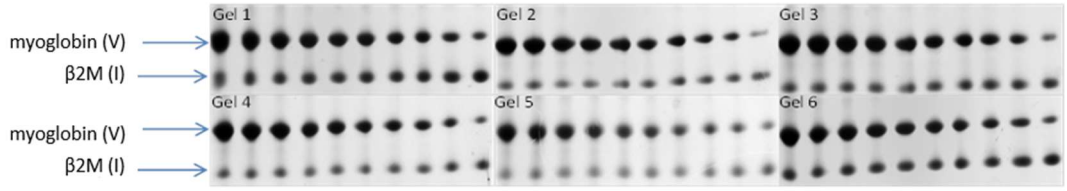
The two dyes stain by different methods and allow detection in different ways. Coomassie-based staining involves the binding of a coloured dye to the protein by a combination of electrostatic and Van der Walls interactions and is known to stain proteins well, particularly those containing basic residues[144]. Due to its preference for basic residues Coomassie can provide different levels of staining for different proteins depending on their amino acid composition but as we are interested in quantifying different concentrations of the same protein this need not concern us. SYPRO red stain is a fluorescent dye which binds to the SDS coat proteins acquire during SDS-PAGE[143]. I feel that the SYPRO red staining, by virtue of being a light emitting stain in contrast to being an absorbing stain like Coomassie, probably did in theory provide a result that would give better staining over a wide range of protein concentration. This is due to the way that the camera records the result; with a light emitting stain it is possible to adjust the exposure so that the brightest band is not overexposed and the other bands will then be less bright. Brighter bands should emit more light if they contain more protein. Whereas with a Coomassie stain once a band is stained to its maximum darkness (where no light can be transmitted through that band) it cannot be determined whether it has just reached that level or has surpassed it, and what you are measuring becomes the size of the band. In practice however there is no evidence that the SYPRO red is either more sensitive or has a greater range of linearity and in fact as shown by the error bars the variation between gels was much greater than for Coomassie staining. This might be because the staining with SYPRO red tended to be very uneven across the gel with a varying intensity in the background staining and the occasional bright spot that did not seem to be a protein band. You can see this in Figure 3.12A, gel 2 shows an example of

uneven staining across the gel, gel 1 shows a “bright spot” which is in the middle of a protein band, gel 4 has examples of both problems. These problems meant that while in theory the SYPRO red had the ability to provide a more accurate assay; in practice Coomassie was often better because the staining was more likely to be even across the gel and no artefacts in the background staining would interfere with the densitometry. Consequently, in future assays, the gels were stained with SYPRO red and if the staining was even with few artefacts these were used; if the SYPRO red staining was poor the gels were re-stained with Instant Blue and these used instead for quantification. It should be noted that while the Beer-Lambert law says that the relationship between the signal and the amount of protein being measured is linear and so it would be expected that in Figure 3.13 and Figure 3.14 that the line of best fit should pass through the origin. The calibration graphs show that this method over-estimated the volume for small values, this is likely due to errors in estimating the baseline. However this problem can be overcome by careful selection of the concentration of protein used in degradation assays to ensure that the majority of the assay will take place over the linear range.

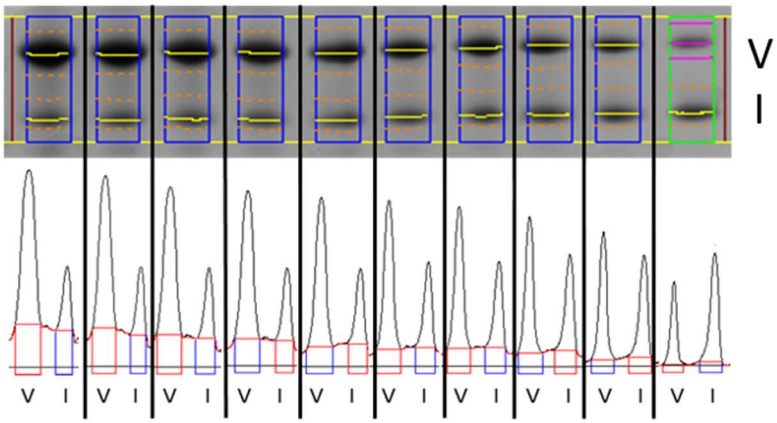
A



B



C



D

Protein mass /ug	I raw vol.	V raw vol.	Relative V raw vol.	V corrected vol.	Relative V corrected vol.
3	21,498,890	53,854,952	1.00	53,854,952	1.00
2.625	21,830,096	51,216,064	0.95	50,439,014	0.94
2.25	21,451,516	46,293,528	0.86	46,395,764	0.86
1.875	23,738,886	42,418,512	0.79	38,415,911	0.71
1.5	23,412,064	38,239,832	0.71	35,114,971	0.65
1.25	25,308,810	37,734,028	0.70	32,053,649	0.60
1	26,187,120	33,306,758	0.62	27,343,913	0.51
0.75	28,408,766	30,933,696	0.57	23,409,680	0.43
0.5	33,486,128	27,346,482	0.51	17,557,091	0.33
0.25	36,625,868	18,735,870	0.35	10,997,703	0.20

Figure 3.12: SDS-PAGE gels used to determine the linear range over which protein sample concentration could be accurately quantified. Each gel stained with (A) SYPRO red and (B) Instant Blue. (C) is a composite image of various screen-grabs from the Gene Tools software showing the bands picked on the gel image. Yellow lines are the maximum intensity and dashed orange lines are the boundaries of the band, shown below the gel image are the intensity profiles measured for each band with V being the upper variant band and I being the lower invariant band. The volume under each intensity profile is then calculated and the values for the variant bands normalised as described in section 2.8.1. (D) shows a comparison between the raw volumes and relative (to 1) raw volumes and corrected volumes for the data shown in C. The relative corrected volumes were then averaged and plotted as shown in Figures 3.13 and 3.14.

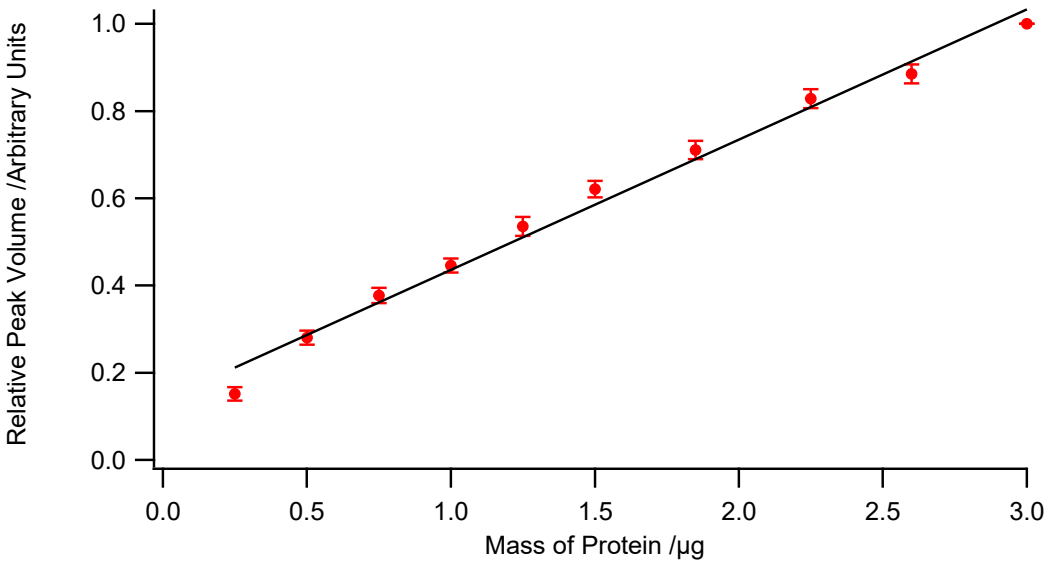


Figure 3.13: Analysis of densitometry from loading a known amount of protein when using Coomassie-based stain. Error bars represent the standard error of the mean from the 6 replicate gels. The solid black line is a best-fit to the mean values $r^2 = 0.99$

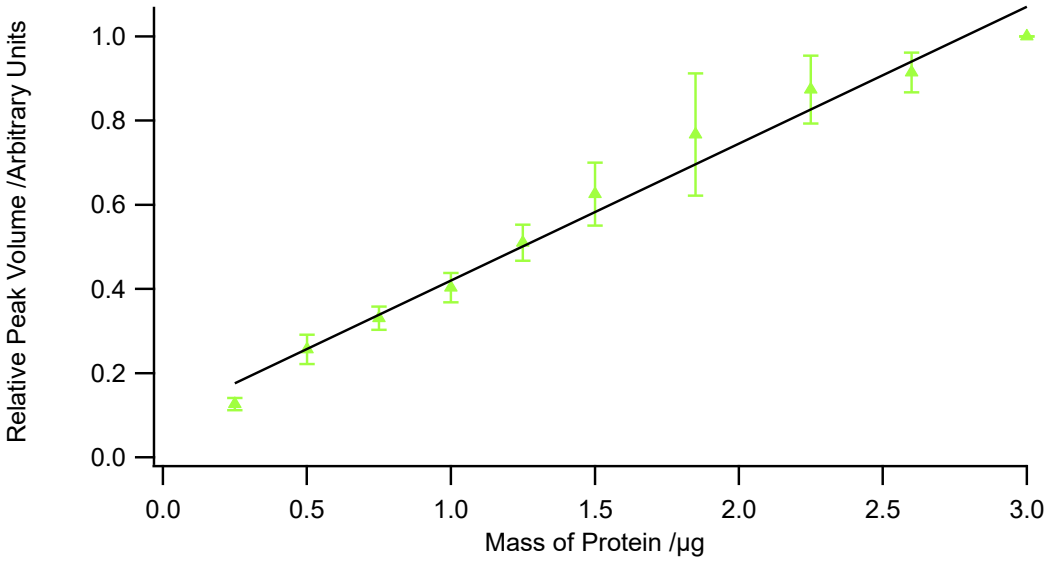


Figure 3.14: Analysis of densitometry from loading a known amount of protein when using SYPRO red stain. Error bars represent the standard error of the mean from the 6 replicate gels. The solid black line is a best-fit to the mean values $r^2 = 0.98$.

3.5 DTT requirement

One potential application of using ClpXP as a biologically relevant force applicator would be investigating the possibility of altering the mechanical strength of a protein for degradation (or a ligand for such a protein) by the introduction or removal of specific disulphide bonds within the protein[35]. While it is of course possible to use protein engineering to add or remove such disulphide bonds by adding or removing cysteines from the protein; a simpler method is to simply make or break these bonds by performing the experiment in a reducing or non-reducing environment. However, upon searching the ClpXP literature it was unclear whether reducing agents are required for ClpXP activity. The only reference was as follows: “Because reducing agents are important for robust ClpXP activity but would cleave disulphides, we used the HaloTag domain to allow an alternative method of covalent linkage to the DNA.”[30]. However no evidence was given for this statement and some studies have been performed which did not mention the use of reducing agents in their experimental protocol. It was decided to investigate the claim that reducing agents were required.

$^{\Delta N}ClpX_6$ (200 nM) was used with $ClpP_{14}$ (600 nM) and 10 μM test substrate (Im9-ssrA, which does not contain disulphide bonds) and degradations were carried out in both reducing (2 mM DTT) and non-reducing conditions. Figure 3.15 shows the raw results from a series of assays including 2 mM DTT and shows the same experiment without DTT. Figure 3.16 shows the compiled results of the reaction, plotting the mean amount of Im9-ssrA remaining (calculated relative to the 10 μM Im9-ssrA present at time = 0).

As can be seen from Figure 3.15, degradation by $^{\Delta N}ClpX_6P$ is still robust irrespective of the presence of a reducing agent, in this case DTT. However the rate of degradation is slowed in the absence of DTT (Figure 3.16, Table 3.2). If it was desired to investigate the mechanical strength of a protein in the presence and absence of disulphide bonds it will be necessary to disrupt them by removing one of the cysteines from the protein by mutagenesis.

The reason degradation is enhanced in reducing conditions is unclear, because the precise mechanism of action of ClpX is not known. It has been reported that the presence of DTT can have effects on protein function not directly related to disulphide formation which may account for this[145,146]. The mechanism by which these effects operates is unknown, and is both less common and less well studied than thiol-disulfide exchange but could include the chelation of metal ions[147], or interactions between DTT and the protein due to DTT's capacity to form two strong and two weak hydrogen bonds which may cause conformational change or steric hindrance[145].

Interestingly two single molecule studies, discussed in the introduction, performed a ClpXP degradation to investigate the stall force and slippage experienced by ClpX as it unfolded and translocated a substrate under distinct redox conditions[4,30]. This study did not investigate the effects of the reducing environment specifically. However, the study that did not use a reducing environment found that ClpX had a lower stall force and greater slippage than the other study which had 1 mM DTT in the reaction mixture.

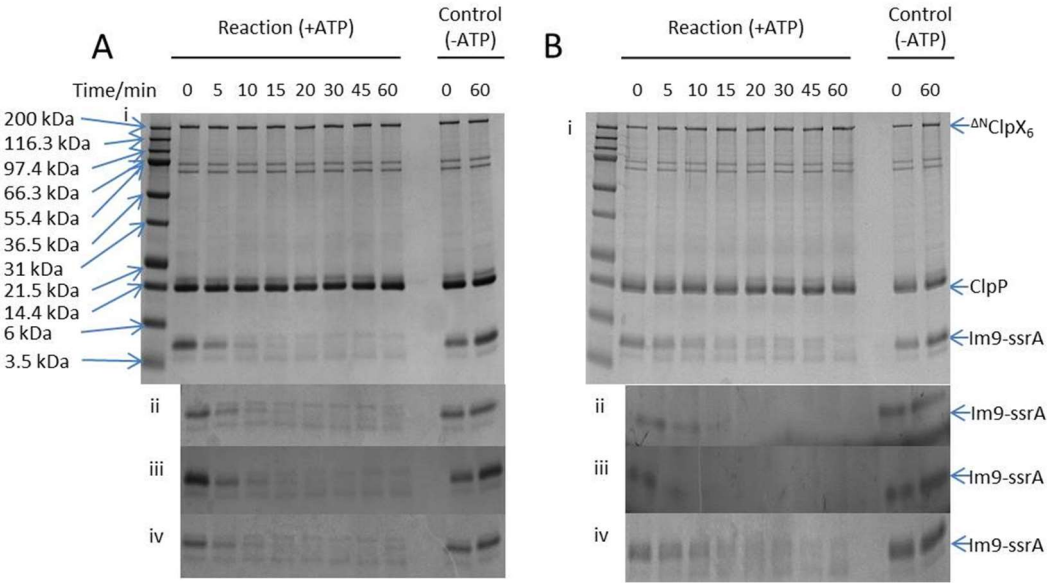


Figure 3.15: SDS-Page showing degradation of Im9-ssrA by $^{\Delta N}ClpX_6P$ in the presence (A) and absence (B) of 2mM DTT. Lanes marked +ATP contained 5 mM ATP (i) shows the complete gel. (ii-iv) show just the Im9-ssrA band from replicate experiments. Some lanes, such as those in B iii and B iv had bands which had completely degraded. If no visible band remained the amount of remaining protein was assumed to be 0.

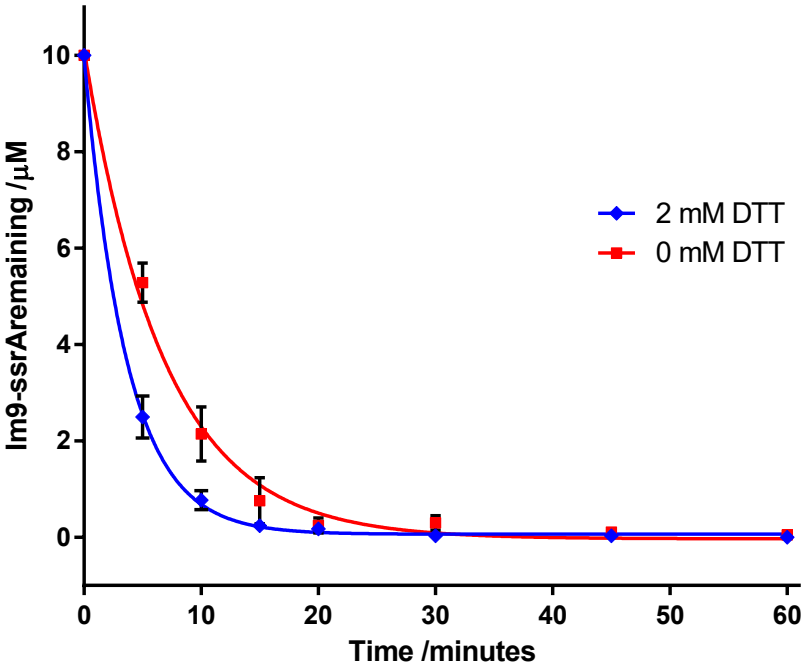


Figure 3.16: Graph of the mean results of $^{\Delta N}ClpX_6P$ degrading Im9-ssrA. Blue diamonds show the reaction containing 2 mM DTT; red squares represent reactions without DTT. Blue and red lines show the result of a mono-exponential fit to the data. Error bars show the standard error. n=3.

Table 3.2: Results of fitting a mono-exponential to the Im9-ssrA degradation data shown in Figure 3.16.

	k /s^{-1}	**R^2 for fit**
2 mM DTT	0.28±0.33	0.99
0 mM DTT	0.15±0.02	0.95

3.6 Substrate switching

Another interesting use of the ClpXP system would be to investigate the importance of directionality of an applied force to protein mechanical strength at biologically accessible loading rates. Since the ssrA tag is fused to the substrate protein at the C-terminus, ClpX can only apply force to the substrate in one geometry. There are however, N-terminal tags which will allow ClpX to apply its unfolding force at the N-terminal of a protein[64]. Unfortunately the $^{\Delta N}ClpX_6$ variant used here is unsuitable for this purpose as the N-terminal tags require the N-terminal zinc binding domain of the ClpX in order to be degraded and this is lacking in the pseudohexamer (see section 1.4.1)[148,149]. Wild type ClpX is able to degrade N-terminally tagged substrates, such as those with a λO degradation tag. The rate of degradation of N-terminally tagged substrates is, however, much slower relative to ssrA-tagged substrates[64]. It was decided to make a mutant ClpX where the positively charged residues of the RKH loop, which normally interact with the negatively charged residues and the C-terminus of the ssrA tag, were replaced with alanine, which has shown to increase the rate of degradation of a λO-tagged substrate[67].

3.6.1 ClpX$^{\Delta RKH}$

The wild type ClpX with an N-terminal hexahistidine affinity tag (available in the laboratory) was altered by site-directed mutagenesis (Quikchange, Stratagene) to the R228A, K229A, and H230A mutant which was named His_6-$ClpX^{\Delta RKH}$. An expression trial was carried out showing successful expression in BL21 (DE3) pLysS *E. coli* and a large scale expression was performed. The

His_6-$ClpX^{\Delta RKH}$ was purified as previously described and the results of the purification are shown in Figure 3.17.

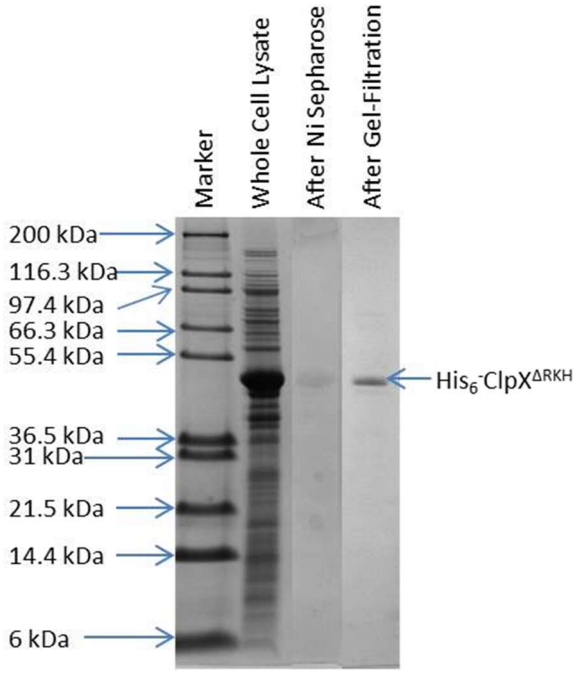


Figure 3.17: Composite summarising the purification of His_6-$ClpX^{\Delta RKH}$. Lanes show the whole cell lysate, the fraction selected from the Ni Sepharose purification and the fraction selected for freezing in aliquots after gel filtration

3.6.2 λO-Arc

It was necessary to produce a λO-tagged substrate for degradation by His_6-$ClpX^{\Delta RKH}$; the substrate chosen was the *E. coli* Arc repressor protein (Arc) (Section 1.7.3, Figure 1.24) as this has previously been shown to be degraded successfully by ClpX mutants. A fusion protein was devised in which the first 12 residues of the λO sequence was added to the N-terminal of Arc and six histidine residues were added to the C-terminal to facilitate purification to produce λO-Arc (Appendix – Section 7.1.3.2). The sequence for this protein was reverse translated using a database of the most commonly used *E. coli* codons[150]. A 5' NdeI and a 3' XhoI restriction site was added to this DNA sequence and a synthetic gene purchased (Eurofins MWG Operon, Germany). The DNA was supplied in a pCR2.1 plasmid and the required sequence was excised by a restriction digest and ligated into a pET23a plasmid for expression. After expression and purification a test degradation was performed to ensure that His_6-$ClpX^{\Delta RKH}$ was capable of degrading λO-Arc. As can be seen from Figure 3.18 His_6-$ClpX^{\Delta RKH}$ and λO-Arc were compatible and the degradation was robust.

Figure 3.19 and Figure 3.20 show the results obtained when ClpX and His_6-$ClpX^{\Delta RKH}$ were each used to degrade both Im9-ssrA and λO-Arc. These results clearly demonstrate that ClpX had a greater activity against ssrA-tagged substrate and a reduced activity against the λO-tagged substrate. By contrast, the situation was reversed for His_6-$ClpX^{\Delta RKH}$, which had good activity against the λO-Arc substrate and no detectable activity against Im9-ssrA. These initial results were encouraging, demonstrating that it was possible to change the specificity of the ClpX by making changes in the RKH loop.

A summary of the results of these experiments are shown in Table 3.3. Generally these data demonstrate that the substrate specificity of ClpX has been switched. The data for His_6-$ClpX^{\Delta RKH}$ vs Im9-ssrA reveals the expected error of the method as it shows an apparent 13% increase in remaining substrate over time. It demonstrates that the ClpX RKH mutation decreases the activity of ClpX against the ssrA-tagged Im9 substrate compared to WT and increases activity against the λO-tagged Arc substrate.

However, several problems were identified during these experiments. Firstly it appeared that ClpX was degrading over the experimental time-course, which is most apparent in the gel of His_6-$ClpX^{\Delta RKH}$ versus λO-Arc (Figure 3.19D), but is also detectable in the other gels shown in Figure 3.19. These figures show that His_6-$ClpX^{\Delta RKH}$ appeared to be degrading, yielding a cleavage product. This may be due to the presence of an extended length of polypeptide at the N-terminus of His_6-$ClpX^{\Delta RKH}$ (His_6 tag). This unstructured sequence can then be engaged by $ClpXP^{RKH}$ leading to autodegradation[151]. This effect could be ameliorated by addition of an excess of His_6-$ClpX^{\Delta RKH}$. In terms of a quantitative assay, however, this is not ideal situation as the enzyme concentration changes in addition to the substrate.

It is interesting to speculate on what might be happening at a molecular level. Previous work in the Brockwell group has shown that the conversion of ClpX into the cleaved product reduces the degradation activity of the ClpXP[152] it is not known, however, whether the ClpX is degraded while part of the complex or as a free monomer in solution. As ClpX is not ssrA tagged the former is most likely. It is also not known whether the ClpXP complex will disassociate after degrading one of its member ClpX subunits or whether it becomes

inactive or even if it remains active but becomes less efficient as it is known that a ClpX ring containing one or more inactivated subunits will retain some activity[74]. It is not known whether the ClpXP complex can dynamically assemble/disassemble under the assay conditions though as proteins are not static entities it is likely that the complex exists in equilibrium with monomeric and partially assembled complexes. I believe the most likely outcome is that a ClpX ring which contains a partly degraded ClpX monomer will disassociate into its individual parts, which will become free to reassemble into a new ClpX ring which most likely will not include any of the partly degraded ClpX as the changes it undergoes will make it unlikely to maintain the required features for ring formation. However no compelling evidence exists to support this speculation.

It would also have been possible to try the experiments omitting either the ClpX or ClpP component to check for possible contaminant protease activity that was contributing to the ClpX degradation, but this was not done.

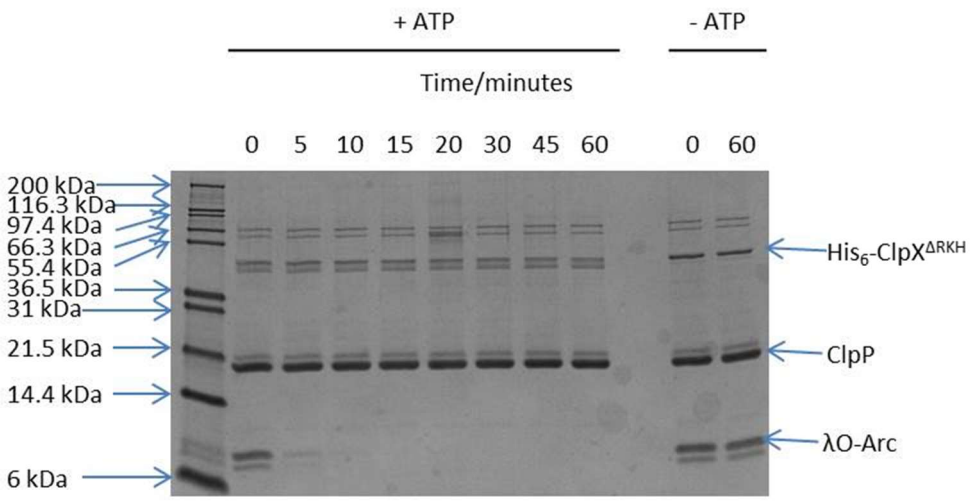


Figure 3.18: SDS-PAGE gel showing the degradation of λO-Arc by His_6-$ClpX^{\Delta RKH}$ P. ATP concentration was 5 mM. For lanes where the amount of remaining substrate was too small to be detected the amount was assumed to be 0.

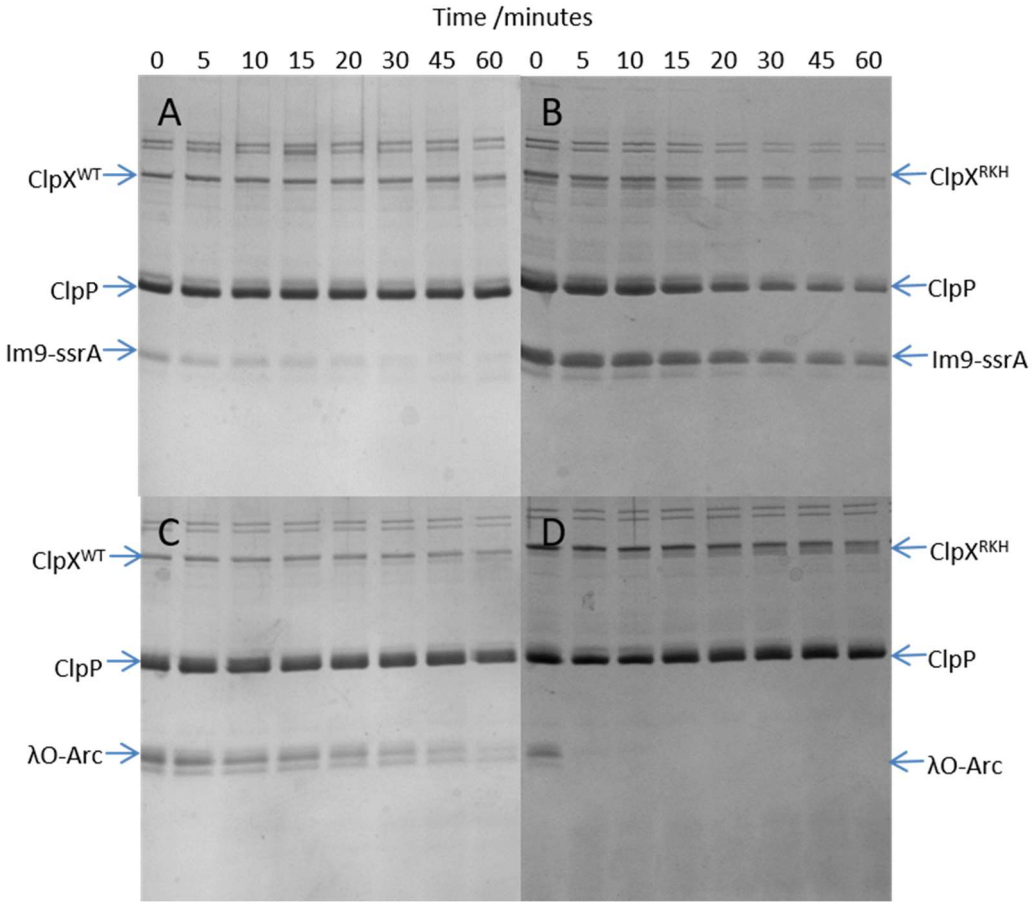


Figure 3.19: Gel composite showing the degradation of λO-Arc or Im9-ssrA by either ClpXP or His_6-$ClpX^{\Delta RKH}$ P. A: ClpX vs. Im9-ssrA. B: His_6-$ClpX^{\Delta RKH}$ vs. Im9-ssrA. C: ClpX vs. λO-Arc. D: His_6-$ClpX^{\Delta RKH}$ vs. λO-Arc. All experiments used an ATP concentration of 5 mM.

