UNIVERSITY OF LEEDS

Understanding the Mechanical Properties of the *Staphylococcus aureus* Biofilm-associated Protein, SasG

Alexander Jonathan Edward Bruce

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

The University of Leeds

Astbury Centre for Structural Molecular Biology

January 2022

Acknowledgements

Firstly, I would like to thank my supervisor Professor David Brockwell for guiding me through this PhD. You have brought me out of my comfort zone, allowing me to flourish as a scientist. I would also like to express my gratitude to those in Sheffield, especially my co-supervisor, Dr. William 'Mack' Durham, who helped me along the way. I am immensely grateful to Nasir Khan for sorting out all my purchases, helping with equipment and of course, feeding me most days of the week in my final year! I must also thank Sophie Cussons for fixing every piece of equipment I may/may not have broken? They just don't make ÄKTAs like they used to… Of course, I would like to thank everyone (past and present) in the Radford-Brockwell-Calabrese labs and in Astbury who has helped me individually throughout this PhD, but that would be thesis-length in itself. However, I would especially like to thank Dr. Willis for sharing his fountain of scientific knowledge on a daily basis. Also, Dr. Schiffrin for introducing me to the world of Python programming language and helping me write/debug scripts. James for those late night lab chats, and everyone else.

I would also like to thank Dr. Bezzubova, who gave me the chance to intern at Jones Day and introduced me to the exciting world of intellectual property. I learnt so much and thoroughly enjoyed every day at the Munich office.

I would like to thank the support of my family, especially my Mum and Dad, who have been there for me at every stage of this exciting journey. I am incredibly lucky and I am so grateful for them. My deepest appreciation goes to my other half, Kate. I am forever in your debt for all your patience, understanding, enthusiasm (for pretty much anything) and endless love throughout my PhD.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ΔL_C	Change in contour length
Å	Angstrom
Aap	Accumulation-associated protein
AEX	Anion exchange
AFM	Atomic force microscope
CD	Circular dichroism
CEX	Cation exchange
Clf	Clumping factor
CLM	Collagen-like motif
Co^{2+}	Cobalt ion
CTM	Charged triple mutant
CWA	Cell-wall anchored
DLL	Dock, lock and latch
DNA	Deoxyribonucleic acid
DTPA	Diethylenetriamine pentaacetate
E. coli	*Escherichia coli*
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EPS	Extracellular polysaccharide
FACTS	Fast analytical continuum treatment of solvation
Far-UV	Far-ultra violet
Fg	Fibrinogen
Fn	Fibronectin
FnBPs	Fibronectin binding proteins
FWHM	Full width at half maximum
FX	Force-extension
GA	Gibson Assembly®
GB1	Protein G
GG	Golden Gate
H_6	Hexahistidine tag
HT	High tension
ICAM-1	Intercellular adhesion molecule-1
Ig	Immunoglobulin
IMAC	Immobilised metal affinity chromatography
K	Kelvin
LB	Lysogeny broth

L_C	Contour length
LC-MS	Liquid chromatography mass spectrometry
Lor	Loricrin
MBP	Maltose-binding protein
MC	Monte Carlo
MCS	Multiple cloning site
MD	Molecular dynamics
MSCRAMMS	Microbial surface component recognising adhesive matrix molecules
NEAT	Near iron transporter
nm	Nanometer
nN	NanoNewton
OI	Osteogenesis imperfecta
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
pN	PicoNewton
PSGL-1	P-selectin glycoprotein ligand-1
RCO	Relative contact order
RpfB	Resuscitation-promoting factor B
S. aureus	*Staphylococcus aureus*
S. epidermidis	*Staphylococcus epidermidis*
SasG	Surface protein G
SCFS	Single cell force spectroscopy
ScFv	Single-chain Fv fragment
Sdr	Serine aspartate repeat
SDS-PAGE	Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SEC	Gel filtration chromatography
SEC-MALLS	Gel filtration chromatography with multi-angle laser light scattering
SMFS	Single molecule force spectroscopy
TEV	Tobacco Etch Virus
TS	Transition state
vWF	Von Willebrand factor
WLC	Worm-like chain
Zn^{2+}	Zinc ion

List of Amino Acid Abbreviations

Alanine – Ala – A

Arginine – Arg – R

Asparagine – Asn – N

Aspartic acid – Asp – D

Cysteine – Cys – C

Glutamic acid – Glu – E

Glutamine – Gln – Q

Glycine – Gly – G

Histidine – His – H

Isoleucine – Ile – I

Leucine – Leu – L

Lysine – Lys – K

Methionine – Met – M

Phenylalanine – Phe – F

Proline – Pro – P

Serine – Ser – S

Threonine – Thr – T

Tryptophan – Trp – W

Tyrosine – Tyr – Y

Valine – Val – V

Abstract

In nature bacteria typically exist on substrates as multicellular communities encapsulated in extracellular polysaccharide substance (EPS), commonly known as biofilms utilising a repertoire of cell surface proteins. The cell wall anchored (CWA) *Staphylococcus aureus* (*S. aureus*) surface protein G (SasG) B domain is thought to play a role in both the formation and maintenance of biofilm structure. However, *in vitro* evidence is inconclusive. Here we utilise a single-molecule force spectroscopy (SMFS) protein display system to first determine whether there is homophilic* bonding of SasG in the presence of Zn^{2+} and the molecular basis for this. We establish that the B domain of SasG does participate in Zn^{2+}-induced dimerisation and believe this is through the use of pleomorphic coordination of Zn^{2+} ions (verified using a variant designed to reduce the Zn^{2+} coordination capacity).

In addition, as biofilms form at a solid:liquid interface, cell surface proteins are directly subject to external hydrodynamic forces. The ability of the bacterial cell-wall protein structures to withstand or respond to these forces as a mechanical cue will determine the success of a bacterial colonisation. The B domain of SasG comprises E and G5 repeats which have remarkable mechanical stability, however, the origin of this unusual mechanical phenotype is unknown. Through the use of protein engineering we create a homo-polyprotein system suitable for representing the B domain of SasG in SMFS experiments. By disrupting structures and bonding patterns through residue substitution, we uncover a novel mechanical protein motif – the collagen-like region of E and G5 domains. Furthermore, we demonstrate the main force-bearing region are the 'mechanical clamps', comprising of tandem arrays of long stretches of hydrogen bonds and side chain packing interactions, as recently predicted.

These results provide *in vitro* evidence for the Zn^{2+}-dependent aggregation of SasG in biofilm formation and provide a novel target for therapeutics. Furthermore, insights on novel mechanical structures adds to the toolbox for the rational design of designer proteins requiring a level of mechanical strength for function.

---

* Homogeneous protein-protein interactions

1. Introduction

In 1675 Antonie van Leeuwenhoek, a Dutch cloth merchant, ground glass lenses to construct a magnifying lens for examining weaves of cloth. Upon looking at a sample of rain water he observed it to be teeming with tiny living things, which he termed 'animalcules'[1]. The discovery was noted in his highly celebrated 1677 paper, the 'letter on the protozoa'[2]. In later years, he gave unequivocal descriptions of bacteria[3], which still hold true today (Figure 1.1). Since Leeuwenhoeks initial discoveries, it has been and will be an exciting journey of understanding, fighting and exploiting bacteria for our benefit[4].

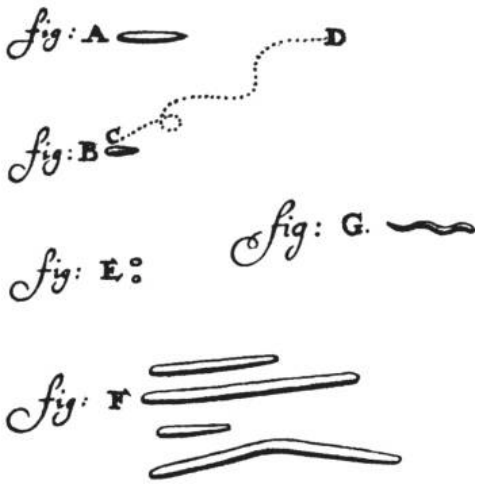

Figure 1.1 – Hand-drawn micrograph of 'animacules' from Leeuwenhoek's mouth. The dotted line portrays movement. Taken from reference[3].

1.1 Biofilms

Bacterial adaption resulting from genetic recombination events, mutations, acquisition of new genetic material and the regulated flexibility of gene expression (phenotypic plasticity) is the cornerstone of the fitness and survival of bacteria. One significant and clinically relevant example of bacterial adaption is the ability to shift from planktonic free-moving cells to their sessile biofilm counterparts. The inadvertent discovery of biofilms can be accredited to Van Leeuwenhoek, where

he observed microorganisms on tooth surfaces[3]. However, it was not until Heukelekian and Heller discovered that bacterial growth and activity was enhanced by surface attachment (the 'bottle effect')[5] and Zobell's observation that the population of bacteria on surfaces was substantially higher than in the surrounding medium[6] that it became clearer that bacteria had an affinity for surface growth. The examination of biofilms had to wait until 1969 when Jones and colleagues visualised biofilms from trickling filters of a wastewater plant using electron microscopy (Figure 1.2)[7]. Based on observations of cell morphology, these researchers were able to show that the biofilm was composed of a variety of microorganisms and determined these cells were encased in polysaccharide-like material through use of the stain, Ruthenium red.

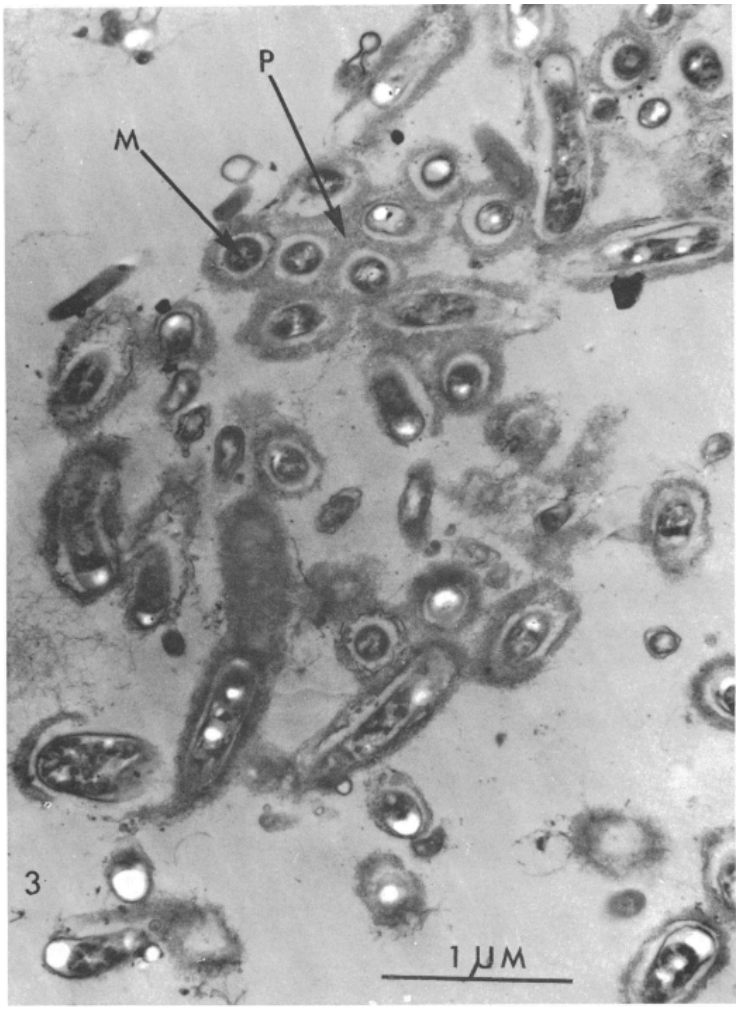

Figure 1.2 – One of the first published images of a biofilm ('slime layer') ever imaged. Electron micrograph of a 9-day sample with lead citrate post stain. M are microbes and P is polysaacharide-like material. Taken from reference[7].

Biofilms are commonly defined as a highly structured aggregate of microorganisms (predominantly mixed-species[8]) on a substrate/interface/each other, in which cells are frequently embedded within a self-produced matrix of EPS and exhibit an altered phenotype with respect to growth rate and gene transcription in comparison to planktonic cells[9]. Most cells in both mobile biofilms (flocs) and surface attached biofilms are in contact with one another, hence the term 'aggregate'[10]. Bacteria in their planktonic form have high cell growth, reproduction rate and expression of virulence factors, whereas their sessile biofilm counterparts are quite the opposite[11]. Surprisingly, the natural and predominant state appears to be the latter and has been determined to be one of the most ubiquitous and successful modes of life on Earth[12]. There are several incentives for biofilm formation, including the means to colonise a favourable niche, as a stress response and to exploit the cooperative benefits as a community[13].

Biofilm formation is a complex developmental process with involvement of over 100 genes/proteins[14] and follows a common five-step developmental pathway: reversible attachment, irreversible attachment, microcolony formation (accumulation/aggregation), maturation and dispersion (Figure 1.3). When planktonic bacteria are within 10-20 nm of the surface, their negative charges are repelled by those found on most environmental surfaces[15]. This repulsion can be overcome with attractive non-specific forces (Van der Waals, electrostatic and hydrophobic), allowing the bacteria to reversibly attach. If the cell does not dissociate, the transition from reversible to irreversible attachment may occur. For gram-positive bacterial species colonising humans, this irreversible interaction is mainly mediated through attachment to matrix proteins such as fibrinogen (Fg), Fibronectin (Fn) by microbial surface components recognising adhesive matrix molecules (MSCRAMMS)[11].

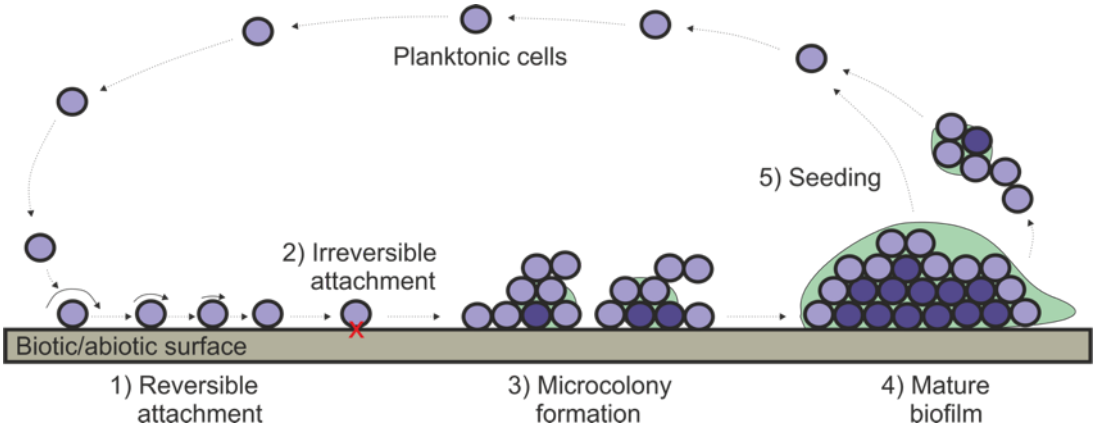

Figure 1.3 – Schematic of the five major stages of the developmental process of biofilm formation. Bacteria cells with a planktonic phenotype (light purple) come into contact with a biotic/abiotic surface resulting in 1) reversible attachment mediated by non-covalent interactions. This attachment may become 2) irreversible if the bacteria engages in adhesin-ligand interactions (displayed as a red cross) through proteins such as MSCRAMMS. 3) Cells begin to aggregate together utilising EPS (green) and/or CWA proteins for intercellular contact to form microcolonies. Cells deeper in the layers begin to display their phenotypic plasticity and become more dormant – biofilm phenotype (dark purple). 4) The biofilm carries on growing, more cells become dormant and extracellular DNA, proteins, lipids and EPS are further released into the matrix to establish a mature biofilm. 5) Eventually, cells may be signalled to revert back to their planktonic phenotype and they will go onto seed other sites. Or mechanical abrasion/increase in shear flow will cause partial sloughing of the biofilm, also promoting seeding of further sites.

Following irreversible attachment, the cells begin to multiply resulting in monolayer formation. These cells express biofilm-related proteins several fold higher than their planktonic counterparts. The expression of these proteins promotes cell-to-cell adhesion and drives microcolony (aggregate) formation, which may be further strengthened by the production of EPS. Some bacterial species are unable to produce EPS (*icaADBC*-negative)[16–19] and utilise multifunctional CWA proteins to aggregate in the presence of metal ions (e.g. Zn^{2+})[20,21].

The biofilm recruits cells, from the same species and others, creating a typically heterogeneous community. As the biofilm matures into > 100 layers, it can develop into a ‘mushroom’ shape structure with pores and channels throughout permitting the transport of oxygen and essential nutrients between the microcolonies[22]. Simultaneously, the entrapped bacteria further secrete various proteins, lipids, extracellular DNA, polysaccharides *etc.* into the biofilm.

This resulting matrix provides structure to the biofilm, promoting steep chemical gradients, enhanced horizontal gene transfer and cell-to-cell communication[10,23]. The resulting biofilm is highly dynamic, with the ability to respond to local changing conditions triggering highly coordinated gene

expression resulting in changing activities, properties and structures[24]. Following mature biofilm formation, bacteria begin reverting back to their planktonic state and are released from the biofilm to seed other sites. This occurs due to a myriad of factors, including, but not limited to intense competition, outgrown population and/or lack of nutrients[11].

It is now recognised that biofilm formation is a major part of many bacterial diseases, including prosthetic infections, infective endocarditis, osteomyelitis and many more[9]. These are usually chronic and antimicrobial resistant, both due to the protective function of the EPS matrix and/or altered growth rates[10]. Furthermore, they colonise water systems, food and beverage product lines, dairy plants and more resulting in large economic burdens.

Surprisingly, bacteria preferentially form biofilms in very high shear environments and planktonic bacteria have been observed to adhere to surfaces and initiate biofilm formation under shear forces[9] suggesting that environmental mechanical forces play a key role in the physiology of biofilm formation.

1.2 Biology Takes its Cues from Physics

In nature bacteria typically exist on substrates rather than in the bulk liquid of their environment[25,26]. This interface environment is mechanically distinct from that of the bulk liquid and in turn, the cell-surface structures are directly subject to a plethora of external mechanical forces during biofilm formation. This begins with the transition from bulk fluid to surface-association where they experience contact forces with biotic surfaces as bacteria colonise the airways, intestine, urinary tract and blood vasculature, or the abiotic surfaces of riverbeds, sewage systems and catheters. These forces will initially be compressive (inwards) (Figure 1.4A) as the bacteria lands and is pressed against the surface. During/after attachment, the bacteria must resist the shear forces created by high fluid flow parallel to the substrate in both the biotic and abiotic environments. Here, the hydrodynamic forces are tangential to the attaching/attached bacterium, resulting in shear forces acting to prevent attachment and displace bound bacteria (Figure 1.4B). These forces act tensile (outwards) as a bacterium is pulled from the surface (Figure 1.4C)[27]. During the

accumulation/aggregation stage (Figure 1.4D), these tensile forces continue to act to detach cells from one another (Figure 1.4E). Furthermore, as bacteria grow within elastic materials such as the biofilm matrix (Figure 1.4F), they are subject to the rheology of their surroundings which may act to pull them apart or compress them (Figure 1.4G)[28]. In what way the bonds/intrinsic structure of the proteins respond to these external mechanical forces will determine the success of colonisation.

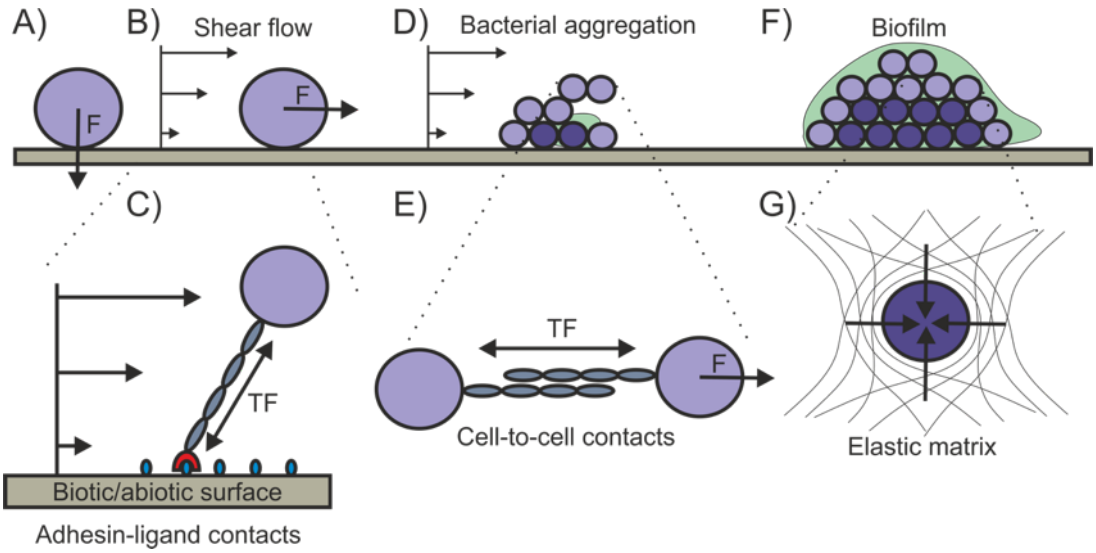

Figure 1.4 – Schematic to display the forces exerted on bacteria during biofilm formation. A) As bacteria come into contact with the biotic/abiotic surface, they are subject to compressive downwards forces. B) Once cells are associated with the surface they are subject to shear flow delivering hydrodynamic forces at a tangent to the bacteria, which C) induce tensile forces (TF) throughout molecular complexes. D) During bacterial aggregation shear flow continues to mechanically stress cells as they form intercellular contacts, resulting in E) TF across proteins involved in cell-to-cell contacts. F) Bacterial cells growing within a biofilm are subject to G) compressive forces as the growing cells mechanically deforms the surrounding elastic matrix.

1.3 The Application of Force to Study Biological Systems

During all stages of biofilm formation, force will be loaded onto bonds and propagated through protein structures. The stability of these proteins are typically investigated by chemical and thermal denaturation[29] and complex binding events are usually studied by bulk ensemble techniques such as microscale thermophoresis, isothermal titration calorimetry or surface plasmon resonance. These experiments are carried out at equilibrium under zero external force *in vitro*, a paradox when they may be subject to unsteady mechanical stimuli *in vivo*. The former may lead to ambiguous or

overlooked conclusions about structure-function relationships, especially in those protein systems whose functional role is to withstand or respond to an external load applied locally, not globally.

One method to directly study the importance of force in mechanically loaded protein systems is by the mechanical denaturation of single molecules/complexes or by mechanically perturbing their environment *in vitro*[30–32]. Two of the most popular techniques to achieve this are Atomic force microscope (AFM) force spectroscopy and optical tweezers (Table 1.1), however, techniques such as magnetic tweezers and recently developed acoustic force spectroscopy are also employed[33–36].

	AFM	Optical Tweezers
Temporal resolution (ms)	< 1*	0.1
Spatial resolution (nm)	0.1 (vertical)	0.1
Force range (pN)	< 1* – 10000	0.02 – 250
Applications	(un)folding / dissociation	(un)folding / (un)binding processive motors
Limitations	High noise level, only rupture forces, no direct measurement of separation	Narrow force range, photodamage and local heating from laser

Table 1.1 – Comparison of the key parameters, features and limitations of AFM and optical tweezers methods. As the techniques have varied underlying methods for exerting force onto single molecules, each technique is applied to differently suited biomolecular systems. *Using recently developed 'Warhammer' cantilevers by Edwards and Colleagues[37]. Adapted from reference[33].

1.3.1 AFM Force Spectroscopy

Since the invention of the AFM in 1986[38] as primarily a high-resolution imaging tool, it has since been complemented by the development of force-mode AFM. Force-mode has become an increasingly popular biophysical tool for studying the effects of force on protein systems, both as single molecule (SMFS) and on single cells (SCFS), providing information inaccessible by other methods[33,34,39–41]. In the SMFS experimental setup, the force response of a protein or complex is measured as a function of extension as the distance between two well-defined attachment points is increased, allowing interactions stabilising a structure or complex to be explored. This is owed to the AFMs high dynamic range (<1-10000 pN), highly sensitive piconewton (pN) force resolution (theoretically 2-10 pN for standard tips), measurement of forces exerted with sub-nanometre (nm) spatial resolution, and the ability to operate under physiological conditions[32,33,42–45]. New

technological advances have allowed improved experimental ability with drift reduced to 100 pm over tens of minutes and sub-pN force resolution by engineering cantilevers[46–48]. Furthermore, the ability to probe protein unfolding and protein-protein/protein-ligand unbinding up to the retraction velocities exploited in molecular dynamics (MD) simulations with sub-microsecond time resolution has been achieved with high speed force spectroscopy utilising ultra-short cantilevers[49].

1.3.1.1 AFM Experimental Set Up

An AFM comprises of two essential components: An optical head and a XY stage scanner (Figure 1.5). Attached to the optical head is a gold coated silicon nitride cantilever comprising of a triangular stylus on the tip of a flexible lever, which acts as a force-sensitive probe. A piezo-electric device controls the approach and retraction of the cantilever, with respect to the surface, at a constant velocity/force by method of piezo expansion. In force-mode AFM, the cantilever approaches and pushes hard against the surface until a deflection threshold is reached. The cantilever is withdrawn at a constant velocity/force away from the surface, defining the force loaded onto a single protein or bond. A focussed diode laser is reflected off the cantilever tips gold surface onto a four quadrant position-sensitive photodiode (converts photons into a proportional photocurrent) detector. Cantilever deflection causes vertical displacement of the laser spot, which can be quantified by comparing the photocurrent produced by each quadrant. The resulting voltage difference is converted to distance in nm (x) from the cantilever tip deflection sensitivity (nm/V) determined during cantilever calibration. As the cantilever behaves as a Hookean spring, with a calculable spring constant (k), the difference in distance can be converted into force (F) according to Hooke's law (Equation 1.1)[31]:

$$F = kx \tag{1.1}$$

Hooke's law states that the restoring force is equal to the negative of the displacement times a constant (spring constant). When there is a linear force exerted, the resisting, restoring force is proportional to the exerted force.

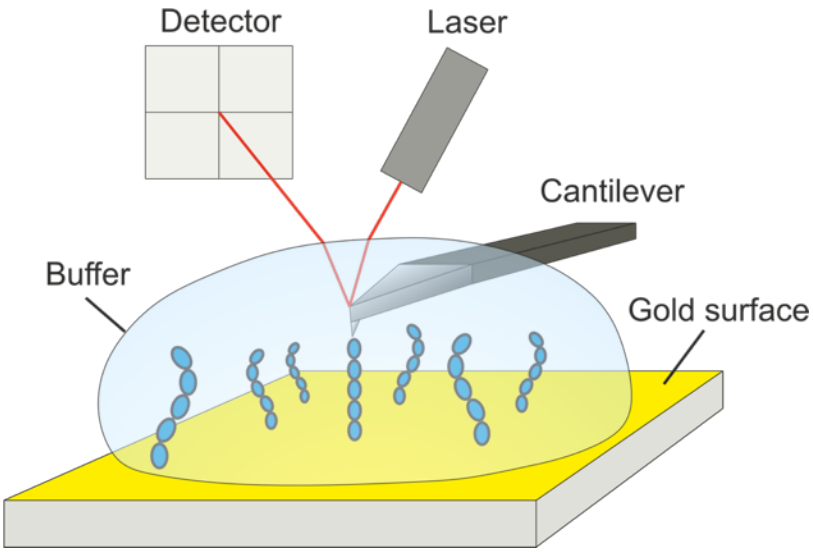

Figure 1.5 – Schematic of a typical AFM setup for polyprotein unfolding assays. The gold surface is decorated in polyproteins (blue beads) via gold-thiol bonding. A laser is positioned at the tip of the cantilever probe, which detects any deviations (bending) of the cantilever in response to entropic restoration forces of protein domains under mechanical tension.

1.3.1.2 Determining Protein Mechanical Strength Using SMFS

SMFS has been established as a powerful technique for determining the mechanical behaviour of protein structures, measuring forces that occur upon unfolding and refolding of single protein domains. Over 20 years ago, the first mechanical unfolding experiments were performed on the giant muscle protein titin using AFM force spectroscopy[50]. Although this paved the path for SMFS of polyproteins, data analysis was complicated due to naturally occurring polyproteins being heterogeneous in composition leading to subtly different mechanical properties between domains. To improve the statistical evaluation of these naturally occurring polyproteins, homo-polyproteins of identical repeats or hetero-polyproteins (chimeric) with reference domains (domains of known mechanical properties) with short peptide linkers are engineered[30]. An advantage of using polyproteins over monomeric domains is the signature 'sawtooth' force-extension (FX) profile, which distinguishes bone fide unfolding events from non-specific events or sample contaminants.

During a constant velocity SMFS experiment on a polyprotein construct, the cantilever tip will retract from the sample surface at a constant retraction speed. If the tip has interacted with one of the polyproteins bound to the gold surface, the domains will unfold on a successive domain-to-domain

basis producing a distinctive equally spaced 'saw-teeth' force-extension profile with the apices of the teeth corresponding to the force at which each domain unfolds (Figure 1.6)[51]. For heterogeneous polyproteins, the unfolding events may not be equally spaced if the domains are significantly different in size. Each tooth presents a non-linear rising edge, displaying an exponential-like relationship between extension and force as the most compliant section (linkers/unfolded domains) of the polypeptide chain behaves as an entropic spring. The steps involved in obtaining a FX trace when retracting the tip from the surface at a constant velocity is described in Figure 1.6. This resulting sawtooth profile provides a mechanical fingerprint in SMFS experiments.

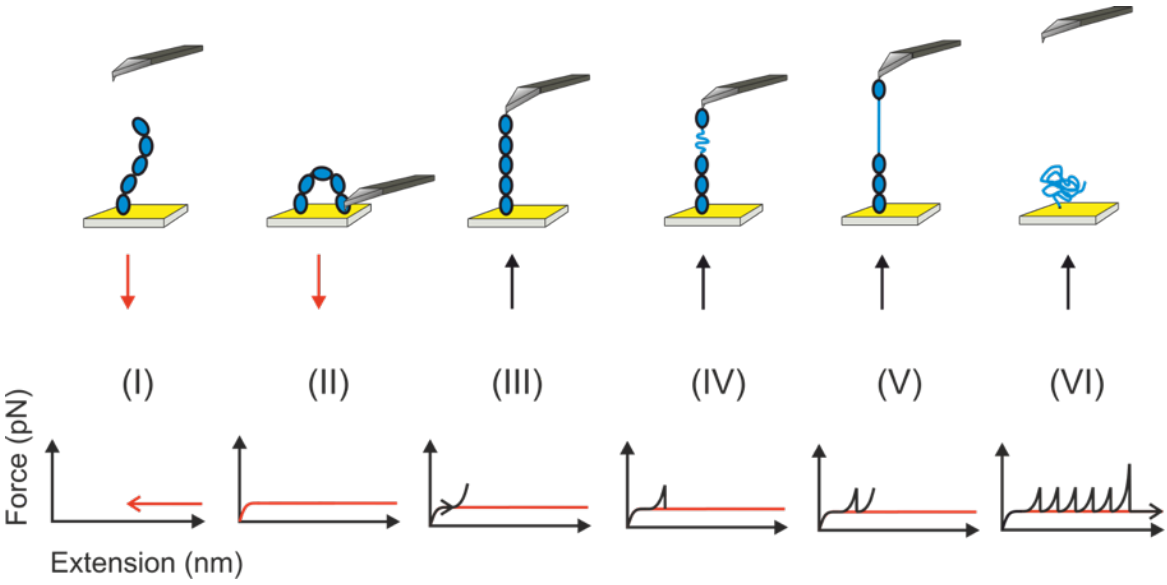

Figure 1.6 – Schematic showing the steps in obtaining a FX sawtooth profile for polyprotein unfolding at a constant retraction velocity. The features of the single-molecule diagram correspond to the force-extension profile underneath. In force-extension mode, the cantilever is moved towards the surface under piezoelectric control (I) and presses into the surface causing a change in deflection (II). Once a threshold is reached the cantilever is retracted at a constant speed and if a protein has adsorbed onto the cantilever tip, an entropic restoring force is observable as the domain resists the forced decrease in entropy (due to preference of a polypeptide chain to form a random coil[52]) as the distance between the tip and the surface increases (III). The chain is extended to a point where the force exerted on the folded domain reaches a point capable of permitting thermally activated transitions and unfolding of the domain occurs causing an abrupt change in force. This vertical linear decrease in force is because the domain unfolding occurs at a rate faster than the extension rate. (IV). This unfolded domain will continue to unravel adding to the effective length of the chain until fully stretched out, where an entropic restoring force will be observed again as force is applied across another domain (V). After all five domains have unfolded, the protein will detach from the tip resulting in a usually large detachment peak.

1.3.1.3 Probing Intermolecular Interactions Using SMFS

SMFS is also an indispensable technique for examining force-dependent properties of biomolecular complexes as it provides the estimation of unbinding forces (Table 1.2), dissociation rate constants, energy landscapes, the length of complex before dissociation and can differentiate between bond types (catch-bond, slip bond *etc.*). Furthermore, SCFS enables the quantitative study of cell bonding mechanisms under physiological conditions *in situ*. A typical SMFS experiment to determine the mechanical properties of protein:protein or protein:ligand complexes requires the binding partners to be covalently attached to the cantilever tip and surface through the use of linkers. These are typically heterobifunctional polyethylene glycol (PEG) linkers with two functional groups (e.g. NHS ester and maleimide) for protein/ligand immobilisation to the substrate and cantilever tip. These linkers provide defined strong covalent attachment, reduction in non-specific protein sticking, enables substrate flexibility and provides a characteristic force profile when stretched[53].

Molecular partners	**Retraction velocity (nms^{-1})**	**Dissociation force (pN)**	**Reference**
Avidin/biotin	5000	173 ± 19	54
Streptavidin/biotin	5000	326 ± 19	54
Human serum albumin (HSA)/anti-HSA	200	244 ± 22	55
Intercellular adhesion molecule-1 (ICAM-1)/anti-ICAM-1 antibody (Ab)	4680	100 ± 50	56
Ab single-chain Fv (ScFv) fragment/fluorescein	1000	50 ± 4	57
P-selectin/P-selectin glycoprotein ligand-1 (PSGL-1)	2800	~ 168	58

Table 1.2 – Early examples of complex dissociation events studied by AFM force spectroscopy. Values taken from reference[32].

The SMFS procedure involves bringing the AFM probe into contact with the surface until a defined threshold is reached in an attempt to form a complex between the binding partners (Figure 1.7). Upon retraction of the AFM probe, if an interaction has been formed, an entropic restoring force increasing in an exponential-like fashion will be observed. When the complex dissociates, the force

sharply drops to zero. The force at the apice of the exponential-like curve is the force at which complex dissociation occurs and is usually a distribution of forces as complex dissociation under force is a stochastic process.

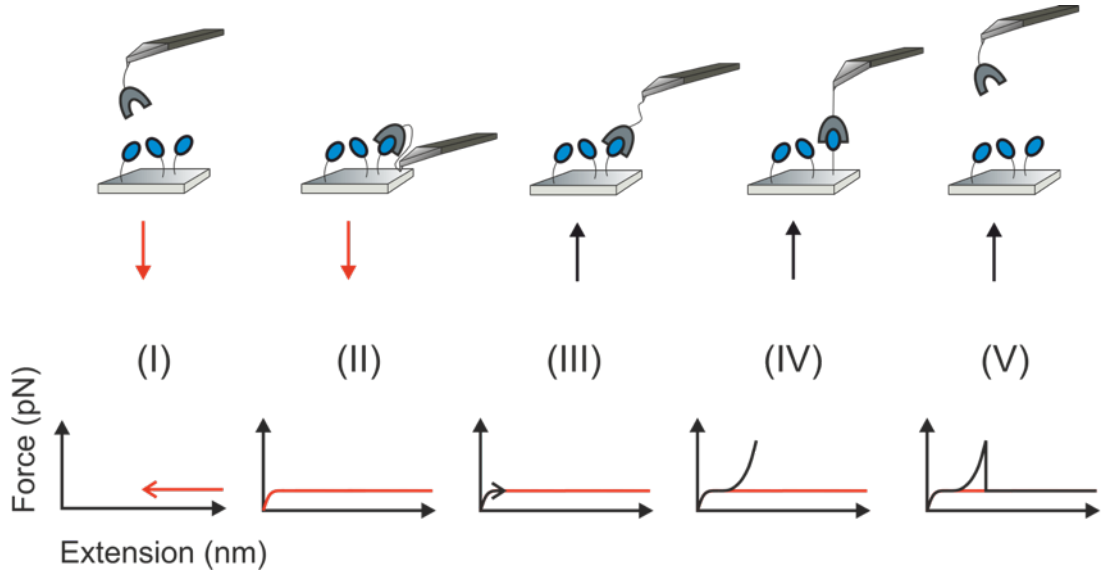

Figure 1.7 – Schematic of a typical SMFS protein:protein or receptor:ligand experiment to measure complex dissociation at a constant retraction velocity. The cantilever tip and silicon nitride surface are functionalised with binding partners, coloured in grey and blue. I) The cantilever is brought towards the functionalised surface (red line) II) until a defined deflection is reached, which brings the binding partners into contact with one another. III) The cantilever is retracted from the surface at a constant velocity and IV) If a complex is formed, an entropic restoring force is produced as the complex is stretched causing a bend in the cantilever until V) complex dissociation.

1.3.2 The Worm-Like Chain Model For Polymer Extension

As discussed earlier, a polypeptide chain (or PEG linker) acts as an entropic spring, producing a restoring force when stretched. This elasticity follows a nonlinear relationship which can be described by the Worm-Like Chain (WLC) model of polymer elasticity; an extension of the Freely-jointed chain model describing the semi-flexible behaviour of polymers subject to an external force[31,59]. The WLC continuum model denotes a polymer as a continuous flexible chain of static segments of fixed inter-segment angles with the ability to rotate freely about their vertices[52]. This model was empirically derived from the stretching of DNA[60,61] and predicts the entropic restoring force (F) of the polymer chain at any extension (x) using Equation 1.2:

$$F(x) = \frac{k_B\,T}{p}\left(0.25\left(1-\frac{x}{\mathrm{L_c}}\right)^{-2} - 0.25 + \frac{x}{\mathrm{L_c}}\right)$$

1.2

Where, parameter *p* is the persistence length (length of static components – an amino acid, 4.0 $\pm$ 0.2 Å for proteins[62]) and L_C is the contour length.

The WLC model has been extensively used to delineate protein domain unfolding events[31] and protein:protein interactions[63,64] as the FX profiles are well defined by the model. When this model is fit to the rising edge of the signature exponential-like curve during complex dissociation or protein unfolding, it is possible to determine the L_C (Figure 1.8). The L_C is valuable for both protein unfolding and complex dissociation studies. It is the predicted maximum extension of the polypeptide chain and is proportional to the length of the complex at dissociation or to the total number of 'unravelled' amino acids at protein unfolding events.

For polyprotein unfolding studies, quantification of this parameter for each unfolding event allows the change in contour length (ΔL_C), the unfolded length of the polyprotein minus the length of the folded protein, to be calculated (Figure 1.8B). This is related to the number of amino acids 'trapped' in the native structure, and thus provides a value equal to the number of amino acids 'released' during an unfolding event. Comparison of this value to a known 3D structure of a protein allows assignment of an unfolding peak to a particular protein.

Fitting the WLC model to the leading edge predicts how force will be loaded onto the polymer chain as a function of extension. Theoretically, the domain can unfold at any point along this extension due to the stochastic nature of thermal fluctuations. If a domain were to unfold at a low force (corresponds to a low polypeptide chain extension), the distance to the next unfolding event in a polyprotein (peak-to-peak distance) would be larger than average or the unfolding distance shorter than average/predicted[44]. As the WLC model extrapolates to full extension at each polyprotein unfolding event, the ΔL_C is therefore not equal to the peak-to-peak distance.

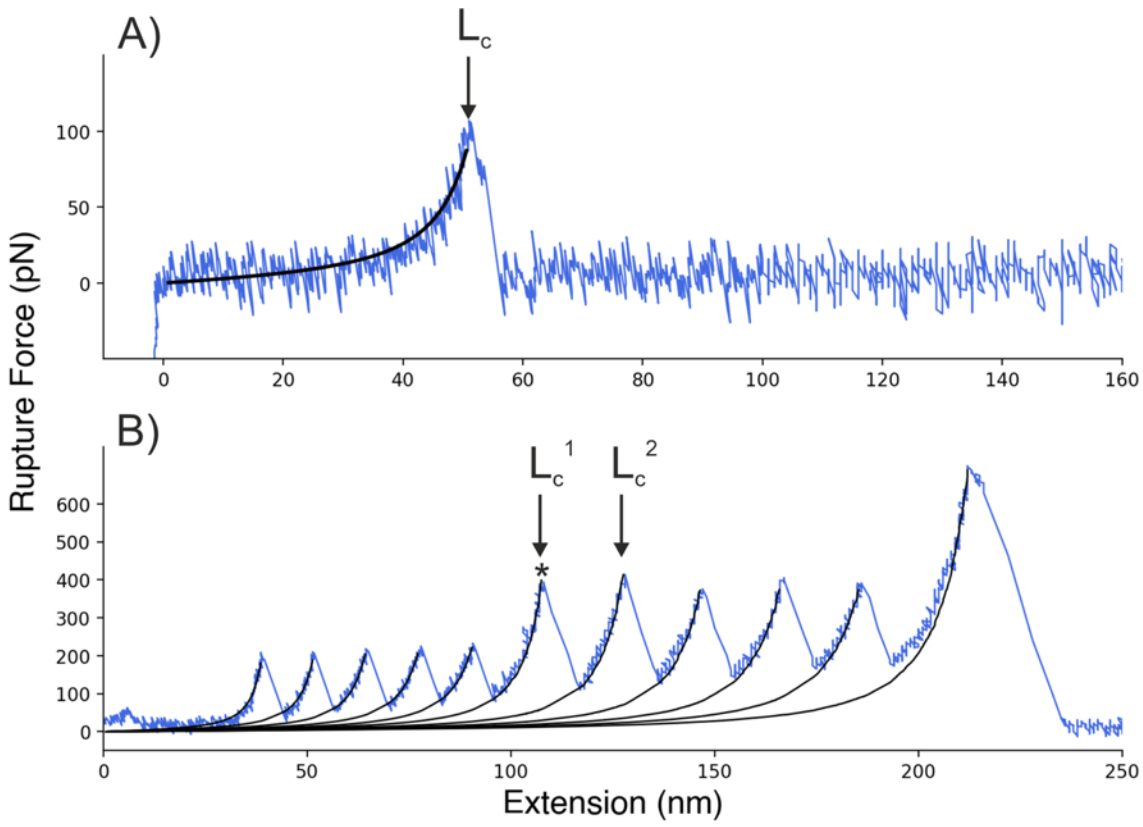

Figure 1.8 – The WLC model can be used to model the entropic response of a polypeptide chain to force during protein unfolding and complex dissociation. A) SMFS complex dissociation example from this thesis at a retraction velocity of 1000 nms^{-1}. WLC fit as a black line giving the L_C. B) SMFS E-G5[2] pentamer mechanical unfolding at a retraction velocity of 1500 nms^{-1} from this thesis. WLC model as thin black lines, with a held persistence length of 3.8 Å. ΔL_C of peak* is calculated from the subtraction of L_C^1 from L_C^2 and corresponds to the number of amino acids 'released' during that unfolding event.

When probing interaction events, the L_C allows identification of the specificity of an interaction, as the expected L_C can be estimated from structural information. Changes in the predicted L_C values can indicate remodelling of the binding partners[64], or report on different interactions/binding regimes (Figure 1.9). Similar to polyprotein unfolding, the predicted L_C is typically different from the measured distance at complex rupture due to the stochastic nature of complex dissociation under force.

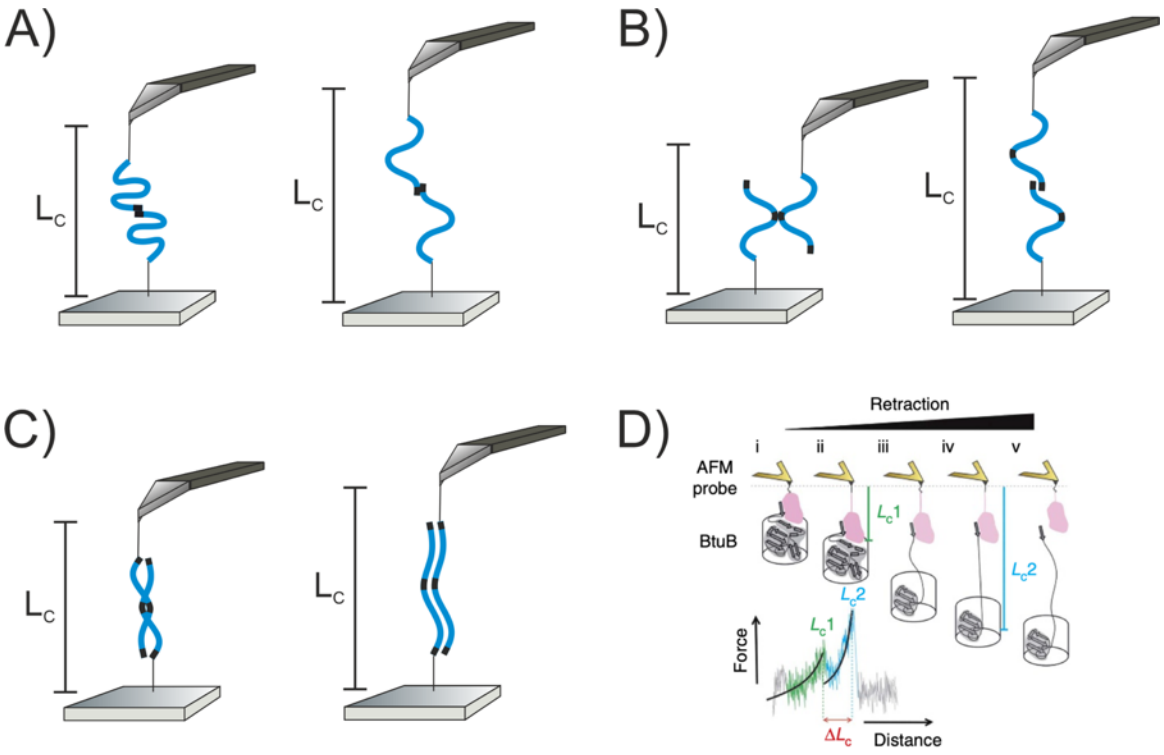

Figure 1.9 – Properties of interactions and proteins the L_C may report on. A) Distinct L_C values may result from different conformations (such as collapsed and extended) of a protein with the same interaction region (black). B) Binding partners may have multiple interaction sites which will yield distinct L_C values. C) The conformation of the complex may result in different L_C values, such as two interaction partners tightly coiled around one another in comparison to the partners binding side-to-side. D) An experimental example of the remodelling of proteins prior to complex dissociation taken from reference[64]. A double L_C is observed when BtuB is subject to external force, corresponding to unfolding of part of the plug domain (light grey shaded region) and then dissociation of BtuB (open cylinder) from TonB (pink). ΔL_C corresponds to the amount of protein that unfolds under force.

In addition, discrepancies in the fitting of the WLC model can elucidate intermediate unfolding events during protein unfolding. During I27 polyprotein unfolding the rising phases of the force peaks display a deviation (shoulder) from the WLC fitting, which gets less apparent as event number increases (Figure 1.10)[65]. This shoulder was attributed to the hydrogen bonds between A and B β-strands simultaneously breaking in all of the folded domains leading to an extension of 6.6 Å per domain. Hence, the shoulder decreases in size as the event number increases because the number of folded domains decreases.

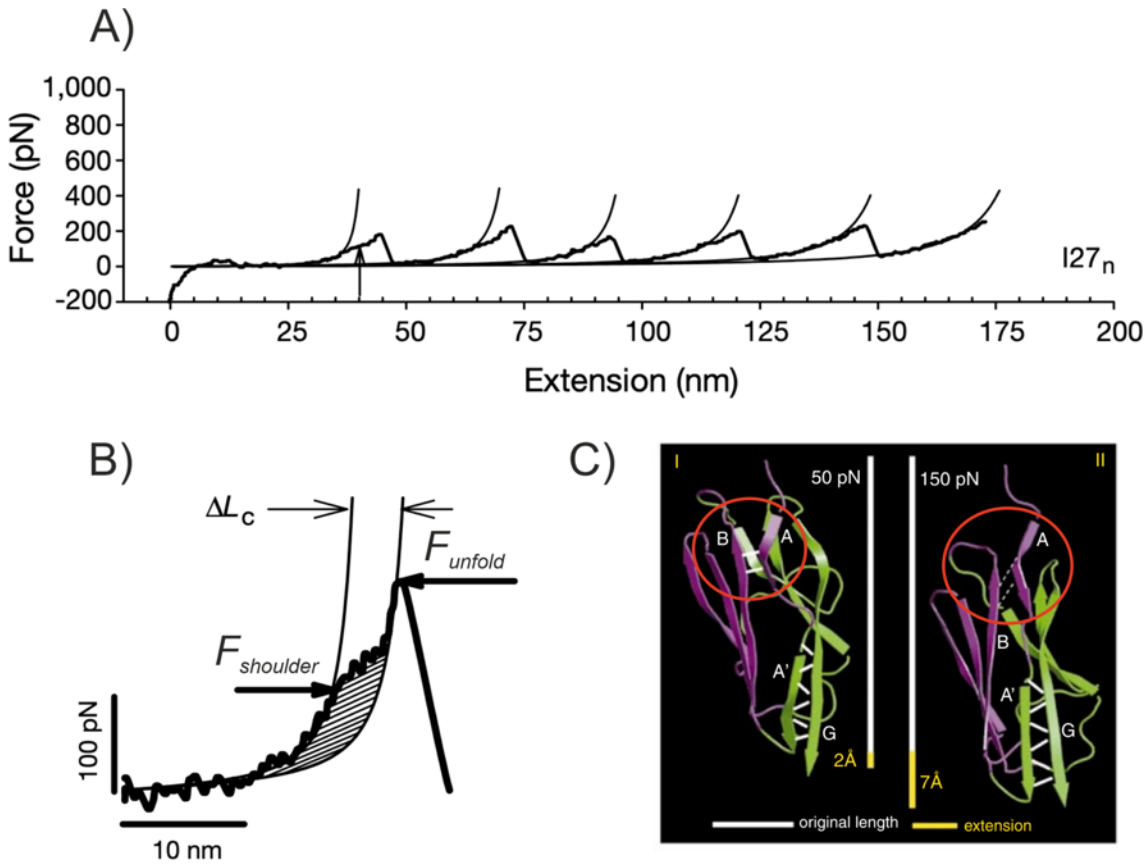

Figure 1.10 – Intermediate unfolding events during protein unfolding can be determined through discrepancies of the WLC model fit. A) Force peaks corresponding to the sequential unfolding of the I27 domain of human cardiac titin, displaying shoulder-like deviations (arrow marking the initial peak shoulder) on the WLC model fit. B) FX profile of the first force peak of the sawtooth pattern displaying the shoulder. Subtracting the $F_{shoulder}$ L_C from the F_{unfold} L_C gives an ΔL_C value equal to the total structural extension to the intermediate for all folded domains. C) The structure of I27 displaying the hydrogen bonding between the A and B β-strands (solid white lines, panel I) shearing under 150 pN of force (dashed white lines, panel II) to give a total extension of ~ 7 Å to the intermediate structure. Figure adapted from reference[65].

1.3.3 Exploring the Underlying Energy Landscape During Mechanical Perturbation – The Bell-Evans-Ritchie Model

SMFS unfolding of a protein and protein:protein/protein:ligand unbinding can be considered a two-state process; a low-energy folded state and a high-energy unfolded state, with a 1D energy barrier, termed the mechanical transition state (TS), separating the states and is defined by the reaction coordinate, which in an SMFS experiment is the end-to-end length of a protein (Figure 1.11). Although the assumption of a 1D reaction coordinate simplifies physical reality, it allows us to describe the mechanical unfolding of a protein/domain or dissociation of a complex.

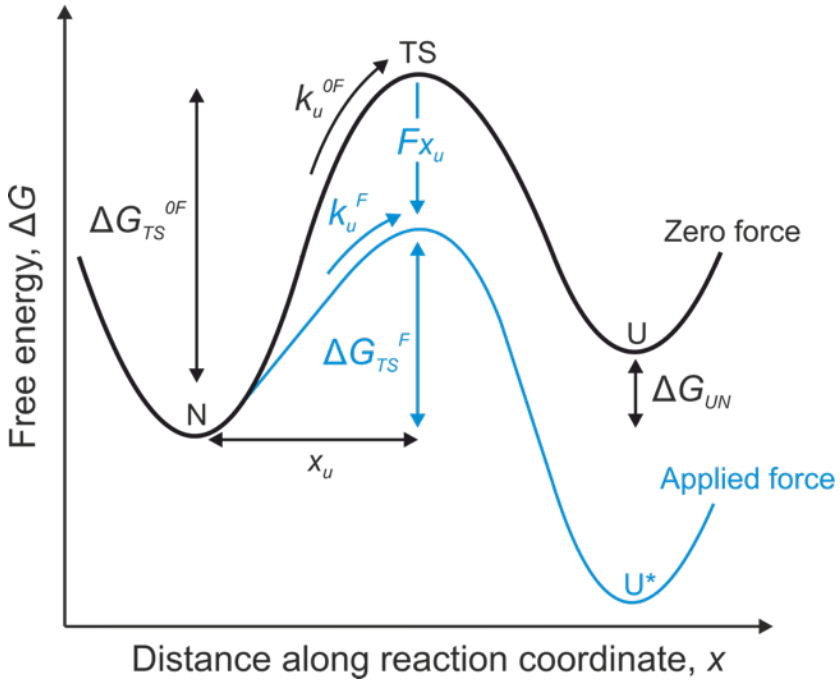

Figure 1.11 – Schematic displaying the underlying free energy landscape of protein unfolding under force. A typical free energy landscape of a two-state protein folding pathway leading from a native, folded state to an unfolded state via a high energy transition state (ΔG_{TS}^{0F}). Under zero external force, spontaneous crossing of the transition state energy barrier occurs at unfolding rate k_u^{0F} (or k_{off}^{0F} for complex dissociation). The distance from the folded state well to the transition state is x_u along the reaction coordinate (end-to-end extension). The application of external force tilts the energy landscape (blue) by a function of Fx_u, both lowering the energy barrier and stabilising the unfolded population (U*).

Applied force acts as a local denaturant and according to the Bell model can be described as a tilting of the energy landscape as displayed in (Figure 1.11)[66]. It assumes the rate of unfolding is described by a van't Hoff Arrhenius-like equation, where the rate of reaction is dependent on the force. Under equilibrium conditions with no external force, the majority of the protein will populate the low-energy folded/bound native state. To reach the high energy unfolded state, the protein must first traverse the highest energy state, the TS. The height of this barrier (ΔG_{TS}^{0F}) defines the rate the protein will unfold spontaneously due to thermal fluctuations. When an external force (F) is applied onto the polypeptide chain (increasing the distance between two fixed points e.g. the proteins termini), the energy landscape is tilted along the mechanical coordinate by a function of $F(cos\theta)x$, where x is the distance reaction coordinate from the native well and θ is the angle between the molecular coordinate and applied force (Figure 1.11). This stabilises the TS and unfolded state, by lowering the energy barrier to unfolding/dissociation (ΔG_{TS}^{0F}) by a function of Fx_u to give ΔG_{TS}^F ($\Delta G_{TS}^{0F} - Fx_u$), where x_u is distance between the native state and TS. The exponential increase in

the unfolding rate constant in the presence of force (k_u^F) can be described using the analytical equation derived by Evans and Ritchie (Equation 1.3)[66].

$$k_u^F = Aexp^{-\frac{(\Delta G_{TS}^F - Fx_u)}{(k_B T)}} = k_u^{0F} exp^{\frac{Fx_u}{k_B T}}$$

1.3

Where k_B is Boltzmann's constant, T is the temperature, A is the attempt frequency and k_u^{0F} is the unfolding rate constant in the absence of force. There is no effect on the native state, however, the unfolded state is energetically more favourable than the native state[67]. It is assumed there is no change in shape of the energy landscape[68]. In turn, there is an increase in the probability of thermally activated free-energy barrier crossing attempts from the native to the unfolded well. Thus, forced protein unfolding by SMFS occurs as a result of a balance between thermal and mechanical forces and is therefore a stochastic process.

Different retraction velocities provide different time windows for thermally driven barrier crossing attempts to occur under application of external mechanical load. For example, for protein unfolding: at low retraction velocities (low loading rates), the protein/complex will spend more time under force allowing more thermally activated barrier crossing attempts, thus a higher probability of crossing the transition state to unfolding/dissociation. In turn this leads to a lower unfolding force and the opposite is true of higher retraction velocities (Figure 1.12). It is the retraction velocity dependence which enables the underlying features of the unfolding energy landscape to be extracted from data using the Bell-Evans-Ritchie model.

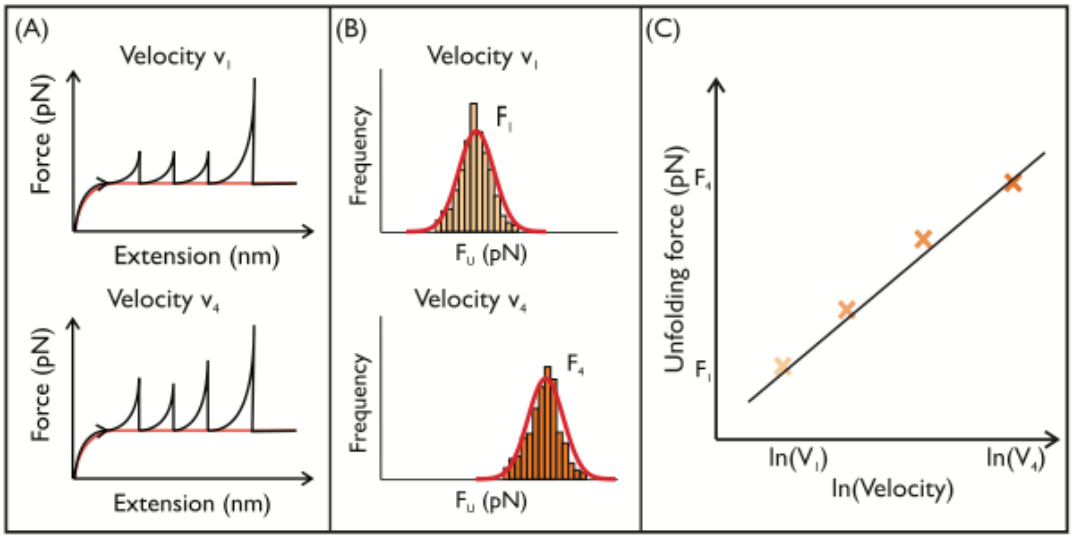

Figure 1.12 – SMFS force-extension experiments are generally carried out at a variety of different retraction velocities. A) Fabricated FX profiles of a polyprotein unfolding to give a sawtooth profile at a low (V_1) and high (V_4) retraction velocity. The higher retraction velocity displays an obvious increase in peak height. B) These forces are plotted as a histogram and Gaussian model is fit to them to extract the modal rupture force. The modal rupture force typically increases with the increase of retraction velocity. C) The modal or mean rupture force at a range of retraction velocities can be plotted as a function of the natural logarithm of retraction velocity and fitted with a linear regression. Downstream simulations (Monte Carlo) permit the underlying features of the unfolding energy landscape to be extracted from this data. Figure taken from reference[31].

This relationship between unfolding force and loading rate depends on x_u, also described as the 'malleability' of the protein, or the distance from the native well to the TS barrier. This is described by Equation 1.4 (the Bell-Evans-Ritchie model), where the rupture force (F_R) is proportional to the logarithm of loading rate (R)[66,67]:

$$F_R = \left(\frac{k_B T}{x_u}\right) \ln\left(\frac{R x_u}{k_u^{0F} k_B T}\right)$$

1.4

Where k_B is the Bolzmann's constant, T is temperature in Kelvin and k_u^{0F} is the spontaneous protein unfolding rate at zero force. R is defined as the product of retraction velocity and the cantilever spring constant defined by Equation 1.5:

$$R = \frac{\mathrm{d}F}{\mathrm{d}t} = kv$$

1.5

Where v is the retraction velocity and k is the cantilevers spring constant. However, for polyprotein unfolding, the loading rate does not remain constant throughout an experiment[69]. During an SMFS polyprotein unfolding experiment, both the compliance (determined by stiffness of the scaffold holding the domain during extension and the flexible cantilever) and number of folded domains changes during the experimental time course. Figure 1.13 illustrates these competing effects which influences the loading rate applied to each folded domain and therefore precludes the direct calculation of loading rates from polyproteins. To circumvent this, the unfolding force is studied as a function of retraction velocity (Figure 1.12C), which is fitted with a Monte Carlo (MC) model to extract parameters of the underlying energy landscape (k_u^{0F} and x_u) by taking into consideration the effects the domain number and compliance effect have on the loading rate[50,70].

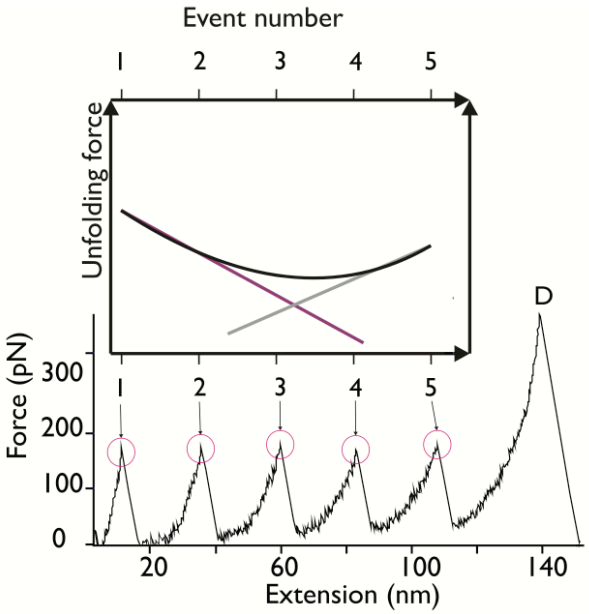

Figure 1.13 – The unfolding history effect on the observed unfolding force. A FX profile of an I27 pentameric polyprotein, with the event numbers numbered 1 – 5 and the detachment peak labelled as D. 1 and 5 correspond to the first and last I27 unfolding event, respectively. Inset is a schematic illustrating how the rupture force changes with event number due to two competing effects: domain number and compliance effect. As the event number increases, the number of domains left folded decreases and the probability for thermally activated crosses of the barrier decreases, which increases the observed unfolding force (grey linear line). However, as the number of unfolded domains increase, the length, and therefore, compliance of the chain increases. As the chain extends, more time is spent at a lower force, resulting in a decreased observed unfolding force (purple linear line). Competition between compliance and domain number effects results in a non-linear relationship between event number and rupture force. This relationship will only hold for a polyprotein where the domains unfold individually. Figure taken from reference[31].

1.3.4 What Governs the Mechanical Stability of Proteins?

We have discussed that the competing construct and compliance effects influence the rupture force of domains, but what are the intrinsic properties that have evolved for mechanical strength?

1.3.4.1 Protein Topology and Mechanical Strength

Since the development of force-mode AFM, the force-extension behaviour of a number of proteins has been studied and consequently mechanical strengths ranked. The key determinants for mechanical strength is the secondary structure content and arrangement. The general trend appears to be that α-helical proteins are the most mechanically labile, followed by α/β structures, with proteins exhibiting a high β-sheet content being the most mechanically stable[71]. This trend between mechanostability and secondary structure has also been described by a simple elastic model[72]. Furthermore, a common topology is observed in most mechanically strong proteins studied to date: The immunoglobulin (Ig)-like, β-sandwich fold[73]. However, new mechanically strong topologies, albeit similar to the Ig-like fold, are being discovered such as the β-braid[74,75]. A selection of proteins ranked by their secondary structure content with their measured rupture force at specific retraction velocities is given in Table 1.3. This general trend holds when the N- and C-termini are aligned with the direction of force.

Protein	Rupture force (pN)	Retraction velocity (nms^{-1})	SCOP class	SCOP fold	Reference
SdrG B1 domain	> 2000	400-6400	All β	Ig-like β sandwich	[76]
Aap G5 sub domain	475 ± 48	1000	All β	β-braid	[75]
SasG G5 sub domain	379 ± 28	1500	All β	β-braid	[74]
Aap E sub domain	312 ± 45	1000	All β	β-braid	[75]
SasG E sub domain	229 ± 27	1500	All β	β-braid	[74]
I27	204 ± 26	400-600	All β	Ig-like β sandwich	[29]
c7A	480 ± 14	400	All β	Ig-like β sandwich	[77]
c1C	425 ± 9	400	All β	Ig-like β sandwich	[77]
I1	127 ± 18	400-600	All β	Ig-like β sandwich	[78]
1FNIII	220 ± 44	600	All β	Ig-like β sandwich	[79]
10FNIII	74 ± 20	600	All β	Ig-like β sandwich	[79]
13FNIII	89 ± 18	600	All β	Ig-like β sandwich	[79]
*Tm*CSP	78 ± 2	400	All β	OB-fold	[71]
Barnase	~ 70	100-500	α/β	Microbial Ribonuclease	[80]
Protein L	152 ± 5	700	α/β	β-grasp	[70]
Protein G (GB1)	184 ± 41	400	α/β	β-grasp	[81]
Ubiquitin	203 ± 35	400	α/β	β-grasp	[82]
Spectrin	60 & 80*	3000	All α	Spectrin repeat-like	[83]
Calmodulin	< 15	600	All α	EF Hand-like	[51]

Table 1.3 – Table of rupture force values, retraction velocities, and SCOP classifications of a number of mechanically characterised proteins. Fold classification was obtained from a search of the SCOP class of the protein[84]. *60 & 80 pN for spectrin occurred due to short and long elongation events. β-braid fold comprises three interlaced strands capped at both ends by a triple stranded β-sheet. *Tm*CSP: *Thermotoga maritima* Cold Shock Protein.

1.3.4.2 Direction of Applied Force

Despite the general trend that secondary structural elements and their topology are a key determinant of the mechanostability of proteins, it cannot explain the broad range of force responses observed for proteins of similar structures. For example, FNIII domains with almost

identical structures vary significantly in their mechanical properties (from 75-220 pN at 600 nms^{-1})[79], suggesting protein mechanical strength may be modulated in other ways. Early SMFS experimental studies[51] and theoretical studies[85,86] suggested the difference in mechanostability of proteins sharing a similar topology could be due to the number, position and geometry of hydrogen bonds with respect to direction of force application.

However, it is difficult to determine the effect of extension geometry on mechanostability using different proteins due to distinct topologies, sequence, kinetic and thermodynamic stabilities. A number of groundwork studies including, but not limited to, those by Carrion-Vazquez and colleagues[82] and Brockwell and colleagues[87] circumvented these problems with techniques that allowed distinct pulling geometries to be applied to ubiquitin and E2lip3, respectively. Carrion-Vazquez and colleagues achieved this by pulling ubiquitin between the N- and C-termini or between a lys48 and a cysteine residue, resulting in unfolding forces of 203 pN (at a retraction velocity of 400 nms^{-1}) and 85 pN (at a retraction velocity of 300 nms^{-1}). The differences were determined to be due to the direction of the hydrogen bonds with respect to the applied force. By controlling the points of extension, Brockwell and colleagues observed that when force was applied parallel ('sheer' geometry – Figure 1.14A) to the β-strands of E2lip3 an unfolding force of 187 pN at 700 nms^{-1} was observed, however, when force was applied perpendicular ('peeling' geometry – Figure 1.14B) to the β-strands of the protein, the mechanical response of the protein became undetectable by AFM (Figure 1.14C)[87]. Since 2003, the importance of secondary structure arrangement in relation to the force has been demonstrated for a variety of proteins, such as with green fluorescence protein and was achieved by controlling the pulling geometry by circular permutations (identical topology, but different location of N- and C-termini)[88].

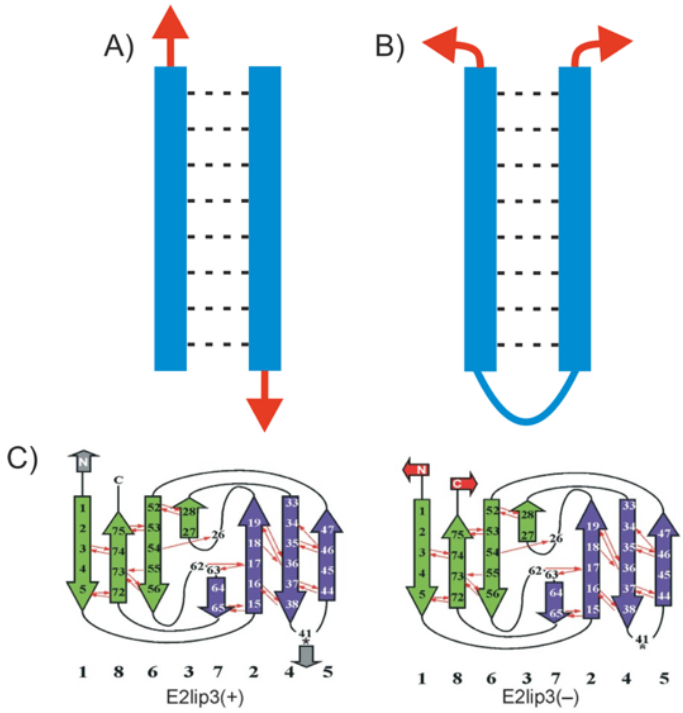

Figure 1.14 – The mechanical strength of a protein depends on the pulling geometry with respect to the hydrogen bonded β-strands. A) The 'sheer' geometry, where the application of force is perpendicular to the hydrogen bonds requiring simultaneous rupture of the hydrogen bonds. B) The 'peeling' geometry, where the force is applied to the N and C-termini of the protein, which causes the hydrogen bonds to sequentially break. C) Force is applied in a sheer geometry to E2lip3(+) which results in an unfolding force of 187 pN, whereas E2lip3(-) terminal strands are in 'peeling' geometry when pulled and its unfolding is undetectable by AFM. Figure C) taken from reference[87].

1.3.4.3 'Mechanical Clamps'

For a protein to be mechanically stable it must have a 'mechanical clamp': a force-bearing region (determined by a particular structure and/or interactions) able to withstand unfolding forces greater than the noise of the instrument[31]. An early example of a 'mechanical clamp' region was first identified by steered molecular dynamics simulations of forced unfolding of the I27 protein and involved two terminal β-strands (A' and G) hydrogen-bonded together (Figure 1.15)[89]. Mutagenesis coupled with experimental SMFS supported this observation[90]. As the protein unfolds in a shear geometry, the simultaneous rupture of these hydrogen bonds was required.

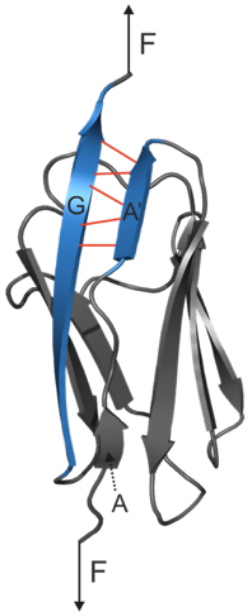

Figure 1.15 – The ability of I27 to withstand force depends mainly on the hydrogen bonded A' G β-strand mechanical clamp. The direction of pulling is shown (black arrows labelled F) to display the shear geometry of the hydrogen bonds (red solid lines). Although the A strand is a terminal strand in shear geometry, it dissociates from the structure early on during the mechanical unfolding and does not contribute to the mechanical clamp.

An extensive systematic study utilising coarse grain mechanical unfolding simulations of 17134 proteins (from the PDB), identified a number of distinctive mechanical clamp regions in proteins with predicted mechanical robustness[91,92]. The clamp motifs were defined according to the hydrogen bonding patterns between secondary structure elements in localised regions within the protein. These are displayed in Figure 1.16.

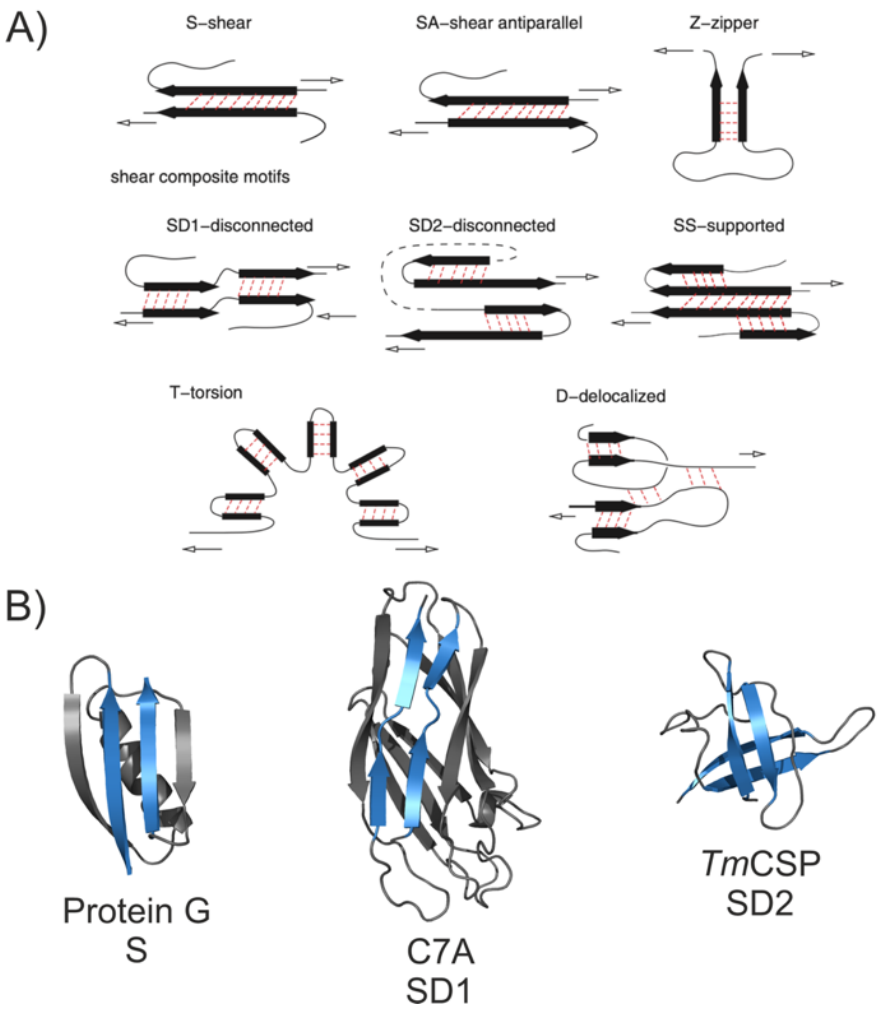

Figure 1.16 – Proteins can contain a force-bearing region, known as a mechanical clamp motif. A) Mechanical clamp motifs as described by Sikora and colleagues[91,92]. Black structure arrows and light arrows indicate β-strands generally (but can be α-helices – black rectangles) and force direction, respectively. Elementary shear motifs (e.g. S and SA) can combine to form shear composite motifs (e.g. SD1, SD2). B) Example structures that satisfy these defined mechanical clamp (blue) patterns, with the protein ID and corresponding clamp type below. Protein G PDB: 1PGA, C7A PDB: 1AOH and *Tm*CSP PDB: 1G6P. Figure adapted from reference[92].

The relative contact order (RCO), is defined as the average sequence distance between residues in contact normalised to the length of the protein[93]. Previously a correlation between the RCO and rupture force of the protein was suggested[94]. More recently, for a larger data size, it was shown that there was no clear trend for all proteins with increasing RCO and rupture force. However, when grouped by secondary structure there was a evident trend for increased mechanical stability for all-β proteins with increased RCO[95]. In the same article x_u values for 25 mechanically characterised proteins vs their corresponding rupture forces were plotted and grouped by their clamp motif revealing a non-linear fit (Figure 1.17). Although there does not appear to be sufficient number of

mechanically characterised proteins to accurately represent clamp groups, an interesting observation was proteins with SD1 motifs were found to exhibit a small x_u and a range of unfolding force values. This indicates a low level of malleability with versatility in mechanical stability. These initial observations are interesting and, as the bank of mechanically characterised proteins increases, this should be revisited.

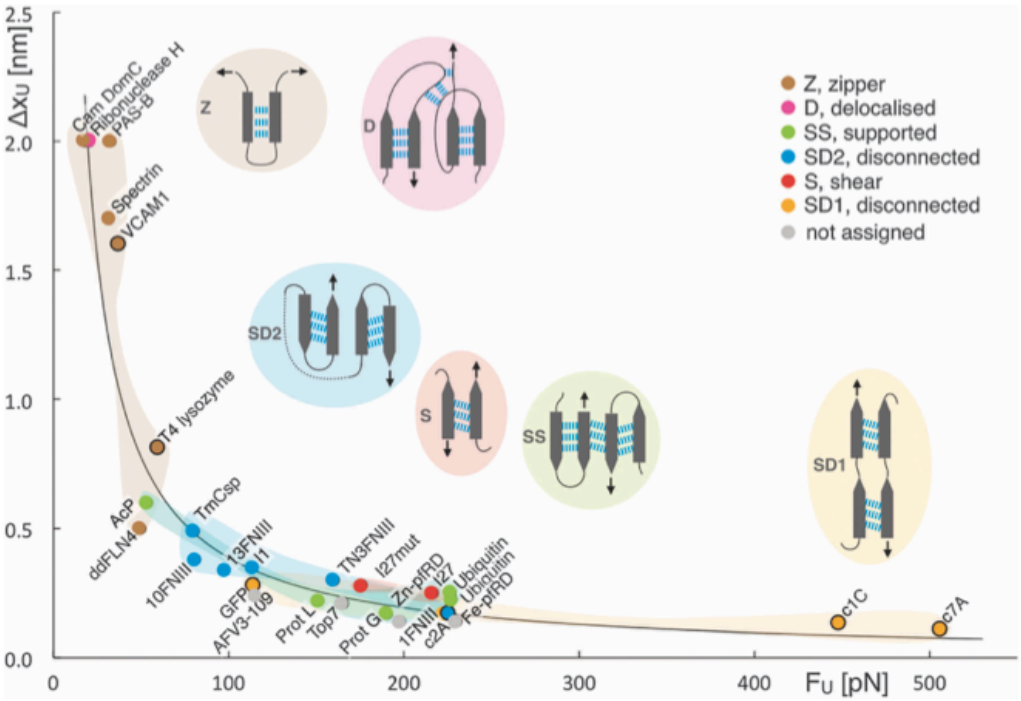

Figure 1.17 – The relationship between x_u and rupture force for 25 mechanically characterised proteins at 600 nms^{-1} grouped by their clamp motif. The data can be described by a non-linear fit following a power law with x_u = 39.4/F$_U$ with an R^2 = 0.91 (goodness of fit). Figure taken from reference[95].

Localised hydrogen bonded regions and their patterns between secondary structure elements have been determined to be responsible for the mechanical strength of proteins so far. However, proteins with a similar structure and containing the same mechanical clamp, but displaying different force responses cannot be explained solely by their hydrogen bonding patterns. Mutational studies[96–98] and simulations[70,85,99] have identified other interactions and non-native interactions important for the force response of proteins.

1.3.5 Point Mutagenesis to Probe Mechanical Stability

Mutagenesis by the method of protein engineering has been a powerful tool for elucidating mechanical clamps[90], non-covalent contributions to mechanical stability[96,97], metal ion-induced mechanostability[76], covalent bonds promoting inextensibility[100] and characterising intermediates[65]. The importance of the hydrogen-bonded 'mechanical clamp' region was first implied by steered molecular dynamics simulations of the I27 protein[89]. Subsequent mutagenesis-SMFS studies on I27 showed mechanical clamp variants to have altered force responses, but hydrophobic core variants showed little difference from wild type[90,101]. Example mechanical clamp mutants are displayed in Figure 1.18. However, TNfn3, the third fibronectin type III domain from human tenascin, shares the same β-sandwich structure consisting of two antiparallel β-sheets with I27, but hydrophobic core variants showed a decrease in mechanical stability[102]. This was further compounded when the mechanically weak fnIII domain, FNfn10, was engineered to have increased mechanical strength by replacing the hydrophobic core with the core of the homologous mechanically stronger fnIII domain, TNfn3[103]. The opposite was carried out, where the hydrophobic core of the mechanically weak domain (FNfn10) was grafted onto the strong fnIII domain, TNfn3, leading to a weaker FNfn10-like mechanical strength.

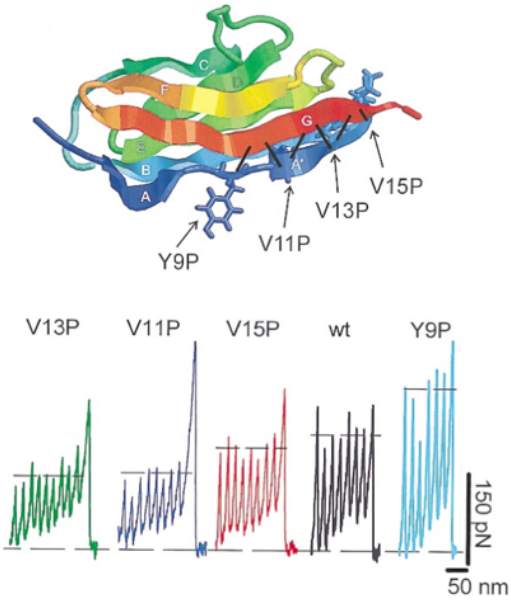

Figure 1.18 – Point mutations in I27 alter the mechanical stability. Structure of I27 displaying the hydrogen bonding between A' and G β-strands. FX profiles of the wild type and four proline substitution variants, three (V11P, V13P and V15P) of which displayed diminished mechanical strength and one (Y9P) showed increased mechanical stability. Figure taken from reference[90].

In the structurally similar proteins, Protein L and GB1 (Figure 1.19), the hydrophobic core has been observed to play a role in tailoring the mechanical strength of their hydrogen bonded clamps[96,97]. Their mechanical unfolding occurs by the shearing of one sub-domain motif from the rest of the structure. Sadler and colleagues identified a mechanical rheostat in the hydrophobic core of Protein L through mutational SMFS studies[97]. A single conservative hydrophobic volume reduction mutation, I60V, resulted in a 36 pN decrease at a retraction velocity of 447 nms^{-1} (wild type: 144 pN) in mechanical stability with little change in the thermodynamic stability of the protein. Two other mutants were created aiming to increase the mechanostability by enhancing hydrophobic contacts via enhancement of inter-sub-domain interactions and/or cavity filling of the hydrophobic core only. This led to an increased unfolding force of 76 (I60F) and 13 pN (L10F) at a retraction velocity of 447 nms^{-1}, respectively. These were both coupled with increased thermodynamic stability. Mutational SMFS studies on GB1 have also illustrated the importance of hydrophobic core packing in mechanostability[96]. Wang and Li observed that three mutations (F30L, Y45L and F52L) across this interface led to a significant decrease (50 – 90 pN) in the unfolding force of GB1 at a 400 nms^{-1} retraction velocity (wild type: 183 pN)[96]. It was determined that these mutations disrupted the hydrophobic core packing and in turn provided evidence that interactions mediated by hydrophobic residues at shearing interfaces are important for mechanical stability (Figure 1.19).

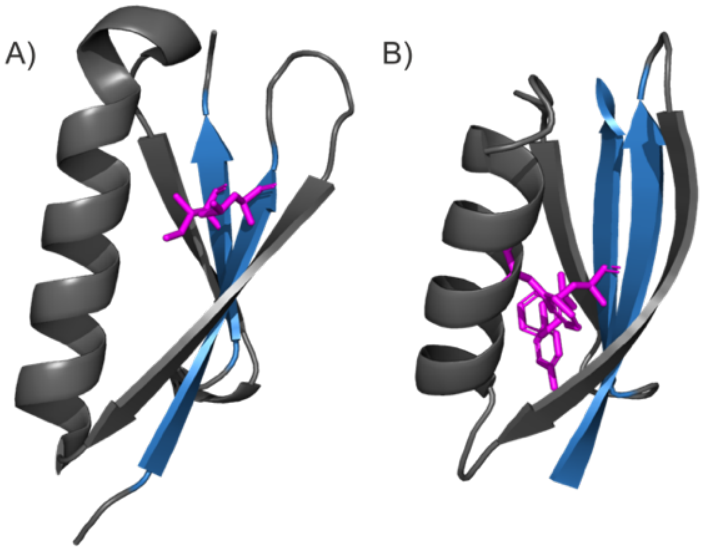

Figure 1.19 – The hydrophobic cores of Protein L and Protein G (GB1) are mechanical rheostats. Structures of A) Protein L (1HZ6) with the side chains of hydrophobic residues L10 and I60 utilised in reference[97] displayed and B) Protein G (1PGA) with the side chains of hydrophobic residues F30, Y45 and F52 utilised in reference[96]. Hydrophobic side chains mutated in these studies displayed (magenta). The shearing interface is coloured in blue.

Mutations have also been exploited to remove metal binding sites which can be important for mechanical stability[76]. Milles and colleagues created a number of variants affecting the three Ca^{2+} binding sites in the B domain of SdrG and observed a range of mechanical phenotypes. All phenotypes exhibited reduced unfolding forces compared to the wild type, and was suggested to be due to Ca^{2+} ions mediating electrostatic contacts between charged residues across the hydrogen bonded mechanical clamp which locks them in shear geometry. This agreed with the 'calcium clamp' hypothesis from Oude Vrielink and colleagues where calcium was observed to mechanostabilise the extender domains of *Mp*AFP and *Mh*Lap adhesion proteins from *Marinomonas primoryensis* and *Marinobacter hydrocarbonoclasticus*, respectively[104]. SdrG B domain mechanostability is covered in depth in Section 1.4.3.

1.3.5.1 Effects of Mutations on the Underlying Energy Landscape

The average rupture force can be plotted as a function of the logarithm of retraction velocity and should be described by a linear fit according to the Bell-Evans-Ritchie model (Section 1.3.3). The gradient of the linear fit helps inform us about basic features of the underlying 1D energy landscape of mechanical protein unfolding, such as the spontaneous protein unfolding rate at zero force (k_u^{0F}) and the distance between the folded and transition state in the mechanical unfolding reaction coordinate (x_u). MC simulations are commonly utilised to obtain these parameters from the experimental retraction velocity dependence on rupture force. The parameter x_u tells us how far the protein can be 'stretched' from its native structure before it unfolds (Figure 1.20) i.e. the malleability of the protein[95]. As the non-covalent forces that maintain a proteins structure are weak, most proteins are 'mechanically soft', meaning that a while the TS activation barrier can be relatively high, under moderate forces, the protein can be extended yet maintain a near-native like state[105]. If the external force were to be removed from the protein prior to an unfolding event (crossing the TS), the protein will likely return back to the native state. The malleability of a protein is simply the 'amount' the protein can be physically deformed without breaking or unfolding under mechanical force.

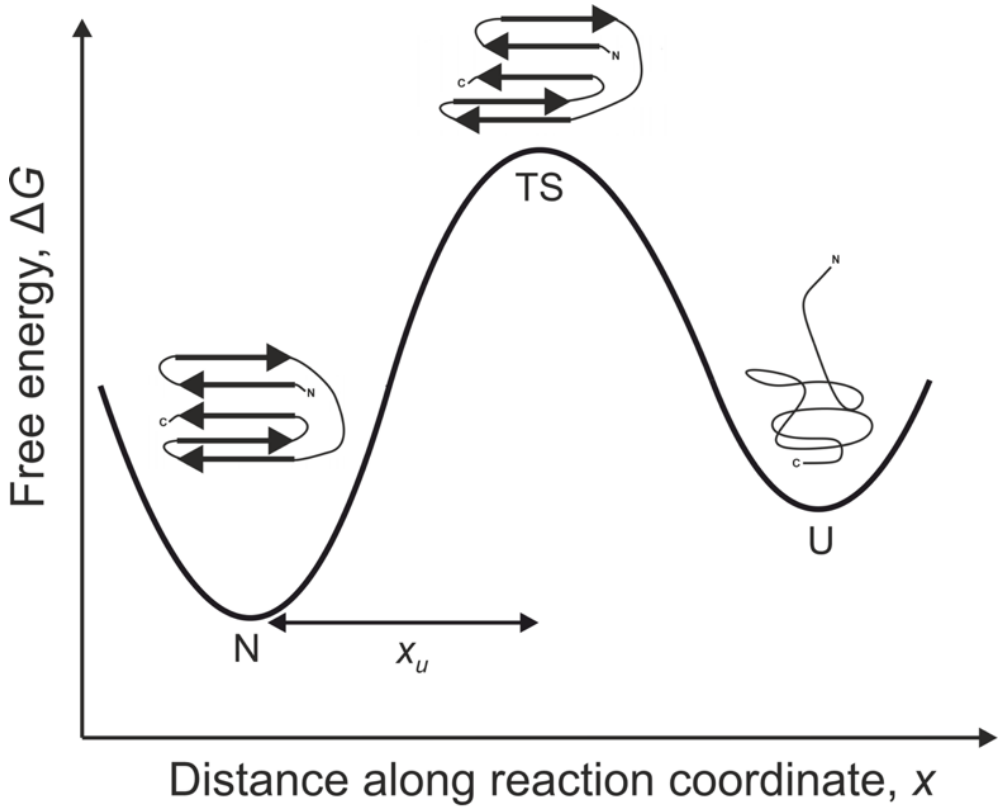

Figure 1.20 – Schematic of a simplified free energy diagram of mechanical protein unfolding displaying protein structures along the reaction coordinate. The x_u distance is a measure of the amount of deformation of the structure from native prior to the highest energy state observed prior to unfolding – the TS.

Mutations in protein structure may not only affect the mechanical stability of the protein, but also have the potential to change the relationship between the retraction velocity and unfolding force. This change in gradient reflects a change in x_u indicating the variant has become either more or less mechanically 'soft'[95]. A change in the 'mechanical softness' of a protein can be due to a change in temperature[106] or change in interaction patterns[98].

The importance of salt bridge networks for the mechanical softness of proteins have been revealed through point mutagenesis coupled with SMFS. Tych and colleagues mutated *Bacillus subtilis* Cold Shock Protein (*Bs*CSP) to create a charged triple mutant (CTM) mimicking the ionic network of the hyperthermophilic protein *Tm*CSP[98]. They observed an increase in mechanical softness of CTM compared to the wild-type *Bs*CSP (Figure 1.21). This study suggested longer range interactions, such as salt bridges, may contribute to the malleability of a protein under force.

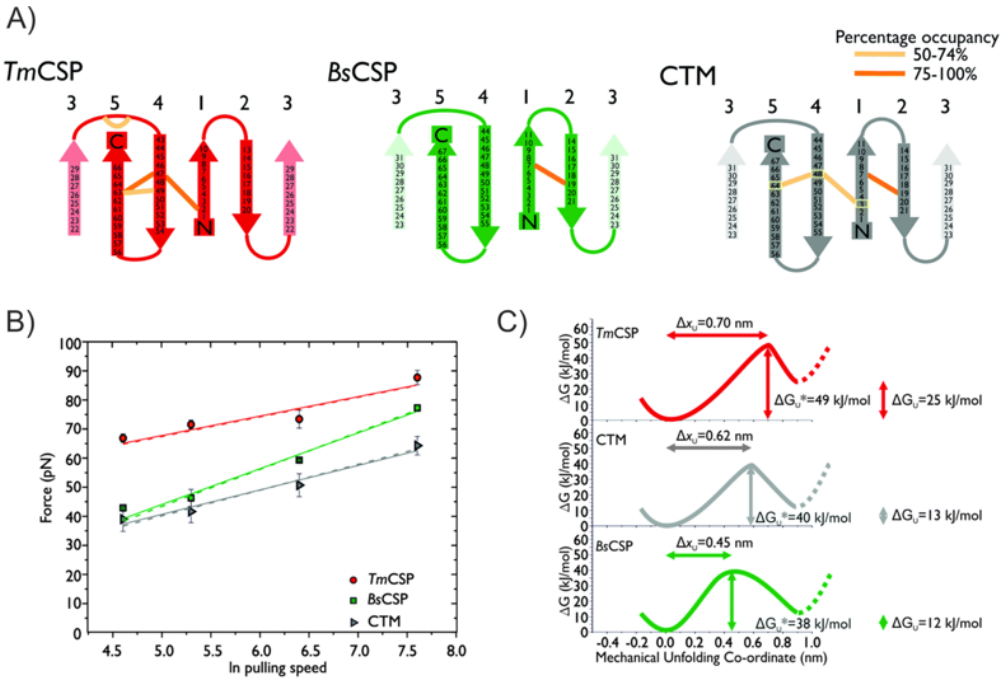

Figure 1.21 – Ionic interactions contribute to the mechanical properties of proteins. A) Topology diagram of *Tm*CSP (red), *Bs*CSP (green) and CTM (grey), with any salt bridges with an occupancy of > 50 % (orange) over five 200 ns simulations. The mutations in CTM are highlighted and were designed to 'graft' the ionic interactions of *Tm*CSP onto *Bs*CSP. B) Force plotted as the logarithm of pulling speed. C) An increase in mechanical softness was observed with the new grafted protein, CTM. Figure taken from reference[98].

1.3.5.2 Covalent Interactions and Mechanostability

In addition to non-covalent interactions, covalent interactions have been observed to play a major role in the mechanostability of proteins. The prototypical example is the *Streptococcus pyogenes* major pilin Spy0128, which comprises of two β-sheet sub domains (Figure 1.22) and assembles into a single linear homopolymer called pili. These pili have been determined to be crucial for mechanical process such as cell adhesion in colonisation and infection[107]. Utilising SMFS, the Spy0128 domain was shown to be inextensible by mechanical forces up to 800 pN[108]. This was postulated to arise from isopeptide bonds* across the hydrogen bonded mechanical clamps (Figure 1.22). Upon abrogating the formation of these isopeptide bonds through Lys to Ala mutations, both sub domains displayed unfolding forces of 200 - 250 pN at a retraction velocity of 400 nms^{-1}. Interestingly, isopeptide bonds have been found to form almost exclusively between the terminal

* Amide bond formed between the side chains of Lys and Asp/Asn residues

strands of pilin subunits across several bacterial species[109–111], which may suggest a ubiquitous and highly-evolved method for mechanostability.

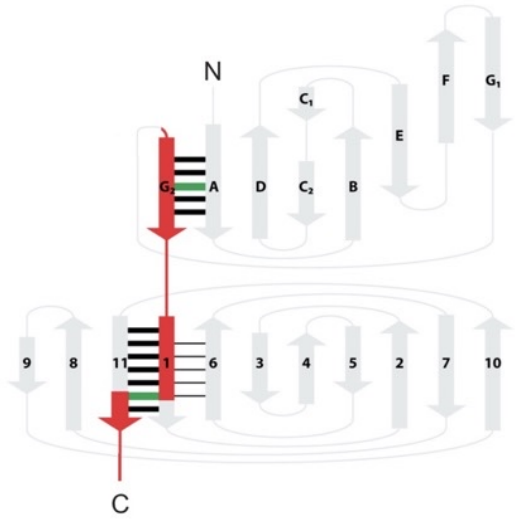

Figure 1.22 – Spy0128 is mechanically inextensible due to the formation of isopeptide bonds. Topology diagram of a Spy0128 domain consisting of two sub domains. Preventing formation of isopeptide bonds (green lines) results in both sub domains displaying typical unfolding behaviour of force-bearing proteins, probably due to the hydrogen bonded terminal strand mechanical clamps. Figure adapted from reference[108].