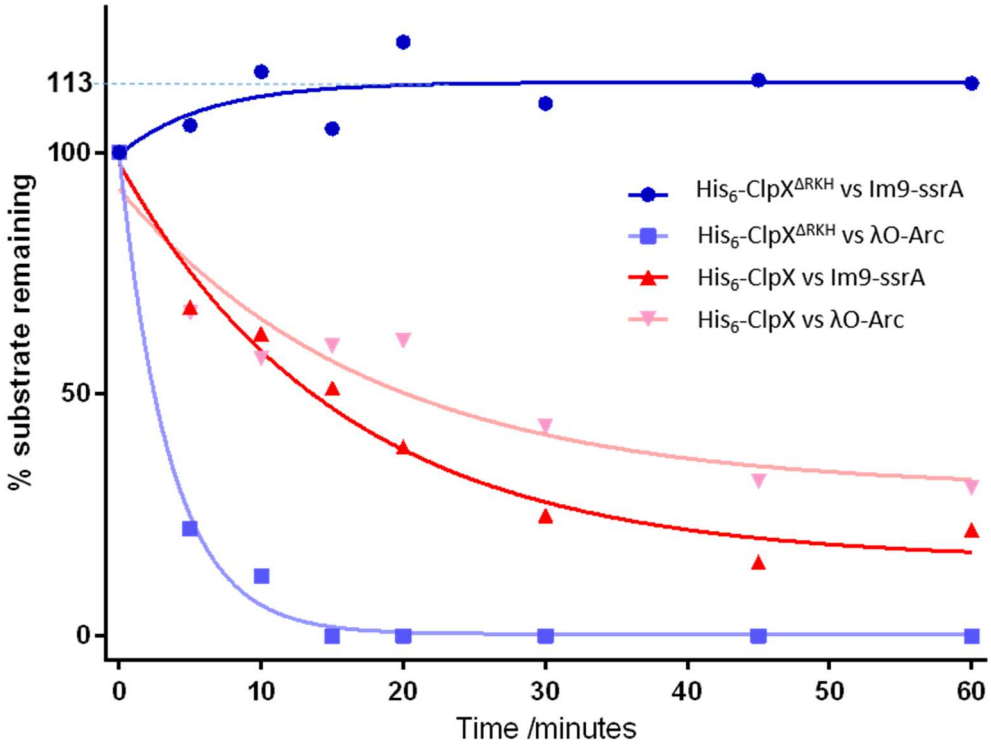


Figure 3.20: Graph of the densitometry results obtained from the results shown in Figure 3.19. % Substrate remaining is the amount of the initial 10 μM substrate remaining after each time interval. Each experiment was performed once.

Table 3.3: showing the results of fitting mono-exponentials to the data shown in Figure 3.20.

	k /s^{-1}	R^2 for fit
His_6-$ClpX^{\Delta RKH}$ vs Im9-ssrA	0.16±0.18	0.49
His_6-$ClpX^{\Delta RKH}$ vs λO-Arc	0.28±0.027	0.99
His_6-ClpX vs Im9-ssrA	0.063±0.011	0.97
His_6-ClpX vs λO-Arc	0.056±0.024	0.89

3.6.3 λO-E9

In an attempt to devise a better test substrate for His_6-$ClpX^{\Delta RKH}$ it was decided to utilise another protein which has been used within this group: the DNAse domain of the *E. coli* defence protein colicin E9. Since E9 is toxic to *E. coli* it is co-expressed alongside the immunity protein Im9. The interaction between E9:Im9 is one of the tightest known ($K_d = 10^{-16}$ M). In the laboratory, E9 is co-expressed with Im9, which inhibits E9 activity by binding to the DNAse domain thus preventing lethality[19].

PCR followed by restriction enzyme digestion and ligation into pET23 was used to add the λO tag to the N-terminus of E9 as described in Section 2.2.13. After confirmation of the DNA sequence (Appendix – Section 7.1.3) pET23λOE9 was transformed into BL21 [DE3] *E. coli* cells. However, no transformants were obtained. A subsequent attempt at transformation, alongside a positive control (pUC19) was attempted using a freshly prepared batch of competent cells along with a positive control. The bacteria transformed with the positive control grew but the ones transformed with pET23λOE9 did not. It was concluded that adding the λO tag to the E9 could be interfering with the E9:Im9 binding and causing toxicity to the *E. coli* cells. An E9 variant which had an extended N-terminal sequence (residues 1-16 of the translocation (T) domain of E9 and a TEV cleavage site), known to express well was obtained from the group of Colin Kleanthous at the University of York. The λO tag was added to the variant (Appendix – Section 7.1.3) as before and the resulting sequence confirmed. This new variant was known as λO-(T1-16)-TEV-E9. pET23λO-(T1-16)-TEV-E9 was transformed into BL21 [DE3] *E. coli* and again failed to produce any colonies. However when it was transformed

into BL21 [DE3] pLysS, colonies were obtained. This may have been due to the pLysS which expresses T7 lysozyme, which further suppresses basal expression of T7 RNA polymerase prior to induction, stabilizing recombinants encoding target proteins that affect cell growth and viability[153] reducing any toxic effects of the protein.

After a small scale trial confirmed protein expression, a large scale (10 l) culture was carried out. The protein was purified via the hexahistidine tag on the Im9 (section 2.2.13). Briefly, cleared lysate was applied to a Ni sepharose column to immobilise the Im9:E9 complex. E9 was eluted by denaturing the protein complex with an increasing concentration of the denaturant guanidinium hydrochloride. The E9 was then further purified by size-exclusion chromatography. Results of protein purification are shown in Figure 3.21.

The purified λO-(T1-16)-TEV-E9 was degraded by the His_6-$ClpX^{\Delta RKH}$ (Figure 3.22) but the autodegradation was found to be exacerbated with this more robust substrate.

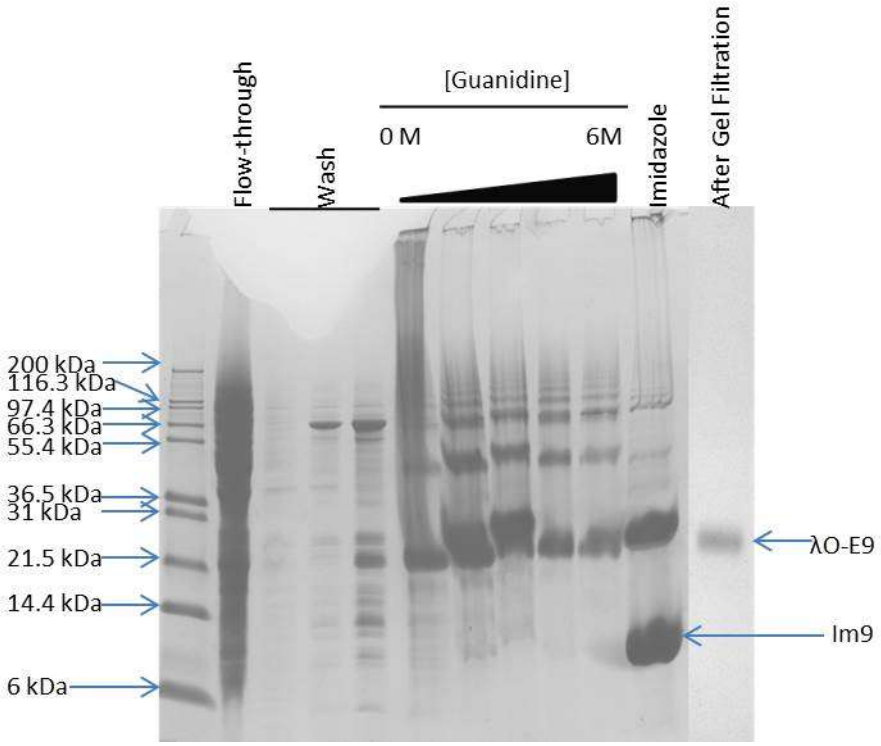


Figure 3.21: Composite gel showing the purification of λO-(T1-16)-TEV-E9. Flow through: cleared lysate which did not bind to the column. Wash: washing column with binding buffer. [Guanidine] shows the elution of the E9 in an increasing concentration of guanidine hydrochloride. Imidazole: Elution of Im9 with wash buffer. After gel filtration: purified E9 after size-exclusion chromatography.

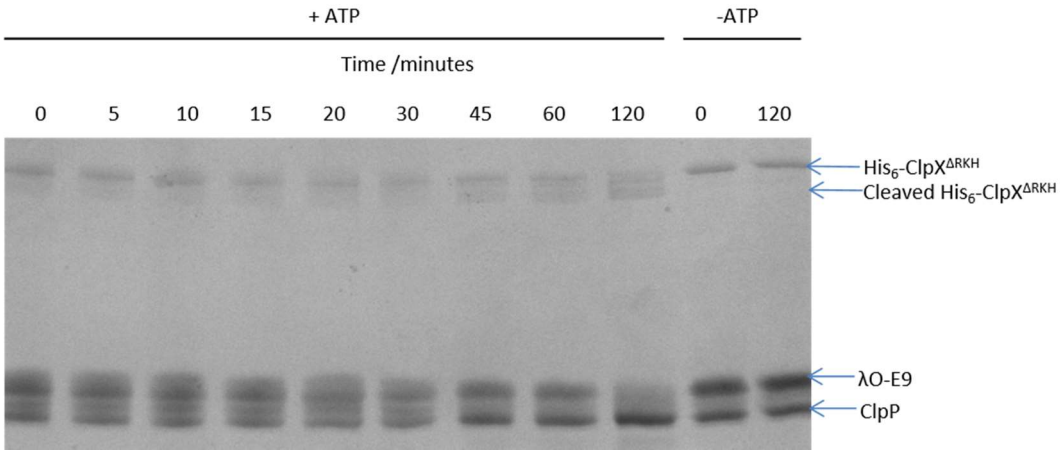


Figure 3.22: SDS-PAGE Gel showing the degradation of λO-(T1-16)-TEV-E9 by His_6-$ClpX^{\Delta RKH}$ P. Lanes labelled +ATP contained 5 mM ATP, lanes labelled –ATP were the negative control containing no ATP.

3.6.4 Thrombin cleaved His_6-$ClpX^{\Delta RKH}$

It was decided to use a slightly different variant of $ClpX^{\Delta RKH}$ with a thrombin cleavage site between the hexahistidine tag and the ClpX to try and resolve this problem. A his-tagged monomeric wild-type ClpX with the thrombin cleavage site, available in the laboratory, was mutated to R228A, K229A, H230A using the Quikchange site directed mutagenesis kit (Section 2.2.4). This was expressed and purified as described for His_6-ClpX (Section 2.6.4 and 3.3.1) A thrombin CleanCleave kit (Sigma) was used to remove the His tag as described in Section 2.6.5. Unfortunately as shown in Figure 3.23A, the protein was no longer in solution after overnight incubation. Another attempt was made with a shorter incubation followed immediately by purification of the protein by gel filtration (Superdex 75 10/300 GL). Only a very small peak was observed eluting from the column and when the protein in this peak was visualised using SDS-PAGE an insignificant amount of protein was present (Figure 3.23B).

While it may have been possible to improve the situation by generating further variants of ClpX (each of the three mutations used in the $ClpX^{\Delta RKH}$ variant have been shown to improve the degradation of λO-tagged proteins when used separately[67]) it was decided not to pursue further optimisation of the ClpXP system. Instead it would be more useful to use the experimentally tractable $^{\Delta N}ClpX_6$ variant to explore the mechanical properties of polyglutamine containing proteins.

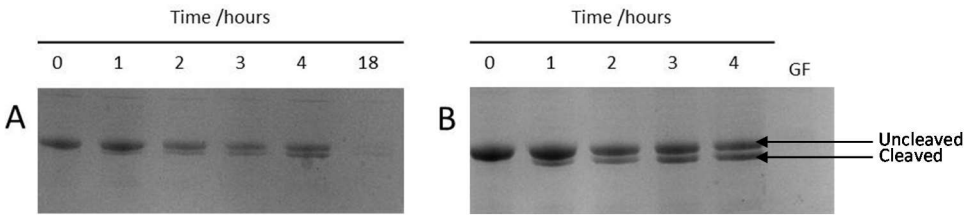


Figure 3.23: Thrombin cleaved His_6-$ClpX^{\Delta RKH}$ to $ClpX^{\Delta RKH}$ after incubation with thrombin. The protein was incubated with the beads at room temperature for 4 hours then: **A:** incubated for a further 14 hours at 4 °C, or **B:** immediately purified by size-exclusion chromatography at 4 °C. Markers on the right show the expected positions for the protein before (Uncleaved) and after (Cleaved) removal of the His tag by thrombin cleavage.

4 Investigation of polyglutamine containing proteins

4.1 Introduction

As discussed in section 1.6 there is an interesting unresolved question related to the mechanical stability of proteins containing polyglutamine repeats. When the stability of soluble protein constructs containing polyQ is investigated by AFM it has been found to be either very mechanically stable[32], or to have access to a hyperstable conformation[33]. How this mechanical stability is related to the deleterious effects of polyQ aggregation *in vivo* remains unresolved. In order to examine whether polyQ constructs were resistant to the actions of biological unfoldases it was decided to compare the rate of degradation (unfolding), using the ClpXP assay detailed in Chapter 3, with mechanical unfolding data obtained using AFM on the same system.

4.2 Aims

The aims of this chapter are to design and produce a protein containing a polyglutamine repeat sequence and a control protein lacking that region. These will then be characterised using a variety of biochemical techniques, investigated using AFM to see if the earlier results can be replicated and finally subjected to degradation by ClpXP to determine what insights we can find using this more biologically relevant technique.

4.3 Protein substrate production

It was necessary to design and produce a protein that could be used in experiments to determine the properties of polyQ (Figure 4.1). It was decided to produce a trimeric I27 polyprotein with two cysteine residues at the N-terminus and an ssrA degradation tag at the C-terminus ($(I27)_3$-ssrA). This protein would be useful for both AFM experiments (which could be compared to earlier experiments using $I27^{154}$) and degradation assays. The N-terminal cysteines would allow it to be tethered to a gold surface and the multiple I27 domains would produce a distinctive fingerprint facilitating the analysis of AFM force-extension data. (Section 1.3). Once characterised, a cassette would be inserted at the DNA level (at the SpeI site, Figure 4.1) containing various polyQ repeat lengths. These constructs would then be characterised by a variety of techniques and compared to the I27 only control.

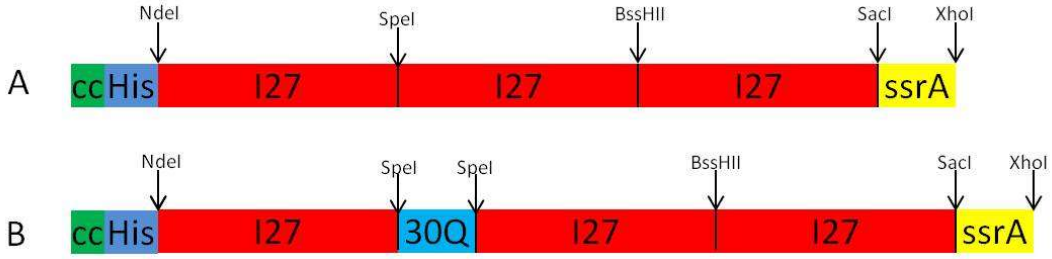


Figure 4.1: Design of the polyprotein to be used in these experiments. A: The control protein B: The protein containing a polyQ domain. Endonuclease restriction sites defining each cassette are shown (arrows). CC: double cysteine N-terminal motif. His: $(His)_6$ affinity tag. ssrA: ssrA degradation tag.

4.3.1 DNA manipulation

4.3.1.1 I27 trimer

As there was no suitable protein available it was necessary to produce one. The strategy to create these constructs is summarised in Figure 4.2. First PCR was used to generate I27 domain 1 – an I27 domain, with a 5' NdeI restriction endonuclease site and a 3' multiple cloning site (MCS) encoding consecutive SpeI, BssHII and SacI restriction sites. After ligation into the plasmid containing the ssrA tag, domains 2 and 3 would be ligated in sequentially (Figure 4.2 D, E).

The starting point for this construct was three plasmids: two plasmids, each containing one copy of I27, one flanked by SpeI and BssHII restriction sites (suitable for use as domain 2) and the other suitable for use as domains 1 and 3 after mutagenesis to insert the necessary restriction sites, and a pET23a plasmid containing an ssrA tag at the end of another protein (E2Lip3) which contained a 5' NdeI restriction site, a SacI restriction site between E2Lip3 and the ssrA tag and a 3' XhoI restriction site (Figure 4.2A). E2Lip3 could be excised by restriction enzyme digest to allow ligation of the I27 sequences into the construct. PCR mutagenesis (Section 2.2.3) was used to produce domain 1 (NdeI site 5' to I27 and a 3' MCS containing SpeI, BssHII and SacI) and domain 3 (SacI and BssHII sites 5' and 3' to I27 respectively). The E2Lip3 gene was excised from the plasmid containing the ssrA tag with NdeI and SacI, dephosphorylated and purified on an agarose gel as described in Sections 2.2.8 and 2.2.9. The same restriction digest was performed on I27 domain 1 and this was also gel purified. After gel extraction of the insert and vector, a ligation reaction was performed to produce a plasmid containing I27

domain 1 with a 3’ ssrA tag and containing the appropriate restriction sites to allow ligation of domains 2 and 3 by a similar method. Restriction digests and DNA sequencing were carried out at each stage to confirm the insert was present and had the correct sequence. Figure 4.3 shows the results of a final restriction digest confirming the presence of 3 I27 genes. DNA sequencing confirmed that the correct sequence was present.

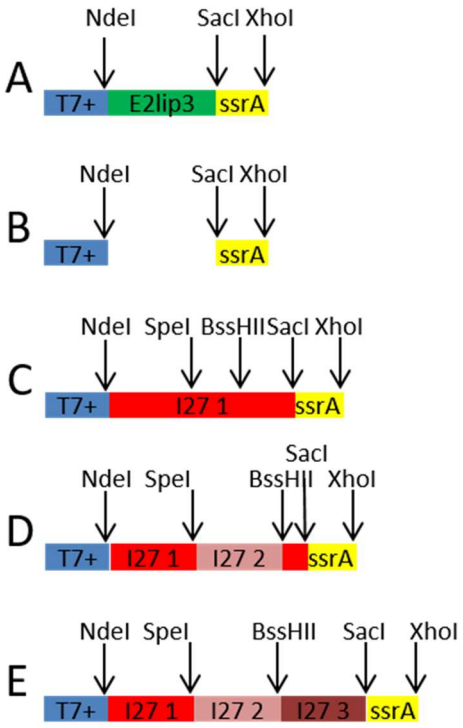


Figure 4.2: Summary of the required ligations to produce the $(I27)_3$-ssrA control construct (domains not shown to scale). The construct is contained in a pET23 derived plasmid which is omitted for simplicity. T7+ is the T7 promoter sequence used to overexpress the protein. E2lip3 and I27 are the protein domains of the same name and ssrA is the ssrA degradation tag. **A**: The parent construct containing E2lip3 with a C-terminal ssrA tag. **B**: The same construct after digestion by NdeI and SacI. **C**: The construct after ligation of I27 domain 1 containing the additional required restriction sites. **D**: The construct after the ligation of the second I27 domain. E. The construct after the ligation of the third I27 domain.

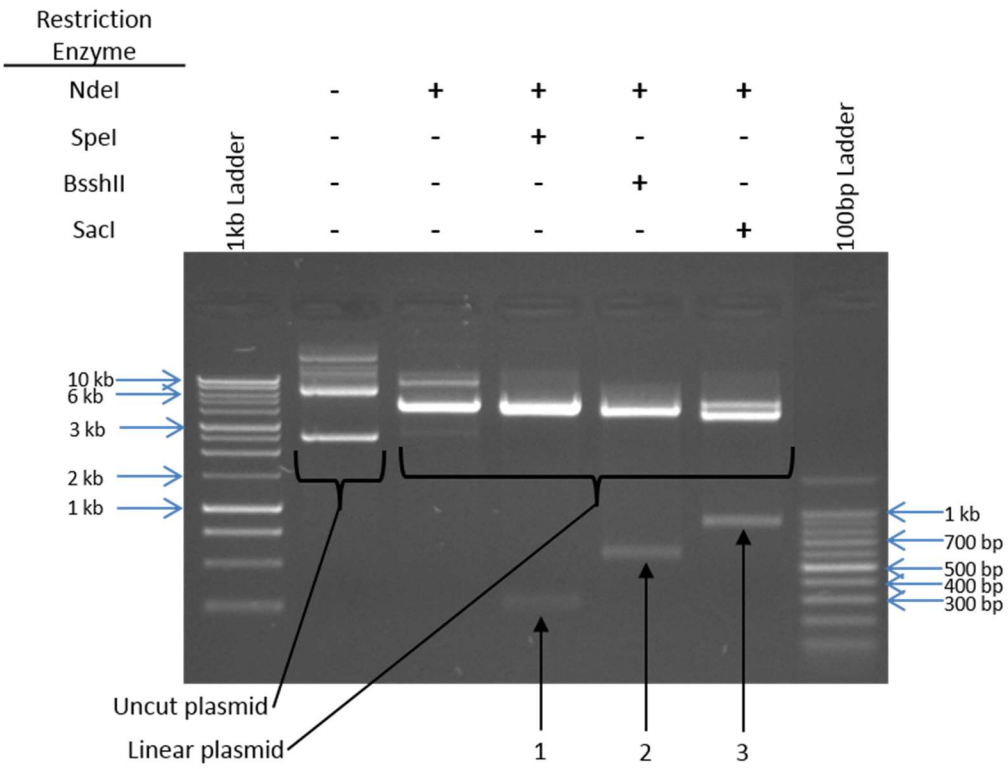


Figure 4.3: Agarose gel showing the results of a restriction digest confirming the presence of the required inserts. The bands labelled Uncut plasmid show a mixture of topological isomers such as supercoiled, linear and nicked DNA. The bands labelled Linear plasmid show the result of the circular plasmid after cutting by a restriction enzyme. 1, 2 and 3 are bands corresponding to the expected size of a DNA fragment with one, two or three I27 domains respectively.

4.3.1.2 Polyglutamine region

Glutamine is encoded by only two codons: CAA and CAG. The resultant highly repetitive sequences pose problems for the construction of proteins containing pre-determined polyglutamine repeat lengths. The method chosen to produce the different length polyQ sequences was based on that of Alexandrov *et al.* 2008.[155] This method was chosen because Alexanndrov *et al.* had successfully used it to produce polyQ sequences of varying lengths in the range we were interested in and it was straightforward to perform using commercially supplied DNA constructs. Three pairs of complementary DNA oligomers were designed (see Appendix – Section 7.2.1) to provide duplex DNA terminators and elongators. The terminators contained the necessary restriction site and had one overhanging end, the elongators had two overhanging ends compatible with each other and the terminators. After annealing of the oligomers to produce a duplex DNA fragment, the oligomers were mixed in a ratio of 1:10 terminators:elongators and ligated by T4 DNA ligase (Figure 4.4).

The initial strategy was to design a 5' terminator with both SpeI and BssHII restriction sites and two sets of 3' terminators; one with a SpeI restriction site and one with a BssHII restriction site (Figure 4.5 A, B). These could be used to produce two polyQ sequences that could be inserted in between I27 domains in the control protein.

After ligation, a ladder of different length DNA products was obtained resolved using a 3% *(w/v)* agarose gel as in Figure 4.6. The ligated mixture was purified by use of a gel extraction kit (Qiagen) to remove proteins and small DNA fragments. After "A-tailing", the ligation product was inserted into the

pre-cleaved T-tailed plasmid pGEM-T. The products of the second ligation step were transformed into *E. coli* SURE2. Blue/white screening (Section 2.2.7) was employed to select candidate colonies for growth and plasmid extraction. Restriction digests to determine the presence and estimated lengths of the inserted polyQ sequence in the plasmid were of mixed success. Sequencing showed that polyQ sequences of various lengths had been successfully created but that the restriction sites were no longer present. The terminators were redesigned minimising the amount of secondary structure that was likely to form using an online prediction tool[156]. The redesigned terminators (Appendix – Section 7.2.1) allowed the production of several successful polyQ sequences (15Q, 17Q, 30Q) flanked with SpeI restriction sites. The production of a construct which had a BssHII restriction site at both ends of the sequence was attempted (Figure 4.5 D), but this was not successful.

It was decided to proceed with the 30Q sequence rather than the 15Q or 17Q sequences as it was long enough that a change in length could be easily detected using AFM and it was felt that a longer sequence would be more likely to exhibit unusual properties. This was ligated into the $(I27)_3$-ssrA construct to produce I27-Q30-$(I27)_2$-ssrA. Other ratios of elongators and terminators were assessed for their ability to produce polyQ sequences of differing lengths but none produced a result substantially different from that shown in Figure 4.6.

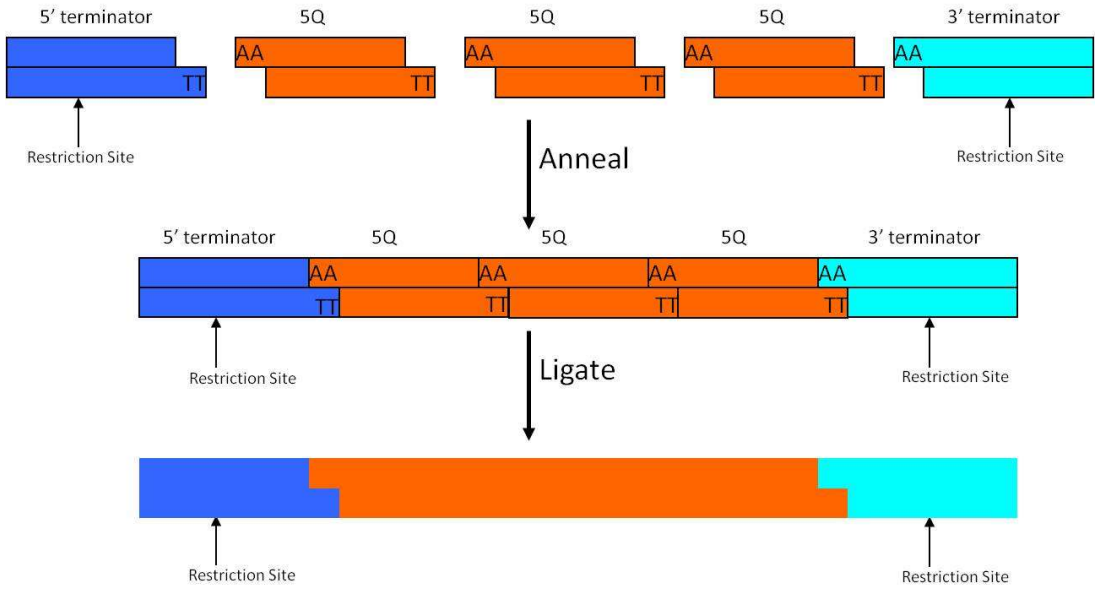


Figure 4.4: A schematic diagram of the process used to produce a polyQ sequence suitable for ligation into the control construct.

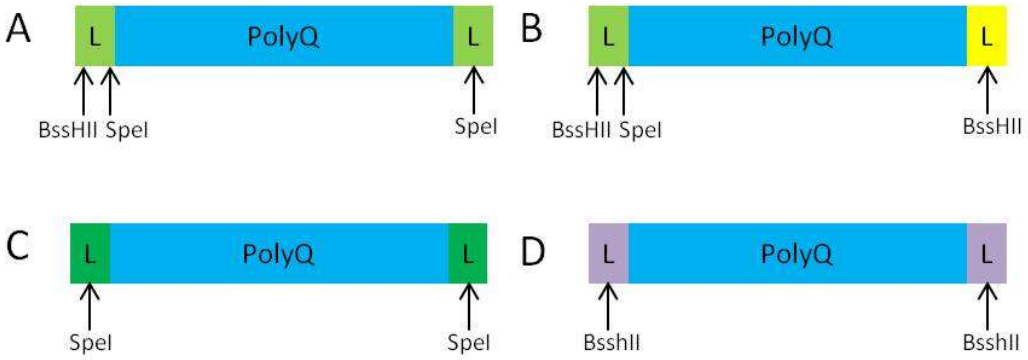


Figure 4.5: Summary of the design of polyQ regions for insertion into I27 trimer. **A:** initial design with 5' BssHII and SpeI restriction sites and 3' SpeI restriction site. **B:** initial design with 5' BssHII and SpeI restriction sites and 3' BssHII restriction site. **C:** final design with 5' SpeI restriction site and 3' SpeI restriction site. **D:** final design with 5' BsshII restriction site and 3' BsshII restriction site

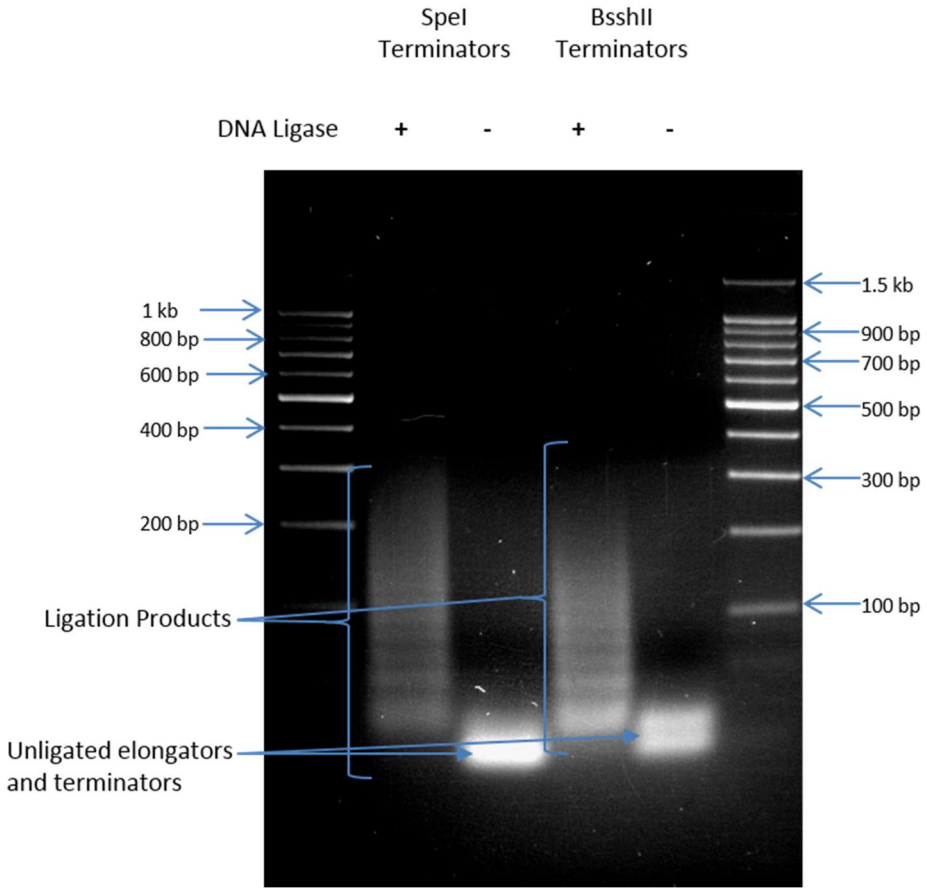


Figure 4.6: Agarose gel showing the results of the ligation of a 1:10 mixture of terminators and elongators. Lanes containing DNA ligase show a ladder of different length ligation products. Lanes without DNA ligase show only the unligated elongators and terminators.