1.3.5.3 Interpreting Effects of Mutation on Mechanical Transition State Structure

Mutational analysis has proved very useful for characterising TS (ϕ-value analysis) along the protein folding pathway in ensemble experiments using chemical denaturants[112]. More specifically, ϕ-value analysis determines which native contacts are formed at the TS of protein folding. ϕ-value analysis has been adapted to determine the extent of native contact preservation at the TS in the mechanical unfolding pathway of mutants[113]. The ϕ-value equation for protein folding (ϕ_F^F) is given in Equation 1.6 and is defined as the ratio of change in the barrier height of the transition barrier of mechanical protein unfolding ($\Delta\Delta G_{TS-N}^{WT-MUT}$) to the change in free energy difference between the folded and unfolded state of the protein ($\Delta\Delta G_{UN}$):

$$\phi_F^F = 1 - \frac{\Delta\Delta G_{TS-N}^{WT-MUT}}{\Delta\Delta G_{UN}}$$

1.6

This allows the structure of the environment around the mutated residue at the transition state to be determined, with a value of 1 indicating the environment is fully native. This was initially applied to I27 using two mutants[113] which led to an extensive mechanical ϕ-value analysis revealing additional side chain interactions involved in the mechanical stability[101], in contrast to earlier work indicating the mechanical stability was entirely due to the hydrogen bonded clamp. Protein L and GB1 have also undergone mechanical ϕ-value analysis, which revealed partially structured transition states for the hydrophobic core mutant residues[96,97]. Unlike partial ϕ-values from bulk chemical denaturation experiments, mechanical ϕ-values utilise single-molecule experiments and thus have the advantage of partial values being interpreted as partial structuring only.

1.4 Utilising SMFS/SCFS to Understand The Mechanics of Biofilm Formation

1.4.1 *Staphylococcal* CWA Proteins and Host Attachment

During the initial stages of biofilm formation bacteria need to establish firm irreversible adhesion to the substrate in order not to be removed by external forces. The immense importance of this adhesion to host tissue (or other substrates) is reflected by the great number of adhesion-mediating surface proteins expressed by bacteria such as *S. aureus* and *Staphylococcus epidermidis* (*S. epidermidis*). *S. aureus* is decorated with about 20 CWA proteins[114] and *S. epidermidis* with about 12[115]. These CWA proteins are characterised by the presence of an N-terminal signal sequence, followed by a domain(s) for mediating receptor-ligand interactions (adhesin domain), then repeats of a varying number followed by a C-terminal sorting motif for covalently coupling the protein to peptidoglycan[114,115] (Figure 1.23). Non-covalently linked surface proteins may also mediate attachment[116], however, the covalently linked proteins have been more thoroughly investigated and will be discussed here. The proteins in these groups are mostly responsible for binding to human extracellular matrix (ECM) proteins (e.g. Fg and Fn) in the attachment stage of biofilm formation, however, the majority are multifunctional and may have a role in bacterial aggregation, immune evasion and nutrient acquisition[114]. As bacterial adhesion/aggregation events typically occur under external force, these will be focused on here.

Staphylococcal CWA proteins can be split into distinct structural and functional groups, including, but not limited to: MSCRAMMS, three-helical bundle, near iron transporter (NEAT) and G5-E repeat proteins. The MSCRAMM family is reserved for proteins which display the common structure motif (at least two adjacent IgG-like folds) and share a common dock, lock, latch (DLL) or collagen hug ligand binding mechanism[117]. Families of MSCRAMMS includes the clumping factor (Clf)-serine aspartate repeat (Sdr) family, Fibronectin binding proteins (FnBPs) and collagen adhesin (Cna). Family schematics are outlined in Figure 1.23.

Figure 1.23 – Schematics of the major *Staphylococci* families of CWA proteins. A) The MSCRAMMS family are shown at the top with the canonical N2 and N3 or N1 and N2 that engage ligands in the DLL or collagen-hug mechanism, respectively. Each family of MSCRAMMS contain B, SD, Fn binding tandem repeats or a mix of these. Other major families include B) NEAT motif, C) three-helical bundle and D) G5-E. The latter is covered in more depth in Section 1.4.3.1. Each protein has a signal sequence, a sorting signal for the covalent anchoring to the cell wall peptidoglycan and many have a wall-spanning region. A common theme is the presence of tandem repeats between the N-terminal ligand binding domain and the wall-spanning region. To the right of the schematics are *S. aureus* and *S. epidermidis* examples in standard and italicised font, respectively. Figure adapted from reference[117].

Recently, AFM studies have provided compelling evidence that *Staphylococcal* species have evolved MSCRAMMS to tune cell adhesion via external mechanical cues. A prototypical example is the *S. epidermidis* SdrG protein (part of Clf-Sdr family), which binds to Fg via a dock, lock and latch (DLL) mechanism mediated by the A domain[118] (Figure 1.24). It involves initial docking of the Fg via β-strand complementation with the N3 G' strand in the binding cleft – a trench created by the orientation of the N2 and N3 sub domains[118]. Following this, a structural rearrangement at the C-terminus of the N3 sub domain permits residue interaction with the Fg firmly locking it in place. Finally, the protein-ligand complex is latched by β-strand complementation of the newly formed N3 C-terminal G'' β-strand with the N2 E β-strand.

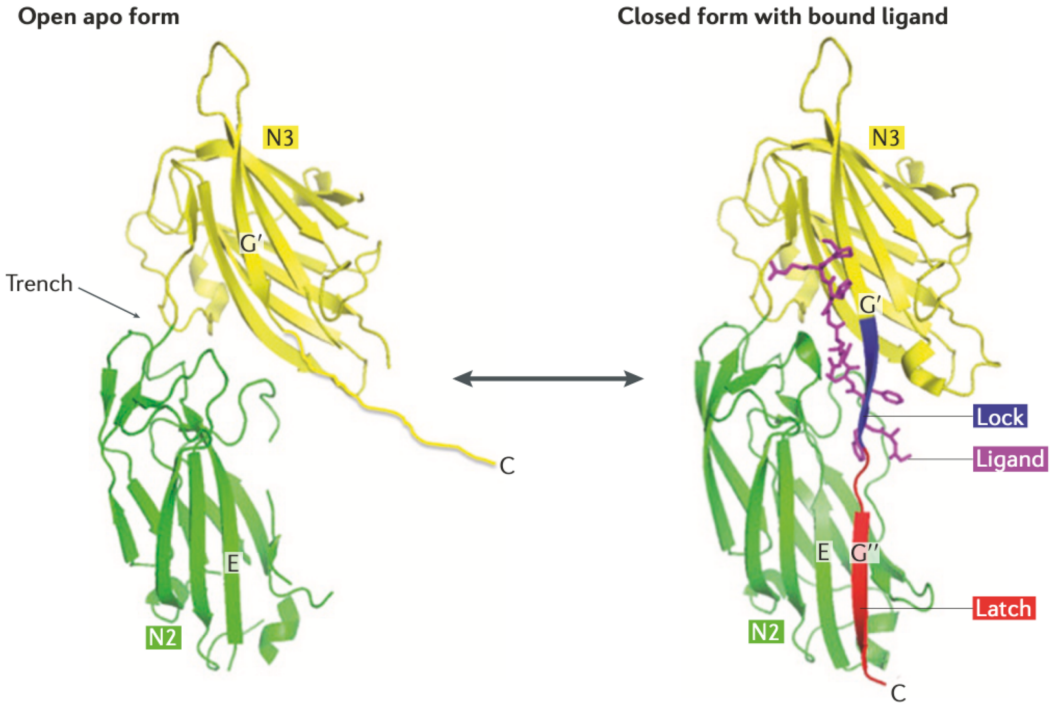

Figure 1.24 – The DLL mechanism. The A region of the MSCRAMM Sdr-Clf family, has a wide trench between N2 and N3 in its apo form. The ligand (e.g. Fg) docks into this trench, prompting conformational changes to lock the ligand in. In the apo form the N3 sub domain displays a disordered C-termini, however, after ligand binding, this region forms the lock (blue) and latch (red), forming a strong ligand-protein complex in its closed form. Figure taken from reference[117].

SCFS (cell probes vs. Fg-functionalised substrates) and SMFS (Fg-functionalised cantilever tips vs. bacterial cells) have been utilised to investigate the SdrG:Fg β-chain interaction in conditions similar to those found *in vivo* (Figure 1.25B & C)[119]. They observed extremely strong adhesion forces of ~ 2 nN at a constant retraction velocity of 1000 nms^{-1}. This incredible mechanical strength

of the ligand-receptor complex arises from the DLL mechanism itself. *In silico* methods showed that when the SdrG:Fg complex is under mechanical stress, the load is distributed over all hydrogen bonds between the backbone of Fg β-chain and the locking strand (confined in a 'corkscrew' geometry) (Figure 1.25D & E), requiring all hydrogen bonds to be broken simultaneously[120]. This is reminiscent of the 'shear' geometry observed in hydrogen bonded terminal mechanical clamps of mechanostable proteins (see Section 1.3.4.2).

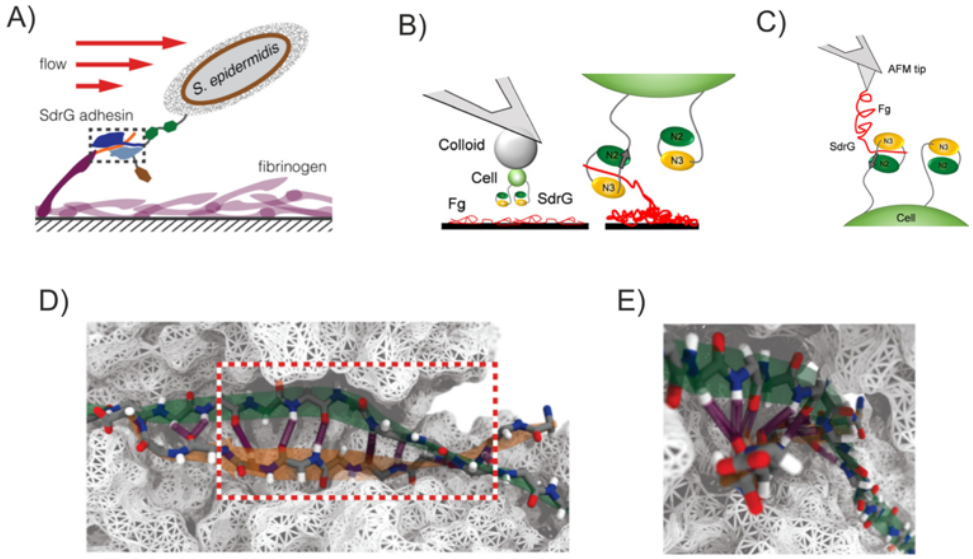

Figure 1.25 – The SdrG:Fg complex achieves its remarkable mechanostability under enormous forces through 'shear' geometry of its ligand. A) SdrG functions to prevent detachment of *S. epidermidis* under high shear flow through forming a complex with the N-terminal peptide of Fg (purple) β-chain (orange). B) *In vivo* SCFS (cell probes vs. Fg-functionalised substrates) and C) *in vitro* SMFS (Fg-functionalised cantilever tips vs. bacterial cells) experimental set ups. D) Hydrogen bonds (purple) contacts with the backbone of Fg β-chain (orange) and locking strand (green) under external load *in silico*. E) Fg β-chain confined in a coiled geometry, similar to a corkscrew in a cork. Peptide back bones are shown as sticks in D) and E). B) and C) taken from reference[119] and A), D) and E) taken from reference[120].

Cna, expressed by *S. aureus,* binds to collagen via it's N-terminal A domain through a mechanism coined the 'collagen hug'[121]. The 'collagen hug' is a variation of the DLL ligand-binding scheme[118]: ligand docking, locking and complex stabilisation by latching. However, the collagen hug utilises the N1 and N2 IgG-like sub domains of Cna, which have corresponding roles to the N2 and N3 sub domains in the DLL mechanism, with the N2 C-terminal extension forming the latch. The long linking peptide between N1 and N2 sub domains creates a hole at the interface for accommodating the monomeric collagen triple-helix rod ligand. The ligand docks into the trench between these sub

domains, prompting a conformational change in the linker locking the collagen in place. This differs from DLL in that it is the linker, not the C-terminal residues from the N3 domain, that locks in the ligand. Following the grasping of collagen in the trench, the ligand is latched in place by β-strand complementation of the C-terminal linker latch of N2 to N1 securing the ligand in place and forming a stable complex. Cna:collagen bonds *in situ* are very strong (~ 1.2 nN with a retraction velocity of 1000 nms^{-1})[122]. These large forces were in the range of those measured for the DLL-based SdrG:Fg interactions *in vivo*[119] implying a strong interaction formed by the collagen hug binding mechanism as observed with DLL. The strong mechanostability of Cna:collagen, may aid the protein in retaining its adhesive function under high shear stresses during initial biofilm stages.

1.4.2 Initial Host Attachment – Holding on Tight(er)

External forces tend to usually prevent bacterial surface attachment and contribute to the removal of bacteria from surfaces[123], however, evidence is growing that external forces can actually mechanoregulate interactions between MSCRAMMS and their cognate ligands. Typically, a receptor ligand bond lifetime decreases with an increase of force[66,67,124]. This is a force-inhibited interaction coined a 'slip bond', as the ligand should slip out of the binding pocket more readily under higher tensile force[124] (Figure 1.26B). These bonds are strongest at the low shear flows, where the applied force is the weakest.

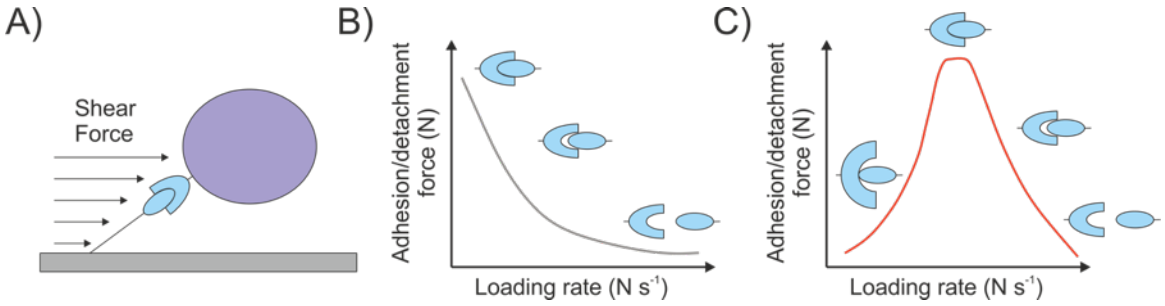

Figure 1.26 – Schematic of shear force and its effect on slip- and catch-bonds. A) Schematic presentation of shear force on a protein:ligand complex on an adhering cell (purple). B) Slip-bonds become weaker/lifetime decreases as the loading rate increases, as the ligand 'slips out' of the binding pocket. C) Catch-bonds increase in strength/lifetime as the loading rate increases, usually due to a force-induced change in their conformation. At a certain loading rate threshold they begin to act as a slip-bond. Reproduced and adapted from reference[125].

However, shear-enhancement of *E. coli* mannose-specific type 1 fimbria (adhesin domain FimH) binding to guinea pig red blood cells in flow chamber experiments[126] and shear stress dependent stick-and-roll adhesion[127] were observed in the early 2000s. This so called catch-bond behaviour was first described by Dembo and colleagues in 1988 and was purely theoretical[124]. It describes the strengthening of a bond in the presence of increasing external mechanical force, until a threshold is reached and then it switches to a slip bond regime[125] (Figure 1.26C). Shortly after FimH force-enhanced adhesion was uncovered, the first molecular detailed evidence of a catch-bond was observed by SMFS studies measuring the response of individual receptor-ligand bonds to tensile force[128]. AFM spectroscopy was utilised to dissociate a bond between P-selectin (on endothelial cells) and PSGL-1 (on leukocyte surfaces), by pulling them away from one another. As the force was increased, the lifetime of the bond followed suit, until the force was excessive and bond dissociation occurred. These initial studies provided solid evidence for the existence of catch-bonds and prompted the search for other catch-bond behaviour in nature.

In addition to the high mechanical strength of the MSCRAMMS-mediated bonds, several *Staphylococcal* MSCRAMMS have been determined to exhibit force-sensitive catch-bond behaviour through SMFS/SCFS studies. A well-documented example is the ClfA:Fg interaction, which was observed to be a force-activated molecular switch by SCFS and SMFS AFM techniques (Figure 1.27A & B)[129].

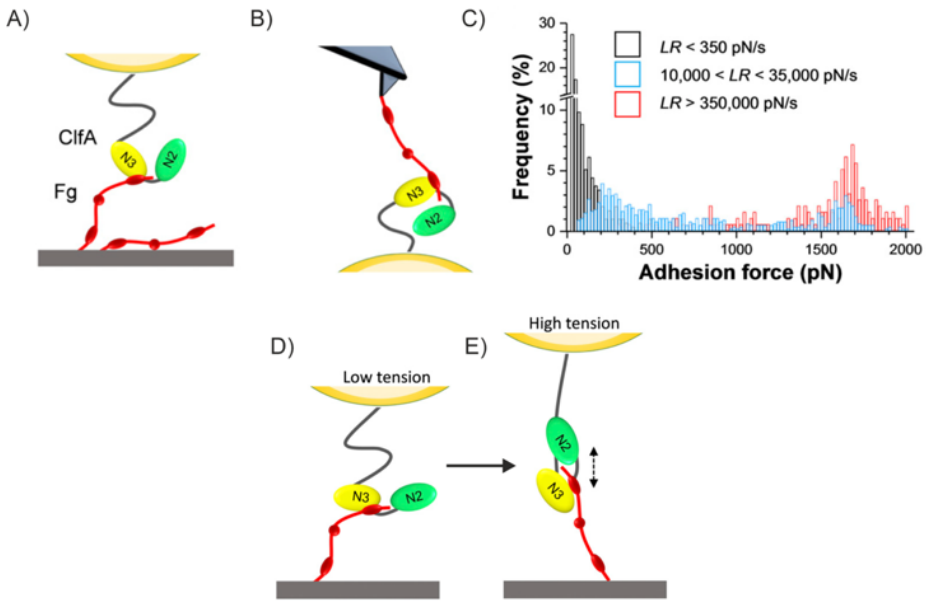

Figure 1.27 – Force nanoscopy of the ClfA:Fg interaction reveals a catch-bond mechanism. Schematic presentation of A) SCFS B) SMFS experimental set ups. C) Discrete ranges of loading rates (LRs) were binned and force distributions plotted as histograms revealed a switch in force response from low (< 350 pN s^{-1}) to high (> 350000 pN s^{-1}) loading rates, respectively. Herman-Bausier and colleagues predicted that at D) low tension Fg weakly binds to the top of the ClfA N3 domain and upon E) high tension, there is an extension and conformational changes of the N2 and N3 sub domains that enable the Fg γ-chain to dock and form strong interactions. Adapted from reference[129].

At low loading rates (low mechanical shear), weak bonds (~ 100 pN) dominated, whilst at higher loading rates (high mechanical shear), stronger bonds (~ 1,500 pN) were favoured (Figure 1.27C). This demonstrates that as force is loaded quickly (high loading rate), the strength of the ClfA:Fg interaction is dramatically enhanced. The increased loading rate is coupled with the observation in a major switch in force distribution rate and rare cases of intermediate forces indicates that the strong bonds are not resulting from the simultaneous rupture of multiple weak interactions. It was postulated that the shear-enhanced strengthening of the ClfA:Fg bond is due to a force-induced conformational change in ClfA, from a weak- to a strong-binding state (Figure 1.27D & E). At low shear forces, Fg binds to the top of the ClfA N3 domain forming weak contacts. At high shear forces conformational changes trigger the DLL mechanism, resulting in a bond strength in the covalent range[130]. Importantly, 'activation' of these strong interactions generally occurs at loading rates greater than 10,000 pN s^{-1}. This is physiologically relevant as bacteria are typically exposed to high shear rates corresponding to 100,000 pN s^{-1} and above[131].

As mentioned, Herman-Bausier and colleagues demonstrated the strength of the Cna:collagen bonds *in vivo* are very strong (~ 1.2 nN)[122]. In the same article, recombinant fragments of the ligand-binding region (N1 and N2 only) displayed moderate binding strength (218 ± 86 pN) (Figure 1.28A) and recombinant (CNA_{31-44}) ligand-binding region (N1, N2 and N3) displayed similar rupture forces (195 ± 108 pN) at a retraction velocity of 1000 nms^{-1}. The Cna:collagen interaction between living bacteria and collagen functionalised cantilever tips displayed single adhesion peaks providing dual force distributions of weak forces (239 ± 107 pN) and strong forces (1166 ± 544 pN) at a retraction velocity of 1000 nms^{-1} (Figure 1.28B). These former bond strengths were in the range of the recombinant CNA_{31-44} binding strengths (~ 220 pN), suggesting they represent the initial hydrophobic interactions between collagen and the residues in the binding trench of the N2 domain. The larger forces were in the range of those measured for the DLL-based SdrG:Fg interactions *in situ*[119] implying a strong interaction formed by the collagen hug binding mechanism as seen with DLL and reliance on the presence of the B region 'stalk'. If the force response was plotted against the loading rate, they may have observed two populations as we have seen with ClfA[129].

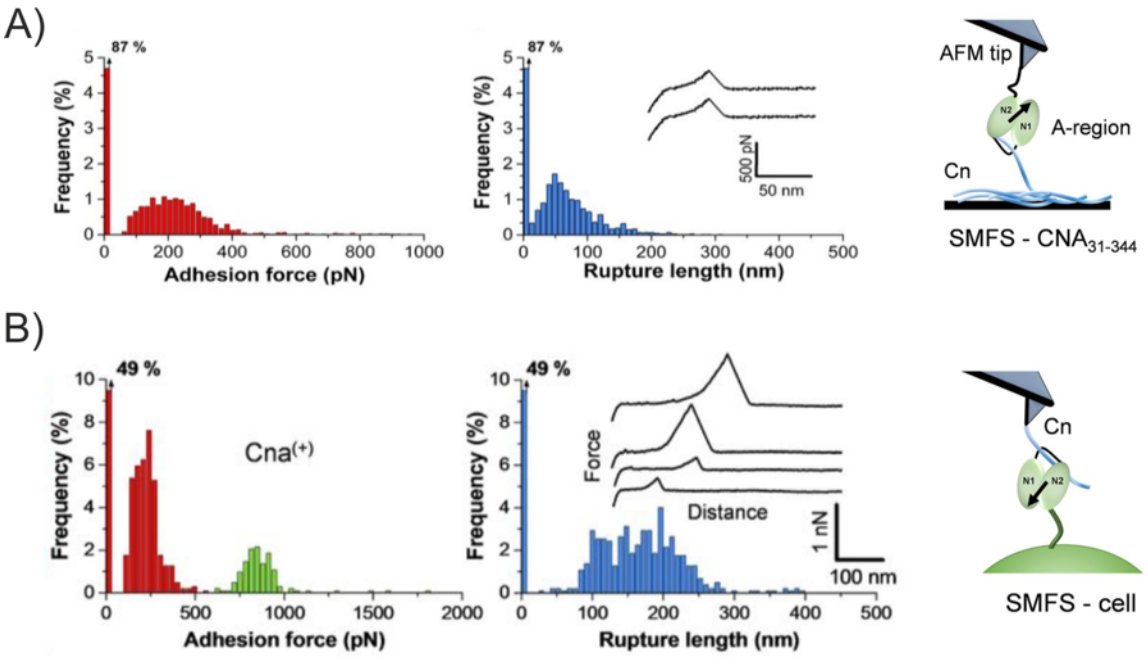

Figure 1.28 – AFM spectroscopy analysis of the Cna:collagen interaction. A) SMFS utilising recombinant CNA_{31-344} displays a unimodal low force response (one repeat displayed). B) SMFS with collagen modified tips with *S. aureus* bacteria expressing full-length Cna ($Cna^{(+)}$) also displayed a bimodal force response (one repeat displayed). N3 is not displayed in the schematics. Adapted from reference[122].

Furthermore, the presence of the B domains in these latter experiments indicates a potential role in the function of strong ligand binding. Interestingly, the FX profiles of the strong force peaks deviated from a WLC model fit (Figure 1.29A). Instead the forces were proportional to distance suggesting linear Hookean spring behaviour. The k_m (stiffness of the molecular complex) was calculated using the following Equation 1.7:

$$k_m = (k \times s)/(1 - s)$$

1.7

Where k is the cantilever spring constant and s is the slope of the linear portion of the raw deflection verses piezo displacement curves.

They noted that the peaks with the strongest forces typically corresponded with the highest k_m values and suggested the B region may stretch and contract (like a nanospring) from the cell wall and modulate the adhesive function of Cna:collagen bonding (Figure 1.29B). The direct correlation between occurrence of strong bonds and spring behaviour implies the B domain plays a functional role as a mechanosensor capable of detecting mechanical forces acting on the cell wall and inducing a stronger Cna:collagen interaction.

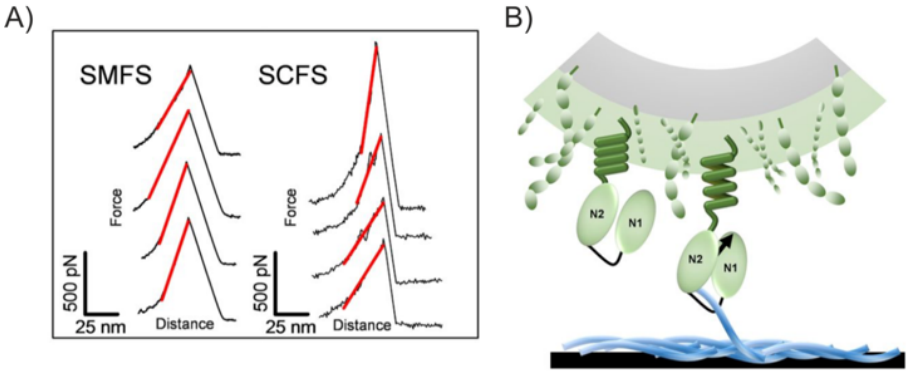

Figure 1.29 – Deviation from the WLC suggests nanospring properties of Cna. A) Representative FX profiles displaying adhesion peaks with a Hookean spring behaviour. B) The B region of Cna may stretch and contract from the cell wall and act as a mechanosensor in turn modulating the function of the adhesin domain. Figure taken from reference[122].

Another interesting example is the CWA LPXTG protein ClfB, part of the same *Staphylococcal* MSCRAMM family as ClfA. Loricrin (Lor), a squamous epithelial cell envelope protein, binds to ClfB via the DLL mechanism[132]. Utilising SCFS (Figure 1.30A) and SMFS (Figure 1.30B), AFM techniques to interrogate the ClfB:Lor interaction, Vitry and colleagues observed enhancement in strength of the ClfB:Lor interaction under increased loading rate[133].

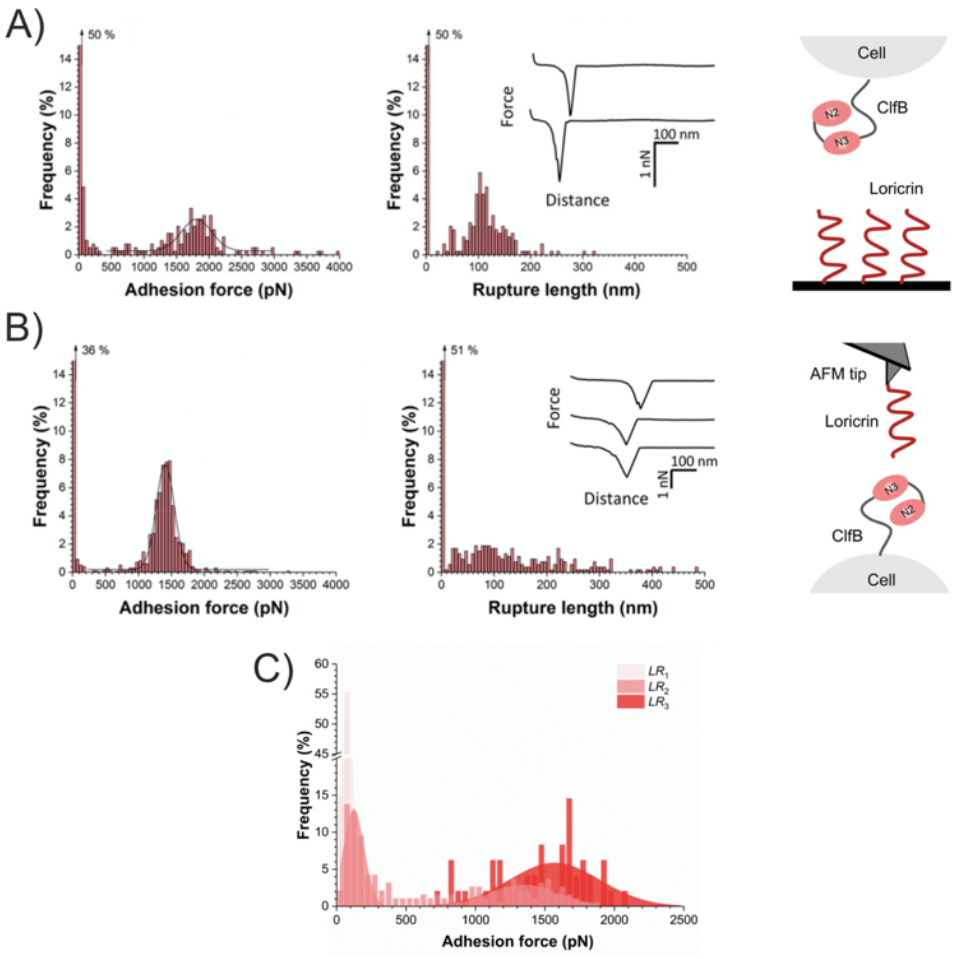

Figure 1.30 – Mechanical force activates the ClfB:Lor interaction. Analysis of the ClfB:Lor interaction by A) SCFS and B) SMFS with force- (left) and rupture length-frequency (middle) histograms (with typical FX profiles shown inset) and schematics displayed (right). C) Force distributions plotted as histograms with highlighted ranges of loading rates ($LR_1 < 350$ pNs^{-1}; $3500 < LR_2 < 10000$ pNs^{-1}; $LR_3 > 35000$ pNs^{-1}). As the loading rates increase, the force distribution is switched, with the probability of forming stronger interactions increasing. Figure adapted from reference[133].

They observed a bimodal force distribution of weak bonds (~ 250 pN) and strong bonds (~ 1,500 pN), occurring at low (< 350 pN s^{-1}) and high loading rates (> 35,000 pN s^{-1}), respectively (Figure 1.30C). As the weak and strong bonds differ by an order of magnitude, and are within the range of

other DLL bonds[129], the strong bonds were determined to not be due to the simultaneous rupture of multiple weak interactions. As with Cna, they explicitly observed weak interactions with AFM tips functionalised with recombinant $ClfB_{N2N3}$ (absent B domain) and Lor substrate by SMFS. In an attempt to determine the molecular origin of this switch in force distribution the molecular stiffness (k_m) of the molecular complex was analysed. Like the previous Cna:collagen example, the k_m values were also observed to increase with increasing rupture forces. The k_m values for both the weak and strong forces were 3.5 $\pm$ 0.1 pN nm^{-1} and 18 $\pm$ 1 pN nm^{-1}, respectively. This implies that stronger forces were associated with an increased molecular stiffness of the complex which is probably due to a change in the conformational state of ClfB. Supporting this notion, the modelling of the adhesion force vs k_m plot (Figure 1.31) using the WLC model, revealed low force level k_m values are consistent with ClfB molecular elasticity, whereas k_m values at a high force level reflect the elasticity of two springs in series; the ClfB protein and cell wall[134]. This data favours a two-state conformation change model in which the ClfB protein switches from a weak-binding to a strong-binding state (activating the DLL binding mechanism) under tensile force.

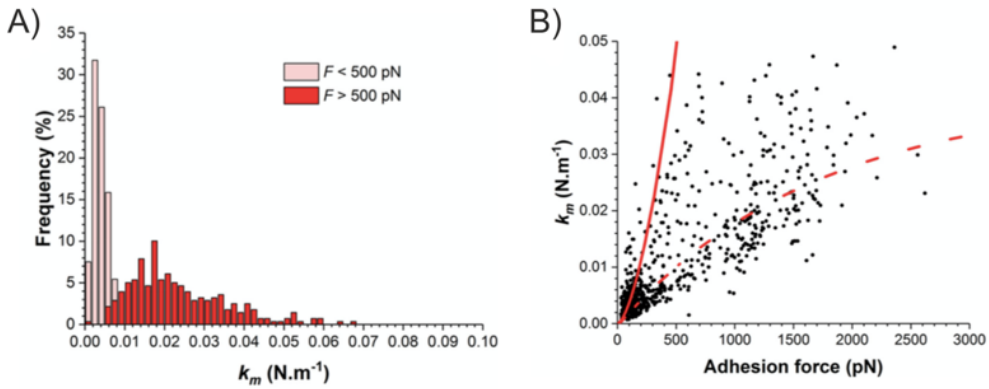

Figure 1.31 – Switch in bond strength relates to an increase in molecular stiffness. (A) Distribution of the spring constants of the molecular complex (k_m) at low (< 500 pN) and high (> 500 pN) force levels. (B) Force versus k_m plot, coupled with WLC model taking into account the elasticity of the ClfB protein only (continuous line) and that of the ClfB protein and the cell wall (dashed line). Figure taken from reference[133].

Other *Staphylococcal* surface proteins displaying force-enhanced catch-bond behaviour and/or mechanostable properties resolved using SMFS or SCFS techniques are listed in Table 1.4.

Staphyloccocal species	Protein:Ligand	Rupture Force (pN)	Technique	Reference
S. pseudintermedius	SpsD:Fg	~ 1800	SCFS	135
S. aureus	IsdB:Vitronectin	~ 150 (low LR) ~ 2000 (high LR)	SMFS	136
S. aureus	SpA:vWF	~ 100 (low LR) ~ 1700 (high LR)	SMFS	137
S. epidermidis	Aap:vWF	~ 300-400 (low LR) ~ 3000 (high LR)	SMFS	75
S. aureus	ClfA:vWF	~ 2000	SCFS	138

Table 1.4 – Examples of *Staphylococcal* proteins and their mechanostable/mechanostable adhesion properties resolved using SMFS or SCFS. Loading rate: LR. All experiments were carried out at a retraction velocity of 1000 nms^{-1}.

1.4.3 Holding Tight – Resisting or Controlled Deformation?

As the adhesin domains of *Staphylococcal* CWA proteins bind with a force within the range of a covalent bond[130], the remaining structure must also be capable of withstanding the extreme mechanical forces enabled by the DLL and collagen-hug binding mechanisms. The proteins involved in adhesion comprise of an N-terminal adhesin domain (involved in DLL/collagen-hug mechanisms) which is usually followed by tandemly arrayed repeats of similar or near-identical domains forming a 'stalk', which is believed to project the adhesin domain into the extracellular space and can be involved in bacterial aggregation. The role of this 'stalk' of domains in pathogen adhesion remains elusive, with suggested roles such as shock dissipators[74] and extendable springs that stretch and contract[139,140].

A prototypical example is the ability of Cna-collagen to withstand high forces (~ 1.2 nN) without unfolding of the spring-like B domain[122]. The B domain of Cna is composed of 1-4 B-repeats (split into two domains with similar secondary structure) depending on the strain[141]. Each domain is predominantly β-sheet in structure with a unique fold that resembles, but is the inverse of, the IgG fold (Figure 1.32). As discussed in Section 1.3.4.1, both β-sheet and Ig-like structure is predominantly associated with mechanically strong proteins, however, forces up to 450 pN are usually sufficient to forcibly unfold these[31]. Like the Cna B-repeat, the major pilin Spy0128 (from *Streptococcus pyogenes*) is mechanically inextensible up to 800 pN by SMFS[108]. This mechanical resilience was found to arise from intramolecular isopeptide bonds which, upon knocking out

became mechanically labile (as discussed in Section 1.3.5.2). As there is evidence of internal isopeptide bonds forming within inverse Ig-like folds (similar to Cna B-repeat sub domains) of gram positive bacterial pili[109,110], Herman-Bausier and colleagues predicted the mechanical inextensibility of the Cna B repeats may result from intramolecular isopeptide bond networks[122]. The location of these predicted isopeptide bonds are displayed in Figure 1.32. The inextensibility of the B region aids the adhesive function of Cna under conditions of high physiological shear stress, thus increasing the chance of colonisation.

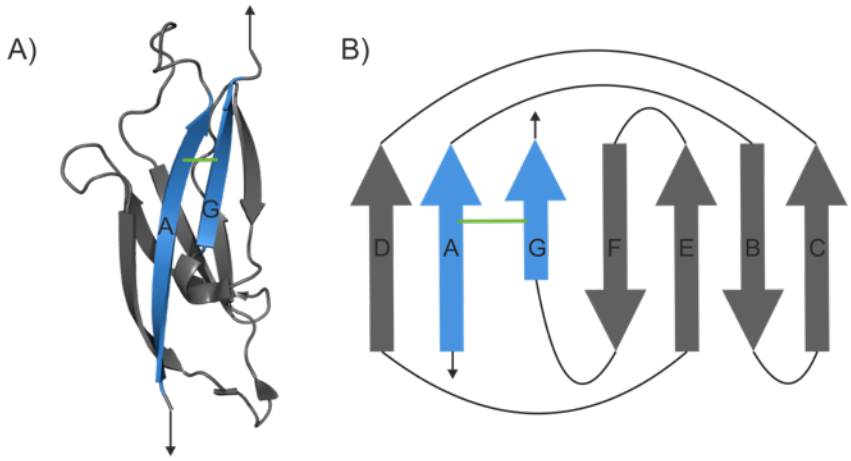

Figure 1.32 – Mechanical inextensibility of the Cna B repeat domain is postulated to arise from an isopeptide bond. A) Structure of 1 Cna B repeat domain, with terminal β-strands (A and G) in blue and direction of force labelled with a solid arrow. The structure is reminiscent to the IgG fold, but the inverse orientation (PDB: 1D2O). B) Topology diagram of a single CnaB repeat with the predicted isopeptide bond (green) locking the terminal β-strands, preventing rupture. Arrows represent β-sheets.

The B domains (B1 and B2) from the *S. epidermidis* SdrG act as a linker between the N-terminal A adhesin domain (involved in the mechanostable DLL mechanism) and the cell-wall anchored C-terminus (Figure 1.33A). Interestingly, the B domains of SdrG have evolved exceptional mechanical stability displaying unfolding forces of > 2 nN at retraction velocities of 400 nms^{-1} upwards[76], similar to that of covalent bond breaking[130]. Milles and colleagues utilised an SMFS system to test the strength of the B1 domain (Figure 1.33B) and observed large rupture events corresponding to the unfolding of the B1 sub domain prior to the dissociation of the adhesin domain-ligand interaction (Figure 1.33C). Interestingly, a second population unfolding at 500 - 800 pN was detected. The addition of ethylenediaminetetraacetic acid (EDTA) caused the lower unfolding population to appear exclusively, suggesting a potential role of metal ions in the extreme mechanostability (Figure

1.33D). Rupture force events > 2 nN were recovered upon the addition of Ca^{2+} ions. It is speculated that the coordination of these Ca^{2+} ions promotes the high mechanostability through electrostatically protecting the hydrogen bonds from breaking and locking them in a shear geometry or serving as a network to divert propagating forces away from the mechanical clamp. In contrast to the Cna B-repeats inextensibility and spring properties, SdrG B-repeats unfold at a similar force to their DLL complex unfolding forces precluding them from being reliable 'shock dissipaters', so their exact role remains unclear. However, it is clear that they are extremely mechanostable, which will help prevent protein perturbation during host attachment and increase the chance of a successful colonisation.

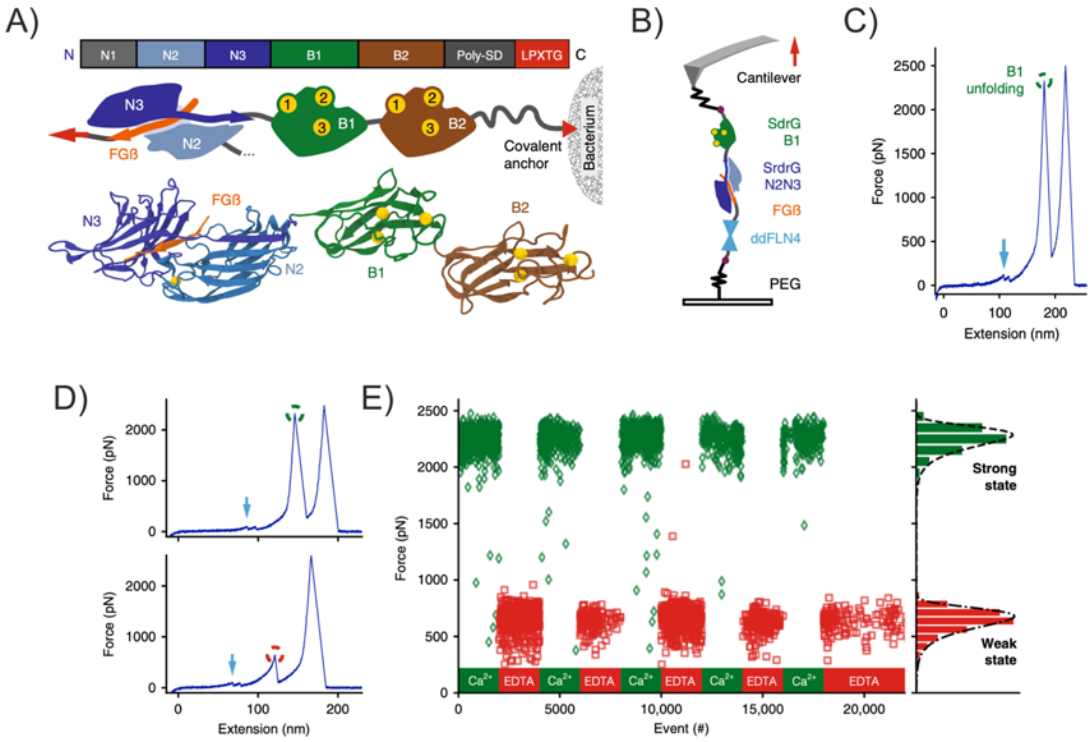

Figure 1.33 – Calcium stabilises the strongest protein fold observed to date.Schematic of SdrG from *S. epidermidis* (top). Schematic of SdrG binding Fg-β strand and the Ca^{2+} coordination sites are displayed in yellow (middle). Cartoon structure of SdrG A, B1 and B2 domains, with Ca^{2+} ions displayed as yellow spheres (bottom). B) SMFS set up utilised in these experiments to probe B1 mechanical strength. C) Typical FX profile of ddFLN4 marker domains unfolding (blue arrow) and B1 unfolding (green broken circle). D) Typical FX profiles in the presence of Ca^{2+} ions (top) and presence of EDTA (bottom) displaying higher and lower mechanical stabilities, respectively. E) Alternate application of Ca^{2+} ions (green diamonds) and EDTA (red diamonds) allows the stronger and weaker stabilities of B1 to be switched between repeatedly. Figure adapted from reference[76].

1.4.3.1 'Periscope Proteins'

'Periscope proteins' are bacterial proteins formed from tandem arrays of highly similar domains to form elongated rods of varying length, which typically serve to project a functional domain distal to the cell surface[142]. Both SasG and Aap, of *S. aureus* and *S. epidermidis*, respectively, are examples of 'periscope proteins' and comprise an A domain promoting attachment to abiotic surfaces[143–145] followed by tandem arrays of highly homologous B-repeats suggested to be involved in Zn^{2+}-mediated bacterial aggregation during biofilm formation[20,144,146] (Figure 1.34). For these proteins to switch from surface attachment to aggregation during biofilm formation, their A domains must be proteolytically cleaved to leave the B-repeats exposed[144,146] to mediate *trans*[*] interactions (Figure 1.34D). In addition to promoting intercellular bonding, Aap in solution has been shown to form amyloid fibers catalysed by Zn^{2+} (Figure 1.34E), which may contribute to biofilm matrix properties[147]. As SasG has the ability to populate the *S. aureus* cell wall at an intermediate to highly dense level[144], SasG will likely be involved in *cis*[**] interactions (Figure 1.34D) due to the proximity of monomers to one another. *Cis* interactions are understood to occur in protein:receptor pairs and have biological roles *in vivo*[148], however, it is unknown whether SasG or Aap form Zn^{2+}-mediated *cis* homophilic interactions *in vivo* and what the benefit of these might be (discussed further in Section 3.3.2).

[*] Interactions between two molecules, expressed on two different cells

[**] Interactions between two molecules expressed on the same cell

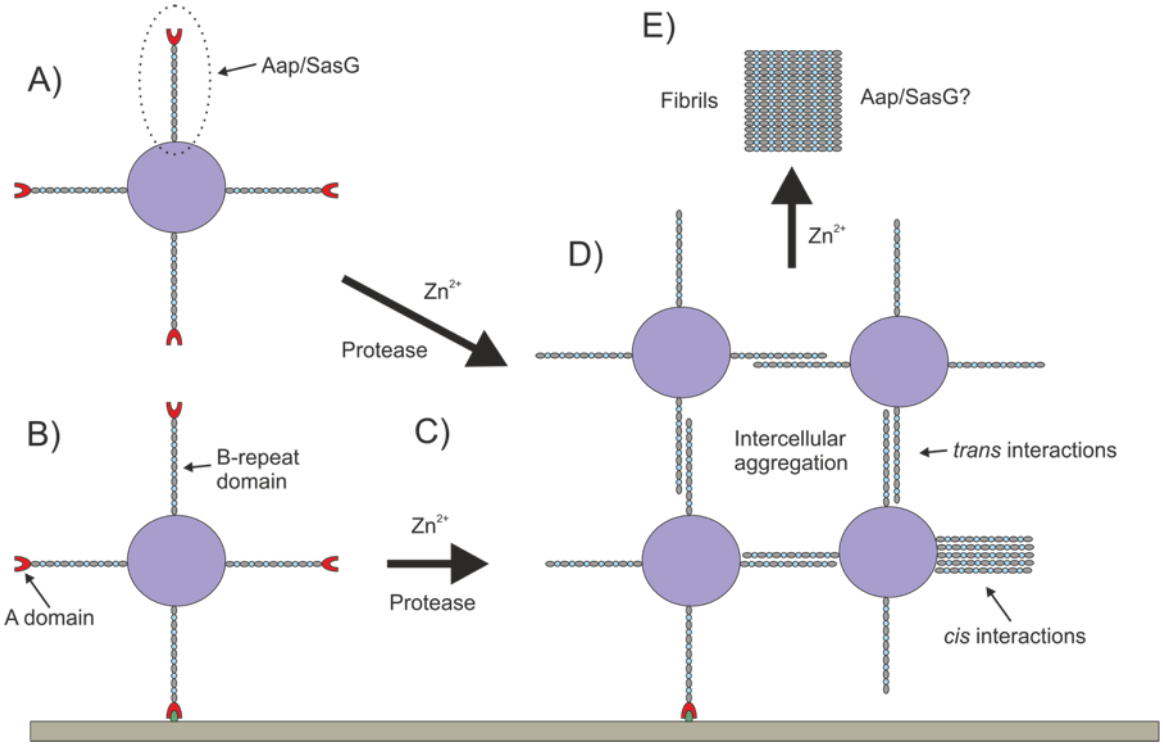

Figure 1.34 – Periscope proteins Aap/SasG promote biofilm formation. Both Aap and SasG are involved in biofilm formation and may take one of two pathways. Aap/SasG will be proteolytically processed (removal of the A domain) in a A) non-bound or B) host-cell attached form. C) Post-proteolysis and in the presence of Zn^{2+}, Aap/SasG may D) promote intercellular aggregation through *trans* homophilic binding events. SasG/Aap may also form *cis* interactions in the presence of Zn^{2+}. E) Aap in solution has been shown to promote fibril formation in the presence of Zn^{2+}.

The N-terminal SasG A domain comprises a unique 157-residue sub domain and a conserved 212-residue sub domain with 59 % sequence identity to the Aap A lectin-like domain[115,145]. High DNA sequence similarity results in high amino acid similarity between B-repeats of 90-100 and 82-91 % for SasG and Aap, respectively[149]. Furthermore, the B-repeat region of SasG and Aap are highly homologous with one another, both on a structural (Figure 1.35B) and at the amino acid level[149]. Each B-repeat comprises two sub domains, E and G5 (Figure 1.35A), which boasts 50 and 34 % similarity between SasG and Aap[149]. The number of these B-repeats varies in number, with 4-14 observed in Aap[150] and 3-10 in SasG[151], with at least 5 required for SasG-mediated *S. aureus* biofilm formation[144]. The E and G5 sub domains of these B-repeats have an unusual flat elongated structure, comprising three interlaced strands (collagen-like regions) capped at both ends by a triple stranded β-sheet (mix of antiparallel and parallel) (Figure 1.35B). As both ligand binding and bacterial aggregation occur under high external shear stress, it is of interest to understand the mechanical properties of these proteins.

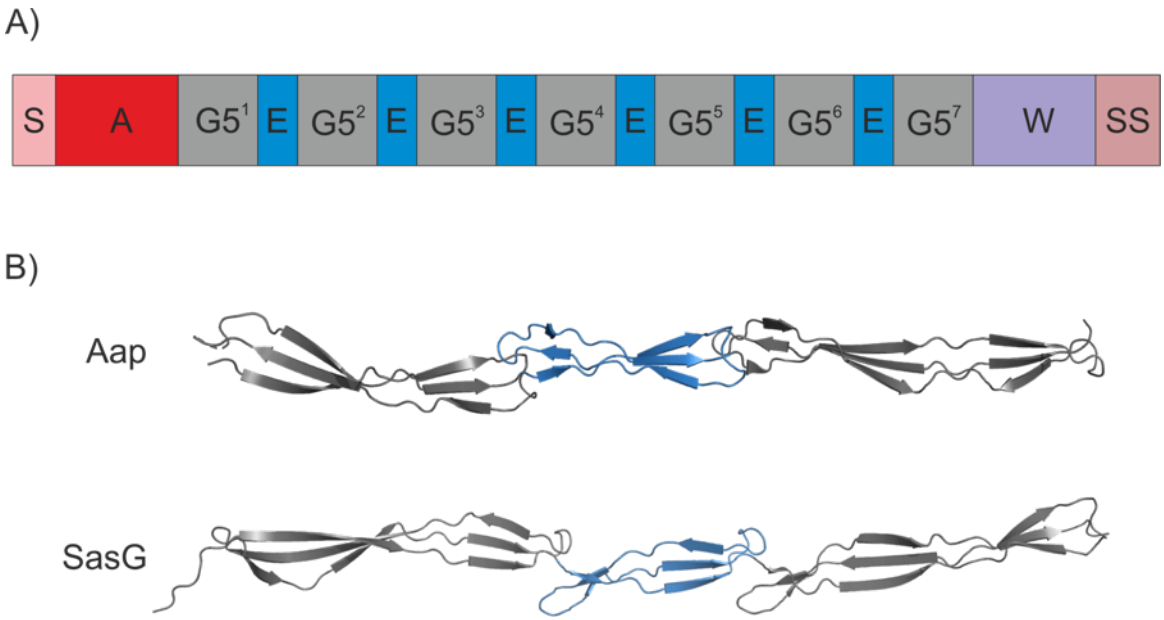

Figure 1.35 – *S. aureus* SasG and *S. epidermidis* Aap are structurally homologous. A) Schematic of full length SasG/Aap. S: Signal sequence, A: A domain, W: Wall spanning region and SS: Sorting signal. B-repeats consist of G5-E. B) Aap (PDB: 4FUP from reference[21]) is structurally homologous to SasG (PDB: 3TIQ from reference[149]). For both schematic and crystal structures the E and G5 sub domains are displayed in blue and grey, respectively.

Utilising SMFS, Gruszka and colleagues mechanically unfolded the B domain of SasG (6.5 B-repeats) and observed a sawtooth profile of six small peaks and seven larger peaks, followed by a detachment peak (Figure 1.36)[74]. The small and large peaks with ΔL_C values of ~ 150 and ~ 220 Å correspond to the E and G5 sub domains, respectively. Both sub domains displayed remarkable mechanostability with unfolding forces of 250 $\pm$ 35 and 420 $\pm$ 36 pN at a retraction velocity of 800 nms^{-1} for the E and G5 sub domains, respectively. In contrast to the cooperative unfolding of the E and G5 sub domains in chemical denaturant[74,152], mechanical perturbation forced the E and G5 sub domains to lose cooperativity and unfold independently of one another.

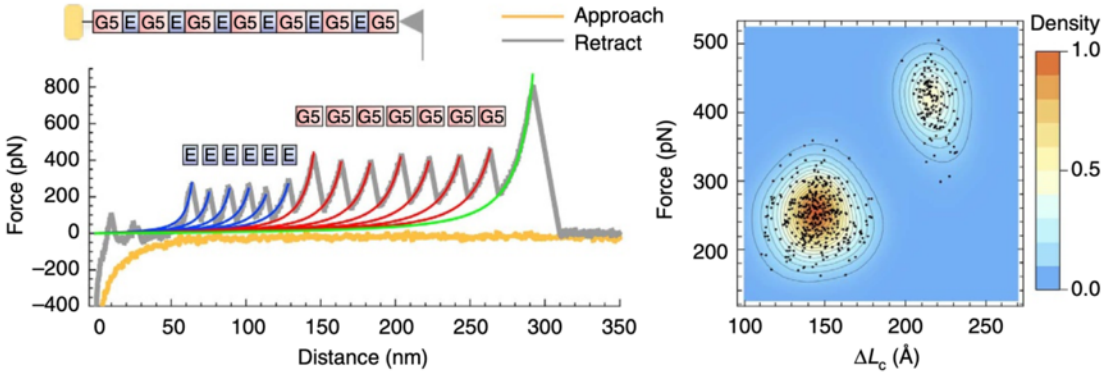

Figure 1.36 – SMFS forced unfolding of the SasG B domain ($G5^1$-$G5^7$) displays remarkable mechanical strength. Schematic of SMFS experimental set up (top, left). Sawtooth profile displaying small and large peaks (bottom left) representing six E (blue) and seven G5 (red) sub domains unfolding. Scatter-contour plot of the unfolding data displaying two populations, resulting from the distinct ΔL_C and mechanical strength between E and G5 sub domains (right). Figure taken from reference[74].

Formosa-Dague and colleagues utilised SCFS in an attempt to distinguish whether SasG engages in receptor-ligand or homophilic bonds in the presence and absence of Zn^{2+}[153]. They observed in the absence of Zn^{2+} ions, cells expressing SasG (A and 8 B domains) displayed minimal self-association. Conversely, upon the addition of Zn^{2+}, cell-to-cell interaction frequency increased to 80-100 % (Figure 1.37A). FX profiles generally displayed multiple peaks and had a maximum adhesion force of 414 $\pm$ 222 pN and a rupture length of 182 $\pm$ 76 nm at a retraction velocity of 1000 nms^{-1}. Upon the addition of EDTA, these multipeak force profiles disappeared, but were restored when Zn^{2+} was reapplied. When Ca^{2+} was utilised in place of Zn^{2+}, adhesive interactions did not reappear, demonstrating the Zn^{2+} specificity of the measured interactions.

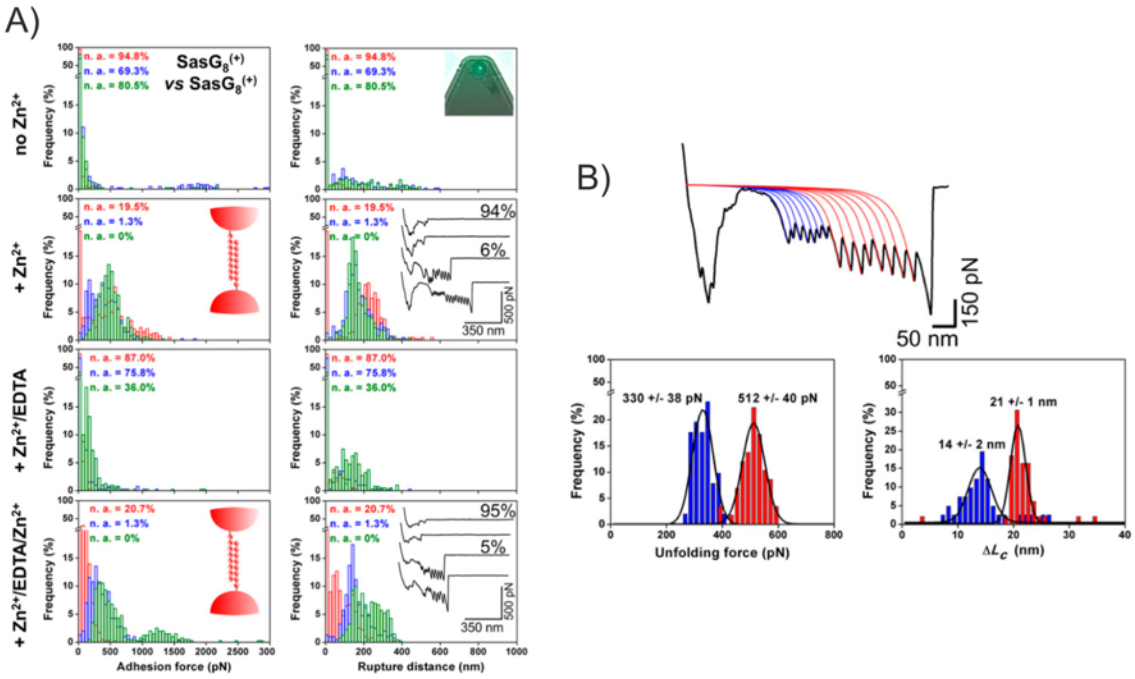

Figure 1.37 – SCFS investigations of SasG Zn^{2+}-induced homophilic bonds at a retraction velocity of 1000 nms^{-1}. A) Adhesion force and rupture distance histograms, with FX profile insets of SCFS experiments of *S. aureus* cells expressing $SasG_8^{(+)}$. Conditions from top to bottom: TBS buffer only → 1 mM Zn^{2+} → 1 mM EDTA → 1 mM Zn^{2+}. There was a large increase in adhesive events in the presence of 1 mM Zn^{2+}, which was abrogated with the addition of 1 mM EDTA and restored upon the second addition of 1 mM Zn^{2+}. B) A small percentage of adhesive events (5-6 %) displayed sawtooth profiles corresponding to the E (blue) and G5 (red) sub domains of SasG unfolding. Figure adapted from reference[153].

A substantial fraction (~ 6 %) of FX profiles displayed sawtooth profiles comprising small and larger peaks (Figure 1.37B) reminiscent of those observed for the purified SasG B domain[74]. As before, the small and large peaks corresponded to the sequential unfolding of the E and G5 domains and displayed ΔLc values of ~ 140 and ~ 210 Å and unfolding forces of 330 ± 38 and 512 ± 40 pN for the E and G5 sub domains, respectively. These unfolding forces of SasG measured *in situ* are in the range of those observed for SasG B domain *in vitro*[74], but the unfolding pattern was slightly different and displayed randomly distributed force peaks – a potential complexity associated with the unfolding of B-repeats involved in homophilic bonding.

Furthermore, they investigated the capability of *S. aureus* cells expressing SasG to form interactions with *S. epidermidis* expressing Aap in the presence and absence of Zn^{2+}. They observed strong adhesion in the presence of Zn^{2+}, with an cell-to-cell interaction frequency of 50-100 % of approach-retract cycles. The FX profiles generally displayed multiple peaks and had a

lower rupture force of 161 ± 79 pN and a rupture length of 168 ± 129 nm with a number of sawtooth profiles present also. This suggests heterophilic bonds can be formed between SasG and Aap and describes how a different species may utilise homologous proteins in order to form a heterogenous community.

More recently, Chantraine and colleagues demonstrated that the A domain of *S. epidermidis* Aap displays catch-bond behaviour using SMFS, with tensile loading strengthening of the Aap:vWF interaction similar to those previously discussed DLL mechanisms, but much higher (~ 3 nN at a retraction velocity of 1000 nms^{-1})[75]. An interesting and perhaps unsurprising observation was the sawtooth profile of ~22 peaks prior to the large A domain:vWF complex dissociation (Figure 1.38). This sawtooth was reminiscent to that observed for SasG by both SMFS[74] and SCFS[153] and represented the sequential unfolding of E and G5 sub domains of the Aap B-repeat region. The E and G5 sub domains unfolded at a force of 312 ± 45 and 475 ± 48 pN with ΔLc values of ~ 140 and ~ 210 Å at a retraction velocity of 1000 nms^{-1}. Interestingly, when cells were treated with mAbs directed against the A or B domain, adhesion was abolished and unfolding events were absent suggesting the requirement of both domains for the interaction.

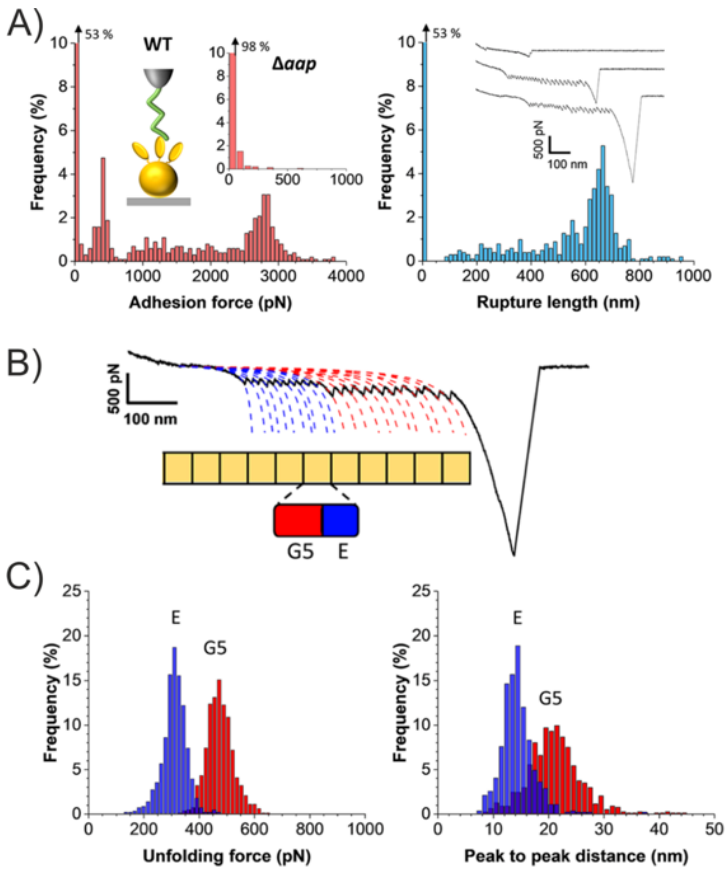

Figure 1.38 – The mechanical properties of Aap A domain:vWF complexes under force. A) The adhesion force (inset displays SMFS schematic and *S. epidermidis* Δ*aap* strain data) and rupture length with representative FX profiles inset (right) obtained by SMFS experiments on *S. epidermidis* coated substrate and cantilever tips functionalised with vWF. B) Typical FX sawtooth profile displaying the small and large peaks corresponding to the sequential unfolding of E and G5 sub domains prior to complex rupture (large ~ 3 nN peak). C) Histograms of unfolding forces and peak-to-peak distances from analysis of the sawtooth profiles. Figure adapted from reference[75].

The remarkable mechanical strength of SasG E and G5 sub domains has been predicted by MD simulations to originate from tandemly-arrayed 'mechanical clamps'[74]. These comprise of an N-terminal β-sheet clamp formed from two anti-parallel β-strands and a C-terminal β-sheet clamp formed from two parallel β-strands (Figure 1.39). These clamps involve long stretches of hydrogen bonding between the β-strands and side-chain packing interactions. As the hydrogen bonds are in a 'shear' geometry to the force vector, the forced unfolding of E and G5 sub domains requires the simultaneous rupture of these hydrogen bonds. The lower mechanostability of E is possibly due to a shorter N-terminal 'mechanical clamp' in comparison to G5. As SasG and Aap are structurally similar and conclusions drawn from SasG can likely be applied to Aap, it has been proposed that the mechanical strength of Aap E and G5 sub domains likely has the same origin[75].

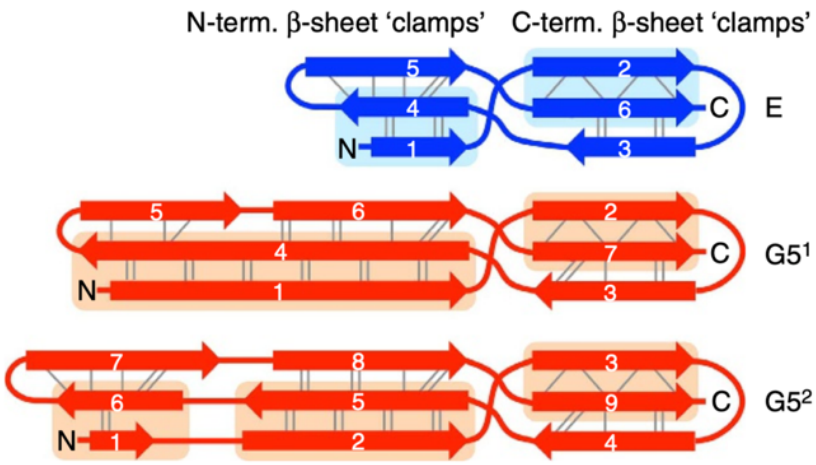

Figure 1.39 – Topology diagram of the postulated 'mechanical clamps' of the E-G5 repeats in SasG as revealed by MD simulations. The 'mechanical clamps' are indicated with shaded boxes. Hydrogen bonds shown as grey solid lines between β-strands. These 'mechanical clamps' are formed from tandemly arrayed long stretches of hydrogen bonds and side-chain packing interactions between β-strands forming an N- and C-terminal clamp. When the E and G5 sub domains are pulled, the hydrogen bonds are in a 'shear' geometry and require simultaneous rupture to unfold the sub domains. Figure taken from reference[74].

Multi-modular structure and remarkable mechanostability are key to the function of SasG and Aap. During biofilm development, the elongated, rod-like shape of the proteins will likely play a key role in projection of the adhesin A domain towards the host for initial attachment (depending on biofilm phenotype) and/or will play a key role in cell-to-cell Zn^{2+}-induced aggregation. The mechanostability of the E and G5 sub domains may allow these contacts to be maintained under physiological shear forces. Furthermore, the E sub domains may act as force 'buffers' relieving mechanical strength, preventing the unfolding of the whole SasG protein. Rapid recovery of the fully folded protein structure may occur after external force subsides, as E sub domains fold readily in the presence of folded G5 sub domains[74,152]. Building on this, Formosa-Dague and colleagues suggested that sequential unfolding of the repeats may expose previously masked residues used for adhesion under force[153]. In contrast, Chantraine and colleagues suggested that the unfolding of the E and G5 sub domains of Aap is part of an allosteric mechanism for shifting the A domain from a weak- to a strong-binding state enabling ultra-strong binding under force[75].

1.4.4 Targeting the Force Sensitivity of *Staphylococcal* CWA Proteins for Antibacterials

SMFS and SCFS studies to date show that both the adhesion and B domain components of *Staphylococcal* CWA proteins appear to be evolved to respond to or resist external force. These B domain(s) not only appear to be mechanically strong (from ~ 250 pN up to > 2 nN), but have distinct functions including, but not limited to, force buffering, protein:protein interactions, spring behaviour and/or allosteric effects. The mechanical stability arises from mechanical clamps formed from long stretches of hydrogen bonds, metal ions and intramolecular covalent bonds. It is possible that the *Staphylococcal* species (and others) have evolved force-dependent adhesion mechanisms to induce surface attachment and resist physical stress at high shear flow as a weapon in their arsenal of virulence factors. Controlling this phenomenon may provide a means for developing antibiofilm strategies. A theoretical example includes the chelation of metal ions, which could have a dual effect of both weakening mechanically strong adhesion proteins stabilised by metal ions[76] and inhibiting metal ion dependent interactions of mechanically strong proteins[21,153] resulting in the failure to stay attached under shear force and limit cell-to-cell aggregation, respectively. Several promising studies have already shown broad-spectrum metal chelators to have inhibitory effects on biofilm formation for several strains of *S. aureus*[154] and act as a potent gram-negative bacterial biofilm disrupter[155]. When the metal ion specificity of biofilm formation is understood in more depth, it may be possible to refine treatment, which could help minimise off-target chelation. In addition, a ground-breaking study by Rivas-Pardo and colleagues demonstrated the use of a short peptide as a ‘mechanical antibiotic’ to block the formation of the isopeptide bond present in Spy0128 which resulted in a mechanically labile protein[156]. Further development of these isopeptide-blocking peptides could result in an alternative therapeutic to antibiotics for gram-positive biofilm-mediated infections.

1.5 Thesis Aims

The aim of the work described in this thesis is to provide molecular insights into the mechanical properties of SasG. SMFS is utilised to investigate both **i)** the potential Zn^{2+}-induced dimerisation/oligomerisation of the SasG B domain and **ii)** the mechanical determinants of the E and G5 sub domains.

There is evidence both for and against the Zn^{2+}-induced dimerisation/oligomerisation of the SasG B domain for driving biofilm formation. The Zn^{2+}-induced homophilic binding of the SasG B domains has been observed *in situ* using SCFS[153]. However, *in vitro* techniques have struggled to detect any dimerisation/oligomerisation of recombinant proteins of up to 2 B-repeats of SasG in the presence of Zn^{2+} ions[149]. Furthermore, SasG E and G5 sub domains display remarkable mechanical strength, however, their structure is unusual in comparison to previously mechanically characterised proteins (lacking an Ig-like fold and a typical hydrophobic core). MD simulations have predicted the mechanical strength arises from tandemly arrayed 'mechanical clamps' comprising of long stretches of hydrogen bonds between β-strands[74]. However, there is currently no experimental evidence supporting this. Furthermore, the E and G5 sub domains comprise collagen-like motifs which have never been investigated using force spectroscopy.

The work in this thesis aims to:

1. Investigate the inconclusive Zn^{2+}-induced oligomerisation of the SasG B domain using SMFS *in vitro* (Chapter 3)
2. Develop a SasG-like SMFS polyprotein system for mechanical unfolding experiments *in vitro* (Chapter 4)
3. Utilise the developed SMFS polyprotein system to unravel the underlying mechanical determinants of the E and G5 sub domains (Chapter 5)

Understanding whether SasG is involved in Zn^{2+}-induced homophilic bonding is an important step for dissecting biofilm formation, which in turn will arm us with the knowledge to design novel strategies to prevent chronic biofilm-mediated infections. Furthermore, dissecting the mechanical determinants of the E and G5 sub domains of SasG will provide fundamental understanding to add

to the toolkit for designer proteins requiring extreme mechanical strength, such as those utilised in biomechanical machinery or as building blocks for mechanically strong biological materials[157].

2. Materials and Methods

2.1 Materials

2.1.1 Chemicals

	Supplier
A	
Acetic Acid, glacial	Fisher Scientific, Loughborough, UK
30 % (v/v) Acrylamide	Severn Biotech, Kidderminster, UK
Agar	Fisher Scientific, Loughborough, UK
Agarose	Melford Laboratories, Suffolk, UK
Ampicillin sodium	Formedium, Norfolk, UK
(3-Aminopropyl)triethoxysilane (APTES)	Sigma Life Sciences, MO, USA
Ammonium persulphate (APS)	Sigma Life Sciences, MO, USA
Ammonium Bicarbonate	Sigma Life Sciences, MO, USA
Ammonium Chloride	Sigma Life Sciences, MO, USA
B	
Benzamidine dihydrochloride	Sigma Life Sciences, MO, USA
C	
Chloroform	Fisher Scientific, Loughborough, UK
Calcium chloride ($CoCl_2$)	Melford Laboratories, Suffolk, UK
Chloramphenicol	Sigma Life Sciences, MO, USA
Cobalt (II) chloride hexahydrate	Sigma Life Sciences, MO, USA
D	
Dithiothreitol (DTT)	Formedium, Norfolk, UK
Dimethyl sulfoxide (DMSO)	Sigma Life Sciences, MO, USA
E	
Ethanol	Sigma Life Sciences, MO, USA
Ethylenediaminetetraacetic acid (EDTA)	Fisher Scientific, Loughborough, UK
G	
Glycerol	Fisher Scientific, Loughborough, UK
D-Glucose anhydrous	Fisher Scientific, Loughborough, UK
H	
30 % (w/w) Hydrogen peroxide	Sigma Life Sciences, MO, USA
37 % (w/w) Hydrochloric acid (HCl)	Fisher Scientific, Loughborough, UK
Hellmanex III	Hellma UK LTD (Essex, UK)
I	
Imidazole	Sigma Life Sciences, MO, USA
Isopropanol	Fisher Scientific, Loughborough, UK
L	
α-Lactose monohydrate	Sigma Life Sciences, MO, USA
Luria broth (LB)	Melford Laboratories, Suffolk, UK
M	
Magnesium sulphate	Sigma Life Sciences, MO, USA
N	
N-Hydroxysuccinimide-PEG_{24}-maleimide ($SM(PEG)_{24}$)	Fisher Scientific, Loughborough, UK
N,N-diisopropylethylamine (DIPEA)	Fisher Scientific, Loughborough, UK
P	
Phenylmethanesufonyl fluoride (PMSF)	Sigma Life Sciences, MO, USA
Potassium chloride	Fisher Scientific, Loughborough, UK
Potassium phosphate monobasic	Fisher Scientific, Loughborough, UK
S	

Sodium chloride (NaCl)	Fisher Scientific, Loughborough, UK
95-98 % Sulphuric acid	Fisher Scientific, Loughborough, UK
Sodium hydroxide (NaOH)	Fisher Scientific, Loughborough, UK
10 % Sodium dodecyl sulphate (SDS)	Severn Biotech, Kidderminster, UK
Sodium phosphate dibasic	Thermo Scientific, Surrey, UK
Sodium phosphate monobasic	Sigma Life Sciences, St. Louis, USA
SYBR™ safe	Thermo Scientific, Surrey, UK
Sodium sulphate	Thermo Scientific, Surrey, UK
T	
Tris	Fisher Scientific, Loughborough, UK
Tryptone	Melford Laboratories, Suffolk, UK
Tetramethylethylenediamine (TEMED)	Sigma Life Sciences, MO, USA
Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)	Sigma Life Sciences, MO, USA
U	
Urea	Fisher Scientific, Loughborough, UK
Y	
Yeast extract	Melford Laboratories, Suffolk, UK
Z	
Zinc chloride ($ZnCl_2$)	Sigma Life Sciences, MO, USA

2.1.2 Gel Ladders and Dyes

100 bp, 1 kb DNA ladder (NEB, Hertfordshire, UK)

Precision Plus Protein™ Dual Xtra prestained protein standard protein ladder (Biorad, CA, USA)

Gel loading dye, purple (6X) (NEB, Hertfordshire, UK)

2.1.3 Centrifuges

2.1.3.1 Bench top

GenFuge 24D (Progen Scientific, London, UK) – for 5-1500 µl sample volumes

Centrifuge 5810 R (Eppendorf, Hamburg, Germany) – for 5-50 ml sample volumes

2.1.3.2 Floor

Beckman Avanti J-26 XP centrifuge (Beckman Coulter, CA, USA) – for 50-1000 ml sample volumes

2.1.4 Incubators

Gallenkamp economy incubator size 1 (Fison Erba Science, Guildford, UK)

New Brunswick™ Innova 44 (Eppendorf, Hamburg, Germany)

Stuart SI600 orbital incubator (Bibby Scientific, Staffordshire, UK)

2.1.5 Kits

QIAquick® PCR purification Kit (QIAGEN Group, Venlo, Netherlands)

QIAprep® Spin Miniprep Kit (QIAGEN Group, Venlo, Netherlands)

2.1.6 Protein Purification Equipment

ÄKTAprime plus (Cytiva, MA, USA)

Superdex™ 75 HiLoad 26/600 (320ml) gel filtration column (Cytiva, MA, USA)

Nickel Sepharose high performance resin (Cytiva, MA, USA)

HisTrap™ FF – 1 ml (Cytiva, MA, USA)

HiTrap™ SP Sepharose FF – 5 ml (Cytiva, MA, USA)

HiTrap™ Q HP – 5 ml (Cytiva, MA, USA)

2.1.7 Spectrophotometers

NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific, MA, USA)

Ultrospec 2100 pro UV-Vis Spectrophotometer (Cytiva, MA, USA)

2.1.8 Circular Dichroism (CD)

Chirascan™ plus CD Spectrometer (Applied Photophysics, UK)

2.1.9 PCR Thermocycler

T100™ Thermal Cycler (BioRad, CA, USA)

2.1.10 AFM

MFP-3D™ Stand Alone AFM (Asylum Research, Buckinghamshire, UK)

2.1.11 AFM Disposables

MLCT AFM probes – Silicon nitride cantilevers with reflective gold tips (Bruker, CA, USA)

Gold coated silicon nitride wafer – Silicon nitride wafer with 100 nm thick gold coating (Platypus technologies, WI, USA)

Silicon nitride wafer (Rockwood Electronic Materials, Alfreton, UK)

2.1.12 Fluorometer

Photon Technology International fluorometer (Ford, West Sussex, UK)

2.1.13 Media For Bacterial Growth

2.1.13.1 Lysogeny Broth (LB) Medium and Agar Plates

LB medium:

Reagent	Concentration (w/v %)
LB	2.5

Contains 10 g tryptone, 5 g yeast extract and 10 g NaCl per litre.

Autoclave at 121 °C 15 psi for 20 minutes.

Agar plates:

Reagent	Concentration (w/v %)
LB	2.5
Agar	1.5

Autoclave at 121 °C 15 psi for 20 minutes. When the agar solution has cooled sufficiently, add the appropriate antibiotic(s) (25 μg/ml chloramphenicol and/or 100 μg/ml of ampicillin) from 1000x stocks: 25 mg/ml chloramphenicol in 100 % ethanol and 100 mg/ml ampicillin sodium in Milli-Q water. 10 ml was then decanted into sterile plates (100 mm x 15 mm) using aseptic technique.

2.1.13.2 Autoinduction Medium (2ZYM-1X LAC)

For protein expression by autoinduction[158].

2ZY Component:

Reagent	Concentration (w/v %)
Yeast Extract	1
Bactotryptone	2

Autoclave at 121 °C 15 psi for 20 minutes.

50X LAC Component:

Reagent	Concentration (w/v %)
Glycerol	0.5
D-Glucose anhydrous	0.05
α-Lactose monohydrate	0.2

Filter sterilise through a 0.22 μm filter.

20X NPSC Component:

Reagent	Molarity (M)
Ammonium chloride	1
Sodium sulphate	0.1
Potassium phosphate monobasic	0.5
Sodium phosphate dibasic	0.5

Autoclave at 121 °C 15 psi for 20 minutes.

To a 2 L baffled flask, add the following components for the final autoinducing expression medium:

Reagent	Volume Added To 500 ml Total (ml)
2ZY Component	465
1 M $MgSO_4$	1
50X LAC	10
20X NPSC	25

Then add the appropriate antibiotic(s) (25 μg/ml chloramphenicol and/or 100 μg/ml of ampicillin) from 1000x stocks.

2.1.14 Buffers

Experimental buffers are described in the methods and/or experimental sections. More general buffers are described below. All buffers were prepared with ultra-pure deionised (Milli-Q) water with a resistivity of 18.2 MΩ·cm.

2.1.14.1 Immobilised Metal Affinity Chromatography (IMAC) Buffer A1

20 mM (Na_2HPO_4-NaH_2PO_4) pH 8.0, 8 M Urea, 500 mM NaCl, 5 mM Imidazole, 2 mM DTT, 1 mM PMSF, 2 mM Benzamidine

2.1.14.2 IMAC Buffer A2

20 mM (Na_2HPO_4-NaH_2PO_4) pH 8.0, 8 M Urea, 500 mM NaCl, 10 mM Imidazole, 2 mM DTT, 1 mM PMSF, 2 mM Benzamidine

2.1.14.3 IMAC Wash Buffer

20 mM (Na_2HPO_4-NaH_2PO_4) pH 8.0, 500 mM NaCl, 10 mM Imidazole, 2 mM DTT, 1 mM PMSF, 2 mM Benzamidine

2.1.14.4 IMAC Elution Buffer

20 mM (Na_2HPO_4-NaH_2PO_4) pH 8.0, 500 mM NaCl, 500 mM Imidazole, 2 mM DTT, 1 mM PMSF, 2 mM Benzamidine

2.1.14.5 IMAC dilution Buffer

20 mM (Na_2HPO_4-NaH_2PO_4) pH 8.0, 500 mM NaCl, 2 mM DTT

2.1.14.6 Anion Exchange (AEX) Wash Buffer

20 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM EDTA, 2 mM DTT

2.1.14.7 AEX Elution Buffer

20 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 2 mM DTT

2.1.14.8 AEX Dilution Buffer

20 mM Tris-HCl pH 8.0, 1 mM EDTA, 2 mM DTT

2.1.14.9 1X PBS, pH 7.4

137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4

2.1.14.10 1X TBS, pH 7.4

20 mM Tris-HCl pH 7.4, 150 mM NaCl

2.1.14.11 50X TAE

50 mM EDTA, 2 M Tris, 1 M Acetic acid

2.1.14.12 SDS-PAGE buffers

SDS-PAGE resolving gel solution – 1.5 M Tris-HCl pH 8.8, 7.5 % (v/v) acrylamide, 0.1 % (v/v) SDS, 0.7 % (w/v) APS, 0.07 % (v/v) TEMED

SDS-PAGE stacking gel solution – 1 M Tris-HCl pH 6.8, 5 % (v/v) acrylamide, 0.07 % (v/v) SDS, 0.32 % (w/v) APS, 0.16 % (v/v) TEMED

SDS-PAGE loading buffer (2X concentrated stock) – 4 % (w/v) SDS, 20 % (v/v) glycerol, 0.2 % (w/v) bromophenol blue, 200 mM DTT

SDS-PAGE cathode buffer (10X concentrated stock) – 1 M Tris-HCl, 1 M Tricine, 1 % (w/v) SDS, pH 8.3

SDS-PAGE anode buffer (10X concentrated stock) – 2 M Tris-HCl, pH 8.8

2.1.14.13 1.5 % (w/v) Agarose Gel

1X TAE, 1.5 % (w/v) agarose, 1X SYBR™ Safe (from 10000x stock stored in dimethyl sulfoxide (DMSO))

2.1.15 Enzymes

Restriction enzymes: BsaI-HF®v2, Q5® DNA polymerase (in Hot Start High-Fidelity 2X Master Mix) and One*Taq*® DNA polymerase (in Quick-Load® 2X Master Mix with Standard Buffer) were purchased from NEB, Hertfordshire, UK. DNAse I (from bovine pancreas) and lysozyme (from chicken egg white) were purchased from Sigma Life Sciences, MO, USA.

2.1.16 Oligonucleotides

All oligonucleotides (primers) were purchased from Eurofins MWG Operon, Ebersberg, Germany.

2.1.17 Constructs

2.1.17.1 pETFPP_1-SasG

H_6-3C-G5^1-E-G5^2-E-G5^3-E-G5^4-E-G5^5-E-G5^6-E-G5^7-Strep-tag(II)-CC was kindly provided by Professor Jennifer Potts (University of Sydney, Australia). This construct comprised of a N-terminal 3C cleavable hexahistidine (H_6) tag and a C-terminal Strep-tag® (II). This construct served as a DNA sequence foundation for creating constructs in this thesis. Schematic of DNA displayed below (ST(II): Strep-tag (II)):

2.1.17.2 H_6-MBP-TEV-SasG

H_6-Maltose-binding protein (MBP) followed by a Tobacco Etch Virus (TEV) cleavage site and G5^1-E-G5^2-E-G5^3-E-G5^4-E-G5^5-E-G5^6-E-G5^7 with C-terminal double cysteines in an expression vector (pMAL-c5X) under the control of a tac promoter. For use in structural characterisation and SMFS studies. Referred to as SasG after purification. Schematic of DNA displayed below:

2.1.17.3 H_6-TEV-SasG-ΔEE

H_6-TEV-G5^1-E-G5^2-E-G5^3-E-G5^4-E-G5^5-E-G5^6-E-G5^7 with C-terminal double cysteines in an expression vector under the control of a T7 promoter. Each E and G5 sub domain contains a glutamic acid to alanine mutation corresponding to E518A and E557A in E-G5^2. The construct

contains a TEV cleavage sequence for H_6 tag removal. Referred to as SasG-ΔEE after purification. Schematic of DNA displayed below with * denoting the E to A mutations:

2.1.17.4 E-G5²

E-G5² monomer (residues 502-629 – UniProt[159] boundaries) from pETFPP_1-SasG cloned into an empty pET14b vector for use as a template for subsequent mutations and cassette assembly. During cassette creation, the 5'-end length of each cassette was increased by 12 bp to match the crystal structure (PDB: 3TIP) boundaries and provide sufficient overhang for further avoidance of faux interdomain interactions (residues 498-629).

2.1.17.5 pWT (E-G5²)₅

Pentameric assembly of E-G5² (residues 498-629) with seven residue linkers between neighbouring E-G5² repeats and C-terminal double cysteines in a pET14b expression vector under the control of a T7 promoter. For use in structural characterisation, equilibrium denaturation and SMFS studies. Referred to as pWT (E-G5²)₅ throughout the text. Variants of this construct included I502P, G517A, V522P, G524A, G527A, P540A, P549A, V550P, V556P, P562A, V580P, G584A, G587A, E588K, K589E, N598A, T601A, E624K. These mutations were located in every E-G5² repeat of the pentamer. Schematic of DNA displayed below (dark purple blocks represent the linkers):

2.1.18 Bacterial strains

2.1.18.1 For DNA Manipulation

NEB® 5-α Competent E. coli cells (NEB, Hertfordshire, UK): Derivative of DH5α with genotype: *fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*

SURE 2 Supercompetent *E. coli* cells (Agilent Technologies, CA, USA): Deficient in *E. coli* genes involved in rearrangement and deletion of DNA when working with repetitive DNA. Genotype: Restriction minus (*McrA-, McrCB-, McrF-, Mrr-, HsdR-*) endonuclease (*endA*) deficient, and recombination (*recB, recJ*) deficient.

2.1.18.2 For Gene Expression and Production of Recombinant Proteins

BL21(DE3) competent *E. coli* cells (Agilent Technologies, CA, USA): B strain lacking *lon* protease and *ompT* outer membrane protease, which can degrade proteins during purification. Genotype: *E. coli* B F^- *dcm ompT hsdS*(r_B^- m_B^-) *gal* λ(DE3)

BL21(DE3)pLysS competent *E. coli* cells (Agilent Technologies, CA, USA): Identical to BL21(DE3) except the cells contain a plasmid (pLysS), which encodes the T7 lysozyme gene (*LysS*). T7 lysozyme binds to T7 polymerase preventing leaky expression. Genotype: *E. coli* B F^- *dcm ompT hsdS*(r_B^- m_B^-) *gal* λ(DE3) [pLysS Cam^r]

2.2 Methods

2.2.1 Simulations

2.2.1.1 Cosolvent Simulations

Cosolvent MD simulations on E-G5[2] (PDB: 3TIP) were carried out by Alexander St John (PhD student, University of Leeds). These cosolvent simulations were carried out with the presence of 112 Zn^{2+} ions (and other negative ions for neutralising the system) as the cosolvent. 50 individual replicas were run for 25 nanoseconds and then averaged over all 50 replicas. The free-energy of Zn^{2+} was calculated by aligning the replica simulations together using the backbone of the protein and then placing a grid over the combined trajectory containing 1x1x1 $Å^3$ voxels[160]. The number of times a Zn^{2+} ion appears within each voxel is counted and divided by the number of frames in the trajectory providing us with the density within said voxel. Utilising Equation 2.1, the probability can be converted into free-energy (ΔG_i):

$$\Delta G_i = -RT \ln\left(\frac{N_i}{N_o}\right)$$

2.1

Where N_i is the density of Zn^{2+} in a voxel and N_o is the density of Zn^{2+} in the absence of any bias from the protein. N_o (or bulk density) is calculated by running a single simulation without the protein, but with the equivalent number of water and cosolvent molecules. This provides us with a free energy map of Zn^{2+} (after ΔG_i is calculated for each voxel) which is loaded onto PyMOL and displays the positions where the protein being there increases or decreases the probability of Zn^{2+} being at a given position.

2.2.1.2 E-G5^2 Forced Unfolding

Forced unfolding of E-G5^2 (PDB: 3TIP) was carried out by Dr. Emanuele Paci (University of Bologna). These simulations were carried out according to the method in reference[74]. Briefly, simulations were performed using a united-force field (CHARMM19) and fast analytical continuum treatment of solvation (FACTS) implicit solvent model. Forced unfolding was simulated by attaching an ideal spring to the N and C atoms of the two termini and retracting them at a constant velocity ($1x10^9$ nms^{-1}) at 0 K (the minimum energy pathway).

2.2.2 Molecular Biology

2.2.2.1 Q5® High-Fidelity PCR

For Q5® PCR, the following reaction was set up in 0.2 ml PCR tubes with a final reaction volume of 25 µl:

Reagent	Volume (µl)
Q5® Hot Start High-Fidelity Master Mix (2X)	12.5
10 µM Forward Primer	1.25
10 µM Reverse Primer	1.25
Template DNA (25 ng/µl)	1
Nuclease-free H_2O	9

The Q5® Hot Start High-Fidelity Master Mix (2X) contains Q5® hot start high-fidelity DNA polymerase at 0.04 units/µl.

The reaction mixture was subjected to the following PCR thermal cycle:

Step	Temperature (°C)	Duration (s)	Cycles
Initial denaturation	98	30	
PCR	98 50-72* 72	10 30 30/kb	X 35
Final extension Hold	72 4	120 ∞	

*Primer annealing temperature.

2.2.2.2 Site-Directed Mutagenesis/Cassette Creation

Site-directed mutagenesis was exploited to delete, add or substitute nucleotides in DNA to create variants, add/delete restriction sites and add cysteine residues to facilitate protein immobilisation for SMFS studies. This was carried out using the NEB Q5® High-Fidelity 2X Master Mix (NEB, Hertfordshire, UK). Primer design was accomplished through use of the NEBaseChanger online tool (https://nebasechanger.neb.com/). The PCR reaction was set up as outlined in Section 2.2.2.1. The PCR product was subjected to phosphorylation, intramolecular ligation and template removal using NEB KLD Enzyme Mix. This reaction was set up in 0.2 ml PCR tubes as outlined below:

Reagent	Volume (µl)
PCR product	1
KLD Reaction buffer (2X)	5
KLD Enzyme Mix (10X)	1
Nuclease-free H_2O	3

After 5 minutes at room temperature, the reaction was transferred to ice and 5 µl was transformed into NEB® *E. coli* 5-α cells as according to Section 2.2.2.5.

The mutagenesis primers for the domestication of pET14b are listed in Table 2.1:

Description	Forward Sequence	Reverse Sequence
MCS BsaI site removal	5'- ATAGGGAGACAACGGTTTC-3'	5'- AGTGAGTCGTATTAATTTC-3'
AmpR gene BsaI site removal	5'-AGCGTGGGTCCCGCGGTATCA-3'	5'- CACCGGCTCCAGATTTATCAG-3'

Table 2.1 – Primers for the domestication of pET14b (pET14b$\Delta bsaI$). Substituted bases in red text.

The mutagenesis primers for E-G5² are listed in Table 2.2:

Mutant	Forward Sequence	Reverse Sequence
I502P	*	*
G517A	5'-TTACCTACCGCTGAGAAGGAA-3'	5'-CTTCGGATCAAACTCATCACGATG-3'
V522P	5'-GAAGGAAGAACCTCCTGGTAAGCCGG-3'	5'-TCACCGGTAGGTAACTTC-3'
G524A	5'-GAAGTTCCTGCTAAGCCGGGTATTAAAA-3'	5'-TTCCTTCTCACCGGTAGG-3'
G527A	5'-GGTAAGCCGGCTATTAAAAACC-3'	5'-AGGAACTTCTTCCTTCTC-3'
V550P	5'-ATACGGTCCTCCTAAAGGCGAC-3'	5'-TTGGTAACGCTGTCAACAG-3'
V556P	5'-CGACAGTATTCCGGAAAAAGAGGAAATC-3'	5'-CCTTTAACAGGACCGTATTTG-3'
P540A	5'-TGTGGTTCGTGCGCCTGTTGA-3'	5'-TCGCCTGTCTCAGGGTTTTTAATAC-3'
P549A	5'-CAAATACGGTGCTGTTAAAGGC-3'	5'-GTAACGCTGTCAACAGGC-3'
P562A	5'-AGAGGAAATCGCGTTCGAAAAAG-3'	5'-TTTTCCACAATACTGTCG-3'
V580P	5'-CACCGAAAAACCGACCCGTGAGGGTC-3'	5'-CCCGGTGCTAAATCAGGA-3'
G584A	5'-ACCCGTGAGGCTCAAAAAGGTG-3'	5'-CACTTTTTCGGTCCCCGG-3'
G587A	5'-GGTCAAAAAGCTGAGAAGACCATTAC-3'	5'-CTCACGGGTCACTTTTTC-3'
E588K	5'-TCAAAAAGGTAAGAAGACCATTACAAC-3'	5'-CCCTCACGGGTCACTTTT-3'
K589E	5'-AAAAGGTGAGGAGACCATTACAACC-3'	5'-TGACCCTCACGGGTCACT-3'
N598A	5'-TACACTGAAAGCCCCGCTGACC-3'	5'-GGGGTTGTAATGGTCTTC-3'
T601A	5'-AAACCCGCTGGCCGGCGAGAT-3'	5'-TTCAGTGTAGGGGTTGTAATGGTCTTCTC-3'
E624K	5'-CGAACTGACCAAATACGGCCC-3'	5'-TTAATAGGGTCCTTTGTGATCTC-3'
E-G5² $\Delta BsaI$	5'-ACGGCCCGGAAACCTGTTGTT-3'	5'-ATTCGGTCAGTTCGTTAATAGGGTCC-3'

Table 2.2 – Primers for E-G5² monomer mutagenesis. Substituted bases in red text. *I502P was created during amplification of cassettes using primers found in Table 2.5.

2.2.2.3 Gibson Assembly®

Gibson Assembly® (GA) is a DNA assembly method employing three enzymatic activities in a single-tube, isothermal reaction (Figure 2.1)[161,162]. A 5' exonuclease generates long overhangs by cleaving nucleotides from the 5' end, exposing complementary nucleotide stretches. Once these are annealed, a DNA polymerase then fills in the gaps of the single strand regions and a DNA ligase seals the nicks and covalently links the DNA fragments together. For cassette creation, primers were designed to amplify the DNA of interest and simultaneously add flanking nucleotides of at least 18 bp complementary to the linearised destination vector.

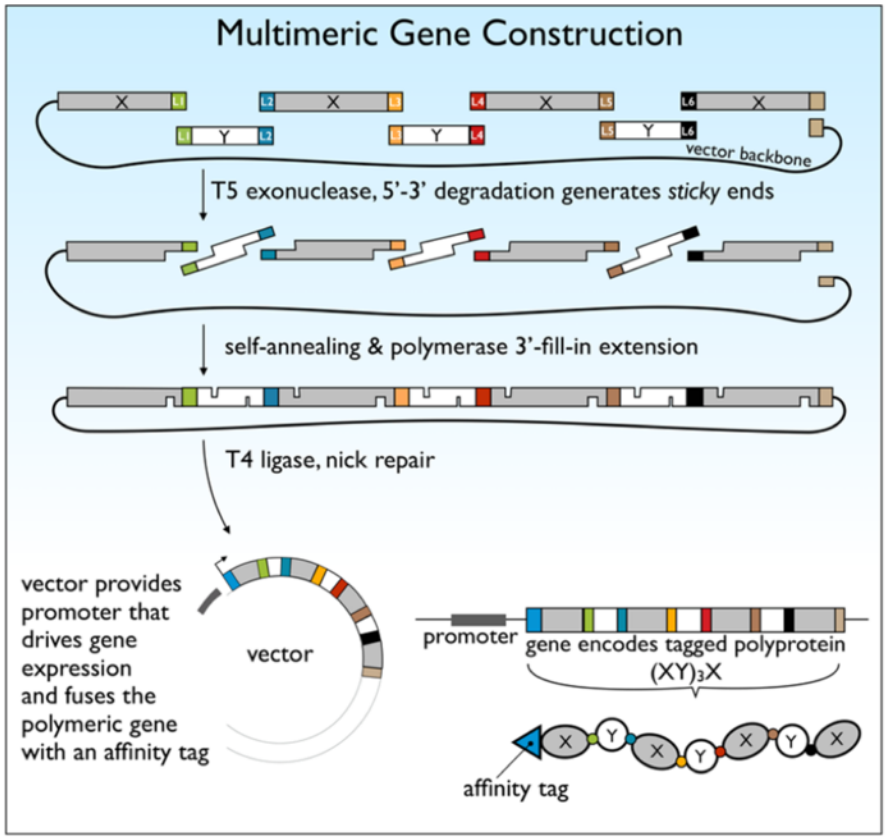

Figure 2.1 – Schematic of the GA work flow for scarless multimeric gene construction. GA can also be utilised to clone a single insert into a vector. Figure taken from reference[163].