4.3.2 Protein production

An expression trial of the control protein ($(I27)_3$-ssrA) was carried out as shown in Figure 4.7. As this protein expressed well and was largely in the soluble fraction, a large scale overexpression was performed. The protein was purified by a three step protocol (nickel-affinity, followed by size-exclusion chromatography and ion exchange (Section 2.6.8)) as shown in Figure 4.8.

The resultant protein (molecular mass = 32,990 Da (expected mass from the sequence)) was almost pure but contained a contaminant of a slightly lower molecular mass (molecular mass ≈ 31,000 Da (from electrophoretic mobility)). A similar contaminant has been observed in all of the ssrA-tagged proteins produced for this study which is believed to be a result of the protein undergoing proteolytic cleavage to remove the ssrA tag. (Mass difference ≈ 2000), mass of residues 279-301 (KVKELLCGAQ AANDENYALAA= 2192)) Various strategies have been attempted to remove or reduce this contaminant but it was always present regardless of the presence of protease inhibitors or undertaking protein purification at low temperatures. While not ideal, the presence of this proteolytic product was tolerated for the following reasons: firstly, the presence of a C-terminally truncated protein would not affect AFM measurements as the cysteines used for immobilisation were N-terminal. Secondly, removal of the ssrA tag renders the protein “invisible” to ClpXP. While its presence would make the real substrate concentration difficult to quantify, these errors would be minimised by estimating the quantities of each species by densitometry and working at higher substrate concentrations.

The protein containing polyQ was produced using a protocol identical to that used for the control protein. Results of the expression trial (Figure 4.9) and purification (Figure 4.10) are shown below.

Once the proteins required had been produced, it was necessary to characterise them to ensure that they were folded and that they were monomeric in solution. This is because the work by Dougan *et al*[32] (which produced the initial result indicating that polyglutamine repeats had extreme mechanical stability) did not perform any controls to ensure the protein they used was monomeric. This may explain their surprising result, if the polyglutamine had begun to aggregate then the AFM experiments may not have been performed on monomeric polyglutamine. If the polyglutamine had begun to aggregate they may have been experimenting on a pre-fibrillar aggregate conformer of polyglutamine or even on polyglutamine amyloid-like fibrils. While the results they obtained (time-distance measurements showing distinct steps as expected for a monomeric I27 concatamer) mean that this is unlikely, it is important that this possibility be ruled out (section 1.6).

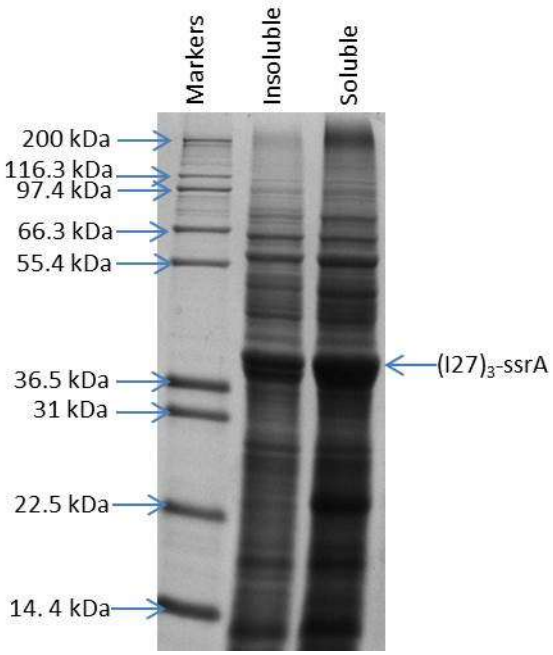


Figure 4.7: expression trial of $(I27)_3$-ssrA. Cells were lysed then centrifuged to separate the soluble and insoluble fractions. Insoluble: re-suspended pellet. Soluble: supernatant

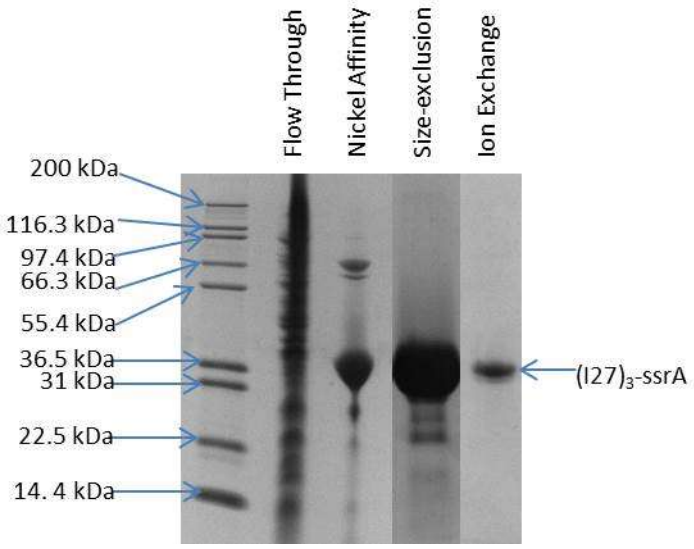


Figure 4.8: SDS-PAGE composite summarising the purification of $(I27)_3$-ssrA. Flow through: cell lysate which failed to bind to the nickel resin. Nickel Affinity: eluate from the nickel resin selected for further purification Size Exclusion: selected fraction after size-exclusion chromatography. Ion exchange: final fraction selected for use after ion exchange chromatography

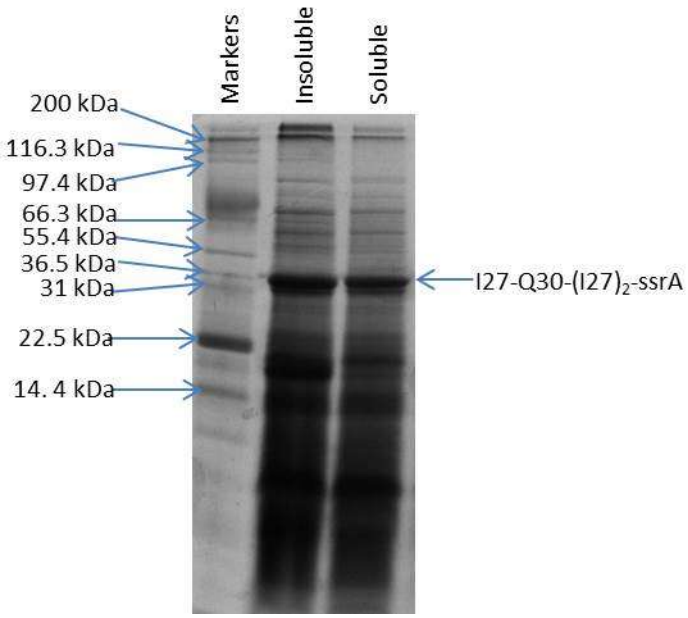


Figure 4.9: expression trial of I27-Q30-$(I27)_2$-ssrA. Insoluble: re-suspended pellet. Soluble: supernatant

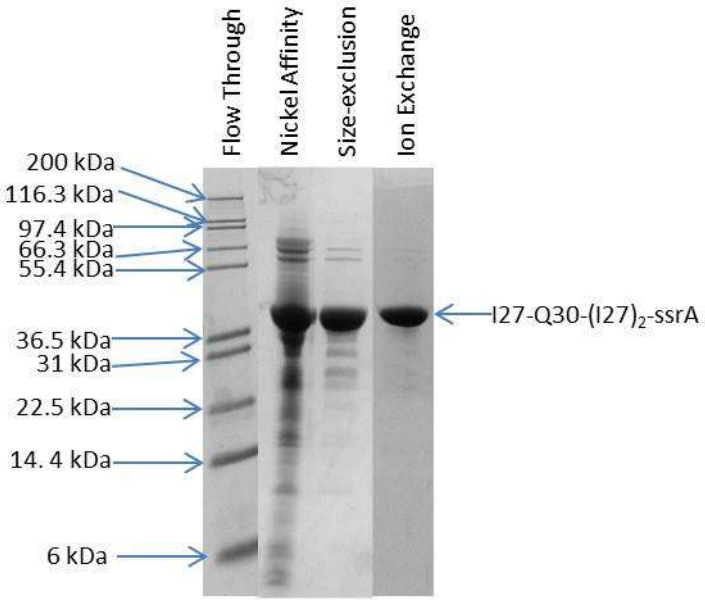


Figure 4.10: SDS-PAGE composite summarising the purification of I27-Q30-$(I27)_2$-ssrA. Nickel Affinity: eluate from nickel resin selected for further purification. Size Exclusion: selected fraction after size-exclusion chromatography. Ion Exchange: fraction selected for use after ion exchange chromatography

4.4 Biophysical characterisation

4.4.1 Circular dichroism spectroscopy

CD is a valuable technique for examining the structure of proteins in solution based on the spectral properties of chiral molecules.

When plane polarised light is separated into left and right circularly polarised components they are absorbed differently by chiral molecules. This results in the non-absorbed light being elliptically polarised[157]. The degree of polarisation can be measured at different wavelengths which gives insight into the three dimensional structure or environment of the molecule studied. While this does not give the same level of structural detail as nuclear magnetic resonance (NMR) or crystallographic techniques it is a relatively quick and easy technique that requires little sample. When applied to biological molecules such as proteins near-ultraviolet (UV) CD gives insight into the tertiary structure of a protein due to the absorption, dipole orientation and the environment surrounding phenylalanine, tyrosine, cysteine and tryptophan amino acids. In this study far-UV CD was used, which can reveal important information about the secondary structure of proteins; allowing an assessment of the degree of alpha-helical, beta-sheet and random coil structure present in a sample[157].

Far-UV CD was used here to assess whether the I27 component of the protein had its expected predominantly beta-sheet structure[42] and whether this was affected by the addition of the 30Q sequence which had been previously shown to have no regular secondary structure[116,126,158]. The experiment used 30 mM protein (≈0.2 mg/ml) in a buffer that was 10 mM Tris-SO_4, 1 mM EDTA,

42 mM K_2SO_4, 0.5 mM TCEP pH 7.6, known as characterisation buffer. This was chosen because it was as close as possible to the buffer used for the degradation experiments while still being compatible with the instruments used for characterisation. The results of the CD measurements are shown in Figure 4.11.

The result for $(I27)_3$-ssrA is similar to that found for $(I27)_5$ and matches the expected result for a mostly beta-sheet protein[157]. By contrast, the spectrum for I27-Q30-$(I27)_2$-ssrA was dominated by the unstructured polyQ region. Because the signal for the Q30 dominated the spectra it was not possible to categorically state that the I27 regions in I27-Q30-$(I27)_2$-ssrA were correctly folded and a further technique was required to verify this.

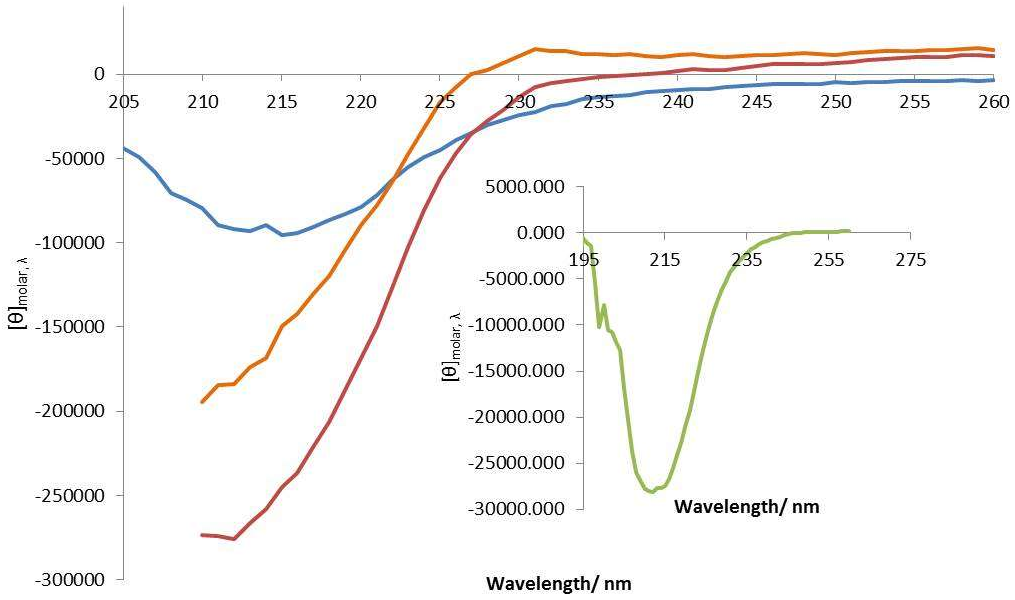


Figure 4.11: Far-UV CD of the control protein $(I27)_3$-ssrA (blue), the polyQ containing protein I27-Q30-$(I27)_2$-ssrA (red) and the spectrum for Q30 calculated by subtracting the blue spectrum from the red spectrum (orange). Inset: Far-UV spectrum for $(I27)_5$ (Toni Hoffman, personal communication).

4.4.2 Tryptophan fluorescence emission spectra

A particularly straightforward method for probing the internal environment of a protein is the use of a fluorophore whose emission wavelength is sensitive to solvent polarity. Fluorescence from the amino acid tryptophan has long been known to be sensitive to the polarity of its local environment and has been used for many studies. Among the properties measured are changes in the fluorescence intensity, wavelength maximum (λ_{max}), band shape, anisotropy, fluorescence lifetimes, and energy transfer. In this study an emission intensity wavelength scan was measured as Trp $\lambda_{max,\ em}$ is sensitive to its local environment, roughly correlating with the amount of solvent exposure of the fluorophore (i.e. whether buried in the hydrophobic core of a protein or not). Since each I27 domain in our protein contains a Trp residue while the polyQ region does not, it is possible to ascertain information about the folded state of the I27 domains by comparing the Trp emission spectra of the assumed folded state in characterisation buffer and the assumed unfolded state in the same buffer with the addition of a denaturant (8 M urea). The results of this experiment are shown in Figure 4.12.

Both the proteins with and without poly Q show a similar shift in λ_{max} and fluorescence intensity upon denaturation, this is a good indication that the Trp in the I27 in both proteins is in a similar state before and after denaturation; which in turn indicated that they both move from a similar folded to unfolded state upon the application of denaturant.

After ascertaining that the I27 domains within the protein appeared to be folded under conditions close to the experimental conditions it was then

necessary to determine whether the sample was monomeric under these conditions.

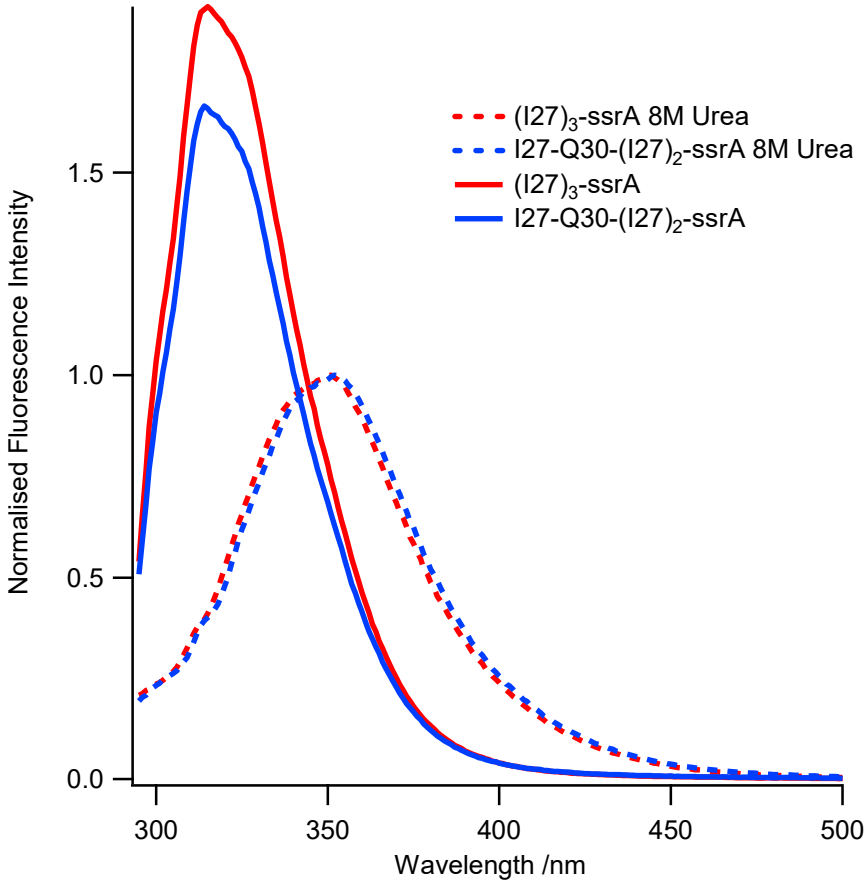


Figure 4.12: Fluorescence emission spectra of the control ($(I27)_3$-ssrA) and polyQ containing protein (I27-Q30-$(I27)_2$-ssrA) in denaturing and non-denaturing conditions. Corrected absorbance is obtained by determining the maximum fluorescence value for the protein in urea then dividing all of the fluorescence values for that protein by the maximum value. This ensures that the maximum normalised fluorescence intensity for the unfolded protein is 1 and adjusts the folded protein's fluorescence relative to the unfolded. The same is then done for the second protein.

4.4.3 Sedimentation velocity analytical ultracentrifugation

In sedimentation velocity AUC analysis of proteins, a high centrifugation speed is used to cause sedimentation of particles over a number of hours. An optical system measures absorbance within the sample chamber and allows the sedimentation rate for the species in the sample to be measured by following the solvent-solute boundary. The control and polyQ containing proteins were analysed by AUC and the sedimentation rates for the various components of the solution were measured. The samples were prepared by dissolving them in characterisation buffer then dialysing them in the same buffer for 24 hours and filtering them before they were sent to the University of Leeds' AUC service (Section 2.10). A summary of the results obtained are shown in Figure 4.13. For both proteins there were some very small components present (<1 S), the bulk of the protein was in a monomeric state (around 2.5 S) and at higher concentrations there was a very small amount of higher order species (>3 S) and some very large species (25 S). However the majority of the protein at the 10 μM concentration used in the following experiments was monomeric. The fact that I27-Q30-$(I27)_2$-ssrA was mostly monodisperse despite the presence of 30Q was not surprising as it is well known that the initial aggregation of polyglutamine repeat containing proteins is controlled by the flanking sequences. The absence of aggregated species also suggests that if mechanically resistant species are identified in AFM experiments that these cannot be due to extraction of the protein from stable fibres.

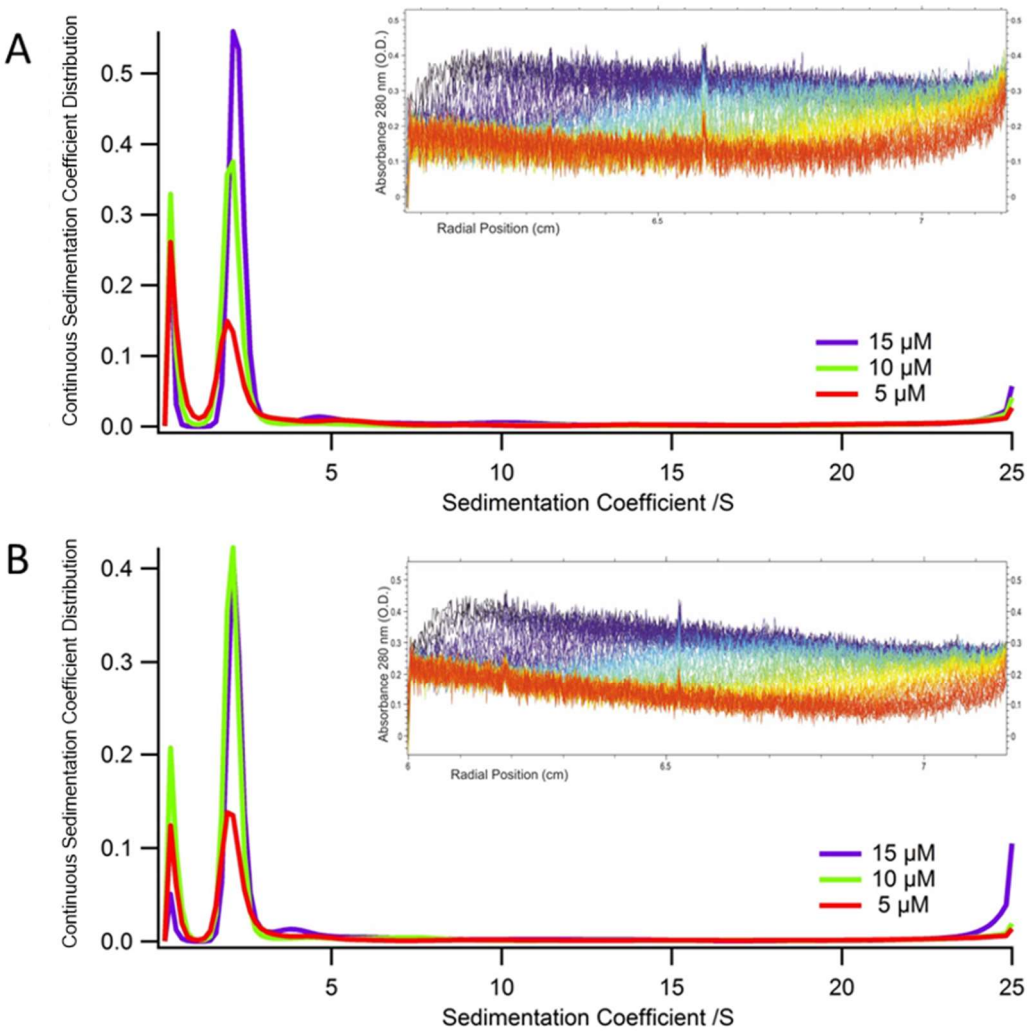


Figure 4.13: Sedimentation velocity AUC of: A - $(I27)_3$-ssrA B - I27-Q30-$(I27)_2$-ssrA. Insets show raw data for protein at 10 mM.

4.5 Dynamic force spectroscopy

Dynamic force spectroscopy was performed on the polyQ construct designed and purified in this study to investigate whether this construct would show the high mechanical strength observed in previous studies[32,33].

4.5.1 Mechanical unfolding of the test protein $(I27)_5$

To validate the experimental protocol, constant velocity mechanical unfolding was first performed on the well characterised $(I27)_5$ construct[32,40] at retraction velocities of 200, 632 and 2000 $nm.s^{-1}$.

Approximately 1000 approach-retract cycles were made at each velocity, aiming to accumulate approximately 100 unfolding events per experiment (experiments with more than 100 good unfolding events had 100 events analysed, chosen at random) and each experiment was performed three times. A typical force-extension profile is shown in Figure 4.14. After filtering of the data to remove blank or unsuitable force-extension profiles (section 2.7.4) force and unfolding distance histograms were generated.

By fitting of a Gaussian distribution to the histograms (Figure 4.15) it was possible to obtain an estimate (from the location of the maxima, χ_0) of the most probable force required for unfolding and the difference in length between the folded and unfolded domain. Force-frequency histograms for $(I27)_5$ unfolded at different retraction velocities are shown in Figure 4.15. They are presented in triplicate as each set of results is from a different experiment which will have a slightly different spring constant and will allow a better assessment of the errors in the experiments than plotting them on a single chart.

It can be seen that the force at which a protein unfolds is proportional to the log of the pulling speed (Figure 4.16) as expected and that the results agree with the previously published results.

The frequency histogram for change of length upon unfolding (ΔL_c) is shown in Figure 4.17. In order to obtain the data for this figure the contour length of each unfolding event was subtracted from the contour length of the subsequent unfolding event. For example it would be possible to obtain four I27 contour lengths from the data shown in Figure 4.14. The length for peak 1 given by the fitting would be subtracted from the length for peak 2, the length for peak 2 subtracted from peak 3, 3 from 4 and 4 from 5. This data can then be plotted as a histogram and a Gaussian fit to the data used to obtain the most likely contour length. The change in contour length was found to be 28.4 ± 0.2nm which agrees well with the published value of 28.1 nm for this construct.[138] The results obtained in this study agreed well with the previously published data especially considering that there is often a discrepancy of up to 20% between different group's results[25].

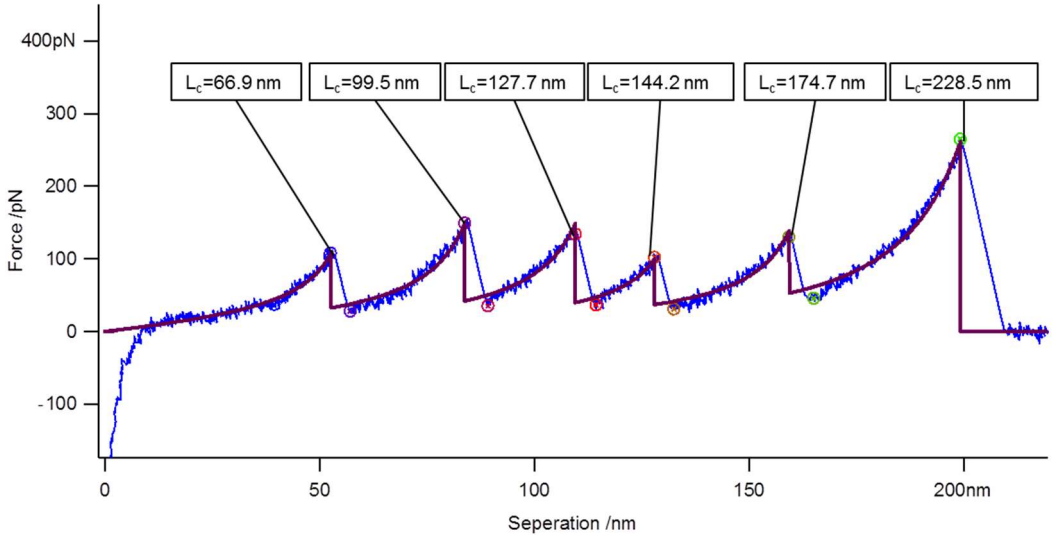


Figure 4.14: A typical force-extension profile for $(I27)_5$ obtained at 632 nm.s^{-1} Peaks 1-5 are I27 domain unfolding events, peak 6 is the pull-off event. Each peak is labelled with the contour length (L_c) calculated by the fitting software. Persistence length was constrained to 400 pm for all events. Note: WLC fitting is only to the rising portion of each peak.

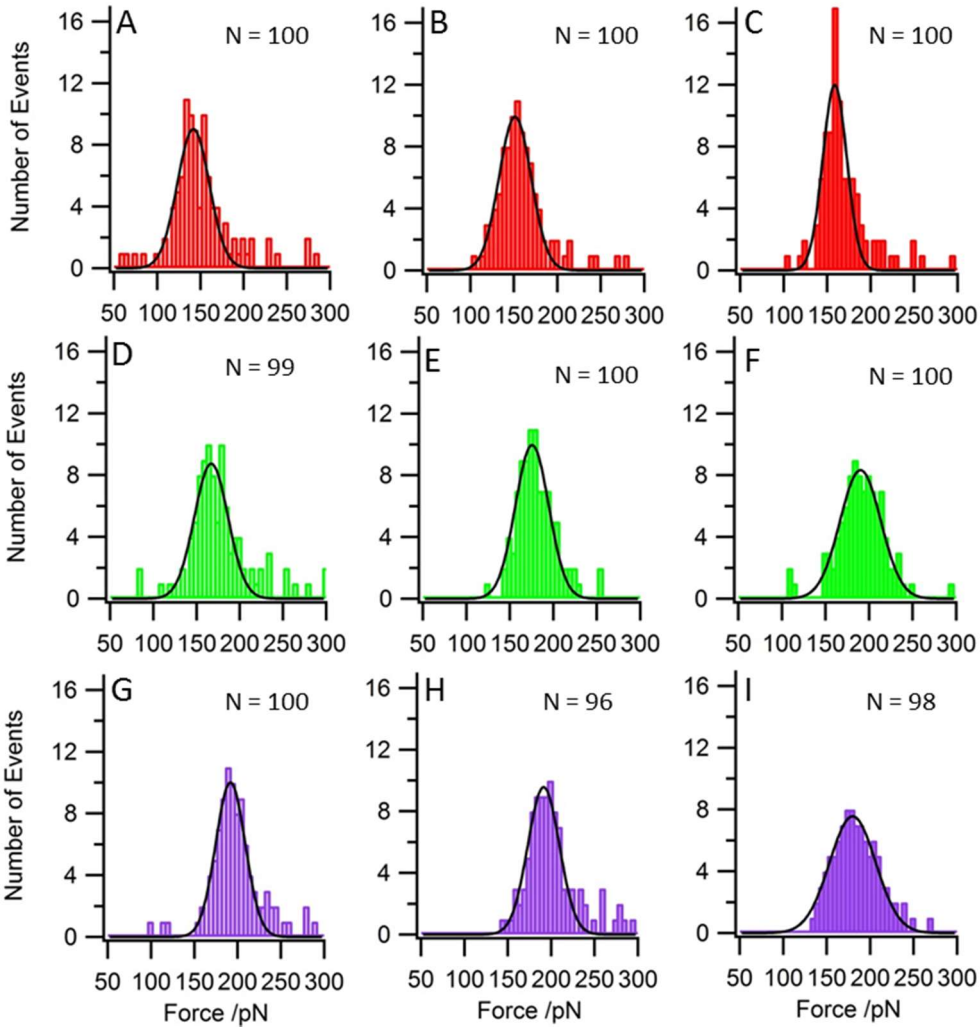


Figure 4.15: Force-frequency histograms generated from $(I27)_5$ constant velocity mechanical unfolding. Top row (A-C)– red 200 $nm.s^{-1}$ Middle row (D-F)– green 632 $nm.s^{-1}$ Bottom row (G-I)– purple 2000 $nm.s^{-1}$ Fits of a Gaussian distribution to each histogram are shown in black lines. Fitting parameters for the Gaussians are given in Table 4.1

Table 4.1: Fitting parameters for the Gaussian distributions shown in Figure 4.15.