Using Q5® PCR, an insert was amplified and the destination vector was linearised as outlined in Section 2.2.2.1. The primers used for monomeric pET14b-E-G5[2] and H_6-MBP-TEV-SasG GA reactions are listed in Table 2.3.

Construct	Process	Forward Sequence	Reverse Sequence
pET14b-E-G5^2	pET14b linearisation and amplification (destination vector)	5'-GGCCTGGTGCCGC-3'	5'-CCATCATCATCATCATCACAGCAGC-3'
	E-G5^2 GA-ready insert	5'-*ATCATCATCATCACAGCAGC*ATCGCAACCGGGCCATC-3'	5'-CTGCCGCGCGGCACCAGGCCTCAACAACA*GGTCTCCGGGCCGTATTC*-3'
H_6-MBP-TEV-SasG	pMAL-c5X-TEV linearisation and amplification (destination vector)	5'-GAATTCCCTGCAGGTAATTAAATAAGC-3'	5'-GGATCCCTGAAAGTACAGGTTT-3'
	SasG GA-ready insert	5'-*GAAAACCTGTACTTTCAGGGATCC*GCACCTAAGACCATCACCGAG-3'	5'-*GCTTATTTAATTACCTGCAGGGAATTC TTAGCAGCA*GGTCTCCGGACCATACTCG-3'

Table 2.3 – DNA primers utilised during Q5® PCR for pET14b-E-G5^2 and pMAL-c5X-MBP-TEV-SasG construct creation. Complementary stretches for annealing during GA are italicised.

For GA, the following reaction was set up in 0.2 ml PCR tubes with a final reaction volume of 20 µl:

Reagent	Assembly reaction
Linearised destination vector (30 ng/μl)	1 µl
Insert in amplicon form	5:1 molar ratio of insert: vector
Gibson Assembly® (2X) Master Mix	10 µl
Nuclease-free H_2O	To 20 µl

The reaction mixture was incubated at 50 °C for 20 minutes (for 1 insert) to promote the activity of the three enzymes. 2 µl of the assembly product was transformed into NEB® 5-α competent *E. coli* cells as outlined in Section 2.2.2.5.

2.2.2.4 Golden Gate Assembly

Golden Gate (GG) assembly is a DNA assembly technique utilising a Type IIS endonuclease, most commonly BsaI, and a T4 DNA ligase in a single tube reaction[164,165]. Primers are designed in such a way that BsaI recognition sites are added distal to the cleavage sites, which upon digestion, leaves complementary overhangs for subsequent annealing. The workflow is outlined in Figure 2.2.

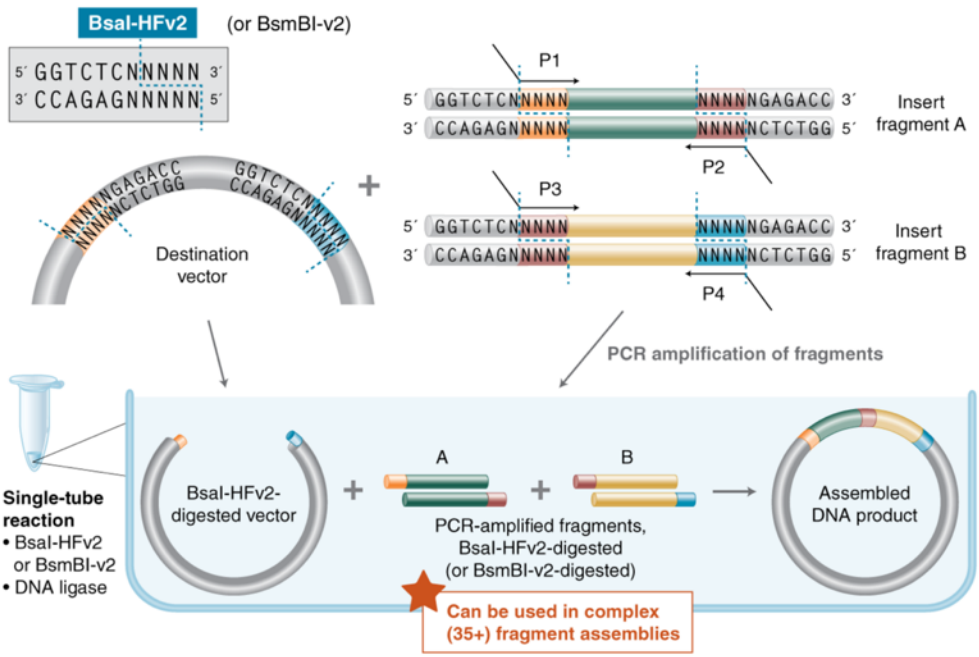

Figure 2.2 – GG assembly workflow utilising the BsaI type IIS endonuclease. BsaI recognises the GGTCTC(N1/N5) recognition site, where GGTCTC represents the recognition/binding site and the N1/N5 indicates the cut site is one base (N1) and five bases (N5) downstream on the top and bottom strands, respectively. This digestion leaves 4 base overhangs to direct the assembly utilising DNA ligase. Taken from reference[166].

The primers utilised in the GG assembly workflow are listed in Table 2.4. I502P forward primers differed slightly as the mutation was found in the annealing region and these are outlined in Table 2.5. Q5® PCR was used for the amplification of the inserts and the linearisation of the destination vector as outlined in Section 2.2.2.1.

Component	Forward Sequence	Reverse Sequence
pET14b destination vector	5'-GGTACT**GGTCTC**G*GACG*AAAGGAAGCTGAGTTGGC-3'	5'-GGTACT**GGTCTC**G*GACG*AAAGGAAGCTGAGTTGGCTG-3'
Cassette 1	5'-GGTACT**GGTCTC**G*ATCA*TCATCACAGCAGCGGCCCGGAAACCATCGCACCGGGCCATC-3'	5'-GGTACT**GGTCTC**G*CCCA*CGCTCAGGGTTTCCGGGCCGTATTC-3'
Cassette 2	5'-GGTACT**GGTCTC**G*TGGG*CGCGACCATTGGCCCGGAAACCATCGCACCGGGCCATC-3'	5'-GGTACT**GGTCTC**G*CAGA*CCAATAACGGTGGTTTCCGGGCCGTATTC-3'
Cassette 3	5'-GGTACT**GGTCTC**C*TCTG*GCGAGCGGCCCGGAAACCATCGCACCGGGCCATC-3'	5'-GGTACT**GGTCTC**C*GTGC*CGCTCAGCGCGGTTTCCGGGCCGTATTC-3'
Cassette 4	5'-GGTACT**GGTCTC**C*GCAC*CATTG	5'-GGTACT**GGTCTC**C*GGCT*ACCGGTAATAACGGTTTCCGGGCCGTATTC-3'

	TGGGCCCGGAAACCATCGCACC GGGCCATC-3’	
Cassette 5	5’-GGTACT**GGTCTC**G*AGCC*TGGCG GGCCCGGAAACCATCGCACCGG GCCATC-3’	5’-GGTACT**GGTCTC**G*CGTC*AACAACA GGTTTCCGGGC-3’

Table 2.4 – Primers for the creation of the pET14b destination vector and cassettes for subsequent GG assembly. Bases in bold represent the BsaI recognition site. Underlined and italicised are the four bases, which become complementary overhangs post BsaI digestion.

Component	**Forward Sequence**
I502P Cassette 1	5’-GGTACT**GGTCTC**G*ATCA*TCATCACAGCAGCGGCCCGGAAACCCCCG CACCGGGCCATC-3’
I502P Cassette 2	5’-GGTACT**GGTCTC**G*TGGG*CGCGACCATTGGCCCGGAAACCCCCGCAC CGGGCCATC-3’
I502P Cassette 3	5’-GGTACT**GGTCTC**C*TCTG*GCGAGCGGCCCGGAAACCCCCGCACCGG GCCATC-3’
I502P Cassette 4	5’-GGTACT**GGTCTC**C*GCAC*CATTGTGGGCCCGGAAACCCCCGCACCGG GCCATC-3’
I502P Cassette 5	5’-GGTACT**GGTCTC**G*AGCC*TGGCG GGCCCGGAAACCCCCGCACCGGGCCATC-3’

Table 2.5 – Forward primers utilised in the creation of I502P cassettes for subsequent GG assembly. Bases in bold represent the BsaI recognition site. Underlined and italicised are the four bases, which become complementary overhangs post BsaI digestion. Red text displays the bases changed to substitute isoleucine for proline.

The linearised pET14b destination vector was either used directly after amplification or circularised (using NEB KLD Enzyme Mix in the same fashion as Section 2.2.2.2) for use at a later date i.e. other variant creation. A segment of the circularised pET14b destination vector and the cassette amplicons for pWT $(E\text{-}G5^{2})_5$ are outlined in Table 2.6.

Component	**Forward Sequence**
Circularised pET14b destination vector – 4632 bp	5’-…AGATATACCATGGGCAGCAGCCATCATC*ATCA*C**GAGACC**AGTACCGG TACT**GGTCTC**G*GACG*AAAGGAAGCTGAGTTGGC… -3’
Cassette 1 – 450 bp	5’-GGTACT**GGTCTC**G*ATCA*TCATCACAGCAGCGGCCCGGAAACCATCGCA CCGGGCCATCGTGATGAGTTTGATCCGAAGTTACCTACCGGTGAGAAGG AAGAAGTTCCTGGTAAGCCGGGTATTAAAAACCCTGAGACAGGCGATGT GGTTCGTCCGCCTGTTGACAGCGTTACCAAATACGGTCCTGTTAAAGGC GACAGTATTGTGGAAAAAGAGGAAATCCCGTTCGAAAAAGAACGCAAAT TTAATCCTGATTTAGCACCGGGCACCGAAAAAGTGACCCGTGAGGGTCA

	AAAAGGTGAGAAGACCATTACAACCCCTACACTGAAAAACCCGCTGACC GGCGAGATCATTAGCAAGGGTGAGAGTAAGGAAGAGATCACAAAGGAC CCTATTAACGAACTGACCGAATACGGCCCGGAAACCCTGAGCG*TGGG*C **GAGACC**AGTACC-3'
Cassette 2 – 451 bp	5'- GGTACT**GGTCTC**G*TGGG*CGCGACCATTGGCCCGGAAACCATCGCACCG GGCCATCGTGATGAGTTTGATCCGAAGTTACCTACCGGTGAGAAGGAAG AAGTTCCTGGTAAGCCGGGTATTAAAAACCCTGAGACAGGCGATGTGGT TCGTCCGCCTGTTGACAGCGTTACCAAATACGGTCCTGTTAAAGGCGAC AGTATTGTGGAAAAAGAGGAAATCCCGTTCGAAAAAGAACGCAAATTTAA TCCTGATTTAGCACCGGGCACCGAAAAAGTGACCCGTGAGGGTCAAAAA GGTGAGAAGACCATTACAACCCCTACACTGAAAAACCCGCTGACCGGCG AGATCATTAGCAAGGGTGAGAGTAAGGAAGAGATCACAAAGGACCCTAT TAACGAACTGACCGAATACGGCCCGGAAACCACCGTTATTGG*TCTG*C**GA GACC**AGTACC -3'
Cassette 3 – 446 bp	5'- GGTACT**GGTCTC**C*TCTG*GCGAGCGGCCCGGAAACCATCGCACCGGGCC ATCGTGATGAGTTTGATCCGAAGTTACCTACCGGTGAGAAGGAAGAAGT TCCTGGTAAGCCGGGTATTAAAAACCCTGAGACAGGCGATGTGGTTCGT CCGCCTGTTGACAGCGTTACCAAATACGGTCCTGTTAAAGGCGACAGTA TTGTGGAAAAAGAGGAAATCCCGTTCGAAAAAGAACGCAAATTTAATCCT GATTTAGCACCGGGCACCGAAAAAGTGACCCGTGAGGGTCAAAAAGGT GAGAAGACCATTACAACCCCTACACTGAAAAACCCGCTGACCGGCGAGA TCATTAGCAAGGGTGAGAGTAAGGAAGAGATCACAAAGGACCCTATTAA CGAACTGACCGAATACGGCCCGGAAACCGCGCTGAGCG*GCAC*G**GAGA CC**AGTACC -3'
Cassette 4 – 449 bp	5'- GGTACT**GGTCTC**C*GCAC*CATTGTGGGCCCGGAAACCATCGCACCGGGC CATCGTGATGAGTTTGATCCGAAGTTACCTACCGGTGAGAAGGAAGAAG TTCCTGGTAAGCCGGGTATTAAAAACCCTGAGACAGGCGATGTGGTTCG TCCGCCTGTTGACAGCGTTACCAAATACGGTCCTGTTAAAGGCGACAGT ATTGTGGAAAAAGAGGAAATCCCGTTCGAAAAAGAACGCAAATTTAATCC TGATTTAGCACCGGGCACCGAAAAAGTGACCCGTGAGGGTCAAAAAGGT GAGAAGACCATTACAACCCCTACACTGAAAAACCCGCTGACCGGCGAGA TCATTAGCAAGGGTGAGAGTAAGGAAGAGATCACAAAGGACCCTATTAA CGAACTGACCGAATACGGCCCGGAAACCGTTATTACCGGT*AGCC*G**GAG ACC**AGTACC -3'
Cassette 5 – 442 bp	5'- GGTACT**GGTCTC**G*AGCC*TGGCGGGCCCGGAAACCATCGCACCGGGCC ATCGTGATGAGTTTGATCCGAAGTTACCTACCGGTGAGAAGGAAGAAGT TCCTGGTAAGCCGGGTATTAAAAACCCTGAGACAGGCGATGTGGTTCGT CCGCCTGTTGACAGCGTTACCAAATACGGTCCTGTTAAAGGCGACAGTA TTGTGGAAAAAGAGGAAATCCCGTTCGAAAAAGAACGCAAATTTAATCCT GATTTAGCACCGGGCACCGAAAAAGTGACCCGTGAGGGTCAAAAAGGT GAGAAGACCATTACAACCCCTACACTGAAAAACCCGCTGACCGGCGAGA TCATTAGCAAGGGTGAGAGTAAGGAAGAGATCACAAAGGACCCTATTAA CGAACTGACCGAATACGGCCCGGAAACCTGTTGTT*GACG*C**GAGACC**AG TACC -3'

Table 2.6 – Circularised pET14b accepting vector and pWT $(E\text{-}G5^2)_5$ cassettes for GG assembly. Bases in bold represent the BsaI recognition site. Underlined and italicised are the four bases, which become complementary overhangs post BsaI digestion.

For SasG-ΔEE, seven synthetic cassettes (outlined in Table 2.7) were designed in-house and purchased from Twist Bioscience, CA, USA. These cassettes were assembled into the circularised pET14b destination vector displayed in Table 2.6, top row.

Component	Forward Sequence
Cassette 1 – 315 bp	5'- ATGGGT**GGTCTC**G*ATCA*TCATCACAGCAGCGAAAACCTGTACTTTCAGGGA TCCGCACCTAAGACCATCACCGAGCTGGCGAAGAAAGTTGAAGAGATTCC TTTTAAAAAAGAACGTAAGTTCAACCCGGATCTGGCACCTGGTACCGAGAA GGTGACACGTGAGGGCCAGAAAGGCGAGAAGACAATCACAACACCTACCT TAAAAAATCCTTTAACAGGCGTGATTATCAGTAAAGGTGAACCGAAAGAGG AGATTACCAAAGACCCGATCAACGAGCTGACAGAGTATGGCCCGG*AAAC*C **GAGACC**ACCCAT-3'
Cassette 2 – 420 bp	5'- ATGGGT**GGTCTC**C*AAAC*CATCGCACCGGGCCATCGTGATGAGTTTGATCC GAAGTTACCTACCGGTGCGAAGGAAGAAGTTCCTGGTAAGCCGGGTATTA AAAACCCTGAGACAGGCGATGTGGTTCGTCCGCCTGTTGACAGCGTTACC AAATACGGTCCTGTTAAAGGCGACAGTATTGTGGCAAAGAGGAAATCCCG TTCGAAAAAGAACGCAAATTTAATCCTGATTTAGCACCGGGCACCGAAAAA GTGACCCGTGAGGGTCAAAAAGGTGAGAAGACCATTACAACCCCTACACT GAAAAACCCGCTGACCGGCGAGATCATTAGCAAGGGTGAGAGTAAGGAAG AGATCACAAAGGACCCTATTAACGAACTGACCGAATACGGCCCGGAAACC A*TTAC*G**GAGACC**ACCCAT-3'
Cassette 3 – 411 bp	5'- ATGGGT**GGTCTC**C*TTAC*CCCGGGTCACCGTGACGAGTTCGATCCTAAGCT GCCTACAGGCGCAAAGGAAGAAGTGCCTGGTAAACCGGGCATCAAGAACC CTGAAACCGGCGACGTGGTTCGTCCGCCGGTTGACAGTGTTACCAAGTAC GGCCCGGTGAAGGGCGATAGCATCGTTGCGAAGGAAGAAATCCCGTTTGA GAAAGAGCGCAAATTCAATCCGGATCTGGCCCCTGGTACCGAGAAGGTTA CCCGCGAAGGCCAAAAGGGTGAAAAAACAATTACAACACCTACACTGAAAA ATCCGCTGACCGGTGTTATTATTAGTAAGGGTGAGCCGAAAGAGGAAATTA CCAAAGATCCGATCAACGAATTAACCGAATACGGTCCGGAAA*CAAT*G**GAG ACC**ACCCAT-3'
Cassette 4 – 418 bp	5'- ATGGGT**GGTCTC**G*CAAT*CACCCCGGGCCATCGCGATGAATTTGATCCTAA ATTACCGACAGGCGCGAAAGAGGAAGTGCCGGGCAAGCCTGGTATTAAGA ATCCGGAGACAGGTGATGTTGTGCGCCCGCCGGTTGATAGCGTGACAAAG TATGGCCCTGTGAAGGGCGACAGCATCGTGGCAAAGAGGAGATCCCGTT CAAAAAGGAGCGCAAATTTAATCCGGACTTAGCACCGGGTACAGAGAAGG TTACCCGTGAAGGTCAAAAAGGCGAGAAGACCATTACCACCCCTACCTTAA AGAACCCTCTGACAGGTGAAATCATCAGTAAAGGCGAAAGCAAAGAAGAG ATCACCAAAGATCCGATCAATGAGTTAACAGAATATGGTCCGGAGACAAT*C ACAC*C**GAGACC**ACCCAT-3'
Cassette 5 – 422 bp	5'- ATGGGT**GGTCTC**G*CACA*CCGGGCCACCGCGATGAGTTTGACCCGAAGCT GCCGACAGGTGCAAAGAAGAGGTGCCTGGCAAACCGGGTATCAAGAACC CGGAAACCGGTGATGTTGTTCGCCCTCCGGTGGATAGTGTGACAAAATAC GGTCCTGTGAAGGGTGATAGTATTGTTGCAAGGAAGAAATTCCGTTTGAA AAAGAGCGTAAGTTCAATCCTGATTTAGCCCCTGGCACAGAGAAAGTTACA

	CGCGAAGGTCAGAAAGGTGAGAAAACAATCACCACCCCGACCCTGAAGAA TCCTTTAACCGGCGAAATCATCAGTAAGGGCGAGAGTAAGGAAGAAATTAC AAAAGACCCTATTAATGAATTAACAGAGTATGGTCCTGAAACAATCACACCT G*GCCA*C**GAGACC**ACCCAT-3’
Cassette 6 – 417 bp	5’- ATGGGT**GGTCTC**G*GCCA*CCGTGACGAATTCGATCCGAAACTGCCTACCGG TGCAAAAGAAGAGGTTCCTGGCAAGCCTGGCATTAAGAACCCGGAAACCG GCGATGTGGTGCGTCCGCCTGTGGACAGTGTTACAAAATATGGCCCGGTG AAAGGCGATAGCATTGTGGCAAAAGAGGAGATTCCTTTTAAGAAGGAGCGT AAATTCAACCCTGACCTGGCCCCGGGTACAGAAAAGGTGACCCGCGAGGG CCAAAAGGGCGAAAAAACCATCACCACACCGACATTAAAAAACCCTTTAAC AGGCGAGATCATCAGCAAAGGTGAGAGCAAAGAAGAAATCACCAAAGACC CGATTAATGAACTGACAGAGTACGGCCCTGAGACAATCACCCCTGGTC*AC CG*C**GAGACC**ACCCAT-3’
Cassette 7 – 408 bp	5’- ATGGGT**GGTCTC**C*ACCG*CGACGAGTTCGACCCTAAGTTACCGACCGGTGC AAAAGAAGAAGTTCCGGGTAAACCTGGCATCAAGAATCCTGAAACCGGCG ACGTTGTTCGCCCTCCGGTGGATAGCGTGACCAAATATGGTCCGGTTAAA GGTGACAGTATCGTGGCGAAGGAAGAGATTCCTTTCGAGAAAGAGCGCAA GTTTAATCCGGACCTGGCCCCTGGCACCGAGAAAGTTACACGCGAGGGCC AGAAGGGTGAAAAGACCATCACAACCCCTACCCTGAAGAACCCGTTAACC GGTGAAATTATCAGCAAGGGTGAAAGTAAAGAGGAGATCACCAAAGATCCT ATCAATGAGCTCACCGAGTATGGTCCGGAAACCTGTTGTT*GACG*G**GAGAC C**ACCCAT-3’

Table 2.7 – SasG-ΔEE synthetic cassettes for GG assembly. Bases in bold represent the BsaI recognition site. Underlined and italicised are the four bases, which become complementary overhangs post BsaI digestion. Bases in red represent base pair substitutions to mutate glutamic acid to alanine.

For the GG assembly, the reagents were assembled in 0.2 ml PCR tubes as follows:

Reagent	**Assembly reaction**
pET14b acceptor vector (75 ng/μl)	1 μl
Inserts in amplicon form (range of concentrations)	2:1 molar ratio of insert: vector
T4 DNA Ligase Buffer (10X)	2 μl
NEB Golden Gate Assembly Mix	1 μl
Nuclease-free H_2O	To 20 μl

The basic thermocycler protocol for a 5-10 insert assembly involves three different temperature steps in the assembly reaction: 37 °C (BsaI-HF®v2 digestion within the temperature range for DNA ligase), 16 °C (DNA ligase optimal temperature) and 60 °C (temperature optimal for BsaI-HF®v2 digestion in the absence of DNA ligase activity).

The thermocycle conditions were as follows:

Step	Temperature (°C)	Duration (min)	Cycles
BsaI digestion	37	1	X 30
DNA ligation	16	1	
BsaI cutting in the absence of DNA ligation	60	5	
Hold	4	∞	

E. coli SURE 2 cells were transformed with 2 µl of the reaction mixture as outlined in Section 2.2.2.5.

2.2.2.5 Transformation of Competent *Escherichia coli* (*E. coli*) Cells

50 µl aliquots of competent cells (both for plasmid production and recombinant protein production) were thawed on ice and inoculated with 2-5 µl of 100 ng/µl plasmid DNA (or varying concentrations of assembly products). The tube was gently agitated and left on ice for 20/30 minutes. Subsequently, the bacteria were heat-shocked at 42 °C for the appropriate time (according to cell type) and placed on ice for a further 2 minutes. Next, 950 µl of pre-warmed LB medium was added to the heat-shocked cells and grown for 1 hour at 37 °C, 200 rpm. After 1 hour, 150 µl of cells were plated onto an LB agar selection plate containing the appropriate antibiotic and incubated at 37 °C overnight.

2.2.2.6 One*Taq*® PCR

One*Taq*® PCR was utilised for the rapid screening of *E. coli* colonies after GA and GG assembly to verify the presence of the correct sized insert/assembly before sequencing. The primers used for the screening of pET14b and pMAL-c5x destination regions are outlined in Table 2.8:

Screen	Forward Sequence	Reverse Sequence
pET14b Screen	5'-GACCACAACGGTTTCCCTCTAGAA-3'	5'-TCAAGACCCGTTTAGAGGCCC-3'
pMAL-c5x screen	5'-TCACACAGGAAACAGCCAGTC-3'	5'-GCTTATTTAATTACCTGCAGGG-3'

Table 2.8 – Primers for colony PCR of both pET14b and pMAL-c5x assemblies and inserts, respectively.

The following reaction mixture was set up in 0.2 ml PCR tubes with a final reaction volume of 20 µl:

Reagent	Volume (µl)
One*Taq*® Quick-Load® (2X) Master Mix with Standard Buffer	10
10 µM Forward Primer	0.4
10 µM Reverse Primer	0.4
Template DNA (25 ng/µl)	1
Nuclease-free H_2O	9

One *E. coli* colony was picked using a fresh 0.5 µl pipette tip and added to the reaction mixture.

This mixture was subjected to the following PCR thermal cycle:

Step	Temperature (°C)	Duration (s)	Cycles
Initial denaturation	94	300	
PCR	94 45-68* 68	10 30 60/kb	X 30
Final extension Hold	72 4	120 ∞	

The PCR products were then subjected to gel electrophoresis (Section 2.2.4.1) to determine the size of the inserts.

2.2.2.7 Preparation of Plasmid DNA for Sequencing from *E.* coli 5-α and SURE 2 Cells

5 ml of sterile LB medium with the appropriate antibiotic was inoculated with one colony from an *E. coli* 5-α or SURE 2 selection plate with the same antibiotic. After 16 hours of incubation at 37 °C, 200 rpm cells were harvested via centrifugation at 4000 rpm, 4 °C for 10 minutes (Micro Centaur Plus, MSE). Subsequently, the plasmids were purified from the harvested cells using the QIAprep® Spin Miniprep Kit (QIAGEN Group, Venlo, Netherlands) following the manufacturer's protocol. The concentration of the plasmids were calculated using a NanoDrop 1000 spectrophotometer (1 mm pathlength) by determining the absorbance at 260 nm (A_{260}) and utilising Equation 2.2:

$$Concentration\ (\mu g ml^{-1}) = 50(\mu g ml^{-1}) \cdot A_{260}$$

2.2

2.2.2.8 Sanger DNA Sequencing

All Sanger DNA Sequencing was carried out by Source Bioscience, Nottingham, UK. Primers utilised for sequencing are outlined in Table 2.9:

Primer	Used For	Sequence
T7F	Pentamer/monomeric variant/H_6-TEV-SasG-ΔEE sequencing	5'-TAATACGACTCACTATAGGG-3'
pMAL_F	H_6-MBP-TEV-SasG Sequencing	5'-TCACACAGGAAACAGCCAGTC-3'
MBP-TEV-SasG_1	H_6-MBP-TEV-SasG Sequencing	5'-TATGGAAAACGCCCAGA-3'
MBP-TEV-SasG_2	H_6-MBP-TEV-SasG & H_6-TEV-SasG-ΔEE sequencing	5'- CACCTAAGACCATCACCGAG-3'
MBP-TEV-SasG_3	H_6-MBP-TEV-SasG & H_6-TEV-SasG-ΔEE sequencing	5'- CTGACCGGCGAGATCATTAGCAAG-3'
MBP-TEV-SasG_4	H_6-MBP-TEV-SasG & H_6-TEV-SasG-ΔEE sequencing	5'-TTTGATCCTAAATTACCGACAGGC-3'
MBP-TEV-SasG_5	H_6-MBP-TEV-SasG & H_6-TEV-SasG-ΔEE sequencing	5'- AAAGAGCGTAAGTTCAATCCTG-3'
MBP-TEV-SasG_6	H_6-MBP-TEV-SasG & H_6-TEV-SasG-ΔEE sequencing	5'- AACCCTTTAACAGGCGAGATCATC-3'
Mid_pentamer	Pentamer Sequencing	5'-AAACCCTGAGCGTGGGC-3'
T7_Term_2	Pentamer Sequencing	5'-GCTTCGCTACTTGGAGCCACTATC-3'

Table 2.9 – Sequencing primers for E-G5^2 monomeric variants, H_6-MBP-TEV-SasG, pWT (E-G5^2)$_5$ and variants thereof (pentamer) and SasG-ΔEE .

2.2.3 Gene Expression, Protein Production and Purification

2.2.3.1.1 Starter Culture

200 ml of sterile LB medium (Section 2.1.13.1) containing the appropriate antibiotic(s) (at a final concentration of 25 µg/ml chloramphenicol and/or 100 µg/ml ampicillin) was inoculated with one colony from an 100 µg/ml ampicillin selection plate with freshly transformed *E. coli* BL21 (DE3) (for H_6-MBP-TEV-SasG) or BL21 (DE3) pLysS (for H_6-TEV-SasG-ΔEE, pWT (E-G5^2)$_5$ and variants thereof). These cultures were incubated for 15 hours at 37 °C, 200 rpm.

2.2.3.1.2 Gene Expression and Production of H_6-MBP-TEV-SasG, H_6-TEV-SasG-ΔEE and pWT $(E\text{-}G5^2)_5$ and Variants Thereof

In a 2 L baffled conical flask, 500 ml of autoinduction medium (Section 2.1.13.2) supplemented with 25 µg/ml chloramphenicol (H_6-TEV-SasG-ΔEE and pWT $(E\text{-}G5^2)_5$ and mutants thereof) and/or 100 µg/ml ampicillin (H_6-TEV-SasG-ΔEE) was inoculated with starter culture to a final OD_{600} of 0.05 and grown at 28 °C, 200 rpm. Cells were harvested after 24 hours via centrifugation at 6,000 x *g* (4 °C) for 20 minutes using a Beckman Avanti J-26 XP centrifuge with a JLA-8.1000 fixed-angle rotor and directly resuspended into IMAC buffer A1 (Section 2.1.14.1) for H_6-MBP-TEV-SasG and H_6-TEV-SasG-ΔEE or A2 (Section 2.1.14.2) for pWT $(E\text{-}G5^2)_5$ and mutants thereof at 50 ml per 10 g of pellet and/or stored at -20 °C. Once resuspended, 1 rice grain of both DNaseI and Lysozyme per 10 g of pellet was added (the latter retains around 60 % of its activity in high denaturant conditions)[167].

2.2.3.1.3 Cell Lysis

For all constructs, cells were lysed by sonication (Sonics Vibra-cell, 6 mm Microtip, Amplitude 75 %) on ice for 3 seconds on / 7 seconds off for a total 'on' time of 90 seconds, twice, with a rest of 5 minutes in between. After cell disruption, the lysate was clarified via centrifugation for 45 minutes at 25,000 rpm, 10 °C (Beckman Avanti J-26 XP centrifuge with a JA-25.50 Fixed-Angle Rotor). The supernatant was collected for subsequent protein purification.

2.2.3.1.4 Purification of H_6-MBP-TEV-SasG and H_6-TEV-SasG-ΔEE

All the following steps were carried out on an ÄKTAprime plus (GE Healthcare) at 25 °C. The clarified supernatant from cell lysis was loaded onto a lab-packed 200 ml pre-charged Ni Sepharose™ Fast Flow column (Cytiva) pre-equilibrated in IMAC wash buffer A1 supplemented with 8 M urea (Section 2.1.14.1) at 0.5-1 ml/min overnight. Subsequent refolding, whilst washing, at 3 ml/min from IMAC buffer A1 into IMAC wash buffer (Section 2.1.14.3) was induced using a 0-100 % gradient over 4 column volumes. The resin was washed in IMAC wash buffer until baseline A_{280} was achieved. The bound protein was eluted isocratically at ~ 50 (10 % IMAC elution buffer – Section 2.1.14.4) and ~ 150 mM (20 % IMAC elution buffer) imidazole concentration for MBP-TEV-SasG and H_6-TEV-SasG-ΔEE, respectively. Only the single peak observed was collected and

concentrated using a Vivaspin® 50000 (H_6-MBP-TEV-SasG) or 30000 MWCO (H_6-TEV-SasG-ΔEE) (Sartorius) for dilution in IMAC wash buffer without imidazole or protease inhibitors (Section 2.1.14.5) to give a final imidazole concentration of 5 mM. A 1 ml HisTrap™ FF column (Cytiva, MA, USA) was pre-equilibrated with IMAC buffer A2 (Section 2.1.14.2) and 20 mg of lab-made H_6-TEV protease was loaded onto it. Once washed in IMAC buffer A1, the construct was loaded and cycled through the TEV-bound column for 15 hours, 25 °C at 0.5 ml/min. The flow through containing the protein without H_6-TEV (SasG-ΔEE) or H_6-MBP-TEV (SasG) was collected and concentrated using a Vivaspin® 30000 MWCO (Sartorius), prior to dilution in AEX wash buffer without NaCl (Section 2.1.14.8) to give a final NaCl concentration of 50 mM. Subsequently, this was loaded, at 0.3 ml/min, onto 2 X 5 ml HiTrap SP HP column (Cytiva) stacked atop of 4 X 5 ml HiTrap Q HP columns (Cytiva) pre-equilibrated with AEX wash buffer (Section 2.1.14.6). The protein-bound columns were then washed with AEX wash buffer, until baseline A_{280} was reached. Following removal of the cation exchange columns, the protein was eluted at a 0-30 % (into AEX elution buffer – Section 2.1.14.7) over 15 column volumes at 5 ml/min. The single peak was concentrated using a Vivaspin® 30000 MWCO (Sartorius). Once centrifuged at 16 xg for 15 minutes to remove any insoluble material or aggregates, further purification was performed by gel filtration using a 320 ml Hiload Superdex™ 26/600 75 column (Cytiva) pre-equilibrated with 1 X PBS, pH 7.4. Fractions were analysed by SDS-PAGE to determine the purity. The purest fractions were pooled together, concentrated to 1 mg/ml, snap frozen in liquid nitrogen and stored at -80 °C for long-term storage.

2.2.3.1.5 Purification of pWT $(E\text{-}G5^2)_5$ and Mutants Thereof

All the following steps were carried out on an ÄKTAprime plus (Cytiva). The clarified supernatant from cell lysis was loaded onto a lab-packed 200 ml precharged Ni Sepharose™ Fast Flow column (Cytiva) pre-equilibrated in IMAC buffer A2 (Section 2.1.14.2) at 0.5-1 ml/min overnight. Subsequent refolding, whilst washing, at 3 ml/min from IMAC buffer A2 into a IMAC wash buffer using a 0-100 % gradient over 4 column volumes (2.1.14.3) until baseline A_{280} was achieved. The bound protein was eluted isocratically at ~ 100 mM imidazole concentration (20 % IMAC elution buffer – Section 2.1.14.4). Only the single peak was collected and dialysed into 10 L of AEX wash buffer (section 2.1.14.6) at 4 °C for 16 hours. Subsequently, this was loaded, at 0.3 ml/min, onto 2

X 5 ml HiTrap SP HP column (Cytiva) stacked atop of 4 X 5 ml HiTrap Q HP columns (Cytiva) pre-equilibrated with AEX wash buffer (Section 2.1.14.6). The columns were then washed with aforementioned buffer, until baseline A_{280} was reached. Following removal of the cation exchange columns, the protein was eluted with a linear 50 to 100 mM NaCl gradient (0 to 30 % gradient into AEX elution buffer – Section 2.1.14.7) over 7.5 column volumes (20 ml) at 5 ml/min. The eluted protein was then concentrated using a Vivaspin® 30000 MWCO (Sartorius). Once centrifuged at 16 xg for 15 minutes to remove any insoluble material or aggregates, further purification was performed with gel filtration at 1 ml/min using a 320 ml Hiload Superdex™ 26/600 75 column (Cytiva) pre-equilibrated with 1 X PBS, pH 7.4. Fractions were analysed by SDS-PAGE to determine the purity. The purest fractions were pooled together, concentrated to 1 mg/ml, snap frozen in liquid nitrogen and stored at -80 °C for long-term storage.

2.2.4 Biochemistry Techniques

2.2.4.1 Gel Electrophoresis

Seals were added to an agarose gel casting tray and a 30-well comb was installed. 1.5 % (w/v) agarose gel solution (Section 2.1.14.13) was freshly prepared and microwaved until all the agarose had dissolved and the solution was clear. SYBR™ Safe was added directly before pouring the gel solution into the sealed casting tray. After the gel had set, the seals were removed and the casting tray and gel was placed into a pre-filled (with fresh 1X TAE buffer) gel box together, ensuring the buffer was covering the whole gel. 2 µl of 6X loading dye (Section 2.1.2) was added to 10 µl of DNA sample(s), which were then loaded into the wells alongside both 100 bp and 1 kb DNA ladders (Section 2.1.2). The gel was run at an initial 100 V (per gel) and then turned up to 150 V once the sample had entered the gel until the bands had migrated 70-80 % down the gel. At this point, the gel was removed from the gel box and analysed using a UV light imaging system (Alliance-Q9, Uvitec, Cambridge, UK).

2.2.4.2 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE stacking and resolving gel solutions (Section 2.1.14.12) were freshly prepared, with tetramethylethylenediamine (TEMED) and ammonium persulphate (APS) added directly before pouring the resolving gel solution into freshly cleaned sealed casting chambers. 100 % ethanol was layered on top of the resolving gel to create an even interface with the stacking gel. Once the resolving gel had set, the ethanol layer was removed and the stacking gel mixture was poured on top. Simultaneously, a 14-well comb was inserted into the top of the casting chamber and allowed to set in place. Once the gel had set, the rubber seals were removed and the casting chamber was inserted into the cathode chamber of an electrophoresis tank. This was then filled with 1X cathode buffer (Section 2.1.14.12) and the anode chamber was filled with 1X anode buffer (Section 2.1.14.12). 10 µl of protein sample was then mixed with an equal volume of 2X SDS loading buffer (with freshly added DTT for each gel ran – Section 2.1.14.12) and boiled at 100 °C for 5 minutes. 10 µl of this solution was loaded into the wells (14 in total) of the stacking gel, alongside 10 µl of pre-stained protein standard protein ladder (Section 2.1.2) in a single well for molecular weight determination. Once the electrodes were connected to a power supply, a voltage of 20 amps per gel was applied until the stain hit the top of the resolving gel. At this point, the voltage was increased to 40 amps per gel until the stain exited the bottom of the gel. The gel was removed and submerged in Quick Coomassie stain (Neo Biotech, Nanterre, France) and incubated on a rocking table until the samples became visible for analysis.

2.2.4.3 Surface Functionalisation (for SMFS)

2.2.4.3.1 Oxidisation of Silicon Nitride AFM Probe and Surface

1 cm^2 silicon nitride surfaces were cut from larger wafers and cleaned/oxidised alongside AFM probes (MLCT). For the cleaning/oxidising stage, both the AFM cantilevers and surfaces were incubated in piranha solution (3:1 ratio of 0.5 M (95 %) sulphuric acid to 30 % (v/v) hydrogen peroxide) for 3 and 30 minutes, respectively. Both were washed in Milli-Q water, dried with N_2 and then placed on a microscope slide inside a Petri dish with a punctured lid. These were subjected to ozone cleaning under a UV lamp at 254 nm (UVlite, UVltec) for 30 minutes.

2.2.4.3.2 Aminosilination of Silicon Nitride AFM Probe and Surface

Cleaned and oxidised surfaces and probes were placed in a desiccator along with 20 µl of N,N-diisopropylethylamine (DIPEA) and 80 µl of (3-Aminopropyl)triethoxysilane (APTES) held in separate 1.5 ml Eppendorf tube lids. The desiccator was evacuated using a vacuum pump for ~ 2 minutes and left to incubate. After 2 hours of incubation, the DIPEA and APTES were removed and the desiccator was flooded with N_2 and left to cure for 48 hours.

2.2.4.3.3 Functionalisation with NHS-PEG_{24}-maleimide Linkers

Aminosilanised surfaces and AFM probes were immersed in ~ 1 ml chloroform containing 20 µl of 250 mM (succinimidyl-[(*N*-maleimidopropionamido)-tetracosaethyleneglycol] ester linkers ($SM(PEG)_{24}$, Sigma-Aldrich) in DMSO. After incubation at room temperature for 1 hour, both surfaces and AFM probes were washed with chloroform, then Milli-Q water and dried with N_2.

2.2.4.3.4 Protein Immobilisation

130 µl of 1 mg/ml protein containing two N-terminal engineered cysteines were deposited on the $SM(PEG)_{24}$ functionalised surface and AFM probe in 1X TBS pH 7.4, 1 mM TCEP. TCEP is required to reduce the disulphide linkage of proteins, consequently leading to a larger amount of protein available for immobilisation. After incubation for 30 minutes at room temperature, unreacted protein was washed from the AFM probe and surface with 1X TBS pH 7.5 $\pm$ 0.5 M NaCl.

2.2.5 Biophysical Techniques

2.2.5.1 Circular Dichroism (CD)

Far UV (180-260 nm) circular dichroism spectroscopy was exploited to obtain secondary structural information of the purified proteins in solution using a Chirascan CD spectrometer (Applied Photophysics®). 200 µl of 0.15-0.2 mg/ml protein solution in 5 mM Na_2HPO_4, pH 7.4 was transferred into a 1 mm pathlength cuvette (Hellma) and incubated at 25 °C before a far-UV spectrum (180-260 nm) was obtained using a 2 nm bandwidth, 1 s timestep (25 °C) with an average of 3 scans per sample (including 3 scans of buffer blanks). For protein in a urea solution, a narrower 210-260 nm spectrum was obtained due to saturation of the detector.

Chirascan spectrometer CD spectra was reported in ellipticity (θ, mdeg) and then converted into molar ellipticity ([θ], deg·cm²/dmol) using Equation 2.3.

$$[\theta] = \frac{\theta}{10 \cdot c \cdot l}$$

2.3

Where c is protein concentration (M) and l is cuvette pathlength (cm). This was then converted into mean residue ellipticity ($[\theta]_{MRE}$, deg·cm²/dmol/peptide bond) using Equation 2.4 to allow comparison between constructs.

$$[\theta]_{\mathrm{MRE}} = \frac{[\theta]}{R - 1}$$

2.4

Where R is the number of residues in the protein. The HT signal (V) as a function of wavelength was reported alongside CD spectra to demonstrate that we were measuring a reliable CD signal as enough photons were reaching the detector (< 600 V).

2.2.5.2 Fluorescence Spectroscopy

For equilibrium denaturation experiments of pWT $(E\text{-}G5^2)_5$ and variants thereof, 1X PBS, pH 7.4 and 1X PBS + 9 M urea, pH 7.4 were prepared alongside 1 mg/ml of the protein being analysed in 1X PBS, pH 7.4. Five stocks of 0.2 mg/ml containing 0, 2, 4, 6, 8 M urea were prepared and used to create 40x 750 µl solutions increasing in urea concentration by 0.2 M (from 0 to 8 M). After equilibration for 16 hours at 25 °C in a circulating water bath (NesLab, MA, USA), the samples were placed in a 1 ml 10 mm pathlength cuvette (Hellma, Southend-on-Sea, UK). Initially, the tertiary fold of the proteins were assessed in the absence of denaturant (1X PBS, pH 7.4) and in the presence of urea. A Photon Technology International (PTI) fluorometer was used to excite the tyrosines at 276 nm and an emission spectrum between 290 and 400 nm was taken (3x repeats and buffer blanks taken).

For the equilibrium denaturation of proteins, a time-drive scan was taken. The samples were excited at 276 nm and the average emission intensity was recorded for 30 seconds (1 second per data point) at 305 nm at 25 °C. The excitation and emission slit widths were set to 1 and 20 nm, respectively. The average signal was calculated for each concentration and plotted as a function of denaturant concentration. A two-state transition chemical denaturant model (Equation 2.5) was fit to the data using Igor Pro 7.02 (Wavemetrics, OR, USA) software.

$$S_{obs} = \frac{(a[D] + b)\left(\frac{\exp(\Delta G_{UN} - M_{UN}[D])}{RT}\right) + (c[D] + d)}{1 + \left(\frac{\exp(\Delta G_{UN} - M_{UN}[D])}{RT}\right)}$$

2.5

Where S_{obs} is the observed signal (relative fluorescence units), $b\ and\ d$ are the native and unfolded y-intercepts respectively, a and c are the native and unfolded baseline gradients respectively, R is the ideal gas constant (8.314 J K^{-1} mol^{-1}), $[D]$ is the denaturant concentration, RT is the temperature (298.15 K), ΔG_{UN} is the denaturant-induced unfolding free-energy (J mol^{-1}) and M_{UN} (J mol^{-1} M^{-1}) is the dependence of ΔG_{UN} on denaturation concentration. Figure 2.3 provides a visual explanation of the fit parameters a, b, c, d and S_{obs} for Equation 2.5.

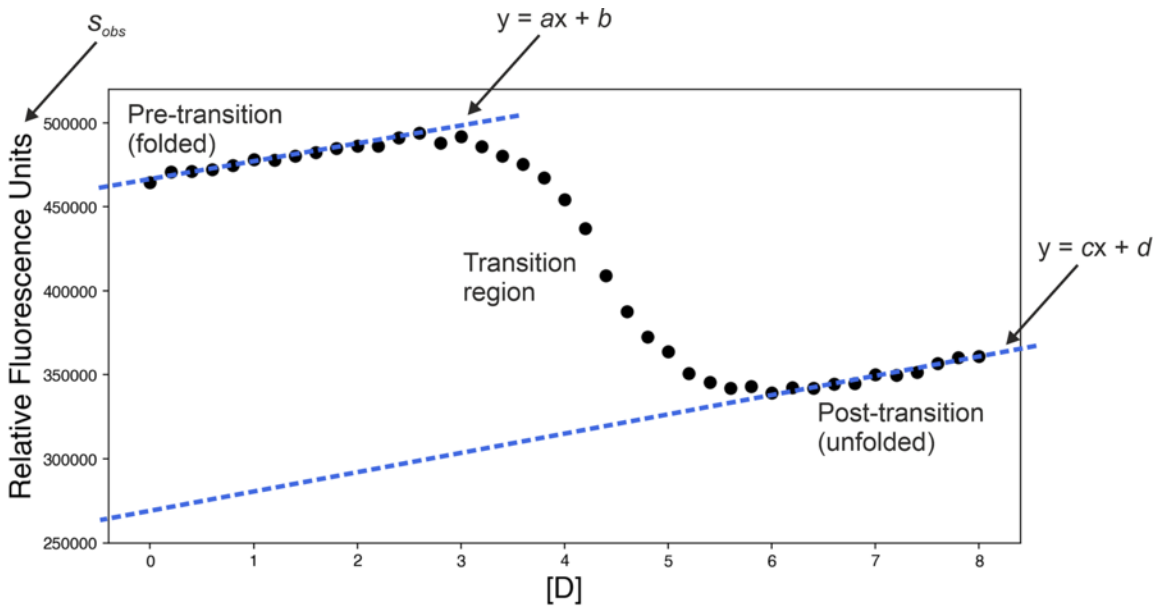

Figure 2.3 – Visual explanation of the extraction of the two-state model fit parameters from a raw protein denaturation curve. S_{obs} is the relative fluorescence units as a function of [D] (black filled circles) and parameters a, b, c and d are determined by fitting straight lines (blue dashed lines) to the pre- and post-transition baselines.

Using Equation 2.6, the S_{obs} and two-state fit was normalised to a fraction of natively folded protein (f_N):

$$f_N = \frac{S_{obs} - (c[D] + d)}{(a[D] + b) - (c[D] + d)} \quad 2.6$$

2.2.5.3 Size Exclusion Chromatography – Multi-angle Laser Light Scattering (SEC-MALLS)

Size exclusion chromatography in combination with multi-angle laser light scattering (SEC-MALLS) was performed on a Wyatt miniDawn TREOS system, equipped with an quasi-elastic light (QELS) scattering detector (Wyatt Optilab T-rEX). A TSKgel® G3000SWXL column (TOSOH, Tokyo, Japan) connected to a Shimadzu HPLC system was pre-equilibrated in the experimental buffer. After baseline collection, 50 μl of 1 mg/ml sample was injected onto the column at 0.75 ml/min. MALLS data was collected for 20 minutes and was processed using the ASTRA 6.1 software (Wyatt

Technology). The $P(\theta)$ (describes the angular dependence of the scattered light of a rod shaped molecule) of SasG was calculated using the rod model as outlined in Equation 2.7:

$$P(\theta) = \left(\frac{1}{u}\right)\int_{0}^{2u} \frac{sin \cdot t}{t} dt - \frac{sin^2 u}{u^2}$$

2.7

Where $u = [(2\pi\, n_O/\lambda_O\, L \sin(\frac{\theta}{2})]$ and L is the rod length (assumed to be much greater than its negligible diameter). The calculated $P(\theta)$ was inserted into Equation 2.8 to calculate the molecular weight (M):

$$\frac{R_\theta}{K^*} = M\, c\, P(\theta)$$

2.8

Where R_θ is the excess Rayleigh ratio of the solution (function of scattering angle (θ) and solute concentration), K^* is an optical constant dependent on the square of the solvent differential refractive index increment (dn/dc) and inverse fourth power of the incident wavelength, and c is the mass concentration of the solute molecules in the solvent.

2.2.5.4 Liquid Chromatography Mass Spectrometry (LC-MS)

LC-MS for recombinant protein mass measurement was carried out by the Biomolecular Mass Spectrometry Facility (University of Leeds, Leeds, UK). Protein desalting and mass analysis was performed by LC-MS using an M-class ACQUITY UPLC (Waters UK, Manchester, UK) interfaced to a Xevo QToF G2-XS mass spectrometer (Waters UK, Manchester, UK). Samples were diluted to 1 µM using 0.1 % trifluoroacetic acid. 1 µl of the 1 µM sample was run on an Acquity UPLC Protein BEH C4 column (300 Å, 1.7 µm, 2.1 mm × 100 mm, Waters, UK) with an Acquity UPLC Protein BEH VanGuard Pre-Column (300 Å, 1.7 µm, 2.1 mm × 5 mm, Waters UK). System flowrate was kept constant at 50 µl/min. Protein sample was loaded on to the trap column in 20% acetonitrile/0.1 % formic acid and washed for 5 min. Following valve switching, the bound protein

was eluted by a gradient of 20-95% solvent B (0.1 % (v/v) formic acid in acetonitrile) in A (0.1 % (v/v) formic acid in water) over 10 min. The column was subsequently washed with 95 % solvent B in A for 5 min before re-equilibration at 20 % solvent B in A ready for the next injection. The mass spectrometer was calibrated using a separate injection of glu-fibrinopeptide. Data were processed using MassLynx 4.2.

2.2.5.5 AFM Force Spectroscopy

2.2.5.5.1 Gold Slide Preparation for Unfolding Experiments

To create the gold surface for gold-thiol protein immobilisation, 1 cm^2 glass surfaces were cut from microscope slides. These were cleaned in the order: water, 2 % (v/v) Hellmanex III, Milli-Q water, 100 % ethanol, Milli-Q water and dried in N_2. A 1:1 ratio mix of epoxy resin part A and B (Epoxy Technology, MA, USA) was deposited in 10 µl spots, per individual glass surface, onto the gold disc (Rockwood Electronic). The clean and dried glass surfaces were deposited at an angle on top of the epoxy spots, ensuring no air bubble formation. The disc was placed on an hotplate (Stuart Scientific, Staffordshire, UK) set to 118 °C and left for at least 1 hour until fixed.

2.2.5.5.2 Cantilever Calibration

The AFM probe, functionalised with protein for SMFS experiments or left bare for protein unfolding experiments, was inserted into a cantilever holder and secured. The silicon surfaces functionalised with protein or the gold surface were attached to a microscope slide with double-sided tape or superglue (Loctite), respectively, and secured to the XY scanner with magnetic bars. A droplet of reaction buffer was applied to the silicon (1X TBS pH 7.4 $\pm$ 0.5 M NaCl) or gold surface (1X PBS, pH 7.4) and was held by surface tension. The AFM probe in the holder was mounted to the MFP-3D head (Figure 2.4) and approached towards the surface (using the front, left rear and right rear leg wheels) until the probe was fully submerged in the droplet. Using the in-built optics, the laser was positioned to the tip of cantilever D (30 pN/nm – manufacturers spring constant) and the deflection was set to 0 using the PD disc. Using the Asylum Research software (MFP version 14), the cantilever was engaged, which causes a Z-piezo voltage maximum (+150). This indicates a full Z-piezo extension and zero surface contact. Using the front wheel on the MFP-3D head, the

cantilever was approached to the surface until the Z voltage was 70: this assures the piezo was in the middle of its Z range (7.5 µm).

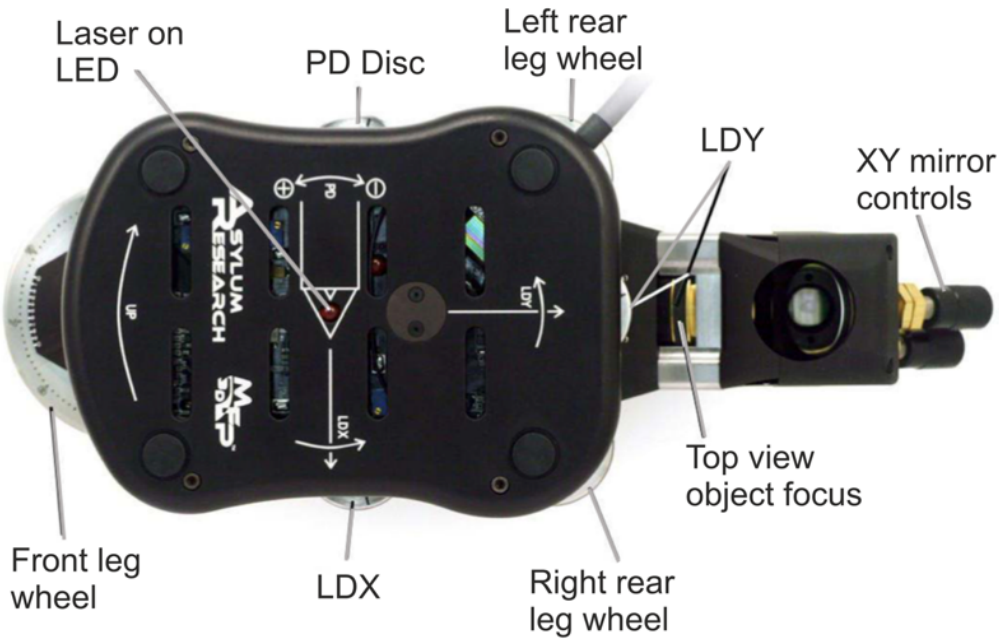

Figure 2.4 – Top view of the MFP-3D AFM head with named controls used in this thesis. LDX and LDY move diode laser along and across cantilever, respectively. Front, left rear and right rear leg wheels control the height of the AFM head relative to the XY scanner (not pictured here). Figure adapted from reference[168].

Spring constant calibration used a non-destructive two-step procedure: determination of the sensitivity of the cantilever (nm/V) from the slope of contact of a single force curve and performing a thermal tune to determine resonant frequency of the cantilever[169]. A single approach-retract cycle was performed with the trigger (amount of deflection the cantilever undergoes before retraction) set to 2.5 nN, which gives a quantifiable deflection slope upon hard contact of the cantilever tip with the surface. The slope of the retract trace contact region (inverse optical level sensitivity) was measured by a linear fit (top to bottom). The virtual deflection baseline was calculated from a linear fit to the horizontal retraction trace. The cantilever was drawn away from the surface by two turns on the front leg wheel (ensures any motion is solely due to thermal fluctuations) and the deflection was set to 0. A thermal tune was carried out to detect the natural thermal fluctuation of the cantilever by performing ~80 frequency sweeps near the resonant frequency of the spring (0-1 MHz)[169]. The drive frequency (the first major resonance peak) was selected and a Lorentzian function was fit.

The area of the thermal fluctuations (P) is used to calculate the spring constant (k) using Equation 2.9:

$$k = \frac{K_B T}{P}$$

2.9

Where T is temperature in Kelvin and K_B is the Boltzmann constant. The spring constant was always within error given by the manufacturer (Bruker) and typically varied from ~ 35-50 pN/nm, with an average of ~40 pN/nm and was re-calculated at the end of the experiment for reliability purposes.

2.2.5.5.3 SMFS Data Collection

2.2.5.5.3.1 Protein:Protein Interactions

The feedback loop trigger point was set to 10 nm (~ 300 pN depending on the spring constant of the cantilever), the start distance was set to 2 µm, retraction distance set to 600 nm, approach velocity was set to 2 μms^{-1} and kept constant. The retraction velocity was set to 1000 nms^{-1} and the sample rate was set to 10 kHz per μms^{-1} velocity. Frequent replenishment of the buffer on the surface was to prevent drying out of the sample. If the buffer was being changed, the surface would be washed in the new buffer > 10 times. 20 μm^2 force maps with 500 (100 x 5 array) approach-retract cycles were taken, with piezo re-positioning (using the XY scanner) between maps to maximise surface coverage and reduce bias.

2.2.5.5.3.2 Protein Unfolding

The protein of interest was immobilised at the C-terminus through covalent bond formation between the sulfhydryl groups of the two C-terminal double cysteines with the gold substrate (gold-thiol bond). From the gold coated disc with epoxy fixed glass surfaces (as outlined Section 2.2.5.5.1), a single glass slide was cleaved off to give a fresh gold substrate. 100 µl of the protein of interest at a concentration of ~0.25 mg/ml in 1X PBS, pH 7.4 was applied to the gold substrate and incubated for 15-30 minutes to allow bond formation. The feedback loop trigger point was set to 2.5 nN, the start distance was set to 2 µm, retraction distance set to 600 nm, approach velocity was set to 2

μms^{-1} and kept constant. Depending on the experiments, five retraction velocities of 200, 800, 1500, 3000 and 5000 nms^{-1} were performed per cantilever tip with sample rates set to 10 kHz per μms^{-1} velocity. Frequent replenishment of the buffer on the surface was to prevent drying out of the sample. 20 μm^2 force maps with 500 (100 x 5 array) approach-retract cycles were taken, with piezo re-positioning (using the XY scanner) between maps to maximise surface coverage.

2.2.5.5.4 Data Processing

All force spectroscopy data were processed using IGOR pro 6.37 with an Asylum Research extension (MFP3DXop v30). The hard contact (x axis – 0 nm) and baseline (y axis – 0 pN) of all the FX retraction traces were manually set on each FX profile before exporting. The WLC model (Section 1.3.2) had to be fit and parameters extracted manually to confirm applicability of the automated data processing Python scripts/software used in this thesis. For each unfolding/unbinding event, the WLC model (fixed persistence length of 0.38 nm) was manually fit by inserting locks at the base of the curve and at the apice. The L_C of protein unfolding events and protein:protein dissociation was calculated by fitting a WLC model to the rising edge of each exponential-like curve when FX profiles were plotted as rupture force as a function of extension.

FX profiles for protein:protein interactions were binned for analysis if: **i)** They exhibited characteristic exponential-like single WLC events and **ii)** The L_C was greater than the length of the $SM(PEG)_{24}$ linkers (20 nm), but smaller than the theoretical maximum for binding if folded. FX (sawtooth) profiles of protein unfolding had different binning criteria depending on the construct. These are outlined in the SMFS sections, but generally required at least 40 % of the number of sub domains present/folded to display unfolding peaks. For a ten sub domain construct, an FX profile would be binned for analysis if four unfolding peaks and a detachment peak (all modelled well by the WLC) were present.

2.2.5.5.5 Processing Protein:Protein Interaction Data Using an Automated Process

Dr Yun Chen (whilst employed at the University of Leeds) developed an automated process for analysing FX data using Python scripting software (utilised in references[170,171]). The .force (force calculated from deflection and spring constant) and .sep (extension) channels from FX data were exported from IGOR pro 6.37 software as .txt files. The horizontal baseline (y axis – 0 pN) was fit to the first 100 nm and the vertical baseline (x axis – 0 nm) was fit to 500 points before the start of the horizontal baseline. If the data deviated from the horizontal baseline, the WLC fit would be attempted (400 fitting attempts). L_C thresholds were set from $20 < L_c < 160$ nm to eliminate events close to the surface (high probability of noise) and higher than the theoretical maximum L_C of folded proteins. The lower rupture force threshold was set to 20 pN as the thermal noise limits of the cantilever prevents detection of unfolding forces much lower than this[44,51]. A complete peak was described in which the entropic restoration part of the peak returns to within 20 pN of the horizontal baseline over 5 nm.

The script cycled through each FX plot and WLC events were fit (as described above). A reduced chi-square value as a quality control parameter was included in the fitting. This was set as a conservative > 10, so real events were not missed. A post-fitting manual screen was carried out to filter any false-positive data. All protein:protein interaction data were analysed utilising this method of data processing.

2.2.5.5.6 Processing Protein Unfolding Data Using Fodis (Force-distance Software)

Manually fitting WLC models to each individual peak of an FX plot using the IGOR pro 6.37 with an Asylum Research extension (MFP3DXop v30) is incredibly labourious and lacks an exportation of fitted parameters option. To semi-automate this process, Fodis, a newly developed software for protein unfolding analysis was utilised[172]. Asylum .force and .sep exported files had to be formatted, using a simple shell script, before uploading into Fodis. The WLC model with a persistence length set to 3.8 Å was fit to peaks above 20 pN outputting the L_C and the rupture force of the fitted peaks. This is exported in a .txt file for downstream analysis.

2.2.5.6 SMFS Data Analysis

Python scripts for the calculation of ΔL_C from the Fodis output .txt file and the automated plotting of histograms and Gaussian fittings (for both protein unfolding and protein:protein interactions) were developed with the help of Dr. Bob Schriffin (University of Leeds).

To visualise the data from protein:protein interactions, datasets were plotted as 2D contour plots (automated using Python scripting software) with bin sizes of 5 nm and 7 pN, for L_C and rupture force, respectively. The same bin sizes were used for Gaussian model fitting to L_C- and force-frequency histograms, which gave the modal values. The mean and the standard deviation (sample) were calculated for any duplicate/triplicate datasets. The percentage of positive hits out of the total number of approach-retract cycles is quoted as the 'hit rate'.

To provide qualitative assessments of shifts in rupture force and ΔL_C, 2D scatter plots of protein unfolding datasets were plotted. For ΔL_C- and force-frequency histograms, bin sizes of 5 Å and 10 pN were utilised, respectively. As before, the automated Gaussian fitting to these provided the modal and full width at half maximum (FWHM) values. The mean and the standard deviation (sample) were calculated for the duplicate/triplicate datasets. The mean was used to calculate the Δforce for variants of pWT $(E\text{-}G5^2)_5$, which was the mean rupture force of the mutant at a particular retraction velocity minus the mean force of pWT $(E\text{-}G5^2)_5$ velocity at the equivalent retraction velocity. The error is the propagated error of standard deviation (sample) of the variables and was calculated according to Equation 2.10:

$$\sigma_x = \sqrt{(\sigma_a)^2 + (\sigma_b)^2}$$

2.10

Where σ_x is the propagated error, σ_a and σ_b are the standard deviations (sample) of the variables.

For the pWT $(E\text{-}G5^2)_5$ (and variants thereof), SasG, SasG-ΔEE constructs, the mean rupture force of triplicate data sets was plotted against the corresponding natural logarithm of retraction velocity, which typically displays a linear relationship. A Python script was developed which automated the

modelling of a linear fit (Equation 2.11) weighted by the inverse of the standard deviation error (sample) of the duplicate/triplicate datasets.

$$y = mx + c$$

2.11

Where m is the gradient and c is the y-intercept. The gradient value helps inform us about basic features of the underlying 1D energy landscape of mechanical protein unfolding and is used in the following mechanical ϕ-value analysis.

2.2.5.7 Mechanical ϕ-value Analysis

ϕ-value analysis is utilised to determine how mutations effect the contacts in the protein at the transition state of mechanical unfolding and is explained in more depth in Section 1.3.5.3[113]. Here we calculate the forced folding (ϕ_F^F).

$\Delta\Delta G_{UN}$ is calculated by measuring the thermodynamic stability (Section 2.2.5.2) of each variant relative to pWT $(E\text{-}G5^2)_5$ and, in turn, assumes the folded ground states of chemical and mechanical denaturation are of similar energy. As the data sets used for this analysis were variants of the pWT $(E\text{-}G5^2)_5$, virtually any loading rate effects would cancel out and the $\Delta\Delta G_{TS-N}^{WT-MUT}$ (change in activation free energy of unfolding) was able to be directly calculated from the mean changes in rupture force and average gradient (Equation 2.12):

$$\Delta\Delta G_{TS-N}^{WT-MUT} = RT(f_{WT} - f_{MUT})/m$$

2.12

Where RT is the product of the molar gas constant and the temperature (2.479 kJ/mol), f_{WT} and f_{MUT} is the mean unfolding forces of pWT $(E\text{-}G5^2)_5$ and variant at 1500 nms^{-1}, respectively. Variants with a gradient (m) agreeable within error of pWT $(E\text{-}G5^2)_5$ were given a fixed average gradient by calculating the mean of all the gradient values. Variants with a deviation in their gradients were calculated using their measured gradients. The latter does, however, preclude direct comparison with pWT $(E\text{-}G5^2)_5$ and other variants as it is likely the elastic properties have changed.

To calculate the forced mechanical folding ϕ-value (ϕ_F^F) Equation 2.13 is utilised:

$$\phi_F^F = 1 - \frac{\Delta\Delta G_{TS-N}^{WT-MUT}}{\Delta\Delta G_{UN}}$$

2.13

Errors were propagated using Equation 2.14:

$$\frac{\sigma_x}{x} = \sqrt{\left(\frac{\sigma_a}{a}\right)^2 + \left(\frac{\sigma_b}{b}\right)^2}$$

2.14

Where σ_x is the propagated error, a and b are measured variables, and σ_a and σ_b are the standard deviations (sample) of the measured variables. Solve for x.

3. Investigating Zinc-mediated Homophilic Interactions of the SasG B-domain

3.1 Objectives

This chapter will begin by introducing the SasG B domain (designated 'SasG' from hereinafter) structure, followed by a characterisation of its biophysical properties including the revisitation of the intriguing mechanical properties. The main focus will be on creating a suitable protein system for SMFS experiments to investigate whether Zn^{2+} mediates SasG dimerisation. The interaction between SasG molecules are investigated in the absence and presence of physiologically-relevant Zn^{2+} concentrations. The creation of a variant, designated SasG-ΔEE, is outlined and its Zn^{2+}-mediated interaction ability is investigated using SMFS.

3.2 Results

3.2.1 Cloning and Purification of SasG

3.2.1.1 Creation of the H_6-MBP-TEV-SasG Fusion Construct DNA

To address the poor yields (< 0.25 mg / 5 L grows) obtained using pETFPP_1-SasG (Section 2.1.17.1) a recombinant MBP-TEV-SasG fusion construct was assembled (schematic displayed in Section 2.1.17.2). This included a TEV cleavage site between MBP and SasG for isolation of SasG during purification. Q5® PCR (as outlined in Section 2.2.2.1) was utilised to amplify the SasG DNA out from pETFPP_1-SasG, whilst simultaneously removing the Strep-tag(II) but retaining the double cysteine and stop codon (primers outlined in Section 2.2.2.2). pMAL-c5X-TEV was linearised and amplified using Q5® PCR (using pMAL-c5X-TEV-POTRA[173] as the vector template). The PCR products were analysed on an agarose gel (1.5 % w/v) stained with SYBR™ safe to confirm successful linearisation of the destination vector and amplification of SasG (Figure 3.1). The SasG amplicon was cloned into the linearised pMAL-c5x vector downstream of the TEV cleavage site using the Gibson Assembly® method (Section 2.2.2.3). 2 µl of the assembly product was transformed (Section 2.2.2.5) into 5-α cells subjected to One*Taq*® (Section 2.2.2.6) colony PCR screening using pMAL-c5X screening primers to confirm an insert of the correct size. The correct sized constructs sequences were confirmed by DNA Sanger sequencing using pMAL-c5X and SasG DNA sequencing primers (Section 2.2.2.8) and the resulting DNA sequence can be found in the Appendix (Section 7.1.1).

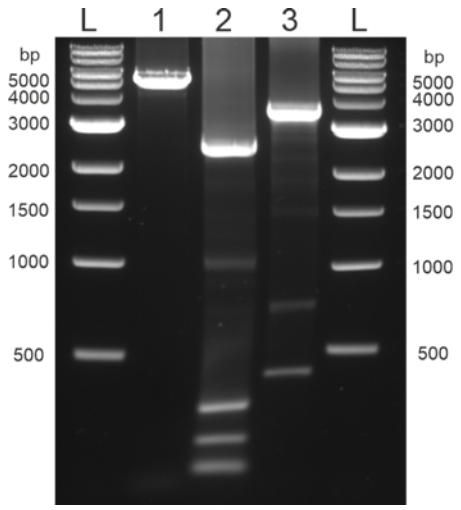

Figure 3.1 – Agarose gel (1.5 % w/v) stained with SYBR™ safe displaying the stages of H_6-MBP-TEV-SasG construct assembly. Lane 1 shows the Q5® PCR amplified and linearised pET-c5X containing MBP and C-terminal TEV cleavage site (5668 bp). Lane 2 shows the Q5® PCR amplified SasG with Gibson Assembly® designed overhangs (2613 bp). Lane 3 shows the product of One*Taq*® colony PCR of assembled H_6-MBP-TEV-SasG (3777 bp) and flanking bases (total: 3866 bp). Shorter length PCR products (lower bands) are observed in lanes 2 and 3 due to non-specific annealing of primers owed to the highly similar DNA between the B-repeats of SasG. L: DNA ladder.

3.2.1.2 Gene Expression, Production and Purification of SasG

H_6-MBP-TEV-SasG under the control of the tac promoter in pMAL-c5X was transformed into *E. coli* BL21 (DE3) cells. Gene expression and protein production was carried out as outlined in Section 2.2.3.1.2. At each step of the purification, chromatograms were recorded and a sample was taken for SDS-PAGE analysis, which are displayed in Figure 3.2. Protein purification was carried out by IMAC using a lab-packed 200 ml pre-charged Ni Sepharose™ Fast Flow column (Cytiva) as described in Section 2.2.3.1.4. After on-column TEV cleavage, the second IMAC step flow through was retained (SasG without H_6-MBP-TEV). Further purification using 2 X 5 ml HiTrap SP HP columns (Cytiva) stacked atop of 4 X 5 ml HiTrap Q HP columns (Cytiva) was carried out. The former columns were utilised to 'pull out' any positively charged proteins from solution prior to anion exchange. The final purification step involved gel filtration (SEC) using a 320 ml Hiload Superdex™ 26/600 75 column (Cytiva). The purity was assessed by SDS-PAGE and LC-MS, which confirmed the successful isolation and identification of SasG. The mass spectra showed a mass of 93939.8 $\pm$ 0.9 as expected (93939.5 Da) for SasG. The yield was increased to ~ 5-8 mg / 5 L grows. The final protein sequence can be found in the Appendix (Section 7.2.1).

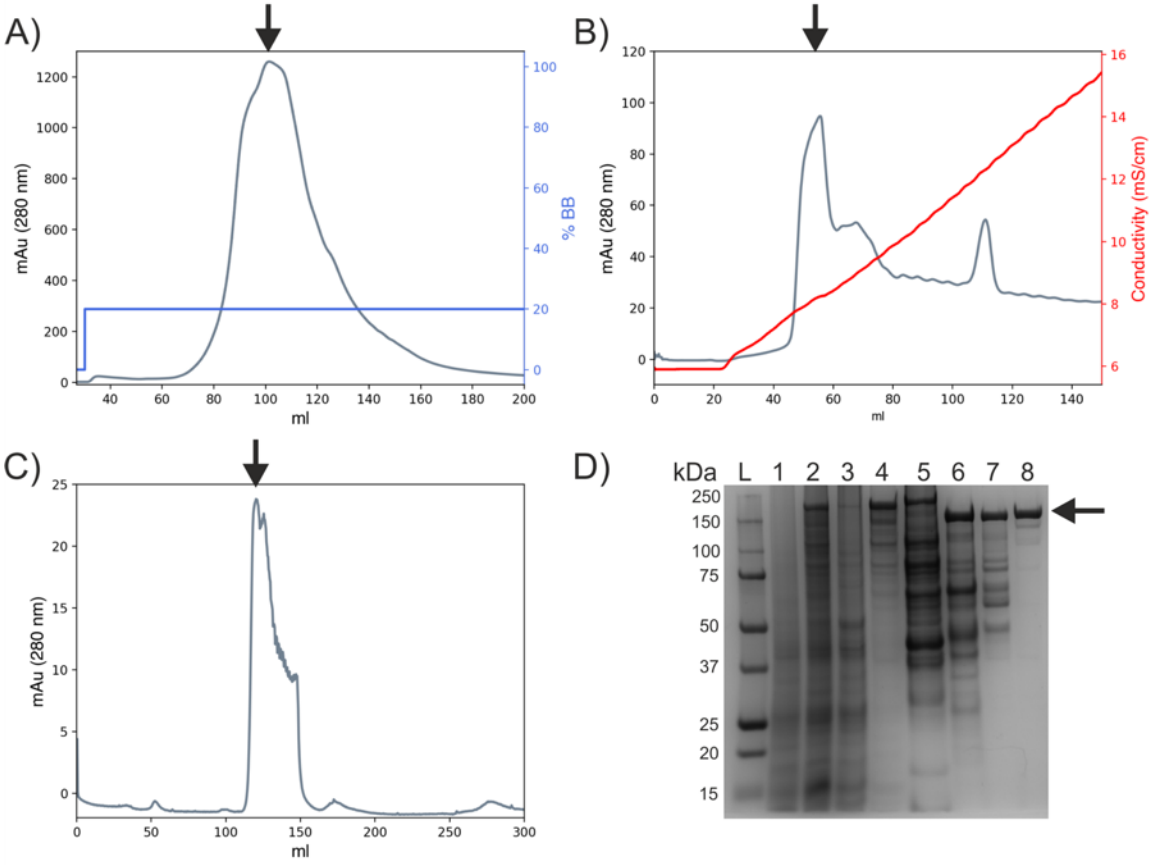

Figure 3.2 – Analysis of the gene expression, production and purification process for SasG from *E. coli* BL21(DE3). Examples of A) the elution chromatogram from an the IMAC purification step, where the blue line represents the elution buffer (Section 2.1.14.4) percentage utilised on the purification system. B) The elution profile from the AEX purification step, where the red line represents the conductivity. C) Chromatogram of the SEC purification step, where pure protein was isolated. Vertical arrows correspond to the peaks which were collected for subsequent purification/isolation and SDS-PAGE analysis. D) SDS-PAGE gel of SasG gene expression, production and purification steps. Lane 1 is protein produced at 0 hours, lane 2 corresponds to protein produced at 24 hours, lane 3 is the pellet (insoluble) fraction and lane 4 is the cytosolic fraction, indicating a soluble protein. Lane 5 is the initial IMAC purification step, lane 6 is the flowthrough from the second IMAC purification step post-TEV treatment, lane 7 is the subsequent AEX purification step and lane 8 is the final SEC purification step, where pure SasG was isolated. Horizontal arrow corresponds to SasG (with H_6-MBP-TEV removed). SasG is ~ 94 kDa in size but gives an apparent mass of > 150 kDa on an SDS-PAGE gel. This anomalous electrophoretic mobility is potentially due to the large number of acidic residues in the E and G5 sub domain repeats (> 20 %) repelling negatively charged SDS molecules as observed in other highly acidic (> 20 %) proteins[174,175].

3.2.2 Spectroscopic Analysis of the SasG Fold

To ensure SasG was correctly folded following purification and in accordance with previously published structures, far-UV CD spectroscopy was carried out as outlined in Section 2.2.5.1.

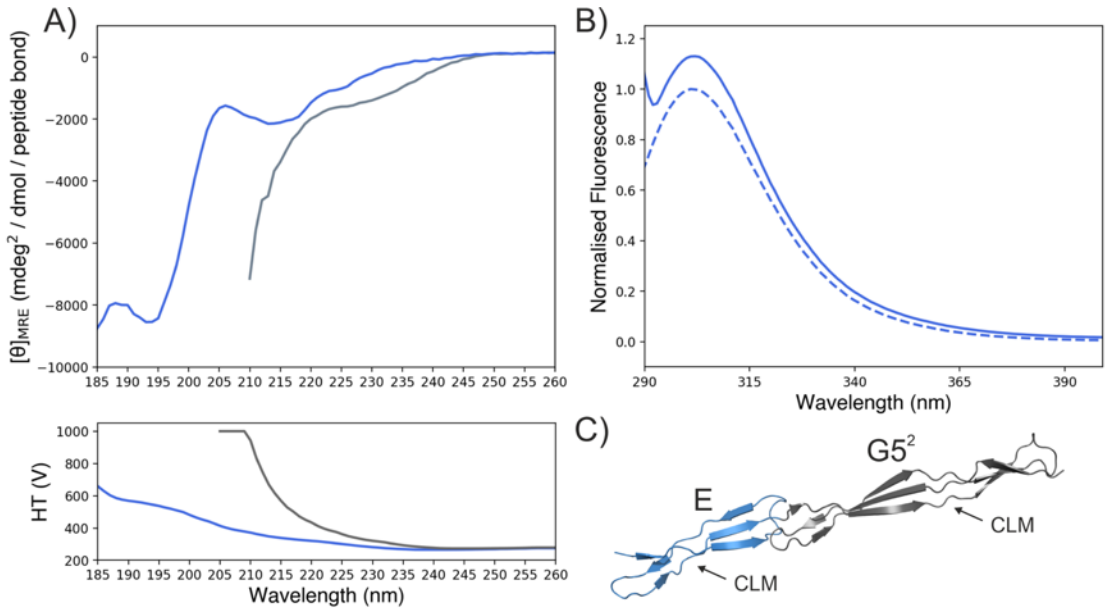

Figure 3.3 – Spectroscopic analysis of the SasG protein fold. A) Far-UV CD spectrum of SasG in the presence (grey) and absence of 6 M urea (blue) with high tension (HT) voltage below. B) Fluorescence emission scans of SasG in the absence (solid) and presence (dashed) of 8 M urea to assess the tertiary structure post-purification. C) Crystal structure of E-G5² (PDB: 3TIP) repeat of SasG displaying major structural features including collagen-like motif (CLM), β-sheet structure and β-turns[149].

Far-UV secondary structure analysis (Figure 3.3A) shows the SasG spectrum to be dominated by a minimum at around 195 nm, the signal for a collagen triple helix[176,177]. Furthermore, a negative minimum at 210-220 nm (characteristic for β-sheets/β-hairpin structure) is observable[178,179]. These signals for a collagen triple helix (E and G5 sub domain contain a collagen-like motif) and β-sheet/β-hairpin topology are consistent with the crystal structures (PDB: 3TIP (Figure 3.3C), 3TIQ and 4WVE)[74,149]. The CD spectrum of full length SasG has not been previously published, however, the spectrum of G5-E-G5 presented by Jasaitis and colleagues displays similar minima to ours[180]. Furthermore, in chemically denaturing conditions (6 M urea), the spectrum is dominated by a signal for a random coil indicating an unstructured polypeptide[178].

To ensure SasG maintained tertiary structure post-purification, intrinsic tyrosine fluorescence was monitored in the presence and absence of 8 M urea (Figure 3.3B) as described in Section 2.2.5.2. 13 tyrosines are found buried in the pseudohydrophobic cores at the G5-E and E-G5 interfaces through the structure. In the absence of urea, SasG has a spectrum typical of a folded protein, with a λ_{max} of 297 nm. Upon addition of 8 M urea, a decrease in fluorescence and red-shift is indicative that the tyrosines are no longer packed in the pseudohydrophobic cores. These spectroscopic analyses suggest that SasG is correctly folded post-purification[180].

3.2.3 SasG is a Mechanically Strong Protein

Previous work by Gruszka and colleagues showed that SasG has remarkable mechanical strength[74]. When SasG was mechanically unfolded using AFM force spectroscopy they observed an FX profile with six small and seven larger unfolding peaks corresponding to the sequential unfolding of E and G5 domains, respectively.

Here we have repeated this study in 1 X PBS, pH 7.4 for comparison with the mechanical properties of the E and G5 sub domains after incorporation into a polyprotein (Chapter 4). SMFS experiments were carried out as outlined in Section 2.2.5.5.3.2. The cantilever tip was retracted from the gold surface at a constant retraction velocity resulting in a FX profile where the rupture force was plotted as a function of the distance between the tip and the gold surface, or extension. This FX profile has the characteristic 'saw-teeth' profile, where each peak ('tooth') reports on the unfolding of a single sub domain. SasG was mechanically unfolded at 5 retraction velocities (200, 800, 1500, 3000 and 5000 nms^{-1}) in triplicate. Consistent with Gruszka and colleagues[74], each retraction velocity displayed six smaller and seven larger unfolding peaks for E and G5 sub domains, respectively (Figure 3.4). As a monomer is extended, sub domains unfold in order of their mechanical resistance and as the E domains are mechanically weaker than the G5 domains, they unfolded first[74].

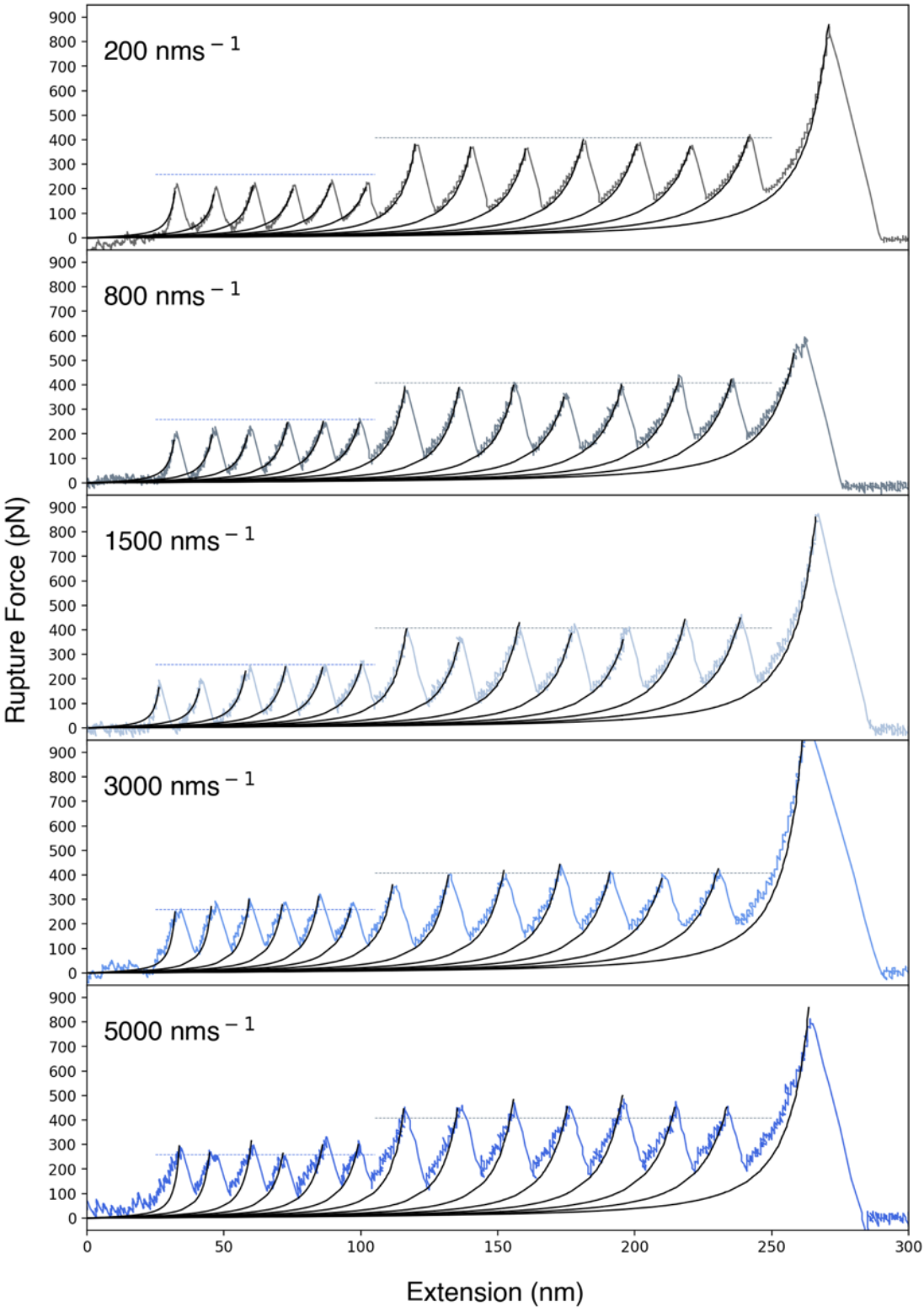

Figure 3.4 – Example SasG FX profiles at a retraction velocity of 200, 800, 1500, 3000 and 5000 nms^{-1} in 1X PBS, pH 7.4 at room temperature. Retraction trace displayed only. The height of the peaks (i.e. rupture force) increases as the retraction velocity increases. WLC model fitting with a persistence length of 0.38 nm in black. Average rupture force of E and G5 domains at 1500 nms^{-1} as blue and grey dotted lines, respectively.

For each mechanical unfolding experiment, traces that contained 6 to 13 full unfolding events which were well described by the WLC chain model were accepted for analysis. From the height of each event apice and L_C values (from the WLC model fit to each unfolding event), the rupture force values were collected and ΔL_C values were calculated, respectively.

The ΔL_C were determined to be 152.3 (ranges from 147.9-157.5) and 218.7 (ranges from 214.4-224.5) Å for SasG E and G5 sub domains, respectively. This is in excellent agreement with the published values of ~ 150 (ranges from 145-154) and ~ 220 (ranges from 216-227) Å for E and G5 sub domains, respectively[74]. Mechanical unfolding forces of 258.4 (ranging from 256.6-276.6) and 408.4 (ranging from 396.2-428.1) pN at a retraction velocity of 1500 nms^{-1} were observed for the E and G5 sub domains, respectively. All unfolding statistics are displayed in Table 3.1 and Table 3.2 for the E and G5 sub domains, respectively.

E **SasG**

Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average ΔL_C (Å)
200	213 102 188	241.5 219.0 228.1	**229.5**	154.5 149.1 151.0	**151.5**
800	117 94 143	267.6 236.1 248.6	**250.7**	153.8 149.0 154.8	**152.5**
1500	87 114 107	276.3 256.6 260.1	**258.4**	154.7 148.8 147.9	**150.5**
3000	79 116 127	285.5 260.0 254.9	**266.8**	157.5 150.0 152.0	**153.2**
5000	71 131 121	293.8 266.3 281.4	**280.5**	157.0 153.2 151.5	**153.9**

Table 3.1 – Summary of rupture force and ΔL_C statistics for SasG E sub domain mechanical unfolding in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode rupture force/ΔL_C are obtained from the Gaussian fits to the histograms. Average is the mean of the mode values at each speed.

G5 **SasG**

Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average ΔL_C (Å)
200	87 117 218	388.6 374.5 386.9	**383.3**	220.9 215.3 218.2	**218.2**
800	133 89 243	422.0 387.7 401.0	**403.6**	218.7 217.7 218.4	**218.3**
1500	82 106 134	433.4 396.2 420.6	**408.4**	222.8 217.2 214.4	**218.1**
3000	102 91 143	428.1 404.1 417.0	**416.4**	222.4 217.0 217.1	**218.8**
5000	124 163 154	450.5 413.3 437.2	**433.7**	224.5 220.4 216.1	**220.3**

Table 3.2 – Summary of rupture force and ΔL_C statistics for SasG G5 sub domain mechanical unfolding in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode rupture force/ΔL_C are obtained from the Gaussian fits to the histograms. Average is the mean of the mode values at each speed.

The published dependence of rupture force on logarithm of retraction velocity of E and G5 sub domains of SasG (in 20 mM Tris, 150 mM NaCl, pH 7.5)[74] was compared to data obtained in this study (in 1X PBS, pH 7.4) (Figure 3.5). While there was no difference in the speed dependence between datasets, the mechanical strength of the E and G5 sub domains were consistently found to be higher in this study. This could be due to a myriad of reasons including, but not limited to, acquisition of fewer unfolding events per pull, errors in spring constant calibration *etc*. However, more importantly, these experiments were carried out in different buffers making direct comparison difficult. Nonetheless, it is clear that SasG E and G5 sub domains have significant mechanical properties.

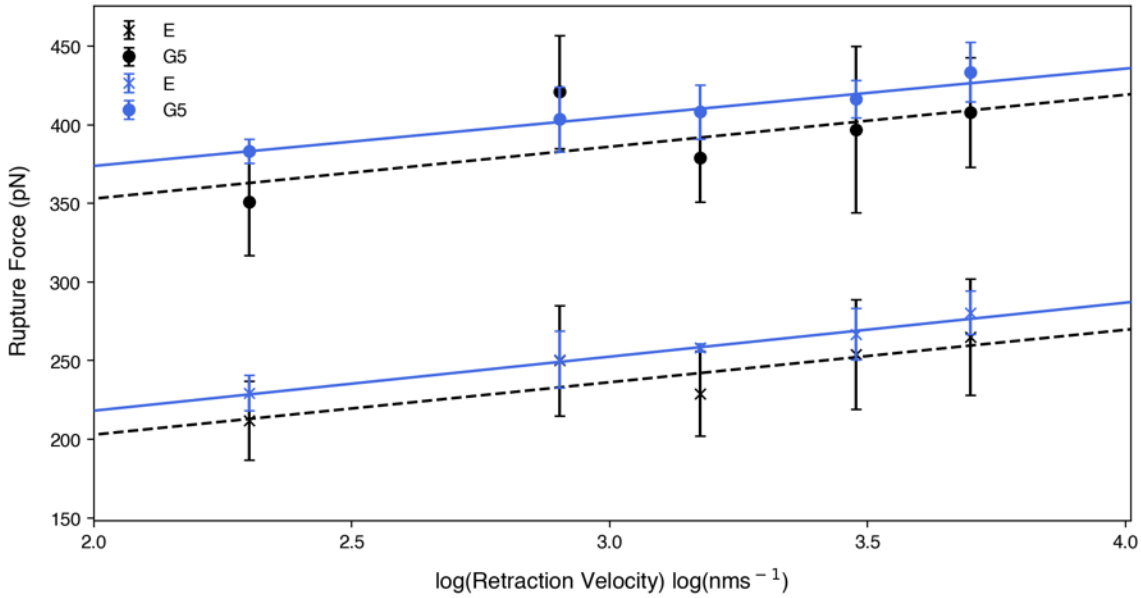

Figure 3.5 – The dependence of rupture force on the logarithm of retraction velocity of SasG sub domains, E and G5. E and G5 data points displayed as crosses and filled circles, respectively. Black and blue data points are taken from reference[74] and from this study, respectively. Blue points and errors are the mean and standard deviation of triplicate datasets from this study. Black points and errors are the mode and standard deviation of triplicate datasets combined. Linear fit is weighted with the inverse of the standard deviation error.

3.2.4 Zinc-induced Oligomerisation of SasG

3.2.4.1 Rationale for Investigation

Zn^{2+} chelation has been observed to inhibit *S. aureus* and *S. epidermidis* biofilm formation, with the addition of 5-20 μM $ZnCl_2$ able to restore the biofilm growth[21]. As Aap, the *S. epidermidis* homolog of SasG, was found to be directly involved in Zn^{2+} dependent biofilm formation, SasG was predicted to have similar properties. This was further investigated by Geoghegan and colleagues, where it was observed that biofilm formation by *S. aureus* SH1000 (pALC2073*SasG*$^{+}$ A^{-}B^{+}) was reduced by the zinc-chelator diethylenetriamine pentaacetate (DTPA) in a concentration-dependent manner[20]. This provided more evidence that it was SasG, more specifically the B-domain, responsible for the Zn^{2+}-induced biofilm formation. However, *in vitro* studies of the Zn^{2+}-mediated interaction of SasG B-domains have shown few promising results, even with Zn^{2+} concentrations being used which are magnitudes higher than are physiologically relevant. More specifically, Geoghegan and colleagues observed dimerisation of both $rB_{2.5}$-GST (2.5X B-repeats, using SPR) and a single B-repeat (using

SEC-MALLS) in the presence of 5 and 10 mM Zn^{2+}, respectively[20]. However, Gruszka and colleagues determined that the dimerisation of the single B-repeats[20] was occurring due to the presence of a non-native histidine residue introduced to facilitate purification[149]. More recently, Formosa-Dague and colleagues observed the clumping of *S. aureus* SH1000 (pALC2073*SasG*$^{+}$) cells in the presence of 1 mM Zn^{2+}[153]. Utilising SCFS they observed interactions between individual SH1000 (pALC2073*SasG*$^{+}$) cells expressing SasG (among other CWA proteins) in the presence of 1 mM Zn^{2+}, which were lost upon the addition of 1 mM EDTA. These Zn^{2+} induced interactions were observed as a combination of single and multiple rupture events (96 % of analysed traces) and saw-tooth profiles (remaining 4 % of analysed traces), with the latter implying multidomain protein unfolding. Furthermore, both cells expressing 1 B-repeat ($SasG_1^{(+)}$) or not expressing SasG at all ($SasG^{(-)}$) were unable to form interactions in the presence of Zn^{2+}. It was also observed that Zn^{2+} alters the structural, mechanical and adhesive properties of the cell surface, which may suggest it is a cell-wall effect. Overall, the understanding of how SasG B-domains are implicated in Zn^{2+}-induced biofilm formation remains elusive.

Here we utilise SMFS to investigate the Zn^{2+}-induced homophilic interactions of SasG and attempt to resolve the molecular details, which to our knowledge has not been attempted before. Zn^{2+} conditions of 10 and 100 μM are utilised, with the former corresponding to blood plasma levels[181] and latter thought to be more relatable to local sites of infection (see Section 3.3.1).

3.2.4.2 SEC-MALLS of SasG in the Presence and Absence of Zn^{2+}

In an attempt to detect any dimerisation/oligomerisation of SasG in solution, SEC-MALLS was carried out in 1X TBS, 2 mM TCEP, pH 7.5 supplemented with 2 mM EDTA or 100 µM $ZnCl_2$. The differential refractive index and molar mass profiles were near identical for both conditions (Figure 3.6). The average molecular weight of SasG in 2 mM EDTA and 100 µM $ZnCl_2$ were calculated utilising the rod model (Section 2.2.5.3) to be 93.0 and 95.6 kDa, respectively. This is in good agreement with the measured mass of 93939.8 ± 0.9 Da and implies SasG is monomeric in solution, regardless of Zn^{2+} presence. This agrees with Gruszka and colleagues results, which found that G5[1], G5[2], E-G5[2], G5[1]-E-G5[2] and E-G5[2]-E-G5[3] constructs remained as monomers in the presence or absence of 5 mM Zn^{2+} when investigated using SEC-MALLS[149]. However, our molar mass profile is not stable and there is presence of a left-hand side shoulder (not observed by Gruszka and colleagues), which calculated higher order species > 500 kDa in size (molar profiles not shown). This may be indicative of oligomers/large aggregates and could suggest SasG is in a monomer-oligomer equilibrium in solution, irrespective of the presence of Zn^{2+}. However, these conditions are vastly different to those found *in vivo* where SasG is anchored at one end to the cell surface[144]. A more physiologically relevant examination is where SasG is tethered at the C-terminal, which is considered next.

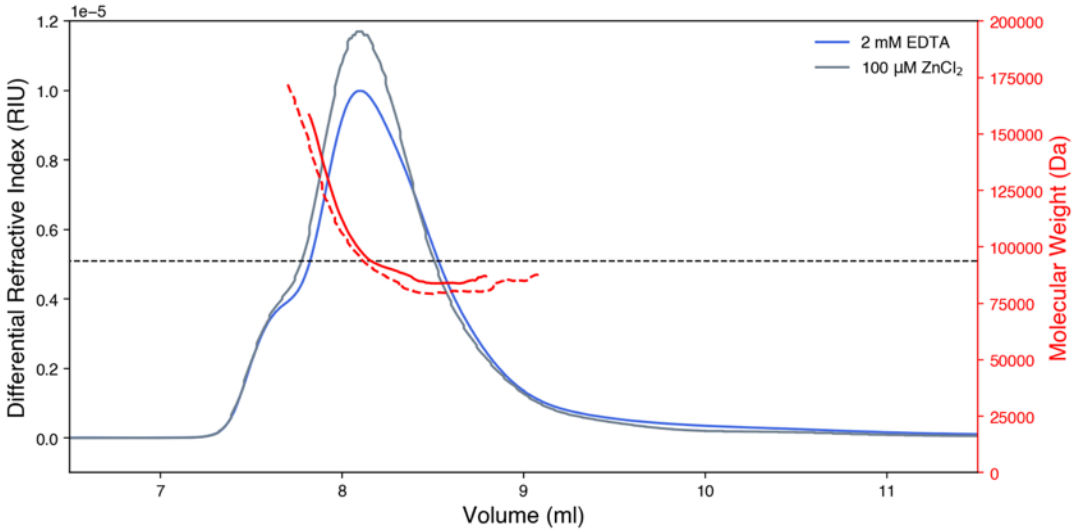

Figure 3.6 – SEC-MALLS of SasG in the presence and absence of 100 µM Zn^{2+}. Differential refractive index plot and calculated molecular weight from the light scattering is shown in a solid and broken red line for 100 µM $ZnCl_2$ and 2 mM EDTA, respectively. The measured molecular weight of SasG is 93939.8 ± 0.9 Da, which is shown as the dashed horizontal black line.

3.2.4.3 Single Molecule Force Spectroscopy

A first aim of this thesis was to design an SMFS system to investigate the potential *trans* homophilic binding of SasG monomers and understand the role of Zn^{2+} in this. A schematic of the SMFS system is outlined in Figure 3.7A. SasG monomers (with C-terminal cysteines) were immobilised onto a silicon nitride AFM probe and silicon nitride substrate using flexible heterobifunctional PEG linkers ($SM(PEG)_{24}$) of ~ 9.5 nm in length. The functionalised tip was brought into contact with the functionalised surface in buffered solution until a 300 pN trigger point was reached (0 s dwell time) and then retracted at 1000 nms^{-1}. Several 20 μm^2 force maps comprising 500 approach-retract cycles were acquired. Between each force map, the cantilever tip was repositioned to ensure sufficient surface coverage and non-bias of areas with a higher hit rate. A typical data set contained 3-4 force maps, with the number of true interaction events depending on the conditions. Interaction events that met the filtering criteria (Section 2.2.5.5.4) were binned for analysis and fitted with the WLC model with a fixed persistence length of 0.38 nm in order to calculate the L_C. Potential *cis* interactions that may occur in this SMFS set up and *in vivo* (see Section 1.4.3.1) are displayed in Figure 3.7B.

3.2.4.3.1 Preliminary SMFS Experiments Investigating Zn^{2+}-associated Complex Formation

An initial experiment utilised 1X TBS (20 mM Tris, 150 mM), pH 7.5 with 2 mM EDTA as the buffer to investigate the interaction between SasG molecules in the absence of Zn^{2+}. EDTA was utilised to ensure any trace metal ions from the purification procedure had been successfully chelated to order to prevent any false positives. This initial condition was then titrated into zinc conditions incrementally from 10 μM to 4 mM $ZnCl_2$ (10 μM, 100 μM, 500 μM, 1 mM, 2 mM and 4 mM). 1500 approach-retract cycles (i.e. three force maps) were taken for each condition with a retraction velocity of 1000 nms^{-1}. The 2 and 4 mM steps are not presented here as it is a far from physiologically relevant concentration of zinc, however, the final step of another 2 mM EDTA (post 4 mM) is still displayed to show the loss of interactions. Example single molecule events are displayed in Figure 3.8 and the contour plots using a bin size of 5 nm and 7 pN for L_C and rupture force, respectively, are displayed in Figure 3.9.

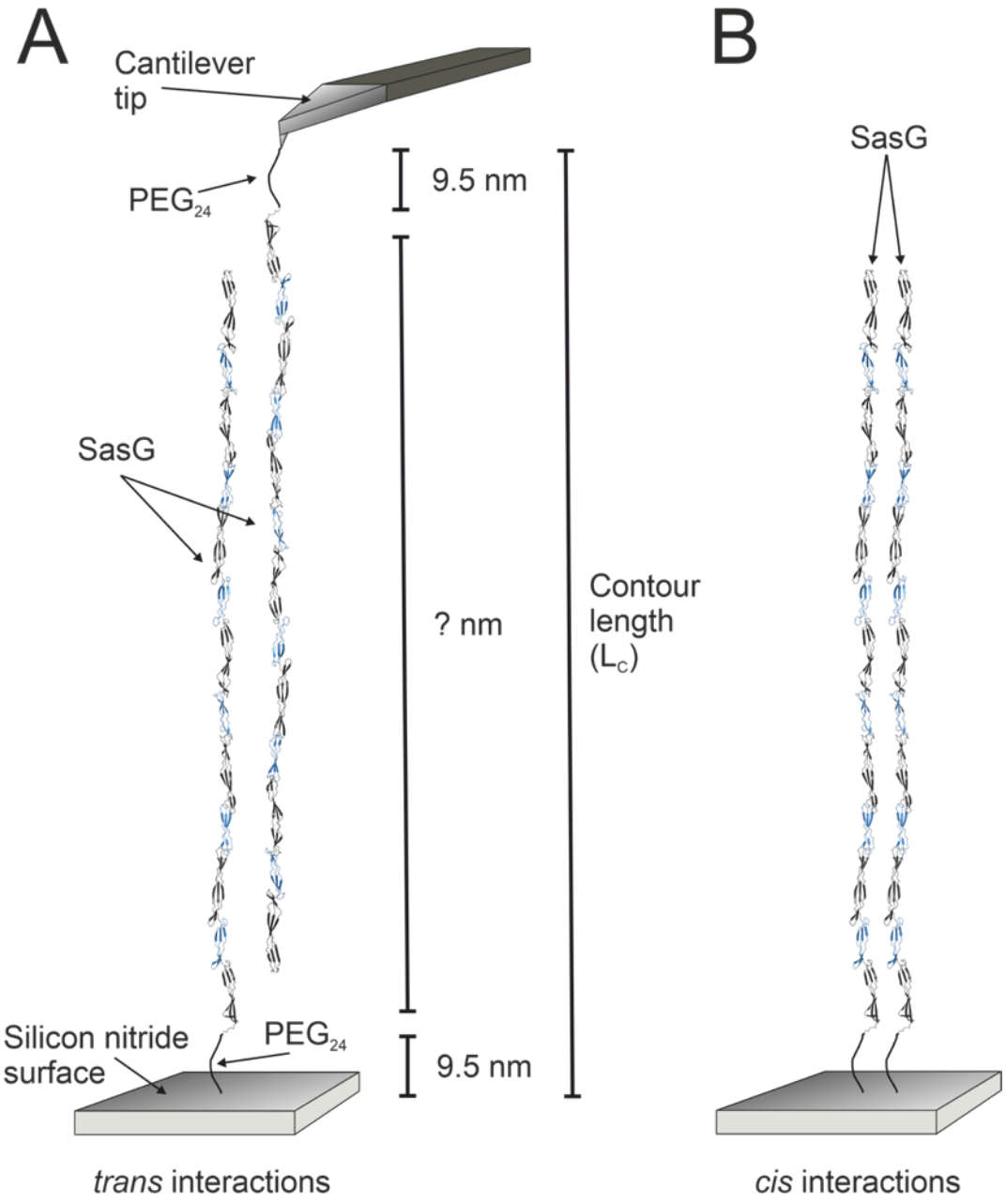

Figure 3.7 – Schematic of SMFS experimental system to investigate the homophilic interactions of SasG monomers. A) SasG molecules are covalently immobilised to the silicon nitride cantilever tip and silicon nitride surface using heterobifunctional SM(PEG)$_{24}$ linkers (black). After the AFM probe has been brought into contact with the surface and SasG molecules have interacted in a *trans* orientation, withdrawing the probe will propagate force across the complex. This force increases until the complex dissociates and at this point the rupture force and L_C can be calculated. B) The potential *cis* interaction, which may also occur as a result of this SMFS set up. The cartoon structure of SasG was generated by iterative superimposition of crystal structures by Dr. Emanuele Paci (University of Bologna) for reference[74].

As Zn^{2+} ions readily bind histidine residues, failure to remove the H_6 tag could possibly result in false-positives[182]. The constructs utilised in this chapter have had their H_6 tags (and any fusion proteins) removed during purification (Section 3.2.1.2 & 3.2.4.5).

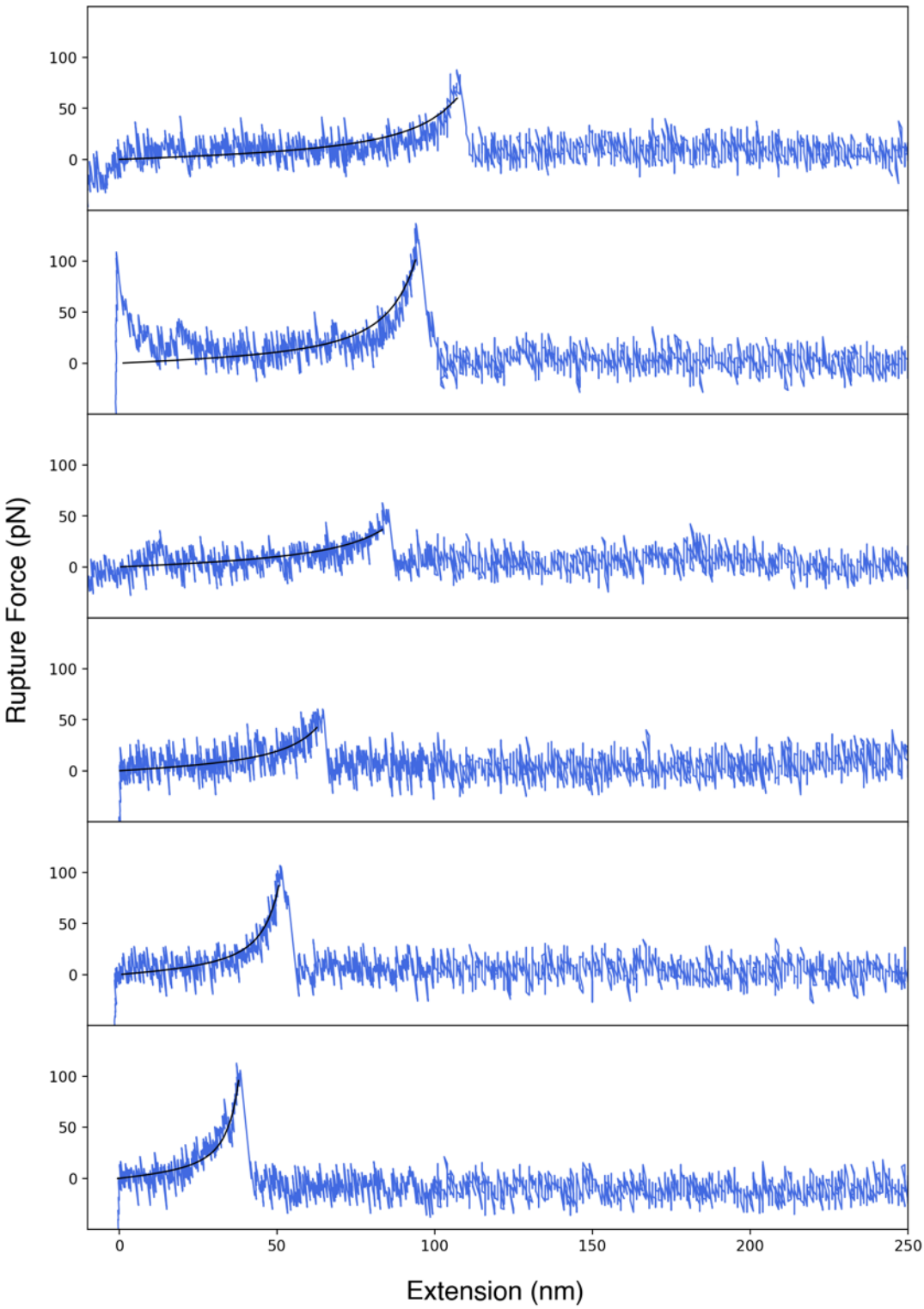

Figure 3.8 – Example FX profiles displaying characteristic single rupture events at a retraction velocity of 1000 nms^{-1}. WLC model fitting with a fixed persistence length of 0.38 nm is displayed as a black solid line. SasG dimerisation events resulted in FX profiles with varying rupture forces and L_C values.

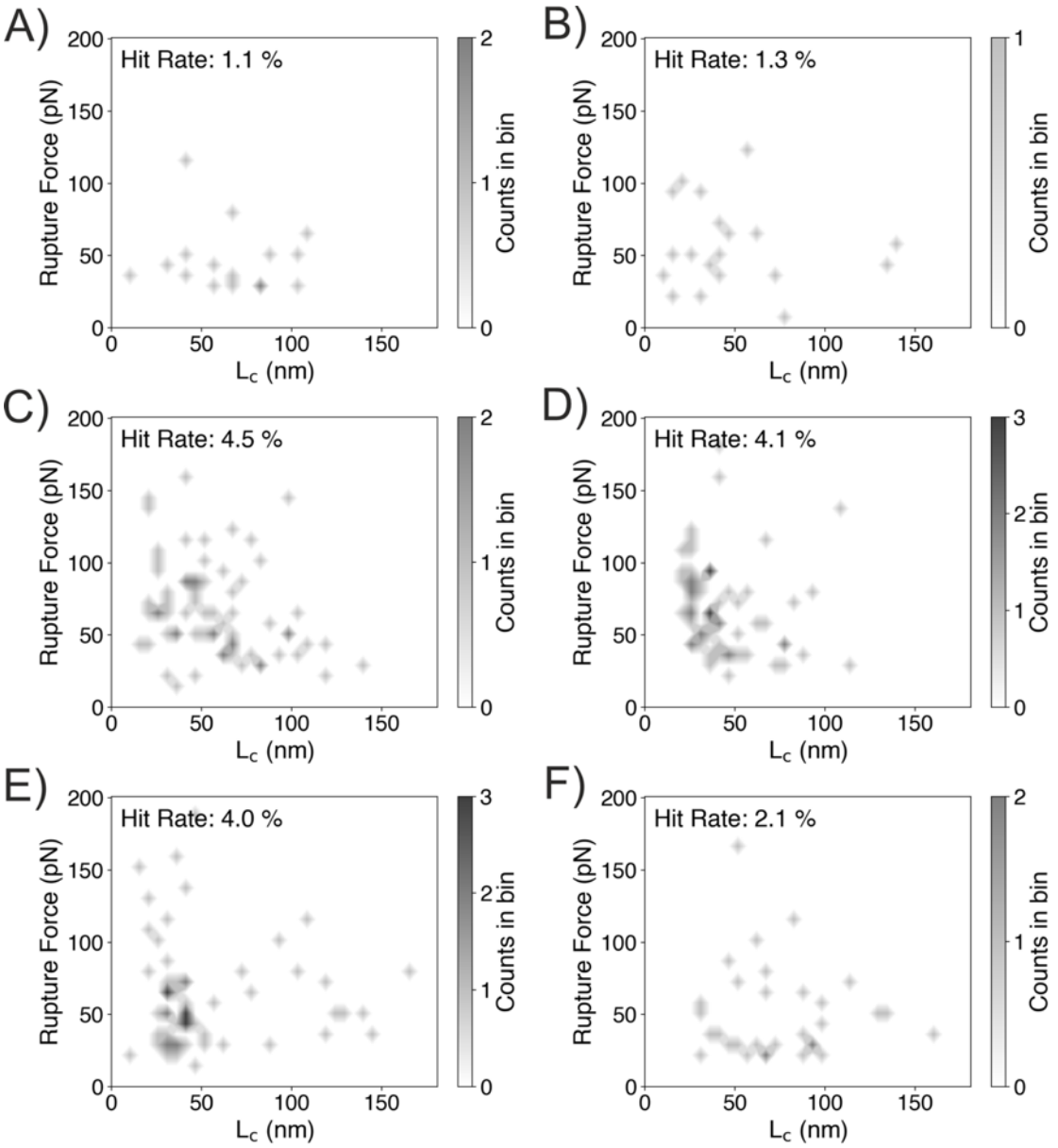

Figure 3.9 – Preliminary zinc titration SMFS experiment in 1X TBS, pH 7.5 + 2 mM EDTA or $ZnCl_2$ in various concentrations at a retraction velocity of 1000 nms^{-1}. The overall number of force-extension cycles for each condition was 1500. A) 2 mM EDTA, B) 10 μM $ZnCl_2$, C) 100 μM $ZnCl_2$, D) 500 μM $ZnCl_2$, E) 1 mM $ZnCl_2$ and F) 2 mM EDTA (post 4 mM $ZnCl_2$ – not displayed here). There is an observable increase in the hit rate and formation of 'hot spots' at $ZnCl_2$ concentrations of 100 μM and above. The marginally higher hit rate in 2 mM EDTA of F) in comparison to A) is probably due to residual Zn^{2+} ions bound to the SasG structure and/or incomplete chelation by EDTA.

The number of analysable curves with more than a single rupture peak was generally < 0.1 % which suggests there are no unfolding events of E or G5 sub domains prior to complex dissociation or a significant number of multiple binding events. As observable from the plots, there is a 4 fold increase in hit rate from 10 μM to 100 μM $ZnCl_2$. Once the zinc concentration had reached 100 μM, there was no further appreciable increase in the hit rate. At higher zinc concentrations the 'hot spot' appears to get more defined, but anything ~ 500 μM and above is less likely to be physiologically relevant. The presence of L_C values > ~ 80 nm (~ 70 nm PyMOL measurement of SasG length in its extended form + 9.5 nm length of an $SM(PEG)_{24}$) gives us confidence that there we are measuring interactions between two SasG molecules. From this preliminary data it was evident that the presence of a threshold concentration (somewhere between 11-100 μM) of Zn^{2+} promotes the dimerisation/oligomerisation of SasG. As we are unable to resolve oligomerisation using SEC-MALLS (Section 3.2.4.2), this suggests the Zn^{2+}-induced interaction observed here is transient in nature.

Unlike DLL mechanisms where hydrogen bonding locks the ligand in place, resulting in shear geometry and in turn very high > 1.5 nN rupture forces (as discussed in Section 1.4.2), we observe lower rupture forces of generally ~ 40 to ~ 125 pN with a few instances near 200 pN. This suggests that our interaction is not mediated by interdomain hydrogen bonding between monomer β-sheet edges in shear geometry and that Zn^{2+} is not electrostatically locking these in place (as observed for some intramolecular mechanical clamps)[76]. In an attempt to disrupt any non-specific ionically-driven interactions and determine whether our Zn^{2+}-mediated interaction is electrostatically driven, we increased the ionic strength of the buffers in the subsequent experiment by adding 0.5 M NaCl.

3.2.4.3.2 SasG Oligomerisation and Zinc

We investigated potential interactions in the presence of 0.5 M NaCl supplemented with 2 mM EDTA, 10 μM and 100 μM $ZnCl_2$ in an attempt to shield non-specific interactions and tease out any Zn^{2+}-coordinated binding. Our preliminary high ionic strength screens displayed more defined 'hot spots' and a lower hit rate. In light of this, the optimised conditions were repeated in triplicate. Each single experiment consisted of 3 force maps of 500 approach-retract cycles and the total number for each condition was 4500. In the presence of 2 mM EDTA or 10 μM $ZnCl_2$ there was a similar hit rate of ~ 1 %, whereas moving into 100 μM $ZnCl_2$ there was a ~ 2.5 fold increase in the hit rate to ~ 2.5 % (Table 3.3). This shows the Zn^{2+}-driven interactions remain in the presence of 0.5 M NaCl indicating they are insensitive to ionic strength.

Condition	Hit rate (%)
2 mM EDTA	1
	0.6
	0.9
10 μM $ZnCl_2$	1.1
	1.1
	1.1
100 μM $ZnCl_2$	2.6
	2.6
	2.4

Table 3.3 – Table displaying the hit rate for each replicate for each condition for SasG dimerisation by SMFS. From 10 μM $ZnCl_2$ to 100 μM $ZnCl_2$ the hit rate increases two fold (students T test $p < 0.05$).

The hit rates observed are low (between 0.6-2.8 %), thus increasing the probability that we are observing single molecule complex dissociation events[183]. Due to this low hit rate, a single contour-histogram graph containing all triplicate data was plotted per condition to observe any 'hot spots', which may be missed if each replicate was plotted separately (Figure 3.10). Upon observation of the contour plot for 100 μM $ZnCl_2$, there is a clear 'hot spot' spreading from ~30-130 nm. In addition, the L_C histogram appeared to display a multimodal distribution, with a dominant and minor L_C population of 63.3 and 111.3 nm, respectively. Gaussian models fit to the force frequency-histogram gave rupture forces of 43.0 and 75.7 pN.

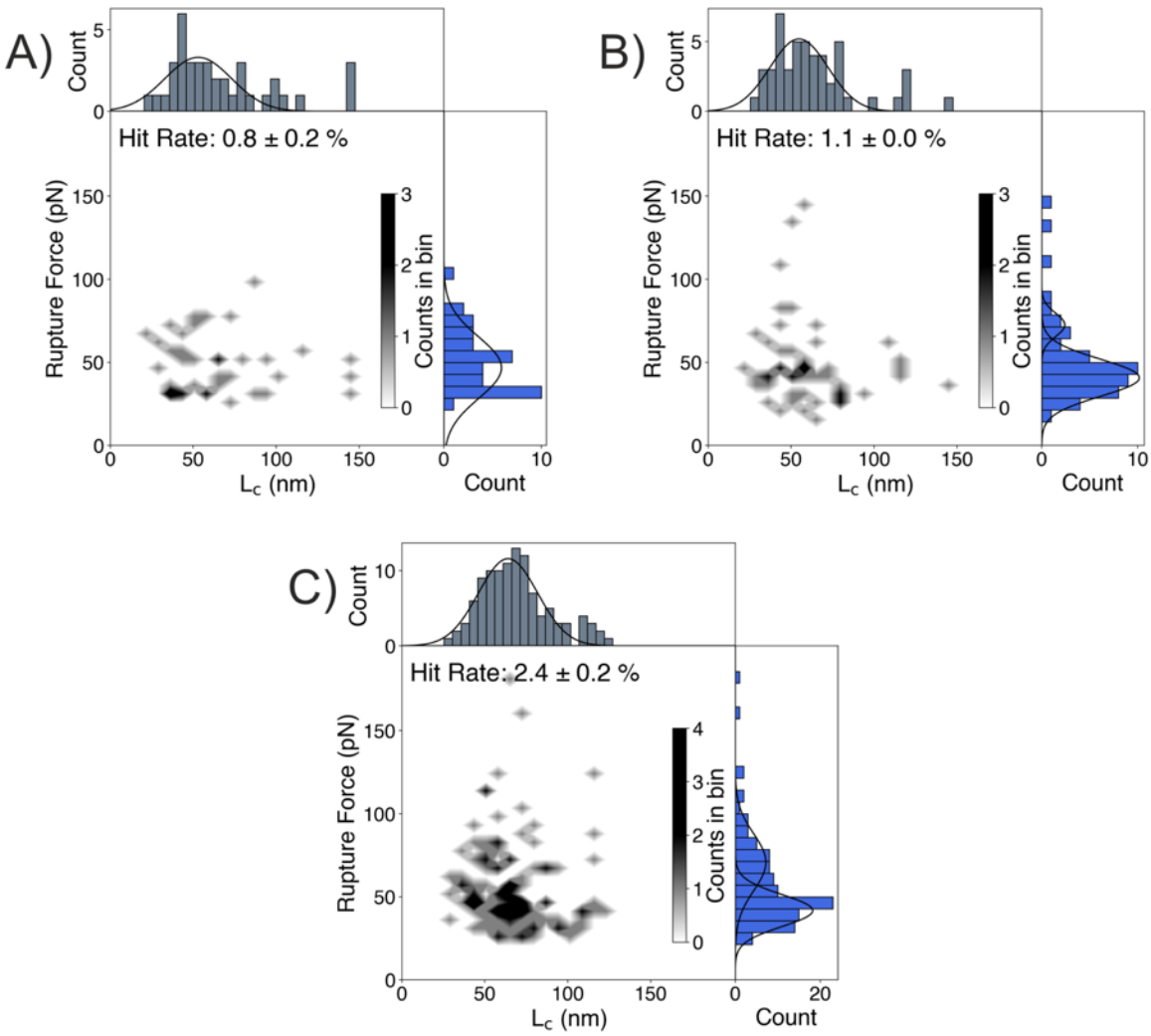

Figure 3.10 – SMFS data from combined triplicate experiments for SasG in 1X TBS + 0.5 M NaCl, pH 7.5 + 2 mM EDTA or $ZnCl_2$ in various concentrations at a retraction velocity of 1000 nms^{-1}. A bin size of 5 nm and 7 pN for the L_C and rupture force, respectively, was utilised for the contour plot and histograms. Contour plots of A) 2 mM EDTA, B) 10 μM and C) 100 μM $ZnCl_2$. The hit rate is displayed inset.

As the rupture force values appear to deviate slightly from a Gaussian distribution, box plots were plotted for each 100 μM $ZnCl_2$ replicate. L_C box plots are also included for comparison with the Gaussian fits. Median values from the box plots (Figure 3.11) are comparable to the modal L_C (of the dominant Gaussian fit) and lower (more dominant) modal rupture force from the histograms. Furthermore, the majority of the data sits within the interquartile range and the medians were similar between replicates. Overall the data indicates we are observing single molecule interactions between SasG molecules with an increase in interactions driven by the presence of 100 μM $ZnCl_2$, which is not effected by an increase in ionic strength.

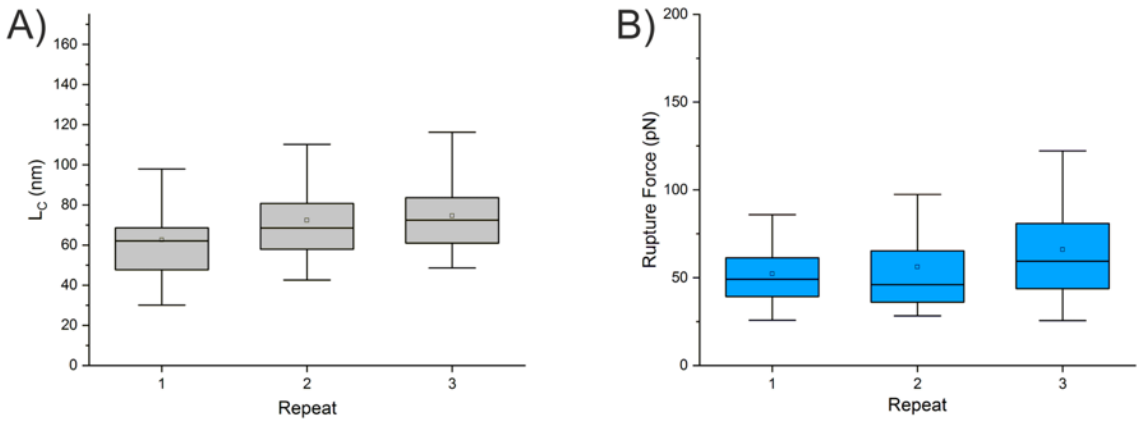

Figure 3.11 – Box plots for L_C and rupture force in 1X TBS, 0.5 M NaCl, 100 μM $ZnCl_2$, pH 7.5. A) median L_C values are 62.7, 68.4 and 73.7 nm for repeat 1, 2 and 3. B) median rupture force values are 49.1, 46.0 and 59.4 pN for repeat 1, 2 and 3. Whiskers represent minimum and maximum values. Horizontal lines and open squares represent median and mean values, respectively. Data shows mostly similar distributions are observed between triplicate experiments.

As it was clear from our data that Zn^{2+} ions promoted complex formation, it was necessary to investigate whether other divalent metal ions were capable of supporting self-association of SasG. As our buffer system contained Tris, a weak chelator of divalent metal ions, we utilised a metal ion with a similar population of metal ion:Tris present at pH 7-8[184] to obviate any concentration-driven differences. Co^{2+} was chosen as it has comparable percentage of Co:Tris with Zn:Tris complex in solution at pH 7-8. In the presence of 100 μM $CoCl_2$ there was an absence of a 'hot spot' and an insensitivity in hit rate to the addition of 2 mM EDTA. When the buffer was replaced with 100 μM $ZnCl_2$, the hit rate tripled and the presence of a 'hot spot' appeared signalling the homophilic zinc-mediated interactions had been re-established (Figure 3.12). The overall lower hit rate in comparison to earlier experiments could be due to technique reflection of the relatively low number of replicates used. However, it is clear that there is no increase in complex formation in 100 μM $CoCl_2$, whereas there is still a 3-fold increase of interactions in 100 μM $ZnCl_2$. This suggests that these SMFS interactions are specific to Zn^{2+} ions as observed beforehand by Formosa-Dague and colleagues by SCFS[153].

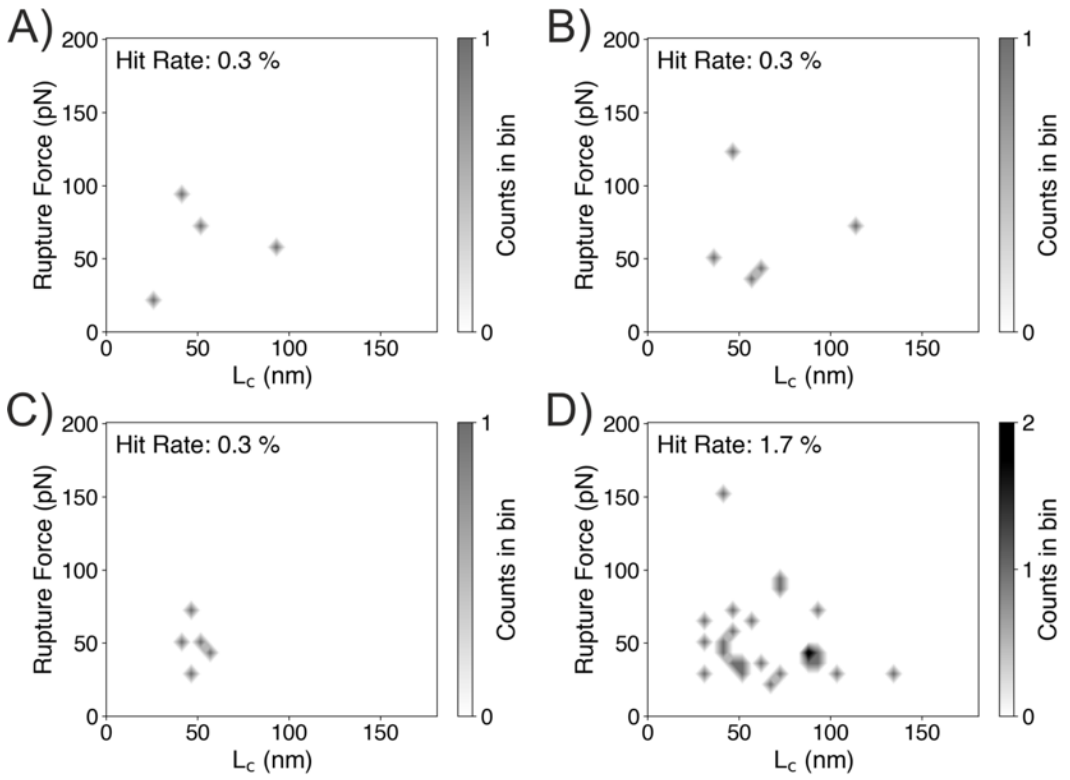

Figure 3.12 – Investigating the metal ion specificity of the SMFS SasG interaction using $CoCl_2$. A bin size of 5 nm and 7 pN for the L_C and rupture force, respectively, was utilised for the contour plots. Base buffer was 1X TBS + 0.5 M NaCl, pH 7.5 and was supplemented with A) 2 mM EDTA, B) 100 μM $CoCl_2$, C) 2 mM EDTA and D) 100 μM $ZnCl_2$. The experiment was carried out from A) to D), collecting 1500 force-extension cycles for each condition. There is no difference in hit rate between 2 mM EDTA and 100 μM $CoCl_2$, however interactions were restored upon addition of 100 μM $ZnCl_2$.

Our L_C data demonstrates that SasG is forming homocomplexes driven by a threshold of Zn^{2+} ions (11 - 100 μM). Interpretation of L_C values is complicated due to two competing effects: I) extensible linker location on the tip and surface yielding apparently distinct L_C values and II) tip geometry affecting complex formation probability (Figure 3.13). The convolution of these two competing effects produces a distribution of apparent L_C values, shorter than the true L_C (L_C^{True})[185]. Our data is further complicated by the length of the constructs (~60 (experimentally determined[74])-70 nm (iterative superimposition of crystal structures) + linker) and potentially multiple binding/coordination sites (Figure 3.13). In an attempt to determine the binding geometry of our SasG complex in the presence of 100 μM $ZnCl_2$ we can look at the maximum binned L_C values for the bimodal gaussian fits as a rough estimate for the L_C^{True} values and compare these to the naïve structure to resolve potential interaction sites. From the 100 μM $ZnCl_2$ L_C histogram (Figure 3.10C) the approximate L_C^{True} values are ~ 100 and ~ 125 nm, for the dominant and minor populations, respectively.

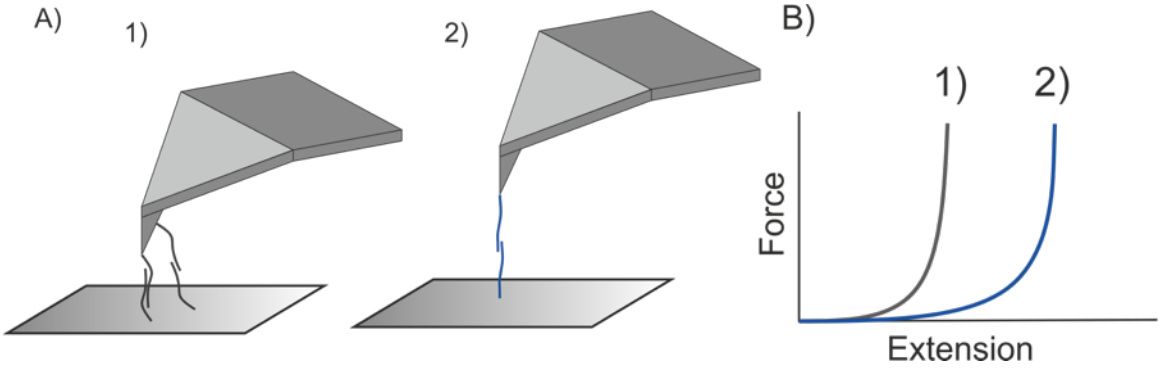

Figure 3.13 – Schematic demonstrating the factors influencing the observed L_C in SMFS dissociation experiments. The observed L_C is dependent on the linker location of both the tip and surface and the binding location of the molecules if there are multiple interaction sites. A) Possible immobilisation positions on the silicon nitride surfaces and tips. B). 1) Protein immobilised on the cantilever distal to the tip interacting with protein immobilised on the substrate at distance to the tip interacting at their termini and a protein attached at the apex of the tip and directly beneath the tip with a different complex scheme give the same L_C. 2) Protein immobilised at the apex of the tip and directly beneath the tip gives the true L_C for a complex scheme involving interaction at the termini of the proteins.

In an attempt to resolve the potential residues involved in coordinating Zn^{2+}, initial cosolvent simulations on E-G5[2] (PDB: 3TIP) in the presence of 114 Zn^{2+} ions were carried out by Alexander St John (PhD student, University of Leeds) as described in Section 2.2.1.1. These simulations revealed multiple negatively charged residues potentially involved in zinc-coordinated interactions: 518E, 520E, 557E, 559E, 583E, 613E and 617D (Figure 3.14).

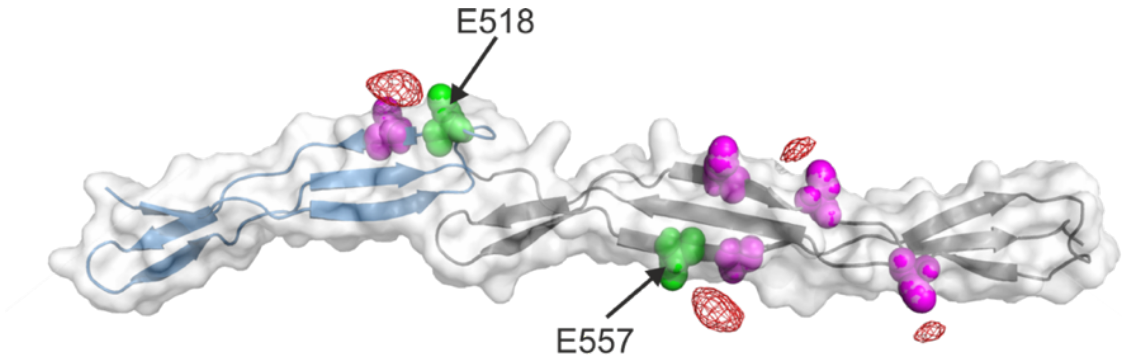

Figure 3.14 – Free energy map of zinc cosolvent simulations. Cosolvent simulations (50 x 25 ns) on E-G5[2] (PDB: 3TIP) in the presence of 114 Zn^{2+} ions. Density of Zn^{2+} ions calculated in 1x1x1 $Å^3$ voxels averaged over the simulation and resulting free energy contoured at -2 kcal mol^{-1} (red mesh). Residues potentially coordinating Zn^{2+} are highlighted in magenta and those which were subject to alanine substitution to produce SasG-ΔEE labelled in green. Simulation carried out by Alexander St John (PhD student, University of Leeds).

Using this information two SasG structures were positioned relative to one another in PyMOL to determine what groups of residues would likely align to allow coordination of Zn^{2+} ions to match the approximate L_C^{True} values (Figure 3.15). Groups of the residues predicted to be involved in Zn^{2+}-coordinated interactions were found proximal to one another in line with our L_C^{True} values (structure set at 70 nm in length). However, there are caveats to this method, which include, but are not limited to, the model is static, we do not know how they interact (coiled around or parallel to one another), if they are fully extended and how loading force may change the complex conformation.

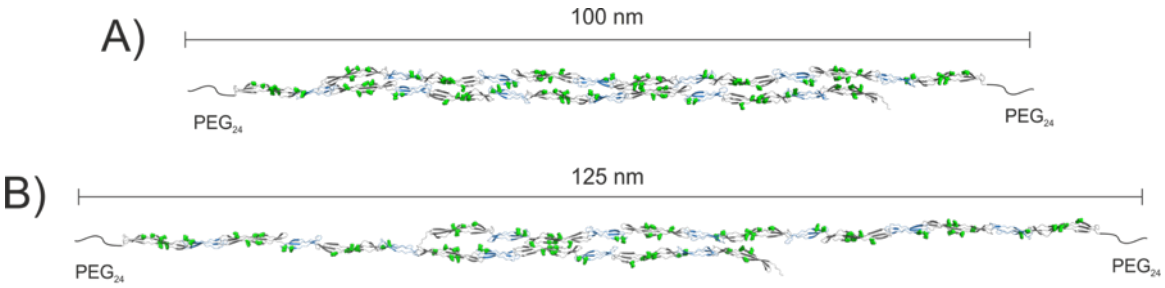

Figure 3.15 – Static modelling of SasG monomers to sample potential geometries of the Zn^{2+}-induced interaction. Residues predicted to coordinate Zn^{2+} ions highlighted as green ball-and-stick. At potential L_C^{True} values of 100 and 125 nm, multiple residues predicted to coordinate Zn^{2+} ions are available and proximal to one another. Made using PyMOL software[186].

In light of the simulations and PyMOL alignment, a variant was designed to remove selected negatively charged side chains predicted to be coordinating Zn^{2+} ions in an attempt to significantly dampen/knock out the Zn^{2+}-induced binding capacity of SasG. This variant, designated SasG-ΔEE, comprised mutations E518A and E557A (of E-G5[2] – Figure 3.14) and the corresponding glutamic acid residues in the remaining 5 E and 6 G5 sub domains.

3.2.4.4 Creation of the H_6-TEV-SasG-ΔEE Construct

The H_6-TEV-SasG-ΔEE gene (Section 2.1.17.3) was constructed via GG assembly (Section 2.2.2.4). Seven contiguous cassettes that together encoded H_6-TEV-SasG-ΔEE were designed for direct use in GG assembly and purchased as synthetic cassettes (Section 2.2.2.4). The pET14b destination vector utilised is discussed in depth in Section 4.2.1.4. After resuspension of the synthetic cassettes, GG assembly was carried out and the product was transformed into SURE 2 *E. coli* cells as described in Section 2.2.2.4 and 2.2.2.5, respectively. Colonies were subjected to One*Taq*® colony PCR (Section 2.2.2.6), using the screening primers outlined in Section 2.2.2.8, to determine which colonies contained an assembly of the correct size. Figure 3.16 displays the linearised vector, cassettes and assembled DNA. Sequences of the correct size were confirmed by DNA sequencing and the full sequence can be found in Appendix (Section 7.1.2).

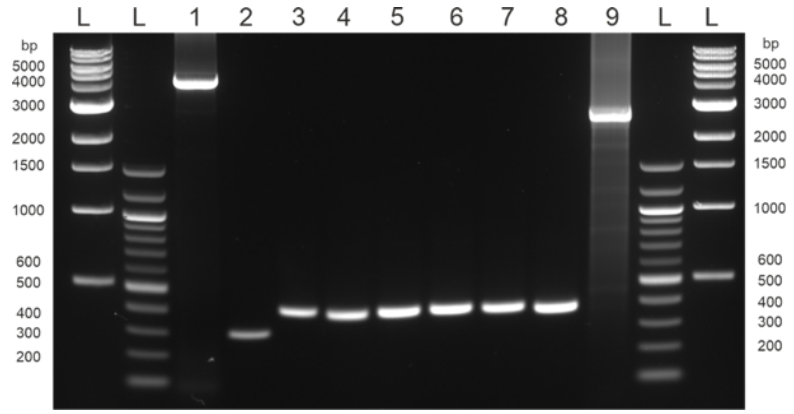

Figure 3.16 – Agarose gel (1.5 % w/v) stained with SYBR-SAFE® showing the assembly of SasG-ΔEE in pET14b. Lane 1 shows the Q5® PCR amplified and linearised pET14b with flanking BsaI sites. (4632 bp). Lanes 2-8 shows the synthetic DNA cassettes 1-7 with flanking BsaI sites (315, 420, 411, 418, 422, 417 and 410 bp). Lane 9 shows the One*Taq*® colony PCR amplification of the assembled construct (2653 bp) and flanking bases (total: 2790 bp). L: DNA ladder.

3.2.4.5 Gene Expression, Production and Purification of SasG-ΔEE

H_6-TEV-SasG-ΔEE under the control of the T7 promoter in pET14b was transformed into *E. coli* BL21 (DE3) pLysS cells. Gene expression and protein production was carried out as outlined in Section 2.2.3.1.2. Although H_6-TEV-SasG-ΔEE is not a fusion as is H_6-MBP-TEV-SasG, the purification process was unchanged, as the H_6-TEV-SasG-ΔEE was required to undergo TEV cleavage to remove the H_6 tag from the construct (leaving SasG-ΔEE after TEV cleavage). After purification, the purity was assessed by SDS-PAGE (Figure 3.17) and LC-MS, which confirmed the successful isolation of SasG-ΔEE. The mass spectra showed a mass of 93186.1 ± 1.3 Da as expected (93185.1 Da) for SasG-ΔEE. The protein sequence can be found in the Appendix (Section 7.2.2). As with SasG the C-terminal contained two cysteine residues for immobilisation in SMFS experiments.

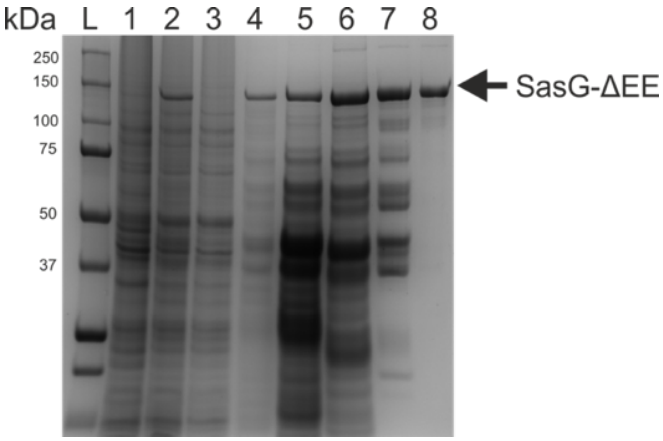

Figure 3.17 – SDS-PAGE gel displaying the gene expression, protein production (of H_6-TEV-SasG-ΔEE) and subsequent purification steps of SasG-ΔEE. Lane 1 is protein produced at 0 hours, lane 2 corresponds to protein produced at 24 hours, lane 3 is the pellet (insoluble) fraction and lane 4 is the cytosolic fraction, indicating a soluble protein. Lane 5 is the initial IMAC purification step, lane 6 is the flowthrough from the second IMAC purification step post-TEV treatment, lane 7 is the subsequent AEX purification step and lane 8 is the final SEC purification step, where pure SasG-ΔEE was isolated. As with SasG, SasG-ΔEE displays an anomalous electrophoretic mobility, however, SasG-ΔEE runs at a slightly lower molecular weight of ~ 140 vs 150 kDa of SasG. This is likely due to the loss of 13 negatively charged side chains in SasG-ΔEE and consequently more negatively charged SDS molecules are able to bind to the structure[174,175].

3.2.4.6 Spectroscopic Analysis of SasG-ΔEE

To determine whether the glutamic acid to alanine substitutions have perturbed the structure of SasG-ΔEE, spectroscopic analyses were carried out.

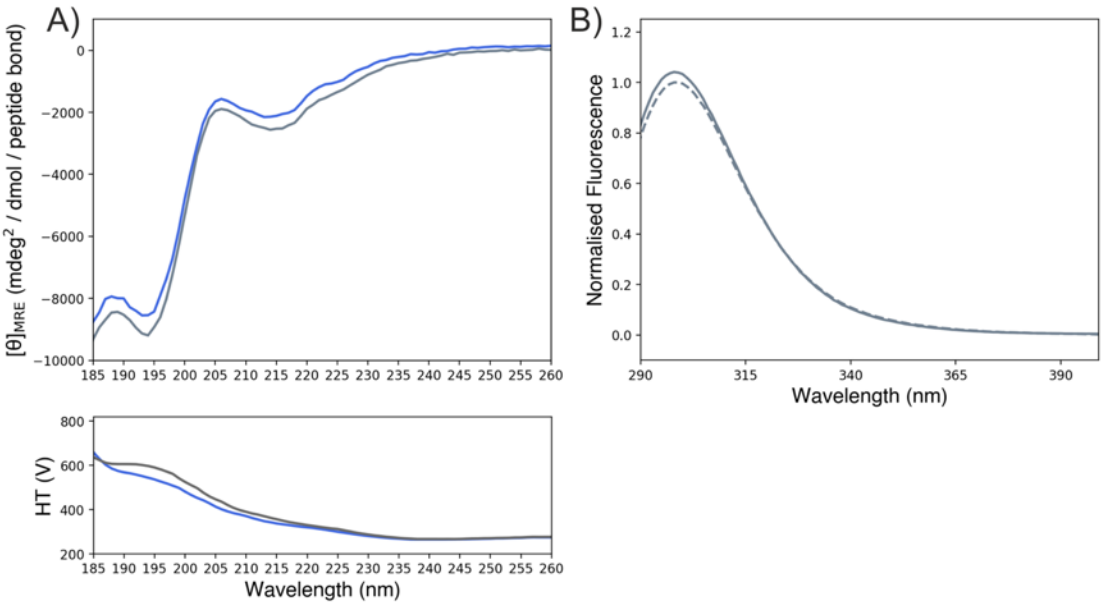

Figure 3.18 – SasG-ΔEE spectroscopic analysis. A) Far-UV CD spectra of SasG-ΔEE (grey) and SasG (blue) with HT voltage below. B) Fluorescence emission scans of SasG-ΔEE in the absence (solid) and presence (dashed) of 4 M urea to assess the tertiary structure.

Upon far-UV secondary structure analysis, the spectrum of SasG-ΔEE displayed analogous minima to the SasG spectrum (Figure 3.18A). This allows us to determine that any changes in Zn^{2+}-induced homophilic binding capacity is due to the substituted residues and not structural alterations. A small, but evident decrease and red shift in fluorescence is observed upon the addition of chemical denaturant, indicating the tyrosines are no longer packed in the pseudohydrophobic cores (Figure 3.18B). These spectroscopic analyses suggest that the secondary structure of SasG-ΔEE is indifferent from SasG and is folded post-purification.

3.2.4.7 Investigating the Zn^{2+}-induced Interaction Potential of SasG-ΔEE

To investigate whether removal of these glutamic acids dampened or abolished Zn^{2+}-induced interactions, an SMFS experiment was designed involving both SasG-ΔEE and SasG. SasG-ΔEE and SasG were individually immobilised on a set of cantilever tips and silicon nitride surfaces. SasG-ΔEE on the tip and surface (-/-) was investigated in the presence of 100 μM $ZnCl_2$. The surface was then swapped with a surface functionalised with SasG (-/+), followed by swapping the tip with a tip functionalised with SasG (+/+) keeping the buffer constant at 100 μM $ZnCl_2$. Finally, the buffer was exchanged to 2 mM EDTA. This was carried out in duplicate (due to SasG-ΔEE yield constraints) with a total of 3500 force-extension cycles for each condition. The resulting contour plots, histograms (both with L_C and rupture force in 5 nm and 7 pN bins, respectively) and schematics are displayed in Figure 3.19. Interactions were observed in all 100 μM $ZnCl_2$ conditions, with a significant decrease observed in 2 mM EDTA. This indicates SasG-ΔEE is still able to partake in Zn^{2+}-induced complex formation.

A hit rate hierarchy is observable: 2 mM EDTA +/+ (0.8 ± 0.4 %) < 100 μM $ZnCl_2$ -/- (1.5 ± 0.1 %) < 100 μM $ZnCl_2$ -/+ (1.8 ± 0.1 %) < 100 μM $ZnCl_2$ +/+ (2.1 ± 0.3 %). This dynamic hit rate with the different constructs may suggest that knocking these glutamic acids out may have reduced the number of residues available for coordinating Zn^{2+}, and in turn there is a loss of coordination schemes. Furthermore, there is no drastic change in distributions between construct conditions, indicating a large variety of binding regimes are still available to SasG-ΔEE. This data indicates SasG-ΔEE is still able to coordinate Zn^{2+} and in turn suggests there is pleomorphism of residues able to coordinate Zn^{2+} ions and distinct coordination schemes.

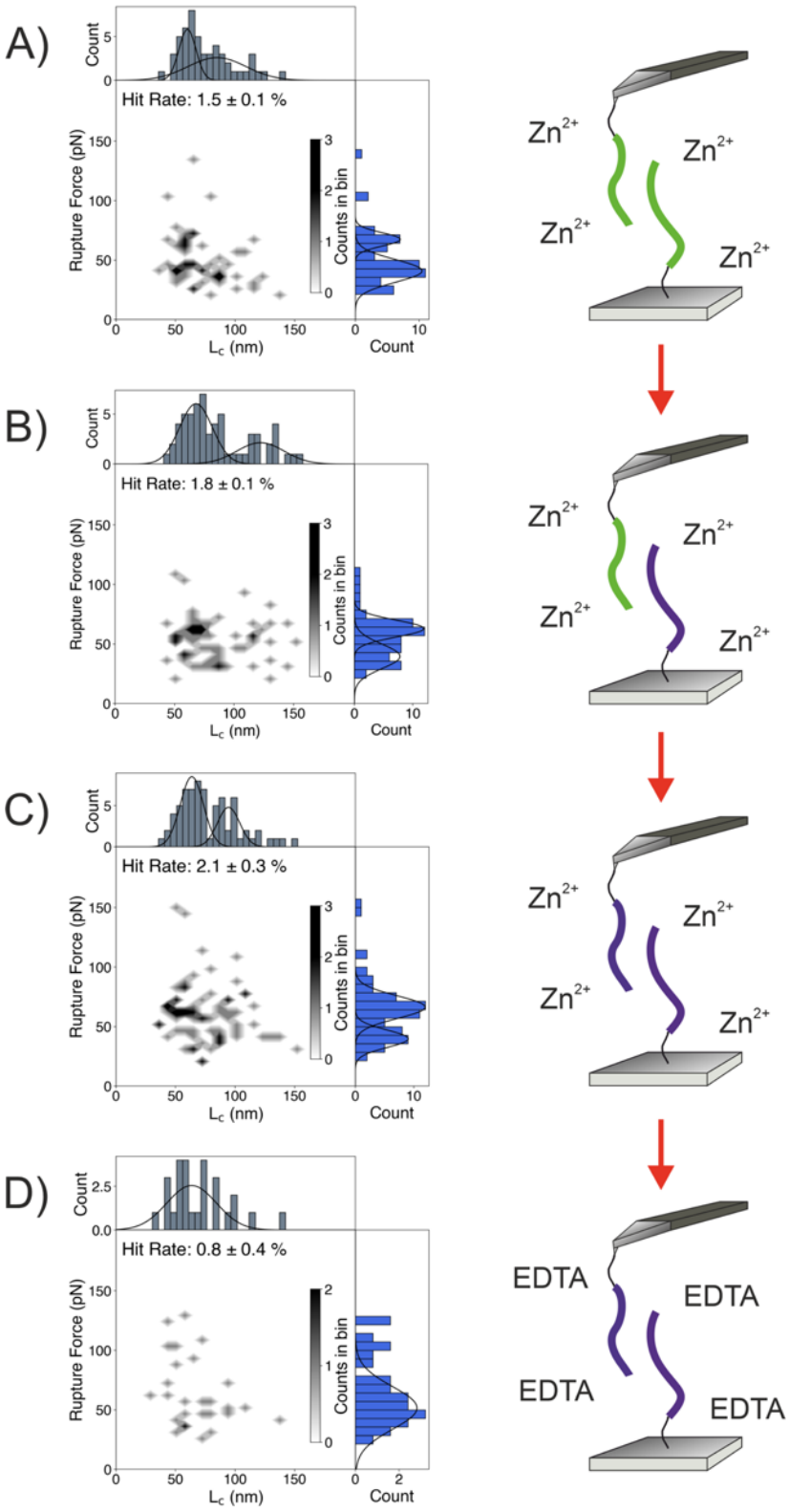

Figure 3.19 – SasG-ΔEE and SasG interactions in the presence and absence of Zn^{2+}. A) SasG-ΔEE on the tip and surface (-/-) in the presence of 100 μM $ZnCl_2$. B) SasG-ΔEE on the tip and SasG on the surface (-/+) in the presence of 100 μM $ZnCl_2$. C) SasG on the tip and surface (+/+) in the presence of 100 μM $ZnCl_2$. D) SasG on the tip and surface (+/+) in the presence of 2 mM EDTA. Schematics of SasG-ΔEE and SasG in green and purple, respectively. Red arrows denote the experimental trajectory. Errors are the SD (sample) of the repeats.

3.2.4.8 Mechanically Unfolding SasG-ΔEE

SasG-ΔEE was mechanically unfolded in 1X PBS, pH 7.4 at five retraction velocities (200, 800, 1500, 3000 and 5000 nms^{-1}) in triplicate (Figure 3.20). The ΔL_C were determined to be 149.4 (ranging from 145.5-153.9) and 215.9 (ranging from 214.2-220.4) Å for the SasG-ΔEE E and G5 sub domains, respectively. This is in good agreement with our previous experimental ΔL_C values observed for our SasG sub domains (Section 3.2.3) and those observed by Gruszka and colleagues[74]. Although the SasG-ΔEE G5 sub domain mechanical properties remained unchanged from SasG, the E sub domain unfolded at a significantly (ANCOVA $p < 0.05$) higher force. Given that hydrophobic contacts at the E-G5[2] interface are potentially involved in the mechanical strength of the E sub domain (see Section 5.3.4.3.1), the inclusion of a non-native hydrophobic residue (Ala) proximal to the E-G5 interfaces in SasG-ΔEE (Figure 3.14) coupled with an increase in mechanical strength may not be surprising. Although an interesting observation, the difference in unfolding force is minor (~ 14.5 pN) and any potential new hydrophobic contacts at the E-G5 interfaces have not altered the secondary structure (Section 3.2.4.6). It is thus unlikely that a change in the force response of the E sub domain is directly responsible for the changes in the Zn^{2+}-induced homophilic binding capacity. All unfolding statistics are found in Section 7.3.1.

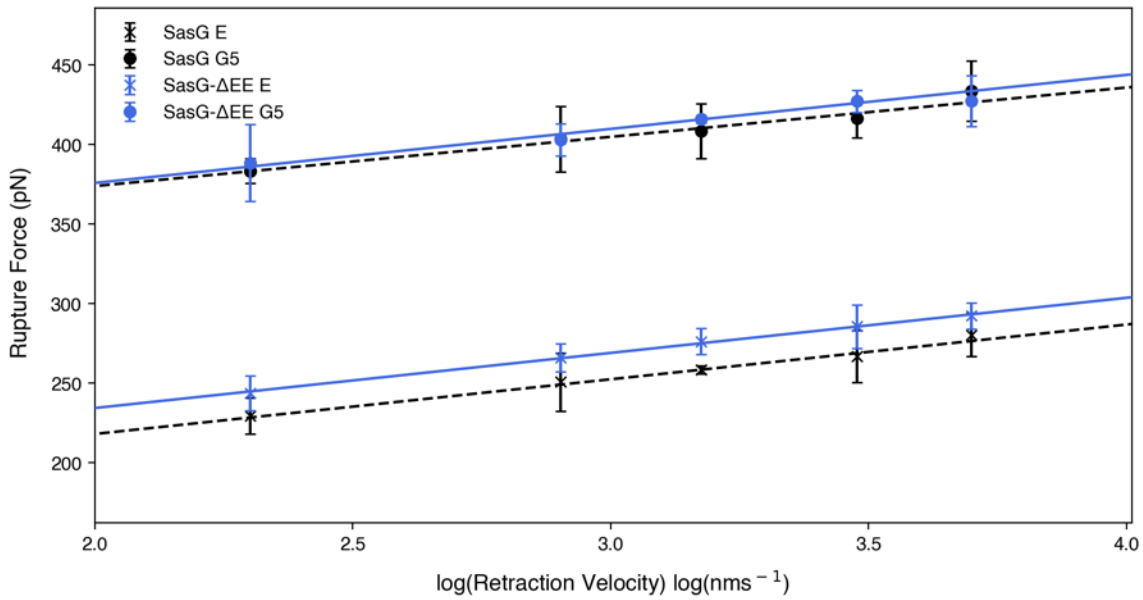

Figure 3.20 – The dependence of rupture force on the logarithm of retraction velocity of SasG-ΔEE sub domains, E and G5. E and G5 data points displayed as crosses and filled circles, respectively. Blue and black points and errors are the mean and standard deviation of triplicate datasets from this study. Linear fit is weighted with the inverse of the standard deviation error.

3.2.4.9 Mechanical Unfolding of SasG in the Presence of Zinc

Metal ions have been observed to stabilise mechanical folds through electrostatically protecting and locking hydrogen bonds in shear geometry[76]. As our data implies Zn^{2+} ions are important for homophilic SasG bonding, we wanted to determine if the effect of Zn^{2+} is twofold: both promoting homophilic binding and capable of altering the mechanical properties of SasG.

The protein was mechanically unfolded in 1X TBS (20 mM Tris, 150 mM NaCl), pH 7.5 supplemented with 100 μM $ZnCl_2$ at five retraction velocities (200, 800, 1500, 3000 and 5000 nms^{-1}) in duplicate. We were unable to use 1X PBS as Zn^{2+} forms complexes with phosphate to form insoluble zinc phosphate precipitates. This experiment was complicated by the regular unfolding of > 2 molecules producing sawtooth profiles of > 13 unfolding peaks (Figure 3.21) suggesting the presence of 100 μM $ZnCl_2$ was promoting association of SasG molecules with one another on the surface (*cis* interactions) prior to interacting with the cantilever tip. Only the FX profiles with ≤ 13 peaks (and a detachment peak) were used for the mechanical unfolding analysis. This is evidence for *in vitro* Zn^{2+}-induced interactions between SasG molecules in the *cis* orientation (as discussed in Section 1.4.3.1 and 3.2.4.3).

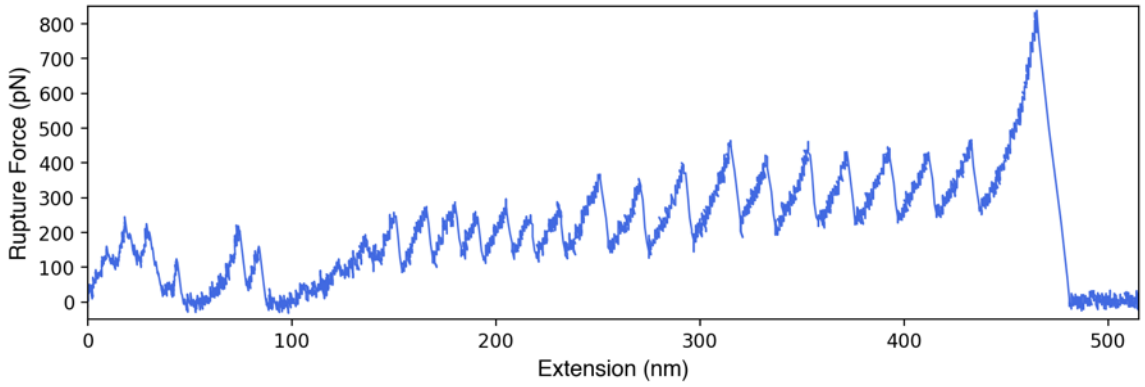

Figure 3.21 – An example SasG sawtooth profile with > 13 unfolding peaks in the presence of 100 μM $ZnCl_2$ when mechanically unfolded at a retraction velocity of 1500 nms^{-1}. These sawtooth profiles suggest Zn^{2+}-induced homophilic interactions between SasG monomers in a *cis* orientation prior to forced protein unfolding.

The mechanical properties of E and G5 sub domains were indifferent to those in the absence of Zn^{2+} (Figure 3.22). The ΔL_C were determined to be 149.9 (ranging from 145.6-156.8) and 216.0 (ranging from 214.9-218.3) Å at 1500 nms^{-1} for the SasG E and G5 sub domains, respectively. This

is in good agreement with the experimental ΔL_C values observed for our SasG sub domains (Section 3.2.3) and as observed by Gruszka and colleagues[74]. All unfolding statistics are found in the Appendix (Section 7.3.2)

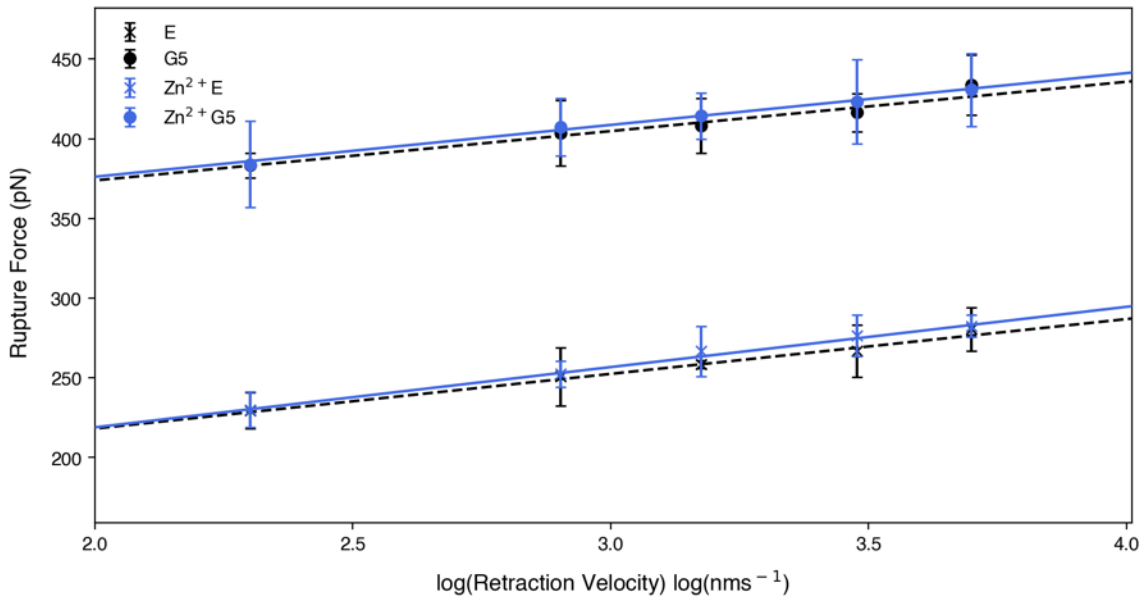

Figure 3.22 – The dependence of rupture force on the logarithm of retraction velocity of SasG sub domains, E and G5 in 100 μM $ZnCl_2$. E and G5 data points displayed as crosses and filled circles, respectively. Black and blue data points are SasG in 1X PBS, pH 7.4 and SasG in 1X TBS, 100 μM $ZnCl_2$, pH 7.5, respectively. Blue points and errors are the mean and standard deviation of triplicate datasets from this study. Linear fit is weighted with the inverse of the standard deviation error.

3.3 Discussion

3.3.1 What is a Physiologically 'Relevant' Zn^{2+} Concentration?

Host cells have intracellular Zn^{2+} concentrations of 100-500 μM, compartmentalised within the cytosol (50 %), nucleus (30-40 %), membranes and with around 20 % complexed by metallothioneins[187–189]. The remaining Zn^{2+} found in the blood serum is at around ~ 10-19 μM[21,181]. During disease pathogenesis, such as biofilm formation, *S. aureus* provokes apoptosis, pyroptosis and necroptosis in a broad spectrum of target cells in a means to establish infection and antagonise (or evade) host immune defences[190]. It would not be surprising, therefore, if there is a large increase in local Zn^{2+} concentration observed at the site of the growing biofilm. This could promote *S. aureus* to transition to the accumulation phase, the phase in which SasG has been implicated in[20].

Consequently, we believe a Zn^{2+} concentration of 100 μM is physiologically relevant and may represent the local concentrations, albeit perhaps acute, observed at the site of infection/biofilm formation.

3.3.2 Zinc-mediated Homophilic Interactions of SasG Resolved by SMFS

SasG was initially implicated in binding to desquamated nasal epithelial cells through the A domain[145]. In the late 2000s, Corrigan and colleagues determined SasG to be implicated in biofilm formation in *ica*-positive and -negative strains, with strains not expressing SasG failing to form biofilms[144]. The necessity of Zn^{2+} for biofilm formation was then observed for MRSA *S. aureus* strain USA300, with the addition of a Zn^{2+}-specific metal chelators (DTPA) inhibiting the formation[21]. Geoghegan and colleagues determined that Zn^{2+}-dependent biofilm formation occurs via the B domain of SasG as biofilm formation by SH1000 (pALC2073*SasG*$^{+}$ A^{-}B^{+}) was reduced by the zinc-chelator DTPA[20]. However, thus far, there has been little experimental evidence that has shown SasG dimerisation in the presence or absence of Zn^{2+} is due to the SasG B domain *in vitro*. Former *in vitro* studies[20,149] have utilised up to 2 B-repeats *in vitro*, which is far below the minimum 5 B-repeats observed to initiate biofilm formation[144]. Here we have utilised SMFS in an attempt to understand the Zn^{2+}-mediated interaction (or lack thereof) between SasG molecules with 6.5 B-repeats *in vitro*.

Our SMFS results are in agreement with Formosa-Dague and colleagues who observed interactions between $SasG_8^{(+)}$ cells in the presence of 1 mM Zn^{2+} using SCFS[153]. $SasG_8^{(-)}$ cells did not show adhesion events with $SasG_8^{(+)}$ cells, implying the events observed were probably not due to the A domain and were potentially mediated by the B domain. They observed a mixture of events, both single and multiple rupture forces (~ 95 % of analysed traces) with an average rupture force and rupture length of 414 ± 222 pN and 182 ± 76 nm, respectively. In contrast, our interactions had an average rupture force of ~ 43-75 pN (with some close to 200 pN) and rupture length of ~ 67-110 nm (L_C). This difference in rupture force is likely to be related to the SCFS technique, where the protein concentration on the cell wall is unregulated and in turn interactions between multiple pairs of SasG are probed simultaneously or complexes of ≥ 3 molecules are rupturing resulting in

an accumulated rupture force. Whereas, in our study the concentration is kept low (1 mg/ml) to prevent this from occurring[183].

Our SEC-MALLS data demonstrates that SasG is mainly monomeric at neutral pH in the presence and absence of Zn^{2+} in solution and is consistent with Gruszka and colleagues investigations on up to two B-repeats[149]. However, our SMFS data suggests that monomeric SasG B domains are interacting in the presence of Zn^{2+} and implies that these Zn^{2+}-mediated interactions are probably transient in nature. Furthermore, interactions appeared insensitive to a sizeable increase in ionic strength of the solution implying these interactions are not electrostatic, but potentially coordinate bonds between residue side chains with a carboxylate side chain group (glutamic and aspartic acid residues) and the Zn^{2+} ions. Covalent bonds were originally found to rupture at ~1-2 nN under external load[130], however, when the dative covalent interaction between Ni^{2+} and His_2 (2 consecutive histidines) was interrogated using SMFS, complex dissociation displayed much smaller rupture forces (100-500 pN)[191]. Both the rupture forces in this study and Formosa-Dague and colleagues published data display forces on a similar scale.

Recently, Formosa-Dague and colleagues showed an absence of interactions when Ca^{2+} was utilised instead of Zn^{2+}[153]. Furthermore, only Zn^{2+} was observed to induce *S. epidermidis* biofilm formation[21] and when the self-association of Aap was studied in solution, only Zn^{2+} and Cu^{2+} (out of several transition metals tested, including Co^{2+}) had the ability to drive self-assembly[192]. Unsurprisingly, our SMFS results corroborate these findings by showing an absence of interactions in 100 μM Co^{2+} that were restored upon the addition of 100 μM Zn^{2+}. The coordination number* (CN) of Zn^{2+} is almost exclusively found with a CN of four (tetrahedral coordination) in metal-protein complexes, however, Co^{2+} prefers a CN of six (octahedral geometry)[193]. This suggests the Zn^{2+}-selectivity observed in our *in vitro* SMFS studies may arise from the ability of SasG side chains to coordinate Zn^{2+} in a tetrahedral geometry.

* The preferred number of coordinating ligands (e.g. amino acid side chains) of a metal ion

The L_C data revealed a great deal of information about the interaction scheme of SasG in the presence of Zn^{2+} and from the contour plots, histograms and box plots it appears plausible that there is one dominant binding regime. However, it is also apparent that there are interactions with larger L_C values indicating that SasG may be able to bind in multiple binding schemes. Using L_C^{True} values obtained from the L_C-frequency histograms, static PyMOL modelling and Zn^{2+} cosolvent simulations, SasG-ΔEE was designed. When investigated by SMFS the variant was observed to retain its ability to participate in Zn^{2+}-induced interactions, albeit to a lesser extent. The hit rate increased when SasG was swapped in to be its binding partner and increased further when the experiment was swapped to SasG only. This data suggests pleomorphism in Zn^{2+} coordinating residues by SasG, which is observed in Aap[192] and further corroborates that the side chains of the predicted glutamic acid residues are contributing to the coordination of Zn^{2+}. Furthermore, SasG-ΔEE displayed an unchanged secondary structure and similar mechanical properties to SasG, suggesting the observed decrease in binding potential is not linked to major structural or mechanical changes.

The coordination of Ca^{2+} by SdrG B domains was shown to stabilise the fold by a factor of four up to 2 nN[76]. To determine whether the effect of Zn^{2+} ions was two-fold, both interaction driving and mechanically stabilising, we mechanically unfolded SasG in the presence of 100 μM Zn^{2+}. We observed no deviation from the absence of Zn^{2+} implying Zn^{2+} is not stabilising any mechanical clamps. However, we did witness a large proportion of FX profiles which comprised of ≥ 13 unfolding events, indicating the monomers may be forming *cis* interactions (as discussed in Section 1.4.3.1 and 3.2.4.3) at the surface and in turn getting 'picked up' together. The schematic in Figure 3.23 suggests a mechanism for this. The observation of these Zn^{2+}-mediated *cis* interactions may explain the relatively low hit rate (< 5 %) observed in this chapter, as a large proportion of monomers may be preoccupied with *cis* interactions and thus unavailable for *trans* interactions.

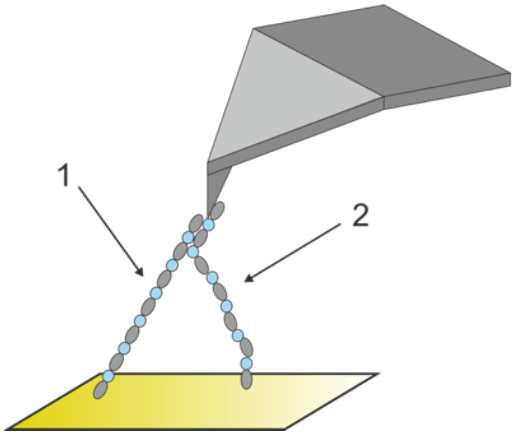

Figure 3.23 – Schematic for a proposed mechanism of observing > 13 unfolding peaks when investigating the mechanical strength of SasG in the presence of Zn^{2+}. Zn^{2+} mediates association between SasG N-termini distantly located from one another on the surface. If the complex associates with the tip, the SasG molecule which is furthest from the path of the cantilever (1) may begin unfolding until the other SasG molecule (2) in the complex elongates and begins unfolding when the cantilever is further away.

The data presented in this chapter provides further evidence that homophilic binding of SasG may occur, but why is Zn^{2+} required? SasG has many features of negatively designed β-sheet proteins, such as continuous H-bonding, β-bulges, very twisted and a switch (collagen-like region) between the sheets of each domain keeping the exposed β-strand edges short[194]. These qualities would imply they have evolved to avoid aggregation. It would thus be tempting to speculate that the requirement of Zn^{2+} ions may be part of the negative design to avoid unwanted edge-to-edge aggregation in the absence of Zn^{2+} ions and as a sensor molecule to drive the bacterial accumulation phase of *S. aureus*. These results provide some further insight to the potential importance and role of Zn^{2+} in biofilm formation. Further addressing this knowledge gap will provide opportunities for novel therapeutic interventions.

4. Development of a SasG Homo-polyprotein for Variant Characterisation by SMFS

4.1 Objectives

One of the main objectives of this thesis was to create a pseudo-WT system to facilitate the construction and mechanical analysis of SasG variants by SMFS. The pseudo-WT (pWT (E-G5^2)$_5$) assembled in this chapter was utilised as a structural, thermodynamical and mechanical foundation for subsequent variant creation to resolve the molecular determinants of the remarkable mechanical strength of SasG (Chapter 5). This chapter will begin by introducing the rationale for choice of a concatenated pentameric E-G5^2 domain construct, followed by protein engineering, expression and purification. Thorough structural and mechanical characterisation of the resulting pWT (E-G5^2)$_5$ homo-polyprotein is outlined.

4.2 Results

4.2.1 Designing, Cloning and Purification of pWT (E-G5^2)$_5$

4.2.1.1 Rationale of a Pentameric Homo-polyprotein of E-G5^2 for Use in SMFS

SasG is a mechanically strong multidomain protein consisting of an A domain and a tandem array of B repeats, composed of E and G5 sub domains (Figure 4.1)[74,149]. The mechanical strength of SasG results from the mechanostability of these E and G5 sub domains, which is predicted to arise from tandemly arrayed stretches of hydrogen bonds[74]. However, the fold of these sub domains is distinct from any mechanically strong protein observed before with regions of interest such as the collagen-like motif and compact interfaces which drives cooperative folding over long distances[152]. This chapter will focus on creating a protein system for use in SMFS to unravel the molecular elements of this remarkable mechanical strength.

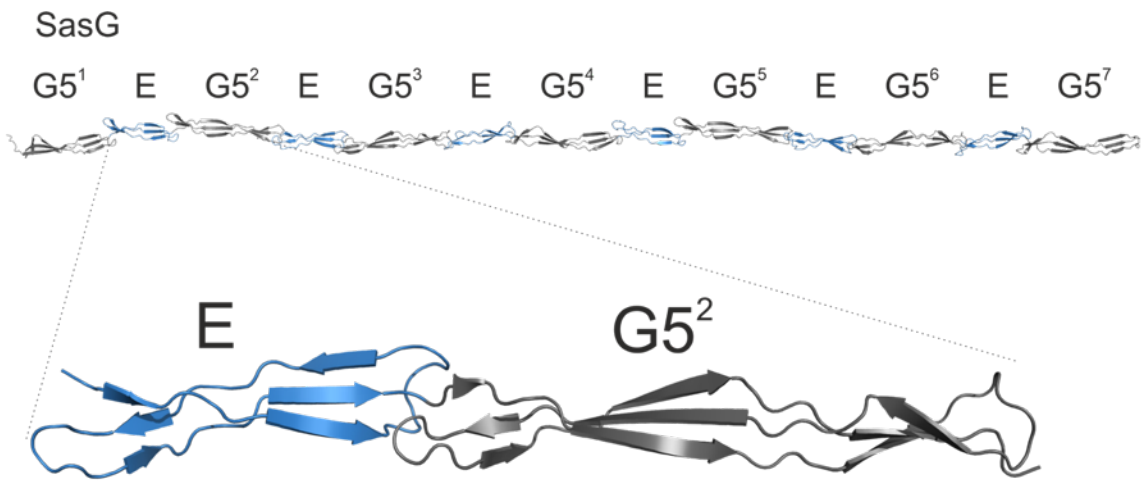

Figure 4.1 – Structural representations of SasG and E-G5^2 B-repeat. SasG comprising one G5^1 sub domain and 6.5 E-G5 B-repeats (SasG structure taken from reference[74]). E-G5^2 (PDB: 3TIP) enlarged.

The method of mutagenesis coupled with SMFS is frequently utilised to determine force-bearing structures and supporting non-covalent/covalent contributions to the mechanostability of proteins[90,96,97,100]. The high levels of DNA sequence similarity between the E-G5 repeats of both the

SasG sequence from *S.aureus* 8325-4 genomic DNA (~ 97 %) and the *E. coli* codon optimised DNA (Section 2.1.17.1) (~ 80 %) precludes direct mutagenesis of the SasG sequence.

To circumvent this problem a decision was made to generate a polyprotein construct (Section 1.3.1.2). These are commonly utilised for SMFS protein unfolding experiments[31] as they provide a clear mechanical fingerprint which can unambiguously distinguish between non-specific adhesion events of the cantilever and surface (problematic when unfolding monomers[195]). Furthermore, due to the many unfolding events per pull there is higher statistical power than using monomeric constructs.

Separate polyproteins of E and G5 or a polyprotein of G5-E were not selected because the E sub domain is disordered when in isolation or following a G5 sub domain (when there is no C-terminal G5 sub domain present)[74,152]. G5-E-G5 was not chosen as this would result in a construct comprising twice as many G5 sub domains than E and the order of sub domains would be dissimilar to SasG (G5-E-G5-G5-E-G5 vs G5-E-G5-E-G5). Therefore, E-G5 was nominated as the E sub domain is folded in the presence of a folded C-terminal G5 sub domain and concatenation would result in a sub domain order (E-G5-E-G5-E-G5) comparable to SasG[74,152]. More specifically, E-G5^2 (first E and second G5 sub domain from the SasG construct – Figure 4.1) was selected as the crystal structure (PDB: 3TIP) is published and we could utilise recent kinetic and thermodynamic analyses carried out by Gruszka and colleagues to guide our variant selection in Chapter 5[152].

Linkers were required to space the E-G5^2 repeats to prevent any faux interdomain interactions, which could affect the thermodynamic stability[196] and potentially the mechanical properties of the resulting construct. To avoid these, unstructured linker regions utilised in our laboratory with recognised desired properties were exploited[163]. These comprised of different palindromic-like seven residue combinations of non-polar (A, I, L and V), uncharged polar (S and T) surrounding a central G. These combinations delivered (i) similar physicochemical properties to one another and (ii) steric freedom to prevent interactions between neighbouring domains. At the DNA level the linkers are distinctly different from one another, which is achieved by switching between the two most frequently utilised amino acid codons in *E. coli*. This generates specific cassette 5' and 3'

termini for simultaneous multi cassette assembly[163]. The chosen peptide linkers and their encoding DNA sequence are displayed in Table 4.1. A double cysteine is grafted onto the C-terminal for immobilisation through gold-thiol bonding prior to mechanical unfolding using SMFS.

Linker	Amino acid sequence	DNA sequence
1	*LSVGATI*	CTGAGCGTGGGCGCGACCATT
2	*TVIGLAS*	ACCGTTATTGGTCTGGCGAGC
3	*ALSGTIV*	GCGCTGAGCGGCACCATTGTG
4	*VITGSLA*	GTTATTACCGGTAGCCTGGCG

Table 4.1 – Peptide linkers utilised for pWT (E-G5^2)$_5$. Peptide and encoding DNA sequences from reference[163].

A structural schematic of our final design, designated **pWT (E-G5^2)$_5$**, is displayed in Figure 4.2. pWT (E-G5^2)$_5$ comprises 5 E and 5 G5^2 sub domains in contrast to 6 E and 7 G5 sub domains in SasG.

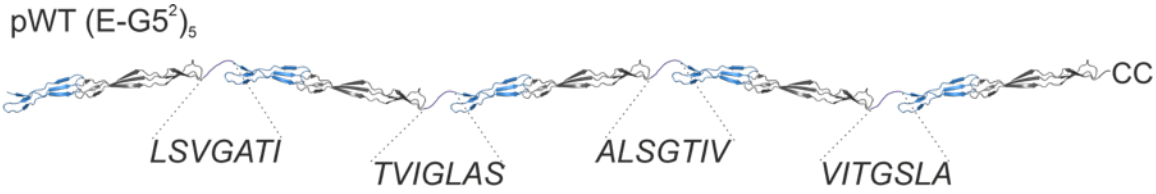

Figure 4.2 – Structural schematic of protein engineered pWT (E-G5^2)$_5$. The C-terminal double cysteine is displayed as CC and the linker peptide sequences are enlarged. E and G5^2 are coloured as blue and dark grey, respectively.

4.2.1.2 Choice of Assembly Method for Engineering pWT (E-G5^2)$_5$ DNA

In an effort to create a rapid system for producing mutant homo-polyproteins in an efficient manner, GA (Section 2.2.2.3) was originally opted for due to its speed and ease of single-step chemical coupling of monomers with the control over length and order of domains[163]. However, off-target assembly of monomer, dimers, trimers, tetramers or incorrectly ordered pentamers occurred frequently, with the correct DNA sequence assembled in 1-2 % of plasmids. After extensive troubleshooting, we switched to GAs modular assembly counterpart GG assembly[164,165]. GG is inherently more robust than GA because it does not rely on the activity of the 5' exonuclease, which has the potential to expose too many nucleotides and reduce the specificity of the overhangs

(especially with homo-polyproteins). Furthermore, GG utilises re-digestion mechanisms simultaneously eliminating the restriction site whilst cloning, this results in scarless assemblies and prevents re-ligation of the destination vector consequently reducing the number of colonies with no insert. In addition assembly efficiency can reach near 100 % with ≤ 12 fragments with high fidelity overhang sets[197]. After our switch to GG, we saw an efficiency reach 90-100 %.

4.2.1.3 Cloning of E-G5^2 for the Creation of a Monomeric Template

In this section the process for creating the pET14b-E-G5^2 construct ('working vector') for use as a template for cassette creation (for pWT (E-G5^2)$_5$ and mutagenesis) is outlined. Q5® PCR (as outlined in Section 2.2.2.1) was used to amplify E-G5^2 out of the SasG DNA (Section 2.1.17.1). Simultaneously, pET14b was linearised and amplified using Q5® PCR and primers outlined in Section 2.2.2.2. The PCR products were analysed on a 1.5 % (w/v) agarose gel stained with SYBR™ safe to confirm successful linearisation of the vector and amplification of the E-G5^2 cassette (Figure 4.3 lanes 1 and 2). The E-G5^2 amplicon was cloned into the linearised pET14b vector in the multiple cloning site (MCS) using the GA method (Section 2.2.2.3). 2 μl of the assembly product was transformed (Section 2.2.2.5) into 5-α cells. Colonies were subjected to One*Taq*® colony PCR (Section 2.2.2.6) to determine those which contained an insert of the correct size (Figure 4.3 Lane 3). The identity of assemblies of the correct size were confirmed by DNA Sanger sequencing (Section 2.2.2.8).

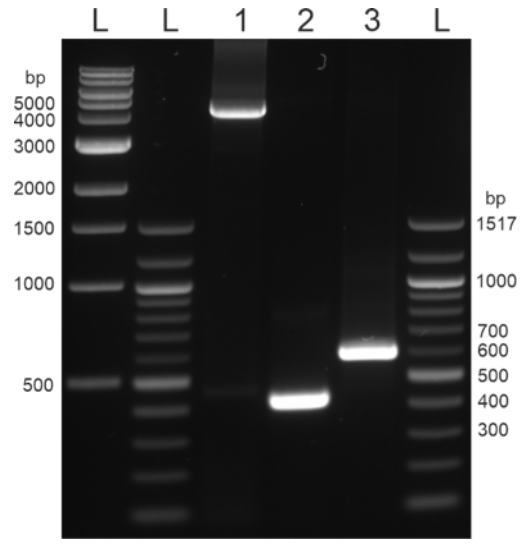

Figure 4.3 – Agarose gel (1.5 % w/v) stained with SYBR™ safe showing the stages of monomeric pET14b-E-G5^2 template creation. Lane 1 shows the Q5® PCR amplified and linearised pET14b (4671 bp). Lane 2 shows the Q5® PCR amplified E-G5^2 with Gibson Assembly® designed overhangs (433 bp). Lane 3 shows the product of One*Taq*® colony PCR of E-G5^2 (620 bp). pET14b colony PCR primers leads to product around 200 bp larger than insert due to distal annealing sites.

4.2.1.4 pET14b Vector Domestication

To create a destination vector for use in GG assembly using BsaI as the type-IIS restriction enzyme of choice, unwanted BsaI recognition sites must be absent, otherwise non-specific cleavage of the destination vector will occur, decreasing the probability of correct assembly. pET14b contains two BsaI recognition sites, one in the ampicillin resistance gene (AmpR) and the other proximal to the MCS. E-G5^2 DNA (from the SasG DNA supplied – Section 2.1.17.1) also contained a 3'-end BsaI recognition site. Q5® PCR was used to introduce silent mutations to remove all unwanted BsaI recognition sites in the vector and E-G5^2 DNA utilising primers outlined in Section 2.2.2.2. This resulted in the constructs pET14bΔ*bsaI* and pET14b-E-G5^{2}Δ*bsaI*. Figure 4.4 displays both the virgin and domesticated vector pET14bΔ*bsaI* treated with BsaI-HF®v2, demonstrating removal of BsaI sites.

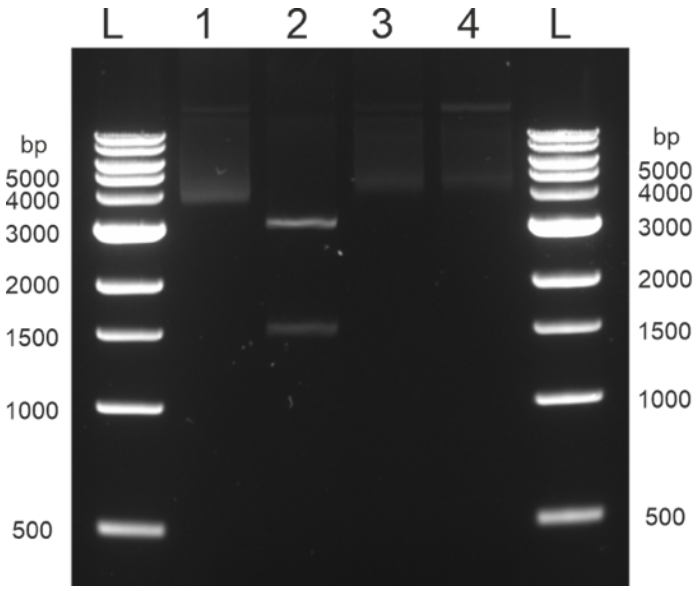

Figure 4.4 – Agarose gel (1.5 % w/v) stained with SYBR-SAFE® showing the presence and absence of BsaI recognition sites in virgin pET14b and pET14bΔ*bsaI*, respectively. Lanes 1 and 2 show the virgin pET14b negative control and the addition of BsaI-HF®v2 enzyme, respectively. Lanes 3 and 4 show the pET14bΔ*bsaI* negative control (Lane 3) and the addition of BsaI-HF®v2 enzyme (Lane 4). Negative control contains no BsaI-HF®v2. L: DNA ladder

4.2.1.5 Assembly of a Pentameric Homo-polyprotein (pWT (E-G5^2)$_5$) Utilising Golden Gate assembly

Q5® PCR was used to generate the destination vector and 5 cassettes of the monomeric E-G5^2 sequence with flanking BsaI recognition sites as described in Section 2.2.2.2. Cassette 5 reverse primer was designed to encode two C-termini cysteines for gold-thiol bonding utilised in SMFS unfolding experiments. After PCR purification of the cassettes and destination vector, assembly and transformation was performed as outlined in Section 2.2.2.5. Colonies were subjected to One*Taq*® colony PCR (Section 2.2.2.6), to determine which contained an assembly of the correct size. Figure 4.5 shows the linearised vector, amplified cassettes and assembled homo-polyprotein DNA.

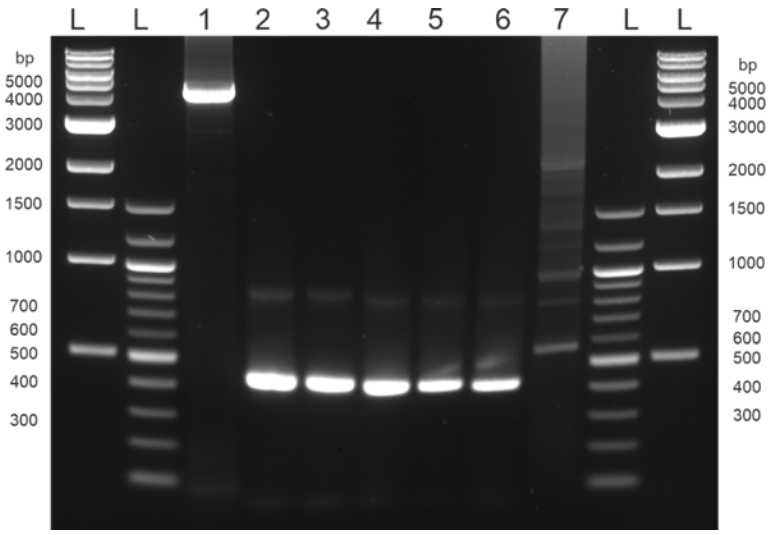

Figure 4.5 – Agarose gel (1.5 % w/v) stained with SYBR-SAFE® showing the assembly of the pWT $(E\text{-}G5^2)_5$ in pET14b. Lane 1 shows the Q5® PCR amplified and linearised pET14b with flanking BsaI sites. (4632 bp). Lanes 2-6 shows the amplified cassettes 1-5 with flanking BsaI sites (450, 451, 446, 449 and 442 bp). Lane 7 shows the One*Taq*® colony PCR amplification of the assembled polyprotein (2109 bp) and flanking bases (total: 2246 bp). Non-specific annealing of screening primers results in multiple bands present in Lane 7. L: DNA ladder.

Sequences of the correct size were confirmed by DNA Sanger sequencing and the DNA sequence is found in the Appendix (Section 7.1.3). The overview of our modular assembly process using GG is described in Figure 4.6.

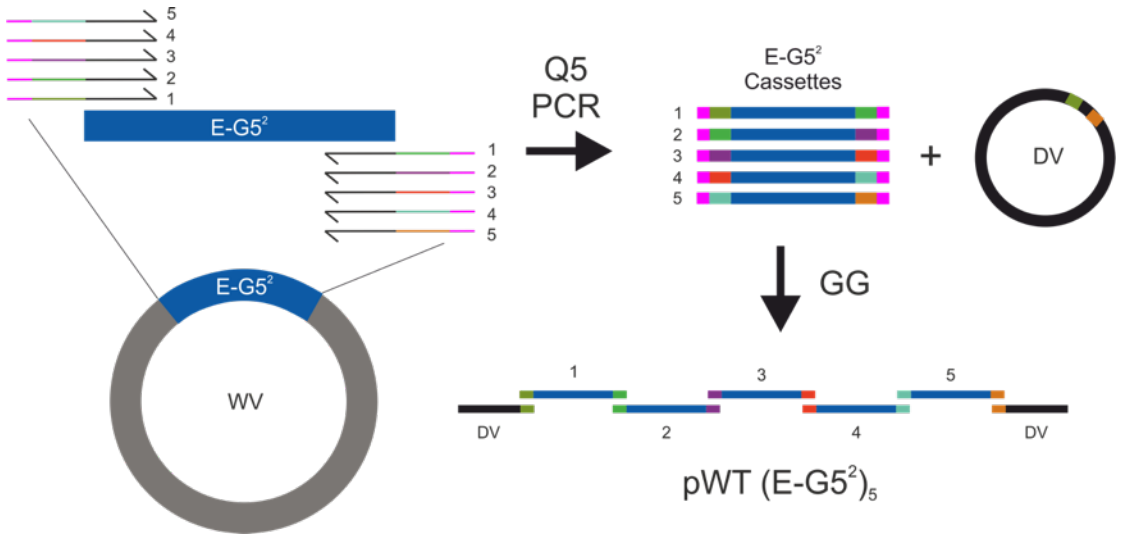

Figure 4.6 – Schematic of the homo-polyprotein modular DNA assembly using GG assembly. E-G5² is amplified out of the 'working vector' (WV) using five different sets of primers to create five E-G5² cassettes with distal BsaI sites (magenta). These sites are removed during cleavage to leave four base overhangs (colour coordinated) allowing simultaneous assembly into the destination vector (DV).

4.2.1.6 Gene Expression, Production and Purification of pWT (E-G5^2)$_5$

pWT (E-G5^2)$_5$ under the control of the T7 promoter in pET14b was transformed into *E. coli* BL21 (DE3) pLysS cells. Gene expression and protein production was carried out as outlined in Section 2.2.3.1.2. At each step of the production and purification a sample was taken for SDS-PAGE analysis (Figure 4.7). Evidence of protein presence in the cytosolic fraction indicated that the protein was soluble. Protein purification was carried out by IMAC using a lab-packed 200 ml precharged Ni Sepharose™ Fast Flow column (Cytiva) as described in Section 2.2.3.1.5. Further purification using 2 X 5 ml HiTrap SP HP columns (Cytiva) stacked atop of 4 X 5 ml HiTrap Q HP columns (Cytiva). The final purification step involved gel filtration using a 320 ml Hiload Superdex™ 26/600 75 column (Cytiva). The purity was assessed by SDS-PAGE (Figure 4.7D, Lane 7) and LC-MS, which confirmed the successful isolation of pWT (E-G5^2)$_5$. The mass spectrum showed a mass of 76095.7 $\pm$ 0.9 Da, around 131 Da smaller than originally predicted. However, this is due to N-terminal Met excision of cytosolic proteins catalysed by the *E. coli* methionyl-aminopeptidase, which is close to 100 % when Gly, as in our case, is the penultimate N-terminal amino acid[198]. The full protein sequence is found in the Appendix (Section 7.2.3.1).