Experiment	χ_0 **/pN**	χ^2	**Width** $(\sqrt{2} \times standard\ deviation)$	**N**
200 nm/s (A)	139.4 ± 1.4	75.3	26.3 ± 2.0	100
200 nm/s (B)	149.0 ± 0.7	20.6	26.1 ± 1.0	100
200 nm/s (C)	156.3 ± 1.0	86.0	19.5 ± 1.4	100
632 nm/s (D)	164.5 ± 1.5	69.5	27.6 ± 2.0	99
632 nm/s (E)	173.1 ± 0.8	28.4	27.1 ± 1.2	100
632 nm/s (F)	187.7 ± 1.0	22.7	33.0 ± 1.4	100
2000 nm/s (G)	189.3 ± 0.8	34.8	24.1 ± 1.2	100
2000 nm/s (H)	188.8 ± 0.9	37.6	25.7 ± 1.3	96
2000 nm/s (I)	177.4 ± 0.8	10.7	37.5 ± 1.1	98

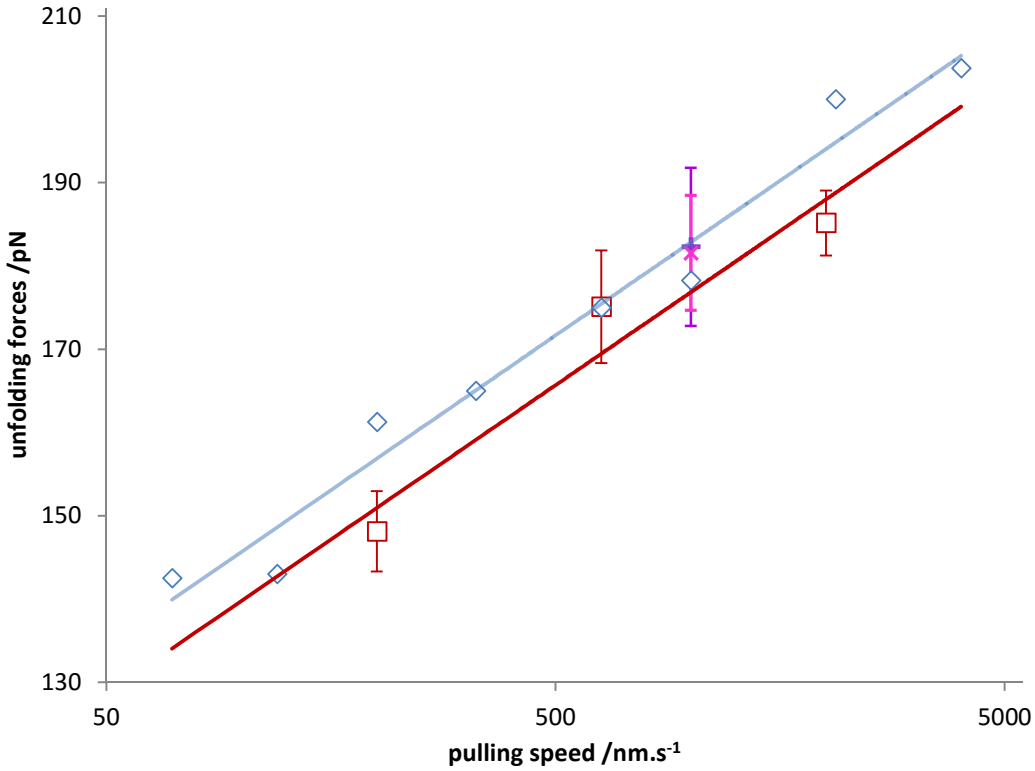


Figure 4.16: unfolding forces calculated from the mechanical unfolding of $(I27)_5$ unfolding forces calculated from the mechanical unfolding of $(I27)_5$. Red Squares: data obtained in this study. Blue Diamonds: data obtained in a previous study[154] Purple +: data for $(I27)_3$-ssrA obtained in section 4.5.2. Pink x: data for I27-Q30-$(I27)_2$-ssrA obtained in section 4.5.2.

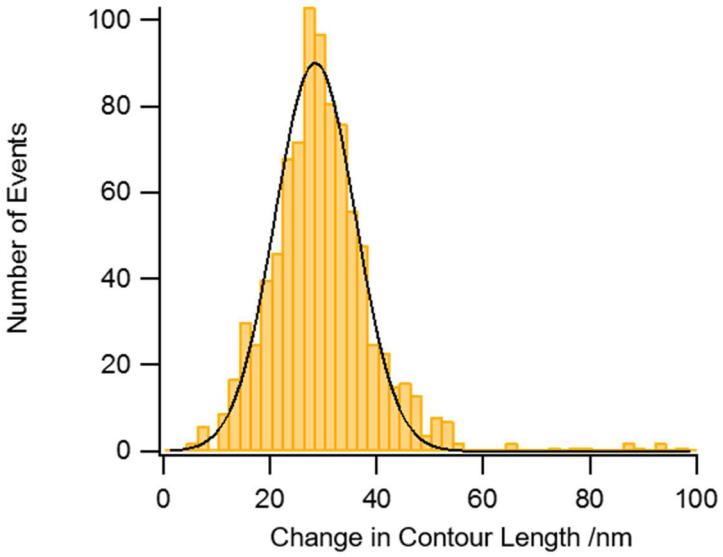


Figure 4.17: Length-frequency histogram generated from $(I27)_5$ constant velocity mechanical unfolding. Fits of a Gaussian distribution to the histogram are shown as a black line. χ^2=1041.2 $\sqrt{2}$×standard deviation = 10.7 ± 0.3 nm χ_0=28.4 nm N=893

4.5.2 Mechanical unfolding of $I(27)_3$-ssrA and I27-Q30-$(I27)_2$-ssrA

Having validated the experimental and data analysis protocols for the well-studied systems, a similar approach was taken for $(I27)_3$-ssrA and I27-Q30-$(I27)_2$-ssrA. Approximately 2000 approach-retract cycles were made at 1000 $nm.s^{-1}$ and the experiment was repeated three times. This pulling speed was chosen because it provides a good balance between the speed at which data can be collected and the number of individual data points collected for each approach-retract cycle. Example force-extension profiles are shown in Figure 4.18. It can be seen that these profiles appear to have three unfolding events and a single unbinding event. There were no examples of force-extension profiles observed that appeared to have a distinct 30Q unfolding event, indicating that either the 30Q was not unfolding or that it was unfolding at a force too low to be detected by the experiment. Since there were no apparent events visible for any 30Q unfolding events, the data needed to be analysed in a different way. It was decided to look only at the contour lengths of the initial unfolding event and final pull-off event for this set of data.

Since this was a novel method of analysing this data it was not possible to directly compare the initial (L_i) and final (L_f) contour lengths to known values to validate the experimental protocol. An initial analysis was carried out using just the unfolding forces and difference in contour lengths for the unfolding I27 domains in the constructs. Since if these were consistent with previous results it would be a good indication that the AFM was set up correctly and providing accurate measurements. The results of this analysis are shown in Figure 4.19 and Figure 4.20, and summarised in Table 4.2 and Table 4.3, the contour lengths obtained (27.3 nm and 26.7 nm) are in good agreement with the

published length of 28.1 nm[138] and the most probable unfolding forces ($(I27)_3$-ssrA: 182.3 pN, I27-Q30-$(I27)_2$-ssrA: 181.6 pN) are in good agreement with previously published data (182.6 pN[154]) and my own experiments (176.8 pN) as shown in Figure 4.16 also agree well with the expected force at that unfolding velocity.

For the mechanical unfolding of Q30 three scenarios were predicted, and are described in Figure 4.12. The data were analysed by looking only at the length of the initial unfolding event and the final extension at the pull-off event where the tip detached from the protein. This would identify whether the Q30 was in a compact globular or expanded form without having to detect individual Q30 unfolding events (see Figure 4.21). Three scenarios could be envisaged: (i) if the Q30 was in an expanded conformation, or too mechanically weak to resist mechanical unfolding it would be expected that the first and final events for I27-Q30-$(I27)_2$-ssrA would be offset by approximately 10.8 nm (30 × 0.36 nm) from the surface in comparison to the control $(I27)_3$-ssrA polyprotein. This is of course an approximation as polyglutamine is intrinsically disordered and has no fixed structure so it is not possible to calculate an exact value. (ii) the polyQ has a resistance to unfolding greater than the maximum force the AFM can apply before the pull-off event, in which case it would be expected that the initial and final unfolding events would close enough to the values for the control protein that the difference would not be detectable by this method. (iii) polyQ would unfold within the force-range of this experiment and a distinct peak for it would be visible in the force-extension profiles recorded.

The experiments yielded much less data in comparison to the experiments described in section 4.5.1 as it was not possible to analyse partial traces. Only

traces which had three distinct unfolding events and a pull-off event were suitable for analysis. As can be seen from the data shown in Figure 4.22 and Figure 4.23 (and summarised in Figure 4.24) the data supports a model whereby polyQ unfolds at a force which is below the thermal noise of the AFM (~18 pN[35]). When comparing the most probable L_i and L_f for the two proteins the values for I27-Q30-$(I27)_2$-ssrA are 10.4 and 10.3 nm longer than those for $(I27)_3$-ssrA respectively. This compares well to the estimated length increase of 10.8 nm.

Although the experimental conditions may not have been ideal due to the low number of I27 domains in the concatamer, it is customary to use AFM to investigate homopolypeptides with five or more domains. An additional complication is the noise close to the surface where the Q30 domain would be observed if it unfolds at low force. It may be possible to overcome this problem by using the carrier-guest strategy developed by the Carrion-Vazquez group[33] (Section 1.6.1.2). To determine whether these results were statistically significant, L_i and L_f values were compared between the two constructs, using a non-paired, two-tailed Student's *t*-test[159] with equal variance. The *t*-test indicates the observed difference in length is significant ($P<0.0001$, $P<0.0005$ respectively) which supports the model that the polyQ is unfolded or mechanically unstable under the experimental conditions employed here.

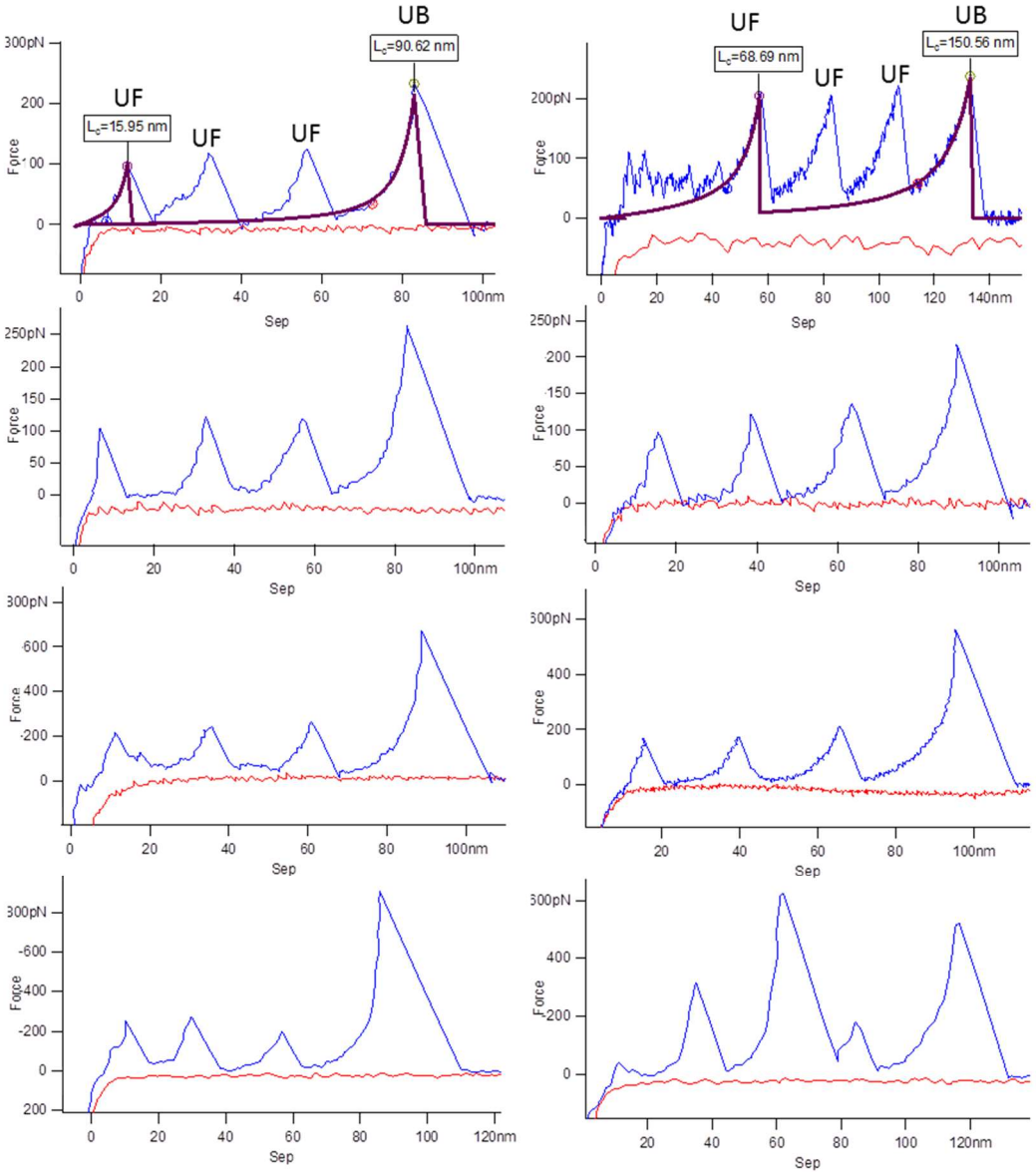


Figure 4.18: Examples of acceptable force-extension profiles obtained for $(I27)_3$-ssrA (left) and I27-Q30-$(I27)_2$-ssrA (right). Red: extension. Blue: retraction. The top row shows WLC fits to the first and last peaks (persistence length = 400 pm) and has the peaks labelled as either UF or UB representing the unfolding of I27 domains or the tip detaching from the polypeptide chain respectively.

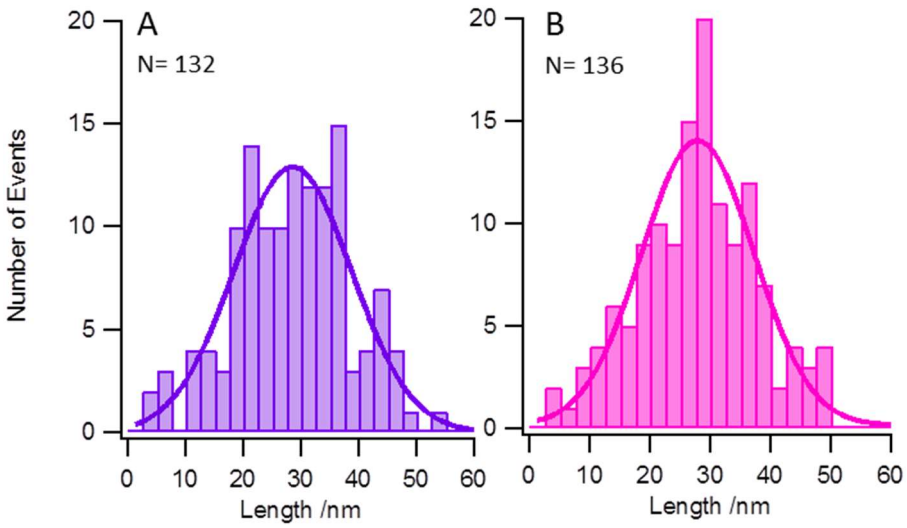


Figure 4.19: Length-frequency histograms generated from $(I27)_3$-ssrA (left) and I27-Q30-$(I27)_2$-ssrA (right) constant velocity mechanical unfolding at 1000 pm/s. Fits of a Gaussian distribution to the histograms are shown as solid lines. Parameters of the fit are shown in Table 4.2.

Table 4.2: Fitting parameters for the Gaussian fits shown in Figure 4.19.

Experiment	χ_0 **/pN**	χ^2	**Width** ($\sqrt{2} \times$ $standard\ deviation$)	**N**
$(I27)_3$-ssrA	27.3±1.0	116.0	14.2 ± 2.4	132
I27-Q30-$(I27)_2$-ssrA	26.7 ± 0.8	96.6	12.2 ± 1.6	136

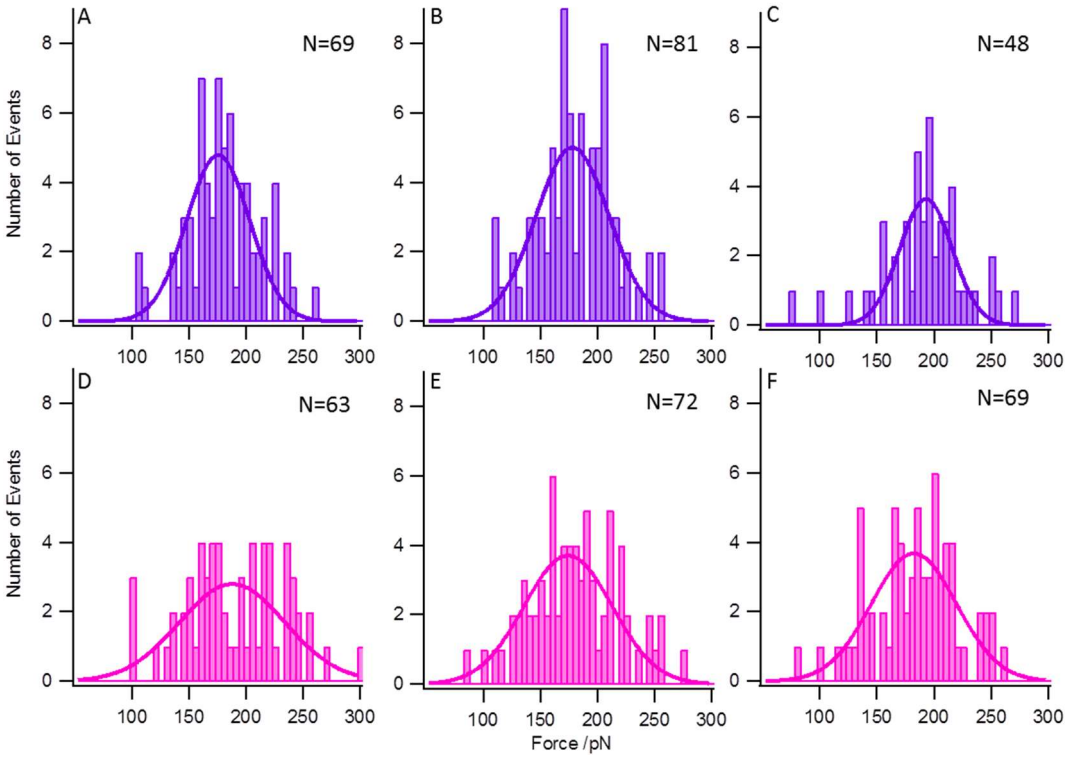


Figure 4.20: Force-frequency histograms generated from $(I27)_3$-ssrA (top row) and I27-Q30-$(I27)_2$-ssrA (bottom row) constant velocity mechanical unfolding at 1000 $nm.s^{-1}$ Fits of a Gaussian distribution to each histogram are shown as a solid line. Fitting parameters for the Gaussians are shown in Table 4.3.

Table 4.3: Fitting parameters for the Gaussians shown in Figure 4.20.

Experiment	χ_0 **/pN**	χ^2	**Width** ($\sqrt{2} \times$ ***standard deviation***)	**N**
$(I27)_3$-ssrA (A)	175.6 ± 3.0	61.0	39.2 ± 4.2	69
$(I27)_3$-ssrA (B)	178.2 ± 3.9	99.6	45.3 ± 5.5	81
$(I27)_3$-ssrA (C)	193.1 ± 2.9	39.0	32.5 ± 4.0	48
I27-Q30-$(I27)_2$-ssrA (D)	187.9 ± 6.5	59.2	65.3 ± 9.3	63
I27-Q30-$(I27)_2$-ssrA (E)	174.3 ± 4.0	46.9	54.6 ± 5.6	72
I27-Q30-$(I27)_2$-ssrA (F)	182.5 ± 4.3	55.2	52.6 ± 6.0	69

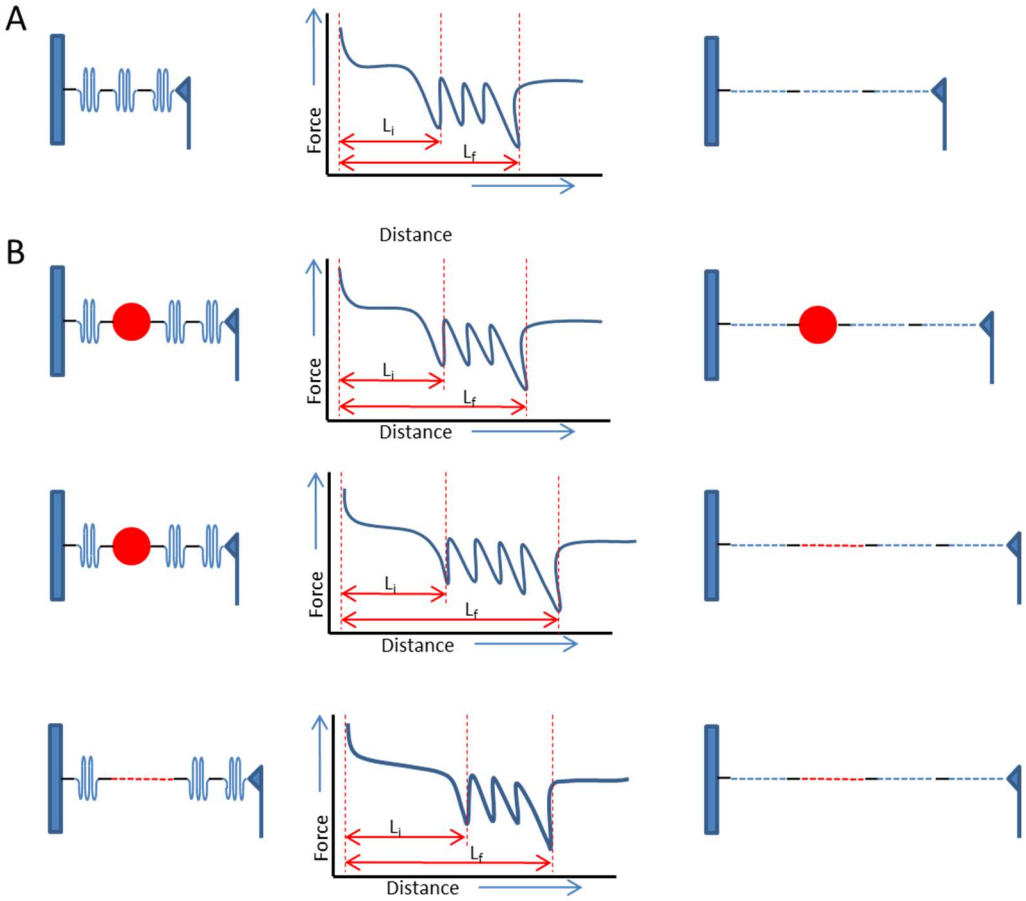


Figure 4.21: A: Expected force-extension profiles for the mechanical unfolding of $(I27)_3$-ssrA. The profile shows three peaks indicating the unfolding of I27 domains and a larger peak representing detachment of the tip from the polyprotein. **B:** Expected force-extension profiles for the three scenarios predicted for the mechanical unfolding of I27-Q30-$(I27)_2$-ssrA. Each profile shows three peaks indicating the unfolding of I27 domains and a larger peak representing detachment of the tip from the polyprotein. They represent the following possible hypotheses: Top - Q30 has extreme mechanical resistance, no peak is observed for its unfolding and the distance to the initial unfolding peak (L_i) and final detachment peak (L_f) are short. Middle: Q30 is mechanically resistant; an additional peak is observed for poly Q unfolding, L_i is short, and L_f is long. Bottom: Q30 is mechanically weak, no peak is observed for unfolding and L_i is long as is L_f.

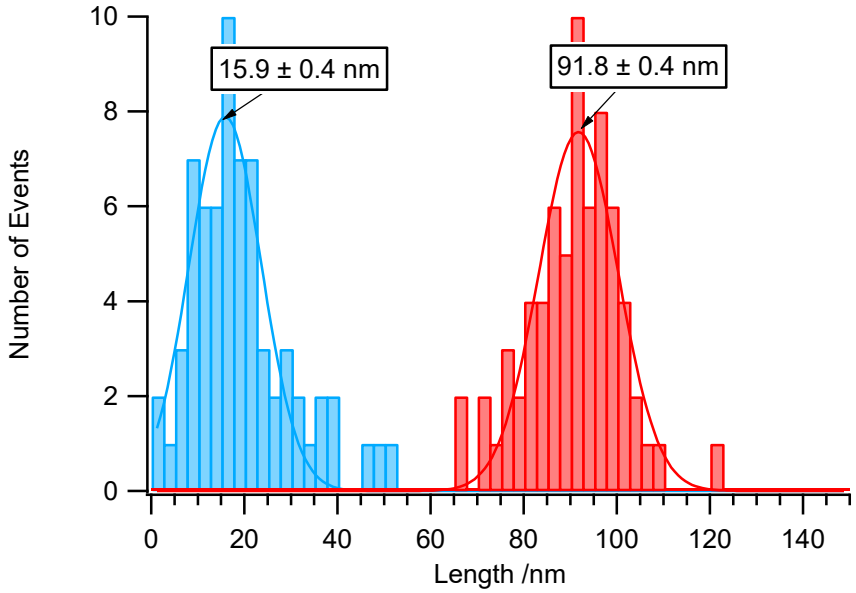


Figure 4.22: Length frequency histogram showing the length value obtained at the initial unfolding event and the final pull-off event for $(I27)_3$-ssrA. Lines are a Gaussian fit to the data and the number shown above each peak is the mean length at the centre of that peak.N=66

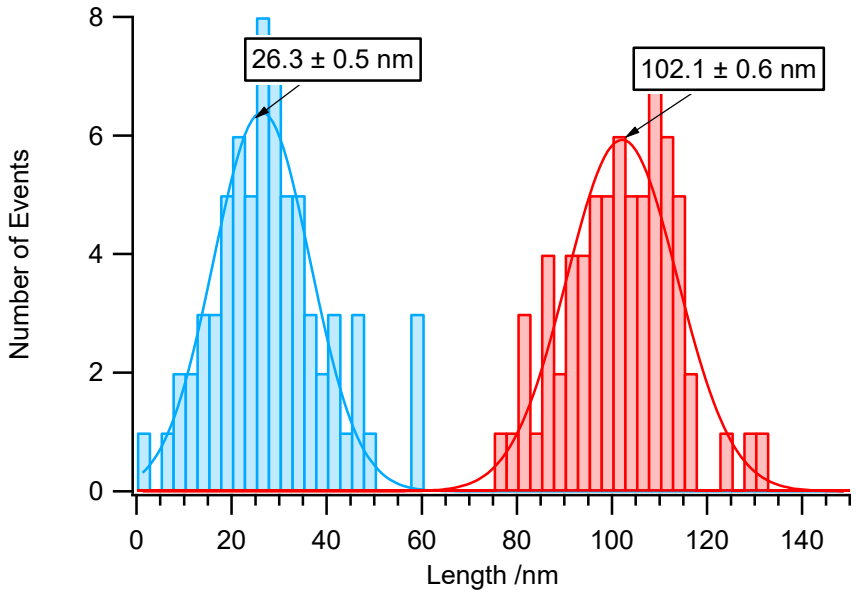


Figure 4.23: Length frequency histogram showing the length value obtained at the initial unfolding event and the final pull-off event for I27-Q30-(I27)2-ssrA. Lines are a Gaussian fit to the data and the number shown above each peak is the mean length at the centre of that peak. N=68

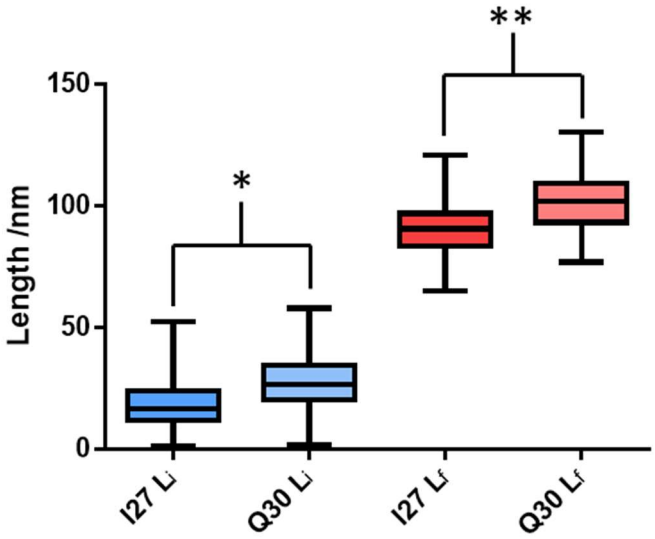


Figure 4.24: A box and whisker plot summarising the data shown in Figure 4.22 and Figure 4.23. For each column the central horizontal line represents the median, the interquartile range is contained within the box and the whiskers display the complete range of the data. * : P<0.0001, ** : P<0.0005.