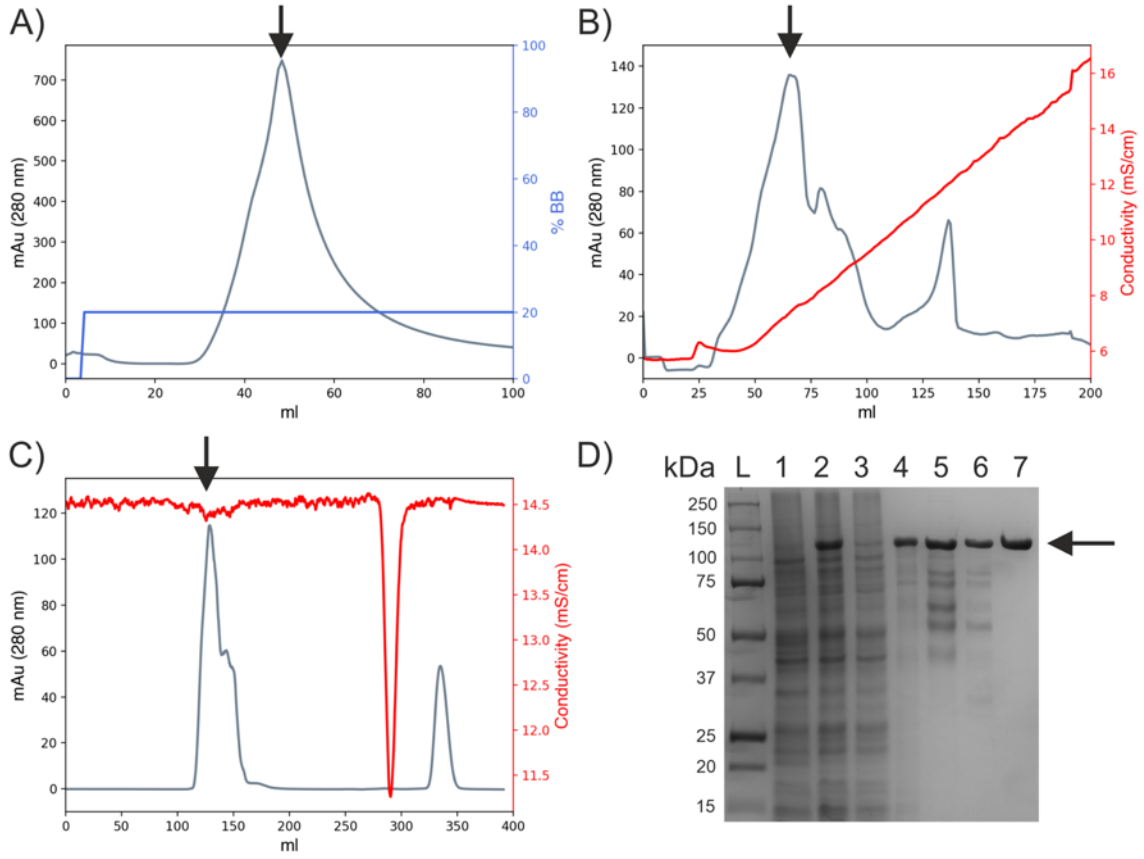

Figure 4.7 – Analysis of the gene expression, production and purification process for pWT $(E\text{-}G5^2)_5$ from *E. coli* BL21 (DE3) pLysS. Examples of A) the elution chromatogram from an the IMAC purification step, where the blue line represents the elution buffer (Section 2.1.14.4) percentage utilised on the purification system. B) The elution profile from the AEX purification step, where the red line represents the conductivity. C) chromatogram of the SEC purification step, where pure protein was isolated. Vertical arrows correspond to the peaks which were collected for subsequent purification/isolation and SDS-PAGE analysis. D) displays an SDS-PAGE of pWT $(E\text{-}G5^2)_5$ protein production and purification steps. Lane 1 is production at 0 hours, lane 2 corresponds to production at 24 hours, lane 3 is the pellet (insoluble) fraction and lane 4 is the cytosolic fraction, indicating a soluble protein. Lane 5 is the initial IMAC purification step, lane 6 is the subsequent AEX purification step and lane 7 is the final SEC purification step, where pure protein was isolated. Horizontal arrow corresponds to pWT $(E\text{-}G5^2)_5$ at the stages of protein production and purification. Similar to SasG and SasG-ΔEE constructs, pWT $(E\text{-}G5^2)_5$ displays anomalous electrophoretic mobility (~ 76 kDa in size but gives an apparent mass of ~ 125 kDa) potentially due to the large number of acidic residues in the $E\text{-}G5^2$ repeats repelling the SDS molecules[174].

4.2.2 Spectroscopic Analysis of the pWT (E-G5²)₅ Protein Fold

To ensure the pWT (E-G5²)₅ construct recapitulated the SasG in secondary structure, far-UV CD spectroscopy was carried out as outlined in Section 2.2.5.1. This was necessary as the mechanical strength of SasG is predicted to be dependent on the topology.

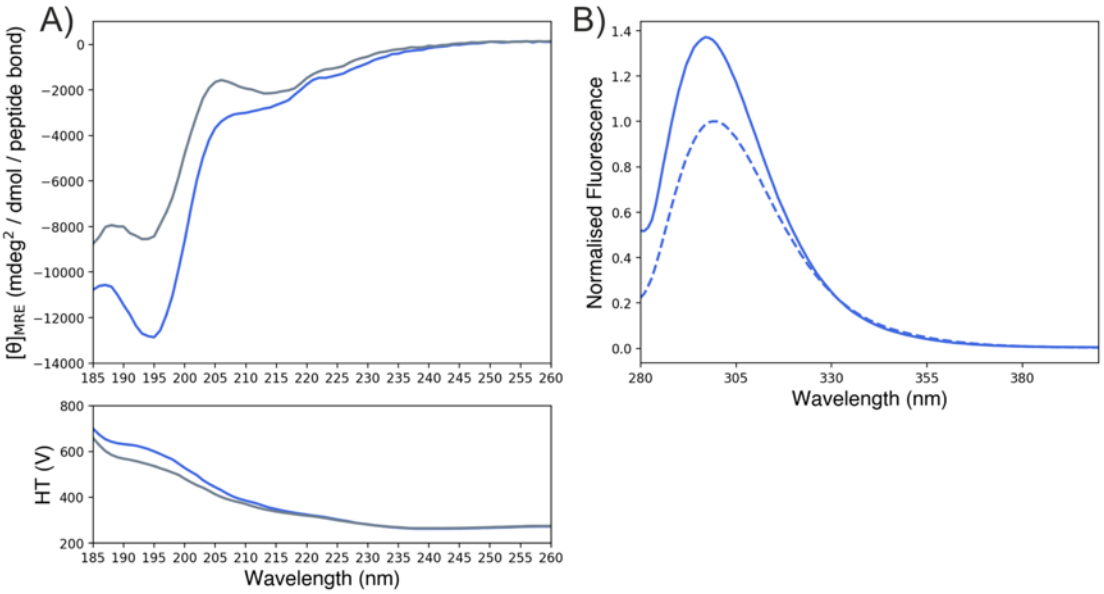

Figure 4.8 – Spectroscopic analysis of the pWT (E-G5²)₅ protein fold. A) Far-UV CD spectra of SasG (grey) and pWT (E-G5²)₅ (blue) with HT voltage displayed below. Beyond a HT of 600 V the spectra begins to become noisy. pWT (E-G5²)₅ shares a similar secondary structure with SasG. B) Fluorescence emission scans of pWT (E-G5²)₅ in the absence (solid) and presence (dashed) of 8 M urea to assess the tertiary structure post-purification.

Consistent with SasG, the far-UV spectrum of pWT (E-G5²)₅ (Figure 4.8A) is dominated by the minimum at around 195 nm, the signal for a collagen triple helix[176,177]. In addition, the observation of a negative minimum at 210-220 nm suggests β-sheet/β-hairpin structure[178,179]. The minimum at 195 nm displays a stronger signal and is slightly red-shifted in comparison to SasG, suggesting random coil signals, possibly due to inclusion of the unstructured linkers.

To ensure that pWT (E-G5²)₅ maintained tertiary structure post-purification, intrinsic tyrosine fluorescence was monitored in the presence and absence of 8 M urea (Figure 4.8B). For pWT (E-G5²)₅, like SasG, tyrosines are found buried in the pseudohydrophobic cores of the interfaces. In E-G5², Y547 is found packed in the E-G5² interface and the other, Y625, is situated at the N-terminus of the G5 domain. In SasG, Y625 would be packed inside a pseudohydrophobic core at

the G5²-E interface. In the absence of urea, pWT (E-G5²)$_5$ has a spectrum typical of a folded protein, with a λ_{max} of 297 nm. Upon addition of 8 M urea, a decrease in fluorescence and a small red-shift (λ_{max} 297 - 299 nm) is indicative that the tyrosines are no longer packed in the pseudohydrophobic cores. These spectroscopic analyses of the pWT (E-G5²)$_5$ construct suggest the topology of SasG has been retained after E-G5² concatenation and is correctly folded post-purification.

4.2.3 Determining the Thermodynamic Stability of pWT (E-G5²)$_5$

The two-state unfolding transition of pWT (E-G5²)$_5$ upon the addition of the chemical denaturant urea was measured to determine the thermodynamic properties (as outlined in Section 2.2.5.2). Briefly, the fluorescence of the construct was monitored by quantifying tyrosine fluorescence emission at 305 nm and fit to a two-state transition model to give the ΔG_{UN} (kJ/mol) and m_{UN} (kJ mol^{-1} M^{-1}) values. Both of these values were identical to those observed for monomeric E-G5² (Table 4.2) suggesting that the domains of the homo-polyprotein were acting independently from one another. Accordingly, the single transition observed in Figure 4.9 indicates unfolding cooperativity between the E and G5² sub domains.

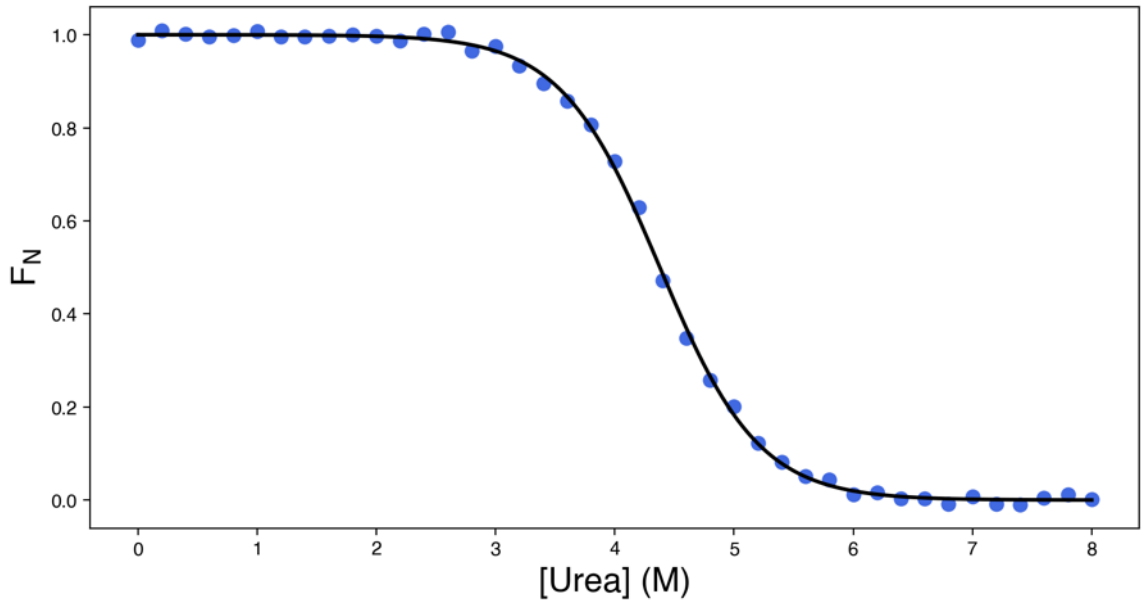

Figure 4.9 – Urea equilibrium denaturation of pWT (E-G5²)$_5$. The data (blue) was normalised using Equation 2.6. Presented as a fraction of folded/unfolded protein (F_N). A two-state model was fit to the data points (black). A single transition indicates cooperativity of the E and G5² domains during unfolding.

Construct	ΔG_{UN} (kJ/mol)	m_{UN} (kJ mol^{-1} M^{-1})
*E-G5^2	26.4 ± 0.8	5.9 ± 0.4
**E-G5^2	26.2 ± 0.8	5.9 ± 0.2
pWT (E-G5^2)$_5$	26.1± 0.8	6.0 ± 0.2

Table 4.2 – Thermodynamic parameters of monomeric E-G5^2 and pWT (E-G5^2)$_5$. *Monomeric E-G5^2 data from reference[74] and **reference[152]. All data was acquired in the same buffer, 1X PBS at 25 °C. The errors were calculated based on the errors on the fits.

4.2.4 Obtaining a Mechanical Fingerprint of pWT (E-G5^2)$_5$

SMFS experiments were carried out as outlined in Section 2.2.5.5.3.2. After the cantilever tip has been brought into contact with the gold surface and a threshold was hit, the cantilever is retracted at a constant retraction velocity resulting in a FX profile with characteristic 'saw-teeth', where each peak (tooth) reports on the unfolding of a single sub domain.

Figure 4.10 shows an example FX profile displaying the complete unfolding of a single pWT (E-G5^2)$_5$ molecule at a constant retraction velocity of 1500 nms^{-1}. As expected, the E sub domains unfolded first as they are mechanically weaker than the G5^2 domains[74] and in turn we observed five smaller peaks followed by five larger peaks. Each peak was described well by the WLC model, which suggests we are observing ten events of protein unfolding followed by one event of detachment of the unfolded polyprotein from the cantilever tip.

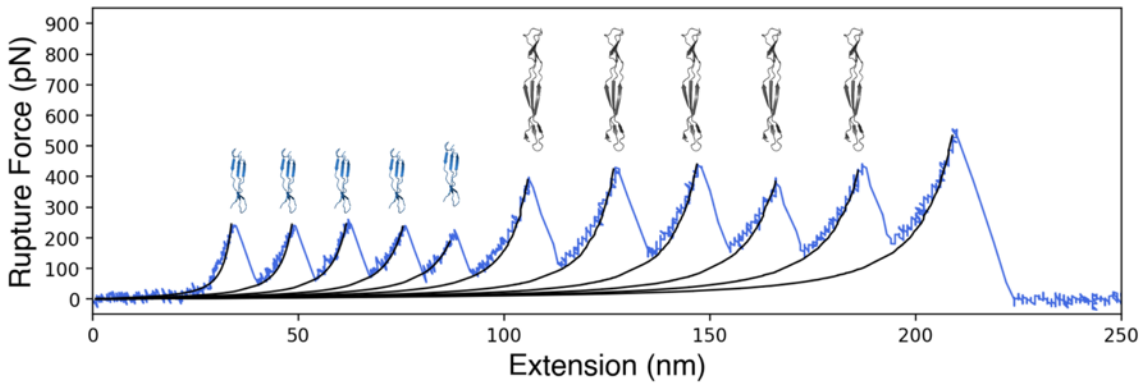

Figure 4.10 – Retraction FX profile of pWT (E-G5^2)$_5$ in 1X PBS, pH 7.4 (room temperature) at a retraction velocity of 1500 nms^{-1}. There are five small peaks (E sub domains) followed by five large peaks (G5^2 sub domains) and finally a large detachment peak. The WLC model fit to each peak is displayed in black. The E and G5^2 sub domains above their corresponding unfolding peaks are shown in blue and grey, respectively.

As discussed in Section 1.3.2, the WLC model allows the extension to be converted into the L_C at each unfolding event. The ΔL_C of a rupture event can be calculated for each unfolding event by subtracting the peaks L_C in question from the following peak L_C. The ΔL_C corresponds to the number of amino acids 'released' during unfolding, which is equal to the length of the polypeptide chain minus the length of the folded protein and in turn permits sub domain assignment to the corresponding unfolding event.

As pWT $(E\text{-}G5^2)_5$ is a distinct construct from SasG, we will outline the process for calculating the predicted ΔL_C. From the crystal structure of $E\text{-}G5^2$ (PDB: 3TIP), the physical length of E can be determined to be 46.8 Å. A caveat of measuring the folded protein in a static state is that we neglect the distance to the mechanical TS, which is why the cited values for the persistence length of an amino acid can vary from protein-to-protein. The persistence length of an amino acid was experimentally determined to be 4.0 $\pm$ 0.2 Å by disulphide bond engineering to control the L_C[62]. We observed using a persistence length of 3.8 Å in the WLC model to fit our data the best (when varying this parameter from 3.8-4.2 Å). Therefore, the theoretical E sub domain polypeptide length can be calculated to be around 197.6 Å (52 amino acids x 3.8 Å). The length of the unfolded protein domain (excluding linkers) minus the N- to C-terminal physical length of the folded protein results in a theoretical ΔL_C of ~ 147 Å for the E sub domain. The physical length of $G5^2$ is measured to be 76.3 Å and the sub domain polypeptide length is around 292.6 Å (77 amino acids x 3.8 Å). Therefore, the ΔL_C would be close to ~ 216.3 Å for $G5^2$ sub domains. The experimental ΔL_C were determined to range from 143.5-151.1 and 210.7-216.7 Å for pWT $(E\text{-}G5^2)_5$ E and $G5^2$ sub domains, respectively. These values were in excellent agreement with both the theoretical values and the experimental values of SasG E (~ 150, ranging from 145-154 Å) and G5 (~ 220, ranging from 216-227 Å) sub domains as observed by Gruszka and colleagues[74] and in this thesis (Section 3.2.3). This indicates that we are observing the unfolding of E and $G5^2$ sub domains and allows the corresponding rupture forces to be correctly assigned.

While the ΔL_C values are indifferent to SasG, we have a distinct construct with both an absence of the G5-E interface (due to the inclusion of linkers) and a difference in the number of sub domains. These may in turn affect the mechanical characteristics[69]. Therefore, the next step is to

mechanically characterise pWT (E-G5^2)$_5$ over several retraction velocities and compare with SasG in matching buffer.

4.2.5 The Mechanical Properties of pWT (E-G5^2)$_5$

In order to assess suitability of pWT (E-G5^2)$_5$ as a surrogate polyprotein system for substituting SasG in mechanical variant analyses, mechanical unfolding experiments in 1X PBS, pH 7.4 were carried out at five retraction velocities (200, 800, 1500, 3000 and 5000 nms^{-1}) in triplicate. An example FX trace for every retraction velocity is displayed in Figure 4.11. For each mechanical unfolding experiment, traces that contained four to ten full unfolding events well described by the WLC chain model were accepted for analysis.

Observation of the maximum number of unfolding events in a single FX profile suggests that all the E-G5^2 domains were folded prior to forced unfolding. From the event apices of these traces, the rupture force values were collected and the WLC model was fit to each event to extract the L_C values for calculation of ΔL_C values. Histograms of the rupture force values (force-frequency) and ΔL_C (ΔL_C-frequency) were plotted with bin sizes of 10 pN and 5 Å, respectively, and Gaussian fits to these provided the mode values. Scatter plots, rupture force-frequency and ΔL_C-frequency histograms of one repeat at all retraction velocities is shown in Figure 4.12. The rupture force and ΔL_C statistics of the triplicate data is displayed in Table 4.3 and Table 4.4.

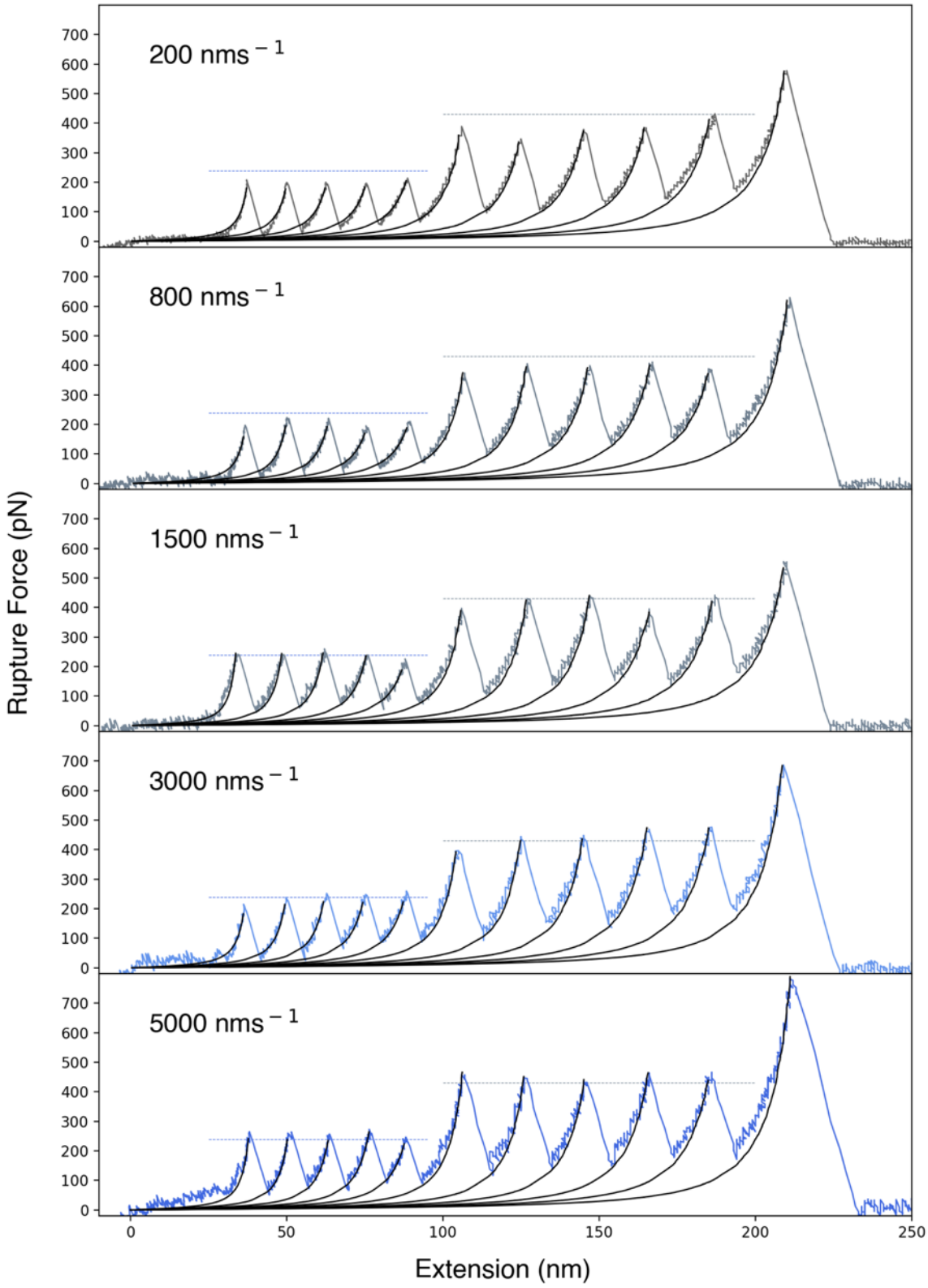

Figure 4.11 – Example pWT $(E\text{-}G5^2)_5$ FX profiles with a retraction velocity of 200, 800, 1500, 3000 and 5000 nms^{-1} in 1X PBS, pH 7.4 at room temperature. Retraction trace displayed only. Noticeable increase in height of the peaks as retraction velocity increases. WLC model fitting with a fixed persistence length of 0.38 nm in black. Average rupture force of E and $G5^2$ domains at 1500 nms^{-1} as blue and grey dotted lines, respectively.

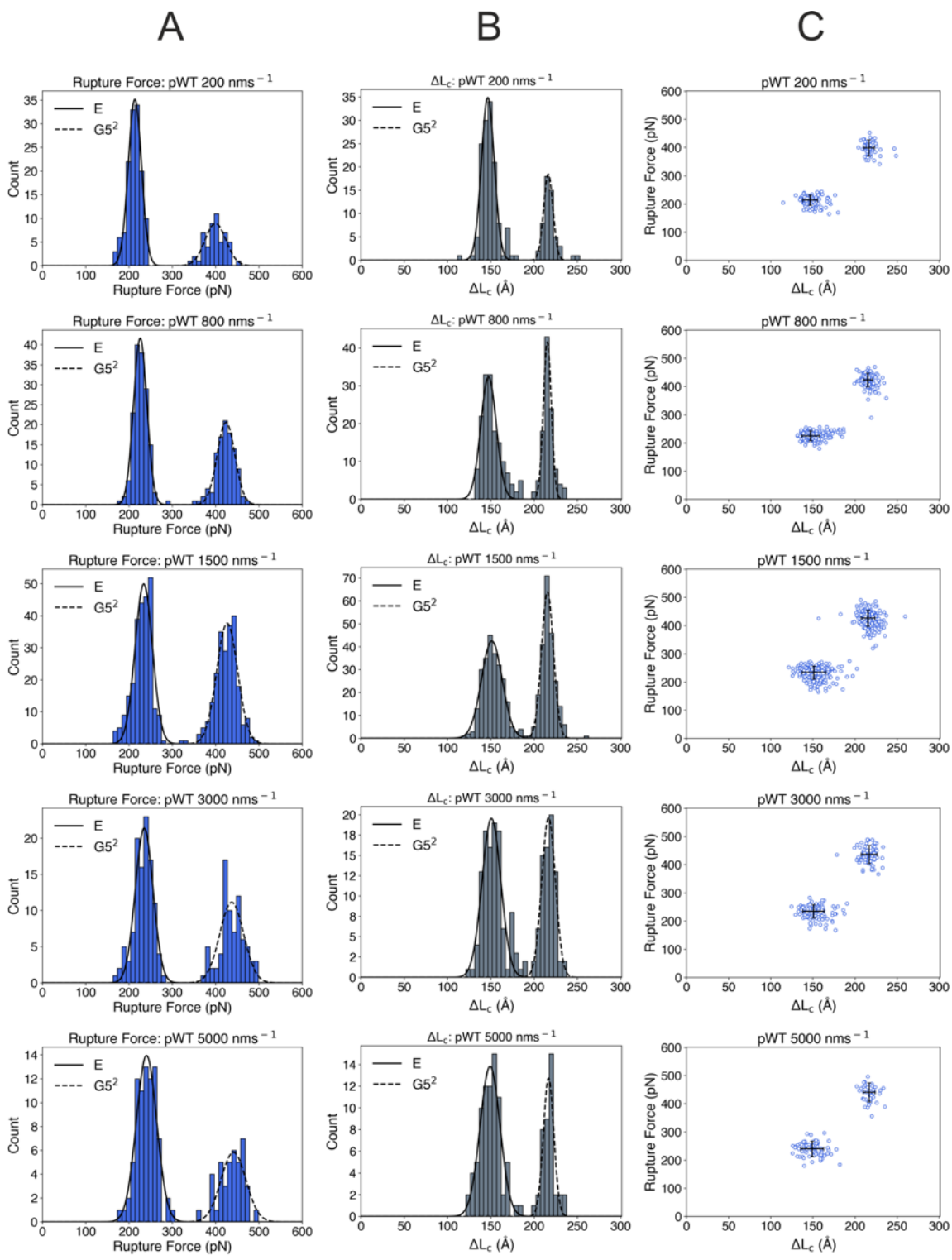

Figure 4.12 – Example pWT $(E\text{-}G5^2)_5$ mechanical unfolding data in 1X PBS, pH 7.4 for one tip at five retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Rupture force increases as retraction velocity increases. Black scatterplot crosshairs are the mode and the FWHM from the corresponding histogram Gaussian model fits.

E **pWT (E-G5²)₅**

Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average ΔL_C (Å)
200	135	213.3	**224.1**	146.5	**146.3**
	218	224.2		146.3	
	116	217.6		146.0	
800	151	225.7	**227.9**	147.3	**146.0**
	150	229.4		146.0	
	171	228.6		144.8	
1500	248	233.6	**238.5**	151.1	**147.7**
	266	243.8		146.1	
	120	238.1		145.9	
3000	108	234.9	**241.2**	150.7	**147.7**
	217	245.9		147.1	
	84	242.9		145.5	
5000	79	240.3	**246.1**	149.1	**146.4**
	130	250.4		146.7	
	100	247.6		143.5	

Table 4.3 – Summary of rupture force and ΔL_C statistics for pWT (E-G5²)₅ E sub domain mechanical unfolding in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode rupture force/ΔL_C are obtained from the Gaussian fits to the histograms. Average is the mean of the mode values at each speed.

G5² **pWT (E-G5²)₅**

Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average ΔL_C (Å)
200	53	399.6	**401.6**	216.3	**215.4**
	81	401.6		215.0	
	41	403.6		214.7	
800	114	423.9	**423.9**	215.2	**214.8**
	83	421.1		215.0	
	66	426.6		214.0	
1500	235	426.5	**431.5**	215.3	**214.8**
	104	429.7		215.6	
	40	438.4		213.4	
3000	79	437.0	**439.0**	216.7	**213.8**
	188	438.7		214.0	
	44	441.4		210.7	
5000	42	441.1	**445.2**	216.7	**214.7**
	60	444.7		213.3	
	68	449.8		214.1	

Table 4.4 – Summary of rupture force and ΔL_C statistics for pWT (E-G5²)₅ G5² sub domain mechanical unfolding in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode rupture force/ΔL_C are obtained from the Gaussian fits to the histograms. Average is the mean of the mode values at each speed.

The pWT (E-G5²)₅ E and G5² average rupture forces were plotted against the logarithm of retraction velocity at five speeds to determine the speed dependence and compared with SasG (data from Section 3.2.3) in a matching buffer (Figure 4.13). As expected, the rupture force for both E and G5² sub domains increases as retraction velocity increases and the data could be well described with a linear fit.

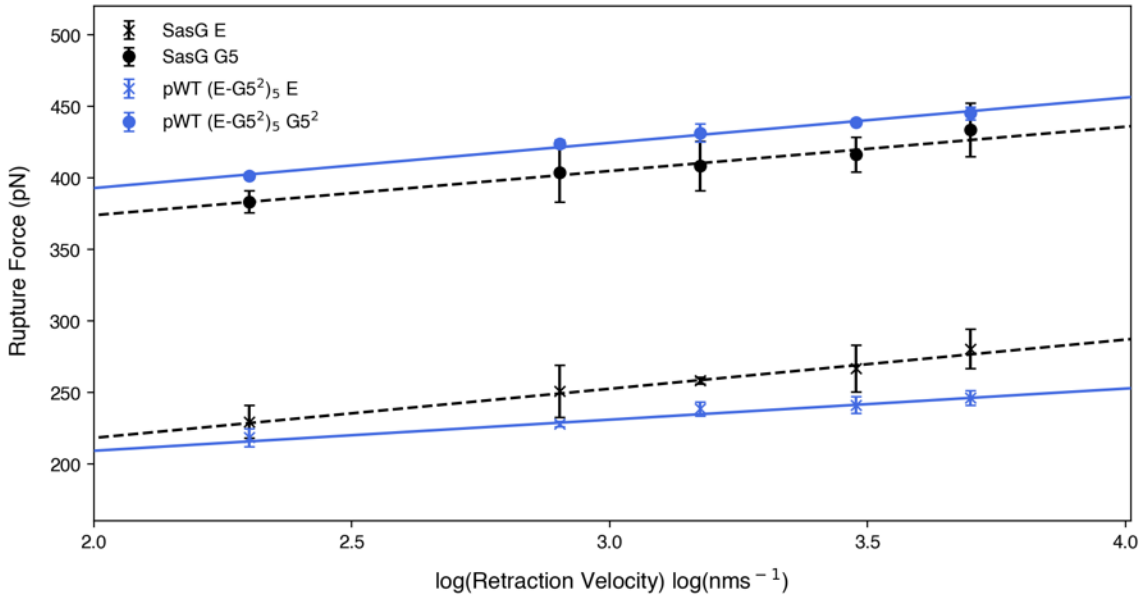

Figure 4.13 – The dependence of rupture force on the logarithm of retraction velocity of E (crosses) and G5²/G5 (filled circles) sub domains of pWT (E-G5²)₅ and SasG. Points and errors are the mean and standard deviation of triplicate datasets. Linear fit is weighted with the inverse of the standard deviation error of the triplicate datasets.

The data shows that the G5² sub domain of pWT (E-G5²)₅ is consistently unfolding at a significantly (ANCOVA $p < 0.05$) higher force than the SasG G5 sub domain (~ 20 pN). Although the E sub domain is unfolding at a similar force at the slowest speed (200 nms^{-1}) there is a gradual, but significant, deviation from SasG E sub domain rupture forces as the retraction velocity increases. As the linear fits of the E sub domain were determined to be significantly different (ANCOVA $p < 0.05$), we can accept the decrease in the gradient. As the gradient of the speed dependence is inversely proportional to the distance to the mechanical TS[44], the decrease in gradient suggests the pWT (E-G5²)₅ E sub domain distance to the mechanical TS has increased. However, as we have changed the supramolecular scaffold and domain number, alterations to the mechanical properties of the sub domains is not unexpected[69].

4.3 Discussion

Guided by crystal structures (PDB: 3TIP & 3TIQ), uniprot boundaries and published ΔL_C values, we successfully isolated E-G5^2 and constructed the pentameric homo-polyprotein, pWT (E-G5^2)$_5$. GG was exploited for the efficient assembly of DNA cassettes, due to its increased efficiency, accuracy and decreased labour expense in comparison to other techniques such as GA or sequential digestion and ligation of cassettes[30,163,197]. The circular dichroism spectrum of pWT (E-G5^2)$_5$ construct indicated analogous secondary structure to SasG, indicating the topology had not been perturbed by the incorporation into a polyprotein. Variations between the spectra are likely due to the incorporation of the disordered linker regions giving rise to random coil signals skewing the spectrum (minima between 195-200 nm)[178]. More specifically, the stronger minima at 195 nm (signal for a collagen structure) with a red-shift in pWT (E-G5^2)$_5$ spectrum is probably due to the accumulation of the random coil signals with signals from the collagen-like region as they occupy overlapping wavelengths[176,177]. This has been observed before, where a linker (GGGSGGGG) between two I27 monomers has resulted in a stronger minima with a slight red-shift in comparison to an I27 monomer in isolation[199]. Retaining the same fold as SasG E and G5 sub domains was essential, as the mechanical strength of these sub domains is believed to be related to their topology[74]. Any deviation from the correct fold could affect the accuracy of the protein system.

SasG is believed to behave as a series of overlapping cooperative units mediated by both the intrinsic instability of the E domains and the stability of the interfaces[74,152]. This is an unusual property of multidomain proteins, which are either usually highly unstable with extremely stable interfaces[200] or have interfaces which impart short-range cooperativity between neighbouring domains only[201]. Interestingly, our pWT (E-G5^2)$_5$ construct shares identical thermodynamical stability parameters, ΔG_{UN} and m_{UN}, with monomeric E-G5^2[74,152]. This indicates that the E-G5^2 repeats in pWT (E-G5^2)$_5$ unfold independently from one another (uncooperatively), but the E and G5^2 sub domains unfold cooperatively. These observations were not unexpected though as the inclusion of seven residue unstructured linkers were expected to distance the neighbouring domains considerably. In turn this prevents a compact G5-E interface, which is a requirement for long-range cooperativity[74]. The effect on the thermodynamic stability of polyprotein domains

distanced by longer linkers (n=8) vs shorter linkers (n=2) has been recently observed in an I27 polyprotein[199]. The shorter linkers were shown to increase $\Delta G_{UN}/m_{UN}$ of the polyprotein in comparison to the longer linkers (displaying monomeric equivalent thermodynamic properties), through predicted interactions of the interfaces and with the short linker. Non-native interactions from linkers or incorrect domain boundaries at interfaces would have had the potential to skew the mechanical properties in downstream SMFS experiments. Our results suggest that our polyprotein construct does not contain any faux interdomain interactions or interactions with the linkers. Furthermore, the monomeric values of the pWT $(E\text{-}G5^2)_5$ circumvent the necessity for monomeric biophysical studies and in turn made construct characterisation rapid and permits comparison with values from monomeric studies[152].

Under mechanical force, the E and $G5^2$ sub domains of pWT $(E\text{-}G5^2)_5$ unfold independently from one another instead of as a cooperative unit[74]. There are two major populations for both rupture force and ΔL_C values, corresponding to the unfolding of the E and $G5^2$ sub domains. The ΔL_C values of E and $G5^2$ did not deviate from those published for SasG E and G5 sub domains[74] and our SasG values (Section 3.2.3). However, both E and $G5^2$ displayed a deviation in their mechanical properties. This is not entirely unexpected though as there are consequences when changes to the supramolecular scaffold and domain number are made[69]. The inclusion of unstructured polypeptide linkers will likely increase the compliance of the construct and in turn sub domains will spend more time at a lower force at the same extension rate. This will result in lower unfolding forces as thermal fluctuations have more time to drive the protein over the mechanical TS. Conversely, there are both fewer E (5 vs 6) and G5 (5 vs 7) sub domains in the pWT $(E\text{-}G5^2)_5$ in comparison to SasG, which will act to increase the unfolding force of the sub domains as there are less attempts per unit of time for the protein to cross the mechanical TS. This is what we observed for the $G5^2$ sub domain. Unexpectedly, both the unfolding force and speed dependence gradient of the E sub domain decreased. As the latter is inversely proportional to the distance to the mechanical TS[44], the decrease in the gradient of the E sub domain indicates the distance to the mechanical TS has been increased. These observations suggest that the compact G5-E interfaces are important for both the mechanical strength and malleability of the E sub domain, which we investigate in Section 5.3.4.3.

Furthermore, we are only sampling the second E sub domain and $G5^2$ and although SasG boasts very high levels of amino acid homology between B-repeats, it is not always 100 %. Subtle differences between sub domains which may get concealed in the statistics when all E and G5 sub domains are being sampled (during SasG unfolding) may accumulate when an identical sub domain is investigated.

The most suitable model for determining the mechanical properties of SasG would be the native construct, however, creating variants would be a time consuming and expensive process. This is circumvented by the creation of the pWT $(E\text{-}G5^2)_5$ protein system, which offers rapid and efficient variant creation and analogous secondary structure. Although there are somewhat altered mechanical properties in comparison to SasG, we are comparing variants of pWT $(E\text{-}G5^2)_5$ so any changes in the mechanical properties should theoretically be inconsequential to our analysis. Thus, the results presented in this chapter have demonstrated and validated a novel construct, pWT $(E\text{-}G5^2)_5$, as an ideal model system to substitute SasG in SMFS studies.

5. Uncovering the Mechanical Determinants of SasG Utilising a Combination of Protein Mutagenesis and SMFS

5.1 Objectives

To our knowledge, the intramolecular interactions and/or structural motifs responsible for the mechanical strength of the SasG B domain have not been experimentally determined, only proposed. The main objective of this chapter was to create variants of pWT $(E\text{-}G5^2)_5$ and resolve changes in mechanical stability using SMFS. More specifically, the proposed hydrogen bonded 'mechanical clamp' region of the E and G5 sub domains, charge networks, the E-$G5^2$ interface and the collagen-like regions were examined for potential force-bearing properties. For each variant the ground state stability at equilibrium was determined using chemical denaturation studies and the forced unfolding kinetics of the local region of substitution was characterised utilising SMFS. Combining these two measurements (mechanical ϕ-value analysis) gives us a powerful insight into the local structure of substitution at the mechanical TS.

5.2 Structural Indicators of Mechanical strength

At first glance, E and G5 sub domains fulfil several qualities of mechanically strong proteins as they are predominantly β-sheet (Figure 5.1A) with long stretches of hydrogen bonds[74]. However, they display an absence of terminal parallel β-strand mechanical clamps and due to their planar β-sheet structure they lack a typical compact hydrophobic core. The published unfolding forces[74] for E and G5 sub domains are ~230 and ~380 pN at a retraction velocity of 1500 nms^{-1}, respectively, and are higher than many of those reported in the literature including I27[29]. Gruszka and colleagues hypothesised this mechanical strength originates from tandemly arrayed mechanical clamps involving long stretches of hydrogen bonds and associated side-chain packing interactions[74] (Figure 5.1C). The difference in mechanostability between the E and G5 sub domains was presumed to be due to the longer length of the N-terminal G5 clamps. These proposed clamps appear reminiscent to a mix of SD1, shear parallel and shear antiparallel clamps[91,92], however, they are separated by

a collagen-like motif responsible for strand swapping[149]. This motif resembles the collagen triple-helical structural motif[202] (Figure 5.1B) and was first described as the β-triple helix-β supersecondary structural motif by Ruggiero and colleagues for the Resuscitation-promoting factor B (RpfB) G5 domain expressed by *Mycobacterium tuberculosis*[203]. Like RpfB, the motif contains a mixed parallel/antiparallel chain arrangement leading to a staggered distorted triple helix.

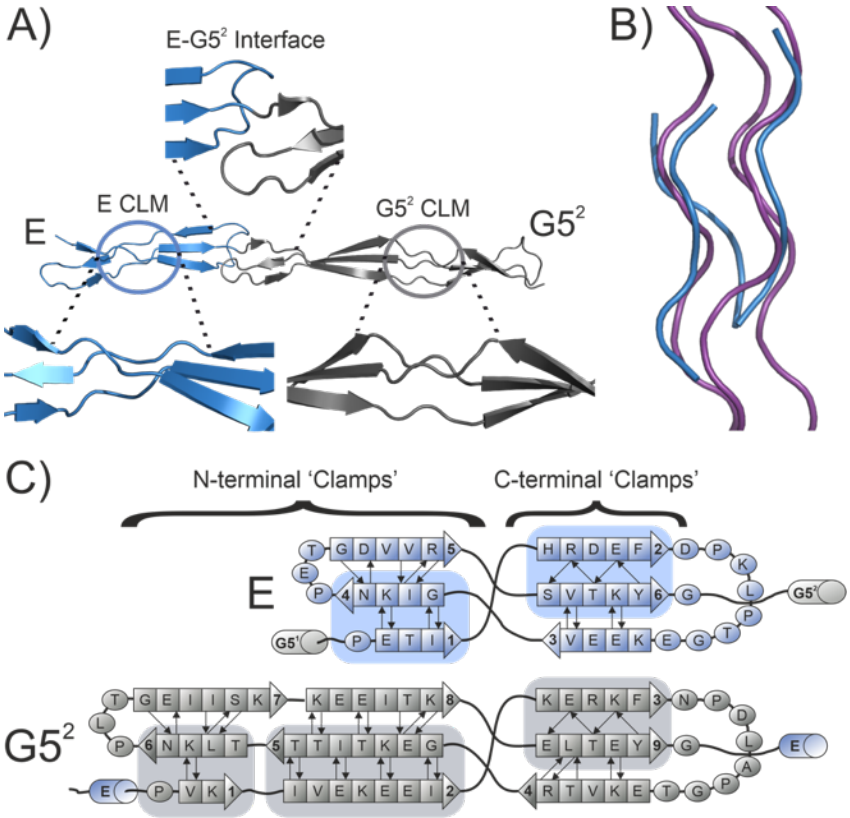

Figure 5.1 – E-G5² structural properties. A) Three dimensional crystal structure of E-G5² (PDB: 3TIP) with collagen-like motifs (CLM) and the E-G5² interface pseudohydrophobic core highlighted. E and G5² sub domains displayed as blue and grey, respectively. B) E collagen-like region (from PDB: 3TIP, coloured in blue) aligned with collagen (PDB: 1BKV, coloured purple) to display the collagen-like homology and the observable staggering of 1/3 strands. C) Topology diagrams of E and G5² sub domains displaying potential β-sheet 'mechanical clamps', which were predicted to be the force-bearing regions of SasG E and G5 sub domains[74].

To delineate the determinants of the remarkable mechanical strength of SasG we first utilised proline scanning mutagenesis in an attempt to block hydrogen bonding and introduce bulges into the β-sheets of the proposed tandemly-arrayed 'mechanical clamps'. Proline was exploited as it is unable to complete the hydrogen bonding network (as it lacks a hydrogen on the α amino group) and causes a bulge due to the main chain steric restrictions of proline in the phi (ϕ) and psi (ψ) conformational space arising from the five-membered ring. Hongbin Li and colleagues exploited the

same methodology coupled with SMFS, which helped to experimentally verify the force-bearing region of I27[90]. The five variants are outlined in Figure 5.2 and comprise I502P, V522P, V550P, V556P and V580P.

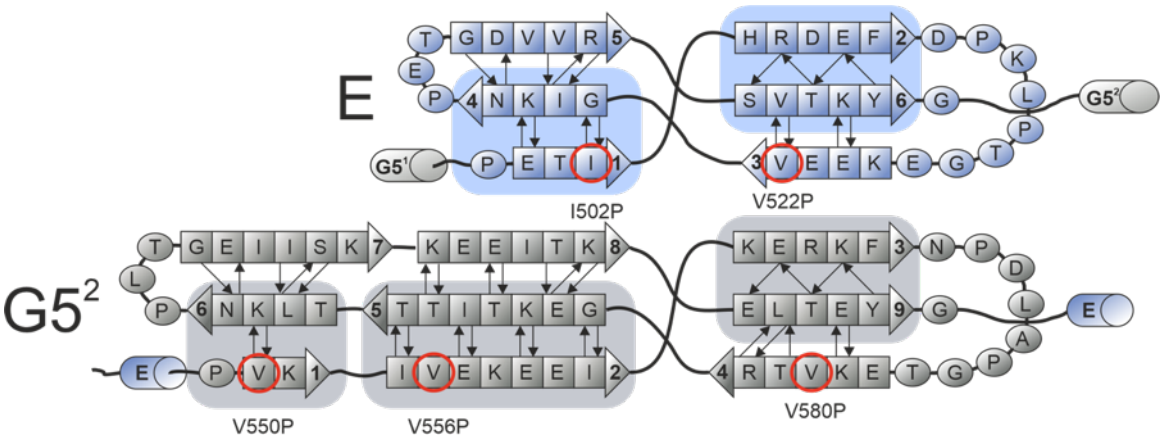

Figure 5.2 – Proline scanning mutagenesis of E-G5² in an attempt to disrupt any hydrogen bonded β-sheet 'mechanical clamps'. Schematic of the secondary structure of E and G5², with the proline substitution mutations circled in red and labelled.

Secondly, alanine scanning mutagenesis of the collagen-like motif was carried out with residue selection guided by Gruszka and colleagues recent thermodynamic and kinetic experimental work on the E-G5² monomer[152]. These included, but are not limited to, the G524A/G527A and G584A/G587A pairs, which are structurally equivalent glycine residues located in one of the PPII-like strands following the typical collagen sequence $X_{aa}Y_{aa}GlyX_{aa}Y_{aa}Gly$ in the E and G5² sub domains, respectively. P540A and P562A are found in non-corresponding E and G5² PPII-like helices, respectively. All six are outlined in Figure 5.3.

A second set of alanine scanning mutations included those close to the interface between E and G5² and included G517A (E C-terminal loop), P549A (located in the two residue GP linker between E-G5²) N598A and T601A (G5² N-terminal loop). The former two were guided by Gruszka and colleagues recent thermodynamic and kinetic experimental work on the E-G5² monomer[152], whereas the latter two were included to disrupt the potential internal hydrogen bonding networks in the G5² N-terminal loop which forms part of the G5² interface. These are outlined in Figure 5.3. The final set of variants were the charge reverse variants. These included E588K, K589E and E624K and were chosen to disrupt side chain interactions in highly charged regions of the 'mechanical clamps'. These are outlined in Figure 5.3 and are located exclusively in the G5² sub domain.

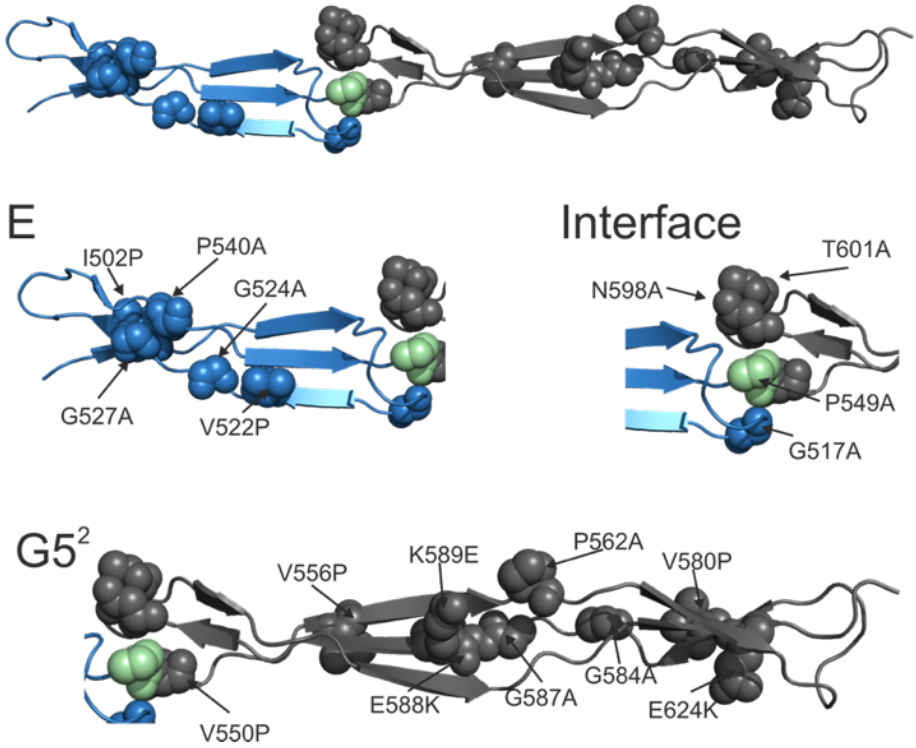

Figure 5.3 – Variants of this study broken into three sections of E-G5². E sub domain variants included I502P, V522P, G524A, G527A and P540A. Interface mutants comprised G517A, P549A, N598A and T601A. G5² variants comprised V550P, V556P, P562A, V580P, G584A, G587A, E588K, K589E and E624K. E and G5² structures coloured in blue and grey, respectively. P549A is coloured in green as it is part of the short GP linker between E and G5² sub domains.

5.3 Results

5.3.1 Cloning, Gene expression, production and Purification of pWT $(E\text{-}G5^2)_5$ Variants

Q5® PCR was used to introduce point mutations into the pET14b-E-G5² ('working vector') DNA sequence to create 18 monomeric mutant template sequences. All primers and mutations are outlined in Section 2.2.2.2, Table 2.2. Once checked by DNA sequencing, the mutant template sequences were used to assemble pentameric homo-polyproteins as outlined with pWT $(E\text{-}G5^2)_5$ (Section 4.2.1.5). Gene expression, protein production and purification were consistent with that of pWT $(E\text{-}G5^2)_5$ (Section 4.2.1.6). At the end of each variant purification, the purity and identity was assessed by SDS-PAGE and LC-MS, respectively, which confirmed the successful isolation of all variants (measured masses found in Table 5.1). The full protein sequences are found in the Appendix (Section 7.2.3).

Variant	Expected Mass (Da)	Measured Mass (Da)
I502P	76014.8	76014.4 ± 0.8
G517A	76165.2	76163.8 ± 0.3
V522P	76085.0	76084.2 ± 1.6
G524A	76165.2	76165.6 ± 1.0
G527A	76165.2	76163.6 ± 2.0
V550P	76085.0	76083.7 ± 2.9
V556P	76085.0	76085.5 ± 1.8
P540A	75964.9	75963.1 ± 0.9
P549A	75964.9	75965.1 ± 0.5
P562A	75964.9	75964.8 ± 0.3
V580P	76085.0	76085.8 ± 2.0
G584A	76165.2	76166.0 ± 1.0
G587A	76165.2	76164.0 ± 0.6
E588K	76090.3	76090.7 ± 0.5
K589E	76099.8	76099.6 ± 0.5
N598A	75879.9	75880.3 ± 0.4
T601A	75944.9	75943.9 ± 0.3
E624K	76090.3	76092.4 ± 3.1

Table 5.1 – Table of pWT $(E\text{-}G5^2)_5$ variants, the expected mass and the measured mass by LC-MS. Every variant was within error of the expected mass (with the N-terminal methionine excised) and machine error (20 ppm: ~ 1.5 Da for proteins of this size).

5.3.2 Spectroscopic Analyses of Variants

Various spectroscopic analyses were carried out to determine if the variants were folded post purification and whether amino acid substitution had perturbed the secondary structure. In order to analyse the effects of amino acid substitution on the secondary structure of the variants, far-UV CD spectroscopy was carried out (Figure 5.4, Figure 5.7, Figure 5.9 and Figure 5.11). As topology is believed to be the main determinant of a proteins mechanical force response, evidence of the presence/absence of secondary structure perturbation can be coupled with SMFS data to report on mechanically important structures. Tyrosine fluorescence emission was utilised to monitor the tertiary fold in the presence and absence of 8 M urea. As with pWT $(E\text{-}G5^2)_5$, excitation at 276 nm and the emission from 280-400 nm was recorded. For all variants there was a clear decrease in fluorescence and small λ_{max} red-shift observed in 8 M urea, suggesting that the tyrosines were more exposed to solvent in the chemical denaturant (Figure 5.5, Figure 5.8, Figure 5.10 and Figure 5.12). This indicates all the variants adopted at least some tertiary structure in the absence of the chemical denaturant.

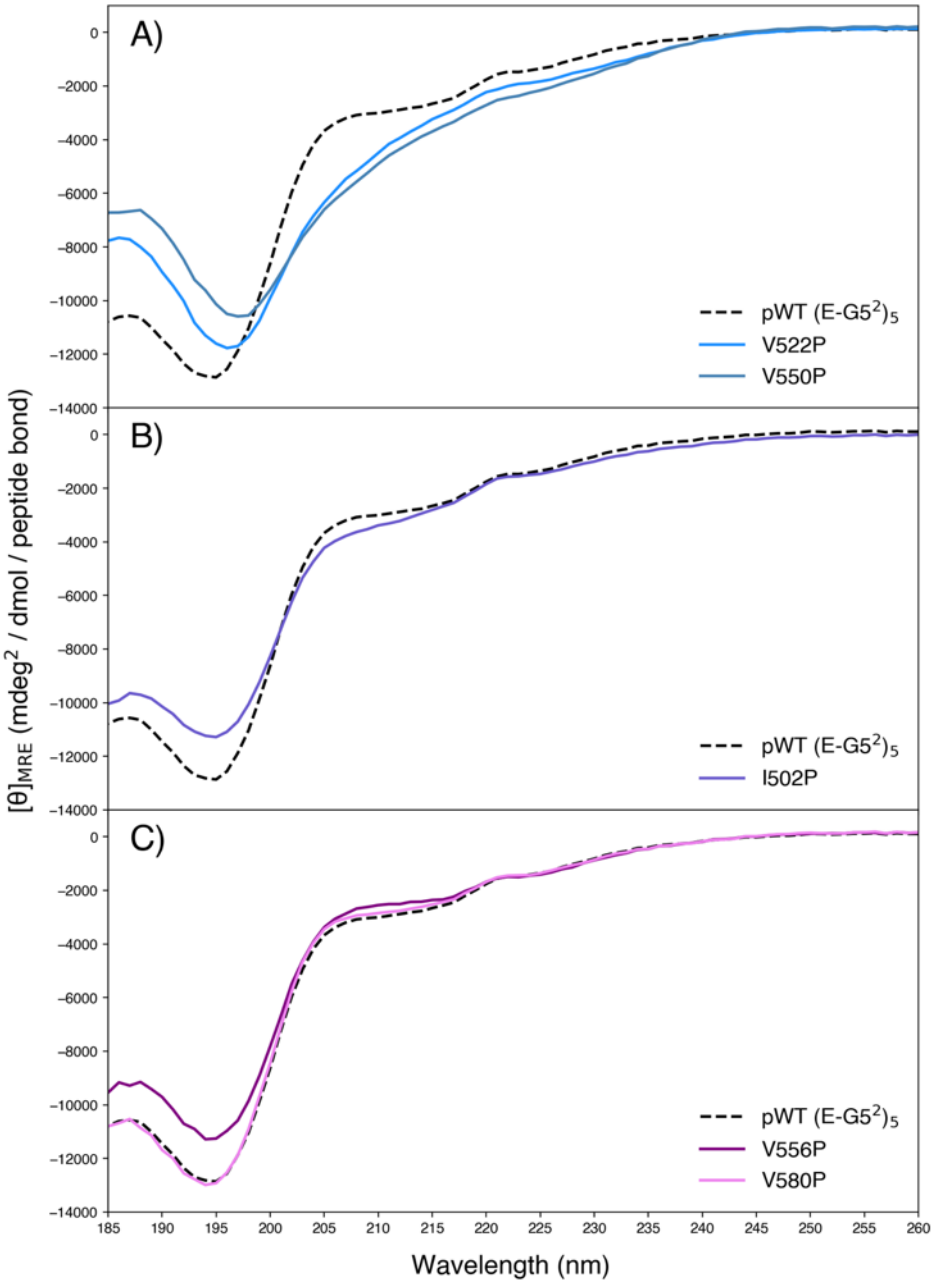

Figure 5.4 – Far-UV CD spectra of the potential 'mechanical clamp' proline variants. A) V522P and V550P, B) I502P. C) V556P and V580P. These spectra show proline mutations either severely (V522P and V550P) or moderately (I502P and V556P) affected the secondary structure. pWT $(E\text{-}G5^2)_5$ displayed in every spectrum for visualisation of secondary structure deviation.

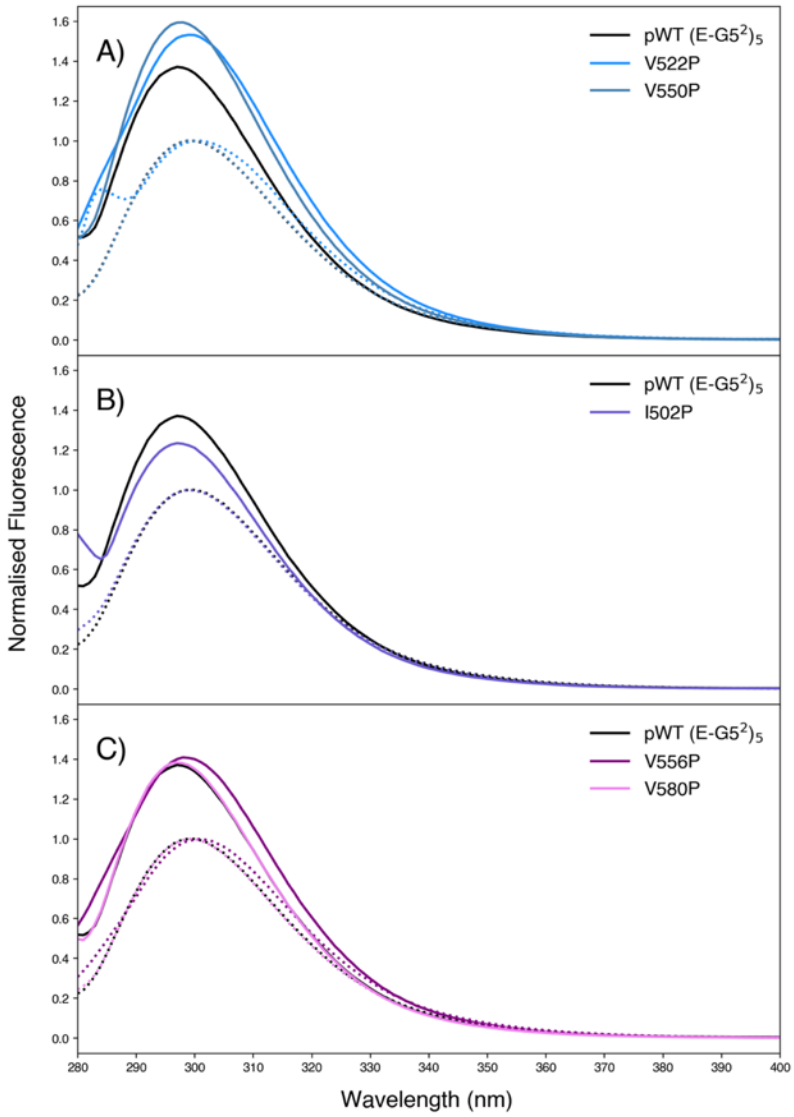

Figure 5.5 – Intrinsic tyrosine emission spectra of the potential 'mechanical clamp' proline variants. A) V522P and V550P, B) I502P and C) V556P and V580P. Absence of urea (solid line) and presence of 8 M urea (dashed lines). pWT $(E\text{-}G5^2)_5$ is included for reference. All variants display tertiary structure, which is lost in the presence of chemical denaturant.

CD spectra of the 'mechanical clamp' variants are displayed in Figure 5.4. Both V522P and V550P display major differences in comparison to pWT $(E\text{-}G5^2)_5$ (Figure 5.4A), with a red-shift in the minimum at around 195 nm (signal for a collagen triple helix) to a longer wavelength of approximately 200 nm indicating random coil structure[176–178]. Furthermore, smoothing of the spectra suggests general loss of structure, such as the loss of the β-sheet and β-turn overlapping minimas (around 205-220 nm)[178,179]. This data suggests these variants are fully unfolded or have significantly unfolded regions. However, Figure 5.5 suggests that all the 'mechanical clamp' proline variants have tertiary structure post-purification, which indicates that V522P and V550P are actually partially

folded in the absence of chemical denaturant. As only G5^2 can fold independently[74,152] we can speculate it is the G5^2 sub domain which is folded in both of these variants.

I502P displays pWT (E-G5^2)$_5$-like structure, however, there is evidence of partial unfolding, with a decrease in the β-sheet signal at 205-215 nm and a slight red-shift of the collagen triple helix minimum towards random coil structure (Figure 5.4B). The accumulation of these effects results as slight smoothing of the spectrum and may imply that the mutation has prevented part of the N-terminal β-sheet from fully forming. This is compounded by a decrease in intensity of the 'folded' structure in comparison to pWT (E-G5^2)$_5$ indicating a larger amount of solvent exposed residues in the absence of chemical denaturant (Figure 5.5). In addition, the FACTS implicit solvent model[204] of I502P predicts an unstructured N-terminal E sub domain (Figure 5.6).

V556P and V580P CD spectra are displayed in Figure 5.4C and are both comparable to pWT (E-G5^2)$_5$, however, V556P displays a slight decrease in the collagen and β-sheet signals, which may indicate a slight overall change in secondary structure.

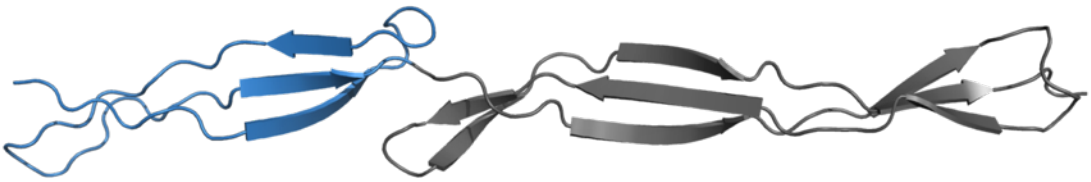

Figure 5.6 – I502P FACTS implicit solvent model predicts the N-terminal of the E sub domain to be unstructured. Created using an Implicit Solvent Modeller[205].

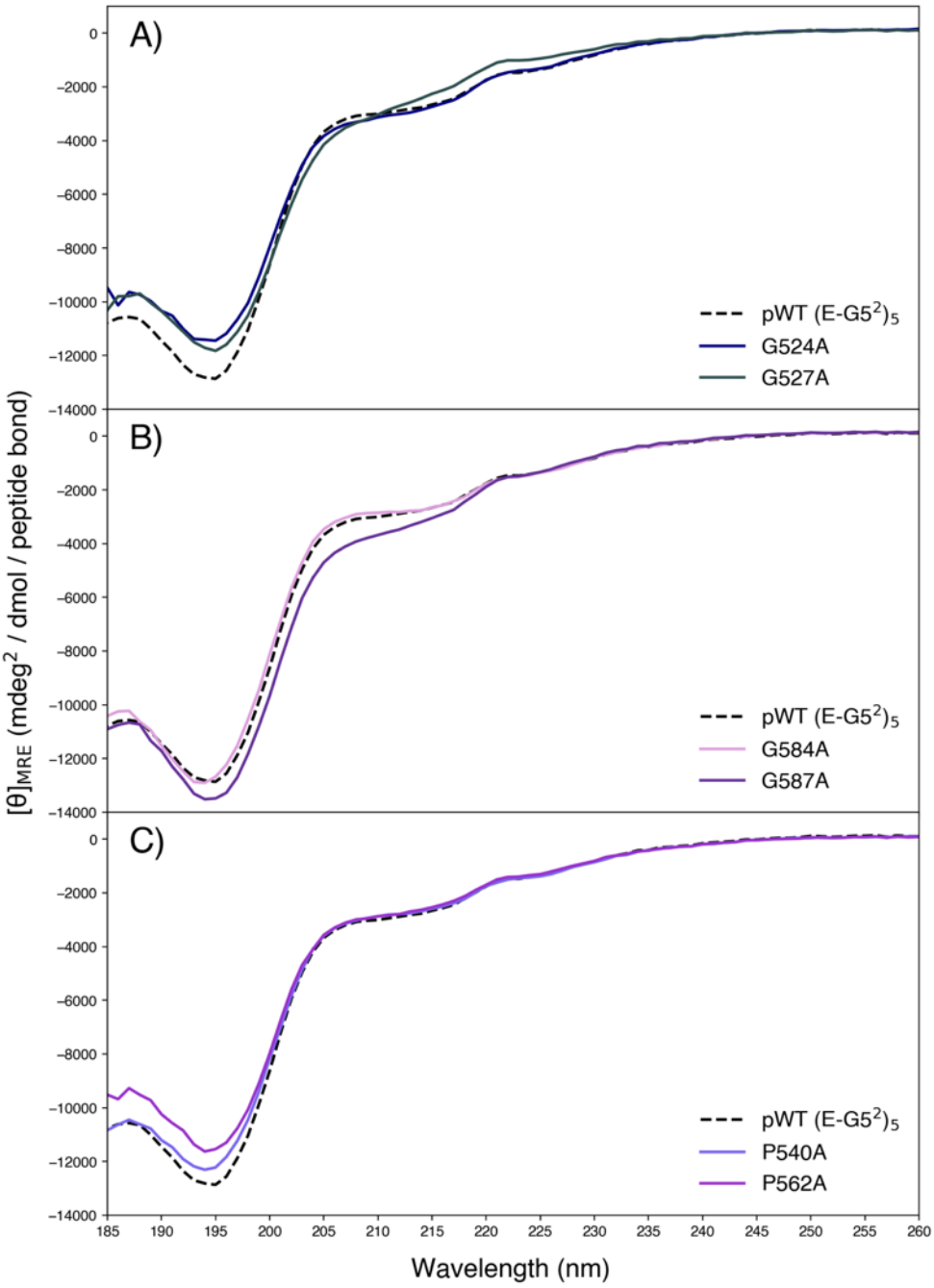

Figure 5.7 – Far-UV CD spectra of the collagen-like motif variants. A) G524A and G527A situated in the E sub domain. B) G584A and G587A situated in the $G5^2$ sub domain. C) P540A and P562A, situated in the E and $G5^2$ sub domains, respectively. pWT $(E\text{-}G5^2)_5$ displayed in every spectrum for visualisation of secondary structure deviation.

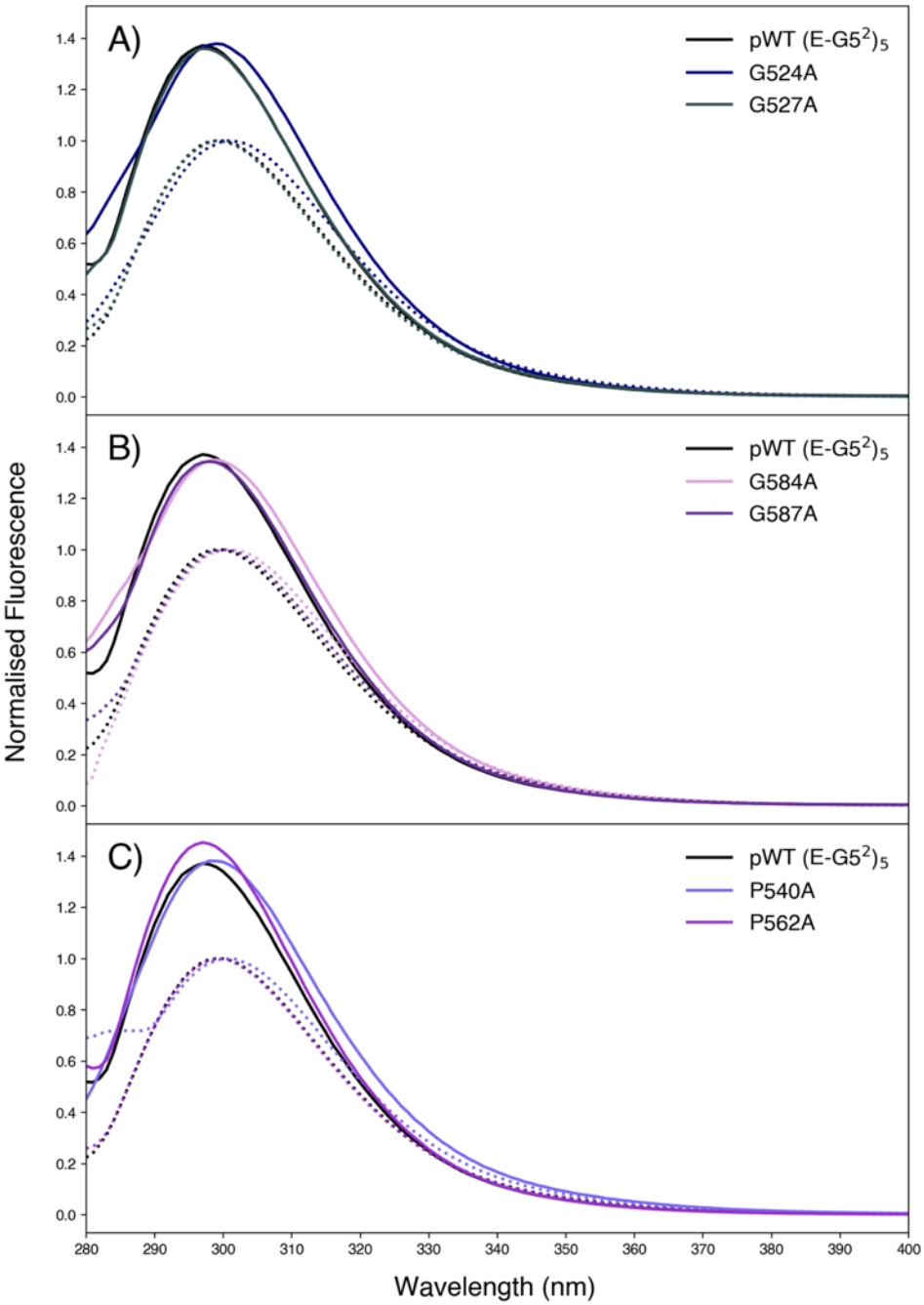

Figure 5.8 – Intrinsic tyrosine emission spectra of the collagen-like region variants. A) G524A and G527A, B) G584A and G587A, and C) P540A and P562A. Absence of urea (solid line) and presence of 8 M urea (dashed lines). pWT $(E\text{-}G5^2)_5$ is included for reference. All variants display tertiary structure, which is lost in the presence of chemical denaturant.

CD spectra of the collagen-like region variants are displayed in Figure 5.7. G524A and G527A of the collagen-like motif variants deviate marginally from pWT $(E\text{-}G5^2)_5$ with a decrease in intensity in the minimum at 195 nm (signal for the collagen triple helix). However, G527A also displays a slight red-shift of said minimum and a smoothing of the 'dip' between 210-215 nm. Interestingly,

G587A displays a similar spectrum to G527A. This change in spectra could suggest perturbation of the collagen-like region structure. The G584A spectrum is analogous to that of pWT $(E\text{-}G5^2)_5$ suggesting similar secondary structure. P540A and P562A have comparable spectra to pWT $(E\text{-}G5^2)_5$ apart from a small decrease in intensity of the minima for the collagen-like region, which may suggest a change in the collagen-like motif structure. Furthermore, P562A displays a minor change in the intensity of the fluorescence emission spectrum, which may indicate a subtle change in tertiary structure (Figure 5.8C).

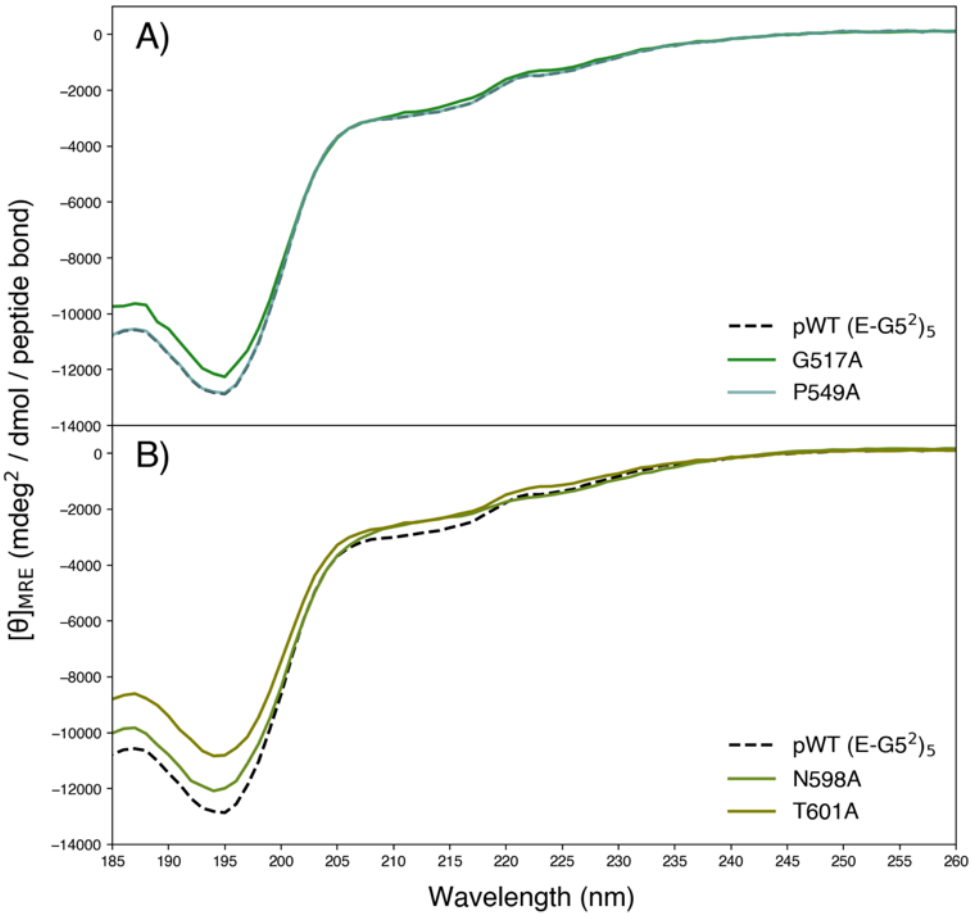

Figure 5.9 – Far-UV CD spectra of the interface variants. A) G517A and P549A.) B) N598A and T601A. These spectra show G517A and P549A have similar secondary structure to pWT $(E\text{-}G5^2)_5$. pWT $(E\text{-}G5^2)_5$ displayed in every spectrum for visualisation of secondary structure deviation.

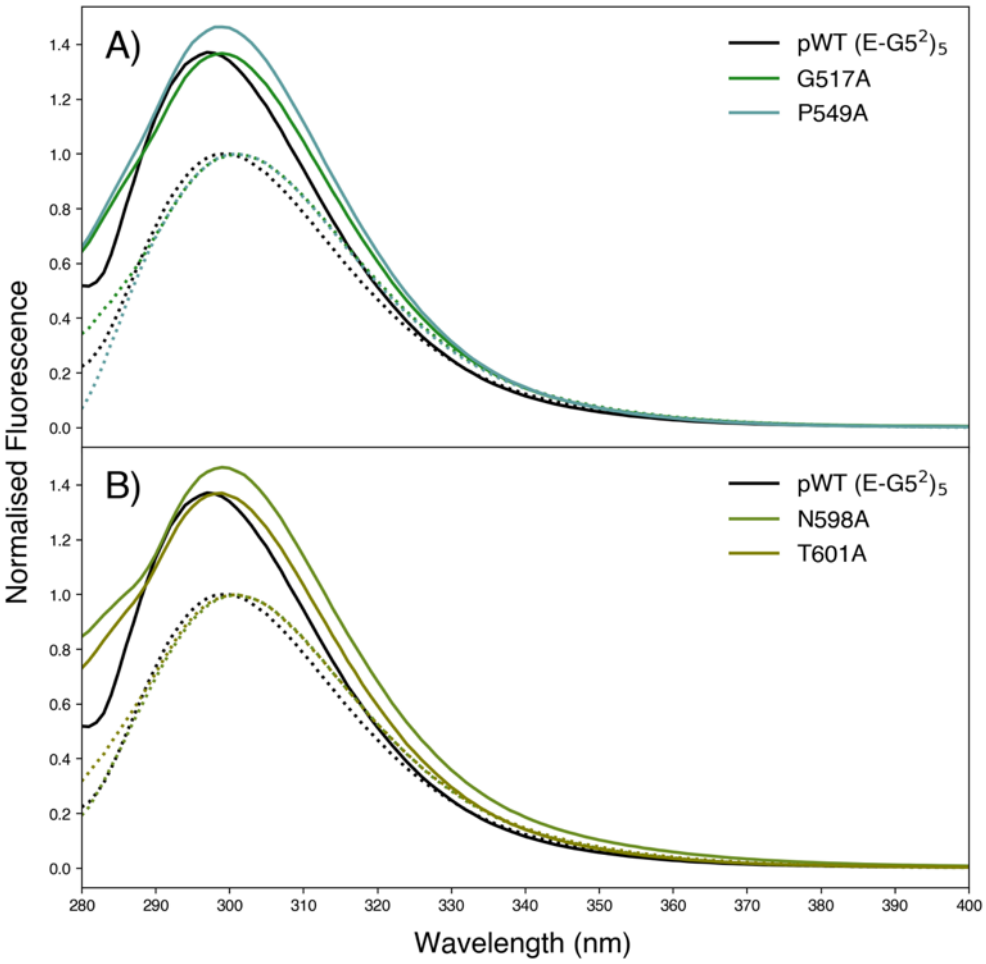

Figure 5.10 – Intrinsic tyrosine emission spectra of the interface variants. A) G517A and P549A and, B) N598A and T601A. Absence of urea (solid line) and presence of 8 M urea (dashed lines). pWT $(E\text{-}G5^2)_5$ is included for reference. All variants display tertiary structure, which is lost in the presence of chemical denaturant.

CD spectra of the interface variants are displayed in Figure 5.9. G517A and P549A display spectra similar and identical to pWT $(E\text{-}G5^2)_5$, respectively. This suggests that the secondary structures of these variants do not deviate from pWT $(E\text{-}G5^2)_5$. However, P549A displays an increase in fluorescence intensity implying a change in the tertiary structure (Figure 5.10A). N598A and T601A display similar spectra to one another, which deviates somewhat from pWT $(E\text{-}G5^2)_5$ with a slight loss of intensity of the negative minima at 210-220 nm (characteristic for β-hairpin structure)[179]. This indicates the $G5^2$ N-terminal loop at the interface may have a reduction in the structuring, with potential consequences on the compaction of the interface. Furthermore, an increase in intensity of fluorescence of the N598A variant suggests a subtle change in tertiary structure (Figure 5.10B).

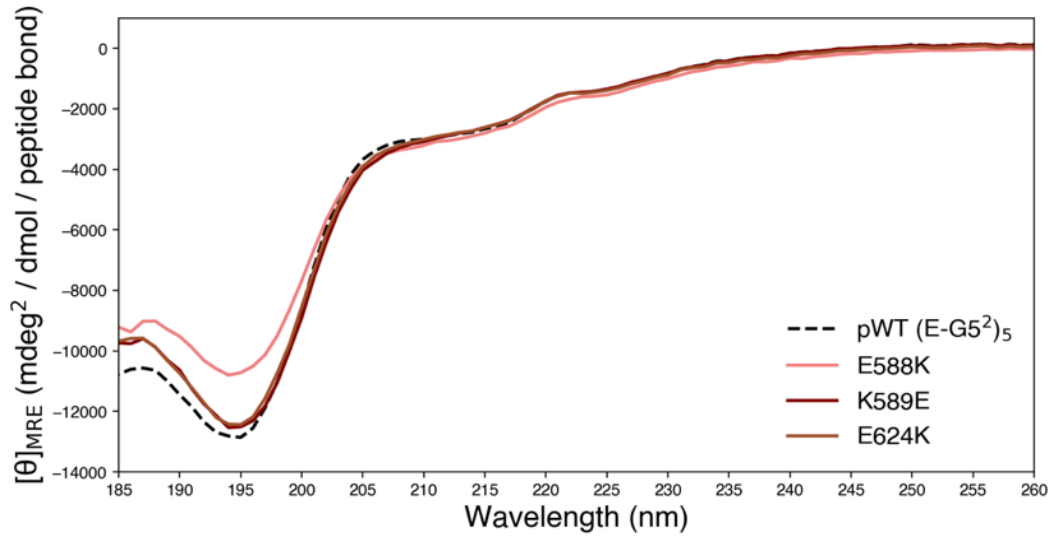

Figure 5.11 – Far-UV CD spectra of charge reverse G5² variants. E588K, K589E and E624K displaying similar shape spectra's to pWT $(E\text{-}G5^2)_5$. pWT $(E\text{-}G5^2)_5$ is displayed for visualisation of secondary structure deviation.

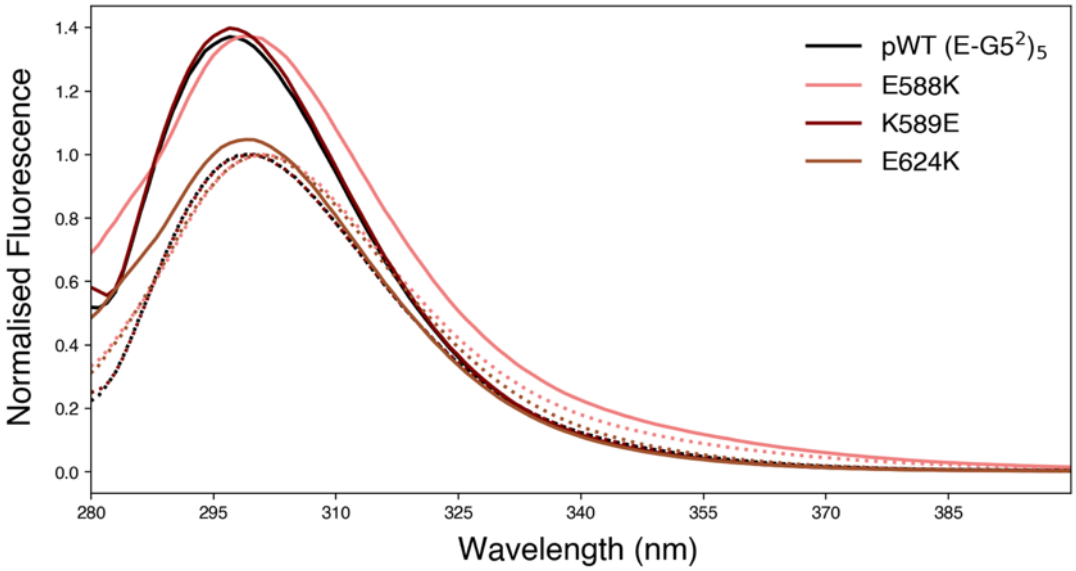

Figure 5.12 – Intrinsic tyrosine emission spectra of charge reverse G5² variants E588K, K589E and E624K. Absence of urea (solid line) and presence of 8 M urea (dashed lines). pWT $(E\text{-}G5^2)_5$ is included for reference. All variants display tertiary structure, which is lost in the presence of chemical denaturant.

Interestingly all three charge swap variants display a similar CD spectrum to pWT $(E\text{-}G5^2)_5$ (Figure 5.11), with the exception of E588K where the minimum corresponding to the triple helix of collagen is slightly decreased in intensity. This could suggest a slight change in the collagen-like motif or a concentration error. E624K reduction in fluorescence in 8 M urea is less obvious (Figure 5.12) due to steep pre- and post-transitions in the raw urea equilibrium data prior to normalising.

This comprehensive spectroscopic data analysis provides evidence that i) the variants have tertiary structure post-purification and ii) there are both pWT $(E\text{-}G5^2)_5$-like and perturbed secondary structures.

5.3.3 The Thermodynamic Stability of Variants

As with pWT $(E\text{-}G5^2)_5$ in Section 4.2.3, the variants were subjected to equilibrium denaturation analysis (Figure 5.13,Figure 5.14, Figure 5.15 and Figure 5.16). The thermodynamic stability was determined by measuring the two-state unfolding transition upon the addition of the chemical denaturant urea. A two-state transition model was fit to the data as described in Section 2.2.5.2. In all cases an excellent fit was obtained and the errors on the resulting parameters, ΔG_{UN} and m_{UN}, (Table 5.2) are the errors of the fit.

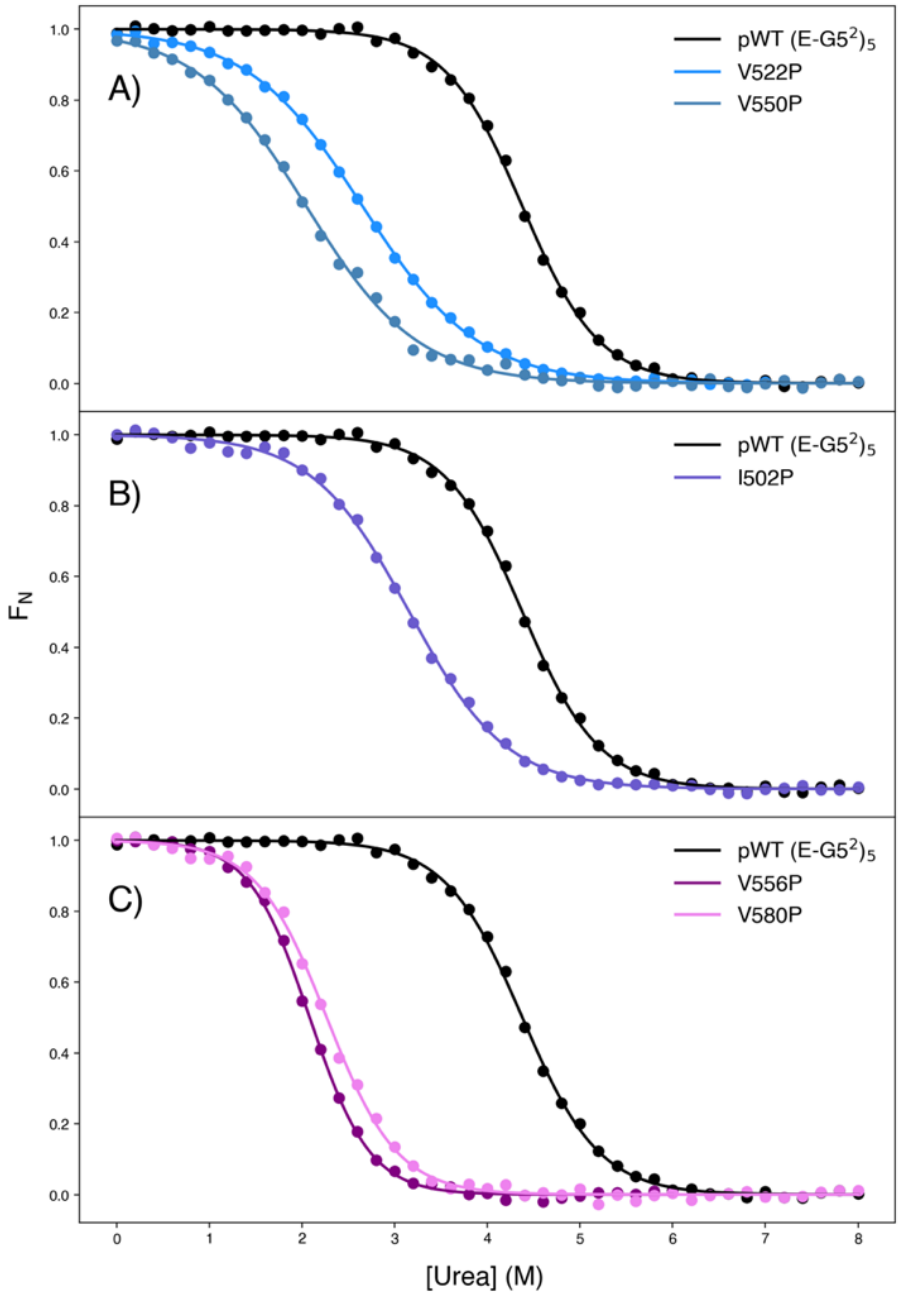

Figure 5.13 – Urea equilibrium denaturation of proposed 'mechanical clamp' mutants I502P, V522P, V550P, V556P and V580P. The data was normalised using Equation 2.6. A) I502P. B) V522P and V550P. C) V556P and V580P. Presented as a fraction of folded/unfolded protein (F_N). A two-state model was fit (line) to the data points (filled circles). pWT $(E\text{-}G5^2)_5$ data included in every plot for visual comparison of variant fit.

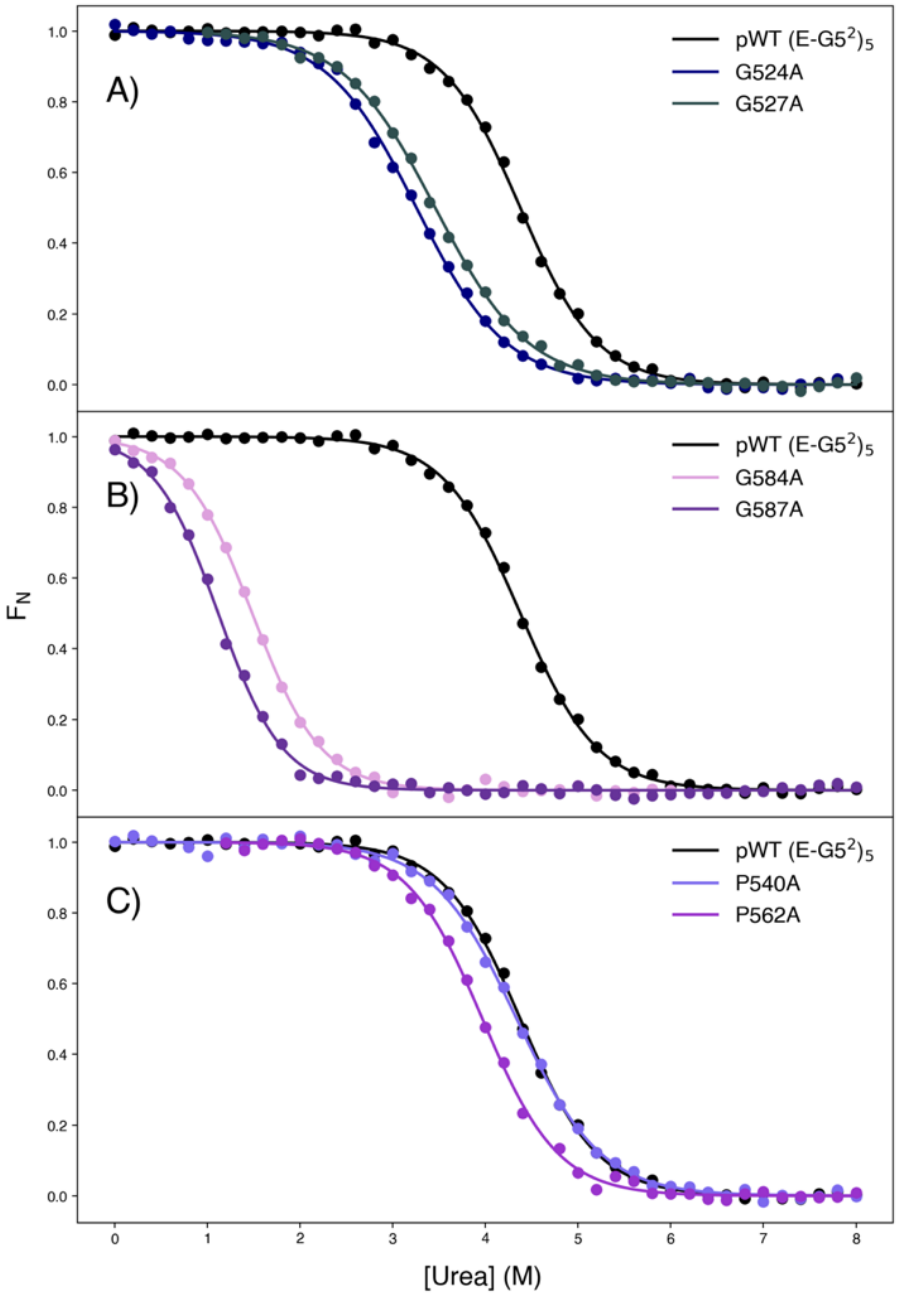

Figure 5.14 – Urea equilibrium denaturation of collagen-like motif variants G524A, G527A, G584A, G587A, P540A and P562A. The data was normalised using Equation 2.6. A) G524A and G527A situated in the E sub domain. B) G584A and G587A situated in the $G5^2$ sub domain. C) P540A and P562A situated in the E and $G5^2$ sub domains, respectively. Presented as a fraction of folded/unfolded protein (F_N). A two-state model was fit (line) to the data points (filled circles). pWT $(E\text{-}G5^2)_5$ data included in every plot for visual comparison of variant fit.

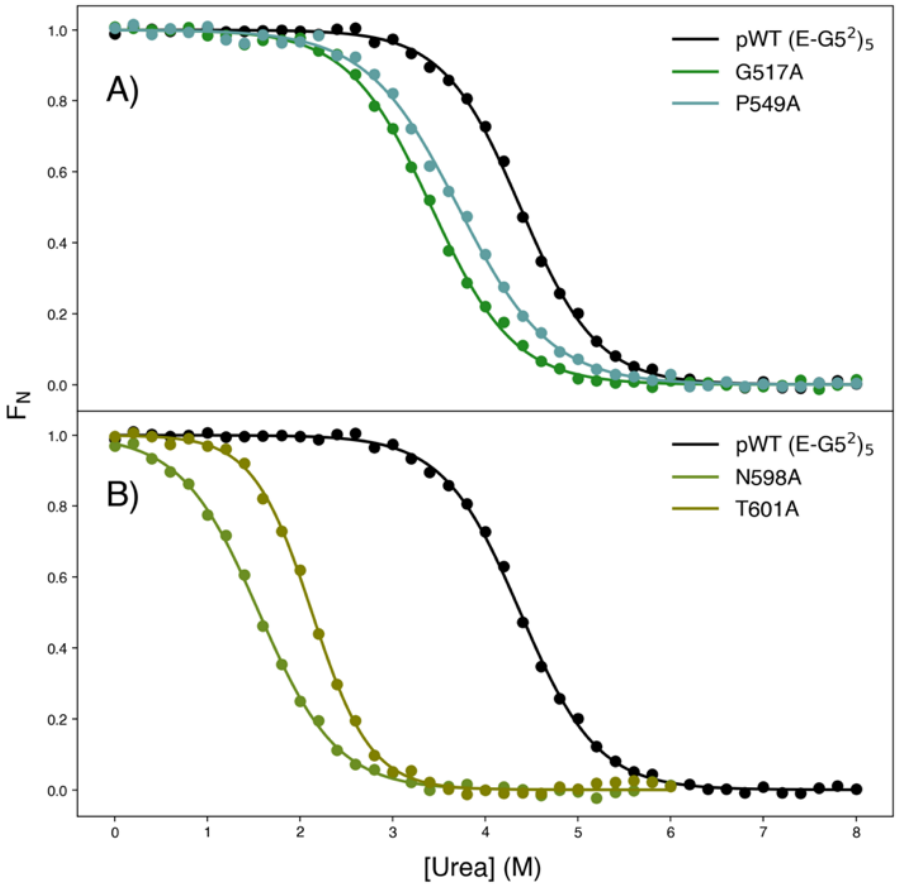

Figure 5.15 – Urea equilibrium denaturation of interface variants G517A, P549A, N598A and T601A. The data was normalised using Equation 2.6. A) G517A and P549A. B) N598A and T601A. Presented as a fraction of folded/unfolded protein (F_N). A two-state model was fit (line) to the data points (filled circles). pWT $(E\text{-}G5^2)_5$ data included in every plot for visual comparison of variant fit.