4.5 $^{\Delta N}ClpX_6P$ Degradation of $(I27)_3$-ssrA and I27-Q30-$(I27)_2$-ssrA

In order to further probe the mechanical stability of proteins containing polyQ repeats, they were degraded by the ClpXP system ($^{\Delta N}ClpX_6$ and ClpP) discussed in Chapter 3. As the rate-limiting step of the ClpXP degradation of a protein is the unfolding step (see introduction). It was hypothesised that if polyQ exhibited an enhanced mechanical stability then $^{\Delta N}ClpX_6P$ would either fail to unfold and degrade it leaving a truncated intermediate (I27-Q30) or it would degrade these constructs more slowly than the control protein without polyQ. The results of these degradations are shown in Figure 4.25. Neither of the degradations had proceeded to completion over the duration of the experiment, longer time points were not recorded because the experimental procedure (described in section 2.8) involved taking samples from the reaction. Increasing the number of samples taken means that either the amount of material used for the experiment is increased, or the number of samples taken at shorter time-points, when the reaction is proceeding more quickly, has to be decreased. While it might have been interesting to see how the reaction progressed to completion, the data collected allowed the quantification of the reaction rates, which was the aim of the experiment. Therefore, expanding the duration of the experiment was not considered necessary.

It can be seen from the SDS-PAGE that there is no detectable intermediate present in the reaction with the polyQ containing protein indicating that ClpXP can completely degrade it. Quantifying the degradation rates by densitometry

(Figure 4.26, Table 4.4) revealed that the polyglutamine containing protein is degraded more quickly than the one which does not contain polyglutamine. This result is difficult to explain in terms of the expected result: that either the degradation rates would be the same or that the polyglutamine containing protein would be degraded more slowly. There are two likely explanations, the first is experimental error. If the change in rate is less than the error involved in the experiment the rates could effectively be the same and the difference observed due to this error. The second is that the polyglutamine is having a destabilising effect on the I27 domains. While it was demonstrated that the I27 domains appeared to be folded under experimental conditions (Section 4.4) they may not have been as stable as in the protein containing no polyQ. It has been suggested that the urea-like side chains of polyQ rich regions can act as a denaturant[160] or it is possible that the importance of the flanking regions to polyQ structure and function are involved (Section 1.5.8). Despite this unexpected result it is clear that the biophysical characterisation together with the ClpXP degradation show that the polyQ containing protein analysed was monodisperse in solution and was not mechanically resistant to unfolding.

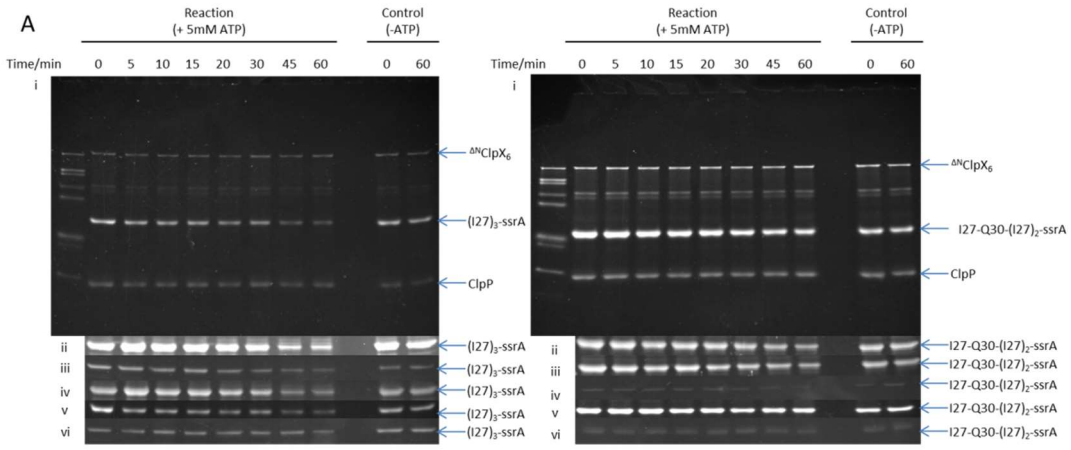


Figure 4.25: SDS-PAGE showing the results of $^{\Delta N}ClpX_6P$ degradation of A: $(I27)_3$-ssrA B: (I27-Q30-$(I27)_2$-ssrA. i: full gels showing all bands ii-vi: repeat experiments showing just the substrate bands. Bands were quantified as described in Section 2.8.1and the results of the quantification are shown in Figure 4.26.

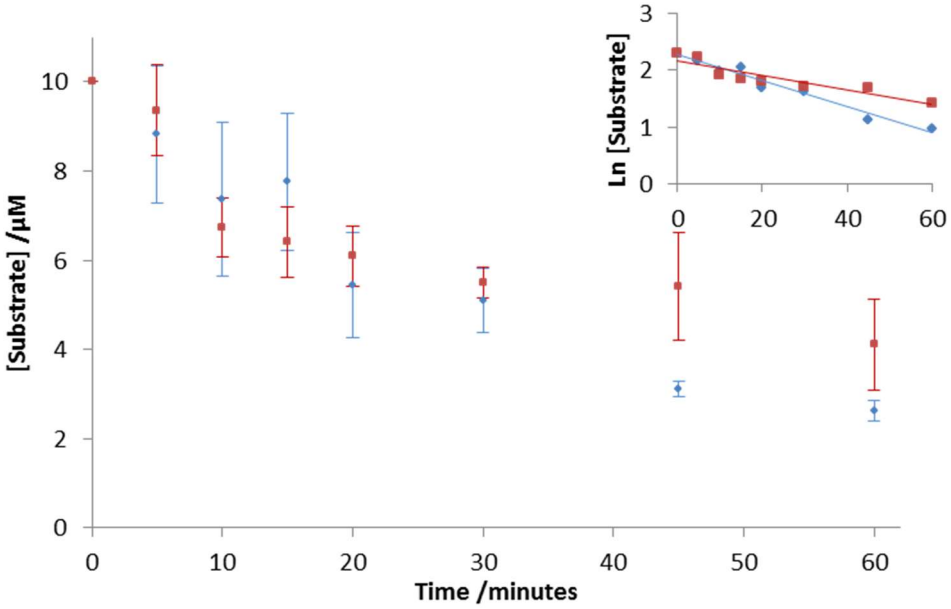


Figure 4.26: Densitometry of the ClpXP degradation of: Red squares: $(I27)_3$-ssrA. Blue diamonds: I27-Q30-$(I27)_2$-ssrA error bars show the standard error. N=6. Inset is the same data plotted as Ln [Substrate] vs Time. Lines are a straight line fit to the data to determine the degradation rate.

Table 4.4: Quantification of degradation rates determined by fitting to Figure 4.26.

	k /s^{-1}	**R^2 for fit**
$(I27)_3$-ssrA.	0.013	0.84
I27-Q30-$(I27)_2$-ssrA	0.023	0.97

5 Discussion

5.1 The use of ClpXP as a method of applying mechanical force

It was demonstrated in Chapter 3 that it is possible to use ClpXP as a tool to investigate the mechanical resistance of a substrate peptide to degradation at biologically accessible force loading rates. The SDS-PAGE assay that was developed is robust and clearly shows the degradation of a protein substrate over the experimental time-course; coupled with densitometry performed on the SDS-PAGE gels the data can be analysed to obtain quantitative data for the degradation rates of different substrates. Table 5.1 shows the results of two such experiments, one to determine whether the presence of DTT affected the rate of ClpXP degradation of Im9-ssrA and one to determine whether a 30Q sequence inserted into a protein affected its degradation rate. These results go some way towards showing the range of different rates which can be resolved as it was possible to measure rates that differed by an order of magnitude, although no attempt to determine the upper and lower rates which could be measured was made. It is reasonable to assume that rates that differed more widely than these could be measured, the times at which samples were taken could easily be altered so that faster reactions could have samples taken at shorter time periods and slower reactions could have longer periods between samples.

Proving that ClpX can operate in different redox environments is an important result because it further expands the range of protein substrates which can be investigated. Proteins with native disulphide bonds, the disruption of which

would take away the biological relevance of the degradation assay performed on them, can now have their mechanical properties investigated in a more native-like environment.

We have also demonstrated that it is possible to fine-tune the substrate specificity of ClpX, while it was not possible to develop a robust assay for a degradation tag other than ssrA I am confident that with further research such an assay could be developed.

It should also be possible to change the conditions under which experiments are performed in order to alter the rate of degradation of the protein substrate which will allow experiments to be fine-tuned to give a measurable change in substrate concentration over a convenient time-course.

However there are some limitations to using ClpXP as a tool to measure mechanical resistance. The most obvious one being that it requires a tagged substrate, which in the most limiting case would allow only the investigation of the mechanical properties of protein expressed in *E. coli* and unfolded from the C-terminus. However many synthesis techniques can overcome this, such as recombinant protein expression in eukaryotic cells which would allow the study of proteins subjected to post-translational modification. The use of techniques such as sulfhydryl-reactive cross-linked chemistry[161,162] or other techniques such as sortase modification[163] could be used to add degradation tags in a site-specific manner.

The major limitation if this technique is that experiments must be performed in conditions under which the ClpXP is stable and active. While the conditions are not particularly onerous for biological samples (200 mM KCl, 10% glycerol *(v/v)*, pH 7.6, 30 °C), and it may be possible to vary them to some extent, they

still require consideration during experimental design and these considerations rule out the possibility of investigating identical proteins under different conditions (temperatures, salt concentrations, denaturant concentrations etc.). It will not be possible to differentiate between the effects of the changes on the mechanical properties of the protein substrate and the effects of changing conditions on the ClpXP unless we can discover a suitable control which could be used to determine the effect of the change in conditions on the ClpXP. For example since ClpXP can degrade a wide variety of substrates, even including ones with non-natural side chains or altered backbone bonds[50], it may be possible to discover a test substrate which had a mechanical stability (measured using an alternative technique such as AFM) that did not change over the range of assay conditions to be investigated, we would then know that any change in degradation rate was due to the effects of the different reaction conditions on the ClpXP. This would allow us to correct the data obtained for the protein we wanted to study by reference to the data obtained for the test protein. Alternatively it may be possible to compare the resting ATP hydrolysis rates of ClpXP under the different conditions and use these as a measure of ClpXP activity under those conditions, although further study would be required to determine whether the resting ATP hydrolysis rate changed proportionally to the degradation rate.

The final limitation relates to batch-to-batch variations when producing ClpXP and the possibility of it becoming less active over time; this will mean that unless ClpXP activity can be normalised, by comparing the degradation of a reference protein perhaps, that it will be difficult to draw firm conclusions by comparing data obtained by different laboratories or over long time periods. To overcome this limitation, we performed experiments comparing different

protein substrates under identical conditions at the same time and using the same batch of ClpXP.

It would be possible to compare the activity of different batches of ClpXP by comparing the degradation rates of a test protein. If several different batches were tested with a single test protein, either one produced in the laboratory and a single batch preparation used to quantify the activity of all further ClpXP produced or a high-quality commercially-available protein that would allow comparison over longer time periods and between different laboratories. This would mean that instead of having ClpXP of a known concentration but unknown activity it would be possible to use an amount of ClpXP known to provide a certain amount of activity. It would be possible to investigate the potential degradation over time of a single batch of ClpXP in the same way. By comparing its activity against a test protein any loss of activity could be detected and its effects accounted for in the experimental design.

Overall the application of mechanical unfolding forces by ClpXP and their analysis using SDS-PAGE is a valuable addition to the toolkit of a scientist interested in the mechanical properties of proteins, particularly because it allows access to biologically relevant forces and loading rates, and replicates the geometry of force application found in chambered proteases *in vivo*, that cannot yet be achieved using other techniques, and makes an excellent complement to dynamic force spectroscopy.

Table 5.1: A summary of protein degradation rates obtained in this thesis.

Substrate protein	Degradation rate (k) /s^{-1}
Im9-ssrA with 2 mM DTT	0.28
Im9-ssrA without DTT	0.15
$(I27)_3$-ssrA.	0.013
I27-Q30-$(I27)_2$-ssrA	0.023

5.2 Mechanical unfolding of polyglutamine repeat containing proteins

I was unable to replicate the work of Dougan et. al[32]. showing that polyQ had extreme mechanical stability, but it is unclear whether this was due to a lack of AFM data or because, as seems apparent from ClpXP degradation, that it did not have any unusual mechanical properties. It is possible that some hyperstable conformations as reported by Hérvas et al[33]. may have been present in our samples, but as they saw only 1% of such conformations at a similar (35Q) polyQ repeat length it is difficult to say whether they were not present or simply too rare to be observed here.

The polyglutamine repeat-length studied (30Q) was degraded by ClpXP without an apparent reduction of the degradation rate compared to a control which did not contain a polyglutamine repeat. This is in contrast to the two DFS studies performed, one of which showed that polyglutamine was extremely mechanically resistant and one of which showed that it had access to a mechanically stable conformation.

I have developed a hypothesis that when studying the force response of polyglutamine the loading rates and directionality of the applied force are very important. This can be thought of as being analogous to Velcro; when two pieces of Velcro are stuck together it is very difficult to separate them by pulling them apart from a direction perpendicular to the joined surfaces, but if they are peeled apart from the ends they can be easily separated. Applying force at a high loading rate by using AFM is analogous to the first situation, but rather than trying to separate Velcro's hooks and eyes you are trying to separate a network of hydrogen bonds which may become locked in a rigid conformation

by the application of force meaning that they must all be broken at the same time in order to unfold the polyglutamine, unfolding with ClpX is analogous to the second situation where the hydrogen bond network is unravelled gradually from the end allowing the bonds to be broken individually as the polyglutamine is unfolded. Since polyglutamine is an intrinsically disordered peptide with no fixed secondary structure it is not possible to determine whether its structure supports such a hypothesis, however molecular dynamics simulations[32], while giving no fixed structure, do support it containing a large number of hydrogen bonds which need to be disrupted in order for it to unfold.

5.3 **Future work**

The obvious follow on from this work is to investigate polyglutamine repeats of different lengths. While the work by Dougan *et al.* found that the mechanical properties of polyglutamine repeats were length-independent, it is clear that the toxic effects and aggregation of polyglutamine are length-dependant and Hervás *et al.* saw an increase in hyperstable events with increasing polyglutamine length. As the repeat length investigated here (30Q) was under the disease length threshold (which varies but is generally considered to be approximately 36Q) it would be worthwhile to investigate longer sequences to determine whether a difference in mechanical strength was observed using the ClpXP assay.

It would also be desirable to re-design the assay so that it could be monitored in a different way. While SDS-PAGE and densitometry is robust and inexpensive it can be seen from Section 3.4.2 that its accuracy is susceptible to loading errors and differences in staining. As mentioned above, it also does

not directly follow unfolding of the substrate. It is also a relatively low-throughput assay, the number of time points which can be measured is limited by: the time taken to physically take samples and mix them with the quenching buffer, and by the desire to avoid differential staining between different SDS-PAGE gels which means all of the samples from one experiment must be analysed on the same gel. In addition removing the samples from the reaction mixture introduces a potential source of error and requires that if more data points are desired a larger reaction volume or an increase in concentration of the reaction components (so that smaller volumes can be removed) is required, this is not an insurmountable problem, but it does add to the complexity of the assay.

To overcome these limitations an assay could be designed which measured the fluorescence of a sample. This would mean that once the substrate was added the reaction mixture would remain undisturbed and many more data points at closer time points could be collected.

Three types of fluorescence assay could be used, illustrated in Figure 5.1, choosing the most appropriate one for the system under investigation: The first would be to use a fluorescent protein downstream of the protein under investigation, this would overcome the limitations of the SDS-PAGE assay but would still not provide a direct measurement of unfolding. The second would be to use an intrinsic fluorophore (such as tryptophan) within the protein under investigation that was known to change environment upon unfolding. This would allow unfolding to be measured directly as well as overcoming the limitations of the SDS-PAGE assay, but would obviously only be suitable for systems containing such an intrinsic fluorophore or to which one could be

added without altering the mechanical properties of the protein. The third would be to make use of fluorescent labelling techniques to add either a FRET pair or a self-quenching fluorophore at locations on the substrate protein where they would be disrupted by its unfolding, while not affecting its mechanical stability.

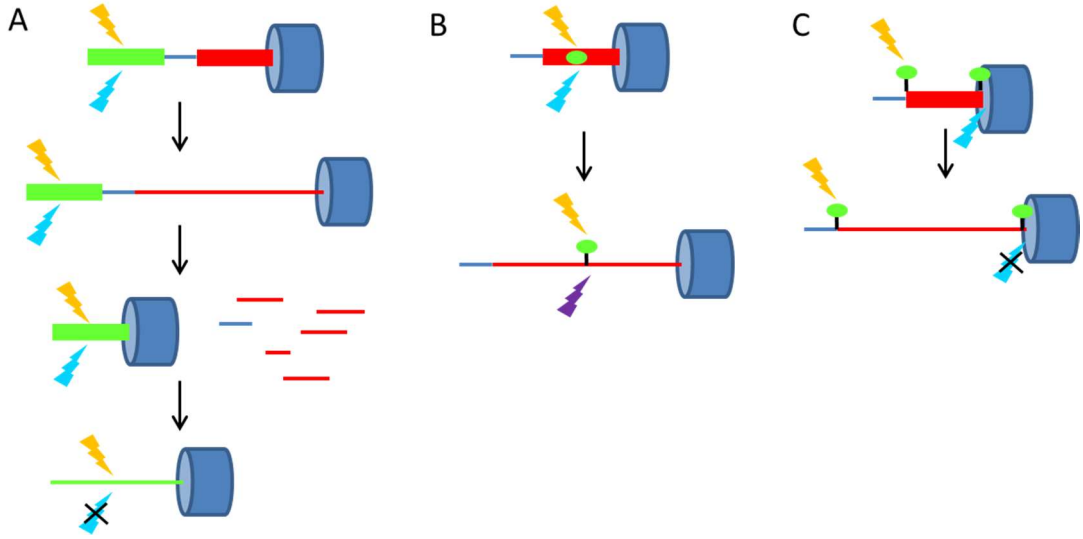


Figure 5.1: Cartoon representation of various methods of using fluorescence to follow a ClpXP (blue cylinder) degradation assay. **A:** A fluorescent protein (green) is added to the protein of interest (red) by a linker (blue line). The assay can be followed by monitoring the fluorescence emission which will end when the fluorescent protein is unfolded. **B:** An intrinsic fluorophore is used to detect unfolding of the protein of interest, the emitted fluorescence will change on unfolding. **C:** A FRET pair is used to detect unfolding, when the protein of interest is unfolded the FRET pair will be too far apart and fluorescence will cease.

6 List of References

1. Newman, R. H., Zhang, J. & Zhu, H. Toward a systems-level view of dynamic phosphorylation networks. *Front. Genet.* **5,** 263 (2014).

2. Alfadda, T. I., Saleh, A. M. A., Houillier, P. & Geibel, J. P. Calcium-sensing receptor 20 years later. *Am. J. Physiol. Cell Physiol.* **307,** C221–C231 (2014).

3. Lim, D. J. & Kalinec, F. Cell and molecular basis of hearing. *Kidney Int. Suppl.* **65,** S104–13 (1998).

4. Maillard, R. a *et al.* ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* **145,** 459–69 (2011).

5. Kaya, M. & Higuchi, H. Stiffness, working stroke, and force of single-myosin molecules in skeletal muscle: elucidation of these mechanical properties via nonlinear elasticity evaluation. *Cell. Mol. Life Sci.* **70,** 4275–92 (2013).

6. Storey, A. & Smith, H. K. Unique aspects of competitive weightlifting: Performance, training and physiology. *Sports Medicine* **42,** 769–90 (2012).

7. Zoldák, G. & Rief, M. Force as a single molecule probe of multidimensional protein energy landscapes. *Curr. Opin. Struct. Biol.* **23,** 48–57 (2012).

8. Fackler, O. T. & Grosse, R. Cell motility through plasma membrane blebbing. *J. Cell Biol.* **181,** 879–84 (2008).

9. Abu Shah, E. & Keren, K. Mechanical forces and feedbacks in cell motility. *Curr. Opin. Cell Biol.* **25,** 550–7 (2013).

10. Fletcher, D. A. & Mullins, R. D. Cell mechanics and the cytoskeleton. *Nature* **463,** 485–92 (2010).

11. Yodh, J. G., Schlierf, M. & Ha, T. Insight into helicase mechanism and function revealed through single-molecule approaches. *Q. Rev. Biophys.* **43,** 185–217 (2010).

12. Sauer, R. T. & Baker, T. A. AAA+ proteases: ATP-fueled machines of protein destruction. *Annu. Rev. Biochem.* **80,** 587–612 (2011).

13. Charras, G. T., Yarrow, J. C., Horton, M. A., Mahadevan, L. & Mitchison, T. J. Non-equilibration of hydrostatic pressure in blebbing cells. *Nature* **435,** 365–9 (2005).

14. Finkelstein, I. J. & Greene, E. C. Molecular traffic jams on DNA. *Annu. Rev. Biophys.* **42,** 241–63 (2013).

15. Huang, S. & Ratliff, K. Mitochondria unfold precursor proteins by unraveling them from their N-termini. *Nat. Struct. Mol. Biol.* **6,** 85–91 (1999).

16. Ferramosca, A. & Zara, V. Biogenesis of mitochondrial carrier proteins: Molecular mechanisms of import into mitochondria. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833,** 494–502 (2013).

17. Marszalek, P. E. & Dufrêne, Y. F. Stretching single polysaccharides and proteins using atomic force microscopy. *Chem. Soc. Rev.* **41,** 3523–34 (2012).

18. Livnah, O., Bayer, E. A., Wilchek, M. & Sussman, J. L. Three-dimensional structures of avidin and the avidin-biotin complex. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 5076–80 (1993).

19. Cascales, E. *et al.* Colicin biology. *Microbiol. Mol. Biol. Rev.* **71,** 158–229 (2007).

20. Thomas, W. E., Trintchina, E., Forero, M., Vogel, V. & Sokurenko, E. V. Bacterial Adhesion to Target Cells Enhanced by Shear Force. *Cell* **109,** 13–923 (2002).

21. Forero, M., Yakovenko, O., Sokurenko, E. V, Thomas, W. E. & Vogel, V. Uncoiling mechanics of Escherichia coli type I fimbriae are optimized for catch bonds. *PLoS Biol.* **4,** e298 (2006).

22. Tchesnokova, V. *et al.* Integrin-like allosteric properties of the catch bond-forming FimH adhesin of Escherichia coli. *J. Biol. Chem.* **283,** 7823–33 (2008).

23. Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. The kinetics of formation of native ribonuclease during oxidation of the reduced

polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.* **47,** 1309–14 (1961).

24. Brockwell, D. J. Probing the mechanical stability of proteins using the atomic force microscope. *Biochem. Soc. Trans.* **35,** 1564–8 (2007).

25. Bustamante, C., Macosko, J. C. & Wuite, G. J. Grabbing the cat by the tail: manipulating molecules one by one. *Nat. Rev. Mol. Cell Biol.* **1,** 130–6 (2000).

26. De Vlaminck, I. & Dekker, C. Recent advances in magnetic tweezers. *Annu. Rev. Biophys.* **41,** 453–72 (2012).

27. Bryant, Z., Oberstrass, F. C. & Basu, A. Recent developments in single-molecule DNA mechanics. *Curr. Opin. Struct. Biol.* **22,** 304–12 (2012).

28. Walter, N. G., Huang, C.-Y., Manzo, A. J. & Sobhy, M. A. Do-it-yourself guide: how to use the modern single-molecule toolkit. *Nat. Methods* **5,** 475–89 (2008).

29. Aubin-Tam, M.-E., Olivares, A. O., Sauer, R. T., Baker, T. A. & Lang, M. J. Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. *Cell* **145,** 257–67 (2011).

30. Hoffmann, T. & Dougan, L. Single molecule force spectroscopy using polyproteins. *Chem. Soc. Rev.* **41,** 4781–96 (2012).

31. Puchner, E. M. & Gaub, H. E. Force and function: probing proteins with AFM-based force spectroscopy. *Curr. Opin. Struct. Biol.* **19,** 605–14 (2009).

32. Dougan, L., Li, J., Badilla, C. L., Berne, B. J. & Fernandez, J. M. Single homopolypeptide chains collapse into mechanically rigid conformations. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 12605–10 (2009).

33. Hervás, R. *et al.* Common features at the start of the neurodegeneration cascade. *PLoS Biol.* **10,** e1001335 (2012).

34. Carrion-Vazquez, M. *et al.* Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl. Acad. Sci. U.S.A.* **96,** 3694–9 (1999).

35. Farrance, O. E. *et al.* A force-activated trip switch triggers rapid dissociation of a colicin from its immunity protein. *PLoS Biol.* **11,**

e1001489 (2013).

36. Best, R. Force mode atomic force microscopy as a tool for protein folding studies. *Anal. Chim. Acta* **479,** 87–105 (2003).

37. Marko, J. J. F. & Siggia, E. D. E. Stretching dna. *Macromolecules* **28,** 8759–70 (1995).

38. Flory, P. J. *Statistical Mechanics of Chain Molecules. The Statistical Distribution of Configuration.* (Wiley Interscience, 1969).

39. Anderson, K., Radford, S., Smith, D. & Brockwell, D. in *Handbook of Molecular Force Spectroscopy* (ed. Noy, A.) 205–51 (Springer, 2008).

40. Crampton, N. & Brockwell, D. J. Unravelling the design principles for single protein mechanical strength. *Curr. Opin. Struct. Biol.* **4,** 508–17 (2010).

41. Merkel, R., Nassoy, P., Leung, A., Ritchie, K. & Evans, E. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature* **397,** 50–3 (1999).

42. Improta, S., Politou, A. S. & Pastore, A. Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure* **4,** 323–37 (1996).

43. Iyer, L. M., Leipe, D. D., Koonin, E. V & Aravind, L. Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.* **146,** 11–31

44. Lupas, A. & Martin, J. AAA proteins. *Curr. Opin. Struct. Biol.* **12,** 746–53 (2002).

45. Zolkiewski, M. A camel passes through the eye of a needle: protein unfolding activity of Clp ATPases. *Mol. Microbiol.* **61,** 1094–100 (2006).

46. Wojtkowiak, D., Georgopoulos, C. & Zylicz, M. Isolation and characterization of ClpX, a new ATP-dependent specificity component of the Clp protease of Escherichia coli. *J. Biol. Chem.* **268,** 22609–17 (1993).

47. Baker, T. A. & Sauer, R. T. ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim. Biophys. Acta* **1,** 15–28 (2011).

48. Kim, Y. I., Burton, R. E., Burton, B. M., Sauer, R. T. & Baker, T. A. Dynamics of substrate denaturation and translocation by the ClpXP degradation machine. *Mol. Cell* **5,** 639–48 (2000).

49. Sauer, R. T. *et al.* Sculpting the proteome with AAA+ proteases and disassembly machines. *Cell* **119,** 9–18 (2004).

50. Barkow, S. R., Levchenko, I., Baker, T. A. & Sauer, R. T. Polypeptide translocation by the AAA+ ClpXP protease machine. *Chem. Biol.* **16,** 605–12 (2009).

51. Martin, A., Baker, T. A. & Sauer, R. T. Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature* **437,** 1115–20 (2005).

52. Martin, A., Baker, T. A. & Sauer, R. T. Pore loops of the AAA+ ClpX machine grip substrates to drive translocation and unfolding. *Nat. Struct. Mol. Biol.* **15,** 1147–51 (2008).

53. Kenniston, J. A., Baker, T. A., Fernandez, J. M. & Sauer, R. T. Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell* **114,** 511–20 (2003).

54. Kenniston, J. A., Burton, R. E., Siddiqui, S. M., Baker, T. A. & Sauer, R. T. Effects of local protein stability and the geometric position of the substrate degradation tag on the efficiency of ClpXP denaturation and degradation. *J. Struct. Biol.* **146,** 130–40 (2004).

55. Alexopoulos, J. A., Guarné, A. & Ortega, J. ClpP: a structurally dynamic protease regulated by AAA+ proteins. *J. Struct. Biol.* **179,** 202–10 (2012).

56. Nyquist, K. & Martin, A. Marching to the beat of the ring: polypeptide translocation by AAA+ proteases. *Trends Biochem. Sci.* **39,** 53–60 (2014).

57. Kim, D. Y. & Kim, K. K. Crystal structure of ClpX molecular chaperone from Helicobacter pylori. *J. Biol. Chem.* **278,** 50664–70 (2003).

58. Donaldson, L. W., Wojtyra, U. & Houry, W. A. Solution structure of the dimeric zinc binding domain of the chaperone ClpX. *J. Biol. Chem.* **278,**

48991–6 (2003).

59. Park, E. Y. *et al.* Structural basis of SspB-tail recognition by the zinc binding domain of ClpX. *J. Mol. Biol.* **367,** 514–26 (2007).