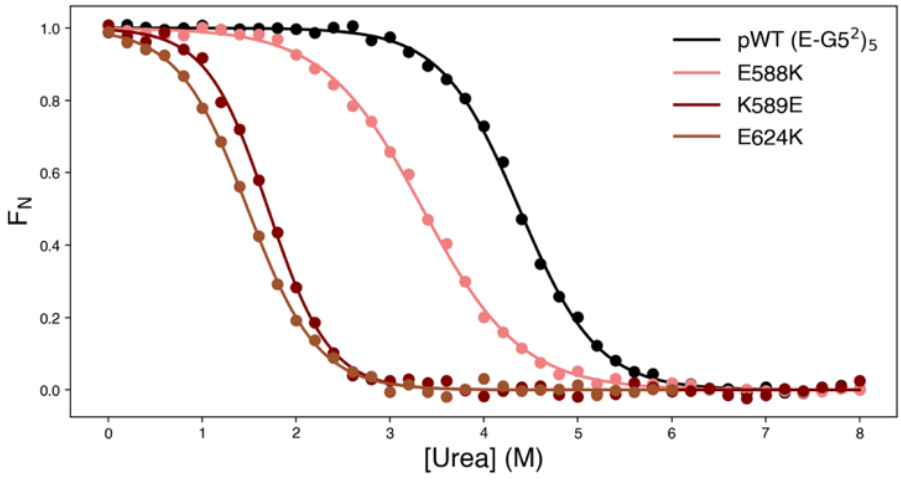

Figure 5.16 – Urea equilibrium denaturation of charge reverse variants E588K, K589E and E624K. The data was normalised using Equation 2.6. Presented as a fraction of folded/unfolded protein (F_N). A two-state model was fit (line) to the data points (filled circles). pWT $(E\text{-}G5^2)_5$ data included in every plot for visual comparison of variant fit.

Variant Group/Sub domain	Construct	ΔG_{UN} (kJ/mol)	m_{UN} (kJ mol^{-1} M^{-1})	$\Delta\Delta G_{UN}$ (kJ/mol)
-	*G5^2	11.7 ± 0.7	4.2 ± 0.2	-
-	*E-G5^2	26.2 ± 0.8	5.9 ± 0.2	-
-	pWT (E-G5^2)$_5$	26.1 ± 0.8	6.0 ± 0.2	-
MC – E	I502P	14.9 ± 0.7	4.7 ± 0.2	11.2 ± 1.1
MC – E	V522P	10.6 ± 0.4	4.0 ± 0.1	15.6 ± 0.9
MC – G5^2	V550P	8.5 ± 0.9	4.1 ± 0.3	17.7 ± 0.9
MC – G5^2	V556P	15.8 ± 0.6	7.6 ± 0.3	10.3 ± 1.0
MC – G5^2	V580P	14.7 ± 0.8	6.5 ± 0.3	11.4 ± 1.2
CLM – E	G524A	16.7 ± 0.8	5.1 ± 0.2	9.4 ± 1.2
CLM – E	G527A	17.1 ± 0.8	5.0 ± 0.2	9.0 ± 1.2
CLM – G5^2	G584A	14.7 ± 0.9	7.0 ± 0.4	11.5 ± 1.2
CLM – G5^2	G587A	8.0 ± 1.0	7.1 ± 0.4	18.2 ± 1.3
CLM – E	P540A	23.7 ± 0.9	5.5 ± 0.2	2.5 ± 1.3
CLM – G5^2	P562A	24.0 ± 1.4	6.1 ± 0.4	2.2 ± 1.6
I – E	G517A	19.0 ± 0.7	5.6 ± 0.2	7.2 ± 1.1
I – GP linker	P549A	18.6 ± 0.9	5.0 ± 0.2	7.6 ± 1.2
I – G5^2	N598A	9.2 ± 0.8	5.9 ± 0.3	17.0 ± 1.2
I – G5^2	T601A	15.8 ± 0.7	7.4 ± 0.3	10.4 ± 1.1
CR – G5^2	E588K	15.9 ± 0.7	4.7 ± 0.2	10.3 ± 1.1
CR – G5^2	K589E	12.9 ± 0.9	7.6 ± 0.4	13.2 ± 1.2
CR – G5^2	E624K	9.9 ± 0.7	6.7 ± 0.3	16.2 ± 1.1

Table 5.2 – Thermodynamic parameters of monomeric *G5^2, *E-G5^2, pWT (E-G5^2)$_5$, and variants thereof. ΔG_{UN} and m_{UN} displayed errors are the errors of the fit. $\Delta\Delta G_{UN}$ calculated from ΔG_{UN} pWT (E-G5^2)$_5$ - ΔG_{UN} variant with the propagated error. MC: 'Mechanical clamp', CLM: collagen-like motif, I: interface and CR: charge reversal. *G5^2 and E-G5^2 values taken from reference[152]. All data presented in this table was carried out in 1X PBS at 25 °C.

5.3.3.1 Discussion

All variants fit well to a two-state transition model with a single transition indicating cooperative unfolding of the E and G5^2 sub domains (for those determined to have both sub domains folded) and the resulting $\Delta\Delta G_{UN}$ values of > 1 kJ/mol for all of the variants suggests they are of a significantly

destabilised nature in comparison to pWT (E-G5^2)$_5$[206]. In addition, we observed both increased and decreased variant m_{UN} values. This is a measure of the amount of solvent accessible surface area exposed upon unfolding[207] and differences may be reporting on changes in structure of the native/denatured state[208,209] or the presence of intermediate states[210,211].

I502P, V522P, G524A, G527A, V550P, P549A and E588K display significantly decreased m_{UN} values (> 10 % difference)[206]. As the CD spectra of mechanical clamp variants (I502P, V522P and V550P), E sub domain collagen-like region variants (G524A and G527A) and E588K display a deviation from pWT (E-G5^2)$_5$ secondary structure in 0 M urea, we can assume it is a change in the structure of the native state causing the observed decrease in the m_{UN} values. For the collagen-like variants, G524A and G527A, this change in native structure may reflect twist relaxation of the collagen-like region, observable in native collagen when glycine is substituted with alanine[212,213].

For variants V522P and V550P, random coil signals in the CD spectra dominate and m_{UN} values are in excellent agreement to G5^2 in isolation[152] which together imply the E sub domain may be fully unstructured. As V550P is located at the N-termini G5^2 sub domain, it is interesting that the E sub domain is unstructured. However, this is not unexpected as Gruszka and colleagues showed the E sub domain will only fold after the interface between E and G5^2 has been formed[152]. As V550P is targeting backbone hydrogen bonding, it is likely the G5^2 N-terminal β-strand is partially unstructured (i.e. 'unlatched' from the structure) and thus unable to drive the formation of the interface. The reduced m_{UN} value of I502P agrees with the spectroscopic analysis and the FACTS implicit solvent model indicating that the E sub domain is partially unstructured in its native state.

The CD spectrum of P549A is virtually identical to pWT (E-G5^2)$_5$, indicating analogous secondary structure at 0 M urea, thus the reduced m_{UN} value could suggest an intermediate is being populated at equilibrium[210,211]. However, P549A is located in the short GP linker between E and G5^2 and the fluorescence emission indicates an altered tertiary structure, it is tempting to speculate that the reduced m_{UN} value is reporting on a change in compactness of the pseudohydrophobic core in the native structure.

Several variants found exclusively in the G5^2 sub domain, V556P, G584A, G587A, T601A, K589E and E624K, displayed increased m_{UN} values which may indicate a change in the denatured structure[208]. These variants also displayed very large $\Delta\Delta G_{UN}$ values of > 10 kJ/mol, suggesting they are significantly destabilised.

Out of 18 variants, all but 3 were found to have both their E and G5^2 sub domains fully folded by both direct and indirect methods. However, as the G5^2 sub domains of V522P and V550P and the majority of the E sub domain of I502P were thought to be folded, all constructs were analysed by SMFS.

5.3.4 SMFS Analysis of Variants

To determine the effect of amino acid substitutions on the mechanical properties of E and G5^2 sub domains, an extensive SMFS study was carried out on the 18 variants of pWT (E-G5^2)$_5$. Each variant was mechanically unfolded in 1X PBS, pH 7.4 at five retraction velocities (200, 800, 1500, 3000 and 5000 nms^{-1}) in triplicate, with the exception of variants V522P and V550P, which were mechanically unfolded at retraction velocities of 200 and 5000 nms^{-1} (ensuring the E sub domains were not unfolding in the noise at the lower speed) for one replicate only. Typical FX profiles for each variant in each group is displayed in Figure 5.17, Figure 5.18, Figure 5.19 and Figure 5.20.

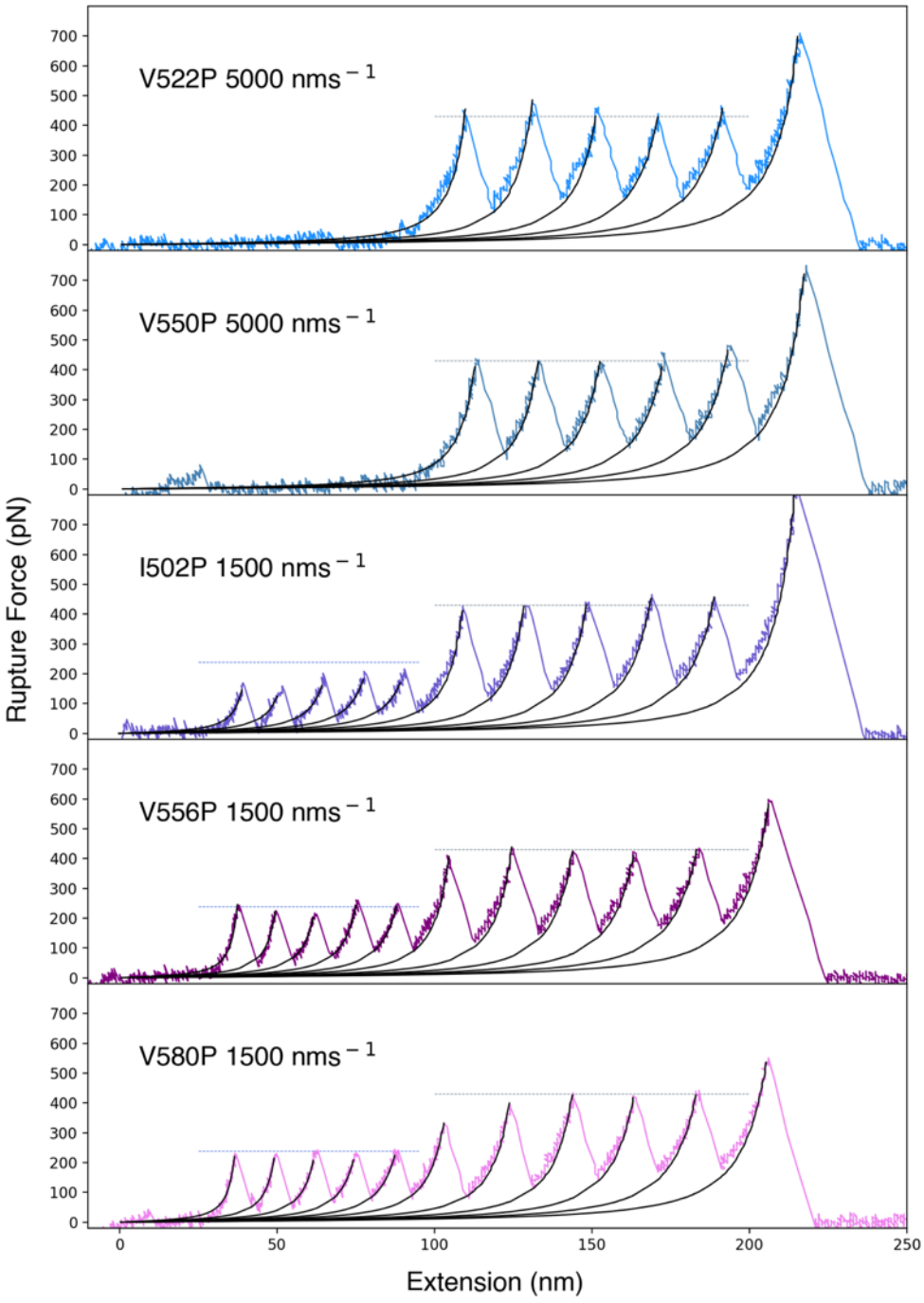

Figure 5.17 – Typical FX profiles of the 'mechanical clamp' variants V522P and V550P at a retraction velocity of 5000 nms^{-1} and I502P, V556P and V580P at a retraction velocity of 1500 nms^{-1} in 1X PBS, pH 7.4 at room temperature. WLC model fitting as black lines. pWT $(E\text{-}G5^2)_5$ average unfolding forces at a retraction velocity of 1500 nms^{-1} for E and $G5^2$ in blue and grey dotted lines, respectively. There is an absence of peaks corresponding to E sub domain unfolding for both V522P and V550P.

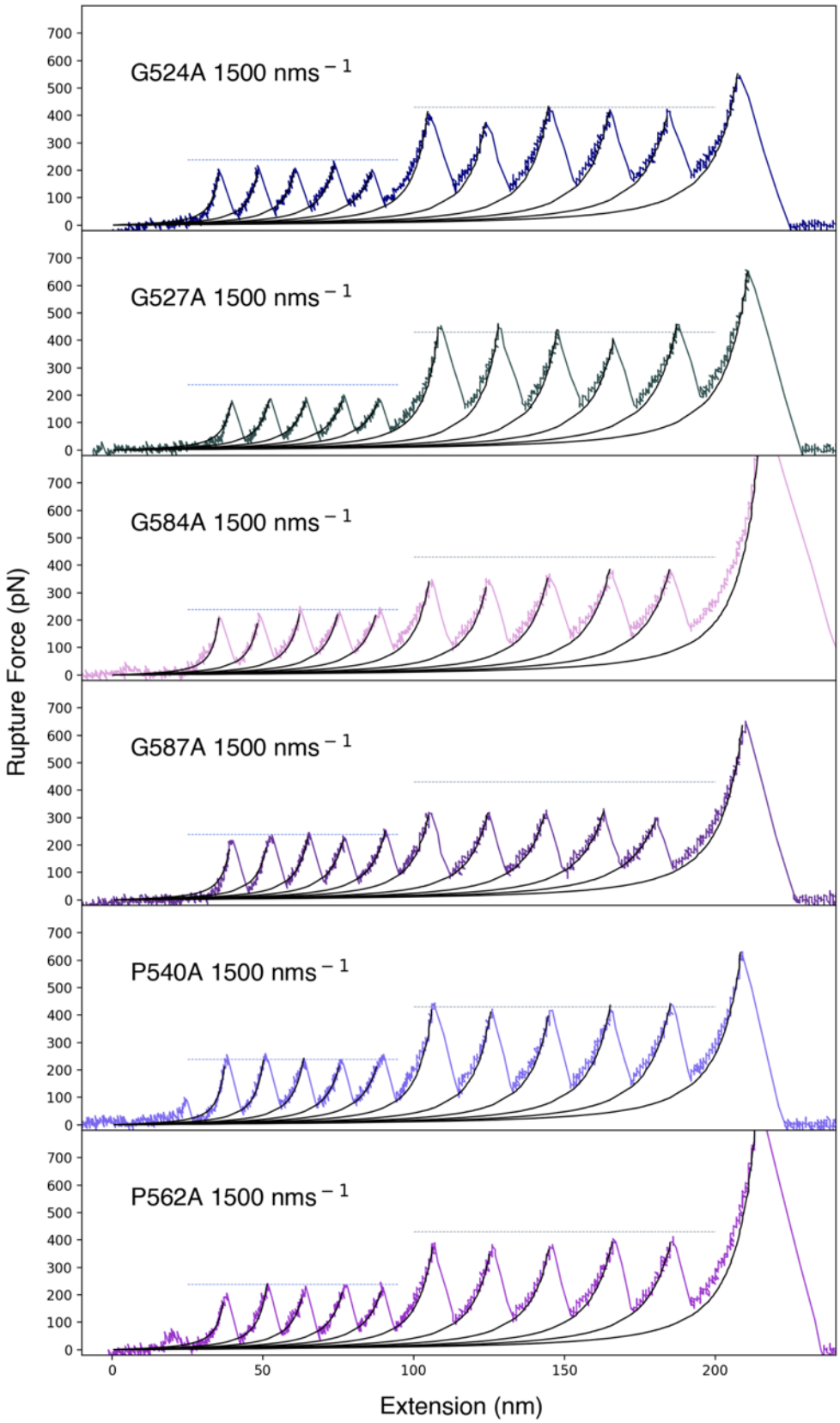

Figure 5.18 – Typical FX profiles of the collagen-like motif variants G524A, G527A, G584A, G587A, P540A and P562A in 1X PBS, pH 7.4 at a retraction velocity of 1500 nms^{-1} at room temperature. WLC model fitting as black lines. pWT $(E\text{-}G5^2)_5$ E and $G5^2$ average unfolding forces at a retraction velocity of 1500 nms^{-1} in blue and grey dotted lines, respectively.

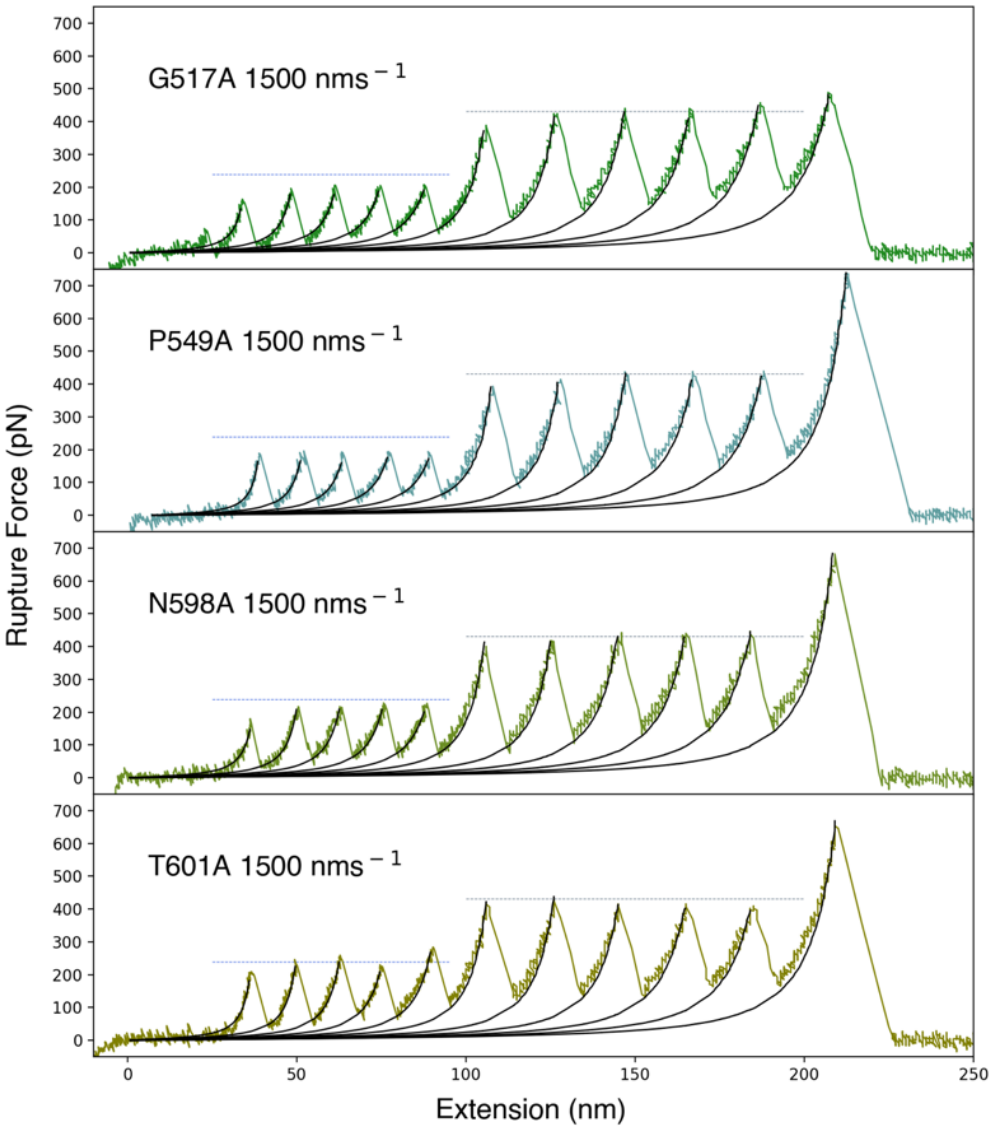

Figure 5.19 – Typical FX profiles of the interface variants G517A, P549A, N598A and T601A in 1X PBS, pH 7.4 at a retraction velocity of 1500 nms^{-1} at room temperature. WLC model fitting as black lines. pWT $(E\text{-}G5^2)_5$ E and $G5^2$ average unfolding forces at a retraction velocity of 1500 nms^{-1} in blue and grey dotted lines, respectively.

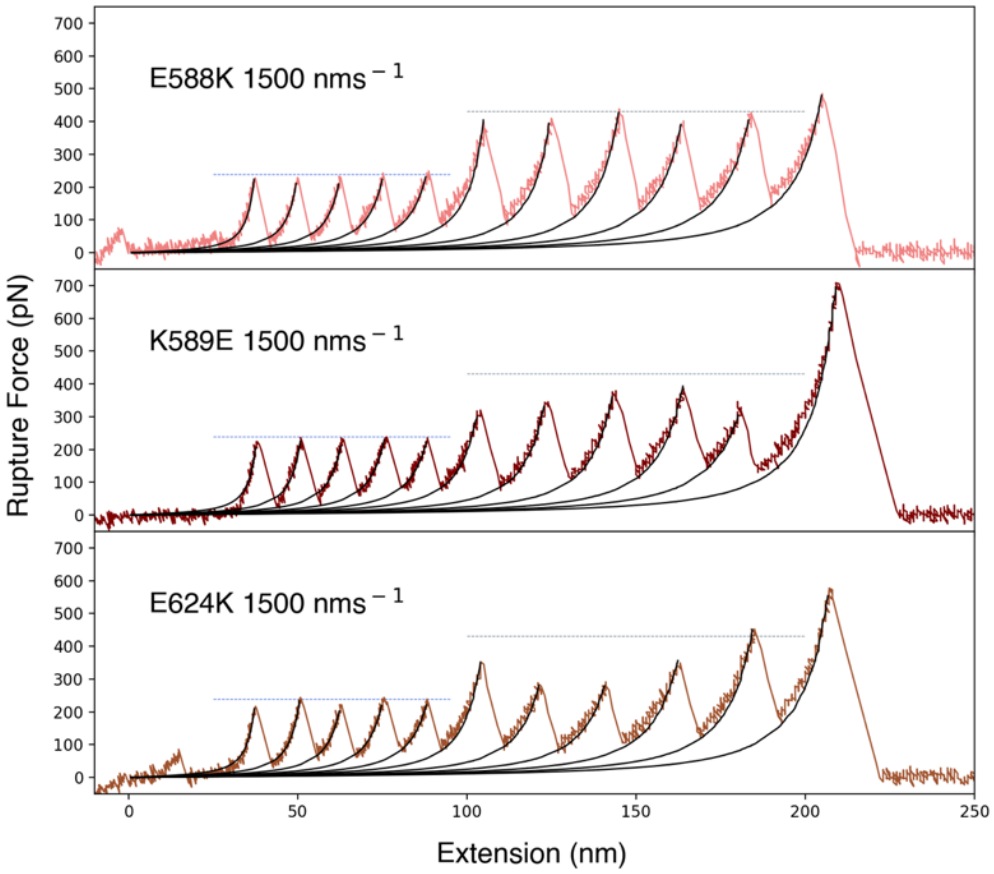

Figure 5.20 – Typical FX profiles of the charge reverse variants E588K, K589E and E624K in 1X PBS, pH 7.4 at a retraction velocity of 1500 nms^{-1} at room temperature. WLC model fitting as black lines. pWT (E-G5^2)$_5$ E and G5^2 average unfolding forces at a retraction velocity of 1500 nms^{-1} in blue and grey dotted lines, respectively.

As with pWT (E-G5^2)$_5$, FX profiles containing 4 to 10 (or 2 to 5 for V522P and V550P) unfolding peaks and a detachment peak were accepted for analysis. From these FX profiles we acquired the rupture force values from the height of the peak apices and predicted L_C values from fitting the rising edge of each peak with the WLC model with a fixed persistence length of 0.38 nm. From the latter we calculated the ΔL_C values for events and values of less than or greater than 180 Å were assigned to E and G5^2 data sets, respectively. As with pWT (E-G5^2)$_5$, histograms of the rupture force values (force-frequency) and ΔL_C (ΔL_C-frequency) were plotted with bin sizes of 10 pN and 5 Å, respectively. Gaussian fits to these provided the modal value and FWHM values, with the latter offering an insight into potential changes of the underlying energy landscape of unfolding[214]. The average of triplicate data provided us with robust rupture force values and ΔL_C values for both the E and G5^2 sub domains, with the latter values displayed in Table 5.3.

Variant Group/Sub domain	Construct	E ΔL_C Range (Å)	G5² ΔL_C Range or modal value** (Å)
MC – E	I502P	140.5-149.5	212.7-216.0
MC – E	V522P*	-	216.9**
MC – G5²	V550P*	-	215.4**
MC – G5²	V556P	143.8-164.1	209.5-219.0
MC – G5²	V580P	144.5-148.4	213.4-217.3
CLM – E	G524A	142.6-150.3	213.3-217.4
CLM – E	G527A	143.3-150.5	212.6-217.7
CLM – G5²	G584A	143.5-148.5	213.6-217.4
CLM – G5²	G587A	140.7-147.9	210.9-217.0
CLM – E	P540A	145.0-153.4	213.4-217.2
CLM – G5²	P562A	143.0-151.5	211.9-216.9
I – E	G517A	142.6-150.6	213.9-216.0
I – E	P549A	145.0-154.4	213.9-229.9
I – G5²	N598A	141.7-150.4	213.0-216.1
I – G5²	T601A	144.5-150.4	211.5-219.1
CR – G5²	E588K	144.0-158.9	212.4-219.9
CR – G5²	K589E	143.6-147.6	210.0-216.8
CR – G5²	E624K	143.0-145.9	214.2-218.5

Table 5.3 – Table of ΔL_C values for each variant mechanically unfolded in 1X PBS, pH 7.4 at room temperature. ΔL_C values are the range of the Gaussian fit modal values of the triplicate repeats, with the exception of V522P and V550P. *V522P and V550P modal values at 5000 nms^{-1}, and the corresponding G5² ΔL_C** is the modal value from the single repeat Gaussian fits. All ΔL_C values are consistent with the pWT $(E\text{-}G5^2)_5$ corresponding domain. MC: 'Mechanical clamp', CLM: collagen-like motif, I: interface and CR: charge reversal.

These values are consistent with pWT $(E\text{-}G5^2)_5$ E and G5² sub domains ΔL_C values of 143.5-151.1 and 210.7-216.7 Å, respectively. This suggests the pathway to unfolding is coarsely the same as a similar amount of structure is being 'released' after the mechanical TS for all variants. Unfolding of the full polyprotein construct, with the maximum number of unfolding peaks for each variant observed at every retraction velocity for every variant suggested the variants were folded. This is with the exception of V522P and V550P where only large peaks corresponding to G5² sub domains unfolding were present.

In the following sections, the SMFS data will be broken down into the variant groups, where an example scatterplot-histogram will be displayed for each variant alongside the speed dependence of rupture force. The latter is the average rupture forces of both E and G5² sub domains plotted against the logarithm of retraction velocity at five retraction velocities and provides an insight into any deviation in the underlying energy landscape for mechanical protein unfolding.

In these analyses a significant change in the mechanical properties of a variant is first determined by performing an ANCOVA ($p < 0.05$) on the linear regression (fit to the speed dependence data) and any decrease in mechanical strength of ≥ 20 pN is noted. A 20 pN change in mechanical unfolding force was chosen as it gives a value above the positional noise limit from cantilevers utilised (~ 15 pN)[44] due to thermally induced cantilever oscillations. A significant change in the gradient can be confirmed by an unpaired t-test using the t-statistic ($p < 0.05$).

Only variants and their respective sub domains with significantly different changes in the mechanical strength or the speed dependence gradient are discussed from hereinafter. Force responses of variants/sub domains not mentioned are to be assumed indifferent from pWT $(E\text{-}G5^2)_5$ and will not be discussed in detail unless necessary. All mechanical unfolding data and example plots can be found in the Appendix in Section 7.3.3 and Section 7.3.4, respectively.

5.3.4.1 'Mechanical Clamp' Variants

Variants I502P, V522P, V550P, V556P and V580P (Figure 5.21) were utilised in an attempt to disrupt the long stretches of main chain hydrogen bonds forming the 'mechanical clamps' of the E and $G5^2$ sub domains predicted by MD simulations[74]. An example scatterplot-histogram of rupture force and ΔL_C data for variants I502P, V556P and V580P and E sub domain non-folders V522P and V550P at a retraction velocity of 1500 nms^{-1} and 5000 nms^{-1}, respectively, are displayed in Figure 5.22.

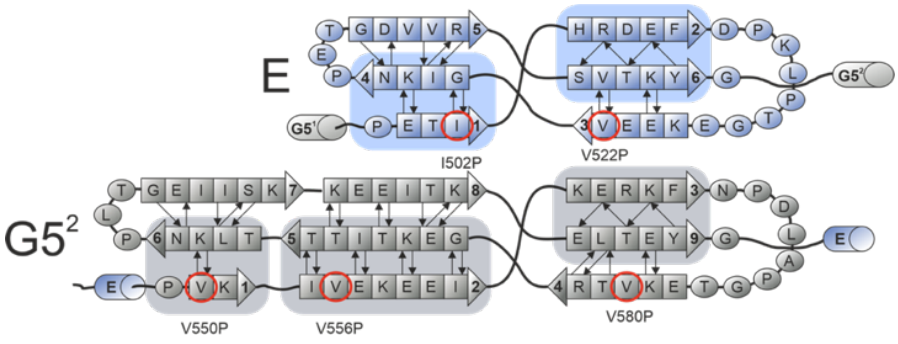

Figure 5.21 – Schematic of the secondary structure of E and $G5^2$ outlining the variants created for proline scanning mutagenesis of the 'mechanical clamps'. 'Mechanical clamps' highlighted in blue and grey for the E and $G5^2$ sub domain, respectively. The proline substitutions are circled in red and labelled.

As expected, Figure 5.22A and Figure 5.22B display a unimodal distribution in both rupture force and ΔL_C values for both variants at a retraction velocity of 5000 nms^{-1}, suggesting only one sub domain is unfolding. The ΔL_C values for both variants were > 180 Å, which indicated we were observing exclusively $G5^2$ sub domain unfolding and the resulting rupture forces were indifferent from pWT $(E\text{-}G5^2)_5$. The SMFS summary for $G5^2$ sub domain unfolding of V522P and V550P is outlined in the Appendix (Section 7.3.3). This complements the spectroscopic and equilibrium analyses that the E domains are unfolded in the native structure of both V522P and V550P and in turn suggests they are not simply unfolding in the noise[51].

Example unfolding statistics for the E and $G5^2$ sub domains for one variant (I502P) are displayed in Table 5.4 (the remaining variants unfolding statistics are found in the appendix – Section 7.3.3).

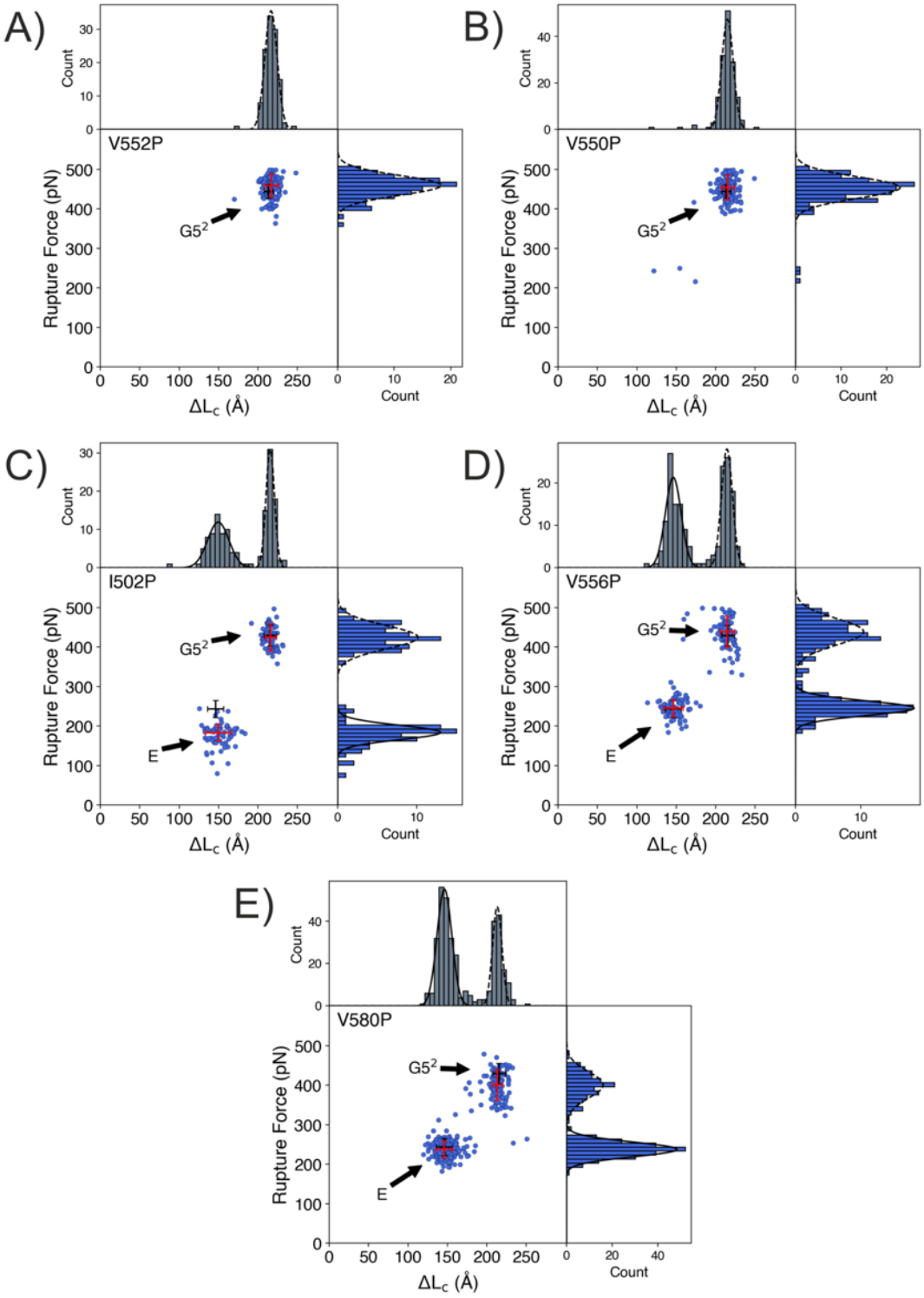

Figure 5.22 – Example rupture force-ΔL_C scatterplot with associated histograms of the 'mechanical clamp' proline variants. A) V522P and B) V550P replicate at a retraction velocity of 5000 nms^{-1}. C) I502P, D) V556P and E) V580P replicates at a retraction velocity of 1500 nms^{-1}. ΔL_C histograms in grey and rupture force histograms in blue with the Gaussian distribution fit as a solid and dashed line for E and $G5^2$ sub domains, respectively. This data displays two distinct populations for both ΔL_C and rupture force indicating two separate sub domains unfolding with different rupture forces with the exception of V522P and V550P, where only one population is present. Red scatterplot crosshairs are the mode and the FWHM from the corresponding marginal histogram Gaussian distribution fits. pWT $(E\text{-}G5^2)_5$ cross hairs from one repeat at the corresponding retraction velocity displayed in black for reference.

Speed (nms^{-1})	I502P E n	Mode Rupture Force (pN)	Average (pN)	$G5^2$ n	Mode Rupture Force (pN)	Average (pN)
200	36	171.9	**174.9**	68	396.0	**406.7**
	17	182.5		20	411.1	
	30	170.4		34	413.1	
800	45	179.8	**179.8**	58	423.1	**424.2**
	79	178.3		67	423.9	
	103	181.3		46	425.6	
1500	66	184.0	**188.2**	73	425.0	**428.3**
	86	187.3		70	433.9	
	131	193.4		41	426.0	
3000	60	192.9	**196.9**	75	447.4	**446.6**
	76	198.1		62	451.3	
	62	199.7		20	441.0	
5000	77	213.2	**207.0**	79	457.2	**458.7**
	88	195.6		44	451.5	
	50	212.2		27	467.4	

Table 5.4 – Summary of rupture force statistics for I502P E and $G5^2$ sub domain mechanical unfolding in 1X PBS, pH 7.4 at room temperature. n is the number of peaks utilised for analysis. Mode rupture force is obtained from the Gaussian fits to the histograms. Average is the mean of the mode values at each speed.

As observed in Table 5.3, the ΔL_C of the E domain remains unchanged from pWT $(E\text{-}G5^2)_5$ indicating the E sub domain structure is similar prior to global unfolding. However, the spectroscopic, equilibrium analyses and implicit solvent model suggests that there is partial unstructuring of the E sub domain which implies this force data may be reporting on an 'unlatched mechanical clamp' to some extent.

The speed dependence of rupture force for both the E and $G5^2$ sub domains of variants I502P, V556P and V580P are displayed in Figure 5.23. The data shows that the I502P E sub domain rupture force is significantly reduced at all retraction velocities, with a mechanical destabilisation of -50.3 $\pm$ 7.0 pN at a retraction velocity of 1500 nms^{-1}. For V580P, the $G5^2$ sub domain displays a deviation in gradient (Figure 5.23B) coupled with a substantial increase in the FWHM values of the Gaussian fittings to the rupture force histograms in comparison to pWT $(E\text{-}G5^2)_5$ (Appendix – Section 7.3.4).

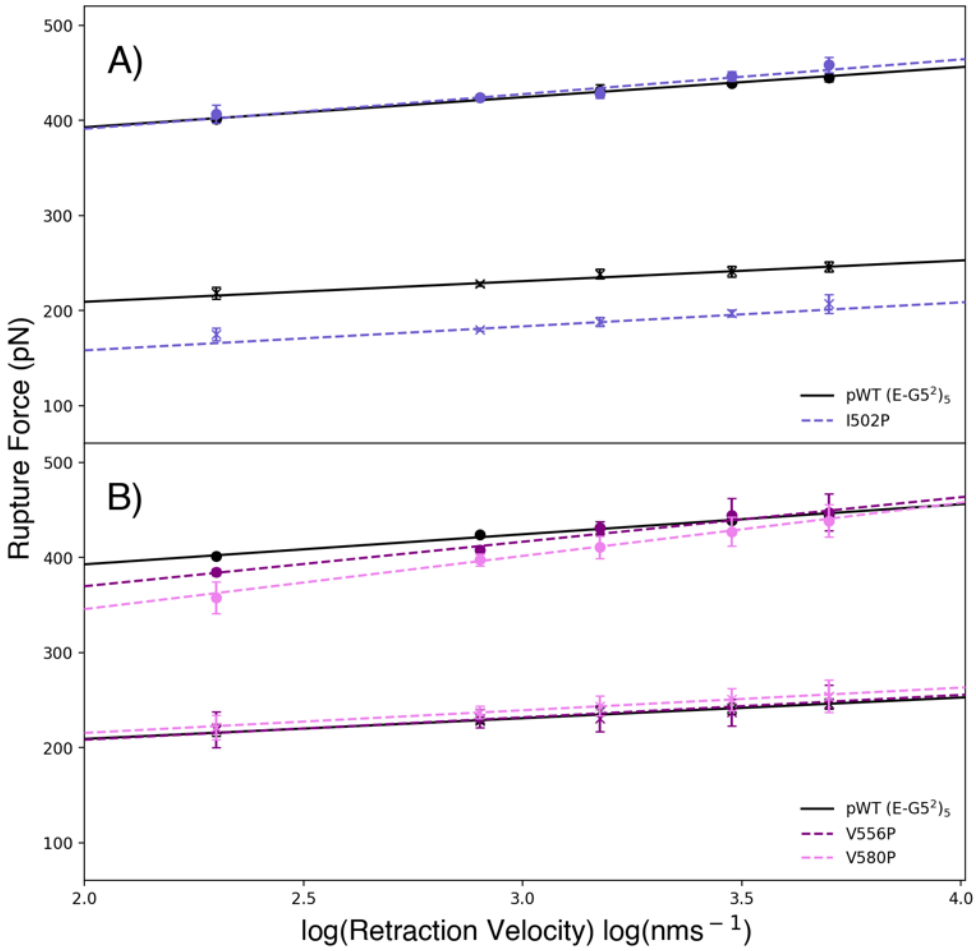

Figure 5.23 – The dependence of rupture force on the logarithm of retraction velocity of the 'mechanical clamp' variants. E (crosses) and G5^2 (filled circles) sub domains of A) I502P and B) V556P and V580P variants. Points and errors are the mean and standard deviation of triplicate datasets. Linear fit is weighted with the inverse of the standard deviation error of the triplicate datasets. There is an evident decrease in mechanical strength of the I502P E sub domain and a change in the gradient of the V580P G5^2 sub domain.

5.3.4.1.1 'Mechanical Clamp' Variant Discussion

β-sheet proline inclusion prevented folding of 3/5 variants and thus precluded the analysis of the force-bearing properties of specific main chain hydrogen bonds of these variants. However, we were still able to extract information from the partially folded I502P E sub domain. The FACTS model predicted the whole N-terminal β-sheet to be unstructured. However, the ΔL_C is within agreement of pWT (E-G5^2)$_5$ suggesting there may be some N-terminal structuring and that our data may be reporting on slight 'unlatching' of the outer β-strand of the N-terminal 'mechanical clamp'. This resulted in a large mechanical destabilisation of the E sub domain (~ 25 %) with a similar

distance to the mechanical TS (x_u) as pWT (E-G5^2)$_5$. This suggests that the interactions of the N-terminal 'mechanical clamp' are of importance to the mechanical strength of the E sub domain.

Valine to proline mutations in mechanical proteins, such as I27, have shown a typically more shallow gradient and thus an increase in distance between the native state and transition state[90]. However, V580P G5^2 displayed a steeper gradient than pWT (E-G5^2)$_5$ G5^2 and a substantial increase in the majority of the rupture force Gaussian fit FWHM values across several retraction velocities, which implies a decrease in the distance to the mechanical TS (x_u) as both the gradient and FWHM are inversely proportional to x_u[44,215]. This suggests the G5^2 sub domain undergoes less physical deformation prior to global unfolding and is more commonly seen with lower temperatures[215]. The unchanged mechanical properties of V556P suggests we have not perturbed any interactions.

5.3.4.2 Collagen-like Motif Variants

The collagen-like motif variants include G524A, G527A, G584A, G587A, P540A and P562A. G524A, G527A and P540A are found in the E sub domain collagen-like motif and G584A, G587A and P562A are found in the G5^2 sub domain collagen-like motif (Figure 5.24). These substitutions were introduced in an attempt to affect the packing of the collagen-like helices. An example scatterplot-histogram of rupture force and ΔL_C data for each variant at a retraction velocity of 1500 nms^{-1} are displayed in Figure 5.25.

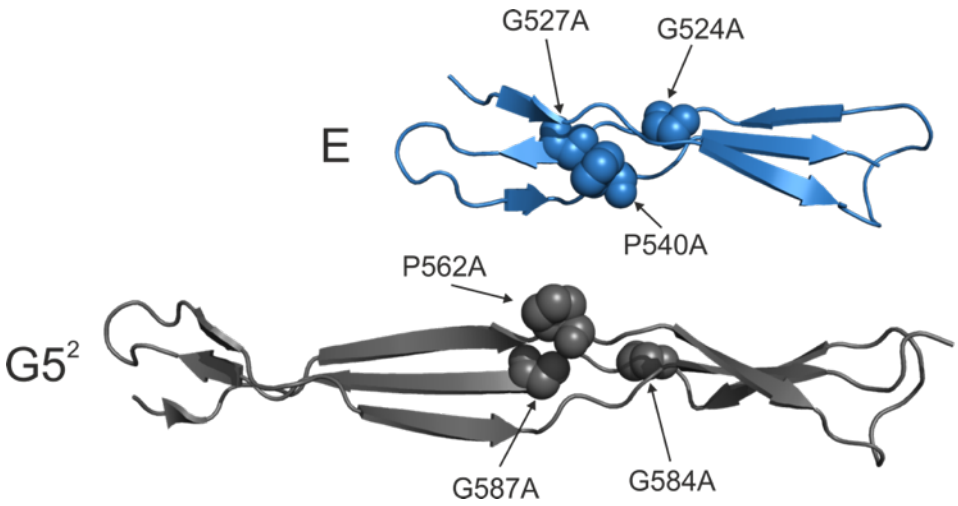

Figure 5.24 – Crystal structure (PDB: 3TIP) annotated with residue spheres to highlight the location of collagen-like region variants discussed in this section. E and G5^2 structures in blue and grey, respectively.

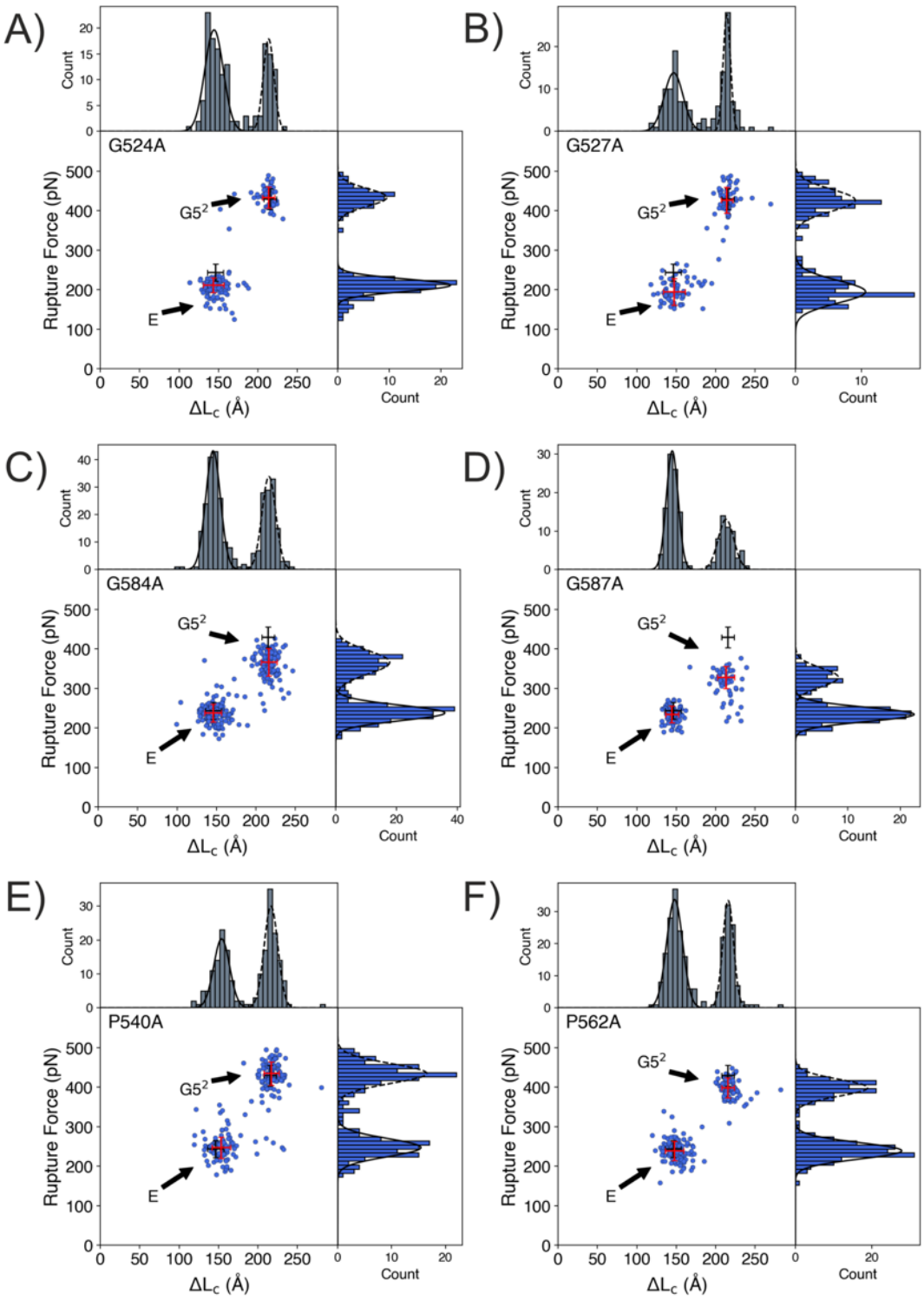

Figure 5.25 – Example rupture force-ΔL_C scatterplot with associated histograms of the collagen-like motif variants. A) G524A, B) G527A, C) G584A, D) G587A, E) P540A and F) P562A replicate at a retraction velocity of 1500 nms^{-1}. ΔL_C histograms in grey and rupture force histograms in blue with the Gaussian distribution fit as a solid and dashed line for E and $G5^2$ sub domains, respectively. This data displays two distinct populations for both ΔL_C and rupture force indicating two separate sub domains unfolding with different rupture forces. Red scatterplot crosshairs are the mode and the FWHM from the corresponding marginal histogram Gaussian distribution fits. pWT $(E\text{-}G5^2)_5$ cross hairs from one repeat at the corresponding retraction velocity displayed in black for reference.

The speed dependence of average rupture force for all collagen-like motif variants is displayed in Figure 5.26.

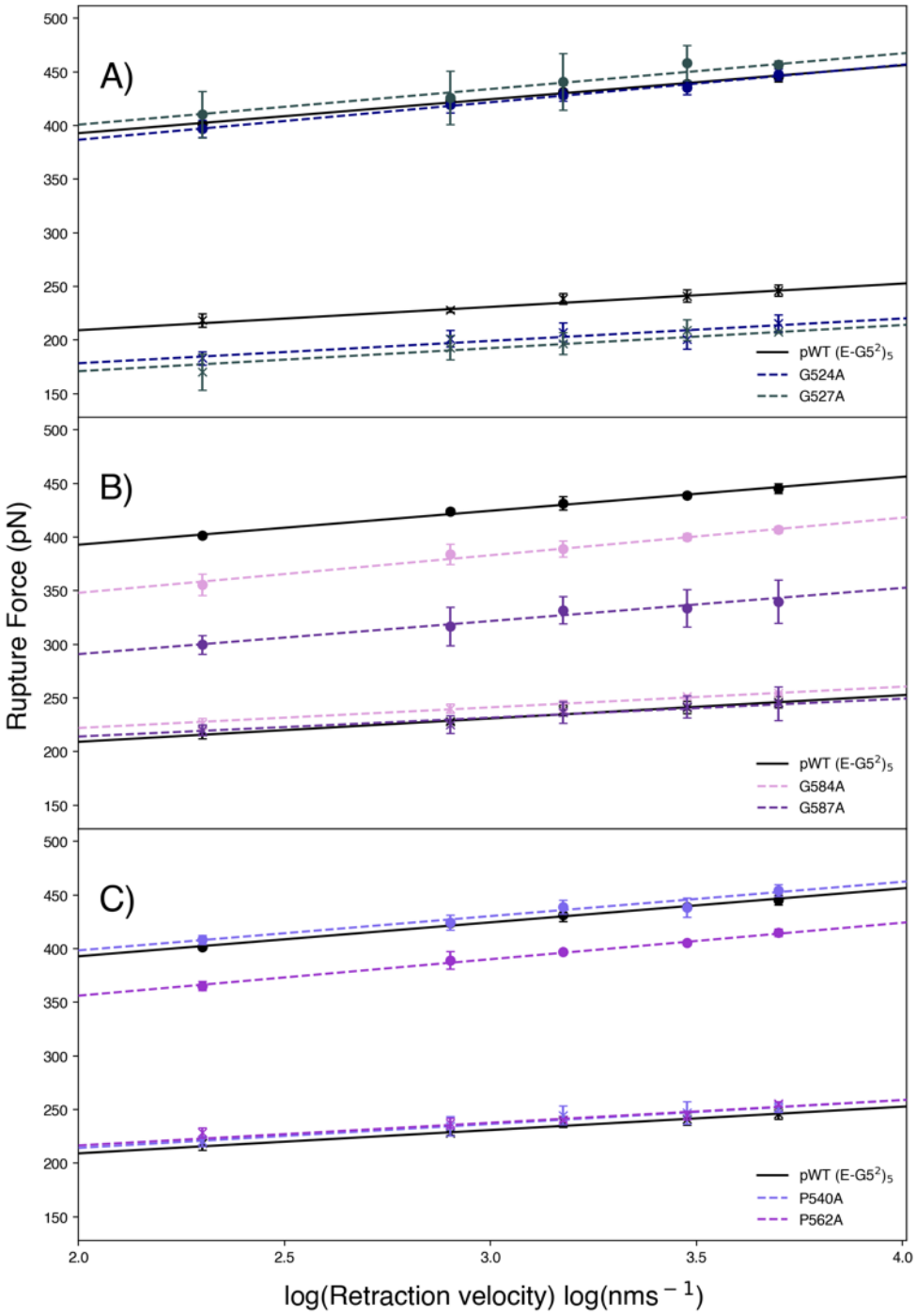

Figure 5.26 – The dependence of rupture force on the logarithm of retraction velocity of the collagen-like motif variants. E (crosses) and G5^2 (filled circles) sub domains of A) G524A and G527A, B) G584A and G587A and C) P540A and P562A variants. Points and errors are the mean and standard deviation of triplicate datasets. Linear fit is weighted with the inverse of the standard deviation error of the triplicate datasets. There is an appreciable decrease in the E sub domain of G524A and G527A and in the G5^2 sub domain of G584A, G587A and P562A.

For the variants with the substitution located in the E sub domain (G524A and G527A) we observed a significant decrease in the mechanical strength of the E sub domain across all retraction velocities, whereas $G5^2$ remained unchanged (Figure 5.26A). Conversely, variants with the mutation located in the $G5^2$ sub domain (G584A, G587A and P562A) displayed a pWT $(E\text{-}G5^2)_5$-like mechanical response of their E sub domains, but a significant decrease in the mechanical response of the $G5^2$ sub domain across all retraction velocities (Figure 5.26B & C). Both sub domains of P540A and all variant gradients remained indifferent to pWT $(E\text{-}G5^2)_5$.

5.3.4.2.1 Collagen-like Motif Variant Discussion

A summary of the force difference ($Force_{MUT} - Force_{pWT}$) for the collagen-like motif variants at 1500 nms^{-1} (Figure 5.27) displays a motif which has been highly mechanically perturbed through alanine scanning mutagenesis.

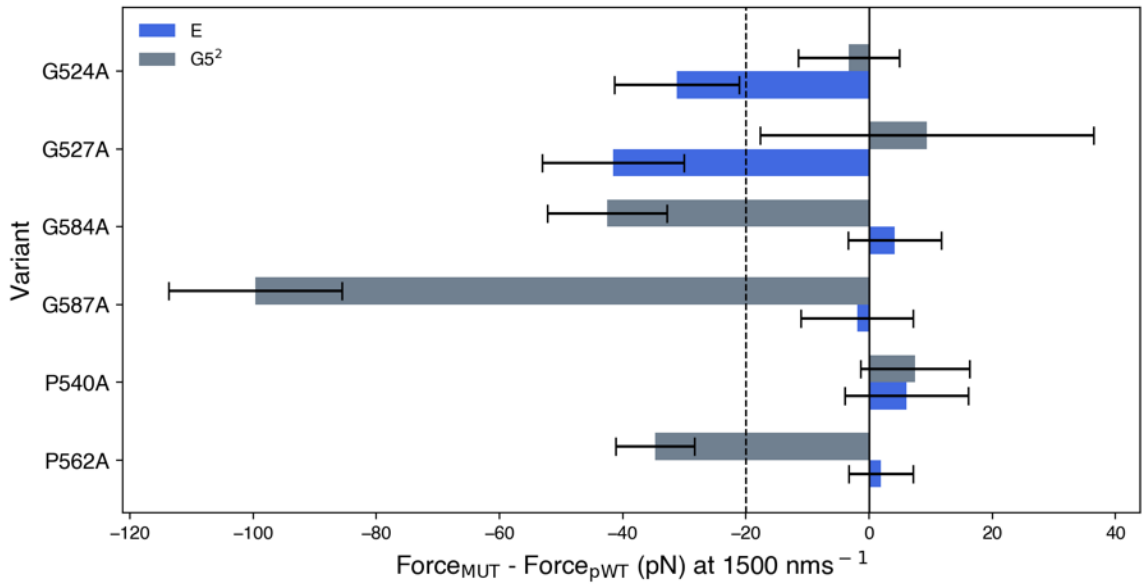

Figure 5.27 – $Force_{MUT} - Force_{pWT}$ of collagen-like motif variants at a retraction velocity of 1500 nms^{-1} to visualise the change in mechanical strength of the sub domains. Black broken line represents a loss of 20 pN in the mechanical strength from pWT $(E\text{-}G5^2)_5$. There is a significant loss of mechanical strength in the E sub domain of G524A and G527A and $G5^2$ sub domain of G584A, G587A and P562A. Errors are the propagated SD (of the triplicate repeats).

The general trend observed is the collagen-like motif variants have a reduced mechanical strength in the sub domain of substitution location with an absence of change in the protein malleability. All of the glycine to alanine variants display a significant decrease in mechanical force of the sub

domain of the substitution location. G524A and G527A display a difference of -31.3 ± 10.1 and -41.6 ± 11.5 pN, respectively, relative to the pWT $(E\text{-}G5^2)_5$ E sub domain at a retraction velocity of 1500 nms^{-1}. Whereas, when G584A and G587A are pulled at 1500 nms^{-1}, a mechanical unfolding force difference in the $G5^2$ sub domain of -42.6 ± 9.7 and -99.7 ± 14.0 pN was observed, respectively.

These large decreases in mechanical strength of the glycine to alanine mutations are mostly coupled with subtle changes in the secondary structure as determined by CD, which may suggest structural perturbation has resulted in reduced force-bearing properties.

Only one out of two of the proline to alanine variants, P562A, displayed an altered mechanical phenotype. When pulled at 1500 nms^{-1} a significant difference of -34.7 ± 6.4 pN in comparison to the pWT $(E\text{-}G5^2)_5$ $G5^2$ rupture force was observed. P562 and P540 are located on structurally distinct strands of the E and $G5^2$ collagen-like regions, respectively (Figure 5.28). The CD spectrum and m_{UN} value of P562A is comparable to pWT $(E\text{-}G5^2)_5$. However, a small change in fluorescence emission intensity for P562A in comparison to pWT $(E\text{-}G5^2)_5$, absent in P540A, indicates there may be a subtle change in the tertiary structure, likely including the collagen-like region. Or it may suggest that force is propagated through the middle strand of the collagen-like region.

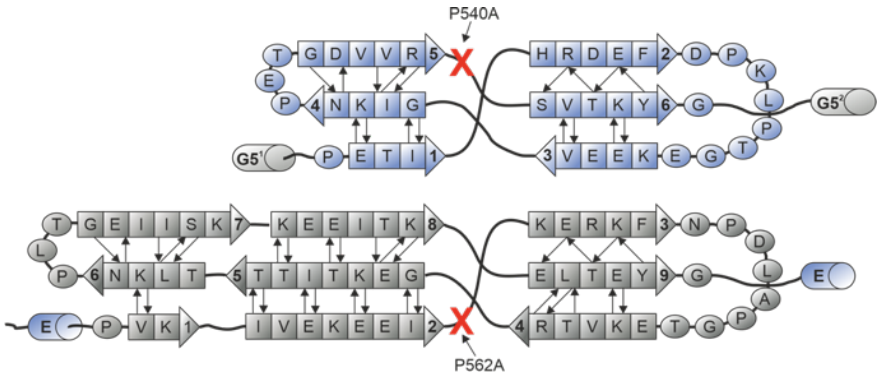

Figure 5.28 – Location of P540A and P562A collagen-like mutations on a schematic of $E\text{-}G5^2$. P540A and P562A are found on non-corresponding strands in the collagen-like regions of E and $G5^2$, respectively. P562A is located on the cross-over strand, whereas P540A is not.

These results denote the collagen-like motif as a region of mechanical interest and will be discussed further on.

5.3.4.3 Interface Variants

The interface variants included alanine substitutions at positions G517, P549, N598 and T601 (Figure 5.29) in an attempt to directly/indirectly affect the contacts of the pseudohydrophobic core (formed due to the contribution of bulky aromatics and longer hydrophobic chains at the interdomain interface). An example scatterplot-histogram of rupture force and ΔL_C data for each variant at a retraction velocity of 1500 nms^{-1} is displayed in Figure 5.30.

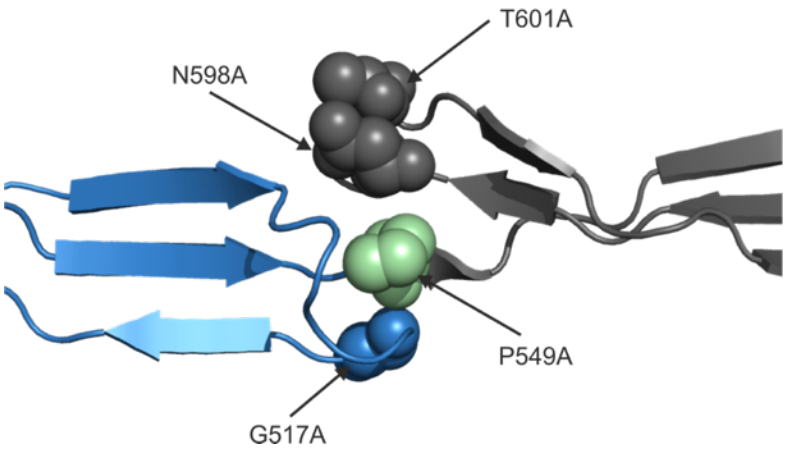

Figure 5.29 – Crystal structure annotated with residue spheres to highlight the location of interface variants discussed in this section. E and $G5^2$ structures in blue and grey, respectively. P549A is displayed in green due to its position in the short GP linker between the E and $G5^2$ sub domains.

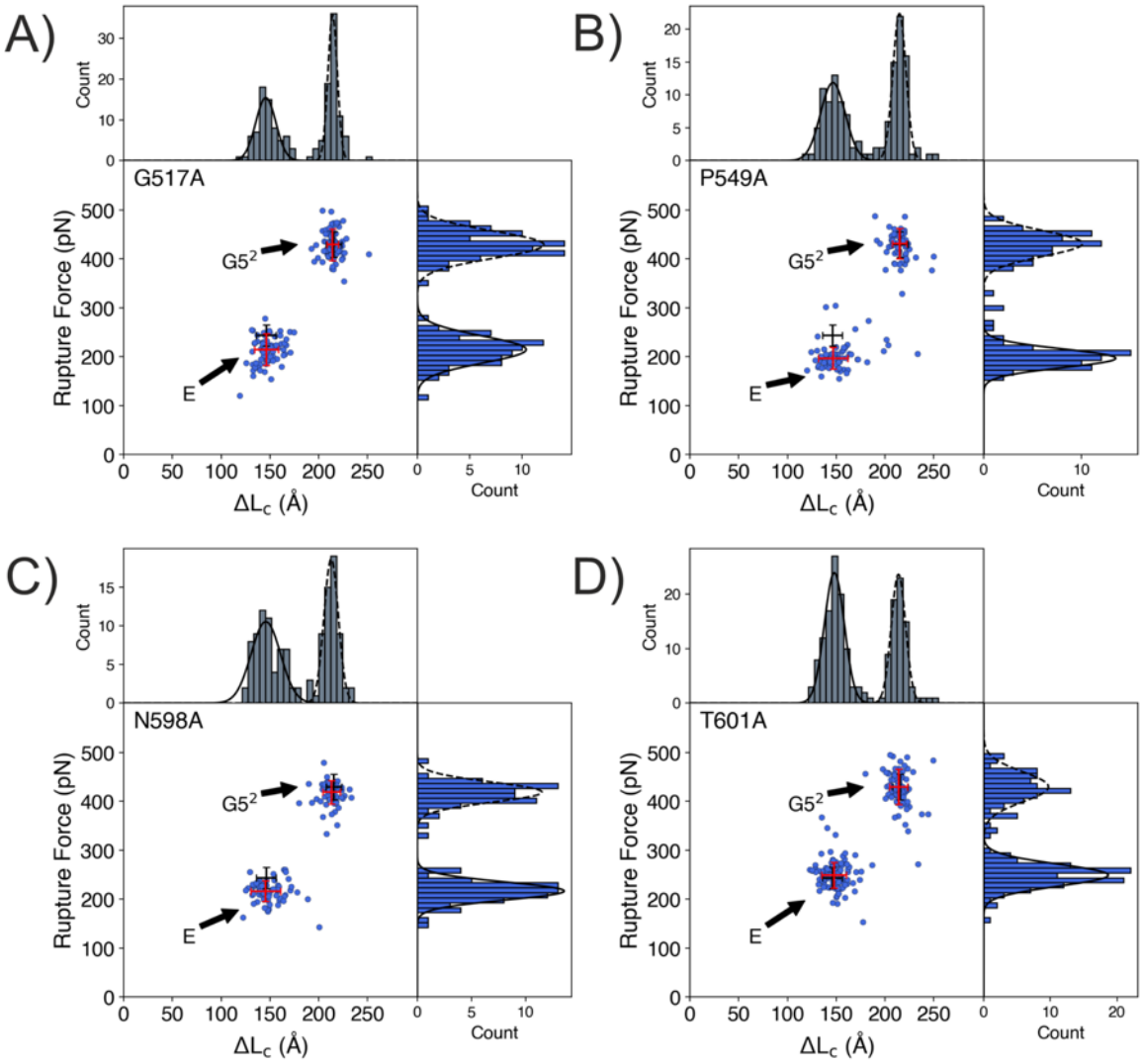

Figure 5.30 – Example rupture force-ΔL_C scatterplot with associated histograms of the interface variants. A) G517A, B) P549A, C) N598A and D) T601A replicate at a retraction velocity of 1500 nms^{-1}. ΔL_C histograms in grey and rupture force histograms in blue with the Gaussian distribution fit as a solid or dashed line for E and $G5^2$ sub domains, respectively. This data displays two distinct populations for both ΔL_C and rupture force indicating two separate sub domains unfolding with different rupture forces. Red scatterplot crosshairs are the mode and the FWHM from the corresponding marginal histogram Gaussian distribution fits. pWT $(E\text{-}G5^2)_5$ cross hairs from one repeat at the corresponding retraction velocity displayed in black for reference.

The speed dependence of average rupture force for all interface motif variants is displayed in Figure 5.31.

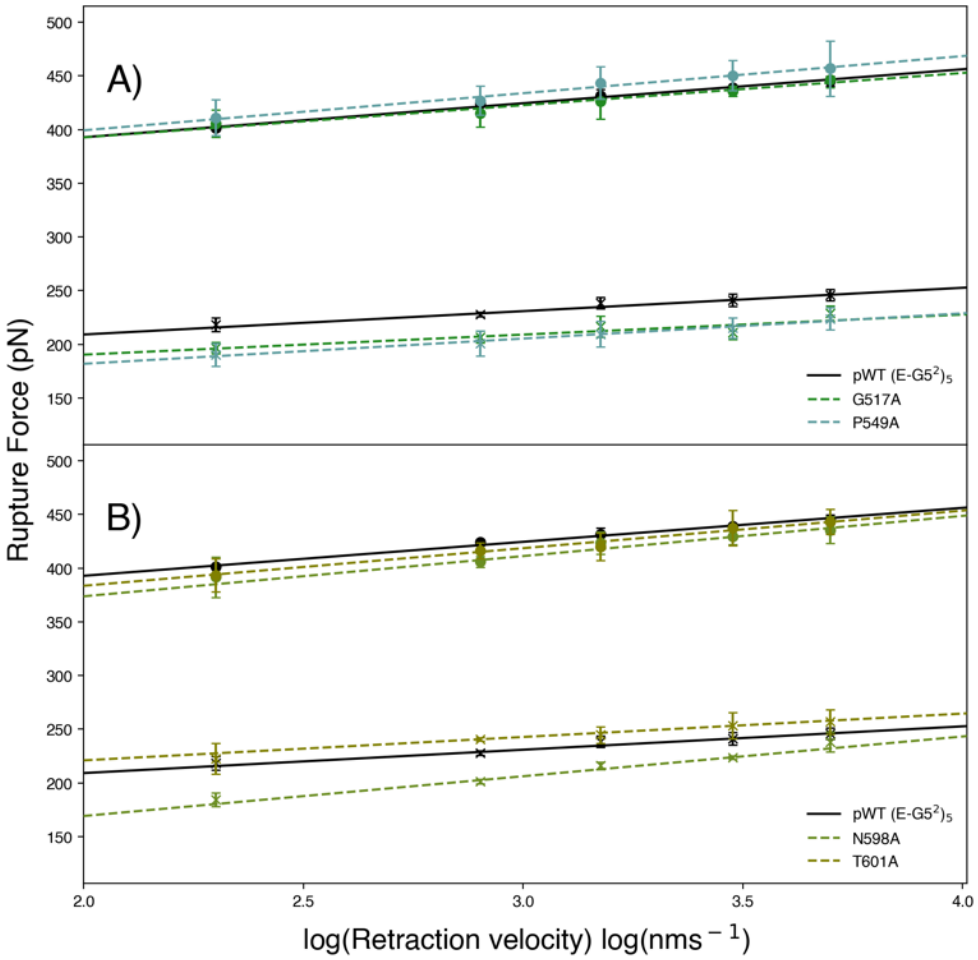

Figure 5.31 – The dependence of rupture force on the logarithm of retraction velocity of the interface variants. E (crosses) and G5^2 (filled circles) sub domains of A) G517A and P549A and B) N598A and T601A variants. Points and errors are the mean and standard deviation of triplicate datasets. Linear fit is weighted with the inverse of the standard deviation error of the triplicate datasets.

Both G517A and P549A displayed a significant decrease in the mechanical strength of the E sub domain across all retraction velocities (Figure 5.31A), with a difference of -22.0 $\pm$ 11.2 and -29.0 $\pm$ 13.0 pN at a retraction velocity of 1500 nms^{-1}, respectively. A significant change in the speed dependence gradient of the E sub domain was observed for N598A (Figure 5.31B). T601A mechanical properties were indifferent to pWT (E-G5^2)$_5$.

5.3.4.3.1 Interface Variant Discussion

Across this variant group, we observed changes in the mechanical properties of the E sub domain only. G517A E sub domain displays a decrease in the mechanical strength, suggesting a mechanically weaker protein. Spectroscopic analyses suggests G517A has both a similar secondary structure and tertiary structure to pWT $(E\text{-}G5^2)_5$, suggesting SMFS is probing a loss of interaction(s). G517 is predicted to form a main chain hydrogen bond with P549, which may be lost if the loop containing G517A subtly changes structure to accommodate the non-native methyl side chain.

Although the CD data indicates P549A has secondary structure analogous to pWT $(E\text{-}G5^2)_5$, subtle differences in the fluorescence emission intensity suggests a change in tertiary structure, presumably at the interface. P549, part of the short GP linker connecting E and $G5^2$, interdigits between F510 and T516 of the E sub domain and is part of the pseudohydrophobic core interacting with P510 of E and P599 and I605 of $G5^2$ (Figure 5.32). P549A may be unable to complete the interactions necessary for the pseudohydrophobic core and in turn results in a markedly destabilised E sub domain.

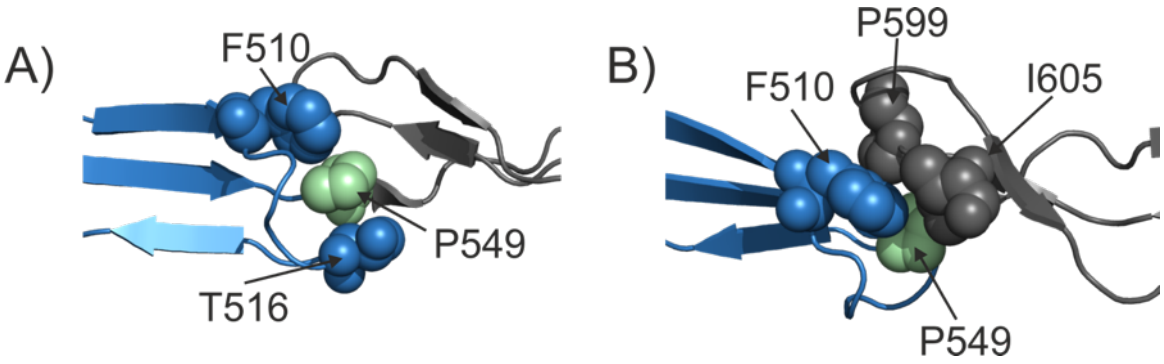

Figure 5.32 – Important residues in the $E\text{-}G5^2$ pseudohydrophobic interface. A) P549 interdigits between F510 and T516 at the interface. B) The pseudohydrophobic core found at the interface comprises of F510, P549, P599 and I605.

N598A is the only variant in this study which affects the mechanical properties of a different sub domain to the one it is located in. The spectroscopic analyses indicates both secondary and tertiary structural changes, with the latter likely affecting the interface.

The importance of hydrophobic contacts in tailoring the mechanical strength has been observed before. In protein L[97] and protein GB1[96] the mechanical properties are mostly determined by the topology of the main chain but are tailored by precise packing of amino acid side chains across the mechanical shearing interface. In our case the hydrophobic contacts are not positioned across the main force-bearing regions, but are situated at the interfaces between mechanical units and loss of these can result in an increase in the speed dependence gradient and/or a decrease in mechanical strength of the E sub domain. An increase in the gradient is usually observed when the TS is bought closer to the native state in the reaction coordinate. In contrast to our perturbation of the pseudohydrophobic core, hydrophobic reduction variant, I60V, of protein L displays an increase in the distance from the native state to the TS suggesting an increase in mechanical 'softness'[97].

Interestingly, the speed dependence gradient of the N598A E sub domain now matches that of the E sub domain in the SasG construct, which has the pseudohydrophobic cores at both the C- and N-terminal, in comparison to the system utilised in this study where the G5-E interface is absent. This indicates that the loss of contacts on both the C- and N-terminal of the E sub domain (in N598A) results in 'levelling out' of the speed dependence and that 'external' hydrophobic interactions govern, to some extent, the malleability of the E sub domain. Furthermore, this could indicate that interactions in the pseudohydrophobic core play a part in the mechanical response of the $G5^2$ sub domain *in vivo*, however, the ability to resolve this is precluded by the nature of this system (and SasG) as the interfaces are broken down prior to $G5^2$ unfolding.

5.3.4.4 Charge Reversal Variants

The charge reversal variants included E588K, K589E and E624K (Figure 5.33) and are positioned in areas with charged networks in an attempt to disrupt any potential side-chain salt bridges contributing to the 'mechanical clamps'. An example scatterplot-histogram of rupture force and ΔL_C data for each variant at a retraction velocity of 1500 nms^{-1} is displayed in Figure 5.34.

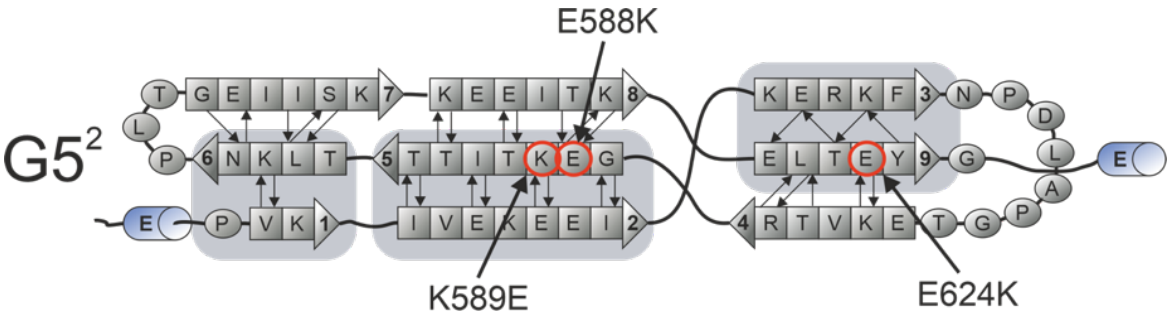

Figure 5.33 – Schematic of the G5² sub domain with the highlighted location of charge reversal variants discussed in this section. The predicted 'mechanical clamps' are highlighted as light grey boxes.

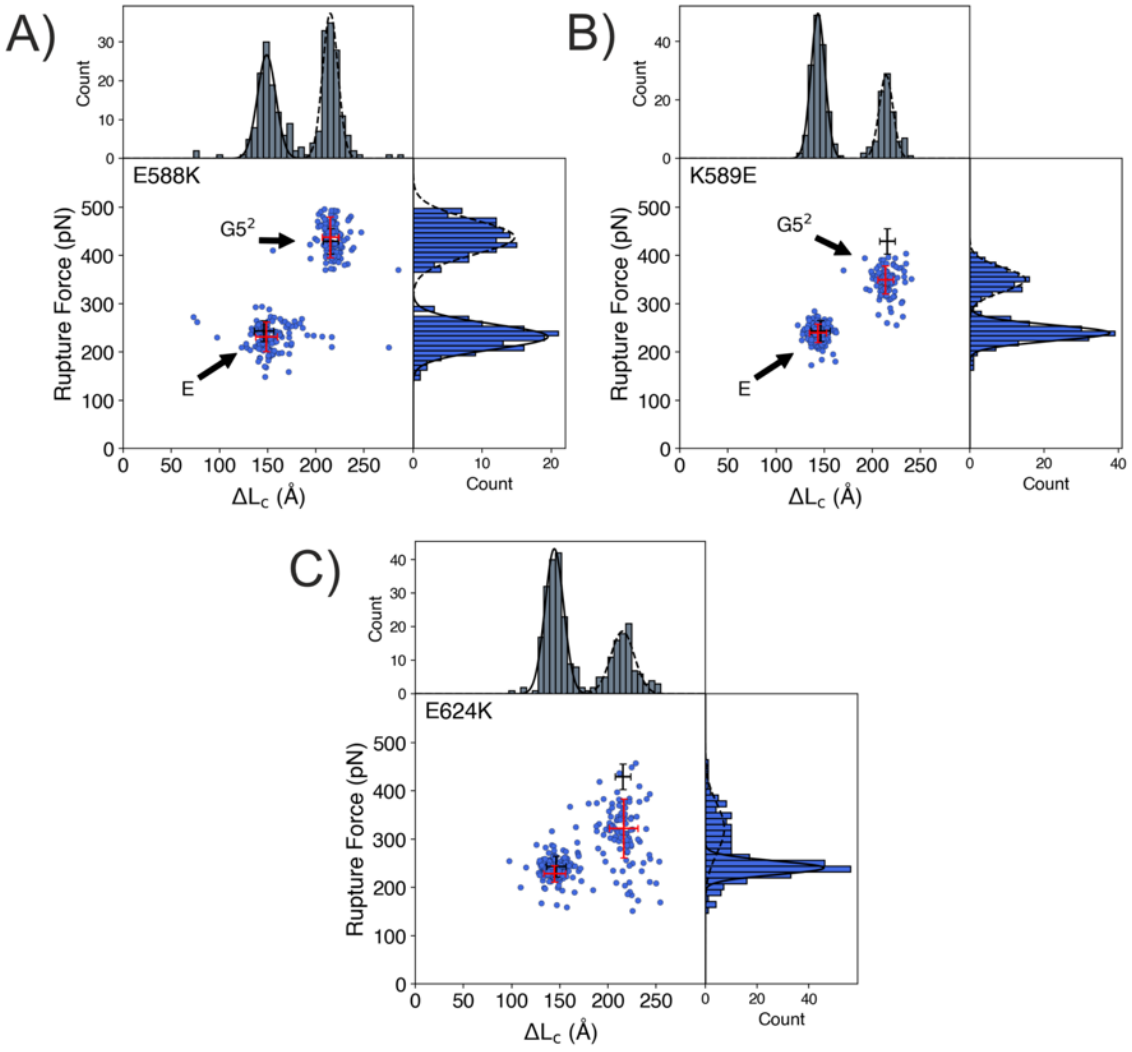

Figure 5.34 – Example rupture force-ΔL_C scatterplot with associated histograms of the charge reversal variants. A) E588K, B) K589E and C) E624K replicate at a retraction velocity of 1500 nms^{-1}. ΔL_C histograms in grey and rupture force histograms in blue with the gaussian distribution fit as a solid or dashed line for E and G5² sub domains, respectively. This data displays two distinct populations for both ΔL_C and rupture force indicating two separate sub domains unfolding with different rupture forces. Red scatterplot crosshairs are the mode and the FWHM from the corresponding marginal histogram Gaussian distribution fits. pWT $(E\text{-}G5^2)_5$ cross hairs from one repeat at the corresponding retraction velocity displayed in black for reference.

The speed dependence of average rupture force for all the charge reversal variants is displayed in Figure 5.35, which displays a varying mechanical response within this variant group. K589E displays a significant decrease in the unfolding force of the $G5^2$ sub domain across all retraction velocities, with a decrease of -73.8 $\pm$ 15.8 pN when pulled at 1500 nms^{-1}. E624K exhibits a considerable increase in gradient (~ 2x as steep) and is substantially mechanically perturbed, with a decrease of -99.2 $\pm$ 11.4 pN at a retraction velocity of 1500 nms^{-1}.

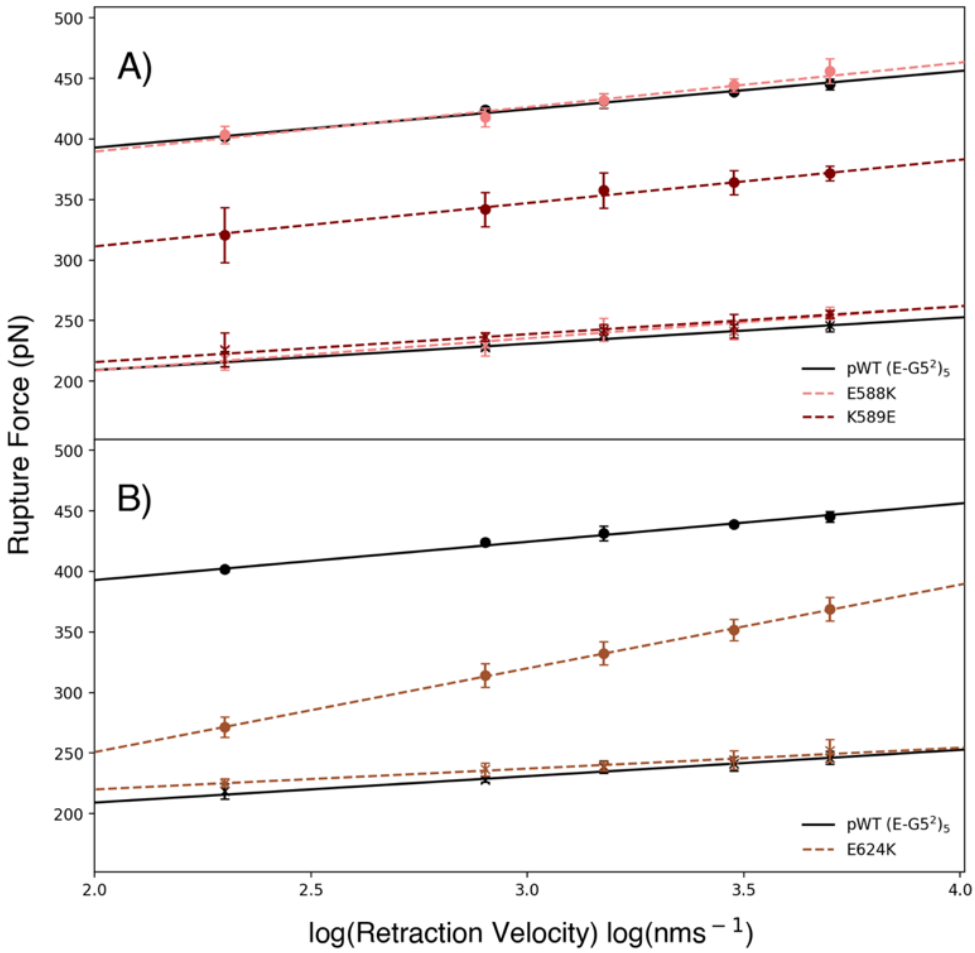

Figure 5.35 – The dependence of rupture force on the logarithm of retraction velocity of the charge reversal variants. E (crosses) and $G5^2$ (filled circles) sub domains of A) E588K and K589E and B) E624K variants. Points and errors are the mean and standard deviation of triplicate datasets. Linear fit is weighted with the inverse of the standard deviation error of the triplicate datasets. E624K $G5^2$ sub domain displays a sizeable change in gradient.

5.3.4.4.1 E624K – A Special Case

In addition to E624K displaying a large decrease in mechanical strength and a substantial change in gradient there is striking amount of heterogeneity of $G5^2$ sub domain unfolding height within and between FX profiles (Figure 5.36).

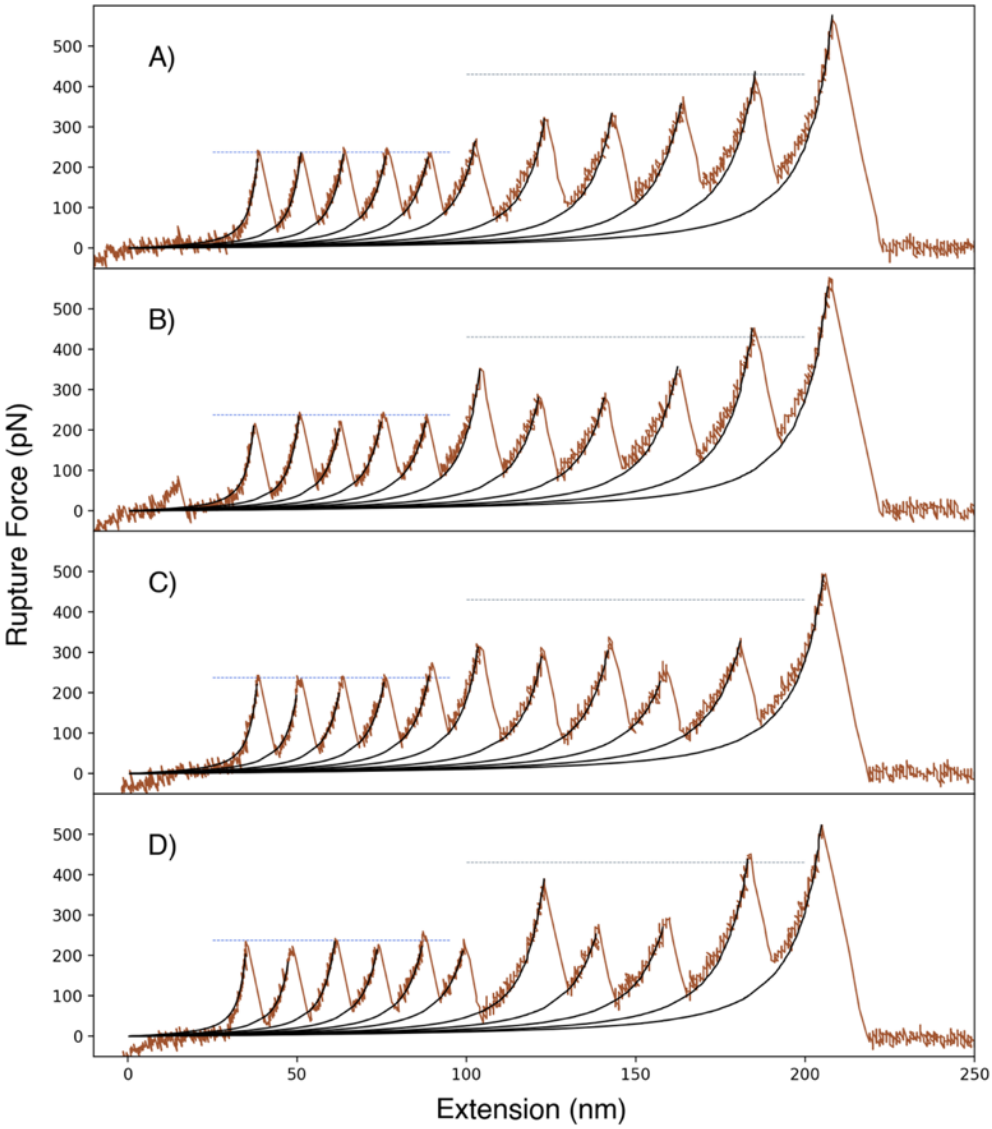

Figure 5.36 – Examples of typical FX profiles of E624K in 1X PBS, pH 7.4 at a retraction velocity of 1500 nms^{-1}. WLC model fitting as black lines. pWT $(E\text{-}G5^2)_5$ E and $G5^2$ average unfolding forces at a retraction velocity of 1500 nms^{-1} in blue and grey dotted lines, respectively. A-D) These FX profiles aim to visualise the heterogeneity of the height of $G5^2$ rupture peak, which vary vastly both between FX profiles and within FX profiles. The E sub domains are unchanged from pWT $(E\text{-}G5^2)_5$.

This heterogeneity results in a large increase in the FWHM of the $G5^2$ rupture force Gaussian fit in comparison to the pWT $(E\text{-}G5^2)_5$ (visualised in Figure 5.37), whereas the E sub domain FWHM

remained indifferent. The average G5^2 rupture force FWHM of pWT (E-G5^2)$_5$ and E624K was ~ 57 and ~ 116 pN, respectively (Appendix – Table 7.37). Although the G5^2 ΔL_C of E624K was agreeable within error to pWT (E-G5^2)$_5$, the average ΔL_C Gaussian fit FWHM was significantly wider for E624K than pWT (E-G5^2)$_5$ (~ 28 vs ~ 15 Å) (Appendix – Table 7.38). Furthermore, the G5^2 rupture force histogram distribution deviates from a strictly unimodal fit typically observed for all other variants (Figure 5.34C and Section 7.3.4). These observations offer strong evidence for a change in the underlying energy landscape of mechanical protein unfolding and/or potentially alternative unfolding pathway(s) for the E624K G5^2 sub domain[44,113,214,215].

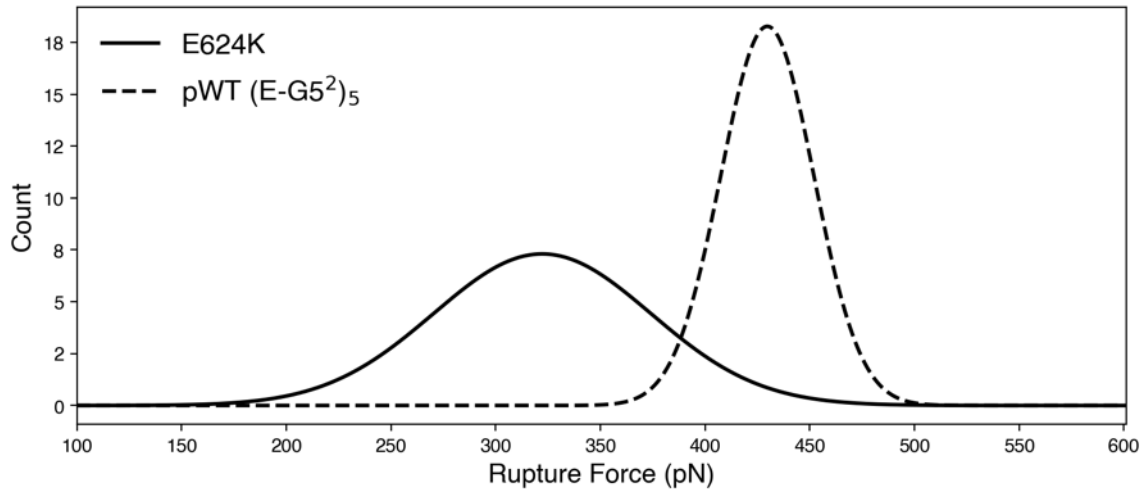

Figure 5.37 – Visualisation of the increase of FWHM of E624K G5^2 rupture force Gaussian fit in comparison to pWT (E-G5^2)$_5$ at a retraction velocity of 1500 nms^{-1}. For these replicates, the FWHM is 122.2 and 52.4 pN for E624K and pWT (E-G5^2)$_5$ G5^2 sub domain unfolding, respectively.

5.3.4.4.2 Reverse Charge Variants Summary

These results indicate significant involvement of the K589 and E624 residues in the mechanical response of the G5^2 sub domain. Due to the position of K589 in the β-sheet and its side chain proximity to oppositely charged residues within the 'mechanical clamp', it is likely it contributes to the side chain packing interactions of the predicted N-terminal 'mechanical clamp' of the G5^2 sub domain[74]. This may suggest that the electrostatic interactions are anchoring these clamps in a shear geometry – an electrostatic locking mechanism. A similar paradigm has been observed before, with electrostatic locking of the SdrG B domain mechanical clamp using Ca^{2+} ions[76]. Interestingly, K589E and E588K are adjoining charged residues with distinct mechanical responses. This is probably

due to the lack of potential electrostatic side chain interactions mediated by E588 across the β-strands of the 'mechanical clamp' in comparison to K589.

E624K has a disordered $G5^2$ sub domain mechanical phenotype with the SMFS data suggesting that it may be utilising alternative unfolding pathway(s). E624K displays an identical secondary structure to pWT $(E\text{-}G5^2)_5$, which implies that it is a loss of interactions, not structure, that is responsible for the mechanical phenotype(s). E624 is situated in a highly charged network and forms side chain hydrogen bonds with the N570 side chain on the outside β-strand of the predicted C-terminal 'mechanical clamp' and T581 on the other outside β-strand (Figure 5.38A). Charge reversal of E to K (FACTS implicit solvent model) predicts the loss of the hydrogen bonding, with presumably a perturbation of the charged network electrostatics (Figure 5.38B). This may destabilise the C-terminal 'mechanical clamp' through loss of a side-chain hydrogen bonds and electrostatic protection, which results in uncontrolled forced unfolding through several pathways. The large reduction in force and altered mechanical properties of E624K supports the hypothesis of a C-terminal 'Mechanical clamp'.

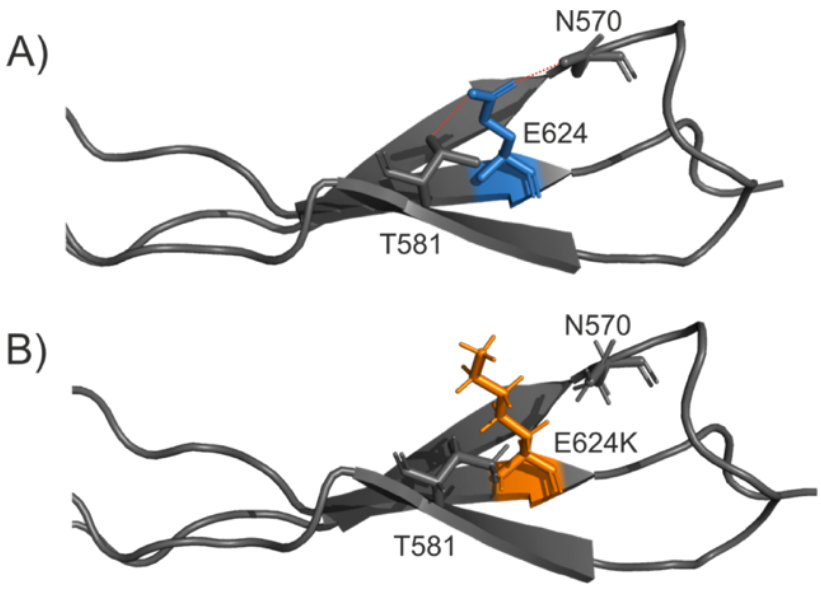

Figure 5.38 – E624 is the central part of a side-chain hydrogen bonded network at the C-terminal of $G5^2$. A) E624 side chain is predicted to form hydrogen bonds with the N570 and T581 side chains, which appears to stabilise the C-termini of the $G5^2$ sub domain by forming a hydrogen bond network across the three β-strands. B) FACTS implicit solvent model of E624K predicting a loss of these stabilising hydrogen bonds.

Interestingly, Gruszka and colleagues observed the Y625W variant of monomeric E-G5[2] (found in the C-terminal G5[2] nucleation site) to fold via a new pathway[152]. Together with our SMFS data, this suggests that interactions in this region are pivotal for both folding and forced unfolding of G5[2].

5.3.5 Mechanical ϕ-value Analysis

Mechanical ϕ-value analysis is utilised to determine the extent of native structure/interaction preservation proximal to the mutated site at the TS of mechanical protein unfolding by comparing the relative effects of mutation on the ground state and TS[113]. The ϕ-value is defined as the ratio of change in the height of the transition barrier of mechanical protein unfolding ($\Delta\Delta G_{TS-N}^{WT-MUT}$) to the change in free energy difference between the folded and unfolded state of the protein ($\Delta\Delta G_{UN}$) as outlined in Section 1.3.5.3. Folding ϕ-values are discussed here (Section 2.2.5.7), which describe the TS for forced folding (ϕ_F^F) and allows the structural preservation of the residues local environment at the mechanical TS to be determined, with a value of 0 (fully unformed) to 1 (fully formed) at the atomic level of individual residues. A partial value suggests that the mutated residue only forms a fraction of its native contacts in the mechanical TS.

MC simulations are usually used to obviate the loading rate effects of compliance and domain number. However, as we are comparing variants of the same construct (same domain number and order of unfolding), these effects will theoretically cancel out and we can utilise the difference in rupture force values in the calculation of the change in activation free energy of unfolding ($\Delta\Delta G_{TS-N}^{WT-MUT} = RT(f_{WT} - f_{MUT})/m$)[113]. A fixed gradient (m) was utilised (average of all gradients which were not significantly different) and we assumed that all mechanical unfolding mechanisms are identical and in turn mechanical unfolding transition states are at the same location on the reaction coordinate with no perturbation of the native or denatured structure.

A caveat of our analysis is that the E sub domain only folds in the presence of an C-terminal folded G5[2] sub domain[74], which precludes direct determination of thermodynamic destabilisation values for the E sub domain. Although G5[2] does fold in isolation, we opted to use thermodynamic stability changes of E-G5[2] sub domains, so all comparisons are relative. Another issue is that E-G5[2] unfolds cooperatively in chemical denaturant, whereas E-G5[2] loses cooperativity under external

mechanical force, so we are unable to decipher individual equilibrium folding parameters for E and G5[2].

All mechanical ϕ-values calculated from a difference in mechanical force at a retraction velocity of 1500 nms^{-1} are outlined in Table 5.5.

Variant / Group	Sub Domain Location	Sub domain E ϕ_F^F	Error	G5[2] ϕ_F^F	Error
I502P / MC	E	**0.5**	± **0.1**	1.0	± 0.1
V556P / MC	G5[2]	0.9	± 0.1	**1.0**	± **0.1**
V580P / MC	G5[2]	1.0	± 0.2	***0.9***	± ***0.1***
G524A / CLM	E	**0.6**	± **0.1**	1.0	± 0.1
G527A / CLM	E	**0.5**	± **0.1**	1.0	± 0.2
P540A / CLM	E	**1.3**	± **0.7**	1.2	± 0.6
P562A / CLM	G5[2]	1.1	± 0.8	**- 0.2**	± **0.1**
G584A / CLM	G5[2]	1.0	± 0.2	**0.7**	± **0.1**
G587A / CLM	G5[2]	1.0	± 0.1	**0.6**	± **0.1**
G517A / I	E	**0.6**	± **0.1**	0.9	± 0.2
P549A / I	GP Linker	0.6	± 0.1	1.1	± 0.2
N598A / I	G5[2]	*0.9*	± *0.1*	**1.0**	± **0.1**
T601A / I	G5[2]	1.1	± 0.2	**0.9**	± **0.1**
E588K / CR	G5[2]	1.0	± 0.2	**1.0**	± **0.1**
K589E / CR	G5[2]	1.0	± 0.2	**0.6**	± **0.1**
E624K / CR	G5[2]	1.0	± 0.1	***0.8***	± ***0.1***

Table 5.5 – Mechanical ϕ-values describing the transition state for forced folding (ϕ_F^F) of the E-G5[2] variants in this thesis. P540A and P562A (underlined) have small $\Delta\Delta G_{UN}$ values, which can cause artefactual ϕ-values[101,216], so these are to be analysed with caution. Italicised numbers are those where there is potentially a change in the x_u, so their gradient is utilised and are included to provide insights into what the structure/contacts at this location may look like at the TS. Values in red are partial ϕ-values and indicate structural perturbation/loss of contacts in the mechanical unfolding transition state. Bolded values are those calculated for the sub domain in which the substitution is located. MC: 'mechanical clamp', CLM: collagen-like motif, I: interface and CR: charge reversal.

Calculated ϕ-values of ~ 1 for the sub domain in the E-G5[2] repeat without the mutation (the control sub domain) acts as a quality control and suggests this is a valid method for the calculation of mechanical ϕ-values for the E and G5[2] sub domains. Absence of values ~ 0, suggests there are no regions entirely unstructured at the mechanical TS.

Typically, nondisruptive deletion mutations of hydrophobic moieties or substitution with glycine are utilised in ϕ-value analysis, however, as the bulk of E and G5^2 is β-sheet in structure and as the majority of residues are surface exposed, more radical mutations of residues are permitted here[112]. The change in gradient of the V580P G5^2 sub domain, N598A E sub domain and E624K G5^2 sub domain precludes direct comparison of mechanical ϕ-values with the pWT (E-G5^2)$_5$ and other variants as a change in gradient indicates a change in x_u[113]. However, they are included for an insight into what the structuring may be like at their TS. The mechanical ϕ-value of 0.8 for the E624K G5^2 sub domain indicates partial unstructuring at this location at the mechanical TS. As SMFS is a single molecule technique partial ϕ-values can only represent partial structuring of the protein (not an alternative TS)[113]. Although in this instance, it is possible to interpret that the ϕ-value may actually be due to an ensemble of multiple parallel pathways to the TS as the SMFS evidence suggests the G5^2 sub domain is unfolding via multiple pathways[113].

Interestingly, the ϕ-value of 0.5 for I502P implies that this residue is in a region that is significantly perturbed in the mechanical TS of the E sub domain. Conversely, MD simulations do not show a major change at the N-terminal 'clamp region' prior to unfolding and the spectroscopic evidence suggests partial unfolding of the structure in the native state, which implies this partial ϕ-value may be an artefact and could be closer to ~ 1[112].

V556P, T601A, E588K and N598A (G5^2) display ϕ-values of ~1, which implies no loss of contacts/structural perturbation. As the 200 and 5000 nms^{-1} rupture force values for the G5^2 sub domain of V550P (E non-folder) are indifferent from pWT (E-G5^2)$_5$, we will assume the gradient (and thus x_u) is unchanged. Using the Gruszka and colleagues G5^2 ΔG_{UN}, we can calculate $\Delta\Delta G_{UN}$ and in turn a rough estimate for the ϕ-value[74]. This was calculated to be ~ 1, indicating no loss of interactions or structure of the G5^2 N-terminal 'mechanical clamp'.

G517A and P549A, mutations found at the E-G5^2 interface, both have a partial ϕ-value of 0.6, which may indicate a loss of contacts or structure at the mechanical TS. The MD of the forced unfolding of E-G5^2 (Section 2.2.1.2) corroborates these values and predicts that both G517 and P549 stay in contact with the E sub domain prior to global unfolding, but lose contact with the G5^2 N-terminal interface (Figure 5.39).

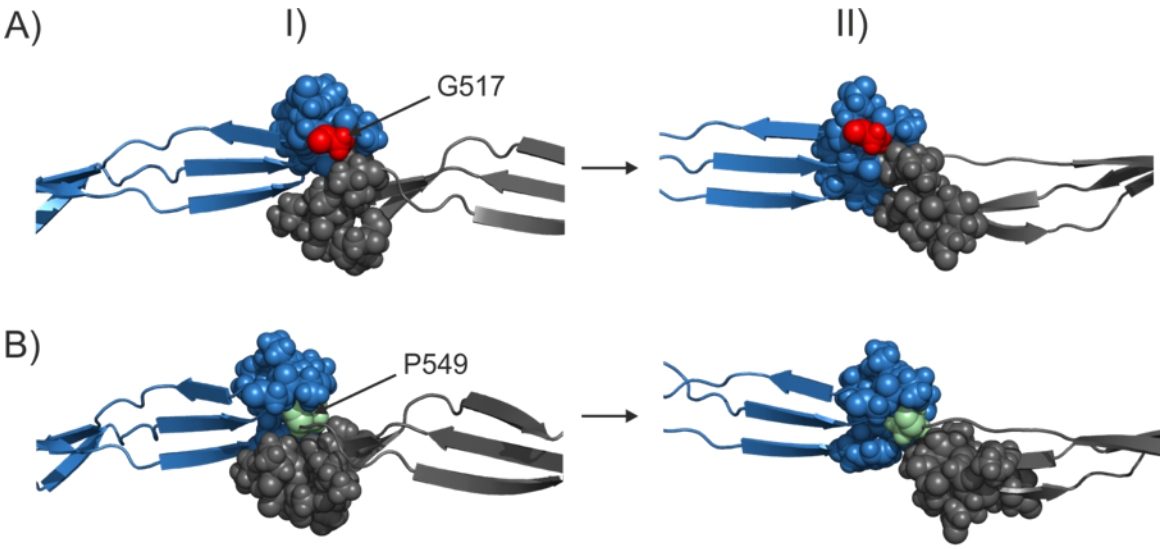

Figure 5.39 – Snapshots of the interface from MD forced unfolding simulations of E-G5^2. A) and B) display snapshots from the MD trajectory with a focus on G517 (red) and P549 (green), respectively. Snapshots are taken at I) zero force and II) prior to E sub domain unfolding. The MD suggests both G517 and P549 lose contacts with the G5^2 N-terminal interface prior to E sub domain unfolding. The partial ϕ-values of G517A and P549A agrees with these observations. MD simulations as described in Section 2.2.1.2.

Generally, the collagen-like motif variants have a decreased unfolding force and no change in their speed dependence, however, spectroscopic analysis does show signs of secondary structure change (probably due to their steric bulk nature) so these partial ϕ-values have to be interpreted with care[112]. However, the partial ϕ-values of 0.6 (G524A), 0.5 (G527A), 0.7 (G584A) and 0.6 (G587A) probably indicates structural perturbation of these sites prior to the mechanical TS. In line with these results, the forced-unfolding MD simulation of E-G5^2 demonstrates the right-handed collagen-like regions (of both the E and G5^2 sub domains) untwisting and elongating (Figure 5.40), prior to global unfolding.

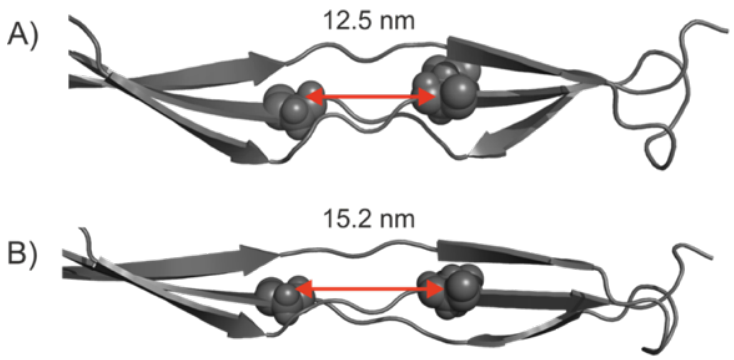

Figure 5.40 – MD trajectory snapshots of the collagen-like region of G5[2] elongating and untwisting during the forced-unfolding simulations on E-G5[2]. Measurement place markers on G587 and E621. A) structure at zero external force displays a native collagen-like structure measuring at 12.5 nm and B) elongated and slightly untwisted collagen-like region measuring at 15.2 nm prior to global unfolding. MD simulations as described in Section 2.2.1.2.

P562A is located in the collagen-like motif and has a ϕ-value of < 0, which may imply the formation of non-native contacts at the TS (destabilising the native state while simultaneously stabilising the TS)[217] or the stabilisation of some structural elements, but destabilising others[218], which may occur due to structural rearrangement of the collagen-like region once loaded with external force. P540A displays a ϕ-value > 1, which may indicate non-native contacts at the TS[217] or stabilisation/destabilisation of both the native state and TS, with a larger effect on the latter[219]. However, these unusual values could be experimental artefacts because of their small $\Delta\Delta G_{UN}$ values[101,216].

This data describes a mechanical TS which exhibits a relatively native overall structure for the β-sheets, but partially non-native for the collagen-like regions and interface.

5.4 Discussion

5.4.1 Chemical vs Mechanical Stability

As mechanical stability is partly based on the height of the activation barrier to unfolding (ΔG_{TS-N}), both a decrease in native state stability or an increase in TS stability due to a residue substitution can lead to a decreased mechanical strength as thermal fluctuations are more likely to drive the protein over the TS with a reduced height of the activation barrier. 9/18 variants were both significantly thermodynamically and mechanically destabilised (at a retraction velocity of 1500 nms-

[1]) in the sub domain of substitution location. These included I502P, G524A and G527A in the E sub domain, G517A and P549A in the interface (reduce in E sub domain mechanostability) and G584A, G587A, K589E, and E624K in the G5[2] sub domain indicating a destabilised ground state (Figure 5.41). Conversely, P562A, located in the G5[2] sub domain was only relatively remotely destabilised, but had a significantly reduced unfolding force, which suggests an increase in TS stability.

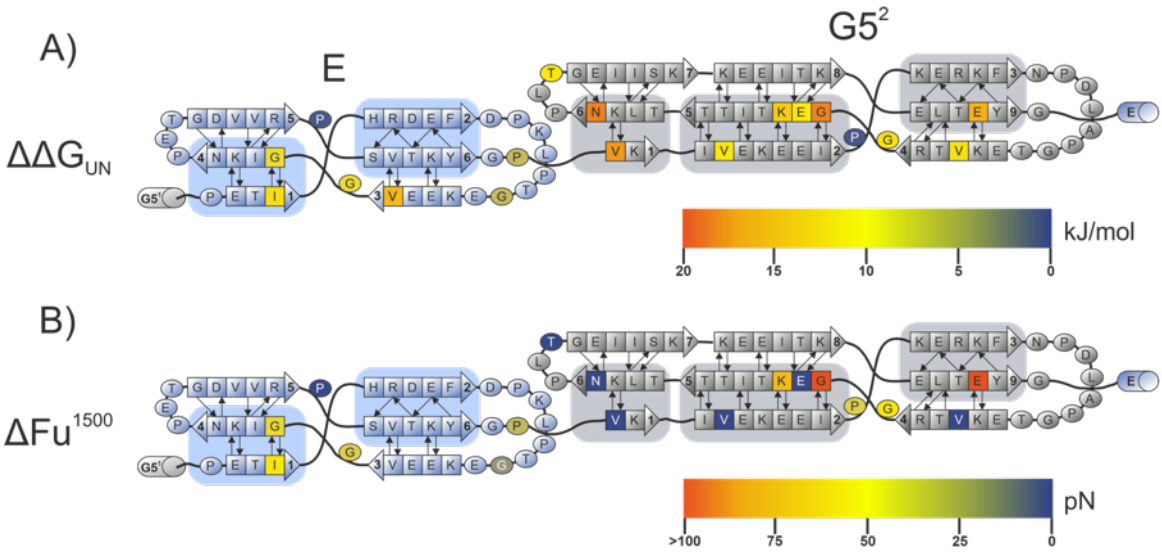

Figure 5.41 – Correlation between thermodynamic and mechanical stability of E-G5[2]. A) $\Delta\Delta G_{UN}$ and B) loss of mechanical strength at a retraction velocity of 1500 nms^{-1} (ΔFu^{1500}). The most mechanically destabilised variants are typically the most thermodynamically destabilised.

In contrast to previous studies[96,97,101], there appears to be a correlation (to an extent) between the thermodynamic and mechanical destabilisation, with the variants with the largest $\Delta\Delta G_{UN}$ values (Figure 5.41A) displaying the largest decreases in mechanostability (G587A and E624K) (Figure 5.41B).

Interestingly, the majority of destabilised variants were located in the collagen-like motif, suggesting a novel structure for protein mechanical strength.

5.4.2 The Collagen-like Region - a Novel Mechanical Motif

The G524/G527 and G584/G587 pairs are structurally equivalent glycine residues located in one (out of three) of the PPII-like strands (making up the collagen-like motifs) and follow the typical collagen sequence $X_{aa}Y_{aa}GlyX_{aa}Y_{aa}Gly$ in the E and G5[2] sub domains, respectively. It is well understood that there are strict sequence restraints on the collagen triple helix requiring glycine every third position for close packing of the chains at the central axis as their side chain is a single hydrogen[220]. The mutation of these glycine residues with a larger residue, such as alanine, leads to Osteogenesis Imperfecta (OI) pathology. OI is a rare disease (brittle bone disease) affecting the connective tissues characterised by extremely fragile bones that break or fracture easily. Bella and colleagues observed the loss of interchain hydrogen bonds and local twist relaxation when glycine was mutated to alanine in a collagen-like peptide[212]. This was due to the local unscrewing of the collagen to accommodate the bulky methyl side chains inside the triple helix (Figure 5.42A). In agreement with this, Punitha and colleagues observed the formation of a local bulge at the substitution site upon the introduction of alanine using MD simulations[213]. The consequential destabilisation was mainly due to a lack of hydrogen bonding and electrostatic repulsions at the site of substitution. When glycine is naively mutated to alanine in G5[2] using PyMOL software[186], it is evident there are steric constraints due to the alanine methyl side chain (Figure 5.42B). In agreement with these observations, our spectroscopic analysis indicates secondary structure changes, potentially due to the collagen-like regions trying to accommodate the bulky methyl side chains and locally untwisting in the process.

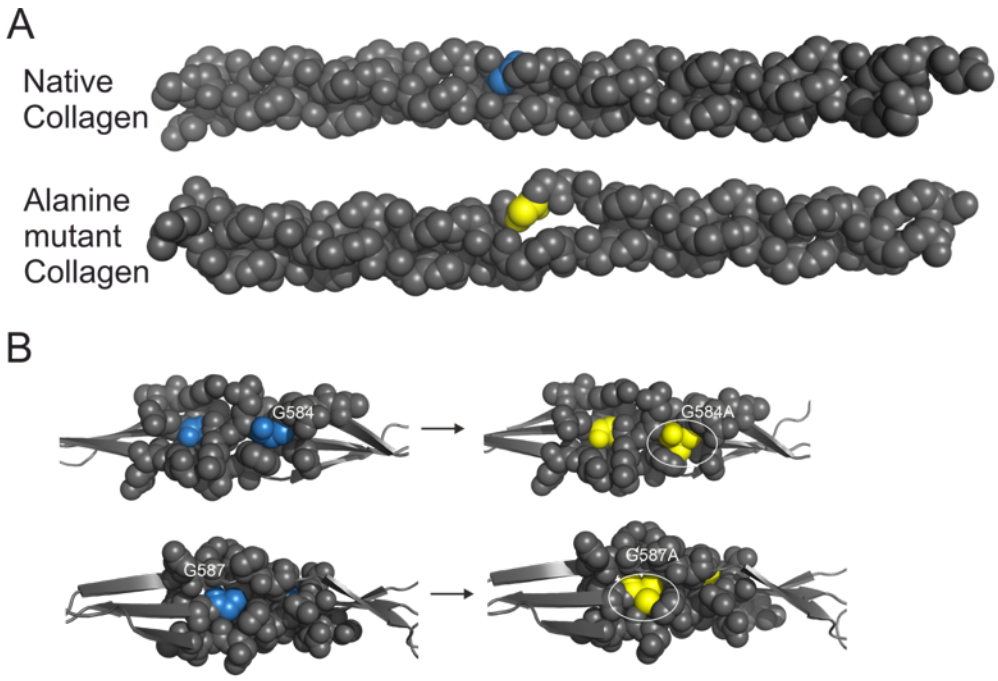

Figure 5.42 – Glycine to alanine mutations in collagen and the collagen-like region of G5[2] of SasG. A) Native collagen structure (PDB: 1BKV[202]) with one of the repeating glycine residues (of $X_{aa}Y_{aa}GlyX_{aa}Y_{aa}Gly$) in blue and mutant collagen structure (PDB: 1CAG[212]) with the same glycine mutated to alanine (yellow). This produces local untwisting and a kink at the site of substitution due to accommodation of the side chain methyl group of alanine in the interior of the triple helix. B) displays two different views of the G5[2] collagen-like motif with the sites of G584A and G587A mutations with native glycine (blue) and mutated to alanine (yellow) on PyMOL[186] to display the steric constrictions (circled) of the additional methyl side chain.

However, this does not explain the large discrepancy in the loss of mechanical strength of G587A (~ 100 pN) in comparison to G584A (~ 40 pN), G524A (~ 30 pN) and G527A (~ 40 pN). G587 is found in a highly charged local sequence environment (KGE) seen in collagenous domains of bacterial proteins that lack hydroxyprolines and have low proline content. In collagen, this KGE sequence contributes a high degree of stability through electrostatic interactions[221]. G587A is reminiscent of a Gly to Ala mutation occurring at position 658 in the 1(I) chain located within a highly charged Lys-Gly-Asp (KGD) local sequence environment, which results in a severe OI pathology[222]. This allowed us to initially speculate that disrupting the charged residues in the KGE sequence found in the G5[2] collagen-like region would lead to a mechanically destabilised phenotype. However, our E588K (KGE_{588}) variant had no effect on the mechanostability, but did display significant reduction in the thermodynamic stability (~ 10 kJ/mol) implying importance for global stability over mechanical stability.

Furthermore, Hyde and colleagues discovered that glycine to alanine substitutions in proline-rich regions of the collagen sequence are far less disruptive to the conformation that those in proline poor regions[223]. As the E collagen-like region is proline rich in comparison to the G5^2 collagen-like region (five vs two proline residues), this may explain the difference in the effect of the G524A/G527A (E) mutations in comparison to the G584A/G587A (G5^2) mutations on the decrease in the mechanical strength of each sub domain.

5.4.3 Mechanical TS Structure

Our mechanical ϕ-value analysis displays a relatively native structure, with the exception of the E-G5^2 interface and collagen-like regions (Figure 5.43A). Partial ϕ-values of G524A, G527A, G584A and G587A implies there is a loss of structure in the collagen-like region at the mechanical TS. In addition, both G517A and P549A display a partial ϕ-value, which implies the interface is broken down prior to global unfolding. MD simulations corroborate these findings and suggest the mechanical TS structure has a detached interface and elongated collagen-like regions, with the former applicable to the E sub domain only, as the interface is broken prior to G5^2 unfolding (Figure 5.43B). The remainder of the structure stays native-like until the mechanical TS, where the structure unfolds rapidly. The MD displays a strand slip 'intermediate event' prior to global unfolding of G5^2 (not shown here), however, our ϕ-values indicate that the strand predicted to slip retains its native contacts at the mechanical TS, suggesting that is not the case.

As the collagen-like region is only partially structured at the mechanical TS, it is unlikely to be a 'mechanical clamp', however, along with interfacial contacts they are still significant for the mechanical properties of the E and G5^2 sub domains. The majority of the residues located in the 'mechanical clamps' display ϕ-values of ~ 1, indicating these contacts are present at the transition state. Our ϕ-value analysis together with the loss of mechanostability of charge reversal variants located in the clamp regions implies the long stretches of hydrogen bonds and their side-chain packing interactions are the main force-bearing structures as predicted by MD simulations[74].

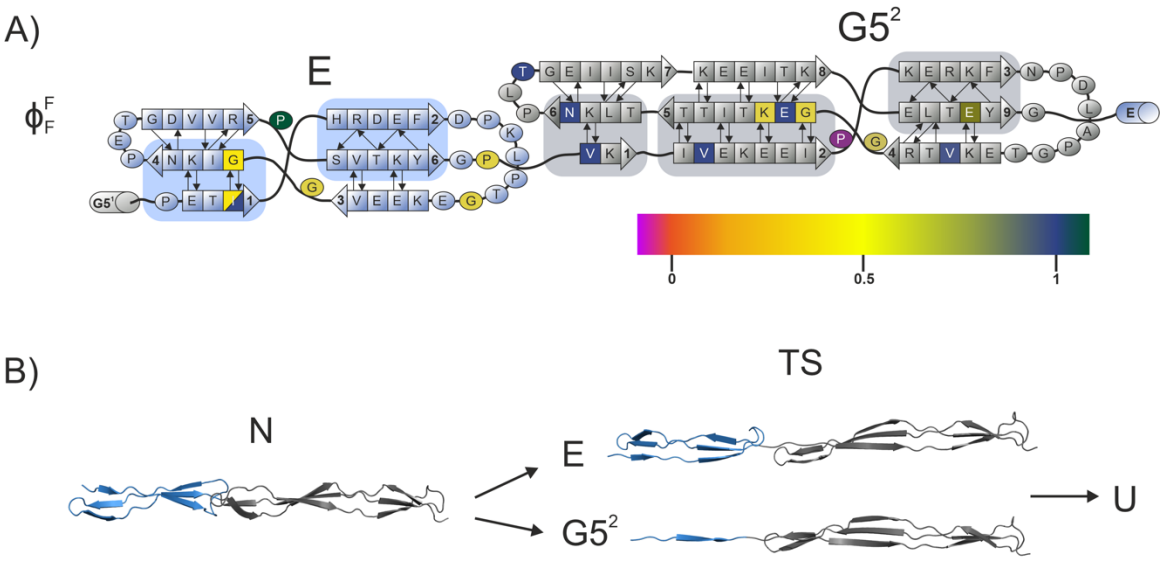

Figure 5.43 – Determining the mechanical TS structures of the E and G5² sub domains of E-G5². A) ϕ_F^F values of the sub domain where the mutation is located in E-G5². I502P displayed as both 0.5 (from calculation), but also as 1 because the variant is partially unfolded in the native state, it is probably giving us a artefactual ϕ_F^F value. B) Speculated mechanical TS structures along the forced unfolding pathway of E and G5². Structure snapshots from the MD trajectory at 0 K (Section 2.2.1.2). The mechanical TS of the E sub domain displays a broken interface and elongated collagen-like regions. The G5² mechanical TS shows an elongated structure, due to collagen-like region extension.

6. Conclusions and Future Directions

6.1 SMFS Reveals the Zn^{2+}-Driven Interactions of SasG

We created a protein system suitable for SMFS to investigate the inconclusive Zn^{2+}-dependent dimerisation/oligomerisation of SasG[20,149]. Our results suggest that a physiologically-relevant concentration of 100 μM Zn^{2+} drives the transient association between SasG molecules *in vitro*. This interaction was absent in the presence of Co^{2+} ions, but was re-established when Zn^{2+} was introduced indicating this interaction shows bias towards Zn^{2+} ions. Cosolvent simulations predicted that the interaction may utilise a set of negatively charged residues, mostly glutamic acid residues, for coordinating Zn^{2+} ions. Considering these predictions, 13 glutamic acid residues (one in each E and G5 sub domain) were substituted with alanine to create SasG-ΔEE. When interrogated using SMFS, a reduced hit rate was observed, suggesting a decrease in available coordinating residues/binding sites and in turn implying a pleomorphic coordination scheme. Lastly, the mechanical properties of SasG in 100 μM Zn^{2+} were indifferent to those in the absence of Zn^{2+}, suggesting that Zn^{2+} is not utilised for mechanical strength. However, FX profiles of > 13 unfolding events were commonly observed, suggesting that SasG molecules may be interacting with one another prior to forced mechanical unfolding.

6.2 The pWT $(E\text{-}G5^2)_5$ System is a Suitable Replacement for SasG for SMFS studies

The SasG B domain E and G5 sub domains display remarkable mechanical stability, with the molecular determinants of this mechanostability predicted by Gruszka and colleagues[74] but not experimentally verified prior to this work. Mutagenesis coupled with SMFS is a powerful technique for unravelling intrinsic elements of mechanical strength, however, the high levels of DNA sequence similarity between E-G5 repeats of the SasG DNA sequence (both native and *E. coli* codon optimised) precludes direct mutagenesis. To circumvent this, GG assembly was utilised to create a rapid workflow of efficient concatenation of $E\text{-}G5^2$ cassettes to create the pWT $(E\text{-}G5^2)_5$ construct. This construct displayed comparable secondary structure to SasG, suggesting the secondary

structure had been retained. Furthermore, the E-G5^2 repeats were observed to behave independently of one another, but the cooperative unfolding of the E and G5^2 sub domain was retained with ΔG_{UN} and m_{UN} values in excellent agreement with published monomeric E-G5^2 values[74,152]. The mechanical properties of pWT (E-G5^2)$_5$ were somewhat altered from SasG, however, this difference in mechanical properties was not unexpected due to the consequence of changing the supramolecular scaffold (addition of seven residue linkers) and number of domains in comparison to SasG[69]. These changes in mechanical properties of the E and G5^2 sub domains are likely trivial, as we will be comparing pWT (E-G5^2)$_5$ with variants of itself and theoretically any differences are likely to cancel out. An interesting observation was the change in the malleability of the E sub domain as this may suggest an involvement of the interface in the mechanical properties of SasG. In turn, we believe pWT (E-G5^2)$_5$, is an ideal model system to substitute SasG in uncovering the molecular determinants of the E and G5 sub domains through mutagenesis and rapid construct assembly coupled with SMFS studies.

6.3 The Mechanical Strength of SasG is Multifaceted

Mutagenesis was utilised to create 18 variants of the pWT (E-G5^2)$_5$ construct, which were subjected to structural characterisation, equilibrium studies of denaturation using urea and single-molecule forced unfolding studies. 15 out of 18 variants were fully folded, with the remaining 3 displaying either a partially or fully unfolded E sub domain.

The ground state stability was compared with the local mechanical unfolding kinetics for 16 out of 18 of the variants to give us an insight into the structure of the mechanical TS (mechanical ϕ-value analysis). The ϕ-value analysis, coupled with the MD forced unfolding simulation, suggests that the TS of both the E and G5^2 sub domains comprises a ‘stretched out’ collagen-like region, native-like β-sheet structure and a loss of structure at the E-G5^2 interface. The latter is unique to the E sub domain TS as the E-G5^2 interface has been broken down prior to initiation of the G5^2 unfolding pathway. Furthermore, our data suggests that it is the MD-predicted ‘mechanical clamps’ and their side chain packing interactions (such as electrostatic interactions) which are the major force-bearing regions prior to global unfolding[74]. This is tailored (for the E sub domain) by contacts

(probably hydrophobic in nature) in the pseudohydrophobic core at the E-G5^2 interface. Unfortunately, due to the E sub domain unfolding and in turn loss of the interface before G5^2 unfolding, we are unable to probe the effect of a disrupted interface on the mechanical phenotype of the G5^2 sub domain using our protein system.

In addition to the hydrogen bonded 'mechanical clamps' with side-chain packing interactions and interfacial interactions, SasG utilises a novel mechanical motif, the collagen-like motif, to impart substantial mechanostability to both the E and G5^2 sub domains. The majority of our collagen-like region variants are glycine to alanine substitutions, which are commonly observed in OI disease pathology. There are strict sequence restraints on the collagen triple helix requiring glycine every third position to permit tight packing of the chains at the central axis[220]. In turn, glycine to alanine substitutions lead to twist relaxation and bulging at the site of substitution as the helix attempts to accommodate the methyl side chain of alanine[212]. We observed a significant decrease in mechanical strength of our glycine to alanine collagen-like region variants. This is a crucial finding as it not only allows us to understand the mechanical determinants of load-bearing bacterial proteins, but has inadvertently offered us an explanation as to why OI patients suffer from weakened tendons and fragile bones – the ability of the collagen triple helix to withstand load on a monomeric scale has been diminished. Furthermore, we now have a novel structure to exploit when tuning the mechanical strength of designer proteins.

6.4 Future Directions

6.4.1.1 How many B-repeats are Required for SasG Homophilic Binding *in vitro*

SasG was determined to only induce biofilm formation with five or more B-repeats, those cells expressing four B-repeats and below did not[144]. Unsurprisingly, Formosa-Dague and colleagues showed that *S. aureus* cells expressing one B-repeat did not adhere to one another[153]. *In vitro* examination of Zn^{2+} induced properties have only been carried out on up to two B-repeats and displayed no dimerisation/oligomerisation[20,149]. It would therefore be of interest to determine the minimum B-repeat requirement *in vitro*, as this may help us understand the conformation of the transient complexes.

6.4.1.2 The Metal Ion Specificity of SasG Homophilic Binding

The SasG homolog, Aap of *S. epidermidis* Zn^{2+} coordination sites accommodate Mn^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+} and Zn^{2+}, however only the latter two can support Aap self-association[192]. Thus, it would be interesting to utilise our protein system to test a number of different metal ions to determine how specific the Zn^{2+}-induced binding of the SasG B domain is *in vitro*.

6.4.1.3 Does SasG Utilise Zn^{2+} Ions to Mediate Heterophilic Binding?

Formosa-Dague and colleagues observed strong adhesion between *S. epidermidis* cells expressing Aap and *S. aureus* cells expressing SasG in the presence of Zn^{2+} utilising SCFS[153]. It would be interesting to see how this translates to *in vitro* utilising SMFS with a cantilever tip functionalised with Aap and the substrate with SasG in the presence of 100 μM Zn^{2}.

6.4.1.4 Can the Zn^{2+}-induced Binding be Abrogated through Mutagenesis?

As SasG-ΔEE displayed a reduced binding capacity, it would be of interest to substitute the remaining predicted residues with alanine in an attempt to abolish the Zn^{2+}-induced binding capacity. This would provide undisputable evidence that SasG is utilising Zn^{2+} to form homophilic bonds. However, when substituting a large number of residues it is likely the surface properties will be significantly altered, which may lead to protein instability and result in change in structure (complicating the results) or an unfolded protein.

6.4.1.5 Utilise MC Simulations to Determine the Underlying Energy Landscape Changes in Our Variants

As our experimental observations indicate a linear fit of speed dependence for the pWT $(E\text{-}G5^{2})_5$ and variants thereof, we can assume the data can be described by the Bell model[66]. The analytical model of this, the Bell-Evans-Ritchie model (Equation 1.3) can be used to derive the unfolding energy landscape parameters under zero external force (x_u and k_u^{0F}) from the loading rate. However, as it is not possible calculate the loading rate from a polyprotein unfolding experiment, due to the competing domain number and compliance effect, a MC model can be fitted to the speed dependence data to extract these underlying energy landscape parameters[50,70]. This would provide us with more information on the underlying energy landscape changes and permit more robust comparison of variants.

6.4.1.6 Can We Determine Whether the Interface Tailors the Mechanical Properties of the G5 Sub Domain?

To investigate the effect of the hydrophobic core on the G5 sub domain, we would require the G5 sub domain to unfold prior to the E sub domain. One way around this would be to utilise a mechanically weak (rupture force < E sub domain) $G5^2$ variant of pWT $(E\text{-}G5^2)_5$ as a new pseudo-WT. However, this may be difficult as we observed the most mechanically destabilised $G5^2$ sub domains (still unfolding at a higher force than the E sub domains) to be the most thermodynamically destabilised. This may suggest the creation of a double-mutant may result in a highly destabilised or unfolded protein. Alternatively, the E sub domain could be engineered to be inextensible by SMFS. For example cysteines could be introduced across the mechanical clamps of the E sub domains to induce disulphide bond formation[62]. This would lock the E sub domain from unfolding and allow the effect of interface variants on the hydrophobic contacts of the G5 sub domain to be investigated.

7. Appendix

7.1 DNA sequences

H_6 displayed in green, start codon in blue, stop codon in red, DNA encoding the TEV cleavage site in orange and the linkers in pWT $(E\text{-}G5^2)_5$ in purple.

7.1.1 H_6-MBP-TEV-SasG

ATGCATCACCATCACCATCACAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAG
TCACCGTTGAGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGA
TGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGT
TGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTACCTGGGATGC
CGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTT
ATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAA
AGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCT
GGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACAT
TAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTG
ATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAA
GGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAG
TGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGC
GTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCG
AAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCC
GTAGCGCTGAAGTCTTACGAGGAAGAGTTGGTGAAAGATCCGCGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCC
GTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAG
ACGCGCAGACTAATTCGAGCTCGAACAACAACAACAATAACAATAACAACAACCTCGGGATC
GAGGGAAGGATTTCACATATGGAAAACCTGTACTTTCAGGGATCCGCACCTAAGACCATCAC
CGAGCTGGAGAAGAAAGTTGAAGAGATTCCTTTTAAAAAAGAACGTAAGTTCAACCCGGATC
TGGCACCTGGTACCGAGAAGGTGACACGTGAGGGCCAGAAAGGCGAGAAGACAATCACAA
CACCTACCTTAAAAAATCCTTTAACAGGCGTGATTATCAGTAAAGGTGAACCGAAAGAGGAG
ATTACCAAAGACCCGATCAACGAGCTGACAGAGTATGGCCCGGAAACCATCGCACCGGGCC
ATCGTGATGAGTTTGATCCGAAGTTACCTACCGGTGAGAAGGAAGAAGTTCCTGGTAAGCCG
GGTATTAAAAACCCTGAGACAGGCGATGTGGTTCGTCCGCCTGTTGACAGCGTTACCAAATA
CGGTCCTGTTAAAGGCGACAGTATTGTGGAAAAAGAGGAAATCCCGTTCGAAAAAGAACGCA
AATTTAATCCTGATTTAGCACCGGGCACCGAAAAAGTGACCCGTGAGGGTCAAAAAGGTGAG
AAGACCATTACAACCCCTACACTGAAAAACCCGCTGACCGGCGAGATCATTAGCAAGGGTGA
GAGTAAGGAAGAGATCACAAAGGACCCTATTAACGAACTGACCGAATACGGCCCGGAGACC
ATTACCCCGGGTCACCGTGACGAGTTCGATCCTAAGCTGCCTACAGGCGAAAAGGAAGAAG
TGCCTGGTAAACCGGGCATCAAGAACCCTGAAACCGGCGACGTGGTTCGTCCGCCGGTTGA
CAGTGTTACCAAGTACGGCCCGGTGAAGGGCGATAGCATCGTTGAGAAGGAAGAAATCCCG
TTTGAGAAAGAGCGCAAATTCAATCCGGATCTGGCCCCTGGTACCGAGAAGGTTACCCGCG
AAGGCCAAAAGGGTGAAAAAACAATTACAACACCTACACTGAAAAATCCGCTGACCGGTGTT
ATTATTAGTAAGGGTGAGCCGAAAGAGGAAATTACCAAAGATCCGATCAACGAATTAACCGA
ATACGGTCCGGAAACAATCACCCCGGGCCATCGCGATGAATTTGATCCTAAATTACCGACAG
GCGAGAAAGAGGAAGTGCCGGGCAAGCCTGGTATTAAGAATCCGGAGACAGGTGATGTTGT
GCGCCCGCCGGTTGATAGCGTGACAAAGTATGGCCCTGTGAAGGGCGACAGCATCGTGGA
AAAAGAGGAGATCCCGTTCAAAAAGGAGCGCAAATTTAATCCGGACTTAGCACCGGGTACAG
AGAAGGTTACCCGTGAAGGTCAAAAAGGCGAGAAGACCATTACCACCCCTACCTTAAAGAAC
CCTCTGACAGGTGAAATCATCAGTAAAGGCGAAAGCAAAGAAGAGATCACCAAAGATCCGAT
CAATGAGTTAACAGAATATGGTCCGGAGACAATCACACCGGGCCACCGCGATGAGTTTGAC
CCGAAGCTGCCGACAGGTGAAAAAGAAGAGGTGCCTGGCAAACCGGGTATCAAGAACCCG
GAGACCGGTGATGTTGTTCGCCCTCCGGTGGATAGTGTGACAAAATACGGTCCTGTGAAGG
GTGATAGTATTGTTGAAAAGGAAGAAATTCCGTTTGAAAAAGAGCGTAAGTTCAATCCTGATT

TAGCCCCTGGCACAGAGAAAGTTACACGCGAAGGTCAGAAAGGTGAGAAAACAATCACCAC
CCCGACCCTGAAGAATCCTTTAACCGGCGAAATCATCAGTAAGGGCGAGAGTAAGGAAGAA
ATTACAAAAGACCCTATTAATGAATTAACAGAGTATGGTCCTGAAACAATCACACCTGGCCAC
CGTGACGAATTCGATCCGAAACTGCCTACCGGTGAAAAAGAAGAGGTTCCTGGCAAGCCTG
GCATTAAGAACCCGGAAACCGGCGATGTGGTGCGTCCGCCTGTGGACAGTGTTACAAAATA
TGGCCCGGTGAAAGGCGATAGCATTGTGGAAAAAGAGGAGATTCCTTTTAAGAAGGAGCGT
AAATTCAACCCTGACCTGGCCCCGGGTACAGAAAAGGTGACCCGCGAGGGCCAAAAGGGC
GAAAAAACCATCACCACACCGACATTAAAAAACCCTTTAACAGGCGAGATCATCAGCAAAGG
TGAGAGCAAAGAAGAAATCACCAAAGACCCGATTAATGAACTGACAGAGTACGGCCCTGAGA
CAATCACCCCTGGTCACCGCGACGAGTTCGACCCTAAGTTACCGACCGGTGAAAAAGAAGA
AGTTCCGGGTAAACCTGGCATCAAGAATCCTGAGACCGGCGACGTTGTTCGCCCTCCGGTG
GATAGCGTGACCAAATATGGTCCGGTTAAAGGTGACAGTATCGTGGAGAAGGAAGAGATTC
CTTTCGAGAAAGAGCGCAAGTTTAATCCGGACCTGGCCCCTGGCACCGAGAAAGTTACACG
CGAGGGCCAGAAGGGTGAAAAGACCATCACAACCCCTACCCTGAAGAACCCGTTAACCGGT
GAAATTATCAGCAAGGGTGAAAGTAAAGAGGAGATCACCAAAGATCCTATCAATGAGCTCAC
CGAGTATGGTCCGGAGACCTGCTGCTAA

7.1.2 H_6-TEV-SasG-ΔEE

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGAAAACCTGTACTTTCAGGGATCCGC
ACCTAAGACCATCACCGAGCTGGCGAAGAAAGTTGAAGAGATTCCTTTTAAAAAAGAACGTA
AGTTCAACCCGGATCTGGCACCTGGTACCGAGAAGGTGACACGTGAGGGCCAGAAAGGCG
AGAAGACAATCACAACACCTACCTTAAAAAATCCTTTAACAGGCGTGATTATCAGTAAAGGTG
AACCGAAAGAGGAGATTACCAAAGACCCGATCAACGAGCTGACAGAGTATGGCCCGGAAAC
CATCGCACCGGGCCATCGTGATGAGTTTGATCCGAAGTTACCTACCGGTGCGAAGGAAGAA
GTTCCTGGTAAGCCGGGTATTAAAAACCCTGAGACAGGCGATGTGGTTCGTCCGCCTGTTG
ACAGCGTTACCAAATACGGTCCTGTTAAAGGCGACAGTATTGTGGCAAAAGAGGAAATCCCG
TTCGAAAAAGAACGCAAATTTAATCCTGATTTAGCACCGGGCACCGAAAAAGTGACCCGTGA
GGGTCAAAAAGGTGAGAAGACCATTACAACCCCTACACTGAAAAACCCGCTGACCGGCGAG
ATCATTAGCAAGGGTGAGAGTAAGGAAGAGATCACAAAGGACCCTATTAACGAACTGACCGA
ATACGGCCCGGAAACCATTACCCCGGGTCACCGTGACGAGTTCGATCCTAAGCTGCCTACA
GGCGCAAAGGAAGAAGTGCCTGGTAAACCGGGCATCAAGAACCCTGAAACCGGCGACGTG
GTTCGTCCGCCGGTTGACAGTGTTACCAAGTACGGCCCGGTGAAGGGCGATAGCATCGTTG
CGAAGGAAGAAATCCCGTTTGAGAAAGAGCGCAAATTCAATCCGGATCTGGCCCCTGGTAC
CGAGAAGGTTACCCGCGAAGGCCAAAAGGGTGAAAAAACAATTACAACACCTACACTGAAAA
ATCCGCTGACCGGTGTTATTATTAGTAAGGGTGAGCCGAAAGAGGAAATTACCAAAGATCCG
ATCAACGAATTAACCGAATACGGTCCGGAAACAATCACCCCGGGCCATCGCGATGAATTTGA
TCCTAAATTACCGACAGGCGCGAAAGAGGAAGTGCCGGGCAAGCCTGGTATTAAGAATCCG
GAGACAGGTGATGTTGTGCGCCCGCCGGTTGATAGCGTGACAAAGTATGGCCCTGTGAAGG
GCGACAGCATCGTGGCAAAAGAGGAGATCCCGTTCAAAAAGGAGCGCAAATTTAATCCGGA
CTTAGCACCGGGTACAGAGAAGGTTACCCGTGAAGGTCAAAAAGGCGAGAAGACCATTACC
ACCCCTACCTTAAAGAACCCTCTGACAGGTGAAATCATCAGTAAAGGCGAAAGCAAAGAAGA
GATCACCAAAGATCCGATCAATGAGTTAACAGAATATGGTCCGGAGACAATCACACCGGGCC
ACCGCGATGAGTTTGACCCGAAGCTGCCGACAGGTGCAAAAGAAGAGGTGCCTGGCAAACC
GGGTATCAAGAACCCGGAAACCGGTGATGTTGTTCGCCCTCCGGTGGATAGTGTGACAAAA
TACGGTCCTGTGAAGGGTGATAGTATTGTTGCAAAGGAAGAAATTCCGTTTGAAAAAGAGCG
TAAGTTCAATCCTGATTTAGCCCCTGGCACAGAGAAAGTTACACGCGAAGGTCAGAAAGGTG
AGAAAACAATCACCACCCCGACCCTGAAGAATCCTTTAACCGGCGAAATCATCAGTAAGGGC
GAGAGTAAGGAAGAAATTACAAAAGACCCTATTAATGAATTAACAGAGTATGGTCCTGAAACA
ATCACACCTGGCCACCGTGACGAATTCGATCCGAAACTGCCTACCGGTGCAAAAGAAGAGG
TTCCTGGCAAGCCTGGCATTAAGAACCCGGAAACCGGCGATGTGGTGCGTCCGCCTGTGGA
CAGTGTTACAAAATATGGCCCGGTGAAAGGCGATAGCATTGTGGCAAAAGAGGAGATTCCTT
TTAAGAAGGAGCGTAAATTCAACCCTGACCTGGCCCCGGGTACAGAAAAGGTGACCCGCGA
GGGCCAAAAGGGCGAAAAAACCATCACCACACCGACATTAAAAAACCCTTTAACAGGCGAGA
TCATCAGCAAAGGTGAGAGCAAAGAAGAAATCACCAAAGACCCGATTAATGAACTGACAGAG
TACGGCCCTGAGACAATCACCCCTGGTCACCGCGACGAGTTCGACCCTAAGTTACCGACCG

GTGCAAAAGAAGAAGTTCCGGGTAAACCTGGCATCAAGAATCCTGAAACCGGCGACGTTGTT
CGCCCTCCGGTGGATAGCGTGACCAAATATGGTCCGGTTAAAGGTGACAGTATCGTGGCGA
AGGAAGAGATTCCTTTCGAGAAAGAGCGCAAGTTTAATCCGGACCTGGCCCCTGGCACCGA
GAAAGTTACACGCGAGGGCCAGAAGGGTGAAAAGACCATCACAACCCCTACCCTGAAGAAC
CCGTTAACCGGTGAAATTATCAGCAAGGGTGAAAGTAAAGAGGAGATCACCAAAGATCCTAT
CAATGAGCTCACCGAGTATGGTCCGGAAACCTGTTGTTGA

7.1.3 pWT (E-G5²)₅

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCGGAAACCATCGCACCGGGCC
ATCGTGATGAGTTTGATCCGAAGTTACCTACCGGTGAGAAGGAAGAAGTTCCTGGTAAGCCG
GGTATTAAAAACCCTGAGACAGGCGATGTGGTTCGTCCGCCTGTTGACAGCGTTACCAAATA
CGGTCCTGTTAAAGGCGACAGTATTGTGGAAAAAGAGGAAATCCCGTTCGAAAAAGAACGCA
AATTTAATCCTGATTTAGCACCGGGCACCGAAAAAGTGACCCGTGAGGGTCAAAAAGGTGAG
AAGACCATTACAACCCCTACACTGAAAAACCCGCTGACCGGCGAGATCATTAGCAAGGGTGA
GAGTAAGGAAGAGATCACAAAGGACCCTATTAACGAACTGACCGAATACGGCCCGGAAACC
CTGAGCGTGGGCGCGACCATTGGCCCGGAAACCATCGCACCGGGCCATCGTGATGAGTTT
GATCCGAAGTTACCTACCGGTGAGAAGGAAGAAGTTCCTGGTAAGCCGGGTATTAAAAACCC
TGAGACAGGCGATGTGGTTCGTCCGCCTGTTGACAGCGTTACCAAATACGGTCCTGTTAAAG
GCGACAGTATTGTGGAAAAAGAGGAAATCCCGTTCGAAAAAGAACGCAAATTTAATCCTGAT
TTAGCACCGGGCACCGAAAAAGTGACCCGTGAGGGTCAAAAAGGTGAGAAGACCATTACAA
CCCCTACACTGAAAAACCCGCTGACCGGCGAGATCATTAGCAAGGGTGAGAGTAAGGAAGA
GATCACAAAGGACCCTATTAACGAACTGACCGAATACGGCCCGGAAACCACCGTTATTGGTC
TGGCGAGCGGCCCGGAAACCATCGCACCGGGCCATCGTGATGAGTTTGATCCGAAGTTACC
TACCGGTGAGAAGGAAGAAGTTCCTGGTAAGCCGGGTATTAAAAACCCTGAGACAGGCGAT
GTGGTTCGTCCGCCTGTTGACAGCGTTACCAAATACGGTCCTGTTAAAGGCGACAGTATTGT
GGAAAAAGAGGAAATCCCGTTCGAAAAAGAACGCAAATTTAATCCTGATTTAGCACCGGGCA
CCGAAAAAGTGACCCGTGAGGGTCAAAAAGGTGAGAAGACCATTACAACCCCTACACTGAAA
AACCCGCTGACCGGCGAGATCATTAGCAAGGGTGAGAGTAAGGAAGAGATCACAAAGGACC
CTATTAACGAACTGACCGAATACGGCCCGGAAACCGCGCTGAGCGGCACCATTGTGGGCCC
GGAAACCATCGCACCGGGCCATCGTGATGAGTTTGATCCGAAGTTACCTACCGGTGAGAAG
GAAGAAGTTCCTGGTAAGCCGGGTATTAAAAACCCTGAGACAGGCGATGTGGTTCGTCCGC
CTGTTGACAGCGTTACCAAATACGGTCCTGTTAAAGGCGACAGTATTGTGGAAAAAGAGGAA
ATCCCGTTCGAAAAAGAACGCAAATTTAATCCTGATTTAGCACCGGGCACCGAAAAAGTGAC
CCGTGAGGGTCAAAAAGGTGAGAAGACCATTACAACCCCTACACTGAAAAACCCGCTGACC
GGCGAGATCATTAGCAAGGGTGAGAGTAAGGAAGAGATCACAAAGGACCCTATTAACGAAC
TGACCGAATACGGCCCGGAAACCGTTATTACCGGTAGCCTGGCGGGCCCGGAAACCATCGC
ACCGGGCCATCGTGATGAGTTTGATCCGAAGTTACCTACCGGTGAGAAGGAAGAAGTTCCT
GGTAAGCCGGGTATTAAAAACCCTGAGACAGGCGATGTGGTTCGTCCGCCTGTTGACAGCG
TTACCAAATACGGTCCTGTTAAAGGCGACAGTATTGTGGAAAAAGAGGAAATCCCGTTCGAA
AAAGAACGCAAATTTAATCCTGATTTAGCACCGGGCACCGAAAAAGTGACCCGTGAGGGTCA
AAAAGGTGAGAAGACCATTACAACCCCTACACTGAAAAACCCGCTGACCGGCGAGATCATTA
GCAAGGGTGAGAGTAAGGAAGAGATCACAAAGGACCCTATTAACGAACTGACCGAATACGG
CCCGGAAACCTGTTGTTGA

7.2 Protein Sequences

The amino acid sequences outlined here are after purification and have undergone cleavage with TEV (SasG and SasG-ΔEE) or methionine excision (pWT (E-G5²)₅ and variants thereof). Amino acid substitutions coloured in red and linker residues in purple.

7.2.1 SasG

GSAPKTITELEKKVEEIPFKKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPK
EEITKDPINELTEYGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGP
VKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKD
PINELTEYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSI
VEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELT
EYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEE
IPFKKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPE
TITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKE
RKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETITPGH
RDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFKKERKFNP
DLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETITPGHRDEFD
PKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPG
TEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.2 SasG-ΔEE

GSAPKTITELAKKVEEIPFKKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPK
EEITKDPINELTEYGPETIAPGHRDEFDPKLPTGAKEEVPGKPGIKNPETGDVVRPPVDSVTKYGP
VKGDSIVAKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKD
PINELTEYGPETITPGHRDEFDPKLPTGAKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSI
VAKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGVIISKGEPKEEITKDPINELT
EYGPETITPGHRDEFDPKLPTGAKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVAKEE
IPFKKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPE
TITPGHRDEFDPKLPTGAKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVAKEEIPFEKE
RKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETITPGH
RDEFDPKLPTGAKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVAKEEIPFKKERKFNP
DLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETITPGHRDEFD
PKLPTGAKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVAKEEIPFEKERKFNPDLAPG
TEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3 pWT $(E\text{-}G5^2)_5$ and variants

7.2.3.1 pWT $(E\text{-}G5^2)_5$

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPV
KGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDP
INELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVT
KYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKE
EITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRP
PVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIIS
KGESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPET
GDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKN
PLTGEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEVPGKP
GIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTIT
TPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.2 I502P

GSSHHHHHHSSGPETPAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGP
VKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKD
PINELTEYGPETLSVGATIGPETPAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDS

VTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGES
KEEITKDPINELTEYGPETTVIGLASGPETPAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVV
RPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGE
IISKGESKEEITKDPINELTEYGPETALSGTIVGPETPAPGHRDEFDPKLPTGEKEEVPGKPGIKNP
ETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLK
NPLTGEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETPAPGHRDEFDPKLPTGEKEEVPGK
PGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKT
ITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.3 G517A

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTAEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPV
KGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDP
INELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTAEKEEVPGKPGIKNPETGDVVRPPVDSVT
KYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKE
EITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTAEKEEVPGKPGIKNPETGDVVRPP
VDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISK
GESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTAEKEEVPGKPGIKNPETGD
VVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLT
GEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTAEKEEVPGKPGIKN
PETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTL
KNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.4 V522P

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEPPGKPGIKNPETGDVVRPPVDSVTKYGPV
KGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDP
INELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEPPGKPGIKNPETGDVVRPPVDSVT
KYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKE
EITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEPPGKPGIKNPETGDVVRP
PVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIIS
KGESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEPPGKPGIKNPET
GDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKN
PLTGEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEPPGKP
GIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTIT
TPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.5 G524A

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEVPAKPGIKNPETGDVVRPPVDSVTKYGPV
KGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDP
INELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEVPAKPGIKNPETGDVVRPPVDSVT
KYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKE
EITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEVPAKPGIKNPETGDVVRPP
VDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISK
GESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEVPAKPGIKNPETGD
VVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLT
GEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEVPAKPGIKN
PETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTL
KNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.6 G527A

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEVPGKPAIKNPETGDVVRPPVDSVTKYGPV
KGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDP
INELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEVPGKPAIKNPETGDVVRPPVDSVT
KYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKE
EITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEVPGKPAIKNPETGDVVRPP
VDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISK
GESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEVPGKPAIKNPETGD
VVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLT
GEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEVPGKPAIKN
PETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTL
KNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.7 P540A

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPAVDSVTKYGPV
KGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDP
INELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPAVDSVT
KYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKE
EITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRP
AVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIIS
KGESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPET
GDVVRPAVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKN
PLTGEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEVPGKP
GIKNPETGDVVRPAVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTIT
TPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.8 P549A

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGAV
KGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDP
INELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVT
KYGAVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKE
EITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRP
PVDSVTKYGAVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIIS
KGESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPET
GDVVRPPVDSVTKYGAVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKN
PLTGEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEVPGKP
GIKNPETGDVVRPPVDSVTKYGAVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTIT
TPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.9 V550P

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPP
KGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDP
INELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVT
KYGPPKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKE
EITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRP
PVDSVTKYGPPKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIIS
KGESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPET
GDVVRPPVDSVTKYGPPKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKN
PLTGEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEVPGKP

GIKNPETGDVVRPPVDSVTKYGPPKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTIT
TPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.10 V556P

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPV
KGDSIPEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDP
INELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVT
KYGPVKGDSIPEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKE
EITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRP
PVDSVTKYGPVKGDSIPEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIIS
KGESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPET
GDVVRPPVDSVTKYGPVKGDSIPEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKN
PLTGEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEVPGKP
GIKNPETGDVVRPPVDSVTKYGPVKGDSIPEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTIT
TPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.11 P562A

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPV
KGDSIVEKEEIAFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDP
INELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVT
KYGPVKGDSIVEKEEIAFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKE
EITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRP
PVDSVTKYGPVKGDSIVEKEEIAFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIIS
KGESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPET
GDVVRPPVDSVTKYGPVKGDSIVEKEEIAFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKN
PLTGEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEVPGKP
GIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIAFEKERKFNPDLAPGTEKVTREGQKGEKTIT
TPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.12 V580P

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPV
KGDSIVEKEEIPFEKERKFNPDLAPGTEKPTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDP
INELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVT
KYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKPTREGQKGEKTITTPTLKNPLTGEIISKGESKE
EITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRP
PVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKPTREGQKGEKTITTPTLKNPLTGEIIS
KGESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPET
GDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKPTREGQKGEKTITTPTLKN
PLTGEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEVPGKP
GIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKPTREGQKGEKTIT
TPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.13 G584A

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPV
KGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREAQKGEKTITTPTLKNPLTGEIISKGESKEEITKDP
INELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVT
KYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREAQKGEKTITTPTLKNPLTGEIISKGESKE
EITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRP
PVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREAQKGEKTITTPTLKNPLTGEIIS

KGESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPET
GDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREAQKGEKTITTPTLKNP
LTGEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEVPGKPGI
KNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREAQKGEKTITT
PTLKNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.14 G587A

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPV
KGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKAEKTITTPTLKNPLTGEIISKGESKEEITKDP
INELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVT
KYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKAEKTITTPTLKNPLTGEIISKGESKE
EITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRP
PVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKAEKTITTPTLKNPLTGEIIS
KGESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPET
GDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKAEKTITTPTLKNP
LTGEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEVPGKPGI
KNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKAEKTITT
PTLKNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.15 E588K

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPV
KGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGKKTITTPTLKNPLTGEIISKGESKEEITKDP
INELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVT
KYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGKKTITTPTLKNPLTGEIISKGESKE
EITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRP
PVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGKKTITTPTLKNPLTGEIIS
KGESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPET
GDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGKKTITTPTLKN
PLTGEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEVPGKP
GIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGKKTIT
TPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.16 K589E

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPV
KGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEETITTPTLKNPLTGEIISKGESKEEITKDP
INELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVT
KYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEETITTPTLKNPLTGEIISKGESKE
EITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRP
PVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEETITTPTLKNPLTGEIIS
KGESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPET
GDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEETITTPTLKN
PLTGEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEVPGKP
GIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEETIT
TPTLKNPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.17 N598A

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPV
KGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKAPLTGEIISKGESKEEITKDP
INELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVT
KYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKAPLTGEIISKGESKE

EITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRP
PVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKAPLTGEIIS
KGESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPET
GDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKA
PLTGEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEVPGKP
GIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTIT
TPTLKAPLTGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.18 T601A

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPV
KGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLAGEIISKGESKEEITKD
PINELTEYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSV
TKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLAGEIISKGESK
EEITKDPINELTEYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVR
PPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLAGEII
SKGESKEEITKDPINELTEYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPET
GDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKN
PLAGEIISKGESKEEITKDPINELTEYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEVPGKP
GIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTIT
TPTLKNPLAGEIISKGESKEEITKDPINELTEYGPETCC

7.2.3.19 E624K

GSSHHHHHHSSGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPV
KGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDP
INELTKYGPETLSVGATIGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVT
KYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKE
EITKDPINELTKYGPETTVIGLASGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRP
PVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIIS
KGESKEEITKDPINELTKYGPETALSGTIVGPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPET
GDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKN
PLTGEIISKGESKEEITKDPINELTKYGPETVITGSLAGPETIAPGHRDEFDPKLPTGEKEEVPGKP
GIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTIT
TPTLKNPLTGEIISKGESKEEITKDPINELTKYGPETCC

7.3 Unfolding Data

7.3.1 SasG-ΔEE

	SasG-ΔEE E				
Speed (nms^{-1})	n	Mode Rupture force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	133 123 90	245.0 232.3 253.9	**243.8**	149.8 148.2 145.5	**147.8**
800	81 182 137	261.5 260.3 276.3	**266.0**	153.5 147.7 148.0	**149.7**
1500	117 75 88	266.8 280.0 282.2	**276.3**	150.8 147.2 147.5	**148.5**
3000	92 63 98	300.2 273.0 284.0	**285.7**	147.5 152.0 150.6	**150.0**
5000	101 121 83	294.7 283.4 299.0	**292.4**	149.1 150.2 153.9	**151.1**

Table 7.1 – Summary of rupture force statistics for SasG-ΔEE E sub domain mechanical unfolding in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode rupture force is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed. SD is the sample standard deviation of the triplicate mode values at each speed.

	SasG-ΔEE G5				
Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	103 39 31	401.0 360.5 403.6	**388.4**	214.2 214.6 214.1	**214.3**
800	79 69 57	411.3 391.8 405.9	**403.0**	217.8 215.3 214.4	**215.8**
1500	61 49 44	413.4 417.9 416.4	**415.9**	216.3 216.9 215.1	**216.1**
3000	90 38 64	435.2 422.2 423.9	**427.1**	215.0 220.4 216.0	**217.1**
5000	68 77 79	441.2 410.0 430.7	**427.3**	214.4 216.7 216.9	**216.0**

Table 7.2 – Summary of rupture force statistics for SasG-ΔEE G5 sub domain mechanical unfolding in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode rupture force is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

7.3.2 SasG with Zn^{2+}

	SasG – 100 μM $ZnCl_2$ E				
Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	58 100	238.1 222.1	**230.1**	148.3 152.0	**150.2**
800	112 61	258.2 246.7	**252.5**	148.5 155.8	**152.2**
1500	101 122	277.9 255.5	**266.7**	148.7 152.5	**150.6**
3000	86 76	285.6 267.3	**276.4**	147.4 145.6	**146.5**
5000	95 140	287.4 277.7	**282.5**	147.9 152.0	**149.9**

Table 7.3 – Summary of rupture force statistics for SasG E sub domain mechanical unfolding in 1X TBS (20 mM Tris, 150 mM NaCl), pH 7.4 supplemented with 100 µM $ZnCl_2$ at room temperature. n is the number of peaks used for analysis. Mode rupture force is obtained from the Gaussian fits to the histograms. Average is the mean of the mode values at each speed.

	SasG – 100 μM $ZnCl_2$ G5				
Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	37 90	403.0 364.8	**383.9**	215.4 217.6	**216.5**
800	118 86	420.1 394.5	**407.3**	215.1 218.3	**216.7**
1500	88 149	424.5 404.1	**414.3**	215.9 217.0	**216.5**
3000	78 94	442.1 404.8	**423.4**	215.4 214.9	**215.2**
5000	55 235	446.9 414.6	**430.7**	213.1 217.0	**215.1**

Table 7.4 – Summary of rupture force statistics for SasG G5 sub domain mechanical unfolding in 1X TBS (20 mM Tris, 150 mM NaCl), pH 7.4 supplemented with 100 µM $ZnCl_2$ at room temperature. n is the number of peaks used for analysis. Mode rupture force is obtained from the Gaussian fits to the histograms. Average is the mean of the mode values at each speed.

7.3.3 Variants

Variant	Retraction velocity (nms^{-1})	n	Modal Rupture Force (pN)	SD (pN)
V522P				
	200	87	402.6	± 51.1
	5000	114	459.7	± 58.6
V550P				
	200	49	408.4	± 59.3
	5000	149	454.1	± 62.9

Table 7.5 – Table displaying the number of events, modal rupture force of 1 repeat of $G5^2$ unfolding in V522P and V550P at retraction velocities of 200 and 5000 nms^{-1}. Carried out in 1X PBS, pH 7.4 at room temperature.

	V556P E				
Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	44	240.7	**219.0**	144.8	**148.2**
	154	211.2		148.7	
	50	205.2		151.0	
800	99	241.0	**230.4**	143.9	**149.1**
	129	227.2		146.7	
	57	223.0		156.5	
1500	90	245.4	**230.5**	146.3	**152.2**
	150	227.5		146.2	
	82	218.7		164.1	
3000	28	255.0	**237.0**	146.8	**149.1**
	91	227.6		150.8	
	56	246.3		149.7	
5000	86	267.8	**254.0**	143.8	**148.4**
	73	246.7		149.1	
	50	247.6		152.4	

Table 7.6 – V556P E sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	**V556P G5²** n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	27	389.1	**384.8**	215.8	**216.9**
	177	382.0		216.7	
	62	383.4		218.2	
800	61	413.0	**408.0**	214.7	**216.5**
	105	404.6		215.6	
	127	406.5		219.0	
1500	86	438.8	**431.0**	214.4	**215.5**
	133	428.3		214.6	
	206	425.7		217.5	
3000	104	463.7	**444.5**	215.9	**215.7**
	134	440.7		216.5	
	78	429.1		214.7	
5000	42	469.8	**447.5**	209.5	**213.6**
	78	436.8		214.6	
	66	435.9		216.7	

Table 7.7 – V556P G5² sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	**V580P E** n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	202	216.8	**221.5**	145.7	**145.4**
	97	235.9		144.5	
	176	211.6		145.9	
800	176	233.2	**237.3**	146.5	**146.6**
	117	244.8		145.8	
	157	234.0		147.5	
1500	223	237.7	**243.8**	145.8	**146.7**
	124	255.8		146.2	
	130	237.8		148.0	
3000	128	242.6	**250.6**	148.4	**147.3**
	102	261.4		146.0	
	210	239.8		147.4	
5000	145	244.0	**254.1**	147.6	**146.7**
	81	273.5		144.6	
	160	244.7		148.0	

Table 7.8 – V580P E sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	**V580P $G5^2$** n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	125 62 112	353.8 376.0 343.5	**357.8**	214.3 217.3 216.1	**215.9**
800	76 73 101	403.9 397.4 391.5	**397.6**	214.3 214.0 215.1	**214.5**
1500	131 77 91	400.9 423.7 407.7	**410.8**	212.9 214.1 215.8	**214.3**
3000	87 68 163	418.1 444.8 418.9	**427.3**	213.4 214.4 215.4	**214.4**
5000	87 70 128	430.4 458.1 427.3	**438.6**	214.9 216.0 214.8	**215.2**

Table 7.9 – V580P $G5^2$ sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	**V580P $G5^2$** Repeat	FWHM (pN)	Average (pN)	**pWT $(E\text{-}G5^2)_5$ $G5^2$** FWHM* (pN)
200	1 2 3	101.1 122.7 113.8	**112.5**	**56.2**
800	1 2 3	52.0 85.5 64.8	**67.5**	**58.3**
1500	1 2 3	78.3 60.1 74.4	**71.0**	**54.1**
3000	1 2 3	70.7 90.9 66.5	**76.0**	**60.5**
5000	1 2 3	86.4 82.3 87.4	**85.4**	**58.6**

Table 7.10 – FWHM values for V580P $G5^2$ sub domain mechanical unfolding in 1X PBS, pH 7.4 at room temperature. FWHM from the gaussian fits of the triplicate repeats and average is the mean of these. SD: standard deviation (sample). The values were generally higher than for pWT $(E\text{-}G5^2)_5$, which may indicate deviation in the kinetic parameters of mechanical unfolding.

Speed (nms^{-1})	G524A E n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	110	188.1	**183.0**	144.5	**145.3**
	42	176.0		148.9	
	56	184.8		142.6	
800	103	206.7	**201.3**	143.3	**146.2**
	75	204.6		148.8	
	113	192.5		146.6	
1500	155	212.8	**207.2**	145.6	**145.5**
	148	197.1		147.1	
	95	211.7		143.6	
3000	115	214.2	**200.5**	146.3	**146.4**
	135	196.9		146.4	
	83	204.1		146.5	
5000	74	224.7	**215.7**	146.8	**147.7**
	96	211.3		150.3	
	94	211.0		146.1	

Table 7.11 – G524A E sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	G524A $G5^2$ n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	70	388.0	**397.0**	217.4	**215.4**
	67	402.2		215.5	
	20	400.7		213.3	
800	66	413.0	**419.5**	214.1	**214.5**
	107	427.9		215.7	
	72	417.5		213.8	
1500	127	427.9	**428.2**	216.3	**215.0**
	147	422.9		215.3	
	55	433.7		213.5	
3000	77	431.9	**435.2**	216.2	**215.0**
	140	431.4		215.1	
	58	442.2		213.7	
5000	45	445.2	**447.6**	216.8	**215.6**
	110	443.9		215.1	
	66	453.8		214.9	

Table 7.12 – G524A $G5^2$ sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

	G527A E				
Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	79 67 38	175.0 165.7 170.4	**170.4**	145.9 143.8 144.6	**144.8**
800	115 64 119	197.1 183.4 198.0	**192.8**	143.0 143.3 150.5	**145.6**
1500	167 127 69	205.7 190.1 195.0	**196.9**	145.0 144.8 147.5	**145.8**
3000	152 129 62	210.0 199.9 218.1	**209.3**	145.1 143.3 150.4	**146.3**
5000	123 123 67	214.8 204.3 202.9	**207.4**	146.6 146.0 148.5	**147.0**

Table 7.13 – G527A E sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

	G527A $G5^2$				
Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	45 46 47	420.9 403.3 406.4	**410.2**	215.4 216.2 217.7	**216.4**
800	37 41 108	442.2 412.0 423.4	**425.9**	212.6 216.0 215.0	**214.5**
1500	58 62 71	459.9 436.3 426.5	**440.9**	214.6 215.1 214.2	**214.6**
3000	40 33 46	476.3 448.6 450.5	**458.5**	214.6 213.9 215.7	**214.7**
5000	34 65 58	478.0 444.2 448.8	**457.0**	213.3 212.7 214.3	**213.5**

Table 7.14 – G527A $G5^2$ sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	**G584A E** n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	127	221.9	**226.0**	144.1	**144.8**
	128	231.6		145.5	
	187	224.4		144.6	
800	144	235.2	**240.2**	145.5	**145.2**
	117	242.5		146.1	
	145	242.9		143.9	
1500	162	242.7	**242.6**	146.7	**146.1**
	174	248.2		145.6	
	145	236.9		146.1	
3000	134	249.1	**251.5**	146.5	**145.4**
	119	250.8		146.2	
	151	252.3		143.5	
5000	175	255.2	**254.4**	145.5	**146.7**
	94	254.7		148.5	
	109	253.4		146.1	

Table 7.15 – G584A E sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	**G584A G5^2** n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	61	344.4	**355.5**	215.7	**215.1**
	105	358.8		214.9	
	143	363.3		214.5	
800	68	373.4	**384.0**	215.9	**215.0**
	109	386.2		214.7	
	100	392.3		214.3	
1500	133	388.4	**389.0**	215.3	**215.9**
	129	381.8		216.5	
	86	396.8		215.8	
3000	56	398.6	**400.1**	215.7	**214.9**
	95	398.0		214.8	
	51	403.7		214.3	
5000	93	404.0	**407.1**	215.6	**215.5**
	98	407.4		217.4	
	66	409.9		213.6	

Table 7.16 – G584A G5^2 sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

	G587A E				
Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	53 81 97	216.3 218.2 225.6	**220.0**	146.2 144.5 140.7	**143.8**
800	77 62 111	215.7 230.9 229.0	**225.2**	146.9 145.1 144.5	**145.5**
1500	43 94 73	227.9 234.4 247.4	**236.6**	145.0 144.7 143.4	**144.3**
3000	57 72 68	230.9 243.8 250.8	**241.8**	144.8 147.9 144.2	**145.6**
5000	59 78 115	228.5 246.3 259.6	**244.8**	144.7 144.1 143.8	**144.2**

Table 7.17 – G587A E sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

	G587A $G5^2$				
Speed (nms^{-1})	n	Mode Rupture force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	44 62 74	300.2 290.5 308.1	**299.6**	216.4 214.7 215.5	**215.5**
800	54 44 79	300.5 313.7 335.9	**316.7**	210.9 217.0 214.3	**214.0**
1500	26 55 48	321.5 328.1 345.9	**331.9**	212.9 212.9 212.2	**212.6**
3000	37 53 48	318.5 329.9 353.0	**333.8**	217.6 211.9 212.8	**214.1**
5000	53 57 58	327.6 328.4 363.0	**339.7**	214.6 215.0 214.1	**214.6**

Table 7.18 – G587A $G5^2$ sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	**P540A E** n	Mode Rupture force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	45 47 65	215.3 221.3 224.6	**220.4**	148.6 145.0 145.8	**146.5**
800	83 56 91	224.6 237.4 241.8	**234.6**	147.6 149.8 146.0	**147.8**
1500	89 76 133	247.3 234.9 251.4	**244.6**	153.4 152.1 145.6	**150.4**
3000	74 73 131	229.6 249.5 244.7	**247.1**	147.9 147.4 147.4	**147.6**
5000	86 108 197	246.7 254.5 253.2	**251.5**	147.0 147.5 146.6	**147.1**

Table 7.19 – P540A E sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	**P540A G5^2** n	Mode Rupture force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	49 21 55	410.0 411.4 403.4	**408.3**	216.1 215.6 217.2	**216.3**
800	78 49 49	416.7 426.9 429.9	**424.5**	216.0 214.1 214.7	**214.9**
1500	114 83 99	434.9 435.8 446.4	**439.0**	217.0 216.1 215.4	**216.2**
3000	70 57 84	428.2 440.7 445.9	**438.3**	214.9 215.3 215.6	**215.2**
5000	87 56 147	448.1 458.5 456.0	**454.2**	215.5 213.4 214.8	**214.6**

Table 7.20 – P540A G5^2 sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

	P562A E				
Speed (nms^{-1})	n	Mode Rupture force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	95 96 56	223.3 229.6 232.4	**228.4**	144.4 143.0 146.5	**144.6**
800	168 142 73	242.8 234.4 234.5	**237.2**	143.3 145.8 151.5	**146.9**
1500	144 107 92	239.1 240.5 241.5	**240.4**	147.7 145.5 149.0	**147.4**
3000	197 171 92	251.3 245.4 243.8	**244.6**	146.8 144.9 150.4	**147.4**
5000	142 158 72	258.0 254.2 254.0	**255.4**	147.6 144.8 149.5	**147.3**

Table 7.21 – P562A E sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

	P562A $G5^2$				
Speed (nms^{-1})	n	Mode Rupture force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	47 46 45	365.5 361.1 369.2	**365.3**	216.3 215.7 216.9	**216.3**
800	84 66 50	398.6 384.1 384.9	**389.2**	214.1 213.6 217.3	**215.0**
1500	101 60 63	398.7 395.3 396.5	**396.8**	215.4 214.3 216.2	**215.3**
3000	73 60 63	407.5 404.6 403.8	**405.3**	215.1 213.3 215.4	**214.6**
5000	55 57 45	411.0 415.3 418.3	**414.9**	211.9 213.2 215.1	**213.4**

Table 7.22 – P562A $G5^2$ sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	**G517A E** n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	89	192.9	**196.6**	148.8	**146.8**
	78	193.9		143.9	
	41	202.9		147.7	
800	87	209.8	**206.7**	145.7	**147.0**
	53	200.2		147.5	
	69	210.0		147.8	
1500	64	227.2	**216.5**	143.3	**145.4**
	83	207.5		146.7	
	70	214.8		146.1	
3000	81	221.8	**210.8**	147.4	**146.0**
	67	212.0		143.7	
	69	209.6		146.8	
5000	100	235.5	**228.7**	149.6	**147.6**
	81	221.0		142.6	
	46	229.4		150.6	

Table 7.23 – G517A E sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	**G517A $G5^2$** n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	105	391.2	**405.5**	216.0	**214.9**
	33	415.3		214.9	
	33	409.9		213.9	
800	66	422.4	**415.4**	215.2	**215.1**
	29	400.5		214.6	
	63	423.4		215.3	
1500	70	440.3	**426.1**	214.9	**214.5**
	58	408.7		214.6	
	82	429.2		214.1	
3000	69	435.9	**436.0**	216.0	**215.2**
	40	431.2		215.3	
	66	440.8		214.1	
5000	90	445.9	**446.6**	215.8	**215.2**
	55	439.2		215.4	
	50	454.6		214.4	

Table 7.24 – G517A $G5^2$ sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	P549A E n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	69	185.2	**190.2**	145.0	**148.3**
	64	202.5		145.6	
	60	183.0		154.4	
800	131	189.5	**200.8**	148.7	**148.0**
	143	212.9		145.6	
	82	200.2		149.7	
1500	61	196.9	**209.5**	146.9	**147.1**
	156	220.5		146.9	
	219	211.2		147.4	
3000	90	208.2	**215.1**	148.4	**149.4**
	120	226.0		146.1	
	71	211.0		153.8	
5000	71	219.7	**223.8**	147.1	**150.1**
	64	235.4		146.9	
	44	216.1		156.5	

Table 7.25 – P549A E sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	P549A G5² n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	20	411.6	**411.1**	213.9	**219.3**
	41	427.5		214.1	
	82	394.1		229.9	
800	70	422.2	**426.9**	214.9	**215.2**
	76	442.1		215.1	
	81	416.3		215.7	
1500	70	431.6	**443.4**	214.8	**215.3**
	78	460.3		214.3	
	136	438.1		216.7	
3000	38	452.2	**450.4**	215.4	**218.3**
	97	463.4		215.0	
	76	435.6		224.5	
5000	43	452.7	**456.8**	214.6	**218.1**
	54	484.2		214.8	
	88	433.5		224.9	

Table 7.26 – P549A G5² sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	N598A E n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	74	191.5	**184.3**	144.7	**146.3**
	47	181.4		147.3	
	43	179.8		146.9	
800	42	203.5	**201.4**	150.4	**148.3**
	101	199.8		148.2	
	82	200.9		146.1	
1500	24	218.8	**216.0**	141.7	**144.9**
	64	216.8		145.5	
	71	212.3		147.5	
3000	78	225.0	**223.6**	146.5	**146.4**
	73	224.7		147.0	
	53	222.4		145.9	
5000	68	246.4	**238.2**	148.4	**147.5**
	116	240.3		147.8	
	53	227.8		146.3	

Table 7.27 – N598A E sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	N598A $G5^2$ n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	126	404.6	**391.6**	214.9	**214.4**
	65	400.3		213.8	
	50	370.1		214.4	
800	89	411.4	**405.9**	214.8	**214.0**
	96	402.6		214.3	
	82	403.5		213.0	
1500	73	435.1	**423.4**	216.1	**215.1**
	60	418.9		213.2	
	97	416.3		215.9	
3000	129	432.5	**429.7**	214.6	**214.5**
	61	435.8		213.6	
	53	420.8		215.4	
5000	82	448.2	**435.5**	215.2	**213.8**
	84	435.4		213.3	
	62	422.9		213.0	

Table 7.28 – N598A $G5^2$ sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

	T601A E				
Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	80	232.6	**222.5**	144.5	**146.8**
	61	228.7		150.4	
	48	206.2		145.6	
800	98	239.1	**240.8**	145.9	**145.7**
	161	242.5		146.3	
	100	240.9		145.0	
1500	105	249.3	**244.6**	147.8	**145.9**
	97	248.8		144.9	
	111	235.6		145.1	
3000	59	263.2	**253.3**	145.4	**146.7**
	107	256.8		149.1	
	81	239.9		145.5	
5000	84	259.2	**256.9**	146.6	**147.2**
	80	266.7		148.2	
	75	244.6		146.7	

Table 7.29 – T601A E sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

	T601A $G5^2$				
Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	32	405.8	**393.4**	214.1	**216.6**
	54	398.2		216.5	
	39	376.3		219.1	
800	57	419.9	**416.3**	214.8	**215.1**
	132	421.1		214.4	
	59	407.9		216.1	
1500	75	429.8	**420.0**	213.9	**213.8**
	59	424.4		214.3	
	57	405.7		213.2	
3000	46	444.3	**437.6**	211.5	**213.5**
	58	449.5		214.4	
	67	418.9		214.8	
5000	49	447.1	**443.2**	213.5	**213.2**
	47	452.6		212.4	
	53	430.1		213.8	

Table 7.30 – T601A $G5^2$ sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

	E588K E				
Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	70	212.3	**218.4**	146.2	**152.5**
	21	213.8		158.9	
	20	229.1		152.4	
800	59	220.4	**230.7**	146.1	**146.6**
	49	231.6		147.5	
	86	240.0		146.1	
1500	117	231.7	**242.6**	148.6	**148.8**
	64	247.2		146.2	
	40	249.0		151.6	
3000	66	237.7	**241.9**	148.2	**147.5**
	32	234.6		145.4	
	41	249.3		149.0	
5000	41	252.1	**256.1**	144.0	**151.5**
	29	254.2		157.5	
	48	262.0		152.9	

Table 7.31 – E588K E sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

	E588K $G5^2$				
Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	89	408.0	**403.5**	216.6	**218.6**
	49	407.3		217.7	
	52	395.3		221.7	
800	60	410.5	**417.9**	212.4	**214.5**
	48	425.6		215.5	
	58	417.6		215.5	
1500	133	437.9	**431.9**	215.2	**215.7**
	60	426.2		214.7	
	53	431.5		217.2	
3000	68	447.9	**444.3**	214.1	**216.5**
	31	438.0		215.4	
	52	446.9		219.9	
5000	45	457.7	**456.2**	218.3	**218.6**
	62	465.5		218.2	
	79	445.5		219.4	

Table 7.32 – E588K $G5^2$ sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

	K589E E				
Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	95	221.2	**225.8**	146.5	**145.7**
	133	241.7		145.4	
	118	214.6		145.2	
800	130	234.9	**236.5**	145.2	**144.6**
	127	240.5		145.1	
	145	234.2		143.6	
1500	118	240.0	**242.3**	144.1	**143.9**
	112	247.4		144.2	
	152	239.6		143.4	
3000	119	241.5	**245.3**	145.2	**145.9**
	120	254.8		147.6	
	88	235.8		144.8	
5000	120	253.8	**255.5**	145.1	**145.7**
	104	259.2		145.4	
	106	253.4		146.6	

Table 7.33 – K589E E sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

	K589E G5²				
Speed (nms^{-1})	n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	96	308.3	**320.6**	216.8	**214.7**
	125	346.9		212.4	
	102	306.7		215.0	
800	51	334.1	**341.8**	212.4	**212.6**
	48	357.9		211.4	
	76	333.4		214.1	
1500	71	348.7	**357.7**	214.2	**213.8**
	103	374.6		213.3	
	94	350.0		213.7	
3000	47	367.0	**364.2**	215.3	**213.6**
	58	372.5		210.0	
	45	353.2		215.5	
5000	88	370.5	**371.7**	213.8	**213.9**
	54	378.4		213.3	
	67	366.1		214.5	

Table 7.34 – K589E G5² sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	**E624K E** n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	67	221.8	**225.1**	142.7	**143.1**
	166	228.7		143.0	
	142	224.9		143.5	
800	141	231.4	**236.5**	143.1	**143.4**
	235	242.2		143.6	
	247	235.8		143.4	
1500	170	234.8	**239.3**	144.3	**144.4**
	178	240.6		144.3	
	195	242.4		144.7	
3000	103	232.1	**244.4**	144.6	**145.2**
	117	247.5		145.9	
	191	241.3		145.0	
5000	83	241.2	**252.0**	143.9	**144.3**
	170	257.2		144.6	
	145	257.5		144.4	

Table 7.35 – E624K E sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

Speed (nms^{-1})	**E624K $G5^2$** n	Mode Rupture Force (pN)	Average (pN)	Mode ΔL_C (Å)	Average (Å)
200	47	264.5	**271.3**	214.9	**216.2**
	101	268.9		217.0	
	108	280.6		216.6	
800	96	315.2	**314.0**	216.8	**217.3**
	145	323.1		218.5	
	175	303.9		216.5	
1500	93	341.2	**332.3**	215.8	**216.2**
	101	322.2		216.1	
	134	333.5		216.6	
3000	65	343.6	**351.8**	216.9	**216.7**
	84	361.0		216.8	
	110	350.8		216.4	
5000	40	357.9	**368.9**	214.2	**214.9**
	90	373.5		216.2	
	92	375.3		214.3	

Table 7.36 – E624K $G5^2$ sub domain mechanical unfolding statistics. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode is obtained from the gaussian fits to the histograms. Average is the mean of the mode values at each speed.

	E624K – Rupture Force G5²			**pWT (E-G5²)₅ G5²**
Speed (nms⁻¹)	Repeat	FWHM (pN)	Average (pN)	FWHM (pN)
200	1 2 3	97.1 115.7 126.1	**113.0**	**56.2**
800	1 2 3	119.7 105.2 127.4	**117.4**	**58.3**
1500	1 2 3	100.8 122.2 102.2	**108.4**	**54.1**
3000	1 2 3	126.5 133.8 107.0	**122.4**	**60.5**
5000	1 2 3	126.8 122.9 113.9	**121.2**	**58.6**

Table 7.37 – Summary of rupture force FWHM statistics for E624K vs pWT $(E\text{-}G5^2)_5$ $G5^2$ sub domain mechanical unfolding. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode FWHM is obtained from the gaussian fits to the rupture force histograms. Average is the mean of the mode values at each speed. The FWHM is significantly larger for E624K than pWT $(E\text{-}G5^2)_5$ at every retraction velocity.

	E624K - ΔL_C G5²			**pWT (E-G5²)₅ G5²**
Speed (nms⁻¹)	Repeat	FWHM (Å)	Average (Å)	Average FWHM (Å)
200	1 2 3	39.2 31.8 29.9	**33.6**	**12.3**
800	1 2 3	22.5 29.1 31.5	**27.7**	**14.9**
1500	1 2 3	20.7 29.5 26.0	**25.4**	**14.4**
3000	1 2 3	18.7 23.4 30.5	**24.2**	**14.3**
5000	1 2 3	31.3 25.7 29.6	**28.9**	**16.8**

Table 7.38 – Summary of ΔL_C FWHM statistics for E624K vs pWT $(E\text{-}G5^2)_5$ $G5^2$ sub domain mechanical unfolding. Carried out in 1X PBS, pH 7.4 at room temperature. n is the number of peaks used for analysis. Mode FWHM is obtained from the gaussian fits to the ΔL_C histograms. Average is the mean of the mode values at each speed. The FWHM is significantly larger for E624K than pWT $(E\text{-}G5^2)_5$ at every retraction velocity.

7.3.4 Example variant repeats

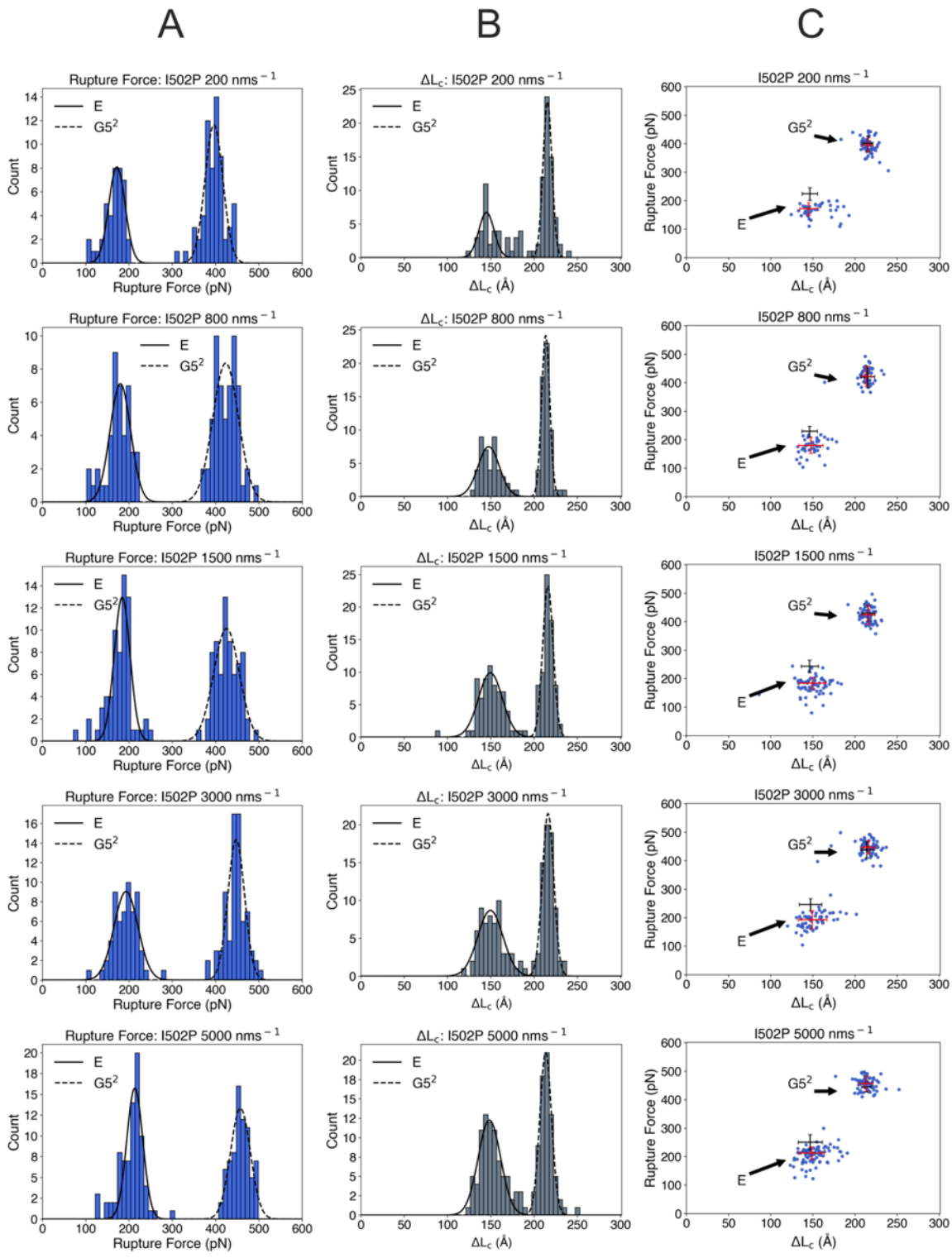

Figure 7.1 – Example I502P mechanical unfolding repeat at multiple retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Red scatterplot crosshairs are the mode and the FWHM from the corresponding histogram gaussian fits. Black crosshairs are the pWT $(E\text{-}G5^2)_5$ equivalent.

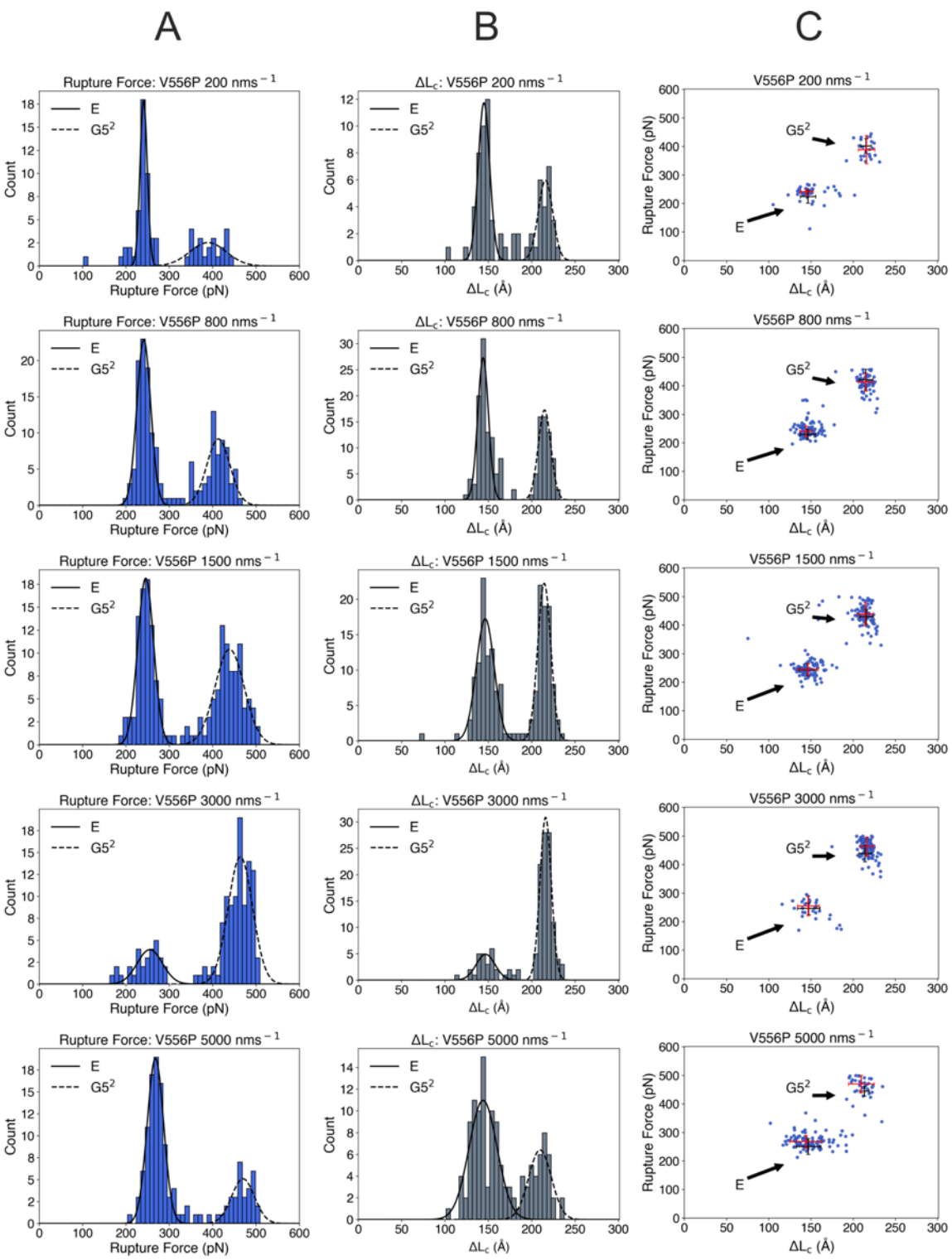

Figure 7.2 – Example V556P mechanical unfolding repeat at multiple retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Red scatterplot crosshairs are the mode and the FWHM from the corresponding histogram gaussian fits. Black crosshairs are the pWT $(E\text{-}G5^2)_5$ equivalent.