60. Grimaud, R., Kessel, M., Beuron, F., Steven, A. C. & Maurizi, M. R. Enzymatic and structural similarities between the Escherichia coli ATP-dependent proteases, ClpXP and ClpAP. *J. Biol. Chem.* **273,** 12476–81 (1998).

61. Stinson, B. M. *et al.* Nucleotide Binding and Conformational Switching in the Hexameric Ring of a AAA+ Machine. *Cell* **153,** 628–39 (2013).

62. Glynn, S. E., Martin, A., Nager, A. R., Baker, T. A. & Sauer, R. T. Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell* **139,** 744–56 (2009).

63. Hayes, C. S., Bose, B. & Sauer, R. T. Stop codons preceded by rare arginine codons are efficient determinants of SsrA tagging in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 3440–5 (2002).

64. Flynn, J. M., Neher, S. B., Kim, Y. I., Sauer, R. T. & Baker, T. A. Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol. Cell* **11,** 671–83 (2003).

65. Neher, S. B. *et al.* Proteomic profiling of ClpXP substrates after DNA damage reveals extensive instability within SOS regulon. *Mol. Cell* **22,** 193–204 (2006).

66. Farrell, C. M., Baker, T. A. & Sauer, R. T. Altered specificity of a AAA+ protease. *Mol. Cell* **25,** 161–6 (2007).

67. Martin, A., Baker, T. A. & Sauer, R. T. Diverse pore loops of the AAA+ ClpX machine mediate unassisted and adaptor-dependent recognition of ssrA-tagged substrates. *Mol. Cell* **29,** 441–50 (2008).

68. Martin, A., Baker, T. A. & Sauer, R. T. Distinct static and dynamic interactions control ATPase-peptidase communication in a AAA+ protease. *Mol. Cell* **27,** 41–52 (2007).

69. Siddiqui, S. M., Sauer, R. T. & Baker, T. A. Role of the processing pore of the ClpX AAA+ ATPase in the recognition and engagement of specific protein substrates. *Genes Dev.* **18,** 369–74 (2004).

70. Wojtyra, U. A., Thibault, G., Tuite, A. & Houry, W. A. The N-terminal zinc binding domain of ClpX is a dimerization domain that modulates the chaperone function. *J. Biol. Chem.* **278,** 48981–90 (2003).

71. Karzai, A. W., Roche, E. D. & Sauer, R. T. The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. **7,** 449–55 (2000).

72. Dougan, D. AAA+ proteins and substrate recognition, it all depends on their partner in crime. *FEBS Lett.* **529,** 6–10 (2002).

73. Levchenko, I., Grant, R. A., Flynn, J. M., Sauer, R. T. & Baker, T. A. Versatile modes of peptide recognition by the AAA+ adaptor protein SspB. *Nat. Struct. Mol. Biol.* **12,** 520–5 (2005).

74. Cordova, J. C. *et al.* Stochastic but highly coordinated protein unfolding and translocation by the ClpXP proteolytic machine. *Cell* **158,** 647–58 (2014).

75. Wang, J., Hartling, J. A. A. & Flanagan, J. M. M. The structure of ClpP at 2.3 A resolution suggests a model for ATP-dependent proteolysis. *Cell* **91,** 447–56 (1997).

76. Brötz-Oesterhelt, H. *et al.* Dysregulation of bacterial proteolytic machinery by a new class of antibiotics. *Nat. Med.* **11,** 1082–7 (2005).

77. Burton, R. E., Siddiqui, S. M., Kim, Y. I., Baker, T. A. & Sauer, R. T. Effects of protein stability and structure on substrate processing by the ClpXP unfolding and degradation machine. *EMBO J.* **20,** 3092–100 (2001).

78. Whitby, F. G. *et al.* Structural basis for the activation of 20S proteasomes by 11S regulators. *Nature* **408,** 115–20 (2000).

79. Thompson, M. W. & Maurizi, M. R. Activity and specificity of Escherichia coli ClpAP protease in cleaving model peptide substrates. *J. Biol. Chem.* **269,** 18201–8 (1994).

80. Thompson, M. W., Singh, S. K. & Maurizi, M. R. Processive degradation of proteins by the ATP-dependent Clp protease from Escherichia coli. Requirement for the multiple array of active sites in ClpP but not ATP hydrolysis. *J. Biol. Chem.* **269,** 18209–15 (1994).

81. Ortega, J., Lee, H., Maurizi, M. & Steven, A. Alternating translocation of protein substrates from both ends of ClpXP protease. *EMBO J.* **21,** 4938–49 (2002).

82. Ortega, J., Singh, S. K., Ishikawa, T., Maurizi, M. R. & Steven, A. C. Visualization of substrate binding and translocation by the ATP-dependent protease, ClpXP. *Mol. Cell* **6,** 1515–21 (2000).

83. Joshi, S. A., Hersch, G. L., Baker, T. A. & Sauer, R. T. Communication between ClpX and ClpP during substrate processing and degradation. *Nat. Struct. Mol. Biol.* **11,** 404–11 (2004).

84. Martin, A., Baker, T. A. & Sauer, R. T. Protein unfolding by a AAA+ protease is dependent on ATP-hydrolysis rates and substrate energy landscapes. *Nat. Struct. Mol. Biol.* **15,** 139–45 (2008).

85. MacDonald, M. E. *et al.* A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72,** 971–83 (1993).

86. Orr, H. T. & Zoghbi, H. Y. Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* **30,** 575–621 (2007).

87. Li, L.-B. B., Yu, Z., Teng, X. & Bonini, N. M. RNA toxicity is a component of ataxin-3 degeneration in Drosophila. *Nature* **453,** 1107–11 (2008).

88. Imarisio, S. *et al.* Huntington's disease: from pathology and genetics to potential therapies. *Biochem. J.* **412,** 191–209 (2008).

89. Butland, S. L. *et al.* CAG-encoded polyglutamine length polymorphism in the human genome. *BMC Genomics* **8,** 126 (2007).

90. Faux, N. G. *et al.* Functional insights from the distribution and role of homopeptide repeat-containing proteins. *Genome Res.* **15,** 537–51 (2005).

91. Yamada, M., Sato, T., Tsuji, S. & Takahashi, H. CAG repeat disorder models and human neuropathology: similarities and differences. *Acta Neuropathol.* **115,** 71–86 (2008).

92. Ribai, P. *et al.* Psychiatric and cognitive difficulties as indicators of juvenile huntington disease onset in 29 patients. *Arch. Neurol.* **64,** 813–

9 (2007).

93. Bauer, P. O. & Nukina, N. The pathogenic mechanisms of polyglutamine diseases and current therapeutic strategies. *J. Neurochem.* **110,** 1737–65 (2009).

94. Zoghbi, H. Y. & Orr, H. T. Glutamine repeats and neurodegeneration. *Annu. Rev. Neurosci.* **23,** 217–47 (2000).

95. La Spada, A. R. & Taylor, J. P. Polyglutamines placed into context. *Neuron* **38,** 681–4 (2003).

96. Ribaï, P. *et al.* Psychiatric and cognitive difficulties as indicators of juvenile huntington disease onset in 29 patients. *Arch. Neurol.* **64,** 813–9 (2007).

97. Gatchel, J. R. & Zoghbi, H. Y. Diseases of unstable repeat expansion: mechanisms and common principles. *Nat. Rev. Genet.* **6,** 743–55 (2005).

98. Squitieri, F., Frati, L., Ciarmiello, A., Lastoria, S. & Quarrell, O. Juvenile Huntington's disease: Does a dosage-effect pathogenic mechanism differ from the classical adult disease? *Mech. Ageing Dev.* **127,** 208–12 (2006).

99. Nance, M. A., Mathias-Hagen, V., Breningstall, G., Wick, M. J. & McGlennen, R. C. Analysis of a very large trinucleotide repeat in a patient with juvenile Huntington's disease. *Neurology* **52,** 392–4 (1999).

100. Williams, A. J. & Paulson, H. L. Polyglutamine neurodegeneration: protein misfolding revisited. *Trends Neurosci.* **31,** 521–8 (2008).

101. Chen, S., Berthelier, V., Yang, W. & Wetzel, R. Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity. *J. Mol. Biol.* **311,** 173–82 (2001).

102. Crick, S. L., Jayaraman, M., Frieden, C., Wetzel, R. & Pappu, R. V. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 16764–9 (2006).

103. Ross, C. A. & Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nat. Med.* **10 Suppl,** S10–7 (2004).

104. Slow, E. J. *et al.* Absence of behavioral abnormalities and neurodegeneration in vivo despite widespread neuronal huntingtin inclusions. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 11402–7 (2005).

105. Kim, M. W., Chelliah, Y., Kim, S. W., Otwinowski, Z. & Bezprozvanny, I. Secondary structure of huntingtin amino-terminal region. *Structure* **17,** 1205–12 (2009).

106. Walters, R. H. & Murphy, R. M. Examining polyglutamine peptide length : A connection between collapsed conformations and increased aggregation. *J. Mol. Biol.* **393,** 978–92 (2009).

107. Pfister, E. L. & Zamore, P. D. Huntington's disease: Silencing a brutal killer. *Exp. Neurol.* **220,** 226–9 (2009).

108. de Fougerolles, A., Vornlocher, H.-P., Maraganore, J. & Lieberman, J. Interfering with disease: a progress report on siRNA-based therapeutics. *Nat. Rev. Drug Discov.* **6,** 443–53 (2007).

109. Deepak, S. *et al.* Polyglutamine homopolymers having 8-45 residues form slablike beta-crystallite assemblies. *Proteins Struct. Funct. Bioinforma.* **61,** 398–411 (2005).

110. Poirier, M. A., Jiang, H. & Ross, C. A. A structure-based analysis of huntingtin mutant polyglutamine aggregation and toxicity: evidence for a compact beta-sheet structure. *Hum. Mol. Genet.* **14,** 765–74 (2005).

111. Perutz, M. F. Glutamine repeats as polar zippers: their role in inherited neurodegenerative disease. *Mol. Med.* **1,** 718–21 (1995).

112. Vitalis, A., Wang, X. & Pappu, R. V. Quantitative characterization of intrinsic disorder in polyglutamine: Insights from analysis based on polymer theories. *Biophys. J.* **93,** 1923–37 (2007).

113. Slepko, N. *et al.* Normal-repeat-length polyglutamine peptides accelerate aggregation nucleation and cytotoxicity of expanded polyglutamine proteins. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 14367–72 (2006).

114. Thakur, A. K. *et al.* Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. *Nat. Struct. Mol. Biol.* **16,** 380–9 (2009).

115. Jayaraman, M., Kodali, R. & Wetzel, R. The impact of ataxin-1-like histidine insertions on polyglutamine aggregation. *Protein Eng. Des. Sel.* **22,** 469–478 (2009).

116. Masino, L., Kelly, G., Leonard, K., Trottier, Y. & Pastore, A. Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins. *FEBS Lett.* **513,** 267–72 (2002).

117. Wetzel, R. Physical chemistry of polyglutamine: intriguing tales of a monotonous sequence. *J. Mol. Biol.* **421,** 466–90 (2012).

118. Bhattacharyya, A. *et al.* Oligoproline effects on polyglutamine conformation and aggregation. *J. Mol. Biol.* **355,** 524–35 (2006).

119. Darnell, G. D., Derryberry, J., Kurutz, J. W. & Meredith, S. C. Mechanism of cis-inhibition of polyQ fibrillation by polyP: PPII oligomers and the hydrophobic effect. *Biophys. J.* **97,** 2295–305 (2009).

120. Raspe, M. *et al.* Mimicking proteasomal release of polyglutamine peptides initiates aggregation and toxicity. *J. Cell Sci.* **122,** 3262–71 (2009).

121. Holmberg, C. I., Staniszewski, K. E., Mensah, K. N., Matouschek, A. & Morimoto, R. I. Inefficient degradation of truncated polyglutamine proteins by the proteasome. *EMBO J.* **23,** 4307–18 (2004).

122. Venkatraman, P., Wetzel, R., Tanaka, M., Nukina, N. & Goldberg, A. L. Eukaryotic proteasomes cannot digest polyglutamine sequences and release them during degradation of polyglutamine-containing proteins. *Mol. Cell* **14,** 95–104 (2004).

123. Li, X., Li, H. & Li, X. J. Intracellular degradation of misfolded proteins in polyglutamine neurodegenerative diseases. *Brain Res. Rev.* **59,** 245–52 (2008).

124. Menzies, F. M. *et al.* Autophagy induction reduces mutant ataxin-3 levels and toxicity in a mouse model of spinocerebellar ataxia type 3. *Brain* **133,** 93–104 (2010).

125. Navon, A. & Ciechanover, A. The 26 S proteasome: from basic mechanisms to drug targeting. *J. Biol. Chem.* **284,** 33713–8 (2009).

126. Wang, X., Vitalis, A., Wyczalkowski, M. A. & Pappu, R. V. Characterizing the conformational ensemble of monomeric polyglutamine. *Proteins Struct. Funct. Bioinforma.* **63,** 297–311 (2006).

127. Maynard, C. J. *et al.* Accumulation of ubiquitin conjugates in a polyglutamine disease model occurs without global ubiquitin/proteasome system impairment. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 13986–91 (2009).

128. Bence, N. F., Sampat, R. M. & Kopito, R. R. Impairment of the ubiquitin-proteasome system by protein aggregation. *Science* **292,** 1552–5 (2001).

129. Ortega, Z. *et al.* Acute polyglutamine expression in inducible mouse model unravels ubiquitin/proteasome system impairment and permanent recovery attributable to aggregate formation. *J. Neurosci.* **30,** 3675–88 (2010).

130. Albrecht, A. & Mundlos, S. The other trinucleotide repeat: polyalanine expansion disorders. *Curr. Opin. Genet. Dev.* **15,** 285–93 (2005).

131. Eichner, T. & Radford, S. E. Understanding the complex mechanisms of β2-microglobulin amyloid assembly. *FEBS J.* **278,** 3868–83 (2011).

132. Kühlmann, U. C., Pommer, A. J., Moore, G. R., James, R. & Kleanthous, C. Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. *J. Mol. Biol.* **301,** 1163–78 (2000).

133. Bonvin, A. M. *et al.* Nuclear magnetic resonance solution structure of the Arc repressor using relaxation matrix calculations. *J. Mol. Biol.* **236,** 328–41 (1994).

134. NEB. Double Digest Finder, NEB. at <http://www.neb.com/nebecomm/DoubleDigestCalculator.asp>

135. Shin, Y. *et al.* Single-molecule denaturation and degradation of proteins by the AAA+ ClpXP protease. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 19340–5 (2009).

136. Burgess, R. R. in *Methods in Enzymology Volume 463* **463,** 331–342 (Elsevier, 2009).

137. Butt, H.-J. & Jaschke, M. Calculation of thermal noise in atomic force microscopy. *Nanotechnology* **6,** 1–7 (1995).

138. Carrion-Vazquez, M., Marszalek, P. E., Oberhauser, A. F. & Fernandez, J. M. Atomic force microscopy captures length phenotypes in single proteins. *Proc. Natl. Acad. Sci. U.S.A.* **96,** 11288–92 (1999).

139. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78,** 1606–19 (2000).

140. Laue, T. Sednterp.

141. BL21(DE3)pLysS Competent Cells. at <http://www.promega.co.uk/products/cloning-and-dna-markers/cloning-tools-and-competent-cells/bacterial-strains-and-competent-cells/bl21_de3_plyss-competent-cells/?__utma=1.431399854.1380028643.1380028643.1380028643.1&__utmb=1.1.10.1380028643&__utmc=1&__utmx=-&__utmz=1.1380028643.1.1.utmcsr=google|utmccn=(organic)|utmcmd=organic|utmctr=(not provided)&__utmv=-&__utmk=233858718>

142. Sørensen, H. P. & Mortensen, K. K. Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microb. Cell Fact.* **4,** 1 (2005).

143. Steinberg, T. H., Jones, L. J., Haugland, R. P. & Singer, V. L. SYPRO orange and SYPRO red protein gel stains: one-step fluorescent staining of denaturing gels for detection of nanogram levels of protein. *Anal. Biochem.* **239,** 223–37 (1996).

144. Fazekas de St Groth, S., Webster, R. G. & Datyner, A. Two new staining procedures for quantitative estimation of proteins on electrophoretic strips. *Biochim. Biophys. Acta* **71,** 377–91 (1963).

145. Alliegro, M. C. Effects of dithiothreitol on protein activity unrelated to thiol-disulfide exchange: for consideration in the analysis of protein function with Cleland's reagent. *Anal. Biochem.* **282,** 102–6 (2000).

146. Littler, D. R. *et al.* The intracellular chloride ion channel protein CLIC1 undergoes a redox-controlled structural transition. *J. Biol. Chem.* **279,** 9298–305 (2004).

147. Zhuo, S. & Dixon, J. E. Effects of sulfhydryl regents on the activity of

lambda Ser/Thr phosphoprotein phosphatase and inhibition of the enzyme by zinc ion. *Protein Eng. Des. Sel.* **10,** 1445–52 (1997).

148. Chowdhury, T., Chien, P., Ebrahim, S., Sauer, R. T. & Baker, T. A. Versatile modes of peptide recognition by the ClpX N domain mediate alternative adaptor-binding specificities in different bacterial species. *Protein Sci.* **19,** 242–54 (2010).

149. Thibault, G. & Houry, W. A. Role of the N-terminal domain of the chaperone ClpX in the recognition and degradation of lambda phage protein O. *J. Phys. Chem. B* **116,** 6717–24 (2012).

150. Stothard P. The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 1102–1104 (2000). at <http://www.bioinformatics.org/sms2/reference.html>

151. Thibault, G., Tsitrin, Y., Davidson, T., Gribun, A. & Houry, W. A. Large nucleotide-dependent movement of the N-terminal domain of the ClpX chaperone. *EMBO J.* **25,** 3367–76 (2006).

152. Trebes, A. Protein Degradation by the ClpXP Enzyme. (University of Leeds, 2009).

153. BL21(DE3)pLysS Competent Cells - EMD4Biosciences | EMD Millipore USA. at <http://www.emdmillipore.com/life-science-research/bl21de3plyss-competent-cells/EMD_BIO-69451/p_ip2b.s1OYzkAAAEjOBl9.zLX>

154. Brockwell, D. J. *et al.* The effect of core destabilization on the mechanical resistance of I27. *Biophys. J.* **83,** 458–472 (2002).

155. Alexandrov, I. M., Vishnevskaya, A. B., Ter-Avanesyan, M. D. & Kushnirov, V. V. Appearance and propagation of polyglutamine-based amyloids in yeast: tyrosine residues enable polymer fragmentation. *J. Biol. Chem.* **283,** 15185–92 (2008).

156. DNA secondary structure prediction. at <http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/dna-form1.cgi>

157. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751,** 119–39 (2005).

158. Miller, J., Rutenber, E. & Muchowski, P. J. Polyglutamine dances the conformational cha-cha-cha. *Structure* **17,** 1151–3 (2009).

159. Student. The probable error of a mean. *Biometrika* **6,** 1–25 (1908).

160. England, J. L. & Kaganovich, D. Polyglutamine shows a urea-like affinity for unfolded cytosolic protein. *FEBS Lett.* **585,** 381–4 (2011).

161. Baca, M., Muir, T. W., Schnoelzer, M. & Kent, S. B. H. Chemical ligation of cysteine-containing peptides: Synthesis of a 22 kDa tethered dimer of HIV-1 protease. *J. Am. Chem. Soc.* **117,** 1881–7 (1995).

162. Smith, M. E. B. *et al.* Protein modification, bioconjugation, and disulfide bridging using bromomaleimides. *J. Am. Chem. Soc.* **132,** 1960–5 (2010).

163. Proft, T. Sortase-mediated protein ligation: an emerging biotechnology tool for protein modification and immobilisation. *Biotechnol. Lett.* **32,** 1–10 (2010).

7 Appendix

7.1 DNA and Amino Acid Sequences used in this work

7.1.1 ClpX Variants

7.1.1.1 Protein: ClpX

Brief Description: Pseudo wild-type ClpX with N-terminal His tag

Plasmid: pET23ClpX

Antibiotic resistance encoded in plasmid: Ampicillin

Number of amino acids: 431

Protein Mass: 47310.0 Da

Calculated extinction coefficient: 16765 $M.cm^{-1}$

DNA sequence:

```
   1  ATGCATCATC ATCATCATCA TATGACCGAT AAACGTAAAG ATGGCAGCGG CAAACTGCTG
  61  TATTGCAGCT TTTGCGGCAA AAGCCAGCAT GAAGTGCGTA AACTGATTGC GGGCCCGAGC
 121  GTGTATATTT GCGATGAATG CGTGGATCTG TGCAACGATA TTATTCGTGA AGAAATTAAA
 181  GAAGTGGCGC CGCATCGTGA ACGTAGCGCG CTGCCGACCC CGCATGAAAT TCGTAACCAT
 241  CTGGATGATT ATGTGATTGG CCAGGAACAG GCGAAAAAAG TGCTGGCGGT GGCGGTGTAT
 301  AACCATTATA AACGTCTGCG TAACGGCGAT ACCAGCAACG GCGTGGAACT GGGCAAAAGC
 361  AACATTCTGC TGATTGGCCC GACCGGCAGC GGCAAAACCC TGCTGGCGGA AACCCTGGCG
 421  CGTCTGCTGG ATGTGCCGTT TACCATGGCG GATGCGACCA CCCTGACCGA AGCGGGCTAT
 481  GTGGGCGAAG ATGTGGAAAA CATTATTCAG AAACTGCTGC AGAAATGCGA TTATGATGTG
 541  CAGAAAGCGC AGCGTGGCAT TGTGTATATT GATGAAATTG ATAAAATTAG CCGTAAAAGC
 601  GATAACCCGA GCATTACCCG TGATGTGAGC GGCGAAGGCG TGCAGCAGGC GCTGCTGAAA
 661  CTGATTGAAG GCACCGTGGC GGCGGTGCCG CCGCAGGGCG GCCGTAAACA TCCGCAGCAG
 721  GAATTTCTGC AGGTGGATAC CAGCAAAATT CTGTTTATTT GCGGCGGCGC GTTTGCGGGC
 781  CTGGATAAAG TGATTAGCCA TCGTGTGGAA ACCGGCAGCG GCATTGGCTT TGGCGCGACC
 841  GTGAAAGCGA AAAGCGATAA AGCGAGCGAA GGCGAACTGC TGGCGCAGGT GGAACCGGAA
 901  GATCTGATTA AATTTGGCCT GATTCCGGAA TTTATTGGCC GTCTGCCGGT GGTGGCGACC
 961  CTGAACGAAC TGAGCGAAGA AGCGCTGATT CAGATTCTGA AAGAACCGAA AAACGCGCTG
1021  ACCAAACAGT ATCAGGCGCT GTTTAACCTG GAAGGCGTGG ATCTGGAATT TCGTGATGAA
1081  GCGCTGGATG CGATTGCGAA AAAAGCGATG GCGCGTAAAA CCGGCGCGCG TGGCCTGCGT
1141  AGCATTGTGG AAGCGGCGCT GCTGGATACC ATGTATGATC TGCCGAGCAT GGAAGATGTG
1201  GAAAAAGTGG TGATTGATGA AAGCGTGATT GATGGCCAGA GCAAACCGCT GCTGATTTAT
1261  GGCAAACCGG AAGCGCAGCA GGCGAGCGGC GAA
```

Protein sequence:

```
MHHHHHHMTD KRKDGSGKLL YCSFCGKSQH EVRKLIAGPS VYICDECVDL CNDIIREEIK
EVAPHRERSA LPTPHEIRNH LDDYVIGQEQ AKKVLAVAVY NHYKRLRNGD TSNGVELGKS
NILLIGPTGS GKTLLAETLA RLLDVPFTMA DATTLTEAGY VGEDVENIIQ KLLQKCDYDV
QKAQRGIVYI DEIDKISRKS DNPSITRDVS GEGVQQALLK LIEGTVAAVP PQGGRKHPQQ
EFLQVDTSKI LFICGGAFAG LDKVISHRVE TGSGIGFGAT VKAKSDKASE GELLAQVEPE
DLIKFGLIPE FIGRLPVVAT LNELSEEALI QILKEPKNAL TKQYQALFNL EGVDLEFRDE
ALDAIAKKAM ARKTGARGLR SIVEAALLDT MYDLPSMEDV EKVVIDESVI DGQSKPLLIY
GKPEAQQASG E
```