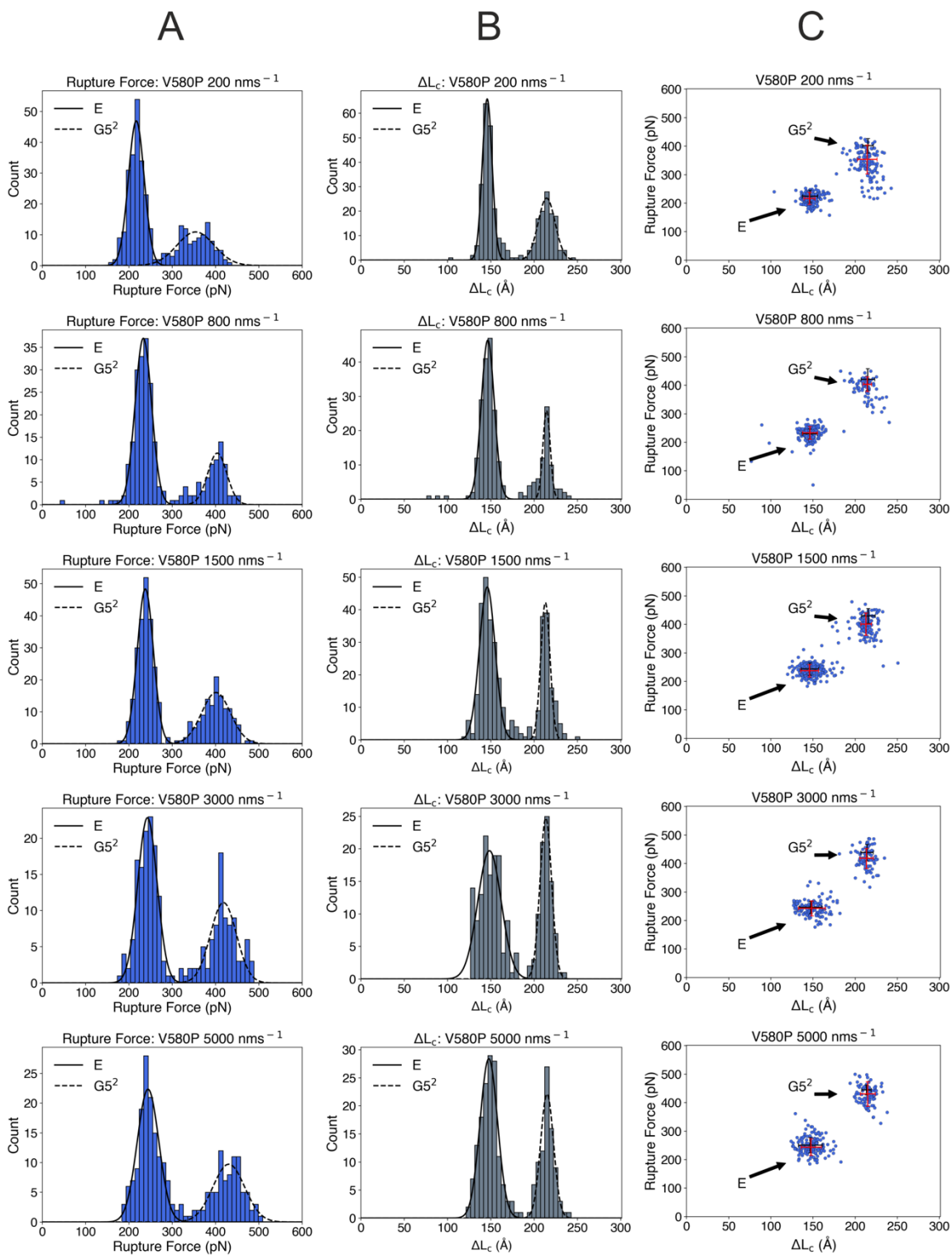

Figure 7.3 – Example V580P mechanical unfolding repeat at multiple retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Red scatterplot crosshairs are the mode and the FWHM from the corresponding histogram gaussian fits. Black crosshairs are the pWT $(E\text{-}G5^2)_5$ equivalent.

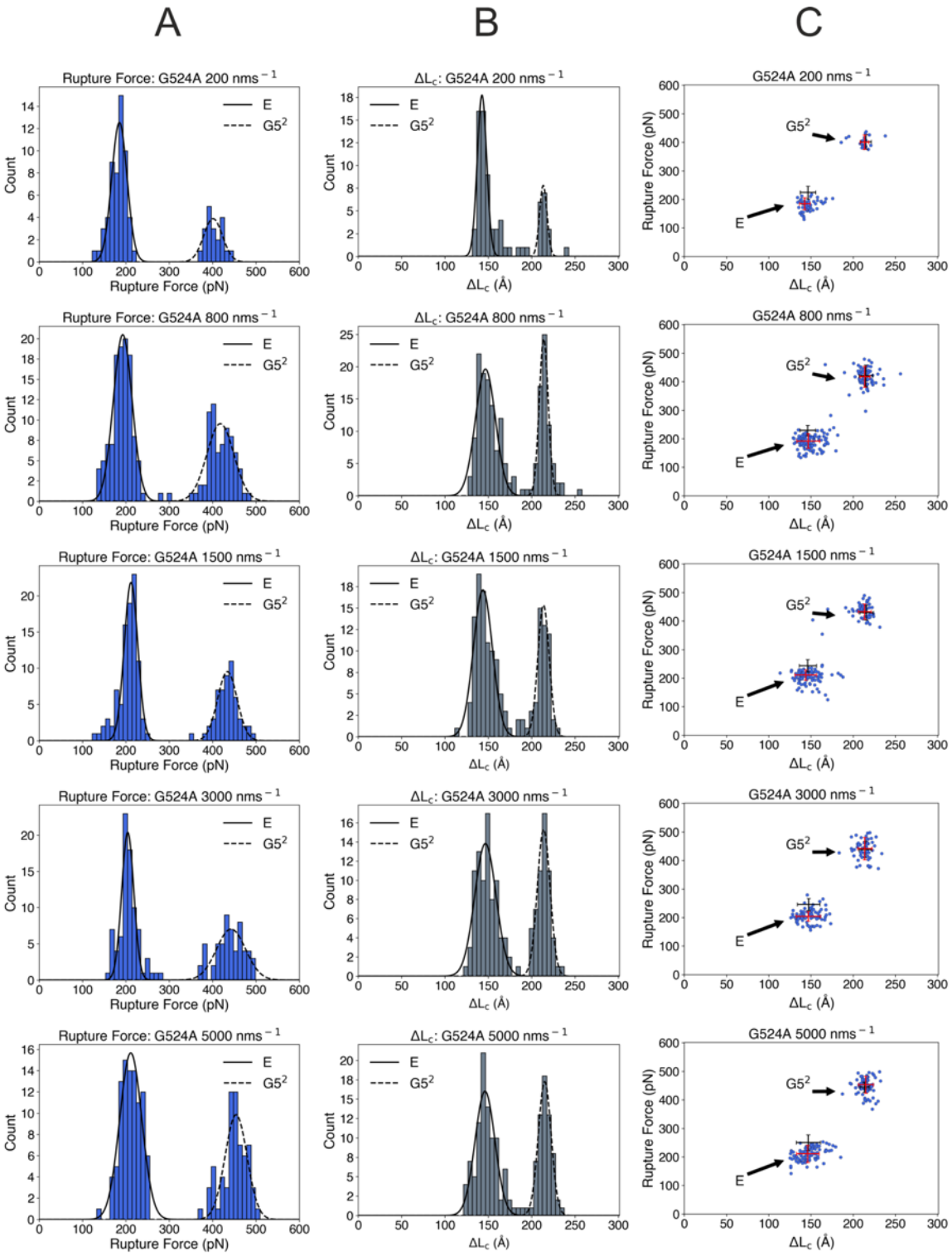

Figure 7.4 – Example G524A mechanical unfolding repeat at multiple retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Red scatterplot crosshairs are the mode and the FWHM from the corresponding histogram gaussian fits. Black crosshairs are the pWT $(E\text{-}G5^2)_5$ equivalent.

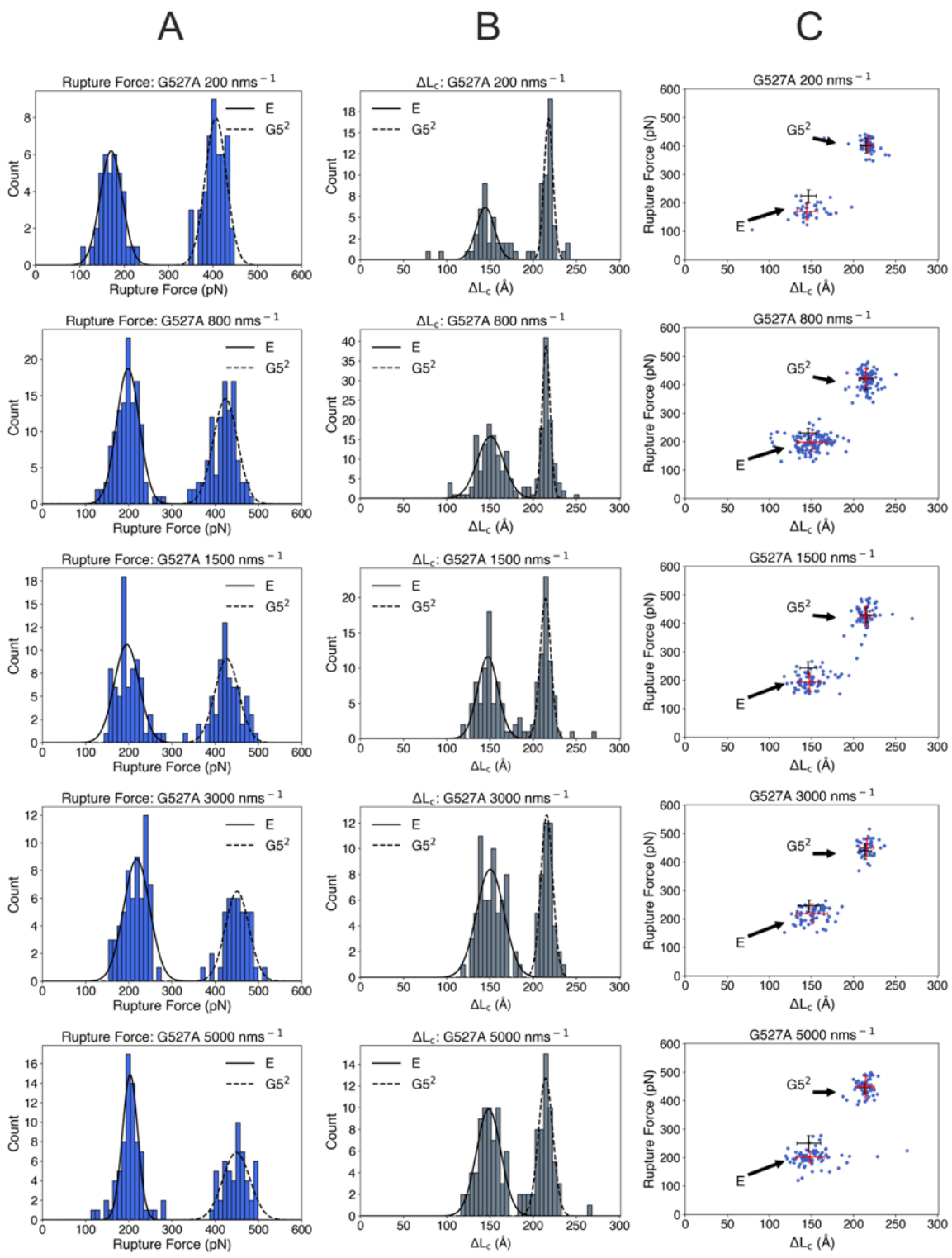

Figure 7.5 – Example G527A mechanical unfolding repeat at multiple retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Red scatterplot crosshairs are the mode and the FWHM from the corresponding histogram gaussian fits. Black crosshairs are the pWT $(E\text{-}G5^2)_5$ equivalent.

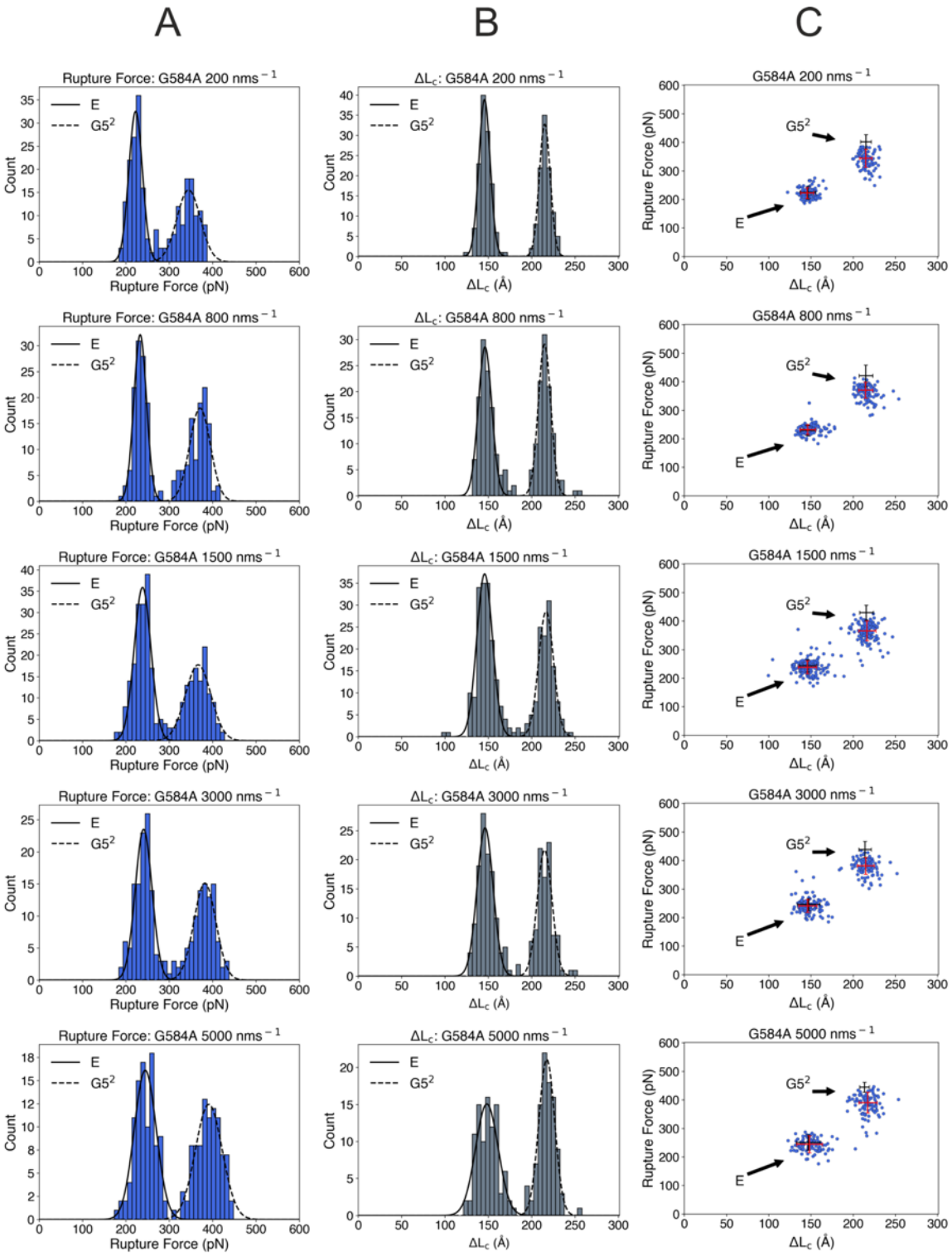

Figure 7.6 – Example G584A mechanical unfolding repeat at multiple retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Red scatterplot crosshairs are the mode and the FWHM from the corresponding histogram gaussian fits. Black crosshairs are the pWT $(E\text{-}G5^2)_5$ equivalent.

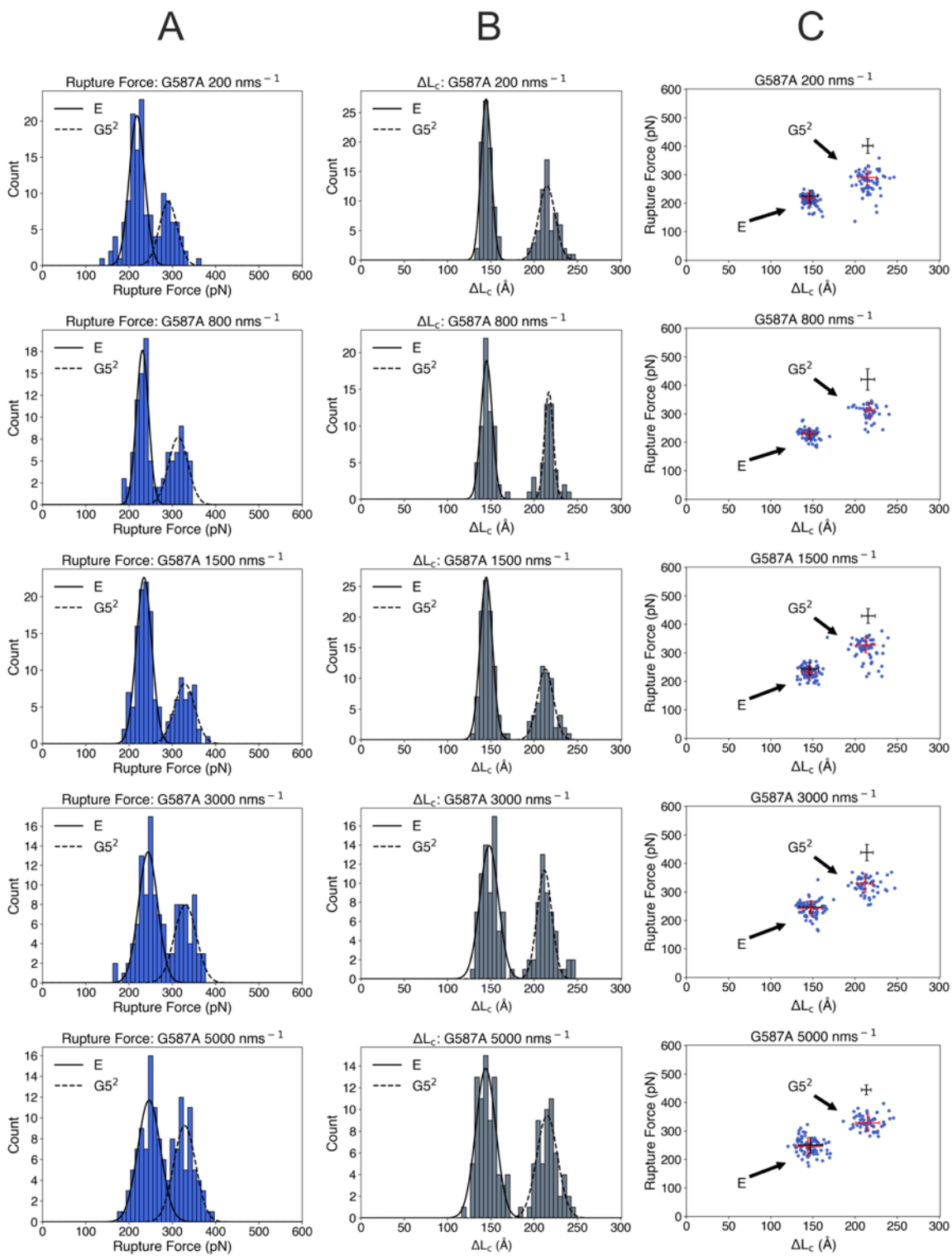

Figure 7.7 – Example G587A mechanical unfolding repeat at multiple retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Red scatterplot crosshairs are the mode and the FWHM from the corresponding histogram gaussian fits. Black crosshairs are the pWT $(E\text{-}G5^2)_5$ equivalent.

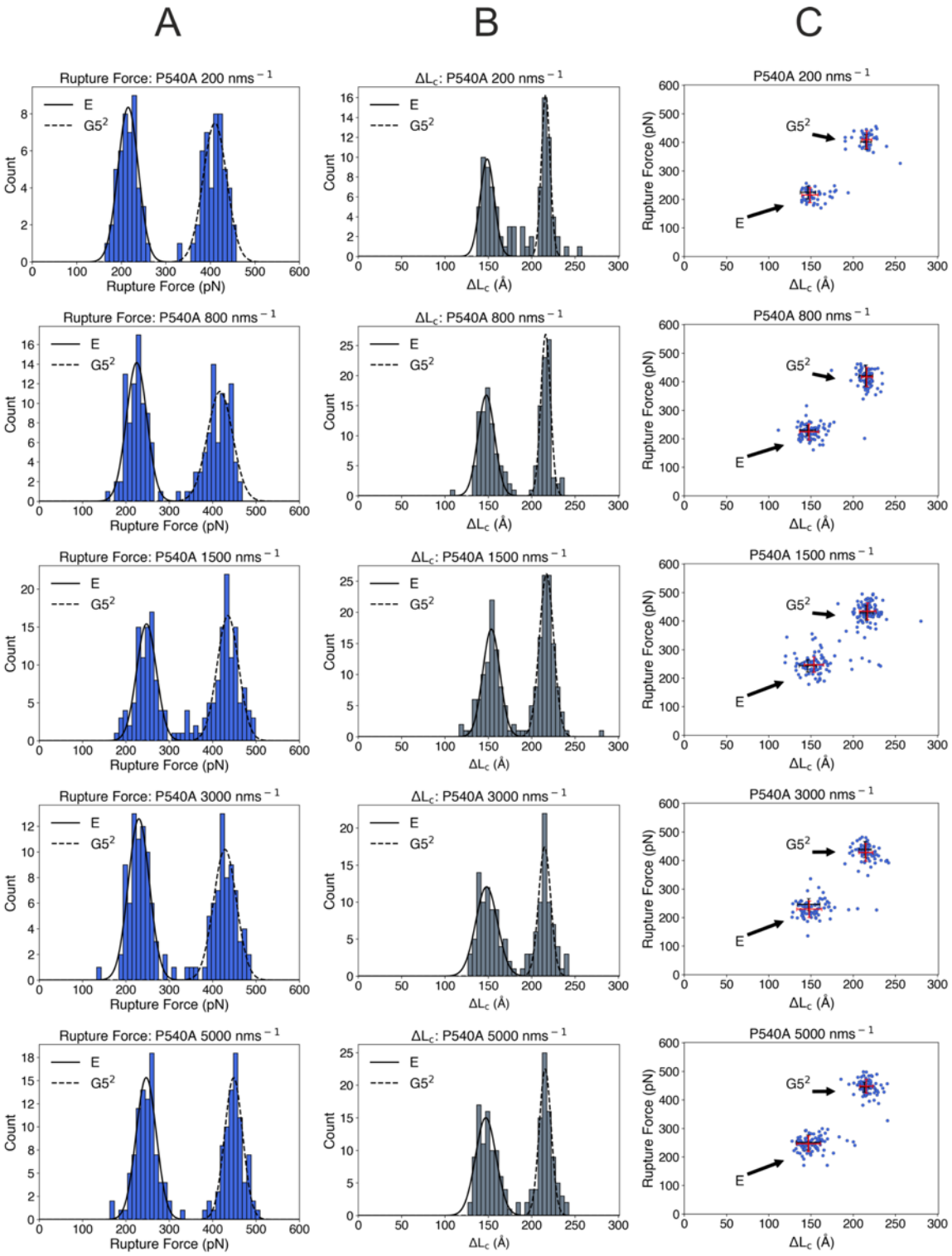

Figure 7.8 – Example P540A mechanical unfolding repeat at multiple retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Red scatterplot crosshairs are the mode and the FWHM from the corresponding histogram gaussian fits. Black crosshairs are the pWT $(E\text{-}G5^2)_5$ equivalent.

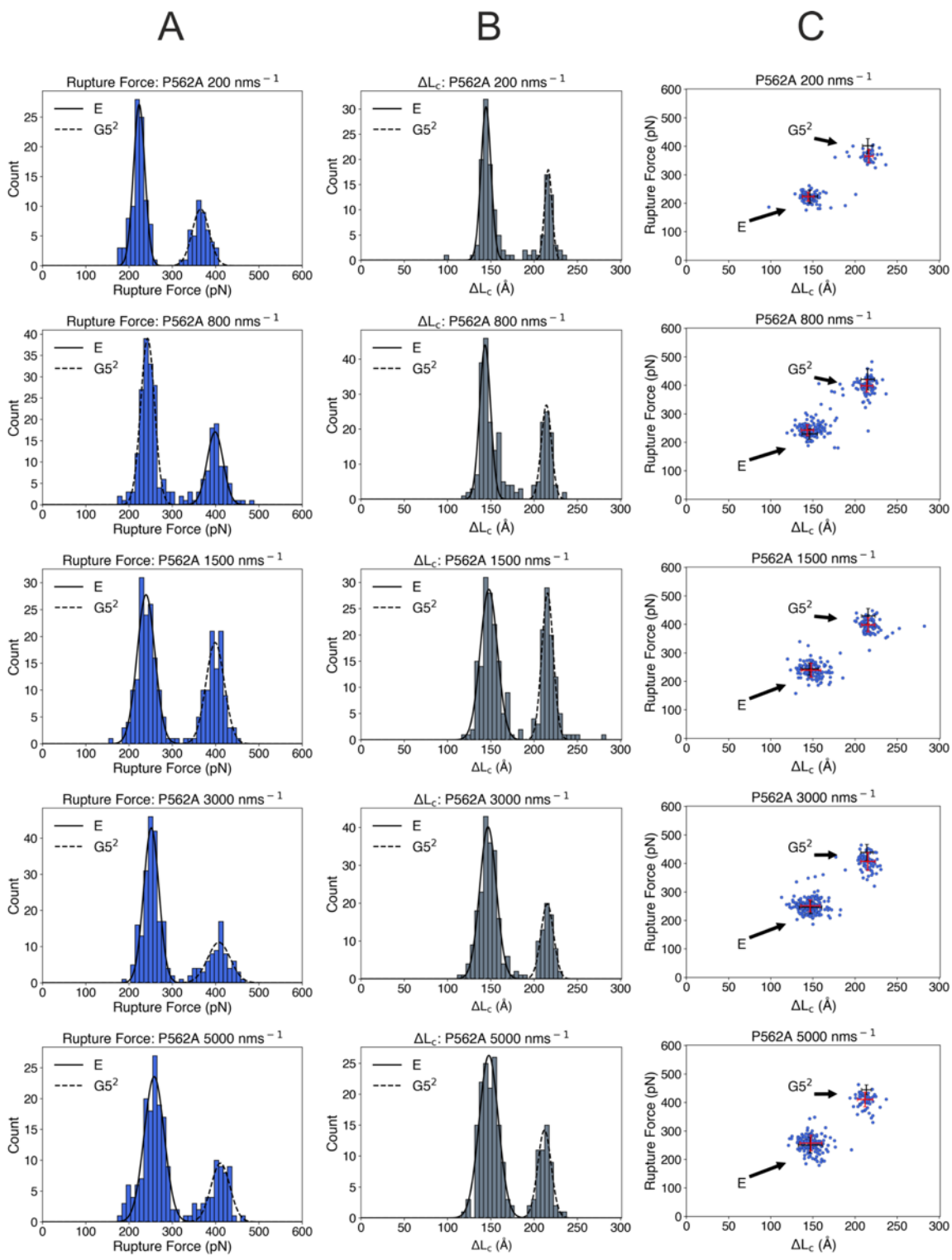

Figure 7.9 – Example P562A mechanical unfolding repeat at multiple retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Red scatterplot crosshairs are the mode and the FWHM from the corresponding histogram gaussian fits. Black crosshairs are the pWT $(E\text{-}G5^2)_5$ equivalent.

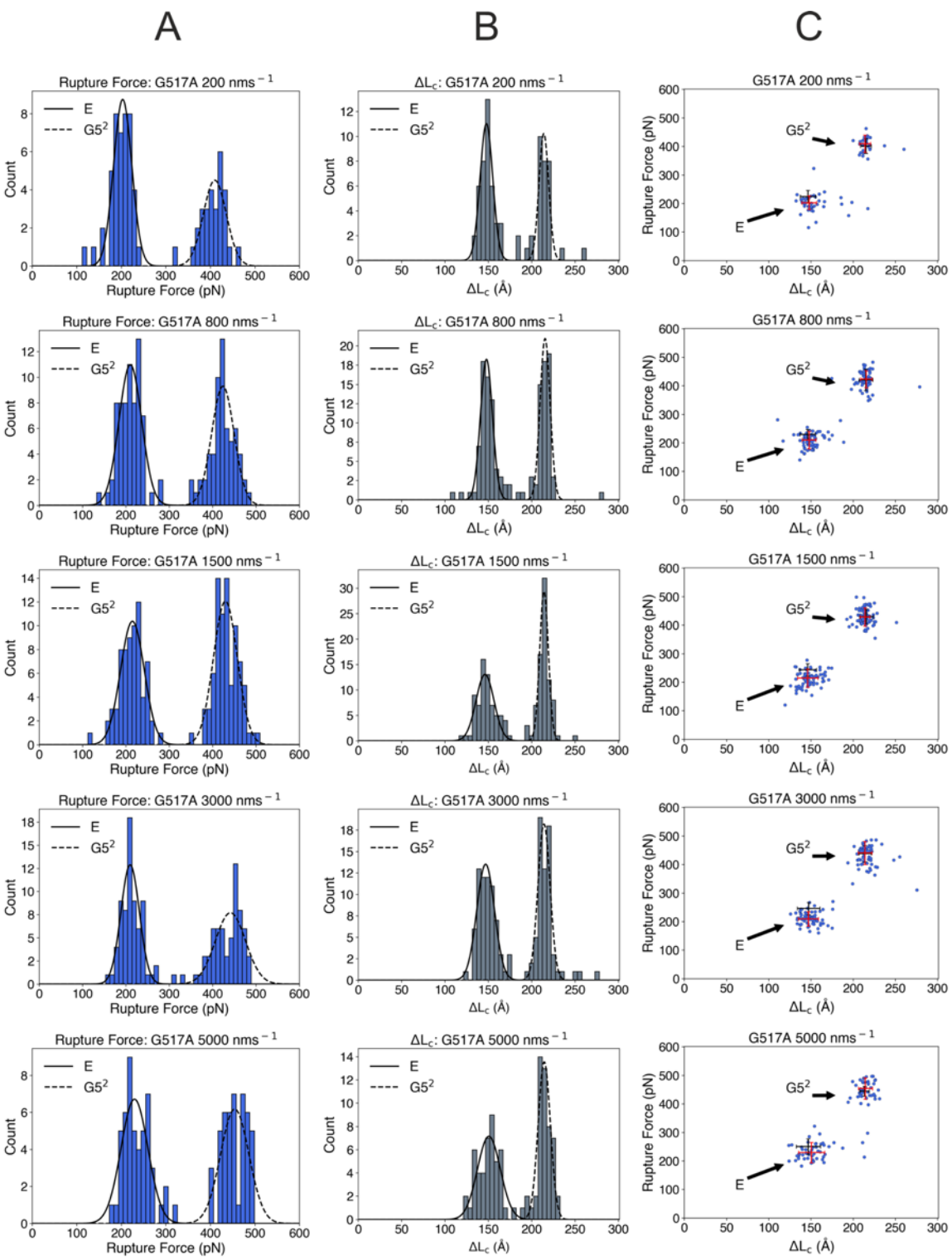

Figure 7.10 – Example G517A mechanical unfolding repeat at multiple retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Red scatterplot crosshairs are the mode and the FWHM from the corresponding histogram gaussian fits. Black crosshairs are the pWT $(E\text{-}G5^2)_5$ equivalent.

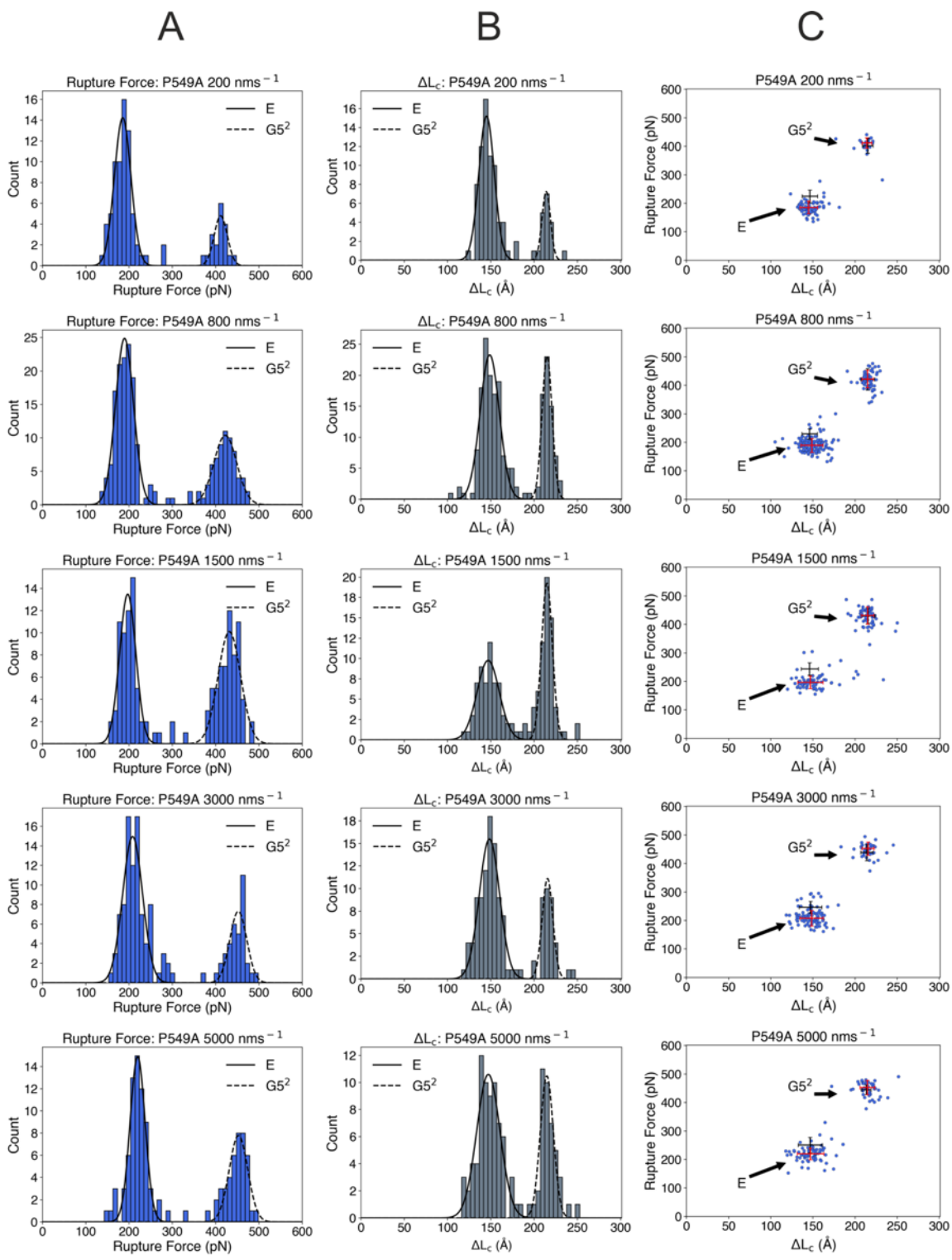

Figure 7.11 – Example P549A mechanical unfolding repeat at multiple retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Red scatterplot crosshairs are the mode and the FWHM from the corresponding histogram gaussian fits. Black crosshairs are the pWT $(E\text{-}G5^2)_5$ equivalent.

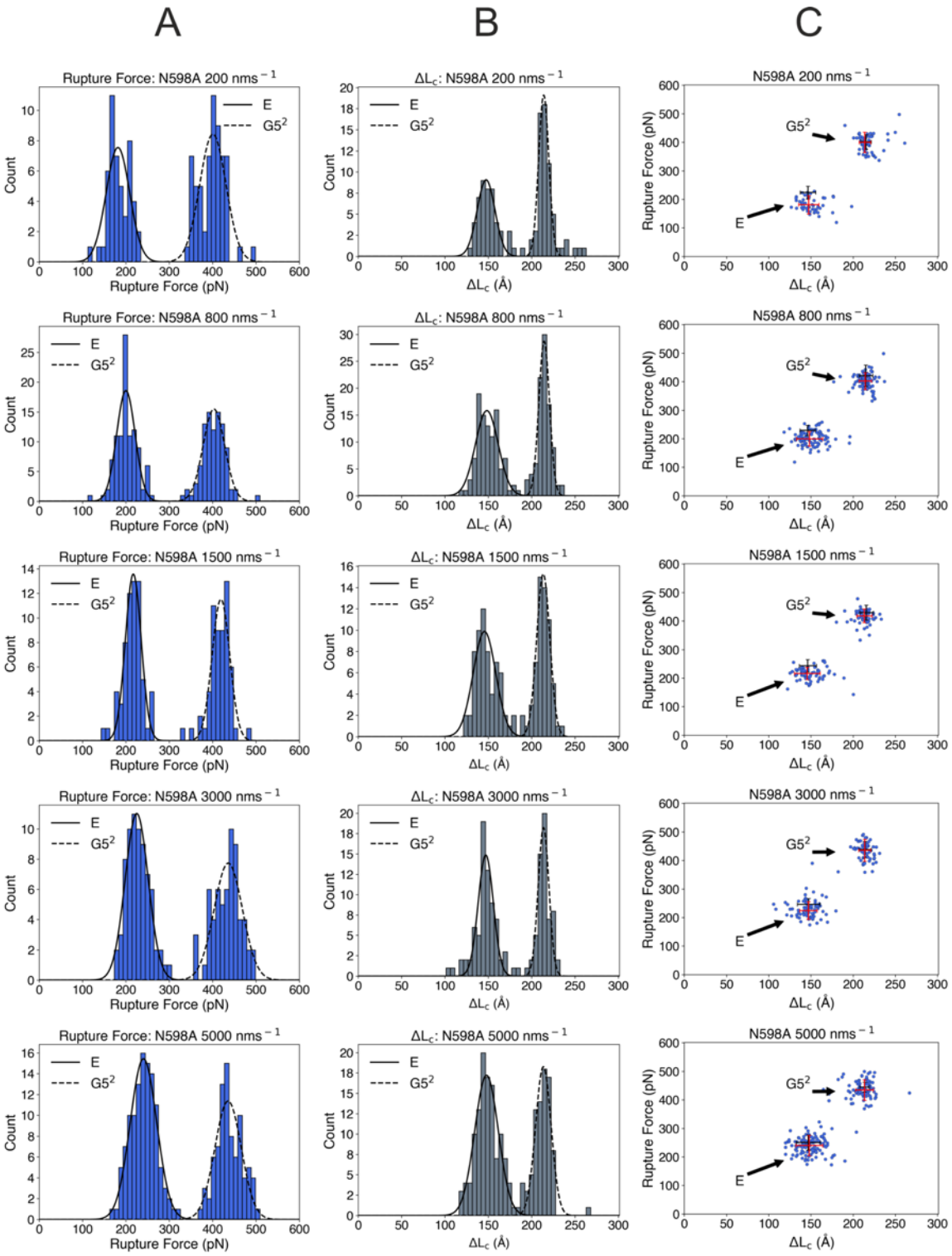

Figure 7.12 – Example N598A mechanical unfolding repeat at multiple retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Red scatterplot crosshairs are the mode and the FWHM from the corresponding histogram gaussian fits. Black crosshairs are the pWT $(E\text{-}G5^2)_5$ equivalent.

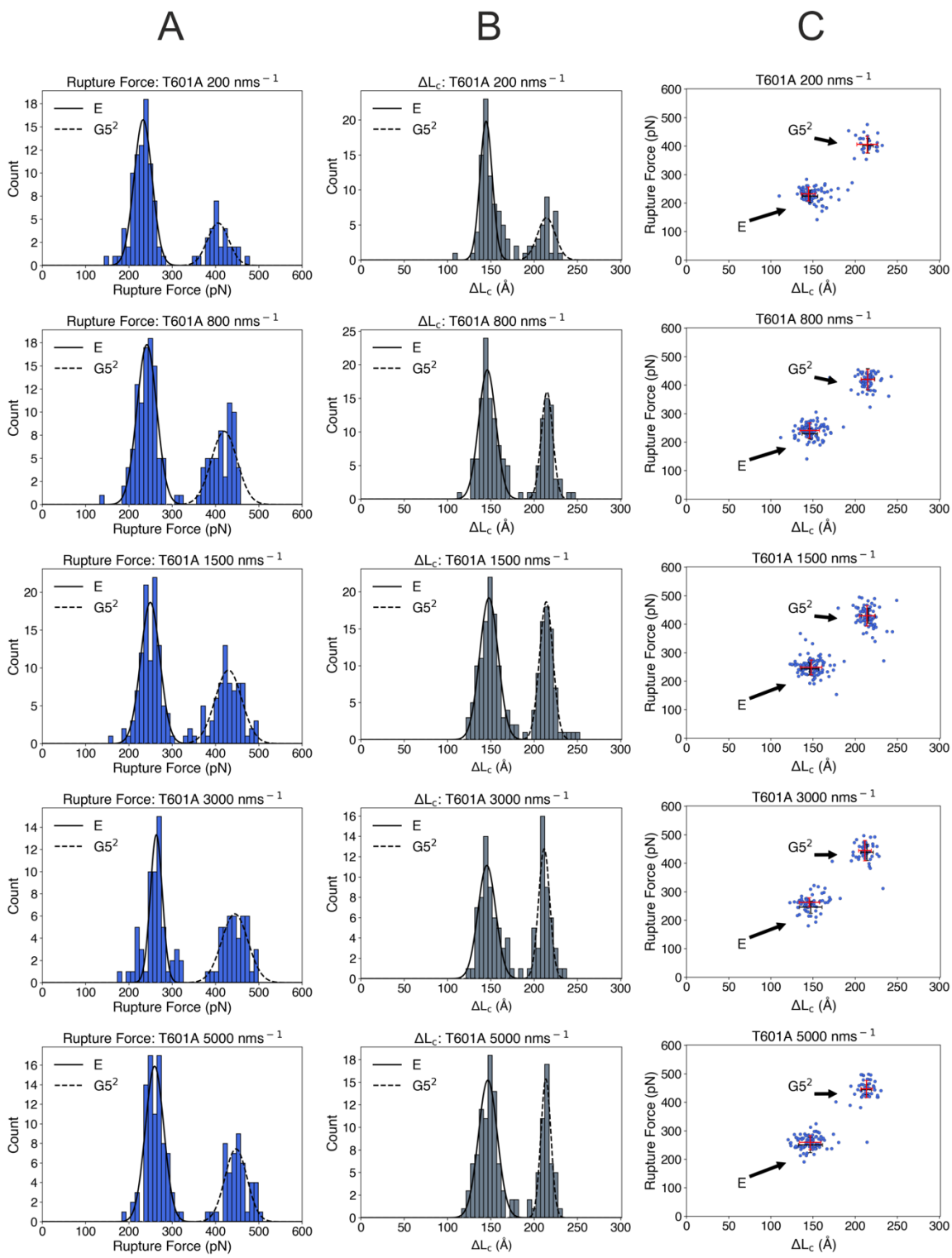

Figure 7.13 – Example T601A mechanical unfolding repeat at multiple retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Red scatterplot crosshairs are the mode and the FWHM from the corresponding histogram gaussian fits. Black crosshairs are the pWT $(E\text{-}G5^2)_5$ equivalent.

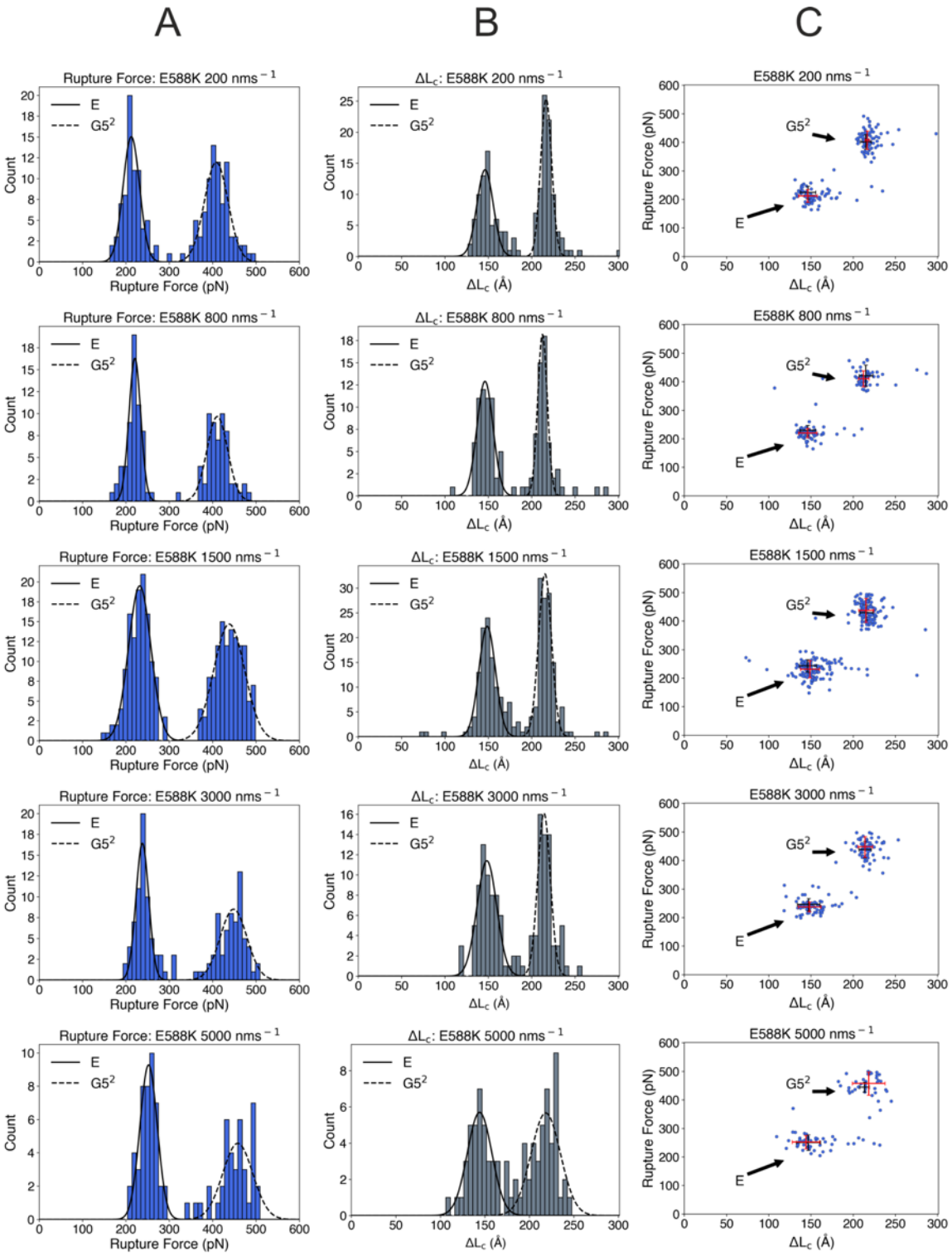

Figure 7.14 – Example E588K mechanical unfolding repeat at multiple retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Red scatterplot crosshairs are the mode and the FWHM from the corresponding histogram gaussian fits. Black crosshairs are the pWT $(E\text{-}G5^2)_5$ equivalent.

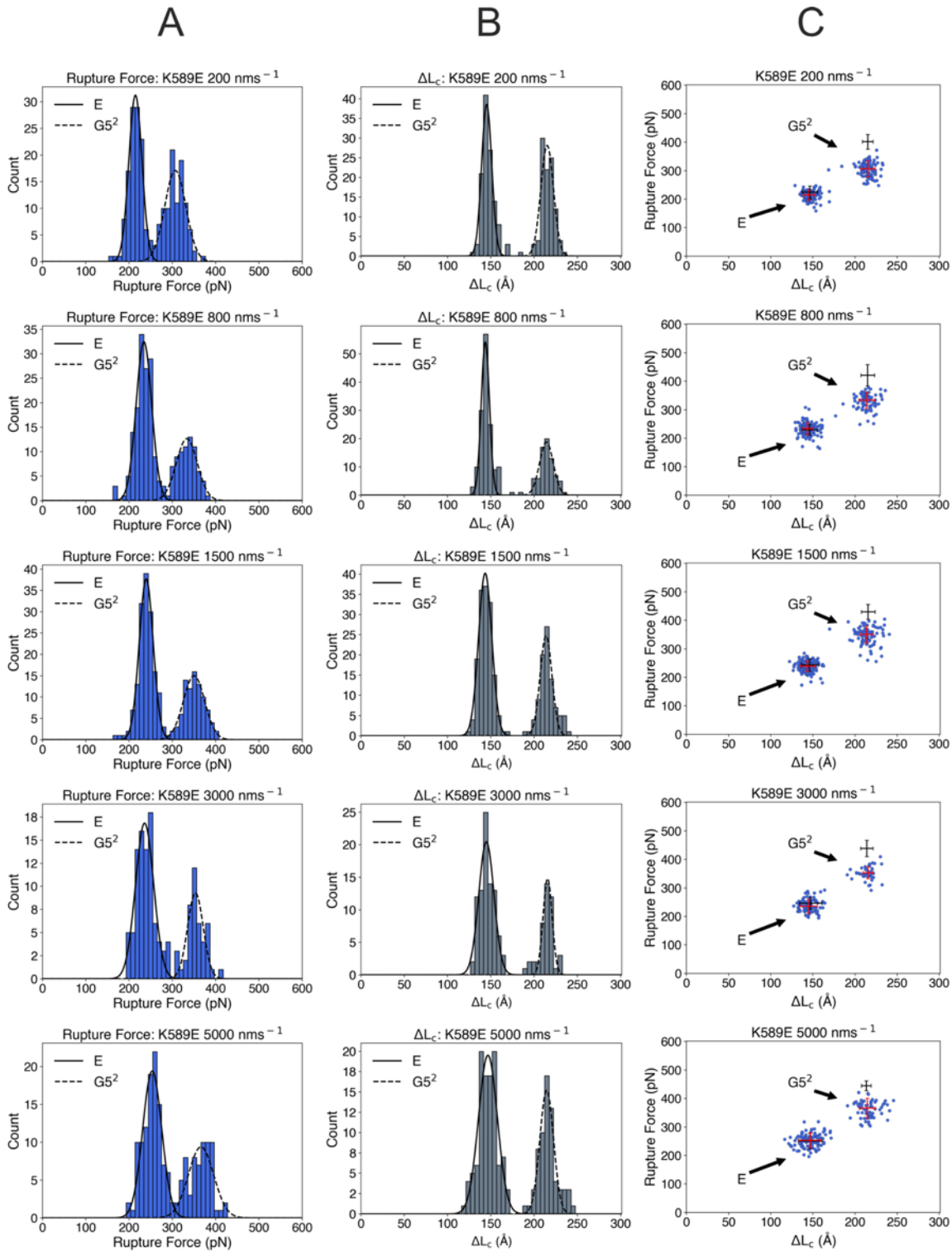

Figure 7.15 – Example K589E mechanical unfolding repeat at multiple retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Red scatterplot crosshairs are the mode and the FWHM from the corresponding histogram gaussian fits. Black crosshairs are the pWT $(E\text{-}G5^2)_5$ equivalent.

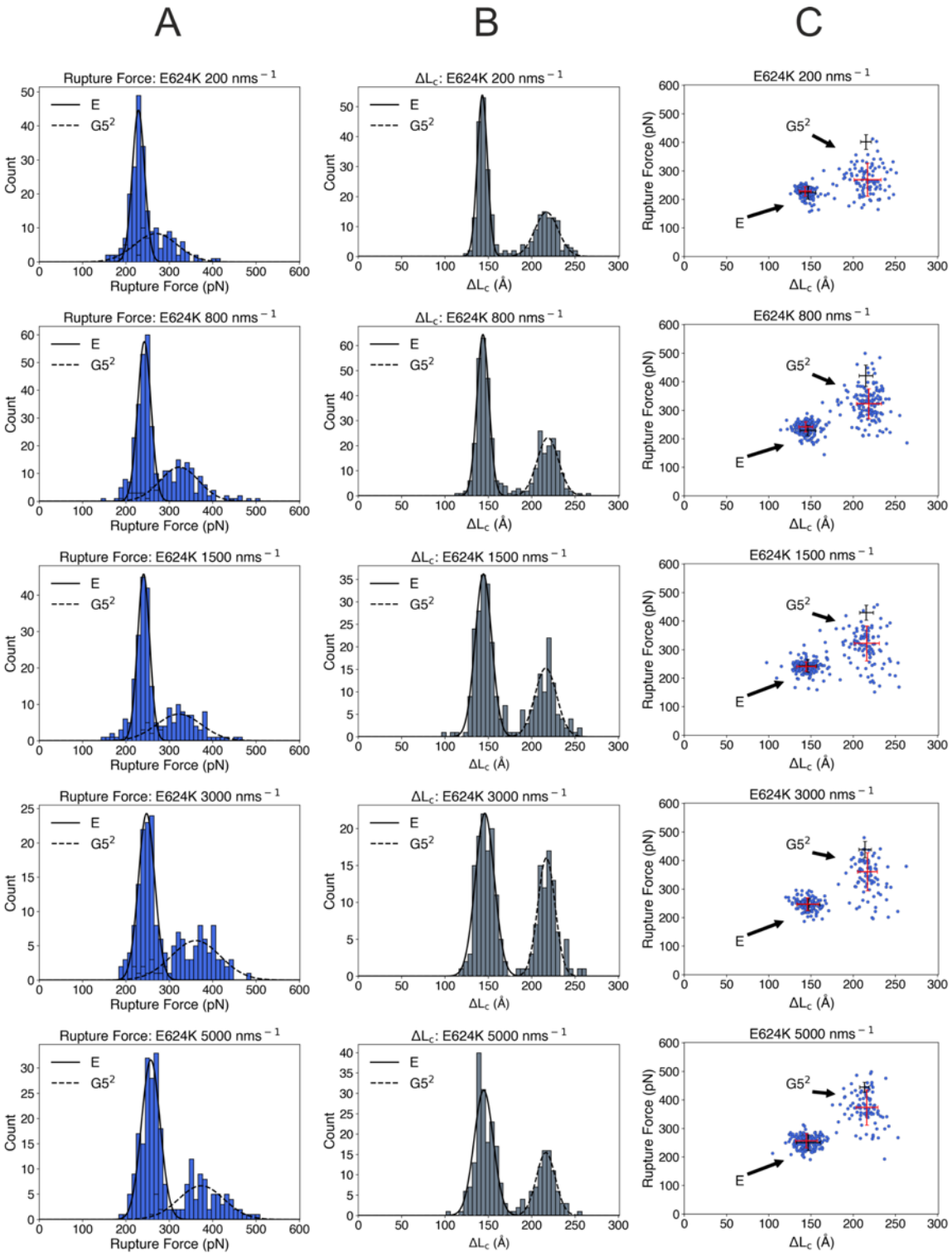

Figure 7.16 – Example E624K mechanical unfolding repeat at multiple retraction velocities. Columns A) and B) display the rupture force- and ΔL_C-frequency histograms, respectively, and column C) displays the scatterplots. From the top to bottom row, the retraction velocity increases from 200-5000 nms^{-1}. Two populations with both distinct rupture force and ΔL_C values are observable at all retraction velocities. Red scatterplot crosshairs are the mode and the FWHM from the corresponding histogram gaussian fits. Black crosshairs are the pWT $(E\text{-}G5^2)_5$ equivalent.

8. Bibliography

1. Lane, N. The unseen world: reflections on Leeuwenhoek (1677) “Concerning little animals.” *Philos Trans R Soc Lond B Biol Sci* vol. 370 (2015).
2. Leeuwenhoek, A. van. Observations, Communicated to the Publisher by Mr. Antony van Leewenhoeck in a Dutch Letter of the 9th Oct. 1676. Here English'd: concerning Little Animals by Him Observed in Rain-Well-Sea. and Snow Water; as Also in Water Wherein Pepper Had Lain Infused. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **12**, 821–831 (1677).
3. Leeuwenhoeck, A. An abstract of a letter from Mr. Anthony Leevvenhoeck at Delft, dated Sep. 17. 1683. Containing some microscopical observations, about animals in the scurf of the teeth, the substance call'd worms in the nose, the cuticula consisting of scales. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **14**, 568–574 (1684).
4. Venkova, T., Yeo, C. C. & Espinosa, M. Editorial: The Good, The Bad, and The Ugly: Multiple Roles of Bacteria in Human Life. *Front Microbiol.* **9**, (2018).
5. Heukelekian, H. & Heller, A. Relation between Food Concentration and Surface for Bacterial Growth. *J Bacteriol* **40**, 547–558 (1940).
6. Zobell, C. E. The Effect of Solid Surfaces upon Bacterial Activity. *J Bacteriol* **46**, 39–56 (1943).
7. Jones, H. C., Roth, I. L. & Sanders, W. M. Electron microscopic study of a slime layer. *J Bacteriol* **99**, 316–325 (1969).
8. Elias, S. & Banin, E. Multi-species biofilms: living with friendly neighbours. *FEMS Microbiol Rev* **36**, 990–1004 (2012).
9. Donlan, R. M. & Costerton, J. W. Biofilms: survival mechanisms of clinically relevant microorganisms. *Clin. Microbiol. Rev.* **15**, 167–193 (2002).
10. Flemming, H. C. *et al.* Biofilms: An emergent form of bacterial life. *Nat Rev Microbiol* vol. 14 563–575 (2016).
11. Rabin, N. *et al.* Biofilm formation mechanisms and targets for developing antibiofilm agents. *Future Med Chem* vol. 7 493–512 (2015).
12. Stoodley, P., Sauer, K., Davies, D. G. & Costerton, J. W. Biofilms as complex differentiated communities. *Annu Rev Microbiol* **56**, 187–209 (2002).
13. Jefferson, K. K. What drives bacteria to produce a biofilm? *FEMS Microbiol Lett* vol. 236 163–173 (2004).
14. Rathinam, N. K., Salem, D. R. & Sani, R. K. Biofilm engineering for improving the performance of microbial electrochemical technologies. in *Biomass, Biofuels, Biochemicals: Microbial Electrochemical Technology: Sustainable Platform for Fuels, Chemicals and Remediation* 315–338 (Elsevier, 2018).
15. Palmer, J., Flint, S. & Brooks, J. Bacterial cell attachment, the beginning of a biofilm. *J Ind Microbiol Biotechnol* **34**, 577–588 (2007).
16. Fitzpatrick, F., Humphreys, H. & O'Gara, J. P. Evidence for icaADBC-independent biofilm development mechanism in methicillin-resistant Staphylococcus aureus clinical isolates. *J Clin Microbiol* **43**, 1973–1976 (2005).
17. Toledo-Arana, A. *et al.* Staphylococcus aureus develops an alternative, ica-independent biofilm in the absence of the arlRS two-component system. *J Bacteriol* **187**, 5318–5329 (2005).
18. Mack, D. Molecular mechanisms of Staphylococcus epidermidis biofilm formation. *J Hosp Infect* **43**, S113–S125 (1999).
19. Heilmann, C. *et al.* Molecular basis of intercellular adhesion in the biofilm-forming Staphylococcus epidermidis. *Mol Microbiol* **20**, 1083–1091 (1996).
20. Geoghegan, J. A. *et al.* Role of surface protein SasG in biofilm formation by Staphylococcus aureus. *J Bacteriol* **192**, 5663–5673 (2010).
21. Conrady, D. G. *et al.* A zinc-dependent adhesion module is responsible for intercellular adhesion in staphylococcal biofilms. *Proc Natl Acad Sci U S A* **105**, 19456–19461 (2008).
22. Karimi, A., Karig, D., Kumar, A. & Ardekani, A. M. Interplay of physical mechanisms and biofilm processes: Review of microfluidic methods. *Lab Chip* vol. 15 23–42 (2015).

23. Costerton, J. W., Geesey, G. G. & Cheng, K. J. How bacteria stick. *Sci Am* **238**, 86–95 (1978).
24. Singer, S. W. *et al.* Posttranslational modification and sequence variation of redox-active proteins correlate with biofilm life cycle in natural microbial communities. *ISME J* **4**, 1398–1409 (2010).
25. Costerton, J. W., Lewandowski, Z., Caldwell, D. E., Korber, D. R. & Lappin-Scott, H. M. Microbial Biofilms. *Annu. Rev. Microbiol.* **49**, 711–745 (1995).
26. Otto, M. Physical stress and bacterial colonisation. *FEMS Microbiol Rev* **38**, 1250–1270 (2014).
27. Dufrêne, Y. F. & Persat, A. Mechanomicrobiology: how bacteria sense and respond to forces. *Nat Rev Microbiol* vol. 18 227–240 (2020).
28. Persat, A. *et al.* The mechanical world of bacteria. *Cell* **161**, 988–997 (2015).
29. Carrion-Vazquez, M. *et al.* Mechanical and chemical unfolding of a single protein: A comparison. *Proc Natl Acad Sci U S A* **96**, 3694–3699 (1999).
30. Hoffmann, T. & Dougan, L. Single molecule force spectroscopy using polyproteins. *Chemical Society Reviews* vol. 41 4781–4696 (2012).
31. Hughes, M. L. & Dougan, L. The physics of pulling polyproteins: A review of single molecule force spectroscopy using the AFM to study protein unfolding. *Rep Prog Phys* **79**, (2016).
32. Lee, C. K., Wang, Y. M., Huang, L. S. & Lin, S. Atomic force microscopy: Determination of unbinding force, off rate and energy barrier for protein-ligand interaction. *Micron* **38**, 446–461 (2007).
33. Chen, Y., Radford, S. E. & Brockwell, D. J. Force-induced remodelling of proteins and their complexes. *Curr Opin Struct Biol* **30**, 89–99 (2015).
34. Zhang, X., Ma, L. & Zhang, Y. High-resolution optical tweezers for single-molecule manipulation. *Yale J Biol Med* **86**, 367–383 (2013).
35. Sitters, G. *et al.* Acoustic force spectroscopy. *Nat Methods* **12**, 47–50 (2015).
36. Sarkar, R. & Rybenkov, V. v. A guide to magnetic tweezers and their applications. *Front. Phys.* **4**, (2016).
37. Edwards, D. T., Faulk, J. K., LeBlanc, M.-A. & Perkins, T. T. Force Spectroscopy with 9-μs Resolution and Sub-pN Stability by Tailoring AFM Cantilever Geometry. *Biophys. J.* **113**, 2595–2600 (2017).
38. Binnig, G., Quate, C. F. & Gerber, C. Atomic force microscope. *Phys Rev Lett* **56**, 930–933 (1986).
39. Dufrêne, Y. F. Towards nanomicrobiology using atomic force microscopy. *Nat Rev Microbiol* **6**, 674–680 (2008).
40. Dufrêne, Y. F. Atomic force microscopy in microbiology: New structural and functional insights into the microbial cell surface. *mBio* **5**, e01363-14 (2014).
41. Crampton, N. & Brockwell, D. J. Unravelling the design principles for single protein mechanical strength. *Curr Opin Struct Biol* **20**, 508–517 (2010).
42. Lee, G. U., Kidwell, D. A. & Colton, R. J. Sensing Discrete Streptavidin–Biotin Interactions with Atomic Force Microscopy. *Langmuir* **10**, 354–357 (1994).
43. Dammer, U. *et al.* Binding strength between cell adhesion proteoglycans measured by atomic force microscopy. *Science (1979)* **267**, 1173–1175 (1995).
44. Anderson, K. L., Radford, S. E., Smith, D. A. & Brockwell, D. J. The Dynamical Response of Proteins Under Force. in *Handbook of Molecular Force Spectroscopy* 202–246 (2008).
45. Hinterdorfer, P. & Dufrêne, Y. F. Detection and localization of single molecular recognition events using atomic force microscopy. *Nat Methods* **3**, 347–355 (2006).
46. King, G. M., Carter, A. R., Churnside, A. B., Eberle, L. S. & Perkins, T. T. Ultrastable atomic force microscopy: Atomic-scale stability and registration in ambient conditions. *Nano Lett* **9**, 1451–1456 (2009).
47. Bull, M. S., Sullan, R. M. A., Li, H. & Perkins, T. T. Improved single molecule force spectroscopy using micromachined cantilevers. *ACS Nano* **8**, 4984–4995 (2014).
48. Edwards, D. T. & Perkins, T. T. Optimizing force spectroscopy by modifying commercial cantilevers: Improved stability, precision, and temporal resolution. *J Struct Biol* **197**, 13–25 (2017).

49. Valotteau, C., Sumbul, F. & Rico, F. High-speed force spectroscopy: microsecond force measurements using ultrashort cantilevers. *Biophys Rev* **11**, 689–699 (2019).
50. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science (1979)* **276**, 1109–1112 (1997).
51. Carrion-Vazquez, M. *et al.* Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Prog Biophys Mol Biol* **74**, 63–91 (2000).
52. Milstein, J. N. & Meiners, J.-C. Worm-Like Chain (WLC) Model. in *Encyclopedia of Biophysics* 2757–2760 (Springer Berlin Heidelberg, 2013).
53. Riener, C. K. *et al.* Simple test system for single molecule recognition force microscopy. *Anal Chim Acta* **479**, 59–75 (2003).
54. Lo, Y.-S. *et al.* Specific Interactions between Biotin and Avidin Studied by Atomic Force Microscopy Using the Poisson Statistical Analysis Method. *Langmuir* **15**, 1373–1382 (1999).
55. Hinterdorfer, P., Baumgartner, W., Gruber, H. J., Schilcher, K. & Schindler, H. Detection and localization of individual antibody-antigen recognition events by atomic force microscopy. *Proc Natl Acad Sci U S A* **93**, 3477–3481 (1996).
56. Willemsen, O. H. *et al.* Simultaneous Height and Adhesion Imaging of Antibody-Antigen Interactions by Atomic Force Microscopy. *Biophys J* **75**, 2220–2228 (1998).
57. Ros, R. *et al.* Antigen binding forces of individually addressed single-chain Fv antibody molecules. *Proc Natl Acad Sci U S A* **95**, 7402 (1998).
58. Fritz, J., Katopodis, A. G., Kolbinger, F. & Anselmetti, D. Force-mediated kinetics of single P-selectin/ligand complexes observed by atomic force microscopy. *Proc Natl Acad Sci U S A* **95**, 12283–12288 (1998).
59. Marantan, A. & Mahadevan, L. Mechanics and statistics of the worm-like chain. *Am J Phys* **86**, 86–94 (2018).
60. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic elasticity of λ-phage DNA. *Science (1979)* **265**, 1599–1600 (1994).
61. Marko, J. F. & Siggia, E. D. Stretching DNA. *Macromolecules* **28**, 8759–8770 (1995).
62. Ainavarapu, S. R. K. *et al.* Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophys J* **92**, 225–233 (2007).
63. 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).
64. Hickman, S. J., Cooper, R. E. M., Bellucci, L., Paci, E. & Brockwell, D. J. Gating of TonB-dependent transporters by substrate-specific forced remodelling. *Nat Commun* **8**, 1–12 (2017).
65. Marszalek, P. E. *et al.* Mechanical unfolding intermediates in titin modules. *Nature* **402**, 100–103 (1999).
66. Bell, G. I. Models for the specific adhesion of cells to cells. *Science (1979)* **200**, 618–627 (1978).
67. Evans, E. & Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophys J* **72**, 1541–1555 (1997).
68. Yew, Z. T., Olmsted, P. D. & Paci, E. Free Energy Landscapes of Proteins: Insights from Mechanical Probes. in *Single Molecule Biophysics: Experiment and Theory* vol. 146 395–417 (Wiley Blackwell, 2011).
69. Zinober, R. C. *et al.* Mechanically unfolding proteins: The effect of unfolding history and the supramolecular scaffold. *Protein Sci* **11**, 2759–2765 (2002).
70. Brockwell, D. J. *et al.* Mechanically unfolding the small, topologically simple protein L. *Biophys J* **89**, 506–519 (2005).
71. Hoffmann, T., Tych, K. M., Brockwell, D. J. & Dougan, L. Single-molecule force spectroscopy identifies a small cold shock protein as being mechanically robust. *J Phys Chem B* **117**, 1819–1826 (2013).
72. Gabovich, A. M. & Li, M. S. Mechanical stability of proteins. *J Chem Phys* **131**, 024121 (2009).
73. Oberhauser, A. F. & Carrión-Vázquez, M. Mechanical biochemistry of proteins one molecule at a time. *J Biol Chem* **283**, 6617–6621 (2008).
74. Gruszka, D. T. *et al.* Cooperative folding of intrinsically disordered domains drives assembly of a strong elongated protein. *Nat Commun* **6**, (2015).

75. Chantraine, C., Mathelié-Guinlet, M., Pietrocola, G., Speziale, P. & Dufrêne, Y. F. AFM Identifies a Protein Complex Involved in Pathogen Adhesion Which Ruptures at Three Nanonewtons. *Nano Lett* **21**, 7595–7601 (2021).
76. Milles, L. F., Unterauer, E. M., Nicolaus, T. & Gaub, H. E. Calcium stabilizes the strongest protein fold. *Nat Commun* **9**, 4764 (2018).
77. Valbuena, A. *et al.* On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proc Natl Acad Sci U S A* **106**, 13791–13796 (2009).
78. Li, H. & Fernandez, J. M. Mechanical Design of the First Proximal Ig Domain of Human Cardiac Titin Revealed by Single Molecule Force Spectroscopy. *J Mol Biol* **334**, 75–86 (2003).
79. Oberhauser, A. F., Badilla-Fernandez, C., Carrion-Vazquez, M. & Fernandez, J. M. The mechanical hierarchies of fibronectin observed with single-molecule AFM. *J Mol Biol* **319**, 433–447 (2002).
80. Best, R. B., Li, B., Steward, A., Daggett, V. & Clarke, J. Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation. *Biophys J* **81**, 2344–2356 (2001).
81. Cao, Y. & Li, H. Polyprotein of GB1 is an ideal artificial elastomeric protein. *Nature Materials* **6**, 109–114 (2007).
82. Carrion-Vazquez, M. *et al.* The mechanical stability of ubiquitin is linkage dependent. *Nat Struct Biol* **10**, 738–743 (2003).
83. Lenne, P.-F., Raae, A. J., Altmann, S. M., Saraste, M. & Hörber, J. K. H. States and transitions during forced unfolding of a single spectrin repeat. *FEBS Lett* **476**, 124–128 (2000).
84. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* **247**, 536–540 (1995).
85. Lu, H. & Schulten, K. Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins* **35**, 453–463 (1999).
86. Klimov, D. K. & Thirumalai, D. Native topology determines force-induced unfolding pathways in globular proteins. *Proc Natl Acad Sci U S A* **97**, 7254 (2000).
87. Brockwell, D. J. *et al.* Pulling geometry defines the mechanical resistance of a β-sheet protein. *Nat Struct Biol* **10**, 731–737 (2003).
88. Dietz, H., Berkemeier, F., Bertz, M. & Rief, M. Anisotropic deformation response of single protein molecules. *Proc Natl Acad Sci U S A* **103**, 12724–12728 (2006).
89. Lu, H., Isralewitz, B., Krammer, A., Vogel, V. & Schulten, K. Unfolding of Titin Immunoglobulin Domains by Steered Molecular Dynamics Simulation. *Biophys J* **75**, 662–671 (1998).
90. Li, H., Carrion-Vazquez, M., Oberhauser, A. F., Marszalek, P. E. & Fernandez, J. M. Point mutations alter the mechanical stability of immunoglobulin modules. *Nat Struct Biol* **7**, 1117–1120 (2000).
91. Sikora, M., Sułkowska, J. I. & Cieplak, M. Mechanical strength of 17,134 model proteins and cysteine slipknots. *PLoS Comput Biol* **5**, (2009).
92. Sikora, M., Sulkowska, J. I., Witkowski, B. S. & Cieplak, M. BSDB: the biomolecule stretching database. *Nucleic Acids Res* **39**, D443–D450 (2011).
93. Plaxco, K. W., Simons, K. T. & Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* **277**, 985–994 (1998).
94. Li, M. S. Secondary structure, mechanical stability, and location of transition state of proteins. *Biophys J* **93**, 2644–2654 (2007).
95. Hoffmann, T., Tych, K. M., Hughes, M. L., Brockwell, D. J. & Dougan, L. Towards design principles for determining the mechanical stability of proteins. *Phys Chem Chem Phys* **15**, 15767–15780 (2013).
96. Bu, T., Wang, H. C. E. & Li, H. Single molecule force spectroscopy reveals critical roles of hydrophobic core packing in determining the mechanical stability of protein GB1. *Langmuir* **28**, 12319–12325 (2012).
97. Sadler, D. P. *et al.* Identification of a Mechanical Rheostat in the Hydrophobic Core of Protein L. *J Mol Biol* **393**, 237–248 (2009).
98. Tych, K. M. *et al.* Tuning protein mechanics through an ionic cluster graft from an extremophilic protein. *Soft Matter* **12**, 2688–2699 (2016).

99. Forman, J. R. *et al.* Non-native interactions are critical for mechanical strength in PKD domains. *Structure* **17**, 1582–1590 (2009).
100. Wang, B., Xiao, S., Edwards, S. A. & Gräter, F. Isopeptide bonds mechanically stabilize spy0128 in bacterial pili. *Biophys J* **104**, 2051–2057 (2013).
101. Best, R. B. *et al.* Mechanical unfolding of a titin Ig domain: structure of transition state revealed by combining atomic force microscopy, protein engineering and molecular dynamics simulations. *J Mol Biol* **330**, 867–877 (2003).
102. Ng, S. P. *et al.* Mechanical unfolding of TNfn3: The unfolding pathway of a fnIII domain probed by protein engineering, AFM and MD simulation. *J Mol Biol* **350**, 776–789 (2005).
103. Ng, S. P., Billings, K. S., Randles, L. G. & Clarke, J. Manipulating the stability of fibronectin type III domains by protein engineering. *Nanotechnology* **19**, 384023 (2008).
104. Oude Vrielink, A. S., Vance, T. D. R., de Jong, A. M., Davies, P. L. & Voets, I. K. Unusually high mechanical stability of bacterial adhesin extender domains having calcium clamps. *PLOS ONE* **12**, e0174682- (2017).
105. Giuseppe, Z. How Soft Is a Protein? A Protein Dynamics Force Constant Measured by Neutron Scattering. *Science (1979)* **288**, 1604–1607 (2000).
106. Tych, K. M., Hoffmann, T., Brockwell, D. J. & Dougan, L. Single molecule force spectroscopy reveals the temperature-dependent robustness and malleability of a hyperthermophilic protein. *Soft Matter* **9**, 9016–9025 (2013).
107. Manetti, A. G. O. *et al.* Streptococcus pyogenes pili promote pharyngeal cell adhesion and biofilm formation. *Mol Microbiol* **64**, 968–983 (2007).
108. Alegre-Cebollada, J., Badilla, C. L. & Fernández, J. M. Isopeptide bonds block the mechanical extension of pili in pathogenic Streptococcus pyogenes. *Journal of Biological Chemistry* **285**, 11235–11242 (2010).
109. Kang, H. J., Paterson, N. G., Gaspar, A. H., Ton-That, H. & Baker, E. N. The Corynebacterium diphtheriae shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds. *Proc Natl Acad Sci U S A* **106**, 16967–16971 (2009).
110. Joo, K. H., Fasséli, C., Fiona, C., Thomas, P. & N, B. E. Stabilizing Isopeptide Bonds Revealed in Gram-Positive Bacterial Pilus Structure. *Science (1979)* **318**, 1625–1628 (2007).
111. Budzik, J. M. *et al.* Intramolecular amide bonds stabilize pili on the surface of bacilli. *PNAS* **106**, 19992–19997 (2009).
112. Fersht, A. R., Matouschek, A. & Serrano, L. The folding of an enzyme: I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* **224**, 771–782 (1992).
113. Best, R. B., Fowler, S. B., Toca-Herrera, J. L. & Clarke, J. A simple method for probing the mechanical unfolding pathway of proteins in detail. *Proc Natl Acad Sci U S A* **99**, 12143–12148 (2002).
114. Foster, T. J. Surface Proteins of Staphylococcus aureus. *Microbiol Spectr* **7**, (2019).
115. Foster, T. J. Surface Proteins of Staphylococcus epidermidis. *Front Microbiol* **11**, (2020).
116. Michael, O. Staphylococcal Biofilms. *Microbiol Spectr* **6**, (2018).
117. Foster, T. J., Geoghegan, J. A., Ganesh, V. K. & Höök, M. Adhesion, invasion and evasion: The many functions of the surface proteins of Staphylococcus aureus. *Nat Rev Microbiol* **12**, 49–62 (2014).
118. Ponnuraj, K. *et al.* A “dock, lock, and latch” structural model for a staphylococcal adhesin binding to fibrinogen. *Cell* **115**, 217–228 (2003).
119. Herman, P. *et al.* The binding force of the staphylococcal adhesin SdrG is remarkably strong. *Molecular Microbiology* **93**, 356–368 (2014).
120. Milles, L. F., Schulten, K., Gaub, H. E. & Bernardi, R. C. Molecular mechanism of extreme mechanostability in a pathogen adhesin. *Science (1979)* **359**, 1527–1533 (2018).
121. Zong, Y. *et al.* A “Collagen Hug” Model for Staphylococcus aureus CNA binding to collagen. *EMBO J* **24**, 4224–4236 (2005).

122. Herman-Bausier, P. *et al.* Mechanical Strength and Inhibition of the Staphylococcus aureus Collagen-Binding Protein Cna. *mBio* **7**, e01529 (2016).
123. Christersson, C. E., Glantz, P. O. & Baier, R. E. Role of temperature and shear forces on microbial detachment. *Scand J Dent Res* **96**, 91–98 (1988).
124. Dembo, M., Torney, D. C., Saxman, K. & Hammer, D. The reaction-limited kinetics of membrane-to-surface adhesion and detachment. *Proc R Soc Lond B Biol Sci* **234**, 55–83 (1988).
125. Sokurenko, E. v, Vogel, V. & Thomas, W. E. Catch-bond mechanism of force-enhanced adhesion: counterintuitive, elusive, but ... widespread? *Cell Host Microbe* **4**, 314–323 (2008).
126. Thomas, W. E., Trintchina, E., Forero, M., Vogel, V. & Sokurenko, E. v. Bacterial adhesion to target cells enhanced by shear force. *Cell* **109**, 913–923 (2002).
127. Thomas, W. E., Nilsson, L. M., Forero, M., Sokurenko, E. v & Vogel, V. Shear-dependent "stick-and-roll" adhesion of type 1 fimbriated Escherichia coli. *Mol Microbiol* **53**, 1545–1557 (2004).
128. Marshall, B. T. *et al.* Direct observation of catch bonds involving cell-adhesion molecules. *Nature* **423**, 190–193 (2003).
129. Herman-Bausier, P. *et al.* Staphylococcus aureus clumping factor A is a force-sensitive molecular switch that activates bacterial adhesion. *Proc Natl Acad Sci U S A* **115**, 5564–5569 (2018).
130. Grandbois, M., Beyer, M., Rief, M., Clausen-Schaumann, H. & Gaub, H. E. How strong is a covalent bond. *Science (1979)* **283**, 1727–1730 (1999).
131. Yakovenko, O. *et al.* FimH forms catch bonds that are enhanced by mechanical force due to allosteric regulation. *J Biol Chem* **283**, 11596–11605 (2008).
132. Mulcahy, M. E. *et al.* Nasal Colonisation by Staphylococcus aureus Depends upon Clumping Factor B Binding to the Squamous Epithelial Cell Envelope Protein Loricrin. *PLoS Pathog* **8**, (2012).
133. Vitry, P. *et al.* Force-induced strengthening of the interaction between Staphylococcus aureus clumping factor B and loricrin. *mBio* **8**, (2017).
134. Alsteens, D. *et al.* Imaging G protein–coupled receptors while quantifying their ligand-binding free-energy landscape. *Nat Methods* **12**, 845–851 (2015).
135. Mathelié-Guinlet, M. *et al.* Force-clamp spectroscopy identifies a catch bond mechanism in a Gram-positive pathogen. *Nat Commun* **11**, 5431 (2020).
136. Mathelié-Guinlet, M., Viela, F., Pietrocola, G., Speziale, P. & Dufrêne, Y. F. Nanonewton forces between Staphylococcus aureus surface protein IsdB and vitronectin. *Nanoscale Adv* **2**, 5728–5736 (2020).
137. Viela, F. *et al.* Binding of staphylococcus aureus protein a to von willebrand factor is regulated by mechanical force. *mBio* **10**, (2019).
138. Viljoen, A. *et al.* Staphylococcus aureus vWF-binding protein triggers a strong interaction between clumping factor A and host vWF. *Commun Biol* **4**, 453 (2021).
139. Deivanayagam, C. C. S. *et al.* Novel fold and assembly of the repetitive B region of the Staphylococcus aureus collagen-binding surface protein. *Structure* **8**, 67–78 (2000).
140. Jemima Beulin, D. S. & Ponnuraj, K. Steered molecular dynamics study reveals insights into the function of the repetitive B region of collagen- and fibrinogen-binding MSCRAMMs. *J Biomol Struct Dyn* **35**, 535–550 (2017).
141. Deivanayagam, C. C. S. *et al.* Novel fold and assembly of the repetitive B region of the Staphylococcus aureus collagen-binding surface protein. *Structure* **8**, 67–78 (2000).
142. Whelan, F. *et al.* Periscope Proteins are variable-length regulators of bacterial cell surface interactions. *Proc Natl Acad Sci U S A* **118**, (2021).
143. P, C. B. *et al.* Role for the A Domain of Unprocessed Accumulation-Associated Protein (Aap) in the Attachment Phase of the Staphylococcus epidermidis Biofilm Phenotype. *J Bacteriol* **196**, 4268–4275 (2014).
144. Corrigan, R. M., Rigby, D., Handley, P. & Foster, T. J. The role of Staphylococcus aureus surface protein SasG in adherence and biofilm formation. *Microbiology (N Y)* **153**, 2435–2446 (2007).
145. Roche, F. M., Meehan, M. & Foster, T. J. The Staphylococcus aureus surface protein SasG and its homologues promote bacterial adherence to human desquamated nasal epithelial cells. *Microbiology (N Y)* **149**, 2759–2767 (2003).

146. Rohde, H. *et al.* Induction of Staphylococcus epidermidis biofilm formation via proteolytic processing of the accumulation-associated protein by staphylococcal and host proteases. *Mol Microbiol* **55**, 1883–1895 (2005).
147. Yarawsky, A. E., Johns, S. L., Schuck, P. & Herr, A. B. The biofilm adhesion protein Aap from Staphylococcus epidermidis forms zinc-dependent amyloid fibers. *J Biol Chem* **295**, 4411–4427 (2020).
148. Held, W. & Mariuzza, R. A. Cis–trans interactions of cell surface receptors: biological roles and structural basis. *Cell. Mol. Life Sci.* **68**, 3469 (2011).
149. Gruszka, D. T. *et al.* Staphylococcal biofilm-forming protein has a contiguous rod-like structure. *Proc Natl Acad Sci U S A* **109**, (2012).
150. Rohde, H. *et al.* Polysaccharide intercellular adhesin or protein factors in biofilm accumulation of Staphylococcus epidermidis and Staphylococcus aureus isolated from prosthetic hip and knee joint infections. *Biomaterials* **28**, 1711–1720 (2007).
151. Roche, F. M. *et al.* Characterization of novel LPXTG-containing proteins of Staphylococcus aureus identified from genome sequences. *Microbiology (N Y)* **149**, (2003).
152. Gruszka, D. T. *et al.* Disorder drives cooperative folding in a multidomain protein. *Proc Natl Acad Sci U S A* **113**, 11841–11846 (2016).
153. Formosa-Dague, C., Speziale, P., Foster, T. J., Geoghegan, J. A. & Dufrêne, Y. F. Zinc-dependent mechanical properties of Staphylococcus aureus biofilm-forming surface protein SasG. *Proc Natl Acad Sci U S A* **113**, 410–415 (2016).
154. Abraham, N. M., Lamlertthon, S., Fowler, V. G. & Jefferson, K. K. Chelating agents exert distinct effects on biofilm formation in Staphylococcus aureus depending on strain background: role for clumping factor B. *J Med Microbiol* **61**, 1062–1070 (2012).
155. Banin, E., Brady, K. M. & Greenberg, P. E. Chelator-Induced Dispersal and Killing of Pseudomonas aeruginosa Cells in a Biofilm. *Appl Environ Microbiol* **72**, 2064–2069 (2006).
156. Rivas-Pardo, J. A., Carmen, B. L., Tapia-Rojo, R., Alonso-Caballero, Á. & Fernández, J. M. Molecular strategy for blocking isopeptide bond formation in nascent pilin proteins. *PNAS* **115**, 9222–9227 (2018).
157. Li, H. & Cao, Y. Protein Mechanics: From Single Molecules to Functional Biomaterials. *Acc Chem Res* **43**, 1331–1341 (2010).
158. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**, 207–234 (2005).
159. Consortium, T. U. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**, D480–D489 (2021).
160. Ferrara, P., Apostolakis, J. & Caflisch, A. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics* **46**, 24–33 (2002).
161. Neb England Biolabs. Gibson Assembly®. (2022).
162. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343–345 (2009).
163. Hoffmann, T. *et al.* Rapid and Robust Polyprotein Production Facilitates Single-Molecule Mechanical Characterization of β-Barrel Assembly Machinery Polypeptide Transport Associated Domains. *ACS Nano* **9**, 8811–8821 (2015).
164. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, e3647–e3647 (2008).
165. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLOS ONE* **4**, (2009).
166. New England Biolabs. NEB® Golden Gate Assembly Kit (BsaI-HF®v2). https://international.neb.com/products/e1601-neb-golden-gate-assembly-mix#Product%20Information (2022).
167. Hédoux, A. *et al.* Influence of urea and guanidine hydrochloride on lysozyme stability and thermal denaturation; a correlation between activity, protein dynamics and conformational changes. *Phys Chem Chem Phys* **12**, 13189–13196 (2010).
168. Asylum Research an Oxford Instruments Company. MFP-3D User Guide. vol. 13 8 (2013).

169. Hutter, J. L. & Bechhoefer, J. Calibration of atomic-force microscope tips. *Rev Sci Instrum* **64**, (1993).
170. Doherty, C. P. A. Pulling apart the intermolecular interactions of the Parkinson's disease linked protein alpha synuclein. (2017).
171. Doherty, C. P. A. *et al.* A peptide-display protein scaffold to facilitate single molecule force studies of aggregation-prone peptides. *Protein Sci* **27**, 1205–1217 (2018).
172. Galvanetto, N., Perissinotto, A., Pedroni, A. & Torre, V. Fodis: Software for Protein Unfolding Analysis. *Biophys J* **114**, 1264–1266 (2018).
173. Schiffrin, R. Roles of periplasmic chaperones and BamA in outer membrane protein folding. (University of Leeds, 2016).
174. Graceffa, P., Jancsó, A. & Mabuchi, K. Modification of acidic residues normalizes sodium dodecyl sulfate-polyacrylamide gel electrophoresis of caldesmon and other proteins that migrate anomalously. *Arch Biochem Biophys* **297**, 46–51 (1992).
175. Alves, V. S., Pimenta, D. C., Sattlegger, E. & Castilho, B. A. Biophysical characterization of Gir2, a highly acidic protein of Saccharomyces cerevisiae with anomalous electrophoretic behaviour. *Biochem Biophys Res Commun* **314**, 229–234 (2004).
176. Lopes, J. L. S., Miles, A. J., Whitmore, L. & Wallace, B. A. Distinct circular dichroism spectroscopic signatures of polyproline II and unordered secondary structures: applications in secondary structure analyses. *Protein Sci* **23**, 1765–1772 (2014).
177. Drzewiecki, K. E., Grisham, D. R., Parmar, A. S., Nanda, V. & Shreiber, D. I. Circular Dichroism Spectroscopy of Collagen Fibrillogenesis: A New Use for an Old Technique. *Biophys J* **111**, 2377–2386 (2016).
178. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat Prot* **1**, 2876–2890 (2006).
179. Zhao, B. *et al.* A Thioether-Stabilized d-Proline–l-Proline-Induced β-Hairpin Peptide of Defensin Segment Increases Its Anti-Candida albicans Ability. *ChemBioChem* **17**, 1416–1420 (2016).
180. Jasaitis, L. *et al.* Rational Design and Self-Assembly of Coiled-Coil Linked SasG Protein Fibrils. *ACS Synth Biol* **9**, 1599–1607 (2020).
181. Wellinghausen, N. & Rink, L. The significance of zinc for leukocyte biology. *J Leukoc Biol* **64**, 571–577 (1998).
182. Alberts, I. L., Nadassy, K. & Wodak, S. J. Analysis of zinc binding sites in protein crystal structures. *Protein Sci* **7**, 1700–1716 (1998).
183. Johnson, K. C. & Thomas, W. E. How Do We Know when Single-Molecule Force Spectroscopy Really Tests Single Bonds? *Biophys J* **114**, 2032–2039 (2018).
184. Fischer, B. E., Häring, U. K., Tribolet, R. & Sigel, H. Metal ion/buffer interactions. Stability of binary and ternary complexes containing 2-amino-2(hydroxymethyl)-1,3-propanediol (Tris) and adenosine 5'-triphosphate (ATP). *Eur J Biochem* **94**, 523–530 (1979).
185. Farrance, O. E., Paci, E., Radford, S. E. & Brockwell, D. J. Extraction of accurate biomolecular parameters from single-molecule force spectroscopy experiments. *ACS Nano* **9**, 1315–1324 (2015).
186. Schrödinger, L. PyMOL. (2021).
187. Tapiero, H. & Tew, K. D. Trace elements in human physiology and pathology: zinc and metallothioneins. *Biomed Pharmacother* **57**, 399–411 (2003).
188. Ma, L., Terwilliger, A. & Maresso, A. W. Iron and zinc exploitation during bacterial pathogenesis. *Metallomics* **7**, 1541–1554 (2015).
189. Vallee, B. L. & Falchuk, K. H. The biochemical basis of zinc physiology. *Physiol Rev* **73**, 79–118 (1993).
190. Missiakas, D. & Winstel, V. Selective Host Cell Death by Staphylococcus aureus: A Strategy for Bacterial Persistence. *Front Immunol* **11**, (2021).
191. Conti, M., Falini, G. & Samorì, B. How Strong Is the Coordination Bond between a Histidine Tag and Ni–Nitrilotriacetate? An Experiment of Mechanochemistry on Single Molecules. *Angew Chem Int Ed Engl* **39**, 215–218 (2000).
192. Chaton, C. T. & Herr, A. B. Defining the metal specificity of a multifunctional biofilm adhesion protein. *Protein Sci* **26**, 1964–1973 (2017).

193. Harding, M. M. Geometry of metal-ligand interactions in proteins. *Acta Cryst.* **D57**, 401–411 (2001).
194. Richardson, J. S. & Richardson, D. C. Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc Natl Acad Sci U S A* **99**, 2754–2759 (2002).
195. Garcia-Manyes, S., Brujic, J., Badilla, C. L. & Fernández, J. M. Force-Clamp Spectroscopy of Single-Protein Monomers Reveals the Individual Unfolding and Folding Pathways of I27 and Ubiquitin. *Biophys J* **93**, 2436–2446 (2007).
196. Rounsevell, R. W. S., Steward, A. & Clarke, J. Biophysical investigations of engineered polyproteins: implications for force data. *Biophys J* **88**, 2022–2029 (2005).
197. Potapov, V. *et al.* Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly. *ACS Synthetic Biology* **7**, 2665–2674 (2018).
198. Hirel, P. H., Schmitter, J. M., Dessen, P., Fayat, G. & Blanquet, S. Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. *Proc Natl Acad Sci U S A* **86**, 8247–8251 (1989).
199. Joshi, T. *et al.* Interdomain linkers tailor the stability of immunoglobulin repeats in polyproteins. *Biochem Biophys Res Commun* **550**, 43–48 (2021).
200. Batey, S., Nickson, A. A. & Clarke, J. Studying the folding of multidomain proteins. *HFSP J* **2**, 365–377 (2008).
201. Aksel, T., Majumdar, A. & Barrick, D. The contribution of entropy, enthalpy, and hydrophobic desolvation to cooperativity in repeat-protein folding. *Structure* **19**, 349–360 (2011).
202. Kramer, R. Z., Bella, J., Mayville, P., Brodsky, B. & Berman, H. M. Sequence dependent conformational variations of collagen triple-helical structure. *Nat Struct Biol* **6**, 454–457 (1999).
203. Ruggiero, A. *et al.* Crystal Structure of the Resuscitation-Promoting Factor ΔDUFRpfB from M. tuberculosis. *J Mol Biol* **385**, 153–162 (2009).
204. Haberthür, U. & Caflisch, A. FACTS: Fast analytical continuum treatment of solvation. *J Comput Chem* **29**, 701–715 (2008).
205. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **29**, 1859–1865 (2008).
206. Friel, C. T., Beddard, G. S. & Radford, S. E. Switching Two-state to Three-state Kinetics in the Helical Protein Im9 via the Optimisation of Stabilising Non-native Interactions by Design. *J Mol Biol* **342**, 261–273 (2004).
207. Myers, J. K., Pace, C. N. & Scholtz, J. M. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* **4**, 2138–2148 (1995).
208. Shortle, D. & Meeker, A. K. Residual structure in large fragments of staphylococcal nuclease: effects of amino acid substitutions. *Biochemistry* **28**, 936–944 (1989).
209. Wrabl, J. & Shortle, D. A model of the changes in denatured state structure underlying m value effects in staphylococcal nuclease. *Nat Struct Biol* **6**, 876–883 (1999).
210. Spudich, G. & Marqusee, S. A Change in the Apparent m Value Reveals a Populated Intermediate under Equilibrium Conditions in Escherichia coli Ribonuclease HI. *Biochemistry* **39**, 11677–11683 (2000).
211. Capaldi, A. P., Ferguson, S. J. & Radford, S. E. The Greek key protein apo-pseudoazurin folds through an obligate on-pathway intermediate. *J Mol Biol* **286**, 1621–1632 (1999).
212. Bella, J., Eaton, M., Brodsky, B. & Berman, H. M. Crystal and molecular structure, of a collagen-like peptide at 1.9 Å resolution. *Science (1979)* **266**, 75–81 (1994).
213. Punitha, V. *et al.* Molecular dynamics investigations on the effect of D amino acid substitution in a triple-helix structure and the stability of collagen. *J Phys Chem B* **113**, 8983–8992 (2009).
214. Evans, E. Probing the relation between force--lifetime--and chemistry in single molecular bonds. *Annu Rev Biophys Biomol Struct* **30**, 105–128 (2001).
215. Schlierf, M. & Rief, M. Temperature Softening of a Protein in Single-molecule Experiments. *J Mol Biol* **354**, 497–503 (2005).
216. Sánchez, I. E. & Kiefhaber, T. Origin of Unusual ϕ-values in Protein Folding: Evidence Against Specific Nucleation Sites. *J Mol Biol* **334**, 1077–1085 (2003).

217. Zarrine-Afsar, A. *et al.* Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. *Proc Natl Acad Sci U S A* **105**, 9999 LP – 10004 (2008).
218. Weikl, T. R. Transition states in protein folding kinetics: modeling phi-values of small beta-sheet proteins. *Biophys J* **94**, 929–937 (2008).
219. Northey, J. G. B., di Nardo, A. A. & Davidson, A. R. Hydrophobic core packing in the SH3 domain folding transition state. *Nat Struct Mol Biol* **9**, 126–130 (2002).
220. Shoulders, M. D. & Raines, R. T. Collagen Structure and Stability. *Annu Rev Biochem* **78**, 929–958 (2009).
221. Persikov, A. v., Ramshaw, J. A. M., Kirkpatrick, A. & Brodsky, B. Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability. *Biochemistry* **44**, 1414–1422 (2005).
222. Xiao, J., Cheng, H., Silva, T., Baum, J. & Brodsky, B. Osteogenesis imperfecta missense mutations in collagen: Structural consequences of a glycine to alanine replacement at a highly charged site. *Biochemistry* **50**, 10771–10780 (2011).
223. Hyde, T. J., Bryan, M. A., Brodsky, B. & Baum, J. Sequence dependence of renucleation after a Gly mutation in model collagen peptides. *J Biol Chem* **281**, 36937–36943 (2006).