7.1.1.2 Protein: $^{\Delta N}ClpX_6$

Brief Description: Pseudohexameric single-chain ClpX varaint

Plasmid: pAYC $^{\Delta N}ClpX_6$

Antibiotic resistance encoded in plasmid: Chloramphenicol

Number of amino acids: 2305

Protein Mass: 247192 Da

Calculated extinction coefficient: 82700 $M.cm^{-1}$

DNA sequence:

```
   1  TAATACGACT CACTATAGGG GAATTGTGAG CGGATAACAA TTCCCCTGTA GAAATAATTT
  61  TGTTTAACTT TAATAAGGAG ATATACCATG GCCGATTACA AAGATGACGA TGACAAAAGC
 121  TCTCATATGA GTGCGCTACC GACGCCGCAT GAAATTCGCA ACCACCTGGA CGATTACGTT
 181  ATCGGCCAGG AACAGGCGAA AAAAGTGCTG GCGGTCGCGG TATACAACCA TTACAAACGT
 241  CTGCGCAACG GCGATACCAG CAATGGCGTC GAGTTGGGCA AAAGTAACAT TCTGCTGATC
 301  GGTCCGACCG GTTCCGGTAA AACGCTGCTG GCTGAAACGC TGGCGCGCCT GCTGGATGTT
 361  CCGTTCACCA TGGCCGACGC GACTACACTG ACCGAAGCCG GTTATGTGGG TGAAGACGTT
 421  GAAAACATCA TTCAGAAGCT GTTGCAGAAA TGCGACTACG ATGTCCAGAA AGCACAGCGT
 481  GGTATTGTCT ACATCGATGA AATCGACAAG ATTTCTCGTA AGTCAGACAA CCCGTCCATT
 541  ACCCGAGACG TTTCCGGTGA AGGCGTACAG CAGGCACTGT TGAAACTGAT CGAAGGTACG
 601  GTAGCTGCTG TTCCACCGCA AGGTGGGCGT AAACATCCGC AGCAGGAATT CTTGCAGGTT
 661  GATACCTCTA AGATCCTGTT TATTTGTGGC GGTGCGTTTG CCGGTCTGGA TAAAGTGATT
 721  TCCCACCGTG TAGAAACCGG CTCCGGCATT GGTTTTGGCG CGACGGTAAA AGCGAAGTCC
 781  GACAAAGCAA GCGAAGGCGA GCTGCTGGCG CAGGTTGAAC CGGAAGATCT GATCAAGTTT
 841  GGTCTTATCC CTGAGTTTAT TGGTCGTCTG CCGGTTGTCG CAACGTTGAA TGAACTGAGC
 901  GAAGAAGCTC TGATTCAGAT CCTCAAAGAG CCGAAAAACG CCCTGACCAA GCAGTATCAG
 961  GCGCTGTTTA ATCTGGAAGG CGTGGATCTG GAATTCCGTG ACGAGGCGCT GGATGCTATC
1021  GCTAAGAAAG CGATGGCGCG TAAAACCGGT GCCCGTGGCC TGCGTTCCAT CGTAGAAGCC
1081  GCACTGCTCG ATACCATGTA CGATCTGCCG TCCATGGAAG ACGTCGAAAA AGTGGTTATC
1141  GACGAGTCGG TAATTGATGG TCAAAGCAAA CCGTTGCTGA TTTATGGCAA GCCGGAAGCG
1201  CAACAGGCAT CTGGTGAGGC CAGCGGGGCC GGCGGTTCTG AAGGTGGCGG CAGCGAAGGT
1261  GGTACCAGCG GGGCCACAAT GAGTGCGCTA CCGACGCCGC ATGAAATTCG CAACCACCTG
1321  GACGATTACG TTATCGGCCA GGAACAGGCG AAAAAAGTGC TGGCGGTCGC GGTATACAAC
1381  CATTACAAAC GTCTGCGCAA CGGCGATACC AGCAATGGCG TCGAGTTGGG CAAAAGTAAC
1441  ATTCTGCTGA TCGGTCCGAC CGGTTCCGGT AAAACGCTGC TGGCTGAAAC GCTGGCGCGC
1501  CTGCTGGATG TTCCGTTCAC CATGGCCGAC GCGACTACAC TGACCGAAGC CGGTTATGTG
1561  GGTGAAGACG TTGAAAACAT CATTCAGAAG CTGTTGCAGA AATGCGACTA CGATGTCCAG
1621  AAAGCACAGC GTGGTATTGT CTACATCGAT GAAATCGACA AGATTTCTCG TAAGTCAGAC
1681  AACCCGTCCA TTACCCGAGA CGTTTCCGGT GAAGGCGTAC AGCAGGCACT GTTGAAACTG
1741  ATCGAAGGTA CGGTAGCTGC TGTTCCACCG CAAGGTGGGC GTAAACATCC GCAGCAGGAA
1801  TTCTTGCAGG TTGATACCTC TAAGATCCTG TTTATTTGTG GCGGTGCGTT TGCCGGTCTG
1861  GATAAAGTGA TTTCCCACCG TGTAGAAACC GGCTCCGGCA TTGGTTTTGG CGCGACGGTA
1921  AAAGCGAAGT CCGACAAAGC AAGCGAAGGC GAGCTGCTGG CGCAGGTTGA ACCGGAAGAT
1981  CTGATCAAGT TTGGTCTTAT CCCTGAGTTT ATTGGTCGTC TGCCGGTTGT CGCAACGTTG
2041  AATGAACTGA GCGAAGAAGC TCTGATTCAG ATCCTCAAAG AGCCGAAAAA CGCCCTGACC
2101  AAGCAGTATC AGGCGCTGTT TAATCTGGAA GGCGTGGATC TGGAATTCCG TGACGAGGCG
2161  CTGGATGCTA TCGCTAAGAA AGCGATGGCG CGTAAAACCG GTGCCCGTGG CCTGCGTTCC
2221  ATCGTAGAAG CCGCACTGCT CGATACCATG TACGATCTGC CGTCCATGGA AGACGTCGAA
2281  AAAGTGGTTA TCGACGAGTC GGTAATTGAT GGTCAAAGCA AACCGTTGCT GATTTATGGC
2341  AAGCCGGAAG CGCAACAGGC ATCTGGTGAG GCCAGCGGGG CCGGCGGTTC TGAAGGTGGC
2401  GGCAGCGAAG GTGGCTCGAG CGGGGCCACA ATGAGTGCGC TACCGACGCC GCATGAAATT
2461  CGCAACCACC TGGACGATTA CGTTATCGGC CAGGAACAGG CGAAAAAAGT GCTGGCGGTC
2521  GCGGTATACA ACCATTACAA ACGTCTGCGC AACGGCGATA CCAGCAATGG CGTCGAGTTG
2581  GGCAAAAGTA ACATTCTGCT GATCGGTCCG ACCGGTTCCG GTAAAACGCT GCTGGCTGAA
2641  ACGCTGGCGC GCCTGCTGGA TGTTCCGTTC ACCATGGCCG ACGCGACTAC ACTGACCGAA
2701  GCCGGTTATG TGGGTGAAGA CGTTGAAAAC ATCATTCAGA AGCTGTTGCA GAAATGCGAC
```

```
2761  TACGATGTCC AGAAAGCACA GCGTGGTATT GTCTACATCG ATGAAATCGA CAAGATTTCT
2821  CGTAAGTCAG ACAACCCGTC CATTACCCGA GACGTTTCCG GTGAAGGCGT ACAGCAGGCA
2881  CTGTTGAAAC TGATCGAAGG TACGGTAGCT GCTGTTCCAC CGCAAGGTGG GCGTAAACAT
2941  CCGCAGCAGG AATTCTTGCA GGTTGATACC TCTAAGATCC TGTTTATTTG TGGCGGTGCG
3001  TTTGCCGGTC TGGATAAAGT GATTTCCCAC CGTGTAGAAA CCGGCTCCGG CATTGGTTTT
3061  GGCGCGACGG TAAAAGCGAA GTCCGACAAA GCAAGCGAAG GCGAGCTGCT GGCGCAGGTT
3121  GAACCGGAAG ATCTGATCAA GTTTGGTCTT ATCCCTGAGT TTATTGGTCG TCTGCCGGTT
3181  GTCGCAACGT TGAATGAACT GAGCGAAGAA GCTCTGATTC AGATCCTCAA AGAGCCGAAA
3241  AACGCCCTGA CCAAGCAGTA TCAGGCGCTG TTTAATCTGG AAGGCGTGGA TCTGGAATTC
3301  CGTGACGAGG CGCTGGATGC TATCGCTAAG AAAGCGATGG CGCGTAAAAC CGGTGCCCGT
3361  GGCCTGCGTT CCATCGTAGA AGCCGCACTG CTCGATACCA TGTACGATCT GCCGTCCATG
3421  GAAGACGTCG AAAAAGTGGT TATCGACGAG TCGGTAATTG ATGGTCAAAG CAAACCGTTG
3481  CTGATTTATG GCAAGCCGGA AGCGCAACAG GCATCTGGTG AGGCCAGCGG GGCCGGCGGT
3541  TCTGAAGGTG GCGGCAGCGA AGGTGGGTCG ACCGGGGCCA CAATGAGTGC GCTACCGACG
3601  CCGCATGAAA TTCGCAACCA CCTGGACGAT TACGTTATCG GCCAGGAACA GGCGAAAAAA
3661  GTGCTGGCGG TCGCGGTATA CAACCATTAC AAACGTCTGC GCAACGGCGA TACCAGCAAT
3721  GGCGTCGAGT TGGGCAAAAG TAACATTCTG CTGATCGGTC CGACCGGTTC CGGTAAAACG
3781  CTGCTGGCTG AAACGCTGGC GCGCCTGCTG GATGTTCCGT TCACCATGGC CGACGCGACT
3841  ACACTGACCG AAGCCGGTTA TGTGGGTGAA GACGTTGAAA ACATCATTCA GAAGCTGTTG
3901  CAGAAATGCG ACTACGATGT CCAGAAAGCA CAGCGTGGTA TTGTCTACAT CGATGAAATC
3961  GACAAGATTT CTCGTAAGTC AGACAACCCG TCCATTACCC GAGACGTTTC CGGTGAAGGC
4021  GTACAGCAGG CACTGTTGAA ACTGATCGAA GGTACGGTAG CTGCTGTTCC ACCGCAAGGT
4081  GGGCGTAAAC ATCCGCAGCA GGAATTCTTG CAGGTTGATA CCTCTAAGAT CCTGTTTATT
4141  TGTGGCGGTG CGTTTGCCGG TCTGGATAAA GTGATTTCCC ACCGTGTAGA AACCGGCTCC
4201  GGCATTGGTT TTGGCGCGAC GGTAAAAGCG AAGTCCGACA AAGCAAGCGA AGGCGAGCTG
4261  CTGGCGCAGG TTGAACCGGA AGATCTGATC AAGTTTGGTC TTATCCCTGA GTTTATTGGT
4321  CGTCTGCCGG TTGTCGCAAC GTTGAATGAA CTGAGCGAAG AAGCTCTGAT TCAGATCCTC
4381  AAAGAGCCGA AAAACGCCCT GACCAAGCAG TATCAGGCGC TGTTTAATCT GGAAGGCGTG
4441  GATCTGGAAT TCCGTGACGA GGCGCTGGAT GCTATCGCTA AGAAAGCGAT GGCGCGTAAA
4501  ACCGGTGCCC GTGGCCTGCG TTCCATCGTA GAAGCCGCAC TGCTCGATAC CATGTACGAT
4561  CTGCCGTCCA TGGAAGACGT CGAAAAAGTG GTTATCGACG AGTCGGTAAT TGATGGTCAA
4621  AGCAAACCGT TGCTGATTTA TGGCAAGCCG GAAGCGCAAC AGGCATCTGG TGAGGCCAGC
4681  GGGGCCGGCG GTTCTGAAGG TGGCGGCAGC GAAGGTGGCA CCTCTGCAGG CACAATGAGT
4741  GCGCTACCGA CGCCGCATGA AATTCGCAAC CACCTGGACG ATTACGTTAT CGGCCAGGAA
4801  CAGGCGAAAA AAGTGCTGGC GGTCGCGGTA TACAACCATT ACAAACGTCT GCGCAACGGC
4861  GATACCAGCA ATGGCGTCGA GTTGGGCAAA AGTAACATTC TGCTGATCGG TCCGACCGGT
4921  TCCGGTAAAA CGCTGCTGGC TGAAACGCTG GCGCGCCTGC TGGATGTTCC GTTCACCATG
4981  GCCGACGCGA CTACACTGAC CGAAGCCGGT TATGTGGGTG AAGACGTTGA AAACATCATT
5041  CAGAAGCTGT TGCAGAAATG CGACTACGAT GTCCAGAAAG CACAGCGTGG TATTGTCTAC
5101  ATCGATGAAA TCGACAAGAT TTCTCGTAAG TCAGACAACC CGTCCATTAC CCGAGACGTT
5161  TCCGGTGAAG GCGTACAGCA GGCACTGTTG AAACTGATCG AAGGTACGGT AGCTGCTGTT
5221  CCACCGCAAG GTGGGCGTAA ACATCCGCAG CAGGAATTCT TGCAGGTTGA TACCTCTAAG
5281  ATCCTGTTTA TTTGTGGCGG TGCGTTTGCC GGTCTGGATA AAGTGATTTC CCACCGTGTA
5341  GAAACCGGCT CCGGCATTGG TTTTGGCGCG ACGGTAAAAG CGAAGTCCGA CAAAGCAAGC
5401  GAAGGCGAGC TGCTGGCGCA GGTTGAACCG GAAGATCTGA TCAAGTTTGG TCTTATCCCT
5461  GAGTTTATTG GTCGTCTGCC GGTTGTCGCA ACGTTGAATG AACTGAGCGA AGAAGCTCTG
5521  ATTCAGATCC TCAAAGAGCC GAAAAACGCC CTGACCAAGC AGTATCAGGC GCTGTTTAAT
5581  CTGGAAGGCG TGGATCTGGA ATTCCGTGAC GAGGCGCTGG ATGCTATCGC TAAGAAAGCG
5641  ATGGCGCGTA AAACCGGTGC CCGTGGCCTG CGTTCCATCG TAGAAGCCGC ACTGCTCGAT
5701  ACCATGTACG ATCTGCCGTC CATGGAAGAC GTCGAAAAAG TGGTTATCGA CGAGTCGGTA
5761  ATTGATGGTC AAAGCAAACC GTTGCTGATT TATGGCAAGC CGGAAGCGCA ACAGGCATCT
5821  GGTGAGGCCA GCGGGGCCGG CGGTTCTGAA GGTGGCGGCA GCGAAGGTGG ATCCAGCGGG
5881  GCCACAATGA GTGCGCTACC GACGCCGCAT GAAATTCGCA ACCACCTGGA CGATTACGTT
5941  ATCGGCCAGG AACAGGCGAA AAAAGTGCTG GCGGTCGCGG TATACAACCA TTACAAACGT
6001  CTGCGCAACG GCGATACCAG CAATGGCGTC GAGTTGGGCA AAAGTAACAT TCTGCTGATC
6061  GGTCCGACCG GTTCCGGTAA AACGCTGCTG GCTGAAACGC TGGCGCGCCT GCTGGATGTT
6121  CCGTTCACCA TGGCCGACGC GACTACACTG ACCGAAGCCG GTTATGTGGG TGAAGACGTT
6181  GAAAACATCA TTCAGAAGCT GTTGCAGAAA TGCGACTACG ATGTCCAGAA AGCACAGCGT
6241  GGTATTGTCT ACATCGATGA AATCGACAAG ATTTCTCGTA AGTCAGACAA CCCGTCCATT
6301  ACCCGAGACG TTTCCGGTGA AGGCGTACAG CAGGCACTGT TGAAACTGAT CGAAGGTACG
6361  GTAGCTGCTG TTCCACCGCA AGGTGGGCGT AAACATCCGC AGCAGGAATT CTTGCAGGTT
6421  GATACCTCTA AGATCCTGTT TATTTGTGGC GGTGCGTTTG CCGGTCTGGA TAAAGTGATT
6481  TCCCACCGTG TAGAAACCGG CTCCGGCATT GGTTTTGGCG CGACGGTAAA AGCGAAGTCC
6541  GACAAAGCAA GCGAAGGCGA GCTGCTGGCG CAGGTTGAAC CGGAAGATCT GATCAAGTTT
6601  GGTCTTATCC CTGAGTTTAT TGGTCGTCTG CCGGTTGTCG CAACGTTGAA TGAACTGAGC
6661  GAAGAAGCTC TGATTCAGAT CCTCAAAGAG CCGAAAAACG CCCTGACCAA GCAGTATCAG
6721  GCGCTGTTTA ATCTGGAAGG CGTGGATCTG GAATTCCGTG ACGAGGCGCT GGATGCTATC
6781  GCTAAGAAAG CGATGGCGCG TAAAACCGGT GCCCGTGGCC TGCGTTCCAT CGTAGAAGCC
6841  GCACTGCTCG ATACCATGTA CGATCTGCCG TCCATGGAAG ACGTCGAAAA AGTGGTTATC
```

```
6901  GACGAGTCGG TAATTGATGG TCAAAGCAAA CCGTTGCTGA TTTATGGCAA GCCGGAAGCG
6961  CAACAGGCAT CTGGTGAATC TAGTCATCAC CATCATCACC ACTAAAAGCT TGCGGCCGCA
7021  CTCGAGTCTG GTAAAGAAAC CGCTGCTGCG AAATTTGAAC GCCAGCACAT GGACTCGTCT
7081  ACTAGCGCAG CTTAATTAAC CTAGGCTGCT GCCACCGCTG AGCAATAACT AGCATAACCC
7141  CTTGGGGCCT CTAAACGGGT CTTG
```

Protein sequence:

```
MADYKDDDDK SSHMSALPTP HEIRNHLDDY VIGQEQAKKV LAVAVYNHYK RLRNGDTSNG
VELGKSNILL IGPTGSGKTL LAETLARLLD VPFTMADATT LTEAGYVGED VENIIQKLLQ
KCDYDVQKAQ RGIVYIDEID KISRKSDNPS ITRDVSGEGV QQALLKLIEG TVAAVPPQGG
RKHPQQEFLQ VDTSKILFIC GGAFAGLDKV ISHRVETGSG IGFGATVKAK SDKASEGELL
AQVEPEDLIK FGLIPEFIGR LPVVATLNEL SEEALIQILK EPKNALTKQY QALFNLEGVD
LEFRDEALDA IAKKAMARKT GARGLRSIVE AALLDTMYDL PSMEDVEKVV IDESVIDGQS
KPLLIYGKPE AQQASGEASG AGGSEGGGSE GGTSGATMSA LPTPHEIRNH LDDYVIGQEQ
AKKVLAVAVY NHYKRLRNGD TSNGVELGKS NILLIGPTGS GKTLLAETLA RLLDVPFTMA
DATTLTEAGY VGEDVENIIQ KLLQKCDYDV QKAQRGIVYI DEIDKISRKS DNPSITRDVS
GEGVQQALLK LIEGTVAAVP PQGGRKHPQQ EFLQVDTSKI LFICGGAFAG LDKVISHRVE
TGSGIGFGAT VKAKSDKASE GELLAQVEPE DLIKFGLIPE FIGRLPVVAT LNELSEEALI
QILKEPKNAL TKQYQALFNL EGVDLEFRDE ALDAIAKKAM ARKTGARGLR SIVEAALLDT
MYDLPSMEDV EKVVIDESVI DGQSKPLLIY GKPEAQQASG EASGAGGSEG GGSEGGSSGA
TMSALPTPHE IRNHLDDYVI GQEQAKKVLA VAVYNHYKRL RNGDTSNGVE LGKSNILLIG
PTGSGKTLLA ETLARLLDVP FTMADATTLT EAGYVGEDVE NIIQKLLQKC DYDVQKAQRG
IVYIDEIDKI SRKSDNPSIT RDVSGEGVQQ ALLKLIEGTV AAVPPQGGRK HPQQEFLQVD
TSKILFICGG AFAGLDKVIS HRVETGSGIG FGATVKAKSD KASEGELLAQ VEPEDLIKFG
LIPEFIGRLP VVATLNELSE EALIQILKEP KNALTKQYQA LFNLEGVDLE FRDEALDAIA
KKAMARKTGA RGLRSIVEAA LLDTMYDLPS MEDVEKVVID ESVIDGQSKP LLIYGKPEAQ
QASGEASGAG GSEGGGSEGG STGATMSALP TPHEIRNHLD DYVIGQEQAK KVLAVAVYNH
YKRLRNGDTS NGVELGKSNI LLIGPTGSGK TLLAETLARL LDVPFTMADA TTLTEAGYVG
EDVENIIQKL LQKCDYDVQK AQRGIVYIDE IDKISRKSDN PSITRDVSGE GVQQALLKLI
EGTVAAVPPQ GGRKHPQQEF LQVDTSKILF ICGGAFAGLD KVISHRVETG SGIGFGATVK
AKSDKASEGE LLAQVEPEDL IKFGLIPEFI GRLPVVATLN ELSEEALIQI LKEPKNALTK
QYQALFNLEG VDLEFRDEAL DAIAKKAMAR KTGARGLRSI VEAALLDTMY DLPSMEDVEK
VVIDESVIDG QSKPLLIYGK PEAQQASGEA SGAGGSEGGG SEGGTSAGTM SALPTPHEIR
NHLDDYVIGQ EQAKKVLAVA VYNHYKRLRN GDTSNGVELG KSNILLIGPT GSGKTLLAET
LARLLDVPFT MADATTLTEA GYVGEDVENI IQKLLQKCDY DVQKAQRGIV YIDEIDKISR
KSDNPSITRD VSGEGVQQAL LKLIEGTVAA VPPQGGRKHP QQEFLQVDTS KILFICGGAF
AGLDKVISHR VETGSGIGFG ATVKAKSDKA SEGELLAQVE PEDLIKFGLI PEFIGRLPVV
ATLNELSEEA LIQILKEPKN ALTKQYQALF NLEGVDLEFR DEALDAIAKK AMARKTGARG
LRSIVEAALL DTMYDLPSME DVEKVVIDES VIDGQSKPLL IYGKPEAQQA SGEASGAGGS
EGGGSEGGSS GATMSALPTP HEIRNHLDDY VIGQEQAKKV LAVAVYNHYK RLRNGDTSNG
VELGKSNILL IGPTGSGKTL LAETLARLLD VPFTMADATT LTEAGYVGED VENIIQKLLQ
KCDYDVQKAQ RGIVYIDEID KISRKSDNPS ITRDVSGEGV QQALLKLIEG TVAAVPPQGG
RKHPQQEFLQ VDTSKILFIC GGAFAGLDKV ISHRVETGSG IGFGATVKAK SDKASEGELL
AQVEPEDLIK FGLIPEFIGR LPVVATLNEL SEEALIQILK EPKNALTKQY QALFNLEGVD
LEFRDEALDA IAKKAMARKT GARGLRSIVE AALLDTMYDL PSMEDVEKVV IDESVIDGQS
KPLLIYGKPE AQQASGESSH HHHHH
```

7.1.1.3 Protein: ClpX$^{\Delta RKH}$

Brief Description: Pseudo wild-type ClpX with N-terminal His tag and R228A, K229A, H230A mutations.

Plasmid: pET23ClpXRKH

Antibiotic resistance encoded in plasmid: Ampicillin

Number of amino acids: 431

Protein Mass: 47101.7 Da

Calculated extinction coefficient: 16765 M.cm^{-1}

DNA sequence:

```
   1  ATGCATCATC ATCATCATCA TATGACCGAT AAACGTAAAG ATGGCAGCGG CAAACTGCTG
  61  TATTGCAGCT TTTGCGGCAA AAGCCAGCAT GAAGTGCGTA AACTGATTGC GGGCCCGAGC
 121  GTGTATATTT GCGATGAATG CGTGGATCTG TGCAACGATA TTATTCGTGA AGAAATTAAA
 181  GAAGTGGCGC CGCATCGTGA ACGTAGCGCG CTGCCGACCC CGCATGAAAT TCGTAACCAT
 241  CTGGATGATT ATGTGATTGG CCAGGAACAG GCGAAAAAAG TGCTGGCGGT GGCGGTGTAT
 301  AACCATTATA AACGTCTGCG TAACGGCGAT ACCAGCAACG GCGTGGAACT GGGCAAAAGC
 361  AACATTCTGC TGATTGGCCC GACCGGCAGC GGCAAAACCC TGCTGGCGGA AACCCTGGCG
 421  CGTCTGCTGG ATGTGCCGTT TACCATGGCG GATGCGACCA CCCTGACCGA AGCGGGCTAT
 481  GTGGGCGAAG ATGTGGAAAA CATTATTCAG AAACTGCTGC AGAAATGCGA TTATGATGTG
 541  CAGAAAGCGC AGCGTGGCAT TGTGTATATT GATGAAATTG ATAAAATTAG CCGTAAAAGC
 601  GATAACCCGA GCATTACCCG TGATGTGAGC GGCGAAGGCG TGCAGCAGGC GCTGCTGAAA
 661  CTGATTGAAG GCACCGTGGC GGCGGTGCCG CCGCAGGGCG GCGCGGCGGC GCCGCAGCAG
 721  GAATTTCTGC AGGTGGATAC CAGCAAAATT CTGTTTATTT GCGGCGGCGC GTTTGCGGGC
 781  CTGGATAAAG TGATTAGCCA TCGTGTGGAA ACCGGCAGCG GCATTGGCTT TGGCGCGACC
 841  GTGAAAGCGA AAAGCGATAA AGCGAGCGAA GGCGAACTGC TGGCGCAGGT GGAACCGGAA
 901  GATCTGATTA AATTTGGCCT GATTCCGGAA TTTATTGGCC GTCTGCCGGT GGTGGCGACC
 961  CTGAACGAAC TGAGCGAAGA AGCGCTGATT CAGATTCTGA AAGAACCGAA AAACGCGCTG
1021  ACCAAACAGT ATCAGGCGCT GTTTAACCTG GAAGGCGTGG ATCTGGAATT TCGTGATGAA
1081  GCGCTGGATG CGATTGCGAA AAAAGCGATG GCGCGTAAAA CCGGCGCGCG TGGCCTGCGT
1141  AGCATTGTGG AAGCGGCGCT GCTGGATACC ATGTATGATC TGCCGAGCAT GGAAGATGTG
1201  GAAAAAGTGG TGATTGATGA AAGCGTGATT GATGGCCAGA GCAAACCGCT GCTGATTTAT
1261  GGCAAACCGG AAGCGCAGCA GGCGAGCGGC GAA
```

Protein sequence:

```
MHHHHHHMTD KRKDGSGKLL YCSFCGKSQH EVRKLIAGPS VYICDECVDL CNDIIREEIK
EVAPHRERSA LPTPHEIRNH LDDYVIGQEQ AKKVLAVAVY NHYKRLRNGD TSNGVELGKS
NILLIGPTGS GKTLLAETLA RLLDVPFTMA DATTLTEAGY VGEDVENIIQ KLLQKCDYDV
QKAQRGIVYI DEIDKISRKS DNPSITRDVS GEGVQQALLK LIEGTVAAVP PQGGAAAPQQ
EFLQVDTSKI LFICGGAFAG LDKVISHRVE TGSGIGFGAT VKAKSDKASE GELLAQVEPE
DLIKFGLIPE FIGRLPVVAT LNELSEEALI QILKEPKNAL TKQYQALFNL EGVDLEFRDE
ALDAIAKKAM ARKTGARGLR SIVEAALLDT MYDLPSMEDV EKVVIDESVI DGQSKPLLIY
GKPEAQQASG E
```

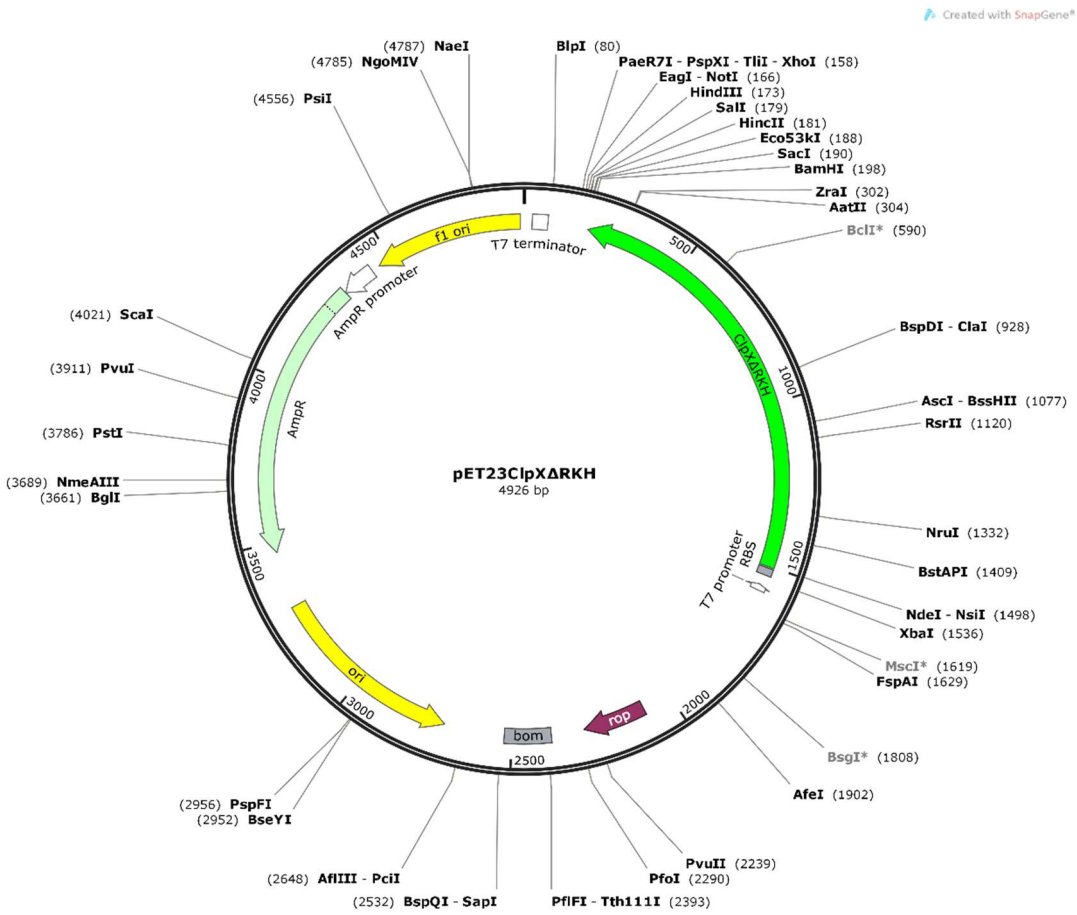


Figure 7.1: Plasmid map of pET23ClpXΔRKH.

7.1.1.4 Protein: ClpX$^{\Delta RKH}$-TC

Brief Description: Pseudo wild-type ClpX with N-terminal His-tag upstream of a thrombin cleavage site and R228A, K229A, H230A mutations.

Plasmid: pET23 ClpXΔRKH-TC

Antibiotic resistance encoded in plasmid: Ampicillin

Number of amino acids: 444

Protein Mass: 48311.0 Da

Calculated extinction coefficient: 16765 M.cm^{-1}

DNA sequence:

```
   1  ATGGGCAGCA GCCATCATCA TCATCATCAT AGCAGCGGCC TGGTGCCGCG CGGCAGCCAT
  61  GGCTCAGGCA AATTGCTGTA TTGCTCTTTT TGCGGCAAAA GCCAGCATGA AGTGCGCAA
 121  CTGATTGCCG GTCCATCCGT GTATATCTGC GACGAATGTG TTGATTTATG TAACGACATC
 181  ATTCGCGAAG AGATTAAAGA AGTTGCACCG CATCGTGAAC GCAGTGCGCT ACCGACGCCG
 241  CATGAAATTC GCAACCACCT GGACGATTAC GTTATCGGCC AGGAACAGGC GAAAAAAGTG
 301  CTGGCGGTCG CGGTATACAA CCATTACAAA CGTCTGCGCA ACGGCGATAC CAGCAATGGC
 361  GTCGAGTTGG GCAAAAGTAA CATTCTGCTG ATCGGTCCGA CCGGTTCCGG TAAAACGCTG
 421  CTGGCTGAAA CGCTGGCGCG CCTGCTGGAT GTTCCGTTCA CCATGGCCGA CGCGACTACA
 481  CTGACCGAAG CCGGTTATGT GGGTGAAGAC GTTGAAAACA TCATTCAGAA GCTGTTGCAG
 541  AAATGCGACT ACGATGTCCA GAAAGCACAG CGTGGTATTG TCTACATCGA TGAAATCGAC
 601  AAGATTTCTC GTAAGTCAGA CAACCCGTCC ATTACCCGAG ACGTTTCCGG TGAAGGCGTA
 661  CAGCAGGCAC TGTTGAAACT GATCGAAGGT ACGGTAGCTG CTGTTCCACC GCAAGGTGGG
 721  CGTAAACATC CGCAGCAGGA ATTCTTGCAG GTTGATACCT CTAAGATCCT GTTTATTTGT
 781  GGCGGTGCGT TTGCCGGTCT GGATAAAGTG ATTTCCCACC GTGTAGAAAC CGGCTCCGGC
 841  ATTGGTTTTG GCGCGACGGT AAAAGCGAAG TCCGACAAAG CAAGCGAAGG CGAGCTGCTG
 901  GCGCAGGTTG AACCGGAAGA TCTGATCAAG TTTGGTCTTA TCCCTGAGTT TATTGGTCGT
 961  CTGCCGGTTG TCGCAACGTT GAATGAACTG AGCGAAGAAG CTCTGATTCA GATCCTCAAA
1021  GAGCCGAAAA ACGCCCTGAC CAAGCAGTAT CAGGCGCTGT TTAATCTGGA AGGCGTGGAT
1081  CTGGAATTCC GTGACGAGGC GCTGGATGCT ATCGCTAAGA AAGCGATGGC GCGTAAAACC
1141  GGTGCCCGTG GCCTGCGTTC CATCGTAGAA GCCGCACTGC TCGATACCAT GTACGATCTG
1201  CCGTCCATGG AAGACGTCGA AAAAGTGGTT ATCGACGAGT CGGTAATTGA TGGTCAAAGC
1261  AAACCGTTGC TGATTTATGG CAAGCCGGAA GCGCAACAGG CATCTGGTGA ATAAGGATCC
1321  CATCACCA
```

Protein sequence:

```
MGSSHHHHHH SSGLVPRGSH MTDKRKDGSG KLLYCSFCGK SQHEVRKLIA GPSVYICDEC
VDLCNDIIRE EIKEVAPHRE RSALPTPHEI RNHLDDYVIG QEQAKKVLAV AVYNHYKRLR
NGDTSNGVEL GKSNILLIGP TGSGKTLLAE TLARLLDVPF TMADATTLTE AGYVGEDVEN
IIQKLLQKCD YDVQKAQRGI VYIDEIDKIS RKSDNPSITR DVSGEGVQQA LLKLIEGTVA
AVPPQGGAAA PQQEFLQVDT SKILFICGGA FAGLDKVISH RVETGSGIGF GATVKAKSDK
ASEGELLAQV EPEDLIKFGL IPEFIGRLPV VATLNELSEE ALIQILKEPK NALTKQYQAL
FNLEGVDLEF RDEALDAIAK KAMARKTGAR GLRSIVEAAL LDTMYDLPSM EDVEKVVIDE
SVIDGQSKPL LIYGKPEAQQ ASGE
```

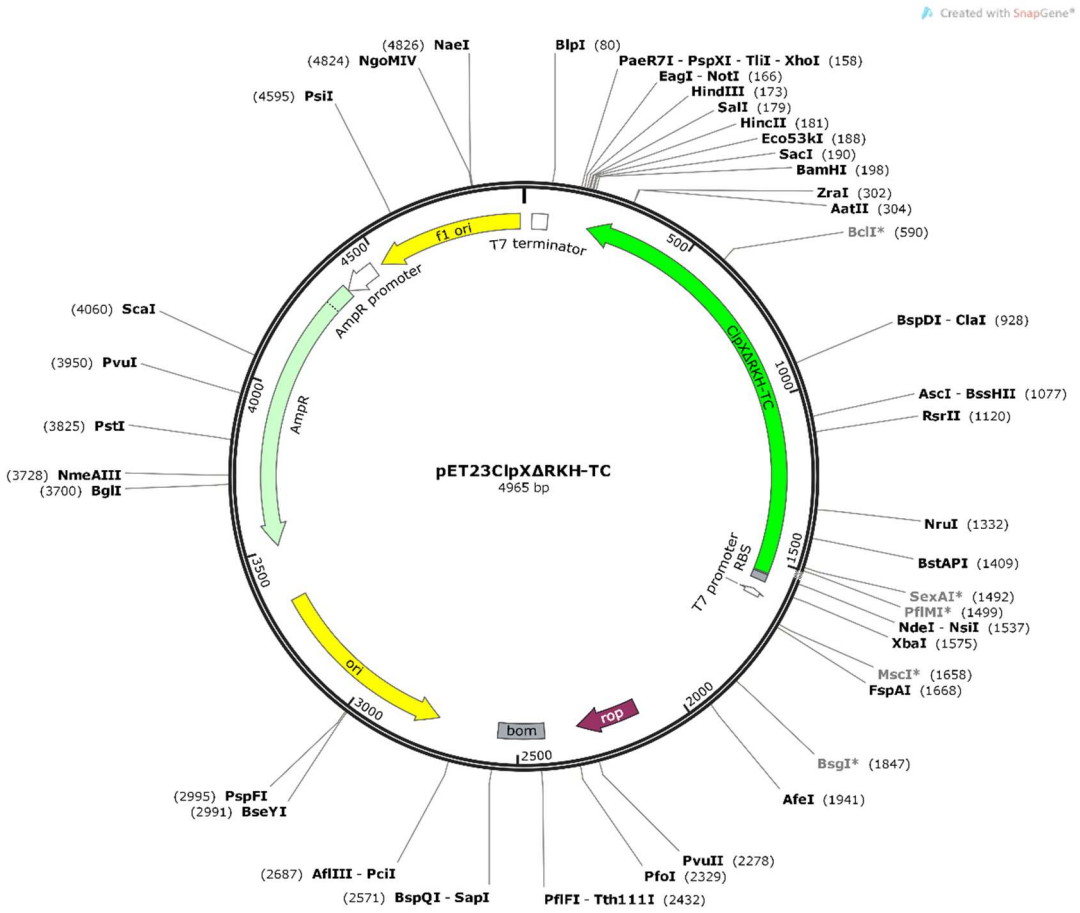


Figure 7.2: Plasmid map of pET23ClpXΔRKH-TC

7.1.2 ClpP

7.1.2.1 Protein: ClpP

Brief Description: Pseudo wild-type ClpP with C-terminal His tag

Plasmid: pET23ClpP

Antibiotic resistance encoded in plasmid: Ampicillin

Number of amino acids: 202

Protein Mass: 22760.1 Da

Calculated extinction coefficient: 9065 $M.cm^{-1}$

DNA sequence:

```
  1  ATATGGCGCT GGTGCCGATG GTCATTGAAC AGACCTCACG CGGTGAGCGC TCTTTTGATA
 61  TCTATTCTCG TCTACTTAAG GAACGCGTCA TTTTTCTGAC TGGCCAGGTT GAAGACCACA
121  TGGCTAACCT GATTGTGGCG CAGATGCTGT TCCTGGAAGC GGAAAACCCA GAAAAAGATA
181  TCTATCTGTA CATTAACTCC CCAGGCGGGG TGATCACTGC CGGGATGTCT ATCTATGACA
241  CCATGCAGTT TATCAAGCCT GATGTCAGCA CCATCTGTAT GGGCCAGGCG GCCTCGATGG
301  GCGCTTTCTT GCTGACCGCA GGGGCAAAAG GTAAACGTTT TTGCCTGCCG AATTCGCGCG
361  TGATGATTCA CCAACCGTTG GGCGGCTACC AGGGCCAGGC GACCGATATC GAAATTCATG
421  CCCGTGAAAT TCTGAAAGTT AAAGGGCGCA TGAATGAACT TATGGCGCTT CATACGGGTC
481  AATCATTAGA ACAGATTGAA CGTGATACCG AGCGCGATCG CTTCCTTTCC GCCCCTGAAG
541  CGGTGGAATA CGGTCTGGTC GATTCGATTC TGACCCATCG TAATTCGCGA CACCATCACC
601  ATCACCATTA ACTCGAG
```

Protein sequence:

```
MALVPMVIEQ TSRGERSFDI YSRLLKERVI FLTGQVEDHM ANLIVAQMLF LEAENPEKDI
YLYINSPGGV ITAGMSIYDT MQFIKPDVST ICMGQAASMG AFLLTAGAKG KRFCLPNSRV
MIHQPLGGYQ GQATDIEIHA REILKVKGRM NELMALHTGQ SLEQIERDTE RDRFLSAPEA
VEYGLVDSIL THRNSRHHHH HH
```

7.1.3 Substrates

7.1.3.1 Protein: Im9-ssrA

Brief Description: E. coli immunity protein Im9 with N-terminal His-tag and C-terminal ssrA tag.

Number of amino acids: 107

Protein Mass: 12012.1 Da

Calculated extinction coefficient: 11460 $M.cm^{-1}$

Protein sequence:

```
MEHHHHHHME LKHSISDYTE AEFLQLVTTI CNADTSSEEE LVKLVTHFEE MTEHPSGSDL
IYYPKEGDDD SPSGIVNTVK QWRAANGKSG FKQGLEAAND ENYALAA
```

7.1.3.2 Protein: λO-Arc-His$_6$

Brief Description: *E. coli*

Plasmid:pET23λOArc-His6

Antibiotic resistance encoded in plasmid: Ampicillin

Number of amino acids: 73

Protein Mass: 8555.8 Da

Calculated extinction coefficient: 6990 M.cm^{-1}

DNA sequence:

```
      1  ATGACCAACA CCGCGAAAAT TCTGAACTTT GGCCGCGCGA GCATGAAAGG CATGAGCAAA
     61  ATGCCGCAGT TTAACCTGCG CTGGCCGCGC GAAGTGCTGG ATCTGGTGCG CAAAGTGGCG
    121  GAAGAAAACG GCCGCAGCGT GAACAGCGAA ATTTATCAGC GCGTGATGGA AAGCTTTAAA
    181  AAAGAAGGCC GCATTGGCGC GCATCATCAT CATCATCAT
```

Protein sequence:

```
MTNTAKILNF GRASMKGMSK MPQFNLRWPR EVLDLVRKVA EENGRSVNSE IYQRVMESFK
KEGRIGAHHH HHH
```

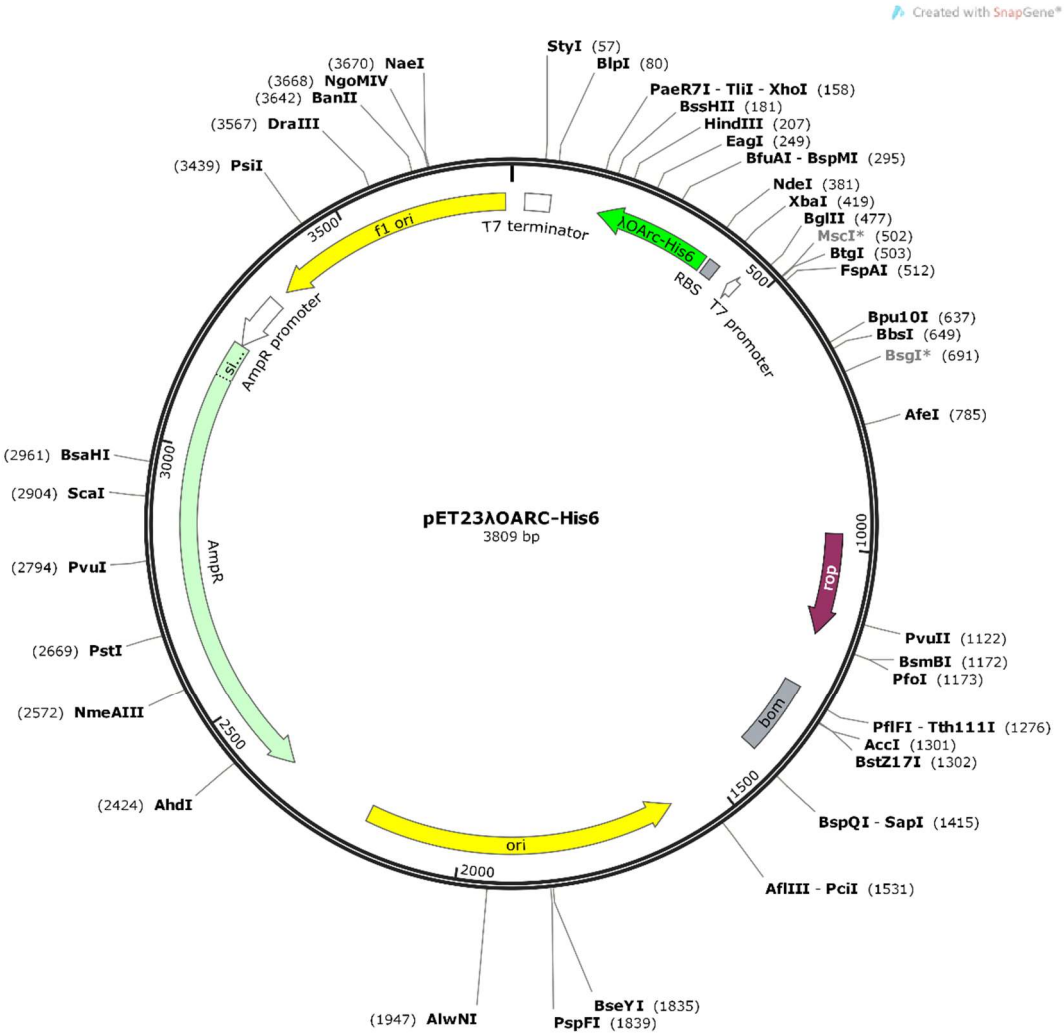


Figure 7.3: Plasmid map of pET23λOARC-His6.

7.1.3.3 Protein: λO-(T 1-16)-TEV-E9

Brief Description: *E. coli* E9 colicin with N-terminal λO-tag and residues 1-16 of the translocation domain located upstream of a Tobacco etch virus (TEV) cleavage site.

Plasmid: pET23λO(T1-16)TEVE9

Antibiotic resistance encoded in plasmid: Ampicillin

Number of amino acids: 170

Protein Mass: 18882.1 Da

Calculated extinction coefficient: 1690 $M.cm^{-1}$

DNA sequence:

```
  1  ATATGACCAA TACCGCGAAA ATTCTGAATT TTGGCCGCAT GAGCGGTGGA GATGGACGCG
 61  GCCATAACAC GGGCGCGCAT AGCACAGAAA ACCTGTATTT CCAGGGCGCC ATGGAGAGTA
121  AACGGAATAA GCCAGGGAAG GCGACAGGTA AAGGTAAACC AGTTGGTGAT AAATGGCTGG
181  ATGATGCAGG TAAAGATTCA GGAGCGCCAA TTCCAGATCG CATTGCTGAT AAGTTGCGTG
241  ATAAAGAATT TAAAAGCTTC GACGATTTTC GGAAGGCTGT ATGGGAAGAG GTGTCGAAAG
301  ATCCTGAGCT TAGTAAAAAT TTAAACCCAA GCAATAAGTC TAGTGTTTCA AAAGGTTATT
361  CTCCGTTTAC TCCAAAGAAT CAACAGGTCG GAGGGAGAAA AGTCTATGAA CTTCATCATG
421  ACAAGCCAAT TAGTCAAGGT GGTGAGGTTT ATGACATGGA TAATATCCGA GTGACTACAC
481  CTAAGCGACA TATCGATATT CACCGAGGTA AGTAAAATGG AACTGAAGCA TAGCATTAGT
541  GATTATACAG AAGCTGAATT TTTACAGCTT GTAACAACAA TTTGTAATGC GGACACTTCC
601  AGTGAAGAAG AACTGGTTAA ATTGGTTACA CACTTTGAGG AAATGACTGA GCACCCTAGT
661  GGTAGTGATT TAATATATTA CCCAAAAGAA GGTGATGATG ACTCACCTTC AGGTATTGTA
721  AACACAGTAA AACAATGGCG AGCCGCTAAC GGTAAGTCAG GATTTAAACA GGGCCTCGAG
781  CACCACCACC ACCACCACTG
```

Protein sequence:

```
MTNTAKILNF GRMSGGDGRG HNTGAHSTEN LYFQGAMESK RNKPGKATGK GKPVGDKWLD
DAGKDSGAPI PDRIADKLRD KEFKSFDDFR KAVWEEVSKD PELSKNLNPS NKSSVSKGYS
PFTPKNQQVG GRKVYELHHD KPISQGGEVY DMDNIRVTTP KRHIDIHRGK
```

This is co-expressed with Im9 from the same plasmid:

```
MELKHSISDY TEAEFLQLVT TICNADTSSE EELVKLVTHF EEMTEHPSGS DLIYYPKEGD
DDSPSGIVNT VKQWRAANGK SGFKQGLEHH HHHH
```

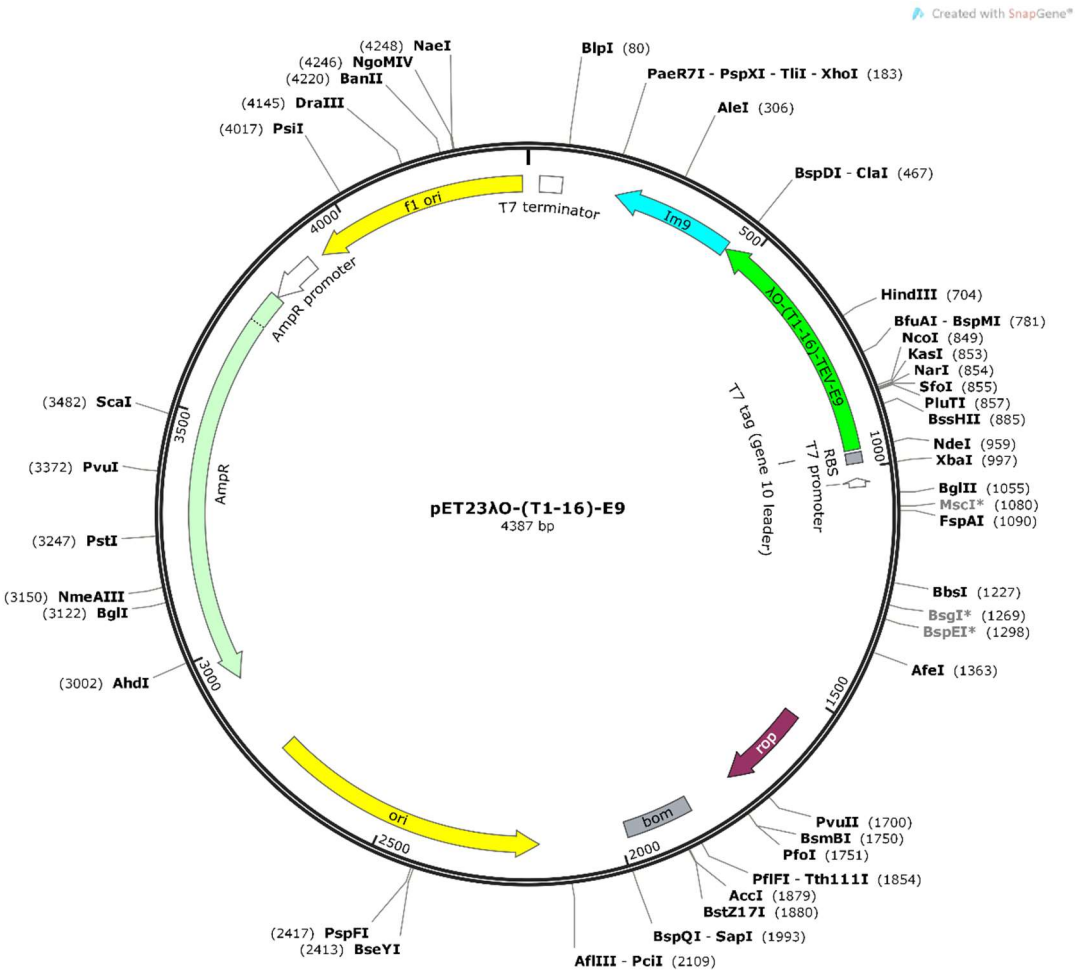


Figure 7.4: Plasmid map of pET23λO-(T1-16)-E9.

7.1.3.4 Protein: $(I27)_3$-ssrA

Brief description: Titin I27 concatamer for use in AFM and ClpXP degradation experiments. With an N-terminal hexahistidine sequence and a C-terminal ssrA tag.

cleavage site.

Plasmid: pET23His6-I27-ssrA

Antibiotic resistance encoded in plasmid: Ampicillin

Number of amino acids: 293

Protein Mass: 31960.3 Da

Calculated extinction coefficient: 22460 $M.cm^{-1}$

DNA sequence:

```
  1  ATGCTAATAG AAGTGGAAAA GCCTCTGTAC GGAGTAGAGG TGTTTGTTGG TGAAACAGCC
 61   CACTTTGAAA TTGAACTTTC TGAACCTGAT GTTCACGGCC AGTGGAAGCT GAAAGGACAG
121   CCTTTGACAG CTTCCCCTGA CTCTGAAATC ATTGAGGATG GAAAGAAGCA TATTCTGATC
181   CTTCATAACT CTCAGCTGGG TATGACAGGA GAGGTTTCCT TCCAGGCTGC TAATGCCAAA
241   TCTGCAGCCA ATCTGAAAGT GAAAGAACTA GTAGAGGCTC GACTAATAGA AGTGGAAAAG
301   CCTCTGTACG GAGTAGAGGT GTTTGTTGGT GAAACAGCCC ACTTTGAAAT TGAACTTTCT
361   GAACCTGATG TTCACGGCCA GTGGAAGCTG AAAGGACAGC CTTTGACAGC TTCCCCTGAC
421   TCTGAAATCA TTGAGGATGG AAAGAAGCAT ATTCTGATCC TTCATAACTC TCAGCTGGGT
481   ATGACAGGAG AGGTTTCCTT CCAGGCTGCT AATGCCAAAT CTGCAGCCAA TCTGAAAGTG
541   AAAGAATTGC TTATCGAAGC GCGCCTAATA GAAGTGGAAA AGCCTCTGTA CGGAGTAGAG
601   GTGTTTGTTG GTGAAACAGC CCACTTTGAA ATTGAACTTT CTGAACCTGA TGTTCACGGC
661   CAGTGGAAGC TGAAAGGACA GCCTTTGACA GCTTCCCCTG ACTCTGAAAT CATTGAGGAT
721   GGAAAGAAGC ATATTCTGAT CCTTCATAAC TCTCAGCTGG GTATGACAGG AGAGGTTTCC
781   TTCCAGGCTG CTAATGCCAA ATCTGCAGCC AATCTGAAAG TGAAAGAATT GCTGTGTGGA
841   GCTCAGGCGG CTAACGACGA AAACTACGCG CTGGCGGCTT AA
```

Protein sequence:

```
MLIEVEKPLY GVEVFVGETA HFEIELSEPD VHGQWKLKGQ PLTASPDSEI IEDGKKHILI
LHNSQLGMTG EVSFQAANAK SAANLKVKEL VEARLIEVEK PLYGVEVFVG ETAHFEIELS
EPDVHGQWKL KGQPLTASPD SEIIEDGKKH ILILHNSQLG MTGEVSFQAA NAKSAANLKV
KELLIEARLI EVEKPLYGVE VFVGETAHFE IELSEPDVHG QWKLKGQPLT ASPDSEIIED
GKKHILILHN SQLGMTGEVS FQAANAKSAA NLKVKELLCG AQAANDENYA LAA
```

Figure 7.5: Plasmid map of pET23His6-I27-ssrA.

7.1.3.5 Protein: I27-Q30-(I27)$_2$-ssrA

Brief description: Titin I27 concatamer for use in AFM and ClpXP degradation experiments with a Q30 section inserted between the second and third I27 domain. With an N-terminal hexahistidine sequence and a C-terminal ssrA tag.

cleavage site.

Plasmid: pET23His6-Q30-ssrA

Antibiotic resistance encoded in plasmid: Ampicillin

Number of amino acids: 326

Protein Mass: 36145.7 Da

Calculated extinction coefficient: 22460 M.cm^{-1}

DNA sequence:

```
  1  ATGCTAATAG AAGTGGAAAA GCCTCTGTAC GGAGTAGAGG TGTTTGTTGG TGAAACAGCC
 61  CACTTTGAAA TTGAACTTTC TGAACCTGAT GTTCACGGCC AGTGGAAGCT GAAAGGACAG
121  CCTTTGACAG CTTCCCCTGA CTCTGAAATC ATTGAGGATG GAAAGAAGCA TATTCTGATC
181  CTTCATAACT CTCAGCTGGG TATGACAGGA GAGGTTTCCT TCCAGGCTGC TAATGCCAAA
241  TCTGCAGCCA ATCTGAAAGT GAAAGAACTA GTAGAACAGC AACAGCAACA ACAGCAACAG
301  CAACAACAGC AACAGCAACA ACAGCAACAG CAACAACAGC AACAGCAACA ACAGCAACAG
361  CAACAACTAG TAGAGGCTCG ACTAATAGAA GTGGAAAAGC CTCTGTACGG AGTAGAGGTG
421  TTTGTTGGTG AAACAGCCCA CTTTGAAATT GAACTTTCTG AACCTGATGT TCACGGCCAG
481  TGGAAGCTGA AAGGACAGCC TTTGACAGCT TCCCCTGACT CTGAAATCAT TGAGGATGGA
541  AAGAAGCATA TTCTGATCCT TCATAACTCT CAGCTGGGTA TGACAGGAGA GGTTTCCTTC
601  CAGGCTGCTA ATGCCAAATC TGCAGCCAAT CTGAAAGTGA AAGAATTGCT TATCGAAGCG
661  CGCCTAATAG AAGTGGAAAA GCCTCTGTAC GGAGTAGAGG TGTTTGTTGG TGAAACAGCC
721  CACTTTGAAA TTGAACTTTC TGAACCTGAT GTTCACGGCC AGTGGAAGCT GAAAGGACAG
781  CCTTTGACAG CTTCCCCTGA CTCTGAAATC ATTGAGGATG GAAAGAAGCA TATTCTGATC
841  CTTCATAACT CTCAGCTGGG TATGACAGGA GAGGTTTCCT TCCAGGCTGC TAATGCCAAA
901  TCTGCAGCCA ATCTGAAAGT GAAAGAATTG CTGTGTGGAG CTCAGGCGGC TAACGACGAA
961  AACTACGCGC TGGCGGCTTA A
```

Protein sequence:

```
 MLIEVEKPLY GVEVFVGETA HFEIELSEPD VHGQWKLKGQ PLTASPDSEI IEDGKKHILI
LHNSQLGMTG EVSFQAANAK SAANLKVKEL VEQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ
QQLVEARLIE VEKPLYGVEV FVGETAHFEI ELSEPDVHGQ WKLKGQPLTA SPDSEIIEDG
KKHILILHNS QLGMTGEVSF QAANAKSAAN LKVKELLIEA RLIEVEKPLY GVEVFVGETA
HFEIELSEPD VHGQWKLKGQ PLTASPDSEI IEDGKKHILI LHNSQLGMTG EVSFQAANAK
SAANLKVKEL LCGAQAANDE NYALAA
```

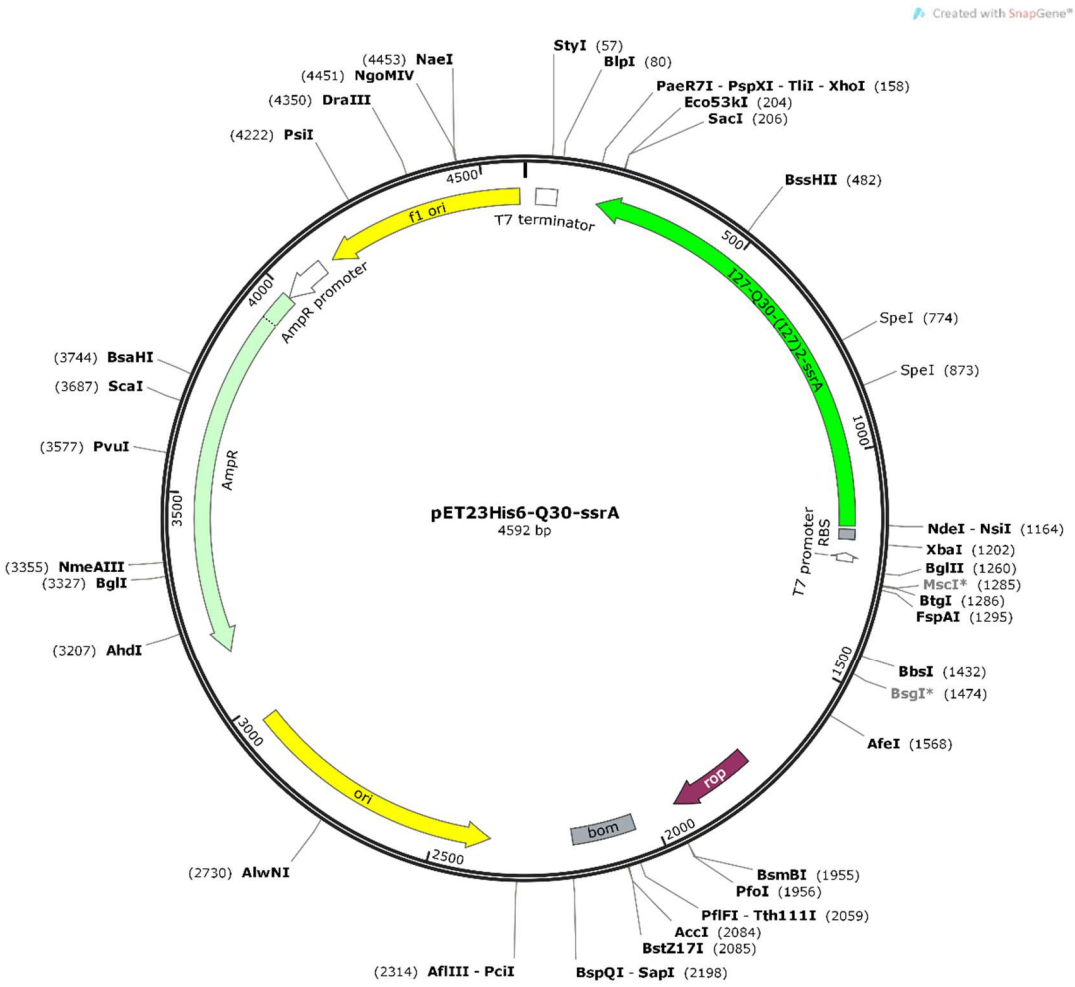


Figure 7.6: Plasmid map of pET23His6-Q30-ssrA.

7.2 PCR and Quikchange Primers

Table 7.1: List of mutagenic PCR primers used in the creation of the I27 concatamer and their purpose.

Primer Sequence	Purpose
5′ ATCCTATGCCATATGCTAATAGAAGTGGAAA AGCCTCT 3′	Add NdeI restriction site to I27 cassette 2 5' end to produce cassette 1
5′ ATCGTAGCAGAGCTCGCGCGCACTAGTTCTTTCACTTT CAGATTGGCTGC 3′	Add MCS (SpeI, BssHII and SacI) to 3' end of cassette 2 to produce cassette 1
5′ GCTACTCATGGCGCGCCTAATAGAAGTGGAAAAGCCTC T 3′	Add BssHII restriction site to 5' end of cassette 2 to produce cassette 3
5′ TGATGCTAGTGAGCTCCACACAGCAATTCTTTCACTTT CAGATTGGC 3′	Add SacI to 3' end of cassette 2 to produce cassette 3

Table 7.2: List of mutagenic PCR primers used in the addition of λO-tag to (T1-16)-TEV-E9

Primer sequence	Purpose
5′- ATCGCATCACATATGACCAATACCGCGAA AATTCTGAATTTTGGCCGCATGAGCGGTG GAGATGGACG-3′	Forward primer to add λO-tag to (T1-16)-TEV-E9
5′-CCGGATCTCAGTGGTGGTGG-3′	Reverse primer to add λO-tag to (T1-16)-TEV-E9

7.2.1 Elongators and Terminators for PolyQ production

Table 7.3: Oligomer sequences for polyQ elongators and terminators. Restriction sites are underlined and complementary overhangs are highlighted in red.

Oligomer Sequence	Purpose
5′-TCAATCAACTAGTAG-3′	5' terminator coding strand, with SpeI restriction site
5′-TTCTACTAGTTGATTGA-3′	5' terminator non-coding strand, with SpeI restriction site
5′-AACTAGTTAGGTAAGTT-3′	3' terminator coding strand, with SpeI restriction site
5′-AACTTACCTAACTAG-3′	3' terminator non-coding strand, with SpeI restriction site
5′-AACAGCAACAGCAAC-3′	Elongator coding strand
5′-TTGTTGCTGTTGCTG-3′	Elongator non-coding strand